



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 10:58 am GMT

PDB ID : 1K28  
Title : The Structure of the Bacteriophage T4 Cell-Puncturing Device  
Authors : Kanamaru, S.; Leiman, P.G.; Kostyuchenko, V.A.; Chipman, P.R.; Mesyanzhinov, V.V.; Arisaka, F.; Rossmann, M.G.  
Deposited on : 2001-09-26  
Resolution : 2.90 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](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/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 : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : trunk28620  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : recalc28949

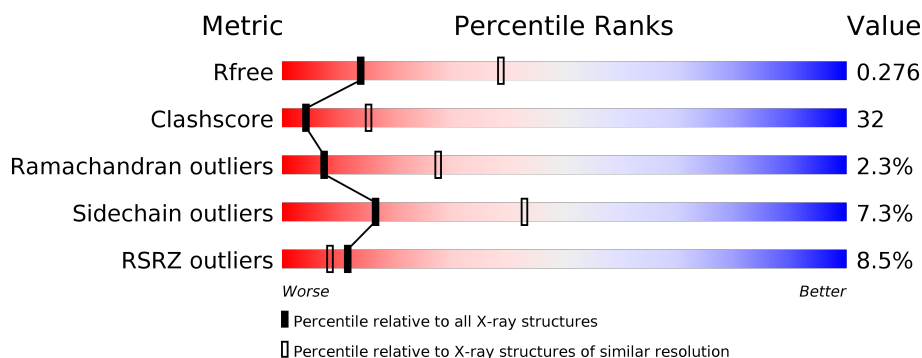
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.90 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))
$R_{free}$	100719	1586 (2.90-2.90)
Clashscore	112137	1807 (2.90-2.90)
Ramachandran outliers	110173	1768 (2.90-2.90)
Sidechain outliers	110143	1770 (2.90-2.90)
RSRZ outliers	101464	1596 (2.90-2.90)

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. 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	584	<div> <div>18%</div> <div>57%</div> <div>33%</div> <div>6%</div> <div>5%</div> </div>
2	D	391	<div> <div>18%</div> <div>38%</div> <div>49%</div> <div>6%</div> <div>7%</div> </div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 7614 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-ASSOCIATED LYSOZYME.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	557	Total	C	N	O	S	0	0	0
			4309	2672	763	853	21			

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	576	SER	-	EXPRESSION TAG	UNP P16009
A	577	VAL	-	EXPRESSION TAG	UNP P16009
A	578	ASP	-	EXPRESSION TAG	UNP P16009
A	579	HIS	-	EXPRESSION TAG	UNP P16009
A	580	HIS	-	EXPRESSION TAG	UNP P16009
A	581	HIS	-	EXPRESSION TAG	UNP P16009
A	582	HIS	-	EXPRESSION TAG	UNP P16009
A	583	HIS	-	EXPRESSION TAG	UNP P16009
A	584	HIS	-	EXPRESSION TAG	UNP P16009

- Molecule 2 is a protein called BASEPLATE STRUCTURAL PROTEIN GP27.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	D	364	Total	C	N	O	S	Se	0	0	0
			2912	1858	478	559	2	15			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	1	MSE	MET	MODIFIED RESIDUE	UNP P17172
D	3	MSE	MET	MODIFIED RESIDUE	UNP P17172
D	40	MSE	MET	MODIFIED RESIDUE	UNP P17172
D	51	MSE	MET	MODIFIED RESIDUE	UNP P17172
D	64	MSE	MET	MODIFIED RESIDUE	UNP P17172
D	135	MSE	MET	MODIFIED RESIDUE	UNP P17172
D	168	ASN	THR	SEE REMARK 999	UNP P17172

*Continued on next page...*

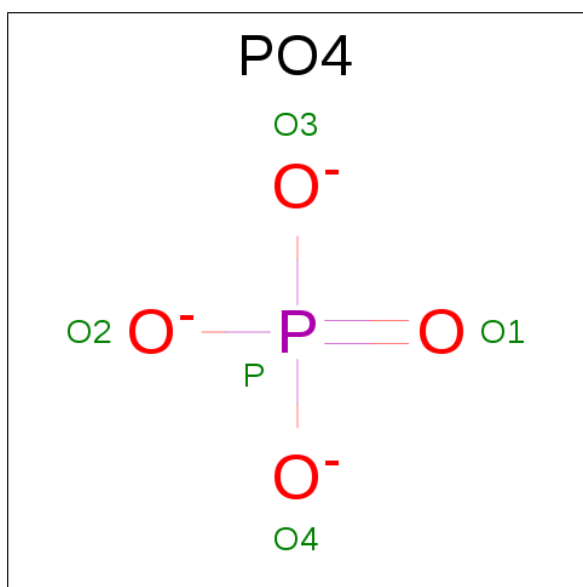
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	193	MSE	MET	MODIFIED RESIDUE	UNP P17172
D	197	MSE	MET	MODIFIED RESIDUE	UNP P17172
D	198	MSE	MET	MODIFIED RESIDUE	UNP P17172
D	202	MSE	MET	MODIFIED RESIDUE	UNP P17172
D	203	MSE	MET	MODIFIED RESIDUE	UNP P17172
D	211	MSE	MET	MODIFIED RESIDUE	UNP P17172
D	248	MSE	MET	MODIFIED RESIDUE	UNP P17172
D	262	SER	GLU	SEE REMARK 999	UNP P17172
D	265	MSE	MET	MODIFIED RESIDUE	UNP P17172
D	286	MSE	MET	MODIFIED RESIDUE	UNP P17172
D	301	MSE	MET	MODIFIED RESIDUE	UNP P17172
D	308	ALA	LEU	SEE REMARK 999	UNP P17172
D	357	MSE	MET	MODIFIED RESIDUE	UNP P17172

- Molecule 3 is POTASSIUM ION (three-letter code: K) (formula: K).

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

- Molecule 4 is PHOSPHATE ION (three-letter code: PO4) (formula: O<sub>4</sub>P).

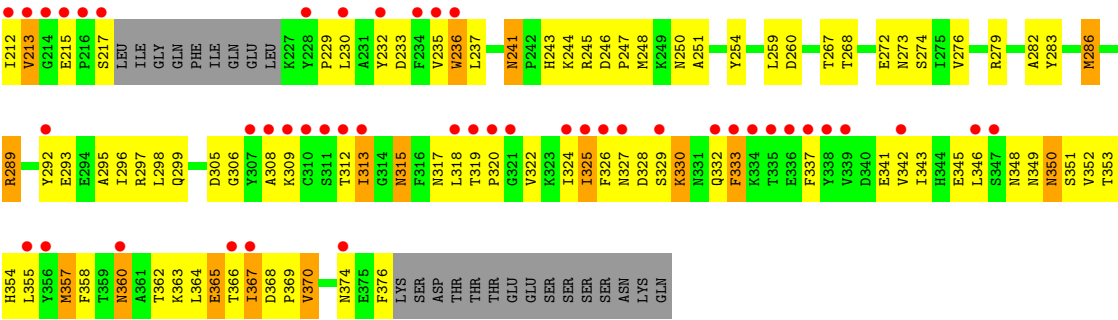


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total O P 3 2 1	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	307	Total 307	O 307	0	0
5	D	82	Total 82	O 82	0	0





## 4 Data and refinement statistics

Property	Value	Source
Space group	H 3 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	139.26Å 139.26Å 382.00Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	50.00 – 2.90 49.72 – 2.90	Depositor EDS
% Data completeness (in resolution range)	5.1 (50.00-2.90) 98.9 (49.72-2.90)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	0.08	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	7.57 (at 2.91Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, $R_{free}$	0.211 , 0.280 0.209 , 0.276	Depositor DCC
$R_{free}$ test set	1602 reflections (5.32%)	DCC
Wilson B-factor (Å <sup>2</sup> )	52.2	Xtriage
Anisotropy	0.794	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 76.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	7614	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	63.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.73% 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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: K, PO4

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	0.39	0/4391	0.65	1/5950 (0.0%)
2	D	0.33	0/2965	0.59	0/3997
All	All	0.36	0/7356	0.63	1/9947 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	575	GLY	N-CA-C	5.25	126.21	113.10

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	4309	0	4202	237	1
2	D	2912	0	2835	227	0
3	A	1	0	0	0	0
4	A	3	0	0	0	1
5	A	307	0	0	6	0
5	D	82	0	0	5	0
All	All	7614	0	7037	460	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 32.

The worst 5 of 460 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:89:LYS:HE2	1:A:89:LYS:H	1.16	1.10
1:A:184:GLU:HG2	1:A:203:GLY:HA3	1.36	1.03
1:A:89:LYS:CE	1:A:89:LYS:H	1.72	1.03
1:A:221:LEU:HD11	1:A:236:ILE:HD11	1.40	1.01
1:A:416:HIS:CD2	1:A:418:THR:HB	1.95	1.01

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)
4:A:586:PO4:P	4:A:586:PO4:O2[2_555]	1.48	0.72
1:A:584:HIS:OXT	1:A:584:HIS:OXT[4_555]	1.87	0.33

## 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	551/584 (94%)	512 (93%)	31 (6%)	8 (2%)	12	39
2	D	360/391 (92%)	295 (82%)	52 (14%)	13 (4%)	4	17
All	All	911/975 (93%)	807 (89%)	83 (9%)	21 (2%)	7	27

5 of 21 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	575	GLY
2	D	77	ALA
2	D	213	VAL

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	232	ASN
2	D	98	ASP

### 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	472/494 (96%)	435 (92%)	37 (8%)	15	39
2	D	324/333 (97%)	303 (94%)	21 (6%)	20	49
All	All	796/827 (96%)	738 (93%)	58 (7%)	16	43

5 of 58 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	418	THR
1	A	519	THR
2	D	333	PHE
1	A	429	ARG
1	A	457	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 37 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	443	ASN
2	D	11	ASN
2	D	354	HIS
1	A	457	ASN
1	A	526	GLN

### 5.3.3 RNA ⓘ

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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 2 ligands modelled in this entry, 1 is monoatomic - leaving 1 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	$\# Z  > 2$
4	PO4	A	586	-	0,2,4	0.00	-	0,1,6	0.00	-

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
4	PO4	A	586	-	-	0/0/0/0	0/0/0/0

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.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	586	PO4	0	1

## 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, 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	557/584 (95%)	0.10	6 (1%) 80 79	22, 45, 71, 103	0
2	D	349/391 (89%)	1.02	71 (20%) 1 1	49, 85, 120, 138	0
All	All	906/975 (92%)	0.46	77 (8%) 11 8	22, 59, 111, 138	0

The worst 5 of 77 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	216	PRO	5.5
2	D	367	ILE	5.1
2	D	99	SER	5.0
2	D	355	LEU	4.7
2	D	103	ASN	4.6

### 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 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. 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	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	K	A	585	1/1	0.91	0.15	-	48,48,48,48	1
4	PO4	A	586	3/5	0.99	0.30	-	24,24,25,95	2

## 6.5 Other polymers [i](#)

There are no such residues in this entry.