



# Full wwPDB X-ray Structure Validation Report ⓘ

May 13, 2020 – 05:29 pm BST

PDB ID : 4UXF  
Title : Crystal structure of the carboxy-terminal region of the bacteriophage T4 proximal long tail fibre protein gp34, P21 native crystal  
Authors : Granell, M.; Alvira, S.; Garcia-Doval, C.; Singh, A.K.; van Raaij, M.J.  
Deposited on : 2014-08-22  
Resolution : 2.00 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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

<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 : 2.11  
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

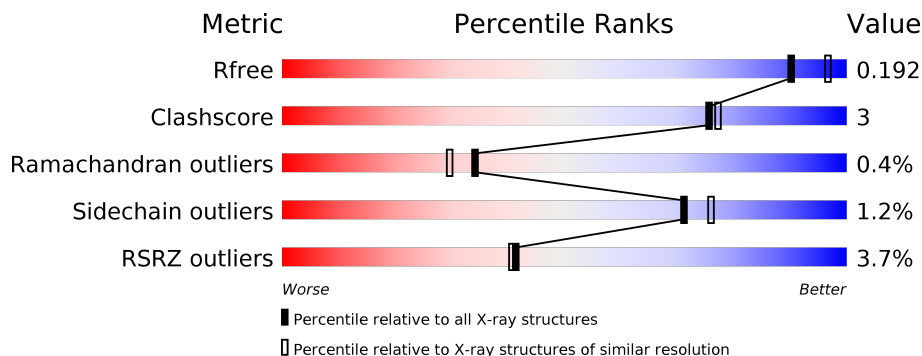
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.00 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))
$R_{free}$	130704	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (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 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	410	 5% 91% 6% •
1	B	410	 2% 90% 6% ••
1	C	410	 4% 90% 6% •

## 2 Entry composition i

There are 3 unique types of molecules in this entry. The entry contains 10873 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 LARGE TAIL FIBER PROTEIN P34.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	397	Total 3048	C 1903	N 537	O 602	S 6	0	2	0
1	B	397	Total 3059	C 1912	N 535	O 605	S 7	0	5	0
1	C	396	Total 3060	C 1915	N 537	O 600	S 8	0	7	0

There are 42 discrepancies between the modelled and reference sequences:

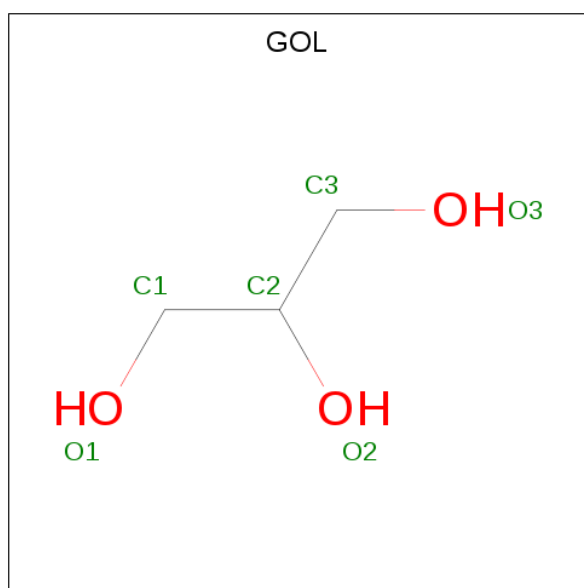
Chain	Residue	Modelled	Actual	Comment	Reference
A	880	MET	-	expression tag	UNP P18771
A	881	GLY	-	expression tag	UNP P18771
A	882	SER	-	expression tag	UNP P18771
A	883	SER	-	expression tag	UNP P18771
A	884	HIS	-	expression tag	UNP P18771
A	885	HIS	-	expression tag	UNP P18771
A	886	HIS	-	expression tag	UNP P18771
A	887	HIS	-	expression tag	UNP P18771
A	888	HIS	-	expression tag	UNP P18771
A	889	HIS	-	expression tag	UNP P18771
A	890	SER	-	expression tag	UNP P18771
A	891	GLN	-	expression tag	UNP P18771
A	892	ASP	-	expression tag	UNP P18771
A	893	PRO	-	expression tag	UNP P18771
B	880	MET	-	expression tag	UNP P18771
B	881	GLY	-	expression tag	UNP P18771
B	882	SER	-	expression tag	UNP P18771
B	883	SER	-	expression tag	UNP P18771
B	884	HIS	-	expression tag	UNP P18771
B	885	HIS	-	expression tag	UNP P18771
B	886	HIS	-	expression tag	UNP P18771
B	887	HIS	-	expression tag	UNP P18771
B	888	HIS	-	expression tag	UNP P18771

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	889	HIS	-	expression tag	UNP P18771
B	890	SER	-	expression tag	UNP P18771
B	891	GLN	-	expression tag	UNP P18771
B	892	ASP	-	expression tag	UNP P18771
B	893	PRO	-	expression tag	UNP P18771
C	880	MET	-	expression tag	UNP P18771
C	881	GLY	-	expression tag	UNP P18771
C	882	SER	-	expression tag	UNP P18771
C	883	SER	-	expression tag	UNP P18771
C	884	HIS	-	expression tag	UNP P18771
C	885	HIS	-	expression tag	UNP P18771
C	886	HIS	-	expression tag	UNP P18771
C	887	HIS	-	expression tag	UNP P18771
C	888	HIS	-	expression tag	UNP P18771
C	889	HIS	-	expression tag	UNP P18771
C	890	SER	-	expression tag	UNP P18771
C	891	GLN	-	expression tag	UNP P18771
C	892	ASP	-	expression tag	UNP P18771
C	893	PRO	-	expression tag	UNP P18771

- Molecule 2 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			6	3	3		
2	A	1	Total	C	O	0	0
			6	3	3		

Continued on next page...

*Continued from previous page...*

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	B	1	Total	C	O	0	0
			6	3	3		
2	B	1	Total	C	O	0	0
			6	3	3		
2	C	1	Total	C	O	0	0
			6	3	3		
2	C	1	Total	C	O	0	0
			6	3	3		
2	C	1	Total	C	O	0	0
			6	3	3		

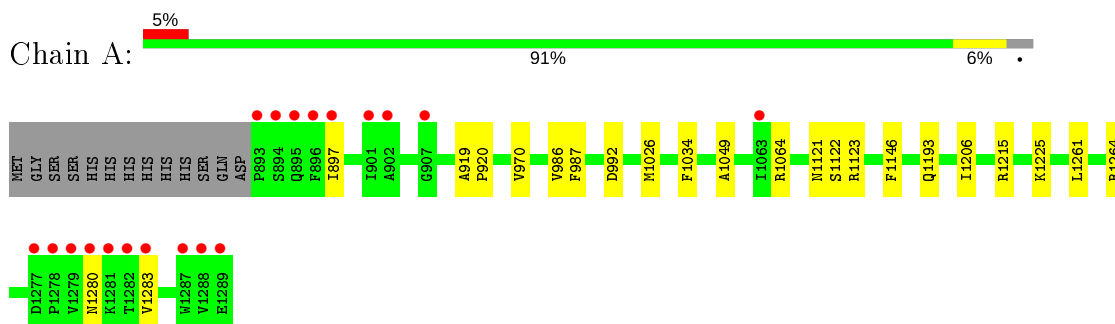
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	697	Total	O	0	0
			697	697		
3	B	578	Total	O	0	0
			578	578		
3	C	389	Total	O	0	0
			389	389		

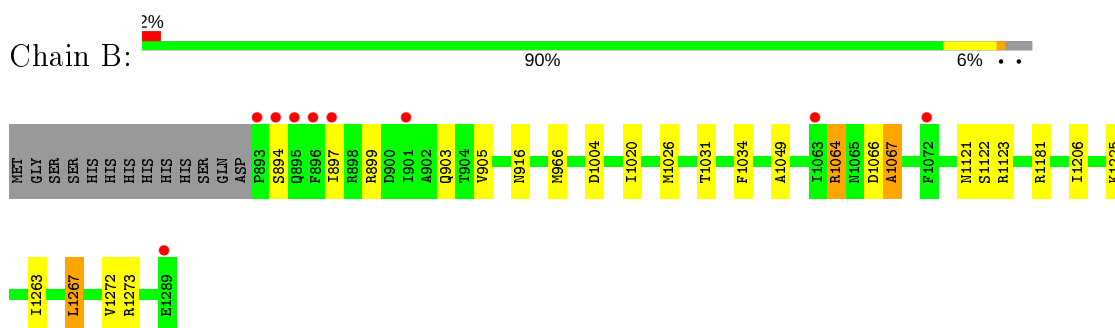
### 3 Residue-property plots [i](#)

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.

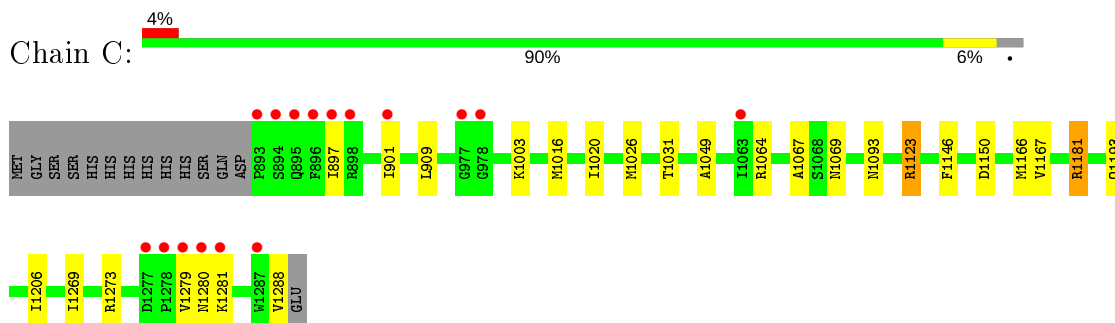
- Molecule 1: LARGE TAIL FIBER PROTEIN P34



- Molecule 1: LARGE TAIL FIBER PROTEIN P34



- Molecule 1: LARGE TAIL FIBER PROTEIN P34



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	92.97Å 76.05Å 116.83Å 90.00° 99.12° 90.00°	Depositor
Resolution (Å)	115.41 – 2.00 29.56 – 2.00	Depositor EDS
% Data completeness (in resolution range)	99.8 (115.41-2.00) 99.9 (29.56-2.00)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.90 (at 2.00Å)	Xtrriage
Refinement program	REFMAC 5.8.0131	Depositor
R, $R_{free}$	0.146 , 0.187 0.156 , 0.192	Depositor DCC
$R_{free}$ test set	2207 reflections (2.03%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	15.7	Xtrriage
Anisotropy	0.393	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 67.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	10873	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	25.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.17% of the height of the origin peak. No significant pseudotranslation is detected.*

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

<sup>2</sup>Theoretical 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

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

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	0.66	0/3122	0.79	2/4244 (0.0%)
1	B	0.67	0/3144	0.82	3/4273 (0.1%)
1	C	0.66	0/3151	0.81	4/4285 (0.1%)
All	All	0.66	0/9417	0.81	9/12802 (0.1%)

There are no bond length outliers.

All (9) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	1181	ARG	NE-CZ-NH1	6.90	123.75	120.30
1	C	1150	ASP	CB-CG-OD1	6.38	124.04	118.30
1	B	1181	ARG	NE-CZ-NH1	5.90	123.25	120.30
1	A	992	ASP	CB-CG-OD2	5.77	123.50	118.30
1	B	1004	ASP	CB-CG-OD1	5.65	123.38	118.30
1	A	1215	ARG	NE-CZ-NH1	5.48	123.04	120.30
1	C	1181	ARG	NE-CZ-NH2	-5.47	117.56	120.30
1	B	1064	ARG	NE-CZ-NH2	-5.17	117.72	120.30
1	C	1123	ARG	NE-CZ-NH1	5.08	122.84	120.30

There are no chirality outliers.

There are no planarity outliers.

### 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.













## 5.8 Polymer linkage issues

There are no chain breaks in this entry.







*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
2	GOL	C	1391	6/6	0.94	0.14	23,29,34,36	0
2	GOL	C	1389	6/6	0.95	0.14	23,27,28,29	0
2	GOL	A	1391	6/6	0.96	0.15	24,30,33,35	0
2	GOL	B	1390	6/6	0.97	0.09	20,23,26,27	0
2	GOL	C	1390	6/6	0.97	0.15	23,24,26,26	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.