



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 17, 2023 – 01:52 PM EDT

PDB ID : 4ZKU
Title : P22 Tail Needle Gp26 crystallized at pH 10.0
Authors : Sankhala, R.S.; Cingolani, G.
Deposited on : 2015-04-30
Resolution : 2.50 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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

MolProbity : 4.02b-467
Xtriage (Phenix) : 1.13
EDS : 2.35.1
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) : 2.35.1

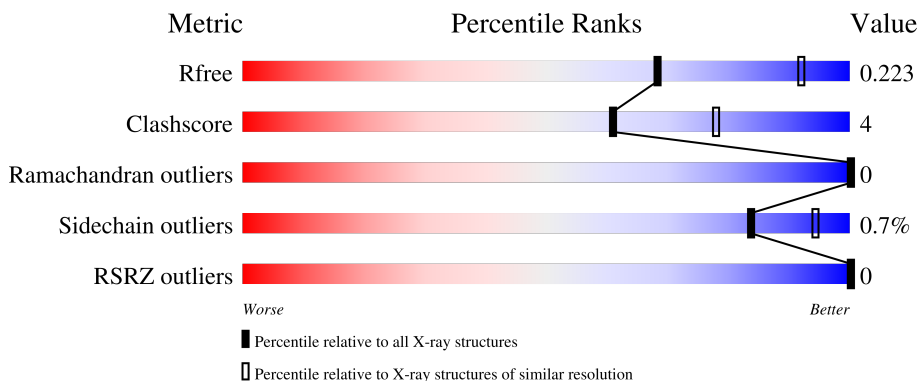
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.50 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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	237	 70% 5% 24%
1	B	237	 69% 7% 24%

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 3154 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 needle protein gp26.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	179	1342	816	242	283	1	0	0	0
1	B	179	1342	816	242	283	1	0	0	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	GLY	-	expression tag	UNP P35837
A	-2	PRO	-	expression tag	UNP P35837
A	-1	GLY	-	expression tag	UNP P35837
A	0	SER	-	expression tag	UNP P35837
A	222	MET	LEU	engineered mutation	UNP P35837
B	-3	GLY	-	expression tag	UNP P35837
B	-2	PRO	-	expression tag	UNP P35837
B	-1	GLY	-	expression tag	UNP P35837
B	0	SER	-	expression tag	UNP P35837
B	222	MET	LEU	engineered mutation	UNP P35837

- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Cl	0	0
			1	1		
2	B	1	Total	Cl	0	0
			1	1		

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

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

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total 1	Ca 1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	219	Total 219	O 219	0	0
4	B	247	Total 247	O 247	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: Tail needle protein gp26

Chain A:  70% 5% 24%

GLY PRO GLY SER MET ALA ASP PRO SER LEU ASN ASN PRO VAL VAL ILE GLN ALA THR ARG LEU ASP ALA SER SER ILE LEU PRO ARG ASN ASN VAL PHE SER LYS SER TYR LEU LEU TYR VAL ILE ALA GLN GLY THR ASP VAL GLY ALA ILE ILE ALA LYS ALA ASN ASN GLU GLY GLN G55 A56

Y57 V105 E109 S114 L115 Q116 A121 R125 Q137 L169 Q173 K220 D233

- Molecule 1: Tail needle protein gp26

Chain B:  69% 7% 24%

GLY PRO GLY SER MET ALA ASP PRO SER LEU ASN ASN PRO VAL VAL ILE GLN ALA THR ARG LEU ASP ALA SER SER ILE LEU PRO ARG ASN ASN VAL PHE SER LYS SER TYR LEU LEU TYR VAL ILE ALA GLN GLY THR ASP VAL GLY ALA ILE ILE ALA LYS ALA ASN ASN GLU GLY GLN G55

K62 E65 K96 S134 K143 T144 A145 L169 K185 D191 S197 D198 T199 N210 A221 D224 R227 D233

4 Data and refinement statistics i

Property	Value	Source
Space group	P 3	Depositor
Cell constants a, b, c, α , β , γ	43.20Å 43.20Å 271.23Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	28.82 – 2.50 28.82 – 2.49	Depositor EDS
% Data completeness (in resolution range)	96.4 (28.82-2.50) 92.5 (28.82-2.49)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	0.08	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.81 (at 2.51Å)	Xtrriage
Refinement program	PHENIX DEV_2016	Depositor
R, R_{free}	0.184 , 0.225 0.184 , 0.223	Depositor DCC
R_{free} test set	957 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å ²)	12.4	Xtrriage
Anisotropy	0.765	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 27.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.468 for -h,-k,l 0.109 for h,-h-k,-l 0.107 for -k,-h,-l	Xtrriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	3154	wwPDB-VP
Average B, all atoms (Å ²)	18.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 14.43% 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

5.1 Standard geometry

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

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	0.21	0/1353	0.38	0/1838
1	B	0.22	0/1353	0.41	0/1838
All	All	0.21	0/2706	0.40	0/3676

There are no bond length outliers.

There are no bond angle outliers.

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.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1342	0	1325	11	0
1	B	1342	0	1325	13	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	219	0	0	8	0
4	B	247	0	0	11	0
All	All	3154	0	2650	24	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 4.

The worst 5 of 24 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:220:LYS:NZ	4:A:402:HOH:O	2.13	0.80
1:B:227:ARG:NH1	4:B:401:HOH:O	2.25	0.67
1:B:96:LYS:HA	4:B:548:HOH:O	1.95	0.67
1:A:57:TYR:N	4:A:404:HOH:O	2.42	0.53
1:B:197:SER:OG	1:B:199:THR:O	2.28	0.50

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	177/237 (75%)	175 (99%)	2 (1%)	0	100	100
1	B	177/237 (75%)	176 (99%)	1 (1%)	0	100	100
All	All	354/474 (75%)	351 (99%)	3 (1%)	0	100	100

There are no Ramachandran outliers to report.

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	142/186 (76%)	141 (99%)	1 (1%)	84	94

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	142/186 (76%)	141 (99%)	1 (1%)	84	94
All	All	284/372 (76%)	282 (99%)	2 (1%)	84	94

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	169	LEU
1	B	169	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	137	GLN

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 4 ligands modelled in this entry, 4 are 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	179/237 (75%)	-0.84	0 100 100	11, 17, 27, 42	0
1	B	179/237 (75%)	-0.81	0 100 100	10, 17, 29, 33	0
All	All	358/474 (75%)	-0.83	0 100 100	10, 17, 28, 42	0

There are no RSRZ outliers to report.

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
3	CA	B	302	1/1	0.98	0.03	22,22,22,22	1
2	CL	B	301	1/1	0.99	0.08	20,20,20,20	1
3	CA	A	302	1/1	1.00	0.02	12,12,12,12	1
2	CL	A	301	1/1	1.00	0.03	9,9,9,9	1

6.5 Other polymers [i](#)

There are no such residues in this entry.