



# wwPDB X-ray Structure Validation Summary Report

Sep 30, 2014 – 01:18 AM EDT

PDB ID : 4OJL  
Title : Crystal Structure of Putative Tailspike Protein (TSP1, orf210) from Escherichia coli O157:H7 Bacteriophage CBA120 in Complex with Glucose  
Authors : Chen, C.; Herzberg, O.  
Deposited on : 2014-01-21  
Resolution : 2.00 Å(reported)

This is a wwPDB validation summary report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

---

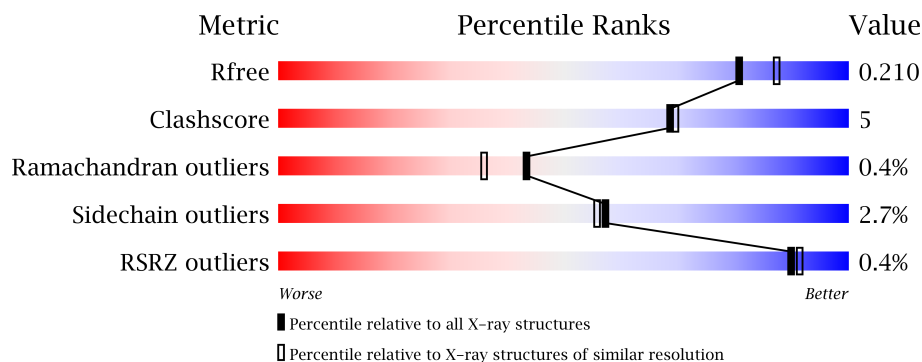
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.16 November 2013
Xtriage (Phenix)	:	dev-1439
EDS	:	stable23828
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable23828

# 1 Overall quality at a glance

The reported resolution of this entry is 2.00 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))
$R_{free}$	66092	4888 (2.00-2.00)
Clashscore	79885	6188 (2.00-2.00)
Ramachandran outliers	78287	6102 (2.00-2.00)
Sidechain outliers	78261	6100 (2.00-2.00)
RSRZ outliers	66119	4890 (2.00-2.00)

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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	776	
1	B	776	
1	C	776	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
2	BGC	A	800	-	X
2	BGC	C	801	-	X

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 19521 atoms, of which 0 are hydrogen and 0 are deuterium.

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.

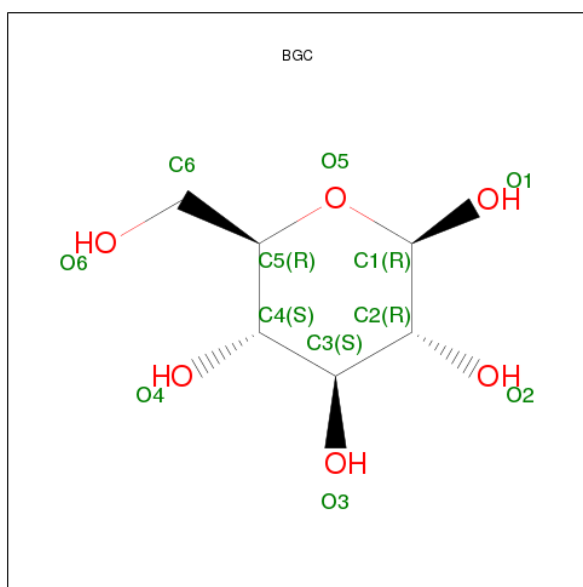
- Molecule 1 is a protein called Tailspike protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	758	Total	C	N	O	S	1	0	0
			5673	3577	954	1123	19			
1	B	756	Total	C	N	O	S	0	0	0
			5653	3568	947	1119	19			
1	C	759	Total	C	N	O	S	1	0	0
			5679	3580	958	1122	19			

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 SUGAR (BETA-D-GLUCOSE) (three-letter code: BGC) (formula: C<sub>6</sub>H<sub>12</sub>O<sub>6</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			12	6	6		
2	B	1	Total	C	O	0	0
			12	6	6		
2	C	1	Total	C	O	0	0
			12	6	6		
2	C	1	Total	C	O	0	0
			12	6	6		

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Zn	0	1
			1	1		

- Molecule 4 is water.

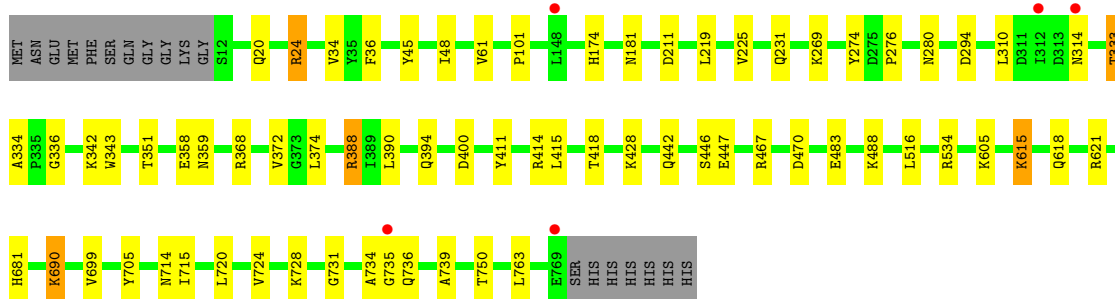
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	756	Total	O	0	0
			756	756		
4	B	841	Total	O	0	0
			841	841		
4	C	870	Total	O	0	0
			870	870		

### 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 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.

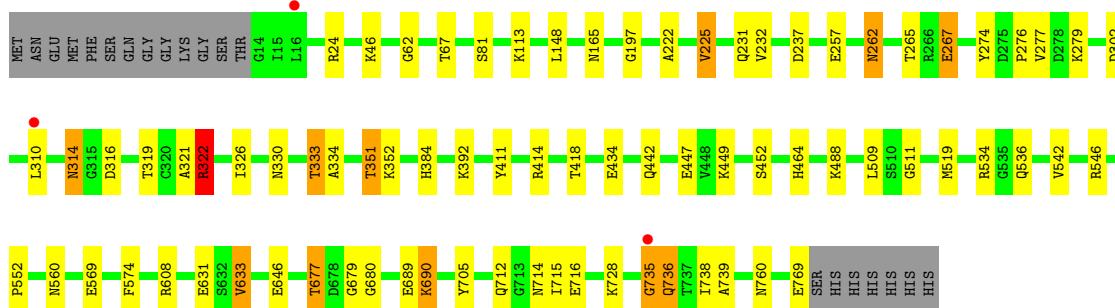
#### • Molecule 1: Tailspike protein

Chain A: 



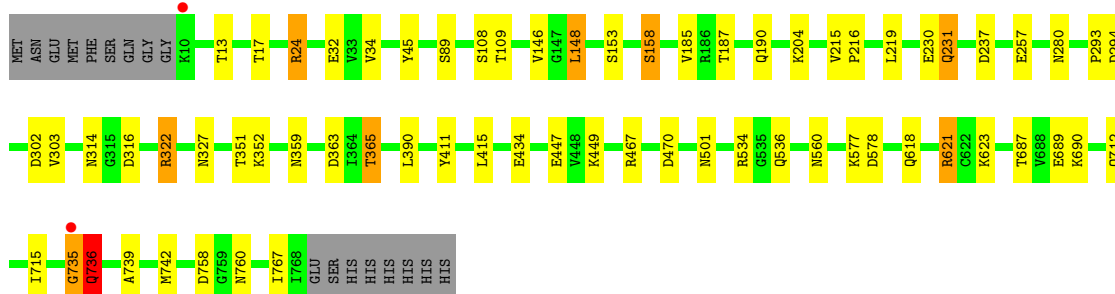
#### • Molecule 1: Tailspike protein

Chain B: 



#### • Molecule 1: Tailspike protein

Chain C: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	123.45Å 152.44Å 171.85Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	43.80 – 2.00 74.85 – 1.95	Depositor EDS
% Data completeness (in resolution range)	96.2 (43.80-2.00) 94.2 (74.85-1.95)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.65 (at 1.95Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.1069)	Depositor
R, $R_{free}$	0.178 , 0.211 0.175 , 0.210	Depositor DCC
$R_{free}$ test set	10573 reflections (5.04%)	DCC
Wilson B-factor (Å <sup>2</sup> )	20.0	Xtriage
Anisotropy	0.315	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 41.1	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 221545 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	19521	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.50% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

## 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, BGC

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	0.36	0/5777	0.66	0/7854
1	B	0.39	0/5757	0.70	2/7828 (0.0%)
1	C	0.39	0/5783	0.68	0/7861
All	All	0.38	0/17317	0.68	2/23543 (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	A	0	1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	322	ARG	NE-CZ-NH1	6.36	123.48	120.30
1	B	322	ARG	NE-CZ-NH2	-5.12	117.74	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	24	ARG	Sidechain

## 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	5673	0	5631	46	0
1	B	5653	0	5600	61	0
1	C	5679	0	5641	58	0
2	A	12	0	12	1	0
2	B	12	0	12	0	0
2	C	24	0	24	0	0
3	A	1	0	0	0	0
4	A	756	0	0	20	1
4	B	841	0	0	27	3
4	C	870	0	0	38	1
All	All	19521	0	16920	162	3

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 5.

The worst 5 of 162 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:C:280:ASN:ND2	4:C:1632:HOH:O	1.82	1.09
1:C:690:LYS:NZ	4:C:1739:HOH:O	1.86	1.07
4:A:1615:HOH:O	1:B:690:LYS:HE3	1.58	1.03
1:B:534:ARG:NH1	4:B:1697:HOH:O	1.90	1.02
1:C:314:ASN:ND2	4:C:1701:HOH:O	1.91	1.02

All (3) 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	Distance(Å)	Clash(Å)
4:B:1303:HOH:O	4:B:1341:HOH:O[2_584]	2.12	0.08
4:A:1626:HOH:O	4:B:1699:HOH:O[4_585]	2.16	0.04
4:B:1350:HOH:O	4:C:1648:HOH:O[2_684]	2.19	0.01



## 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	756/776 (97%)	727 (96%)	26 (3%)	3 (0%)	43	36
1	B	754/776 (97%)	726 (96%)	25 (3%)	3 (0%)	43	36
1	C	757/776 (98%)	728 (96%)	27 (4%)	2 (0%)	50	44
All	All	2267/2328 (97%)	2181 (96%)	78 (3%)	8 (0%)	43	36

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	736	GLN
1	B	736	GLN
1	C	735	GLY
1	C	736	GLN
1	A	735	GLY

### 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	620/637 (97%)	605 (98%)	15 (2%)	61	61
1	B	615/637 (96%)	597 (97%)	18 (3%)	55	52
1	C	620/637 (97%)	603 (97%)	17 (3%)	57	56
All	All	1855/1911 (97%)	1805 (97%)	50 (3%)	57	56

5 of 50 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	326	ILE
1	B	509	LEU
1	C	411	TYR
1	B	333	THR
1	B	411	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	681	HIS
1	C	327	ASN
1	B	328	ASN
1	A	501	ASN
1	B	314	ASN

### 5.3.3 RNA ⓘ

There are no RNA chains 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 5 ligands modelled in this entry, 1 is monoatomic - leaving 4 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	BGC	A	800	-	12,12,12	0.44	0	17,17,17	0.96	0
2	BGC	B	800	-	12,12,12	0.46	0	17,17,17	0.97	0
2	BGC	C	800	-	12,12,12	0.51	0	17,17,17	0.61	0
2	BGC	C	801	-	12,12,12	0.48	0	17,17,17	0.75	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	BGC	A	800	-	-	0/2/22/22	0/1/1/1
2	BGC	B	800	-	-	0/2/22/22	0/1/1/1
2	BGC	C	800	-	-	0/2/22/22	0/1/1/1
2	BGC	C	801	-	-	0/2/22/22	0/1/1/1

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.

## 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 ⓘ

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	758/776 (97%)	-0.34	5 (0%) 84 85	21, 29, 43, 63	1 (0%)
1	B	756/776 (97%)	-0.38	3 (0%) 90 91	18, 25, 39, 56	0
1	C	759/776 (97%)	-0.40	2 (0%) 91 93	18, 25, 36, 58	0
All	All	2273/2328 (97%)	-0.37	10 (0%) 90 91	18, 26, 39, 63	1 (0%)

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	735	GLY	3.9
1	A	769	GLU	3.4
1	A	148	LEU	3.2
1	B	16	LEU	2.6
1	B	735	GLY	2.4

### 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 ⓘ

There are no carbohydrates in this entry.

### 6.4 Ligands ⓘ

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	BGC	C	801	12/12	0.18	12.76	40,51,88,98	0
2	BGC	A	800	12/12	0.14	3.20	25,37,42,43	0
2	BGC	B	800	12/12	0.10	0.47	24,30,33,38	0
3	ZN	A	802[A]	1/1	0.05	-2.09	38,38,38,38	1
2	BGC	C	800	12/12	0.09	-2.21	23,30,32,32	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.