



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2022 – 10:04 AM EST

PDB ID : 7RFO
Title : SeMet Tailspike protein 4 (TSP4) phage CBA120, residues 1-335, obtained in the presence of LiSO₄
Authors : Chao, K.; Shang, X.; Grenfield, J.; Linden, S.B.; Nelson, D.C.; Herzberg, O.
Deposited on : 2021-07-14
Resolution : 3.02 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
Mogul : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.26
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) : 2.26

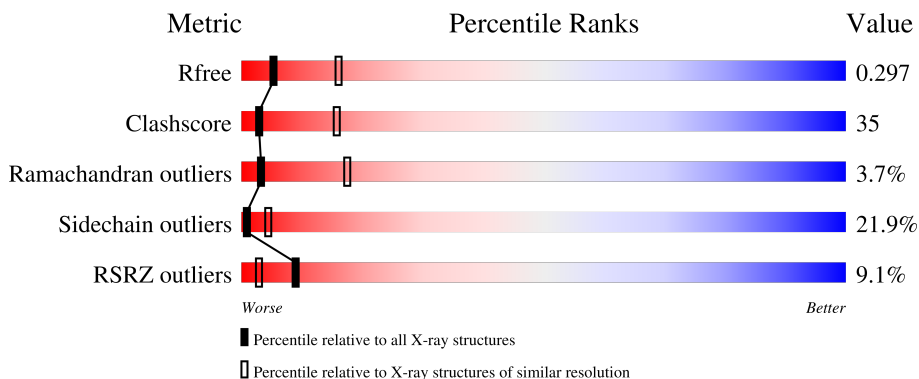
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.02 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	2399 (3.04-3.00)
Clashscore	141614	2734 (3.04-3.00)
Ramachandran outliers	138981	2640 (3.04-3.00)
Sidechain outliers	138945	2643 (3.04-3.00)
RSRZ outliers	127900	2287 (3.04-3.00)

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	341	
1	B	341	
1	C	341	

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 7077 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 Tailspike protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	329	2327	1445	388	488	1	5	0	0	0
1	B	330	2387	1488	388	505	1	5	0	1	0
1	C	324	2363	1475	383	498	1	6	0	0	0

There are 27 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	12	MSE	LEU	engineered mutation	UNP G3M192
A	31	MSE	ILE	engineered mutation	UNP G3M192
A	145	MSE	LEU	engineered mutation	UNP G3M192
A	336	HIS	-	expression tag	UNP G3M192
A	337	HIS	-	expression tag	UNP G3M192
A	338	HIS	-	expression tag	UNP G3M192
A	339	HIS	-	expression tag	UNP G3M192
A	340	HIS	-	expression tag	UNP G3M192
A	341	HIS	-	expression tag	UNP G3M192
B	12	MSE	LEU	engineered mutation	UNP G3M192
B	31	MSE	ILE	engineered mutation	UNP G3M192
B	145	MSE	LEU	engineered mutation	UNP G3M192
B	336	HIS	-	expression tag	UNP G3M192
B	337	HIS	-	expression tag	UNP G3M192
B	338	HIS	-	expression tag	UNP G3M192
B	339	HIS	-	expression tag	UNP G3M192
B	340	HIS	-	expression tag	UNP G3M192
B	341	HIS	-	expression tag	UNP G3M192
C	12	MSE	LEU	engineered mutation	UNP G3M192
C	31	MSE	ILE	engineered mutation	UNP G3M192
C	145	MSE	LEU	engineered mutation	UNP G3M192
C	336	HIS	-	expression tag	UNP G3M192
C	337	HIS	-	expression tag	UNP G3M192

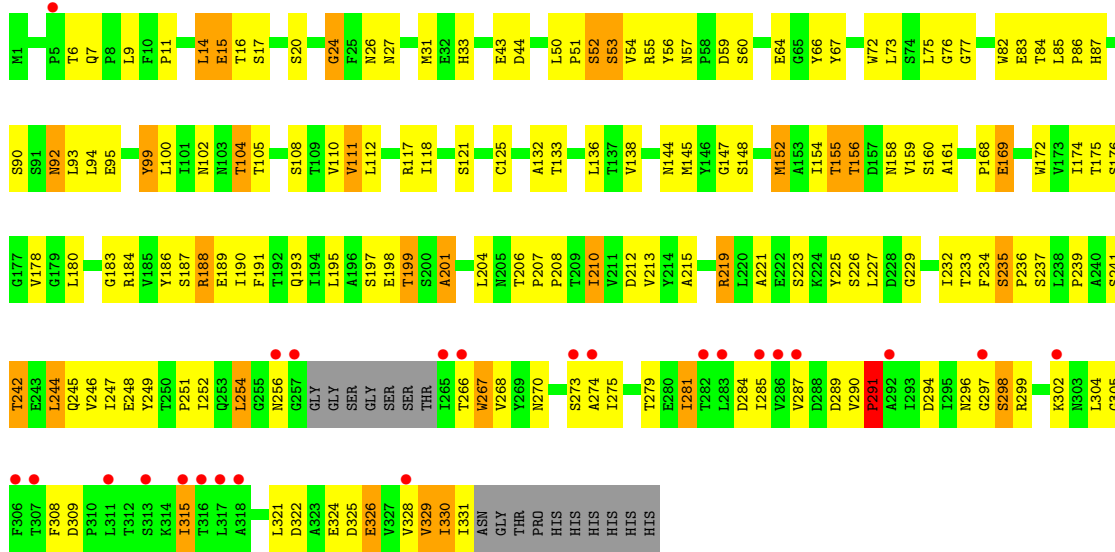
Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	338	HIS	-	expression tag	UNP G3M192
C	339	HIS	-	expression tag	UNP G3M192
C	340	HIS	-	expression tag	UNP G3M192
C	341	HIS	-	expression tag	UNP G3M192

HIS
HIS
HIS
HIS

● Molecule 1: Tailspike protein



4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α , β , γ	67.90Å 67.90Å 607.20Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 3.02 48.01 – 3.02	Depositor EDS
% Data completeness (in resolution range)	94.4 (30.00-3.02) 94.6 (48.01-3.02)	Depositor EDS
R_{merge}	0.23	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.09 (at 3.01Å)	Xtrriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.225 , 0.296 0.237 , 0.297	Depositor DCC
R_{free} test set	1370 reflections (4.86%)	wwPDB-VP
Wilson B-factor (Å ²)	118.2	Xtrriage
Anisotropy	0.224	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 115.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.28$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	7077	wwPDB-VP
Average B, all atoms (Å ²)	141.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.72% 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](#)

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	0.75	0/2374	0.95	3/3250 (0.1%)
1	B	0.75	0/2436	0.96	2/3332 (0.1%)
1	C	0.75	0/2410	0.97	1/3296 (0.0%)
All	All	0.75	0/7220	0.96	6/9878 (0.1%)

There are no bond length outliers.

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	92	ASN	CB-CA-C	6.18	122.76	110.40
1	B	12	MSE	CG-SE-CE	6.16	112.46	98.90
1	A	39	MSE	CG-SE-CE	5.91	111.90	98.90
1	A	12	MSE	CG-SE-CE	5.80	111.67	98.90
1	A	256	ASN	CB-CA-C	5.64	121.67	110.40

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	2327	0	2070	188	0
1	B	2387	0	2163	176	0
1	C	2363	0	2180	173	0
All	All	7077	0	6413	468	0

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 35.

The worst 5 of 468 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:A:152:MSE:HE1	1:A:252:ILE:HG21	1.21	1.19
1:A:287:VAL:HG23	1:A:335:PRO:CB	1.74	1.17
1:B:134:TYR:OH	1:B:218:LYS:HD2	1.44	1.15
1:A:70:GLY:HA3	1:A:73:LEU:HD11	1.24	1.15
1:A:152:MSE:CE	1:A:252:ILE:HG21	1.84	1.07

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	A	325/341 (95%)	274 (84%)	40 (12%)	11 (3%)	3	19
1	B	327/341 (96%)	288 (88%)	29 (9%)	10 (3%)	4	21
1	C	320/341 (94%)	263 (82%)	42 (13%)	15 (5%)	2	13
All	All	972/1023 (95%)	825 (85%)	111 (11%)	36 (4%)	3	17

5 of 36 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	83	GLU
1	A	298	SER
1	B	169	GLU
1	C	188	ARG
1	C	198	GLU

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	235/280 (84%)	164 (70%)	71 (30%)	0	1
1	B	249/280 (89%)	202 (81%)	47 (19%)	1	7
1	C	252/280 (90%)	209 (83%)	43 (17%)	2	9
All	All	736/840 (88%)	575 (78%)	161 (22%)	1	4

5 of 161 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	307	THR
1	C	237	SER
1	C	14	LEU
1	C	133	THR
1	C	275	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	270	ASN
1	B	300	GLN
1	C	158	ASN
1	B	143	ASN
1	B	144	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 monosaccharides 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	324/341 (95%)	0.40	33 (10%) 6 2	84, 132, 234, 313	0
1	B	325/341 (95%)	0.25	31 (9%) 8 2	81, 120, 255, 288	0
1	C	318/341 (93%)	0.38	24 (7%) 14 4	83, 118, 224, 266	0
All	All	967/1023 (94%)	0.34	88 (9%) 9 3	81, 122, 239, 313	0

The worst 5 of 88 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	257	GLY	11.0
1	C	256	ASN	9.3
1	C	318	ALA	7.9
1	B	257	GLY	7.6
1	B	292	ALA	7.1

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 monosaccharides 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.