W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

May 22, 2020 – 08:01 am BST

PDB ID	:	5FB5
Title	:	Crystal structure of the bacteriophage phi29 tail knob protein gp9
Authors	:	Xu, J.W.; Gui, M.; Wang, D.H.; Xiang, Y.
Deposited on	:	2015-12-14
Resolution	:	3.50 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:
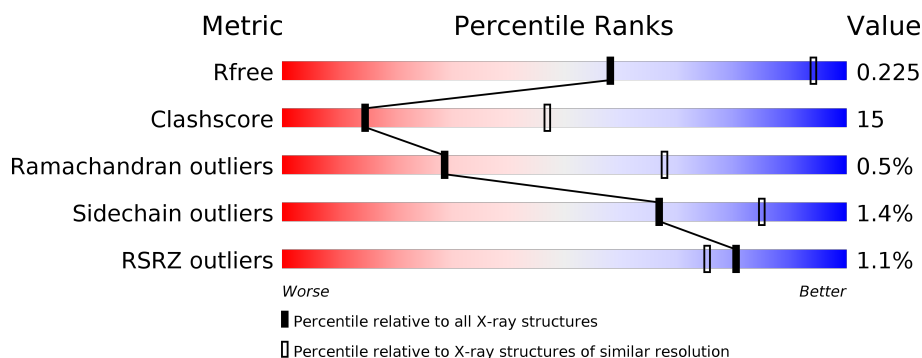
1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.50 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	Percentile Ranks	Value
Rfree		0.225
Clashscore		15
Ramachandran outliers		0.5%
Sidechain outliers		1.4%
RSRZ outliers		1.1%

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	1659 (3.60-3.40)
Clashscore	141614	1036 (3.58-3.42)
Ramachandran outliers	138981	1005 (3.58-3.42)
Sidechain outliers	138945	1006 (3.58-3.42)
RSRZ outliers	127900	1559 (3.60-3.40)

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 >=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% 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	605	72% 19% •• 7%
1	B	605	67% 28% 5%

2 Entry composition ⓘ

There is only 1 type of molecule in this entry. The entry contains 9152 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 Distal tube protein.

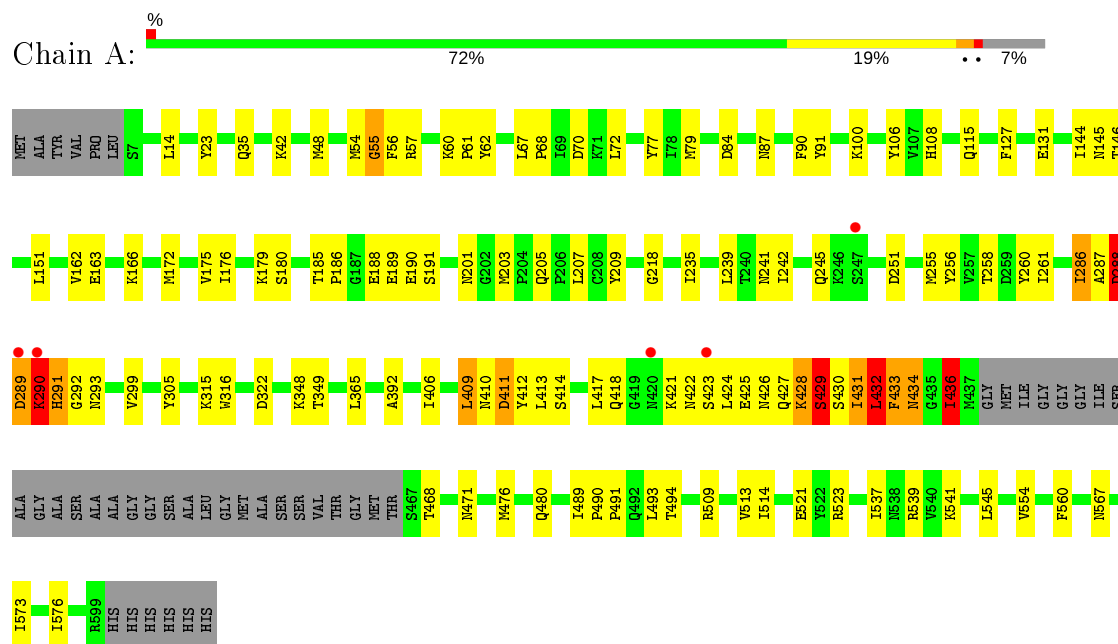| Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	564	Total 4539	C 2883	N 756	O 879	S 21	0	0	0
1	B	576	Total 4613	C 2928	N 768	O 895	S 22	0	0	0

There are 16 discrepancies between the modelled and reference sequences:

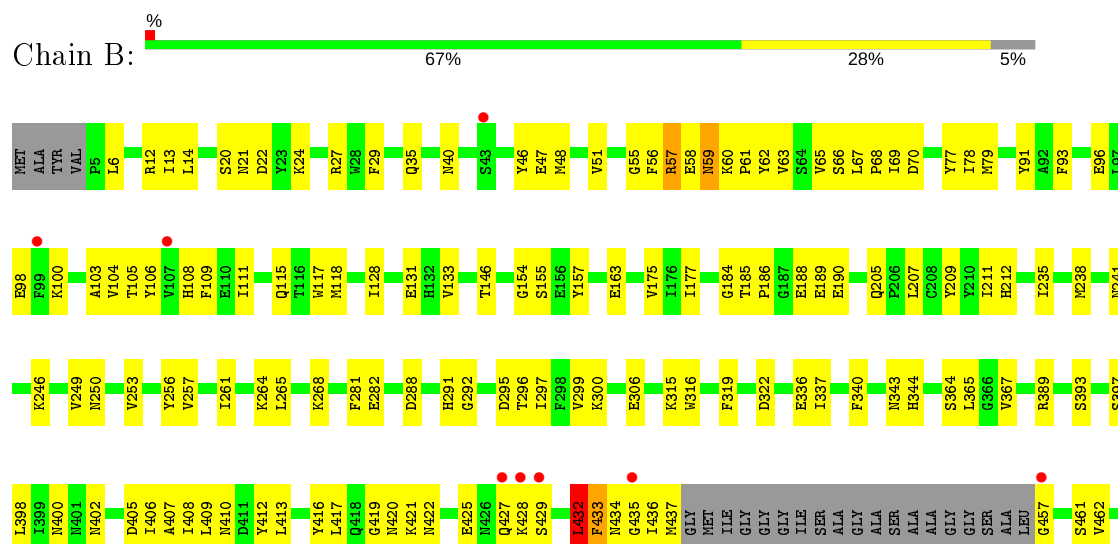| Chain	Residue	Modelled	Actual	Comment	Reference
A	41	SER	ARG	see sequence details	UNP P04331
A	121	ILE	MET	see sequence details	UNP P04331
A	600	HIS	-	expression tag	UNP P04331
A	601	HIS	-	expression tag	UNP P04331
A	602	HIS	-	expression tag	UNP P04331
A	603	HIS	-	expression tag	UNP P04331
A	604	HIS	-	expression tag	UNP P04331
A	605	HIS	-	expression tag	UNP P04331
B	41	SER	ARG	see sequence details	UNP P04331
B	121	ILE	MET	see sequence details	UNP P04331
B	600	HIS	-	expression tag	UNP P04331
B	601	HIS	-	expression tag	UNP P04331
B	602	HIS	-	expression tag	UNP P04331
B	603	HIS	-	expression tag	UNP P04331
B	604	HIS	-	expression tag	UNP P04331
B	605	HIS	-	expression tag	UNP P04331

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 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: Distal tube protein

- Molecule 1: Distal tube protein

M465 T466 M476 T494 N499 T500 A501 F502 G505 V513 I514 K515 K516 E521 Y522 R523 R524 S525 I537 L545 R548 K549 V554 Q555 T556 D564 I565 N566 N567 Q571 R574 T586 D587 N588 I589 G590 L598 R599 HIS HIS HIS HIS HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 3	Depositor
Cell constants a, b, c, α, β, γ	183.45Å 183.45Å 183.45Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.03 – 3.50 49.03 – 3.50	Depositor EDS
% Data completeness (in resolution range)	100.0 (49.03-3.50) 95.8 (49.03-3.50)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	5.09 (at 3.48Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.201 , 0.229 0.205 , 0.225	Depositor DCC
R_{free} test set	1329 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å^2)	67.6	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 37.1	EDS
L-test for twinning[2]	$<|L|> = 0.48, <L^2> = 0.31$	Xtriage
Estimated twinning fraction	0.038 for l,-k,h	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	9152	wwPDB-VP
Average B, all atoms (Å^2)	72.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.79% 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 (i)

5.1 Standard geometry (i)

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	0.30	0/4639	0.60	8/6273 (0.1%)
1	B	0.30	0/4714	0.62	4/6375 (0.1%)
All	All	0.30	0/9353	0.61	12/12648 (0.1%)

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	2

There are no bond length outliers.

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	432	LEU	CA-CB-CG	10.22	138.80	115.30
1	B	429	SER	N-CA-C	8.92	135.08	111.00
1	A	289	ASP	O-C-N	8.80	136.78	122.70
1	B	457	GLY	N-CA-C	8.23	133.68	113.10
1	A	288	ASP	N-CA-C	8.09	132.85	111.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	434	ASN	Peptide
1	A	436	ILE	Peptide

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	4539	0	4404	144	2
1	B	4613	0	4478	157	2
All	All	9152	0	8882	275	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 15.

The worst 5 of 275 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:185:THR:HG22	1:A:186:PRO:CD	1.41	1.50
1:A:185:THR:CG2	1:A:186:PRO:HD2	1.56	1.35
1:A:411:ASP:OD1	1:A:413:LEU:N	1.63	1.32
1:A:55:GLY:HA3	1:A:61:PRO:CA	1.79	1.12
1:A:55:GLY:CA	1:A:61:PRO:HA	1.79	1.12

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)
1:A:428:LYS:NZ	1:B:494:THR:CG2[9_555]	1.89	0.31
1:A:521:GLU:OE1	1:B:146:THR:OG1[9_555]	2.18	0.02

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	560/605 (93%)	529 (94%)	27 (5%)	4 (1%)	22	61
1	B	572/605 (94%)	535 (94%)	35 (6%)	2 (0%)	41	75
All	All	1132/1210 (94%)	1064 (94%)	62 (6%)	6 (0%)	29	68

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	288	ASP
1	A	291	HIS
1	B	57	ARG
1	B	59	ASN
1	A	189	GLU

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	503/528 (95%)	494 (98%)	9 (2%)	59	81
1	B	511/528 (97%)	506 (99%)	5 (1%)	76	88
All	All	1014/1056 (96%)	1000 (99%)	14 (1%)	67	85

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	432	LEU
1	A	433	PHE
1	B	428	LYS
1	A	429	SER
1	B	295	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	426	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	250	ASN
1	B	87	ASN
1	A	291	HIS
1	B	205	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 carbohydrates in this entry.

5.6   Ligand geometry (i)

There are no ligands in this entry.

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	564/605 (93%)	-0.34	5 (0%)	84	79	28, 53, 213, 344	0
1	B	576/605 (95%)	-0.35	8 (1%)	75	69	30, 57, 188, 324	0
All	All	1140/1210 (94%)	-0.35	13 (1%)	80	75	28, 54, 190, 344	0

The worst 5 of 13 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	428	LYS	2.9
1	A	289	ASP	2.8
1	B	427	GLN	2.7
1	A	290	LYS	2.6
1	B	435	GLY	2.5

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.