



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 17, 2026 – 10:53 PM UTC

PDB ID : 4K7H / pdb_00004k7h
Title : Major capsid protein P1 of the Pseudomonas phage phi6
Authors : Boura, E.; Nemecek, D.; Plevka, P.; Steven, C.A.; Hurley, J.H.
Deposited on : 2013-04-17
Resolution : 3.60 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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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

MolProbity	:	4-5-2 with Phenix2.0
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

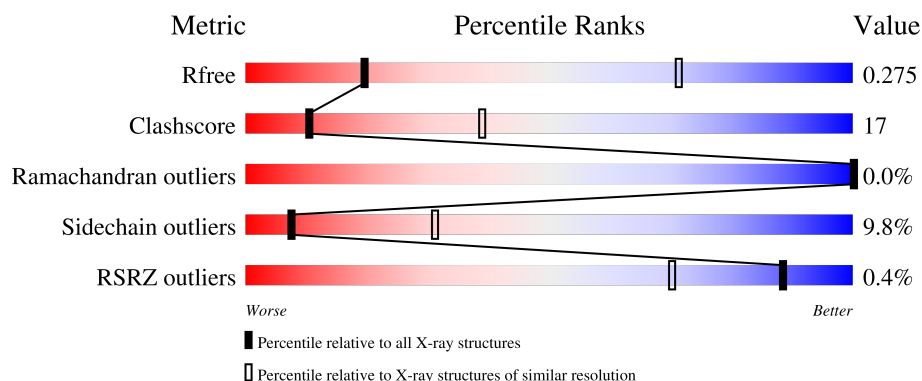
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.60 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))
R_{free}	180053	1747 (3.70-3.50)
Clashscore	190562	1827 (3.70-3.50)
Ramachandran outliers	187476	1773 (3.70-3.50)
Sidechain outliers	187428	1772 (3.70-3.50)
RSRZ outliers	180081	1745 (3.70-3.50)

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 of the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria respectively. 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 $\leq 5\%$. 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.

Mol	Chain	Length	Quality of chain
1	A	775	
1	B	775	
1	C	775	
1	D	775	
1	E	775	

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 29340 atoms, of which 0 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 Major inner protein P1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	760	Total	C	N	O	S	0	0	0
			5868	3709	1032	1105	22			
1	B	760	Total	C	N	O	S	0	0	0
			5868	3709	1032	1105	22			
1	C	760	Total	C	N	O	S	0	0	0
			5868	3709	1032	1105	22			
1	D	760	Total	C	N	O	S	0	0	0
			5868	3709	1032	1105	22			
1	E	760	Total	C	N	O	S	0	0	0
			5868	3709	1032	1105	22			

There are 35 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	expression tag	UNP P11126
A	770	HIS	-	expression tag	UNP P11126
A	771	HIS	-	expression tag	UNP P11126
A	772	HIS	-	expression tag	UNP P11126
A	773	HIS	-	expression tag	UNP P11126
A	774	HIS	-	expression tag	UNP P11126
A	775	HIS	-	expression tag	UNP P11126
B	1	GLY	-	expression tag	UNP P11126
B	770	HIS	-	expression tag	UNP P11126
B	771	HIS	-	expression tag	UNP P11126
B	772	HIS	-	expression tag	UNP P11126
B	773	HIS	-	expression tag	UNP P11126
B	774	HIS	-	expression tag	UNP P11126
B	775	HIS	-	expression tag	UNP P11126
C	1	GLY	-	expression tag	UNP P11126
C	770	HIS	-	expression tag	UNP P11126
C	771	HIS	-	expression tag	UNP P11126
C	772	HIS	-	expression tag	UNP P11126
C	773	HIS	-	expression tag	UNP P11126

Continued on next page...

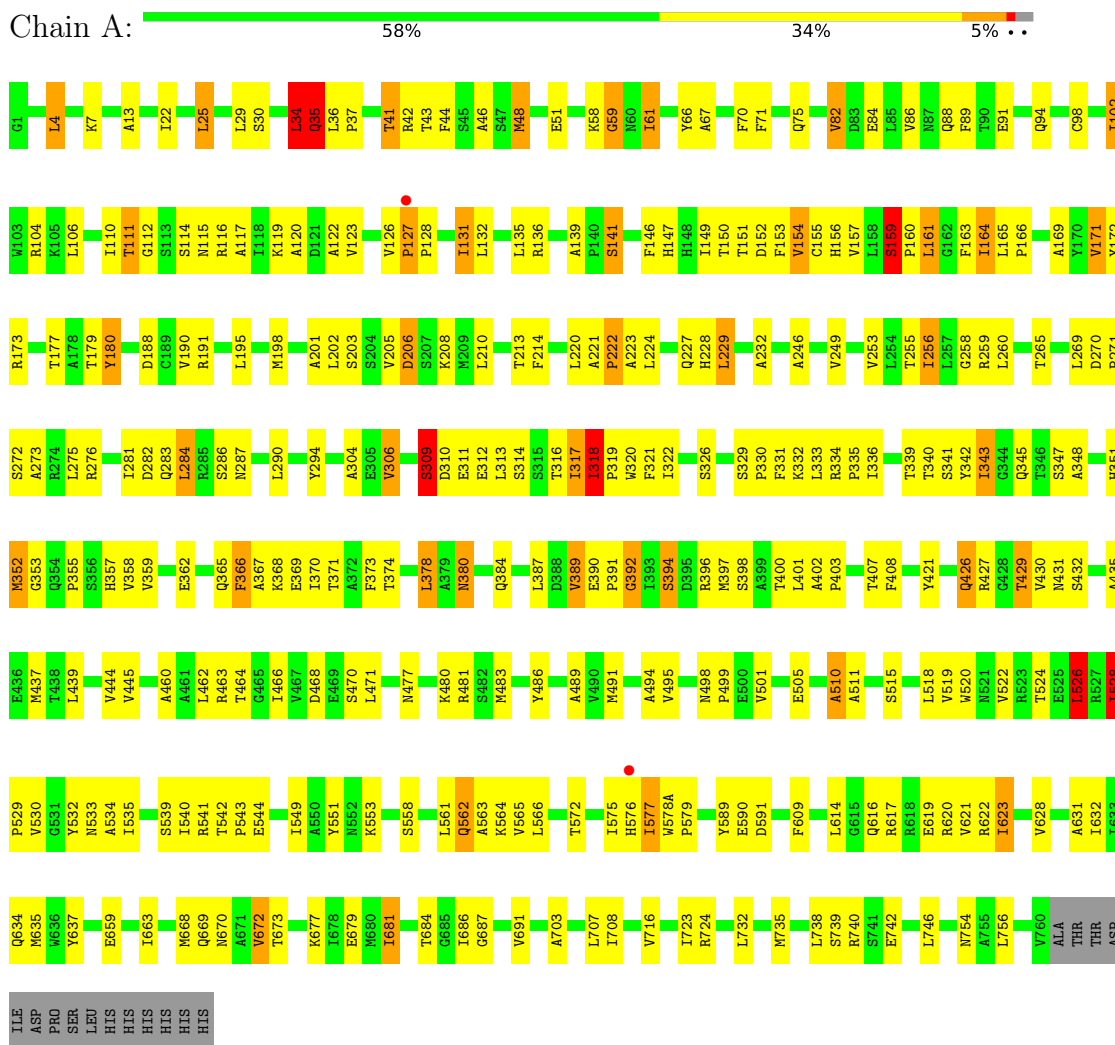
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	774	HIS	-	expression tag	UNP P11126
C	775	HIS	-	expression tag	UNP P11126
D	1	GLY	-	expression tag	UNP P11126
D	770	HIS	-	expression tag	UNP P11126
D	771	HIS	-	expression tag	UNP P11126
D	772	HIS	-	expression tag	UNP P11126
D	773	HIS	-	expression tag	UNP P11126
D	774	HIS	-	expression tag	UNP P11126
D	775	HIS	-	expression tag	UNP P11126
E	1	GLY	-	expression tag	UNP P11126
E	770	HIS	-	expression tag	UNP P11126
E	771	HIS	-	expression tag	UNP P11126
E	772	HIS	-	expression tag	UNP P11126
E	773	HIS	-	expression tag	UNP P11126
E	774	HIS	-	expression tag	UNP P11126
E	775	HIS	-	expression tag	UNP P11126

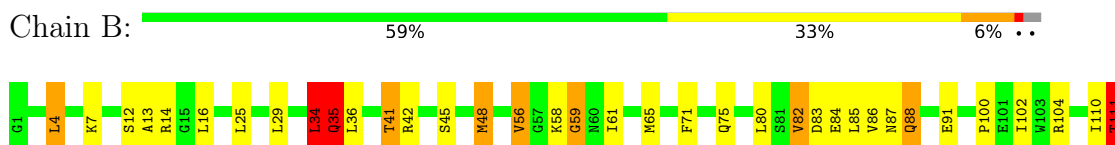
3 Residue-property plots

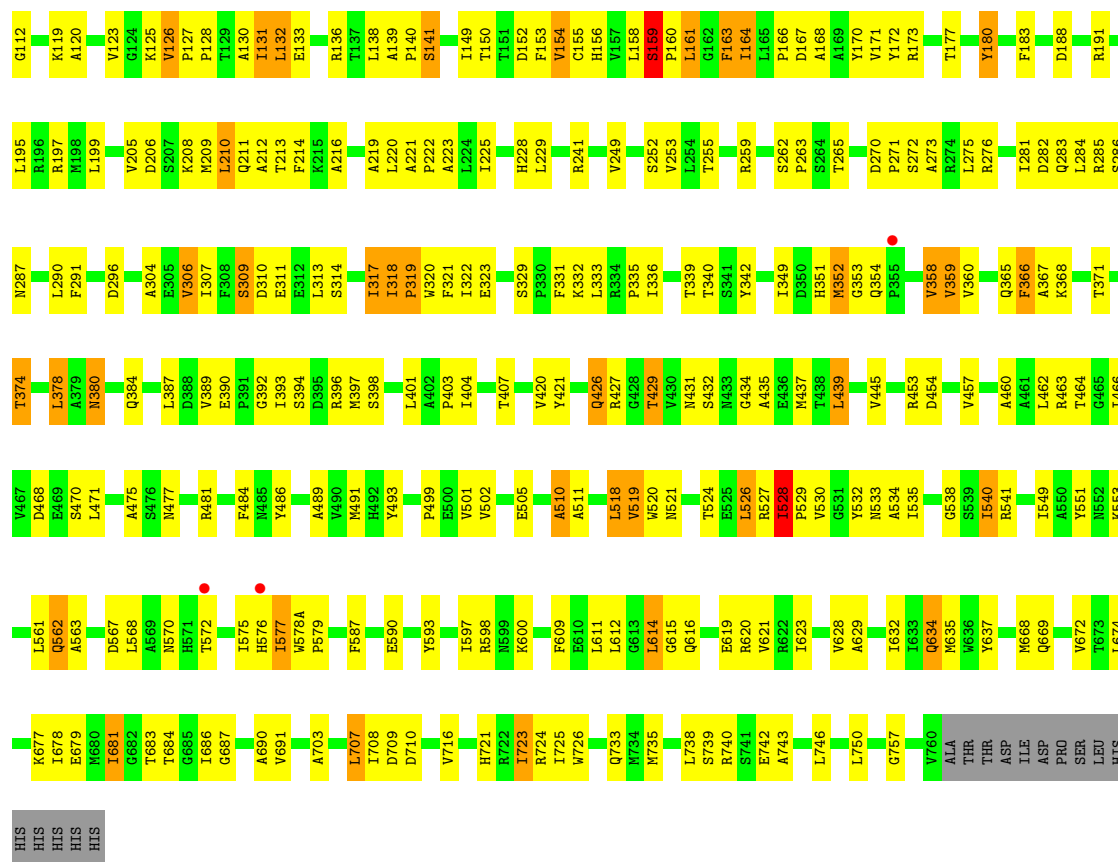
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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.

• Molecule 1: Major inner protein P1



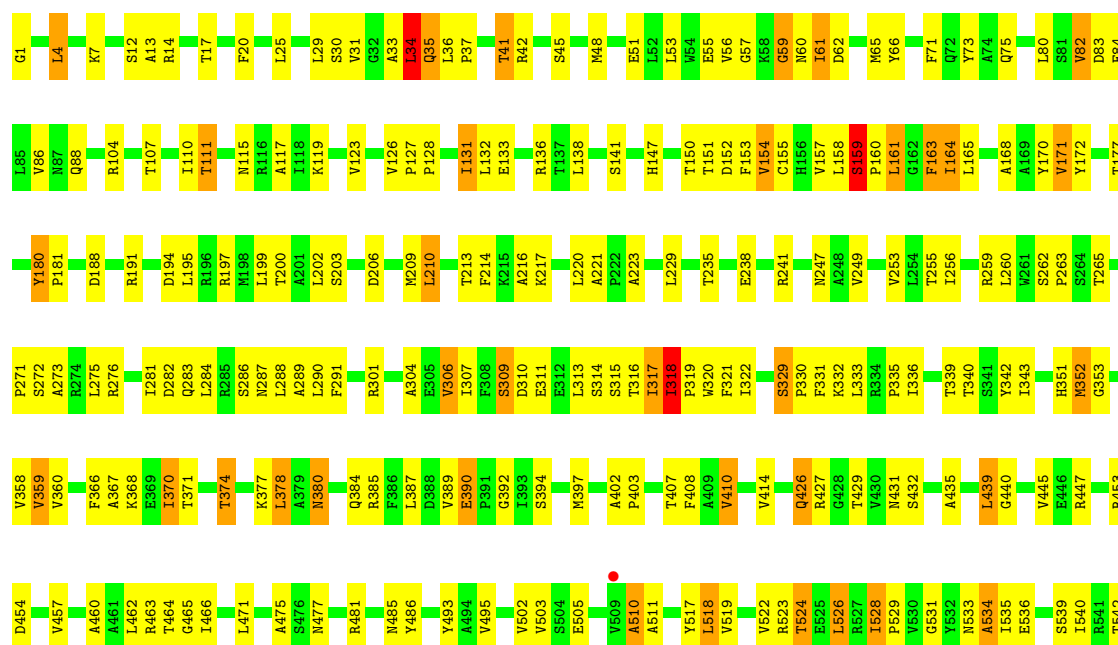
• Molecule 1: Major inner protein P1

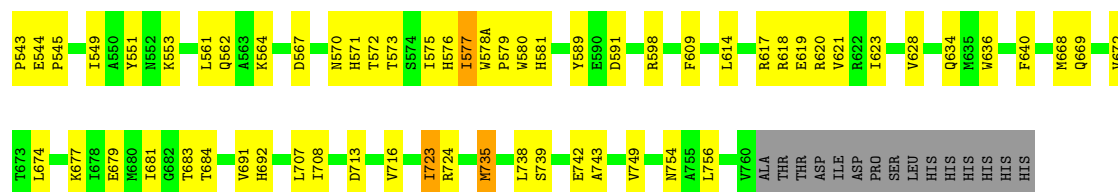




• Molecule 1: Major inner protein P1

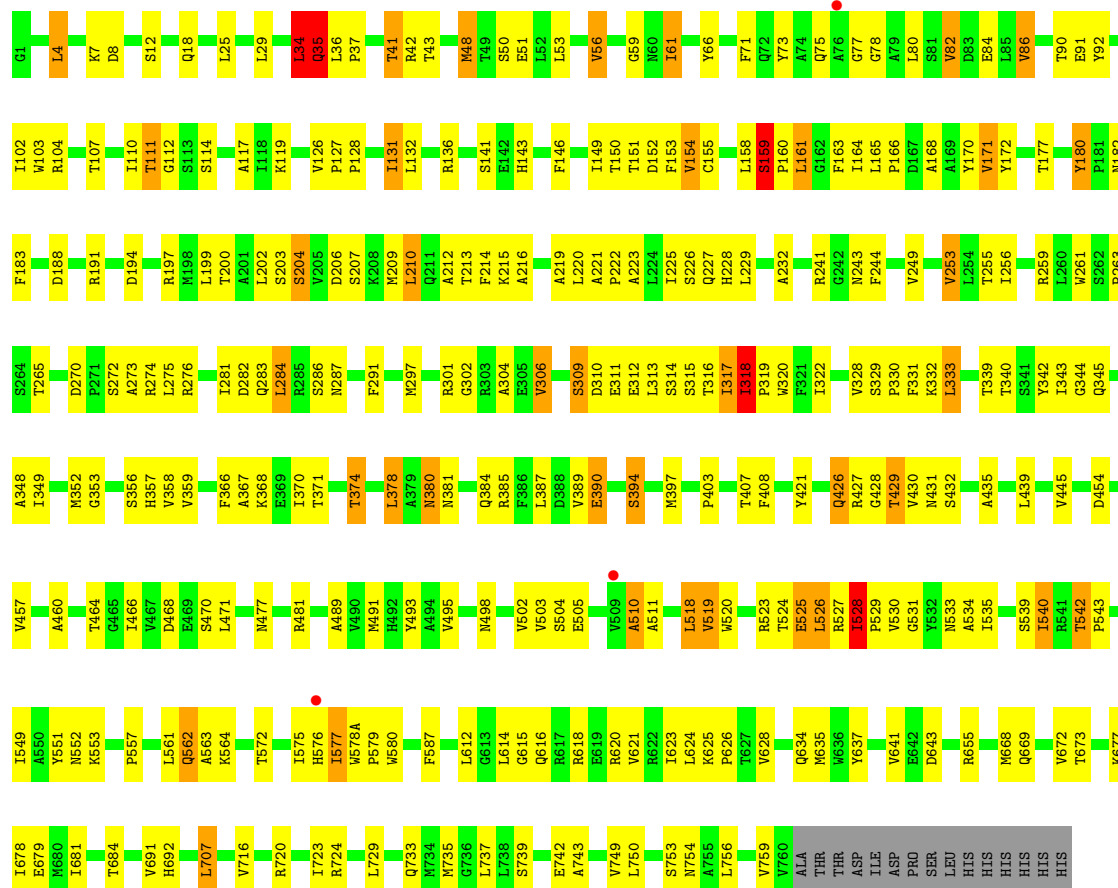
Chain C: 60% 33% 5% .





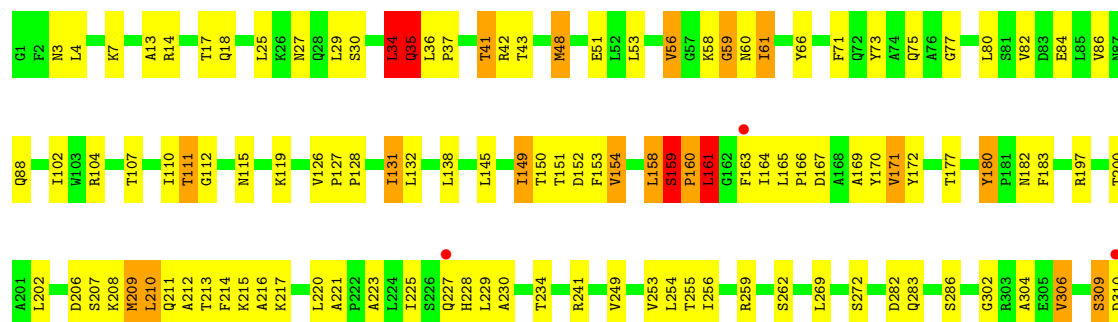
• Molecule 1: Major inner protein P1

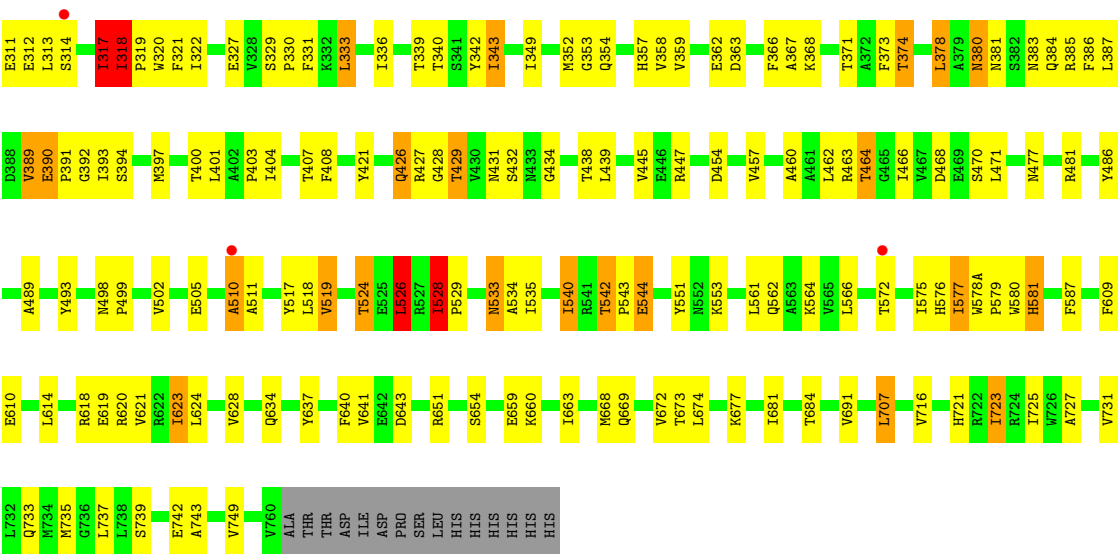
Chain D: 59% 33% 5% ..



• Molecule 1: Major inner protein P1

Chain E: 62% 30% 5% ..





4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α , β , γ	182.59Å 278.85Å 246.47Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.92 – 3.60 40.92 – 3.60	Depositor EDS
% Data completeness (in resolution range)	91.0 (40.92-3.60) 90.9 (40.92-3.60)	Depositor EDS
R_{merge}	0.19	Depositor
R_{sym}	0.19	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.08 (at 3.57Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, R_{free}	0.217 , 0.274 0.221 , 0.275	Depositor DCC
R_{free} test set	3350 reflections (4.58%)	wwPDB-VP
Wilson B-factor (Å ²)	118.3	Xtriage
Anisotropy	0.115	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 155.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	29340	wwPDB-VP
Average B, all atoms (Å ²)	178.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.79% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

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	A	0.53	1/5988 (0.0%)	1.07	30/8146 (0.4%)
1	B	0.54	2/5988 (0.0%)	1.07	37/8146 (0.5%)
1	C	0.54	2/5988 (0.0%)	1.05	22/8146 (0.3%)
1	D	0.57	2/5988 (0.0%)	1.08	29/8146 (0.4%)
1	E	0.58	2/5988 (0.0%)	1.10	37/8146 (0.5%)
All	All	0.55	9/29940 (0.0%)	1.07	155/40730 (0.4%)

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	A	0	4
1	B	0	5
1	C	0	3
1	D	0	5
1	E	0	4
All	All	0	21

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	577	ILE	CA-C	8.19	1.61	1.52
1	E	577	ILE	CA-C	7.85	1.61	1.52
1	C	577	ILE	CA-C	7.83	1.61	1.52
1	A	577	ILE	CA-C	6.36	1.59	1.52
1	D	577	ILE	CA-C	6.35	1.60	1.52

The worst 5 of 155 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	542	THR	CA-C-N	10.01	129.76	118.85

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	542	THR	C-N-CA	10.01	129.76	118.85
1	A	35	GLN	N-CA-C	9.81	126.25	111.04
1	E	35	GLN	N-CA-C	9.81	126.24	111.04
1	A	329	SER	CA-C-N	9.46	129.09	119.82

There are no chirality outliers.

5 of 21 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	159	SER	Peptide
1	A	318	ILE	Peptide
1	A	34	LEU	Peptide
1	A	35	GLN	Peptide
1	B	34	LEU	Peptide

5.2 Too-close contacts

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	A	5868	0	5807	217	0
1	B	5868	0	5807	214	0
1	C	5868	0	5807	192	1
1	D	5868	0	5807	204	2
1	E	5868	0	5807	177	1
All	All	29340	0	29035	985	2

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 17.

The worst 5 of 985 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:51:GLU:H	1:C:564:LYS:HE3	1.32	0.92
1:D:367:ALA:HB1	1:D:368:LYS:HA	1.51	0.91
1:A:51:GLU:H	1:A:564:LYS:HE3	1.36	0.90

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:E:318:ILE:HG22	1:E:320:TRP:HB3	1.54	0.89
1:E:367:ALA:HB1	1:E:368:LYS:HA	1.55	0.86

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:523:ARG:O	1:D:523:ARG:NH2[8_455]	2.12	0.08
1:D:655:ARG:O	1:E:27:ASN:ND2[4_555]	2.18	0.02

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	A	758/775 (98%)	717 (95%)	41 (5%)	0	100	100
1	B	758/775 (98%)	720 (95%)	38 (5%)	0	100	100
1	C	758/775 (98%)	718 (95%)	40 (5%)	0	100	100
1	D	758/775 (98%)	717 (95%)	41 (5%)	0	100	100
1	E	758/775 (98%)	722 (95%)	35 (5%)	1 (0%)	48	79
All	All	3790/3875 (98%)	3594 (95%)	195 (5%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	61	ILE

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	A	618/643 (96%)	559 (90%)	59 (10%)	8	32
1	B	618/643 (96%)	564 (91%)	54 (9%)	9	34
1	C	618/643 (96%)	556 (90%)	62 (10%)	7	30
1	D	618/643 (96%)	555 (90%)	63 (10%)	7	30
1	E	618/643 (96%)	553 (90%)	65 (10%)	6	29
All	All	3090/3215 (96%)	2787 (90%)	303 (10%)	7	31

5 of 303 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	624	LEU
1	E	518	LEU
1	E	4	LEU
1	E	171	VAL
1	E	707	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 24 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	211	GLN
1	D	211	GLN
1	D	182	ASN
1	D	228	HIS
1	A	552	ASN

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

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

5.5 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

There are no ligands in this entry.

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled ‘#RSRZ > 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q < 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å ²)	Q<0.9
1	A	760/775 (98%)	-0.09	2 (0%) 90 75	130, 172, 242, 296	0
1	B	760/775 (98%)	-0.14	3 (0%) 88 70	125, 172, 241, 296	0
1	C	760/775 (98%)	-0.12	1 (0%) 92 85	123, 170, 243, 295	0
1	D	760/775 (98%)	-0.12	3 (0%) 88 70	122, 169, 241, 293	0
1	E	760/775 (98%)	-0.11	6 (0%) 82 58	122, 168, 241, 296	0
All	All	3800/3875 (98%)	-0.12	15 (0%) 88 70	122, 171, 241, 296	0

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	576	HIS	3.1
1	D	76	ALA	2.7
1	E	163	PHE	2.7
1	E	572	THR	2.6
1	D	509	VAL	2.6

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

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

6.3 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.