wwPDB X-ray Structure Validation Summary Report (i)

Jun 19, 2024 – 11:50 AM EDT

PDB ID	:	4OJ6
Title	:	Crystal Structure of a Putative Tailspike Protein (TSP1, orf210) from Escherichia coli O157:H7 Bacteriohage CBA120; Se-Met Protein
Authors	:	Chen, C.; Herzberg, O.
Deposited on	:	2014-01-20
Resolution	:	1.80 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 (i) 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 (i)) were used in the production of this report:

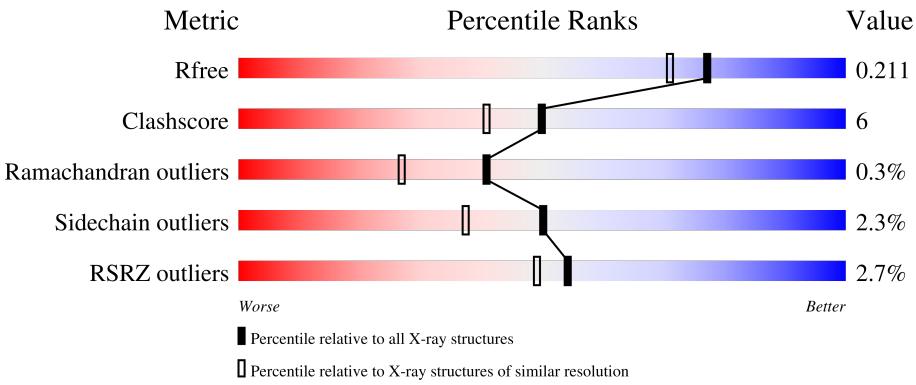
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 1.80 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	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)
RSRZ outliers	127900	5850 (1.80-1.80)

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 <=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	776	2% 87% 10% • •
1	B	776	3% 87% 10% • •
1	C	776	2% 89% 7% • •

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 35715 atoms, of which 16487 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	H	N	O	S	Se			
1	A	758	11158	3563	5517	944	1115	9	10	8	0	0
1	B	756	11112	3550	5488	941	1115	8	10	0	0	0
1	C	753	11082	3532	5482	940	1110	9	9	3	0	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	771	HIS	-	EXPRESSION TAG	UNP G3M189
A	772	HIS	-	EXPRESSION TAG	UNP G3M189
A	773	HIS	-	EXPRESSION TAG	UNP G3M189
A	774	HIS	-	EXPRESSION TAG	UNP G3M189
A	775	HIS	-	EXPRESSION TAG	UNP G3M189
A	776	HIS	-	EXPRESSION TAG	UNP G3M189
B	771	HIS	-	EXPRESSION TAG	UNP G3M189
B	772	HIS	-	EXPRESSION TAG	UNP G3M189
B	773	HIS	-	EXPRESSION TAG	UNP G3M189
B	774	HIS	-	EXPRESSION TAG	UNP G3M189
B	775	HIS	-	EXPRESSION TAG	UNP G3M189
B	776	HIS	-	EXPRESSION TAG	UNP G3M189
C	771	HIS	-	EXPRESSION TAG	UNP G3M189
C	772	HIS	-	EXPRESSION TAG	UNP G3M189
C	773	HIS	-	EXPRESSION TAG	UNP G3M189
C	774	HIS	-	EXPRESSION TAG	UNP G3M189
C	775	HIS	-	EXPRESSION TAG	UNP G3M189
C	776	HIS	-	EXPRESSION TAG	UNP G3M189

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Zn		
2	B	1	1	1	0	1

- Molecule 3 is water.

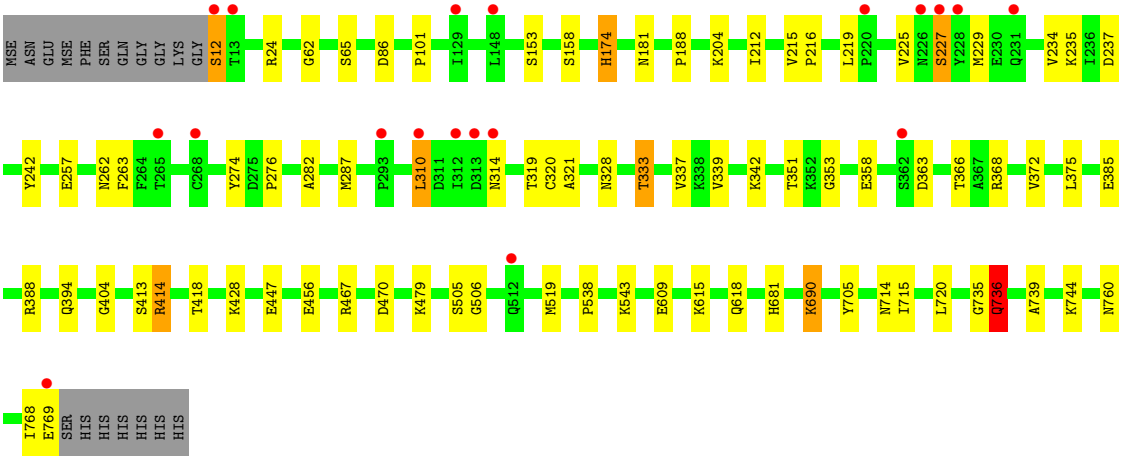
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	707	707	707	0	0
			Total	O		
3	B	828	828	828	0	1
			Total	O		
3	C	827	827	827	0	0

3 Residue-property plots (i)

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.

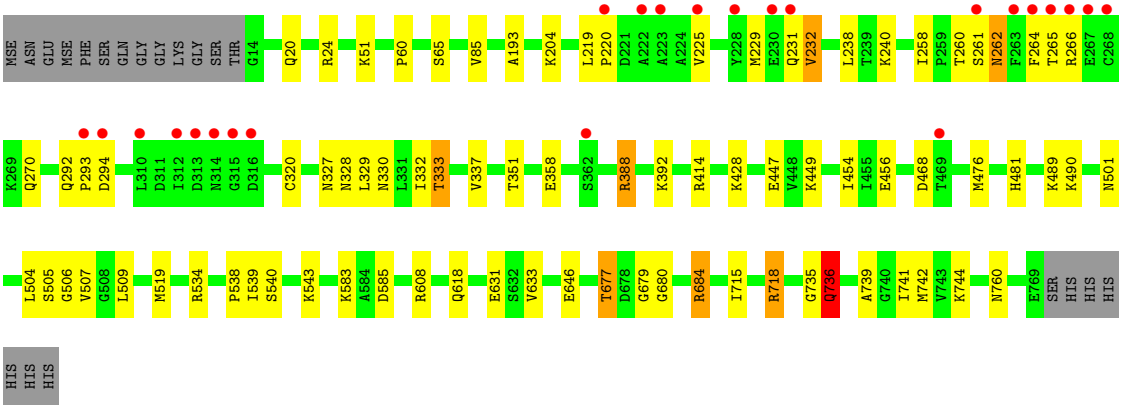
• Molecule 1: Tailspike protein

Chain A:



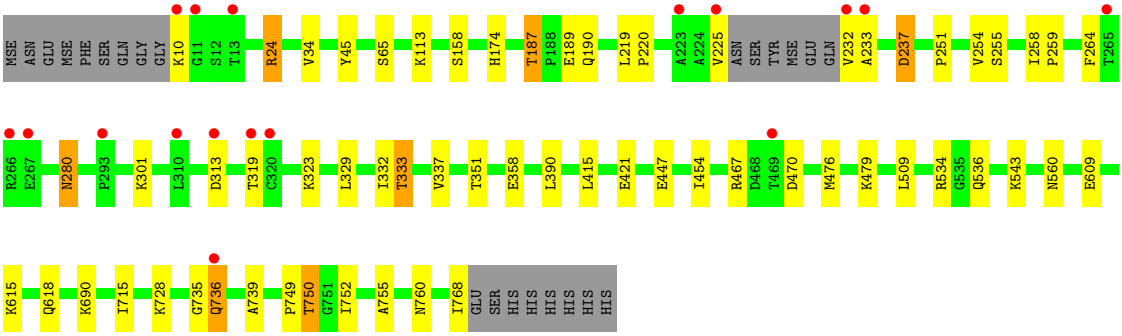
• Molecule 1: Tailspike protein

Chain B:



• Molecule 1: Tailspike protein

Chain C:


MSE
ASN
GLU
MSE
PHE
SER
GLN
GLY
GLY
K10
G11
S12
T13
R24
V34
Y45
S65
K113
S158
H174
T187
P188
E189
Q190
L219
P220
A223
A224
V225
ASN
SER
TYR
MSE
GLU
GLN
V232
A233
D237
P251
V254
S255
I258
P259
F264
T265
R266
E267
N280
P293
K301
L310
D313
T319
C320
K323
L329
I332
T333
V337
T351
E358
L390
L415
E421
E447
I454
R467
D468
T469
D470
M476
K479
L509
R534
G535
Q536
K543
N560
E609
K615
Q618
K690
I715
K728
G735
Q736
A739
P749
T750
G751
I752
A755
N760
I768
GLU
SER
HIS
HIS
HIS
HIS
HIS
HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	123.30Å 153.09Å 171.44Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.85 – 1.80 19.85 – 1.80	Depositor EDS
% Data completeness (in resolution range)	99.7 (19.85-1.80) 99.7 (19.85-1.80)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.81 (at 1.80Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, R_{free}	0.187 , 0.209 0.190 , 0.211	Depositor DCC
R_{free} test set	14894 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	22.0	Xtriage
Anisotropy	0.192	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 50.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	35715	wwPDB-VP
Average B, all atoms (Å^2)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.39% 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, 0.333 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: ZN

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	0.35	0/5735	0.55	0/7784
1	B	0.40	0/5718	0.60	3/7761 (0.0%)
1	C	0.40	0/5692	0.57	0/7726
All	All	0.38	0/17145	0.57	3/23271 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	718	ARG	NE-CZ-NH1	6.75	123.67	120.30
1	B	718	ARG	NE-CZ-NH2	-6.44	117.08	120.30
1	B	684	ARG	NE-CZ-NH2	-6.25	117.17	120.30

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	5641	5517	5576	66	0
1	B	5624	5488	5547	77	1
1	C	5600	5482	5536	61	0
2	B	1	0	0	0	0
3	A	707	0	0	30	1

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	828	0	0	39	1
3	C	827	0	0	34	1
All	All	19228	16487	16659	193	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 6.

The worst 5 of 193 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:B:264:PHE:CE1	3:B:1572:HOH:O	1.77	1.27
1:C:313:ASP:HA	3:C:1595:HOH:O	1.30	1.25
1:B:330:ASN:HB3	3:B:1637:HOH:O	1.09	1.23
1:B:320:CYS:SG	3:B:1666:HOH:O	2.00	1.19
1:C:237:ASP:HB3	3:C:1521:HOH:O	1.42	1.15

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)
3:A:1507:HOH:O	3:C:1564:HOH:O[2_585]	2.13	0.07
1:B:684:ARG:HH22	3:B:1090:HOH:O[2_585]	1.57	0.03

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	756/776 (97%)	725 (96%)	29 (4%)	2 (0%)	41 27
1	B	754/776 (97%)	726 (96%)	25 (3%)	3 (0%)	34 21
1	C	749/776 (96%)	724 (97%)	23 (3%)	2 (0%)	41 27

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	2259/2328 (97%)	2175 (96%)	77 (3%)	7 (0%)	41 27

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	736	GLN
1	B	736	GLN
1	C	736	GLN
1	A	735	GLY
1	B	735	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	A	610/625 (98%)	595 (98%)	15 (2%)	47 34
1	B	608/625 (97%)	593 (98%)	15 (2%)	47 34
1	C	606/625 (97%)	594 (98%)	12 (2%)	55 44
All	All	1824/1875 (97%)	1782 (98%)	42 (2%)	50 37

5 of 42 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	736	GLN
1	C	237	ASP
1	B	741	ILE
1	C	65	SER
1	C	319	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	681	HIS
1	B	328	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	262	ASN
1	A	314	ASN
1	A	174	HIS

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)

Of 1 ligands modelled in this entry, 1 is 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 (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	748/776 (96%)	-0.16	19 (2%)	57	52	18, 31, 50, 70	2 (0%)
1	B	746/776 (96%)	-0.29	24 (3%)	47	41	15, 25, 49, 73	0
1	C	744/776 (95%)	-0.30	17 (2%)	60	56	18, 25, 42, 76	0
All	All	2238/2328 (96%)	-0.25	60 (2%)	54	49	15, 27, 48, 76	2 (0%)

The worst 5 of 60 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	12	SER	8.0
1	C	232	VAL	6.3
1	A	13	THR	6.1
1	B	264	PHE	5.8
1	B	310	LEU	5.6

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 monosaccharides 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(Å^2)	Q<0.9
2	ZN	B	800[A]	1/1	0.99	0.04	26,26,26,26	1

6.5 Other polymers (i)

There are no such residues in this entry.