



wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 05:37 am BST

PDB ID : 4LIN
Title : Exploring the atomic structure and conformational flexibility of a 320 angstrom long engineered viral fiber using X-ray crystallography
Authors : Bhardwaj, A.; Cingolani, G.
Deposited on : 2013-07-02
Resolution : 2.70 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

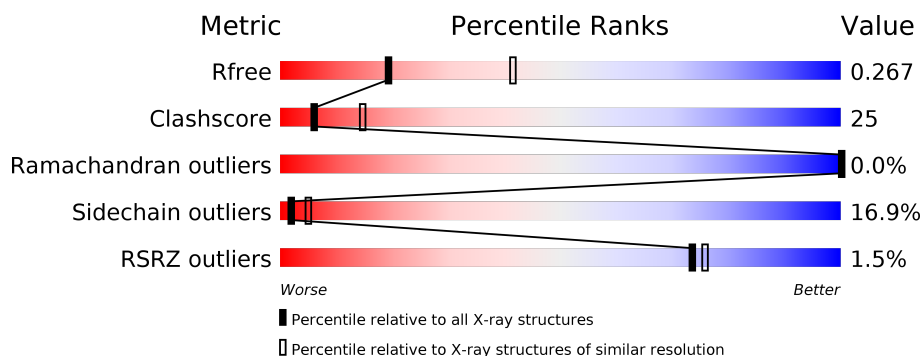
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.70 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}	130704	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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 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	300	<div> <div></div> <div>49%</div> <div>42%</div> <div>5%</div> <div>.</div> </div>
1	B	300	<div> <div>%</div> <div>51%</div> <div>37%</div> <div>8%</div> <div>.</div> </div>
1	C	300	<div> <div></div> <div>50%</div> <div>41%</div> <div>5%</div> <div>.</div> </div>
1	D	300	<div> <div>%</div> <div>50%</div> <div>40%</div> <div>6%</div> <div>.</div> </div>
1	E	300	<div> <div>%</div> <div>52%</div> <div>38%</div> <div>7%</div> <div>.</div> </div>
1	F	300	<div> <div>2%</div> <div>48%</div> <div>44%</div> <div>.</div> <div>.</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	300	
1	H	300	
1	I	300	
1	J	300	
1	K	300	
1	L	300	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	CL	A	1303	-	-	X	-
3	CL	H	1301	-	-	X	-
3	CL	I	1301	-	-	X	X
3	CL	L	1301	-	-	X	-

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 26552 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 Tail needle protein gp26.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	289	Total	C	N	O	S	0	0	0
			2148	1310	386	451	1			
1	B	289	Total	C	N	O	S	0	0	0
			2148	1310	386	451	1			
1	C	289	Total	C	N	O	S	0	0	0
			2148	1310	386	451	1			
1	D	289	Total	C	N	O	S	0	0	0
			2148	1310	386	451	1			
1	E	289	Total	C	N	O	S	0	0	0
			2148	1310	386	451	1			
1	F	289	Total	C	N	O	S	0	0	0
			2148	1310	386	451	1			
1	G	289	Total	C	N	O	S	0	0	0
			2148	1310	386	451	1			
1	H	289	Total	C	N	O	S	0	0	0
			2148	1310	386	451	1			
1	I	289	Total	C	N	O	S	0	0	0
			2148	1310	386	451	1			
1	J	289	Total	C	N	O	S	0	0	0
			2148	1310	386	451	1			
1	K	289	Total	C	N	O	S	0	0	0
			2148	1310	386	451	1			
1	L	289	Total	C	N	O	S	0	0	0
			2148	1310	386	451	1			

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	G	1	Total	Ca	0	0
			1	1		
2	J	1	Total	Ca	0	0
			1	1		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total 1	Ca 1	0	0
2	E	1	Total 1	Ca 1	0	0

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	D	1	Total 1	Cl 1	0	0
3	K	1	Total 1	Cl 1	0	0
3	E	1	Total 1	Cl 1	0	0
3	H	1	Total 1	Cl 1	0	0
3	I	1	Total 1	Cl 1	0	0
3	A	2	Total 2	Cl 2	0	0
3	L	1	Total 1	Cl 1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	70	Total 70	O 70	0	0
4	B	81	Total 81	O 81	0	0
4	C	54	Total 54	O 54	0	0
4	D	68	Total 68	O 68	0	0
4	E	61	Total 61	O 61	0	0
4	F	70	Total 70	O 70	0	0
4	G	58	Total 58	O 58	0	0
4	H	54	Total 54	O 54	0	0

Continued on next page...

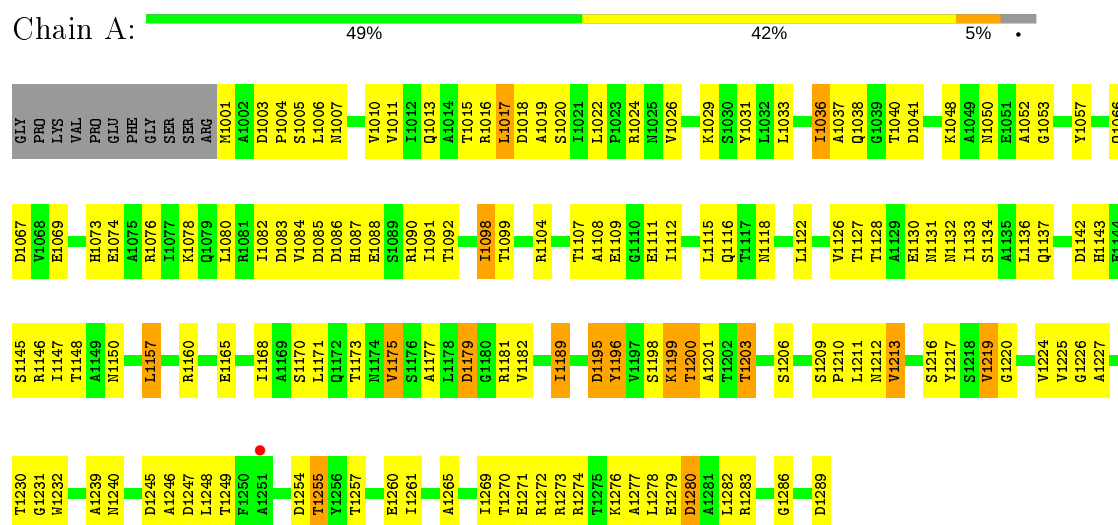
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	I	61	Total 61	O 61	0	0
4	J	67	Total 67	O 67	0	0
4	K	53	Total 53	O 53	0	0
4	L	67	Total 67	O 67	0	0

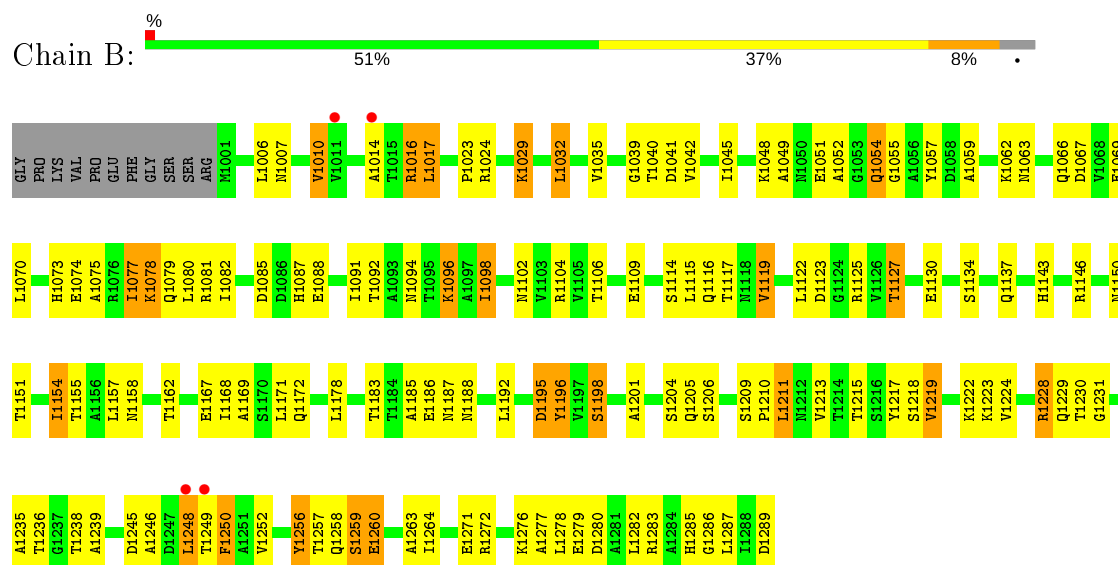
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA 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: Tail needle protein gp26



• Molecule 1: Tail needle protein gp26

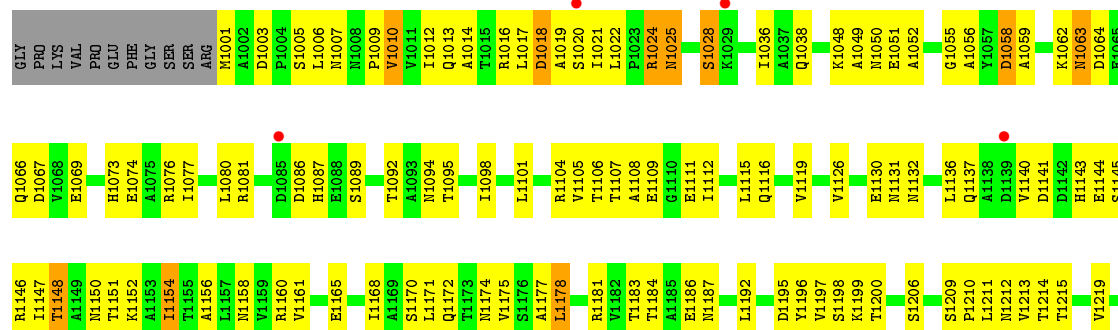


• Molecule 1: Tail needle protein gp26

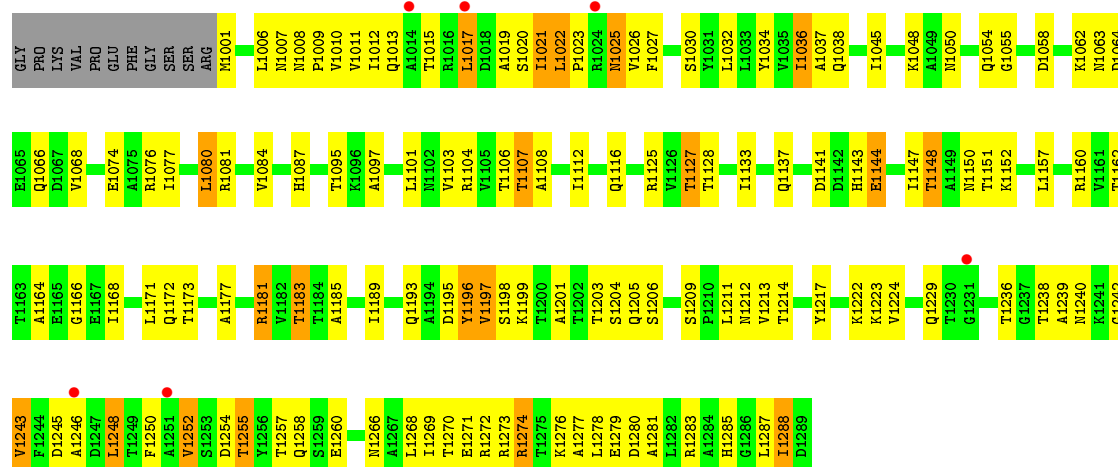




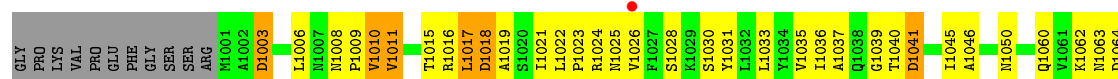
• Molecule 1: Tail needle protein gp26

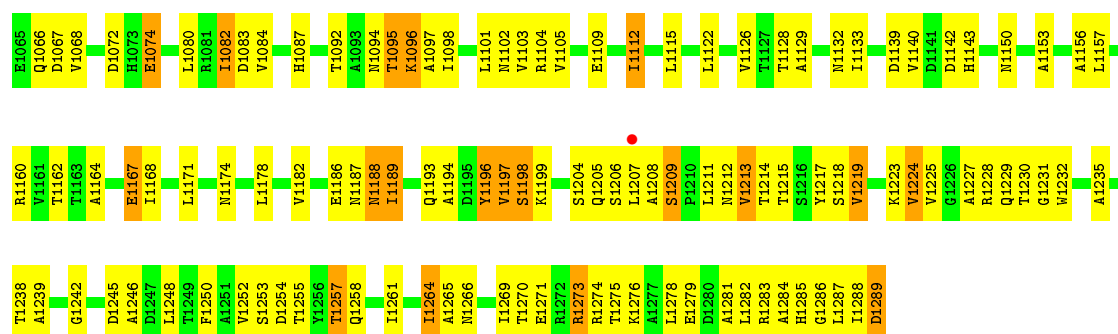


• Molecule 1: Tail needle protein gp26

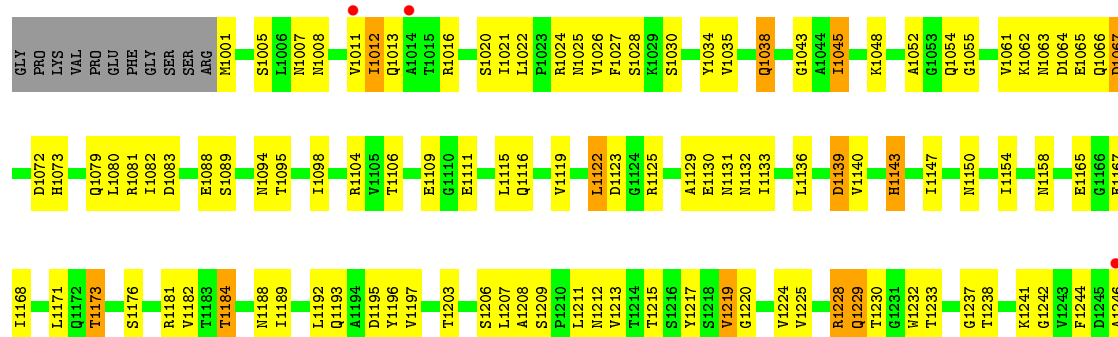


• Molecule 1: Tail needle protein gp26

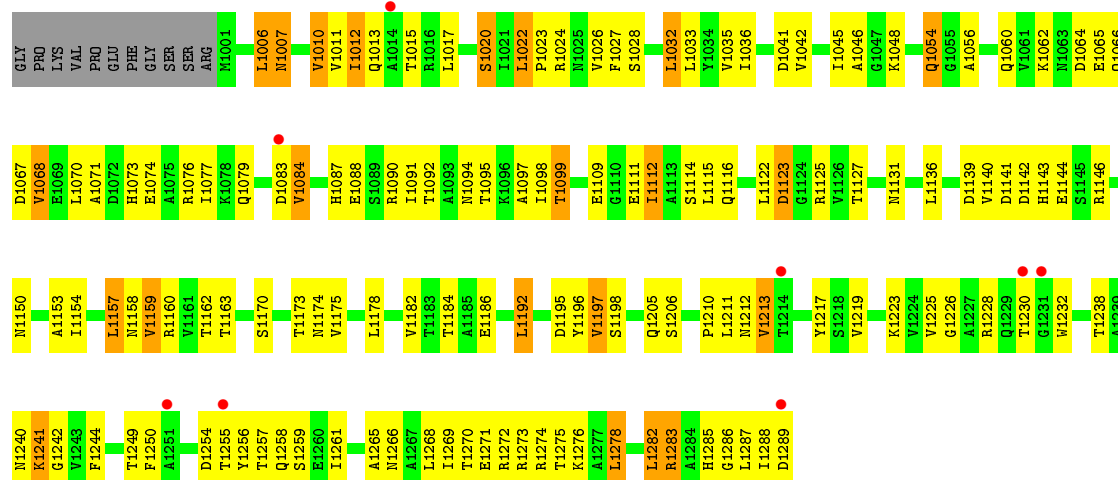




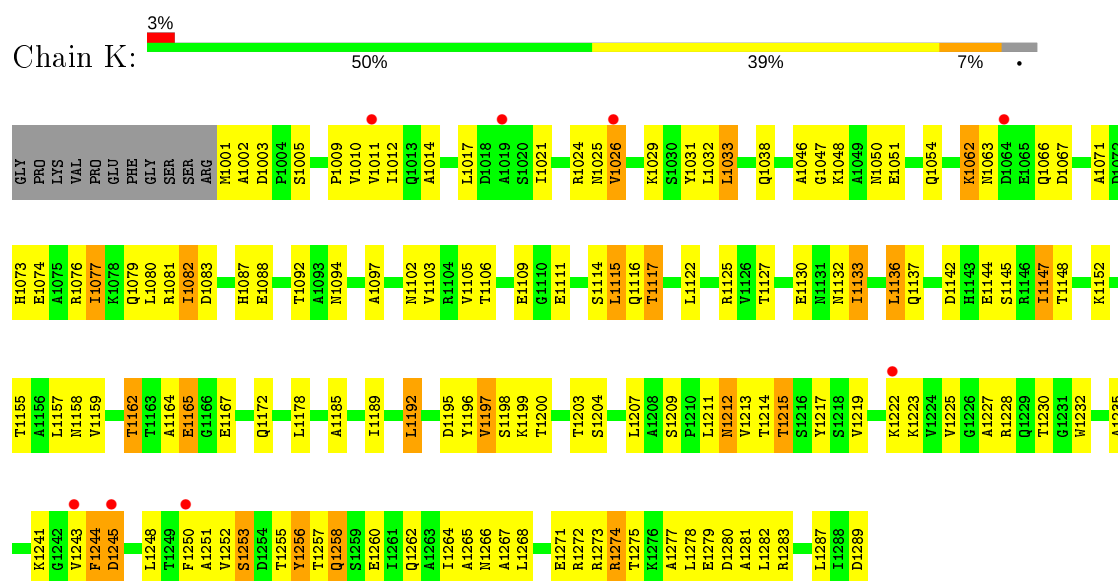
• Molecule 1: Tail needle protein gp26



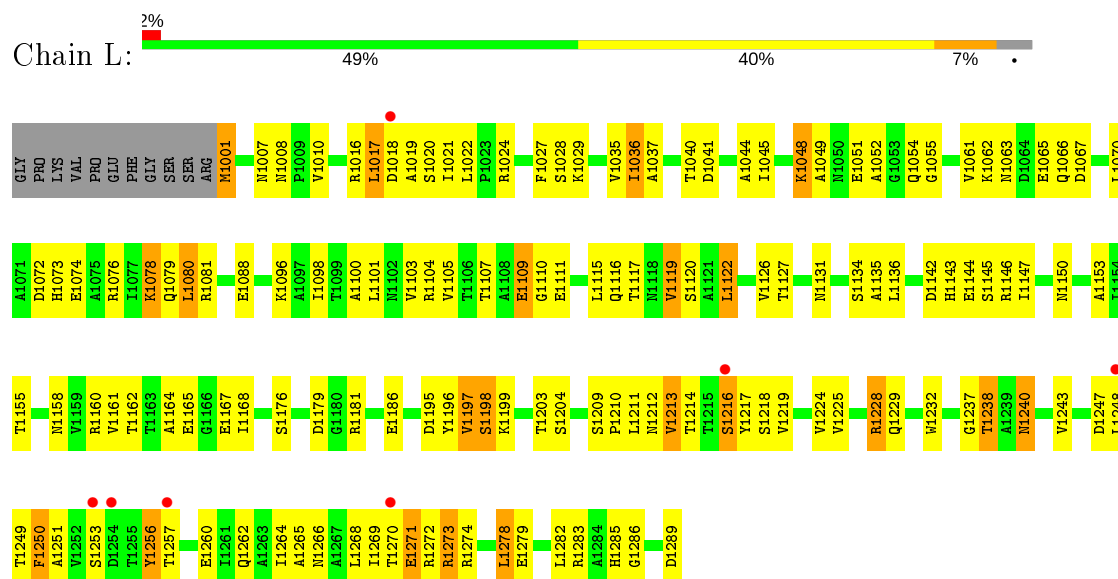
• Molecule 1: Tail needle protein gp26



• Molecule 1: Tail needle protein gp26



• Molecule 1: Tail needle protein gp26



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	38.69Å 147.93Å 151.01Å 87.94° 90.05° 89.95°	Depositor
Resolution (Å)	14.99 – 2.70 29.92 – 2.70	Depositor EDS
% Data completeness (in resolution range)	91.5 (14.99-2.70) 73.2 (29.92-2.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.81 (at 2.68Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.2_1309)	Depositor
R, R_{free}	0.229 , 0.268 0.229 , 0.267	Depositor DCC
R_{free} test set	1766 reflections (2.39%)	wwPDB-VP
Wilson B-factor (Å ²)	32.9	Xtriage
Anisotropy	0.997	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.26 , 21.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.000 for h,-l,k 0.000 for h,l,-k 0.430 for h,-k,-l 0.000 for -h,k,-l 0.000 for -h,-k,l 0.000 for -h,-l,-k 0.000 for -h,l,k	Xtriage
Reported twinning fraction	0.480 for h,-k,-l	Depositor
Outliers	0 of 73933 reflections	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	26552	wwPDB-VP
Average B, all atoms (Å ²)	65.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 18.51% 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 [i](#)

5.1 Standard geometry [i](#)

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

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	0.27	0/2166	0.48	0/2950
1	B	0.26	0/2166	0.51	1/2950 (0.0%)
1	C	0.30	0/2166	0.55	1/2950 (0.0%)
1	D	0.27	0/2166	0.48	0/2950
1	E	0.26	0/2166	0.48	0/2950
1	F	0.26	0/2166	0.52	1/2950 (0.0%)
1	G	0.28	0/2166	0.51	0/2950
1	H	0.30	0/2166	0.52	0/2950
1	I	0.29	0/2166	0.49	0/2950
1	J	0.27	0/2166	0.51	0/2950
1	K	0.27	0/2166	0.52	0/2950
1	L	0.28	0/2166	0.52	0/2950
All	All	0.28	0/25992	0.51	3/35400 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	1248	LEU	CA-CB-CG	5.45	127.83	115.30
1	F	1248	LEU	CA-CB-CG	5.08	126.99	115.30
1	C	1083	ASP	CB-CG-OD1	5.03	122.83	118.30

There are no chirality outliers.

There are no planarity outliers.

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	A	2148	0	2135	133	1
1	B	2148	0	2135	136	0
1	C	2148	0	2135	129	0
1	D	2148	0	2135	137	0
1	E	2148	0	2135	123	1
1	F	2148	0	2135	128	1
1	G	2148	0	2135	126	0
1	H	2148	0	2135	155	0
1	I	2148	0	2134	147	1
1	J	2148	0	2134	165	2
1	K	2148	0	2135	163	1
1	L	2148	0	2134	149	0
2	A	1	0	0	0	0
2	E	1	0	0	0	0
2	G	1	0	0	0	0
2	J	1	0	0	0	0
3	A	2	0	0	3	0
3	D	1	0	0	0	0
3	E	1	0	0	0	0
3	H	1	0	0	2	0
3	I	1	0	0	4	0
3	K	1	0	0	1	0
3	L	1	0	0	2	0
4	A	70	0	0	27	0
4	B	81	0	0	27	2
4	C	54	0	0	17	0
4	D	68	0	0	29	2
4	E	61	0	0	24	0
4	F	70	0	0	21	2
4	G	58	0	0	27	0
4	H	54	0	0	23	0
4	I	61	0	0	30	0
4	J	67	0	0	18	0
4	K	53	0	0	20	0
4	L	67	0	0	25	1
All	All	26552	0	25617	1272	7

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 25.

The worst 5 of 1272 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:J:1265:ALA:HB1	1:K:1250:PHE:CE2	1.41	1.49
1:H:1087:HIS:HD2	4:H:1454:HOH:O	1.08	1.35
1:K:1250:PHE:CE1	1:K:1264:ILE:HA	1.62	1.35
1:J:1217:TYR:CD2	1:J:1225:VAL:HB	1.72	1.23
1:K:1250:PHE:CD1	1:K:1264:ILE:HG12	1.73	1.22

The worst 5 of 7 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:J:1244:PHE:O	4:F:1345:HOH:O[1_455]	1.98	0.22
1:J:1244:PHE:N	4:F:1345:HOH:O[1_455]	2.04	0.16
4:B:1346:HOH:O	4:D:1419:HOH:O[1_446]	2.06	0.14
1:F:1234:ALA:O	4:L:1408:HOH:O[1_655]	2.16	0.04
4:B:1338:HOH:O	4:D:1465:HOH:O[1_446]	2.18	0.02

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	287/300 (96%)	276 (96%)	11 (4%)	0	100	100
1	B	287/300 (96%)	279 (97%)	8 (3%)	0	100	100
1	C	287/300 (96%)	276 (96%)	11 (4%)	0	100	100
1	D	287/300 (96%)	281 (98%)	6 (2%)	0	100	100
1	E	287/300 (96%)	277 (96%)	10 (4%)	0	100	100
1	F	287/300 (96%)	274 (96%)	13 (4%)	0	100	100
1	G	287/300 (96%)	281 (98%)	6 (2%)	0	100	100
1	H	287/300 (96%)	278 (97%)	9 (3%)	0	100	100
1	I	287/300 (96%)	277 (96%)	10 (4%)	0	100	100
1	J	287/300 (96%)	275 (96%)	12 (4%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	K	287/300 (96%)	273 (95%)	14 (5%)	0	100	100
1	L	287/300 (96%)	270 (94%)	16 (6%)	1 (0%)	41	66
All	All	3444/3600 (96%)	3317 (96%)	126 (4%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	L	1019	ALA

5.3.2 Protein sidechains ⓘ

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	229/238 (96%)	193 (84%)	36 (16%)	2	6
1	B	229/238 (96%)	191 (83%)	38 (17%)	2	5
1	C	229/238 (96%)	189 (82%)	40 (18%)	2	4
1	D	229/238 (96%)	196 (86%)	33 (14%)	3	8
1	E	229/238 (96%)	194 (85%)	35 (15%)	2	7
1	F	229/238 (96%)	198 (86%)	31 (14%)	4	9
1	G	229/238 (96%)	193 (84%)	36 (16%)	2	6
1	H	229/238 (96%)	183 (80%)	46 (20%)	1	3
1	I	229/238 (96%)	194 (85%)	35 (15%)	2	7
1	J	229/238 (96%)	188 (82%)	41 (18%)	2	4
1	K	229/238 (96%)	182 (80%)	47 (20%)	1	3
1	L	229/238 (96%)	183 (80%)	46 (20%)	1	3
All	All	2748/2856 (96%)	2284 (83%)	464 (17%)	2	5

5 of 464 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	1257	THR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	H	1064	ASP
1	L	1070	LEU
1	G	1021	ILE
1	G	1197	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 26 such sidechains are listed below:

Mol	Chain	Res	Type
1	G	1038	GLN
1	G	1193	GLN
1	L	1025	ASN
1	G	1054	GLN
1	G	1087	HIS

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 12 ligands modelled in this entry, 12 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 [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 ⓘ

6.1 Protein, DNA and RNA chains ⓘ

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	289/300 (96%)	-0.25	1 (0%) 94 95	29, 61, 89, 114	0
1	B	289/300 (96%)	-0.23	4 (1%) 75 77	29, 59, 80, 96	0
1	C	289/300 (96%)	-0.18	1 (0%) 94 95	26, 64, 92, 122	0
1	D	289/300 (96%)	-0.08	3 (1%) 82 83	26, 73, 101, 118	0
1	E	289/300 (96%)	-0.13	2 (0%) 87 89	28, 64, 93, 115	0
1	F	289/300 (96%)	-0.15	5 (1%) 70 72	28, 60, 98, 124	0
1	G	289/300 (96%)	-0.08	6 (2%) 63 65	33, 67, 100, 141	0
1	H	289/300 (96%)	-0.13	2 (0%) 87 89	32, 68, 99, 114	0
1	I	289/300 (96%)	0.01	5 (1%) 70 72	33, 65, 97, 125	0
1	J	289/300 (96%)	-0.03	8 (2%) 53 54	36, 67, 97, 116	0
1	K	289/300 (96%)	-0.08	8 (2%) 53 54	37, 66, 102, 123	0
1	L	289/300 (96%)	-0.03	7 (2%) 59 60	39, 65, 99, 119	0
All	All	3468/3600 (96%)	-0.12	52 (1%) 73 76	26, 65, 97, 141	0

The worst 5 of 52 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	I	1257	THR	10.9
1	G	1024	ARG	6.0
1	L	1257	THR	5.5
1	L	1248	LEU	4.7
1	J	1251	ALA	4.5

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

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

6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
3	CL	I	1301	1/1	0.60	1.00	171,171,171,171	0
2	CA	J	1301	1/1	0.62	0.22	99,99,99,99	0
3	CL	A	1302	1/1	0.89	0.07	50,50,50,50	0
3	CL	E	1302	1/1	0.91	0.10	61,61,61,61	0
3	CL	D	1301	1/1	0.94	0.09	58,58,58,58	0
3	CL	A	1303	1/1	0.94	0.05	60,60,60,60	0
3	CL	H	1301	1/1	0.94	0.06	38,38,38,38	0
2	CA	E	1301	1/1	0.95	0.09	54,54,54,54	0
3	CL	L	1301	1/1	0.96	0.10	49,49,49,49	0
3	CL	K	1301	1/1	0.97	0.05	52,52,52,52	0
2	CA	A	1301	1/1	0.97	0.12	79,79,79,79	0
2	CA	G	1301	1/1	0.98	0.07	36,36,36,36	0

6.5 Other polymers [i](#)

There are no such residues in this entry.