wwPDB X-ray Structure Validation Summary Report (i)

Oct 11, 2021 – 02:30 AM EDT

PDB ID	:	2POH
Title	:	Structure of Phage P22 Tail Needle gp26
Authors	:	Olia, A.S.; Cingolani, G.
Deposited on	:	2007-04-26
Resolution	:	2.10 Å(reported)

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

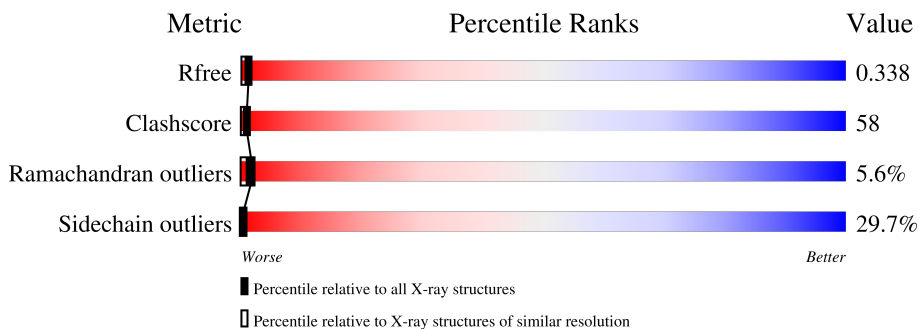|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 2.10 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	Percentile Ranks	Value
Rfree		0.338
Clashscore		58
Ramachandran outliers		5.6%
Sidechain outliers		29.7%

Worse .... *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)

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%

Mol	Chain	Length	Quality of chain
1	A	233	33% 48% 17% •
1	B	233	31% 48% 19% •
1	C	233	33% 47% 17% •
1	D	233	26% 48% 22% •
1	E	233	33% 46% 19% •
1	F	233	28% 48% 22% •

2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 11261 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 Head completion protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	233	Total	C	N	O	Se	0	0	0
			1738	1065	311	360	2			
1	B	233	Total	C	N	O	Se	0	0	0
			1738	1065	311	360	2			
1	C	233	Total	C	N	O	Se	0	0	0
			1738	1065	311	360	2			
1	D	233	Total	C	N	O	Se	0	0	0
			1715	1045	310	358	2			
1	E	233	Total	C	N	O	Se	0	0	0
			1712	1048	308	354	2			
1	F	233	Total	C	N	O	Se	0	0	0
			1734	1062	311	359	2			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	modified residue	UNP Q8LTG5
A	222	MSE	LEU	engineered mutation	UNP Q8LTG5
B	1	MSE	MET	modified residue	UNP Q8LTG5
B	222	MSE	LEU	engineered mutation	UNP Q8LTG5
C	1	MSE	MET	modified residue	UNP Q8LTG5
C	222	MSE	LEU	engineered mutation	UNP Q8LTG5
D	1	MSE	MET	modified residue	UNP Q8LTG5
D	222	MSE	LEU	engineered mutation	UNP Q8LTG5
E	1	MSE	MET	modified residue	UNP Q8LTG5
E	222	MSE	LEU	engineered mutation	UNP Q8LTG5
F	1	MSE	MET	modified residue	UNP Q8LTG5
F	222	MSE	LEU	engineered mutation	UNP Q8LTG5

- Molecule 2 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	155	Total 155	O 155	0	0
2	B	167	Total 167	O 167	0	0
2	C	151	Total 151	O 151	0	0
2	D	124	Total 124	O 124	0	0
2	E	138	Total 138	O 138	0	0
2	F	151	Total 151	O 151	0	0

3 Residue-property plots ⓘ

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. 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. 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.
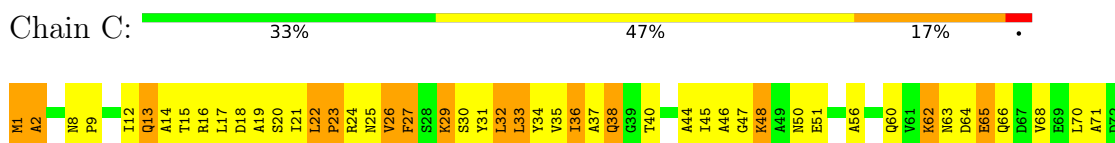
• Molecule 1: Head completion protein
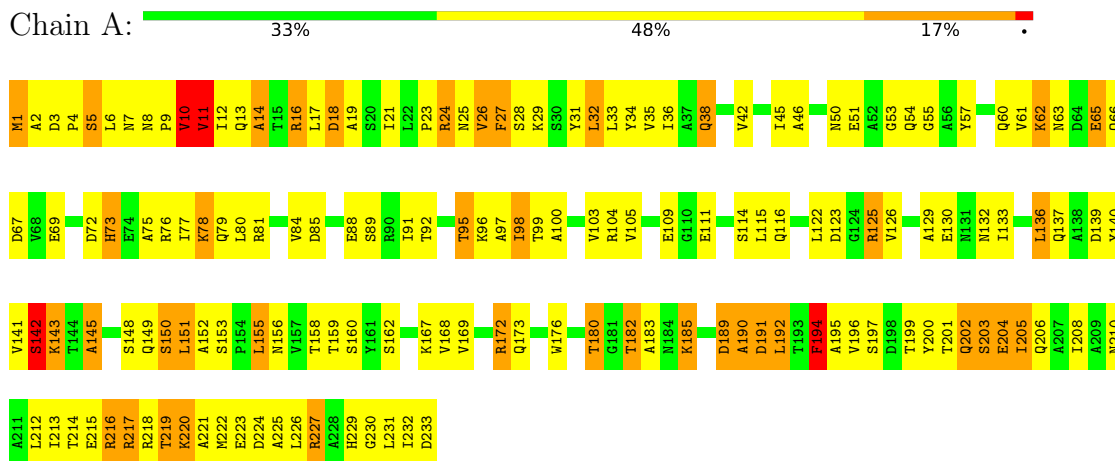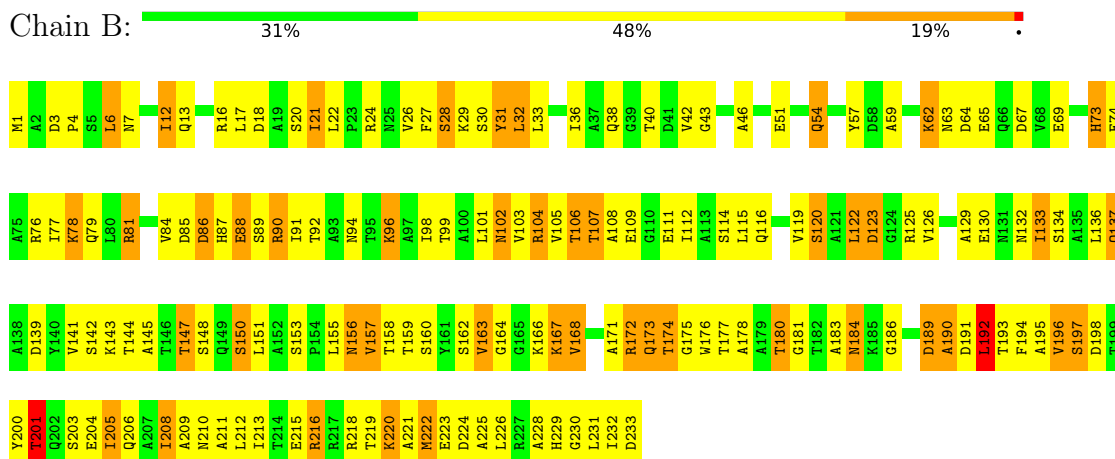
Chain A:

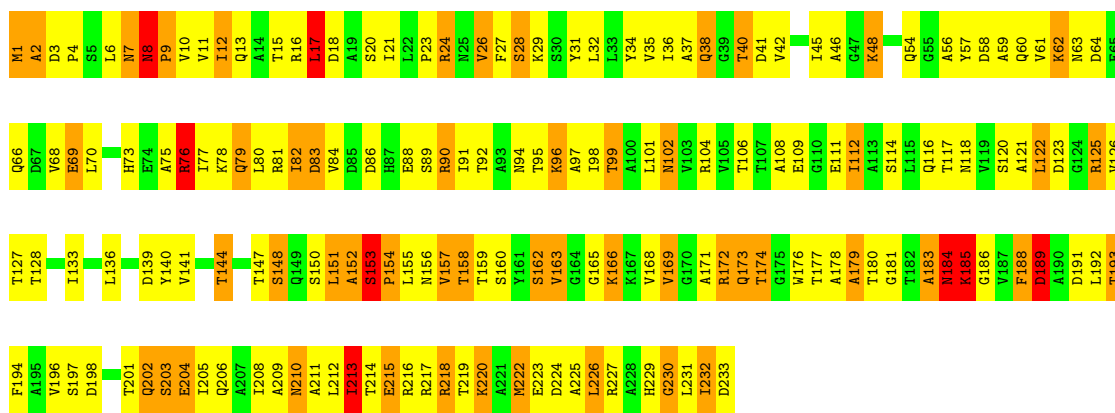• Molecule 1: Head completion protein

Chain B:

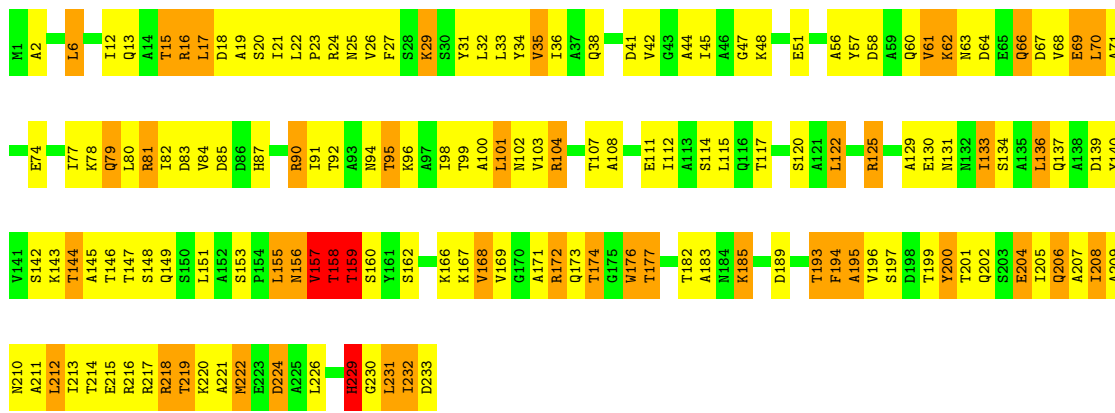• Molecule 1: Head completion protein

Chain C:

• Molecule 1: Head completion protein

Chain D:

• Molecule 1: Head completion protein

Chain E:

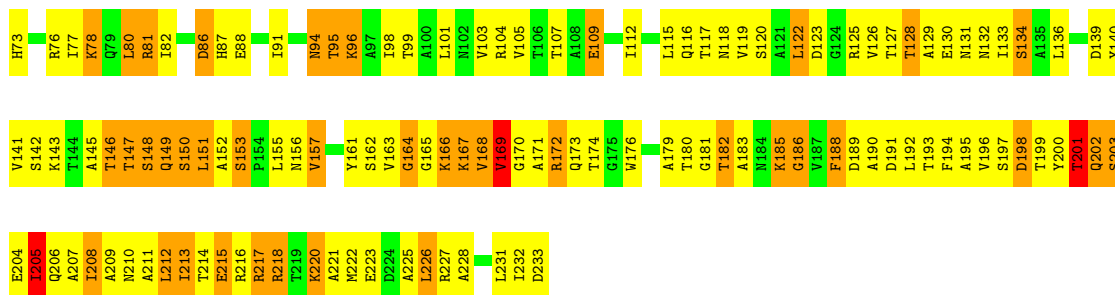

• Molecule 1: Head completion protein

Chain F:
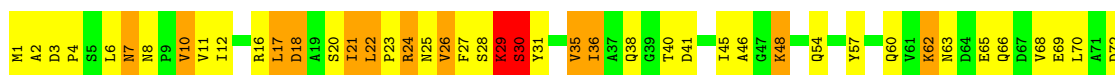
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	40.40Å 114.03Å 171.92Å 90.00° 90.74° 90.00°	Depositor
Resolution (Å)	(Not available) – 2.10 39.21 – 2.00	Depositor EDS
% Data completeness (in resolution range)	94.3 ((Not available)-2.10) 99.0 (39.21-2.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.07	Depositor
$< I/\sigma(I) >$ [1]	2.53 (at 2.00Å)	Xtriage
Refinement program	SHELXL-97	Depositor
R, R_{free}	0.167 , 0.232 0.281 , 0.338	Depositor DCC
R_{free} test set	4967 reflections (4.78%)	wwPDB-VP
Wilson B-factor (Å^2)	27.5	Xtriage
Anisotropy	0.693	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 53.3	EDS
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.023 for h,-k,-l	Xtriage
F_o,F_c correlation	0.88	EDS
Total number of atoms	11261	wwPDB-VP
Average B, all atoms (Å^2)	20.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.34% 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 ⓘ

5.1 Standard geometry ⓘ

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	0.38	0/1754	1.23	6/2384 (0.3%)
1	B	0.35	0/1754	1.03	2/2384 (0.1%)
1	C	0.46	1/1754 (0.1%)	1.09	7/2384 (0.3%)
1	D	0.33	0/1730	0.96	4/2347 (0.2%)
1	E	0.35	0/1728	0.99	5/2347 (0.2%)
1	F	0.34	0/1750	0.98	1/2377 (0.0%)
All	All	0.37	1/10470 (0.0%)	1.05	25/14223 (0.2%)

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	C	0	3
1	D	0	7
1	E	0	3
1	F	0	2
All	All	0	15

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	149	GLN	CD-NE2	11.38	1.61	1.32

The worst 5 of 25 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	172	ARG	NE-CZ-NH2	22.31	131.46	120.30
1	A	172	ARG	NE-CZ-NH1	-21.44	109.58	120.30
1	A	172	ARG	CD-NE-CZ	13.99	143.18	123.60
1	E	229	HIS	O-C-N	-8.77	108.28	123.20
1	D	90	ARG	NE-CZ-NH1	8.40	124.50	120.30

There are no chirality outliers.

5 of 15 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	196	VAL	Peptide
1	C	199	THR	Peptide
1	C	203	SER	Peptide
1	D	1	MSE	Peptide
1	D	2	ALA	Peptide

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	1738	0	1731	235	0
1	B	1738	0	1731	210	0
1	C	1738	0	1731	242	0
1	D	1715	0	1677	317	0
1	E	1712	0	1688	270	0
1	F	1734	0	1721	281	0
2	A	155	0	0	34	0
2	B	167	0	0	32	0
2	C	151	0	0	26	0
2	D	124	0	0	34	0
2	E	138	0	0	20	0
2	F	151	0	0	39	0
All	All	11261	0	10279	1200	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 58.

The worst 5 of 1200 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:D:222:MSE:SE	1:E:222:MSE:HE3	1.31	1.75
1:D:222:MSE:SE	1:E:222:MSE:CE	2.22	1.36
1:E:157:VAL:HG23	1:F:164:GLY:CA	1.54	1.35
1:D:153:SER:OG	1:E:158:THR:HB	1.36	1.25

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:E:157:VAL:CG2	1:E:158:THR:H	1.44	1.23

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	231/233 (99%)	199 (86%)	19 (8%)	13 (6%)	2	0
1	B	231/233 (99%)	211 (91%)	14 (6%)	6 (3%)	5	2
1	C	231/233 (99%)	199 (86%)	18 (8%)	14 (6%)	1	0
1	D	231/233 (99%)	176 (76%)	36 (16%)	19 (8%)	1	0
1	E	231/233 (99%)	204 (88%)	17 (7%)	10 (4%)	2	0
1	F	231/233 (99%)	200 (87%)	16 (7%)	15 (6%)	1	0
All	All	1386/1398 (99%)	1189 (86%)	120 (9%)	77 (6%)	2	0

5 of 77 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	18	ASP
1	A	190	ALA
1	A	191	ASP
1	A	194	PHE
1	A	200	TYR

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	184/182 (101%)	136 (74%)	48 (26%)		0	0
1	B	184/182 (101%)	127 (69%)	57 (31%)		0	0
1	C	184/182 (101%)	136 (74%)	48 (26%)		0	0
1	D	177/182 (97%)	121 (68%)	56 (32%)		0	0
1	E	178/182 (98%)	124 (70%)	54 (30%)		0	0
1	F	183/182 (100%)	122 (67%)	61 (33%)		0	0
All	All	1090/1092 (100%)	766 (70%)	324 (30%)		0	0

5 of 324 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	117	THR
1	F	86	ASP
1	E	146	THR
1	E	224	ASP
1	F	157	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 59 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	38	GLN
1	F	156	ASN
1	E	50	ASN
1	F	131	ASN
1	F	66	GLN

5.3.3 RNA ⓘ

There are no RNA molecules 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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

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 ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2   Non-standard residues in protein, DNA, RNA chains ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3   Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4   Ligands ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5   Other polymers ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.