wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 06:40 PM GMT

PDB ID	:	4M7P
Title	:	Ensemble refinement of protein crystal structure of macrolide glycosyltransferases OleD
Authors	:	Wang, F.; Helmich, K.E.; Xu, W.; Singh, S.; Olmos Jr., J.L.; Martinez iii, E.; Bingman, C.A.; Thorson, J.S.; Phillips Jr., G.N.; Enzyme Discovery for Natural Product Biosynthesis (NatPro)
Deposited on	:	2013-08-12
Resolution	:	1.77 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (i)) were used in the production of this report:

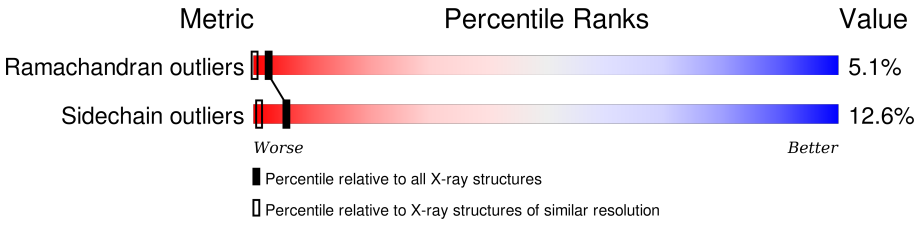
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	FAILED
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.77 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Ramachandran outliers	100387	7570 (1.80-1.76)
Sidechain outliers	100360	7569 (1.80-1.76)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Note EDS failed to run properly.

Mol	Chain	Length	Quality of chain
1	1-A	418	81% 12% • 6%
1	10-A	418	89% 5% 6%
1	11-A	418	88% 6% 6%
1	12-A	418	89% 5% 6%
1	13-A	418	87% 7% 6%
1	14-A	418	90% • 6%
1	15-A	418	90% 5% 6%
1	16-A	418	89% 6% 6%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	17-A	418	90% 6%
1	18-A	418	89% 6% 6%
1	19-A	418	90% 6%
1	2-A	418	88% 6% 6%
1	20-A	418	88% 6% 6%
1	3-A	418	87% 7% 6%
1	4-A	418	90% 5% 6%
1	5-A	418	90% 5% 6%
1	6-A	418	89% 6% 6%
1	7-A	418	90% 6%
1	8-A	418	88% 6% 6%
1	9-A	418	89% 6% 6%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 120618 atoms, of which 57140 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Oleandomycin glycosyltransferase.

Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
1	1-A	395	Total	C	H	N	O	S	Se	0	395	0
			5899	1925	2857	543	567	2	5			
1	2-A	395	Total	C	H	N	O	S	Se	0	395	0
			5899	1925	2857	543	567	2	5			
1	3-A	395	Total	C	H	N	O	S	Se	0	395	0
			5899	1925	2857	543	567	2	5			
1	4-A	395	Total	C	H	N	O	S	Se	0	395	0
			5899	1925	2857	543	567	2	5			
1	5-A	395	Total	C	H	N	O	S	Se	0	395	0
			5899	1925	2857	543	567	2	5			
1	6-A	395	Total	C	H	N	O	S	Se	0	395	0
			5899	1925	2857	543	567	2	5			
1	7-A	395	Total	C	H	N	O	S	Se	0	395	0
			5899	1925	2857	543	567	2	5			
1	8-A	395	Total	C	H	N	O	S	Se	0	395	0
			5899	1925	2857	543	567	2	5			
1	9-A	395	Total	C	H	N	O	S	Se	0	395	0
			5899	1925	2857	543	567	2	5			
1	10-A	395	Total	C	H	N	O	S	Se	0	395	0
			5899	1925	2857	543	567	2	5			
1	11-A	395	Total	C	H	N	O	S	Se	0	395	0
			5899	1925	2857	543	567	2	5			
1	12-A	395	Total	C	H	N	O	S	Se	0	395	0
			5899	1925	2857	543	567	2	5			
1	13-A	395	Total	C	H	N	O	S	Se	0	395	0
			5899	1925	2857	543	567	2	5			
1	14-A	395	Total	C	H	N	O	S	Se	0	395	0
			5899	1925	2857	543	567	2	5			
1	15-A	395	Total	C	H	N	O	S	Se	0	395	0
			5899	1925	2857	543	567	2	5			
1	16-A	395	Total	C	H	N	O	S	Se	0	395	0
			5899	1925	2857	543	567	2	5			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
1	17-A	395	Total	C	H	N	O	S	Se	0	395	0
			5899	1925	2857	543	567	2	5			
1	18-A	395	Total	C	H	N	O	S	Se	0	395	0
			5899	1925	2857	543	567	2	5			
1	19-A	395	Total	C	H	N	O	S	Se	0	395	0
			5899	1925	2857	543	567	2	5			
1	20-A	395	Total	C	H	N	O	S	Se	0	395	0
			5899	1925	2857	543	567	2	5			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	EXPRESSION TAG	UNP Q3HTL6
A	-1	SER	-	EXPRESSION TAG	UNP Q3HTL6
A	0	HIS	-	EXPRESSION TAG	UNP Q3HTL6

- Molecule 2 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	3-A	1	Total	Na	0	1
			1	1		
2	18-A	1	Total	Na	0	1
			1	1		
2	11-A	1	Total	Na	0	1
			1	1		
2	16-A	1	Total	Na	0	1
			1	1		
2	4-A	1	Total	Na	0	1
			1	1		
2	20-A	1	Total	Na	0	1
			1	1		
2	12-A	1	Total	Na	0	1
			1	1		
2	19-A	1	Total	Na	0	1
			1	1		
2	17-A	1	Total	Na	0	1
			1	1		
2	5-A	1	Total	Na	0	1
			1	1		
2	13-A	1	Total	Na	0	1
			1	1		
2	8-A	1	Total	Na	0	1
			1	1		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	1-A	1	Total Na 1 1	0	1
2	6-A	1	Total Na 1 1	0	1
2	14-A	1	Total Na 1 1	0	1
2	2-A	1	Total Na 1 1	0	1
2	10-A	1	Total Na 1 1	0	1
2	9-A	1	Total Na 1 1	0	1
2	7-A	1	Total Na 1 1	0	1
2	15-A	1	Total Na 1 1	0	1

- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	1-A	130	Total O 130 130	0	130
3	2-A	128	Total O 128 128	0	128
3	3-A	141	Total O 141 141	0	141
3	4-A	119	Total O 119 119	0	119
3	5-A	125	Total O 125 125	0	125
3	6-A	120	Total O 120 120	0	120
3	7-A	130	Total O 130 130	0	130
3	8-A	146	Total O 146 146	0	146
3	9-A	135	Total O 135 135	0	135
3	10-A	146	Total O 146 146	0	146
3	11-A	129	Total O 129 129	0	129

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	12-A	136	Total O 136 136	0	136
3	13-A	135	Total O 135 135	0	135
3	14-A	129	Total O 129 129	0	129
3	15-A	121	Total O 121 121	0	121
3	16-A	133	Total O 133 133	0	133
3	17-A	132	Total O 132 132	0	132
3	18-A	126	Total O 126 126	0	126
3	19-A	134	Total O 134 134	0	134
3	20-A	123	Total O 123 123	0	123

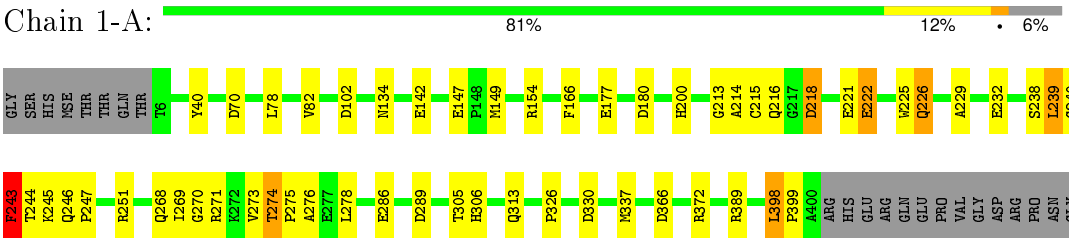
3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

Note EDS failed to run properly.

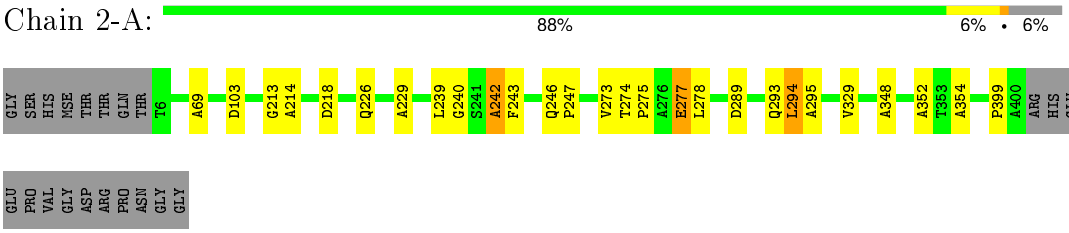
- Molecule 1: Oleandomycin glycosyltransferase

Chain 1-A:



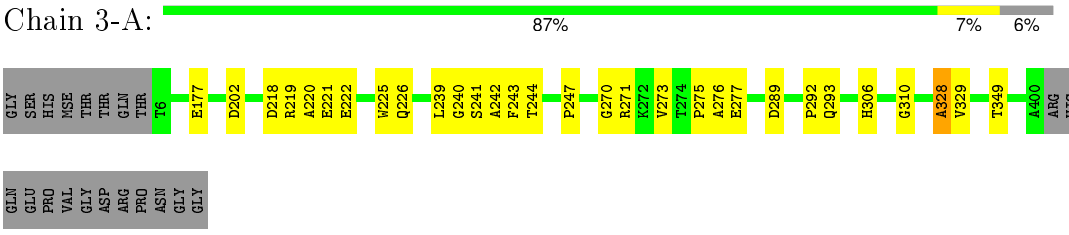
- Molecule 1: Oleandomycin glycosyltransferase

Chain 2-A:



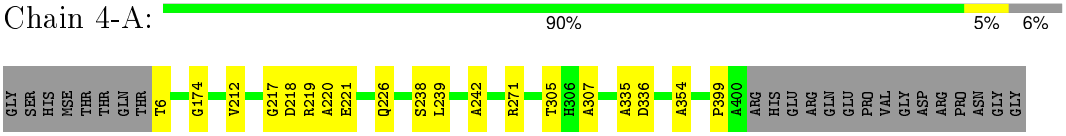
- Molecule 1: Oleandomycin glycosyltransferase

Chain 3-A:



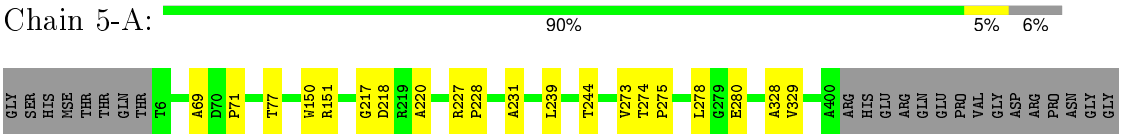
- Molecule 1: Oleandomycin glycosyltransferase

Chain 4-A:



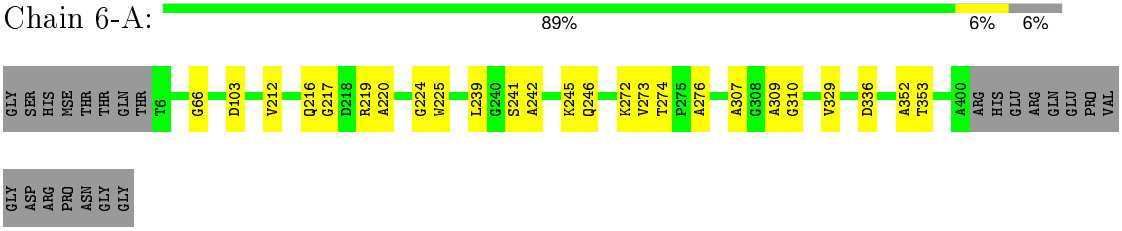
- Molecule 1: Oleandomycin glycosyltransferase

Chain 5-A:



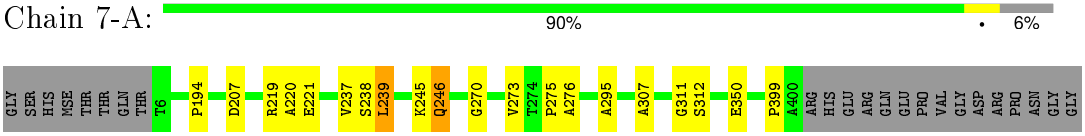
• Molecule 1: Oleandomycin glycosyltransferase

Chain 6-A:



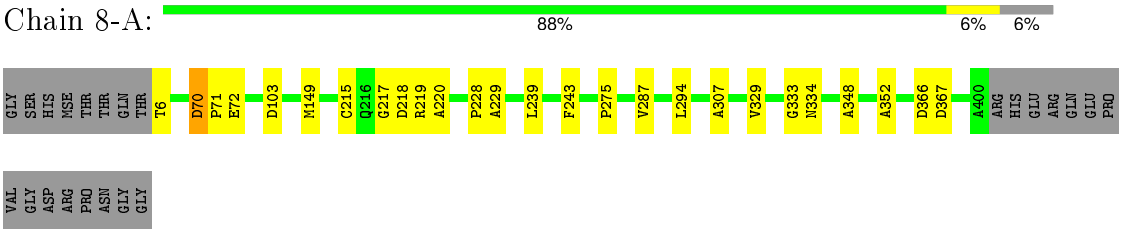
• Molecule 1: Oleandomycin glycosyltransferase

Chain 7-A:



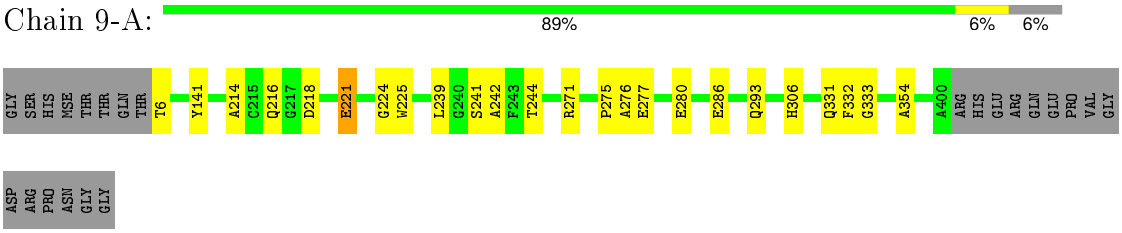
• Molecule 1: Oleandomycin glycosyltransferase

Chain 8-A:



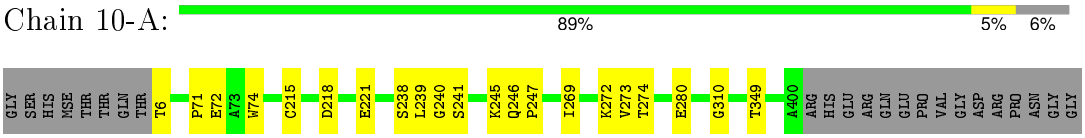
• Molecule 1: Oleandomycin glycosyltransferase

Chain 9-A:



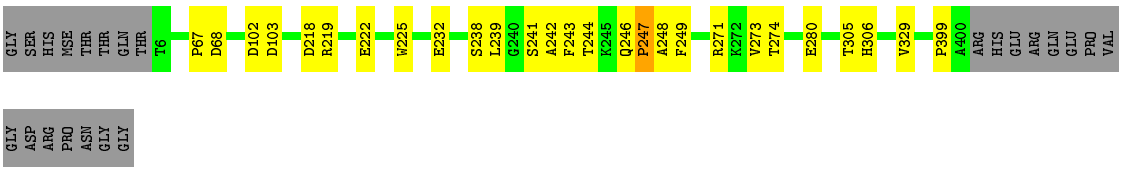
• Molecule 1: Oleandomycin glycosyltransferase

Chain 10-A:



• Molecule 1: Oleandomycin glycosyltransferase

Chain 11-A:

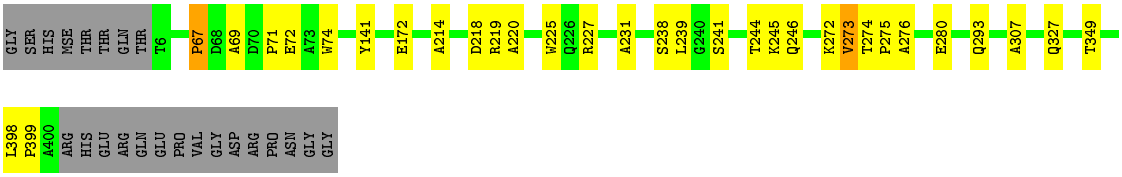


• Molecule 1: Oleandomycin glycosyltransferase

Chain 12-A:

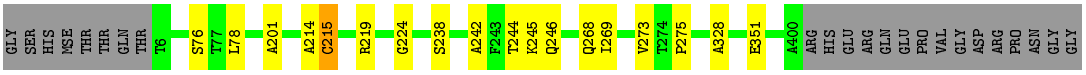
• Molecule 1: Oleandomycin glycosyltransferase

Chain 13-A:



• Molecule 1: Oleandomycin glycosyltransferase

Chain 14-A:

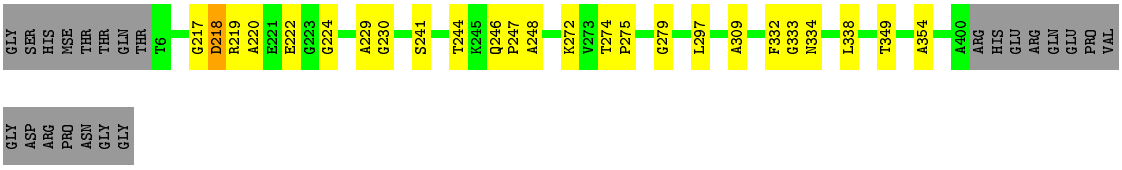


• Molecule 1: Oleandomycin glycosyltransferase

Chain 15-A:

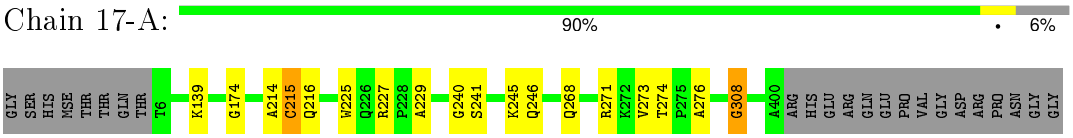
• Molecule 1: Oleandomycin glycosyltransferase

Chain 16-A:



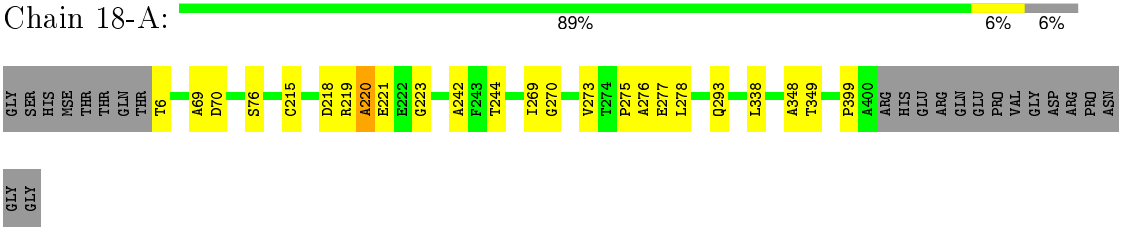
• Molecule 1: Oleandomycin glycosyltransferase

Chain 17-A:



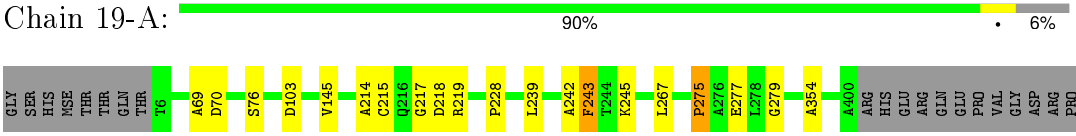
- Molecule 1: Oleandomycin glycosyltransferase

Chain 18-A:



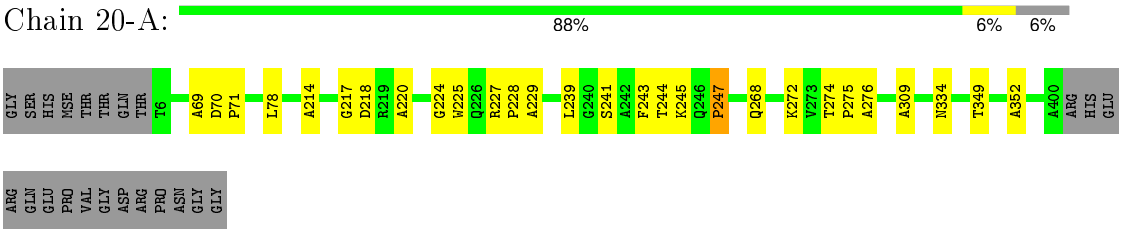
- Molecule 1: Oleandomycin glycosyltransferase

Chain 19-A:



- Molecule 1: Oleandomycin glycosyltransferase

Chain 20-A:



4 Data and refinement statistics (i)

EDS failed to run properly - this section will therefore be incomplete.

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, α, β, γ	124.13Å 124.13Å 67.64Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	42.08 – 1.77	Depositor
% Data completeness (in resolution range)	100.0 (42.08-1.77)	Depositor
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.69 (at 1.77Å)	Xtriage
Refinement program	PHENIX (phenix.ensemble_refinement: dev_1420)	Depositor
R, R_{free}	0.150 , 0.188	Depositor
Wilson B-factor (Å^2)	29.9	Xtriage
Anisotropy	0.297	Xtriage
Estimated twinning fraction	0.022 for h,-h-k,-l	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Outliers	0 of 37890 reflections	Xtriage
Total number of atoms	120618	wwPDB-VP
Average B, all atoms (Å^2)	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.67% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: NA

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$\|Z\|>5$	RMSZ	#$\|Z\|>5$
1	1-A	0.58	0/3115	0.83	2/4251 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	1-A	0	6
1	2-A	0	4
1	3-A	0	5
1	4-A	0	3
1	5-A	0	3
1	6-A	0	2
1	7-A	0	5
1	8-A	0	4
1	9-A	0	2
1	10-A	0	3
1	11-A	0	5
1	12-A	0	5
1	13-A	0	6
1	14-A	0	5
1	15-A	0	5
1	16-A	0	2
1	17-A	0	7
1	18-A	0	6
1	19-A	0	6
1	20-A	0	3
All	All	0	87

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	1-A	239[A]	LEU	CB-CG-CD2	5.63	120.58	111.00
1	1-A	243[A]	PHE	CB-CG-CD1	5.38	124.56	120.80

There are no chirality outliers.

5 of 87 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	1-A	213[A]	GLY	Peptide
1	1-A	215[A]	CYS	Peptide
1	1-A	241[A]	SER	Peptide
1	1-A	268[A]	GLN	Peptide
1	1-A	275[A]	PRO	Peptide

5.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	1-A	3042	2857	2982	0	0
1	2-A	3042	2857	2982	0	0
1	3-A	3042	2857	2982	0	0
1	4-A	3042	2857	2982	0	0
1	5-A	3042	2857	2982	0	0
1	6-A	3042	2857	2982	0	0
1	7-A	3042	2857	2982	0	0
1	8-A	3042	2857	2982	0	0
1	9-A	3042	2857	2982	0	0
1	10-A	3042	2857	2982	0	0
1	11-A	3042	2857	2982	0	0
1	12-A	3042	2857	2982	0	0
1	13-A	3042	2857	2982	0	0
1	14-A	3042	2857	2982	0	0
1	15-A	3042	2857	2982	0	0
1	16-A	3042	2857	2982	0	0
1	17-A	3042	2857	2982	0	0
1	18-A	3042	2857	2982	0	0
1	19-A	3042	2857	2982	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	20-A	3042	2857	2982	0	0
2	1-A	1	0	0	0	0
2	2-A	1	0	0	0	0
2	3-A	1	0	0	0	0
2	4-A	1	0	0	0	0
2	5-A	1	0	0	0	0
2	6-A	1	0	0	0	0
2	7-A	1	0	0	0	0
2	8-A	1	0	0	0	0
2	9-A	1	0	0	0	0
2	10-A	1	0	0	0	0
2	11-A	1	0	0	0	0
2	12-A	1	0	0	0	0
2	13-A	1	0	0	0	0
2	14-A	1	0	0	0	0
2	15-A	1	0	0	0	0
2	16-A	1	0	0	0	0
2	17-A	1	0	0	0	0
2	18-A	1	0	0	0	0
2	19-A	1	0	0	0	0
2	20-A	1	0	0	0	0
3	1-A	130	0	0	0	0
3	2-A	128	0	0	0	0
3	3-A	141	0	0	0	0
3	4-A	119	0	0	0	0
3	5-A	125	0	0	0	0
3	6-A	120	0	0	0	0
3	7-A	130	0	0	0	0
3	8-A	146	0	0	0	0
3	9-A	135	0	0	0	0
3	10-A	146	0	0	0	0
3	11-A	129	0	0	0	0
3	12-A	136	0	0	0	0
3	13-A	135	0	0	0	0
3	14-A	129	0	0	0	0
3	15-A	121	0	0	0	0
3	16-A	133	0	0	0	0
3	17-A	132	0	0	0	0
3	18-A	126	0	0	0	0
3	19-A	134	0	0	0	0
3	20-A	123	0	0	0	0
All	All	63478	57140	59640	0	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). Clashscore could not be calculated for this entry.

There are no clashes within the asymmetric unit.

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	1-A	393/418 (94%)	347 (88%)	27 (7%)	19 (5%)	3	0
1	2-A	393/418 (94%)	337 (86%)	30 (8%)	26 (7%)	1	0
1	3-A	393/418 (94%)	344 (88%)	23 (6%)	26 (7%)	1	0
1	4-A	393/418 (94%)	342 (87%)	35 (9%)	16 (4%)	3	0
1	5-A	393/418 (94%)	345 (88%)	31 (8%)	17 (4%)	3	0
1	6-A	393/418 (94%)	340 (86%)	30 (8%)	23 (6%)	2	0
1	7-A	393/418 (94%)	351 (89%)	25 (6%)	17 (4%)	3	0
1	8-A	393/418 (94%)	332 (84%)	38 (10%)	23 (6%)	2	0
1	9-A	393/418 (94%)	333 (85%)	37 (9%)	23 (6%)	2	0
1	10-A	393/418 (94%)	346 (88%)	29 (7%)	18 (5%)	3	0
1	11-A	393/418 (94%)	341 (87%)	29 (7%)	23 (6%)	2	0
1	12-A	393/418 (94%)	343 (87%)	31 (8%)	19 (5%)	3	0
1	13-A	393/418 (94%)	329 (84%)	36 (9%)	28 (7%)	1	0
1	14-A	393/418 (94%)	344 (88%)	35 (9%)	14 (4%)	4	0
1	15-A	393/418 (94%)	344 (88%)	35 (9%)	14 (4%)	4	0
1	16-A	393/418 (94%)	338 (86%)	31 (8%)	24 (6%)	2	0
1	17-A	393/418 (94%)	344 (88%)	36 (9%)	13 (3%)	5	0
1	18-A	393/418 (94%)	340 (86%)	34 (9%)	19 (5%)	3	0
1	19-A	393/418 (94%)	347 (88%)	30 (8%)	16 (4%)	3	0
1	20-A	393/418 (94%)	337 (86%)	30 (8%)	26 (7%)	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	7860/8360 (94%)	6824 (87%)	632 (8%)	404 (5%)	2 0

5 of 404 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	1-A	214[A]	ALA
1	1-A	218[A]	ASP
1	1-A	222[A]	GLU
1	1-A	232[A]	GLU
1	1-A	240[A]	GLY

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	1-A	309/322 (96%)	270 (87%)	39 (13%)	5 1

5 of 39 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	1-A	226[A]	GLN
1	1-A	244[A]	THR
1	1-A	372[A]	ARG
1	1-A	238[A]	SER
1	1-A	239[A]	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	1-A	264[A]	HIS
1	1-A	339[A]	GLN
1	1-A	293[A]	GLN
1	1-A	81[A]	ASN
1	1-A	331[A]	GLN

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

There are no carbohydrates in this entry.

5.6 Ligand geometry

Of 20 ligands modelled in this entry, 20 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

EDS failed to run properly - this section will therefore be empty.

6.2 Non-standard residues in protein, DNA, RNA chains

EDS failed to run properly - this section will therefore be empty.

6.3 Carbohydrates

EDS failed to run properly - this section will therefore be empty.

6.4 Ligands

EDS failed to run properly - this section will therefore be empty.

6.5 Other polymers

EDS failed to run properly - this section will therefore be empty.