



# Full wwPDB X-ray Structure Validation Report ⓘ

Aug 29, 2020 – 11:32 AM BST

PDB ID : 6L5D  
Title : Crystal structure of the gas vesicle protein GvpF from *Anabaena* sp. PCC7120  
Authors : Cai, K.; Li, Q.  
Deposited on : 2019-10-23  
Resolution : 2.55 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  
Xtriage (Phenix) : 1.13  
EDS : 2.13  
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) : 2.13

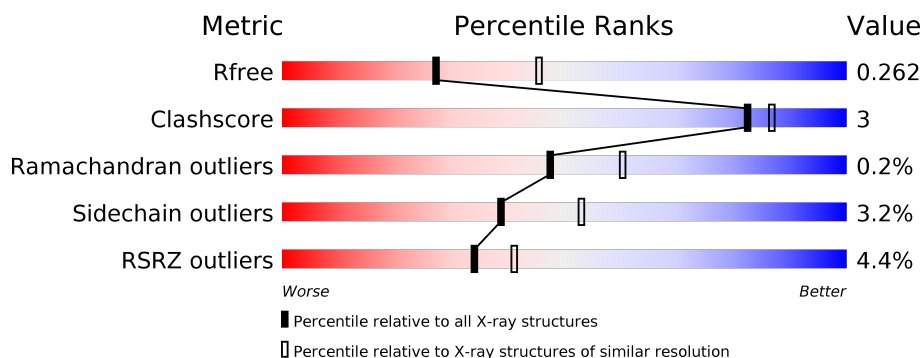
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.55 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	1284 (2.56-2.52)
Clashscore	141614	1332 (2.56-2.52)
Ramachandran outliers	138981	1315 (2.56-2.52)
Sidechain outliers	138945	1315 (2.56-2.52)
RSRZ outliers	127900	1272 (2.56-2.52)

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	250	
1	B	250	

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3990 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 Gas vesicle protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	237	Total	C	N	O	S	0	0	0
			1959	1260	331	363	5			
1	B	244	Total	C	N	O	S	0	0	0
			2003	1286	334	376	7			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	245	HIS	-	expression tag	UNP Q8YUT3
A	246	HIS	-	expression tag	UNP Q8YUT3
A	247	HIS	-	expression tag	UNP Q8YUT3
A	248	HIS	-	expression tag	UNP Q8YUT3
A	249	HIS	-	expression tag	UNP Q8YUT3
A	250	HIS	-	expression tag	UNP Q8YUT3
B	245	HIS	-	expression tag	UNP Q8YUT3
B	246	HIS	-	expression tag	UNP Q8YUT3
B	247	HIS	-	expression tag	UNP Q8YUT3
B	248	HIS	-	expression tag	UNP Q8YUT3
B	249	HIS	-	expression tag	UNP Q8YUT3
B	250	HIS	-	expression tag	UNP Q8YUT3

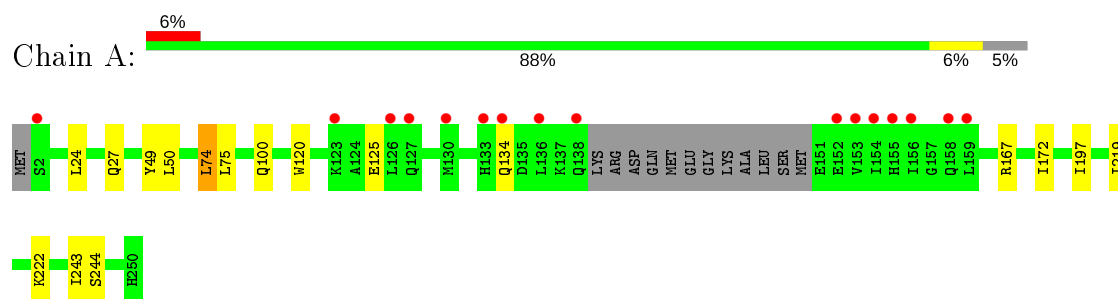
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	13	Total	O	0	0
			13	13		
2	B	15	Total	O	0	0
			15	15		

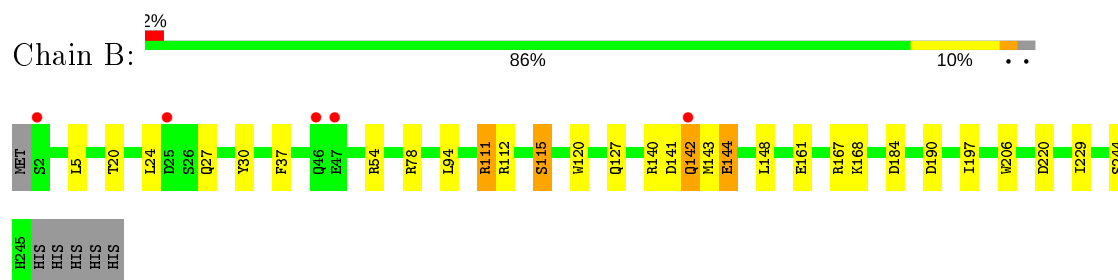
### 3 Residue-property plots [i](#)

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: Gas vesicle protein



- Molecule 1: Gas vesicle protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	96.40Å 147.09Å 106.58Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	44.46 – 2.55 44.46 – 2.55	Depositor EDS
% Data completeness (in resolution range)	99.6 (44.46-2.55) 99.7 (44.46-2.55)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	0.11	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.33 (at 2.54Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
R, $R_{free}$	0.196 , 0.256 0.200 , 0.262	Depositor DCC
$R_{free}$ test set	1268 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	47.2	Xtriage
Anisotropy	0.486	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 28.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	3990	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	53.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.75% 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](#)

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	0.77	0/2006	0.81	1/2713 (0.0%)
1	B	0.81	2/2046 (0.1%)	0.90	4/2763 (0.1%)
All	All	0.79	2/4052 (0.0%)	0.86	5/5476 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	161	GLU	CG-CD	5.73	1.60	1.51
1	B	115	SER	CB-OG	-5.72	1.34	1.42

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	74	LEU	CA-CB-CG	6.78	130.89	115.30
1	B	111	ARG	NE-CZ-NH1	5.77	123.19	120.30
1	B	78	ARG	NE-CZ-NH1	5.27	122.93	120.30
1	B	167	ARG	NE-CZ-NH2	-5.10	117.75	120.30
1	B	54	ARG	NE-CZ-NH2	5.09	122.84	120.30

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	1959	0	1925	7	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	2003	0	1990	13	0
2	A	13	0	0	0	0
2	B	15	0	0	1	0
All	All	3990	0	3915	20	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 3.

All (20) 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:B:141:ASP:O	1:B:143:MET:N	2.01	0.93
1:B:168:LYS:HE3	1:B:197:ILE:HG22	1.59	0.85
1:A:125:GLU:OE2	1:A:167:ARG:NH1	2.11	0.84
1:A:100:GLN:HE22	1:A:244:SER:H	1.46	0.64
1:A:75:LEU:HD22	1:A:243:ILE:HD11	1.83	0.60
1:B:20:THR:HG22	1:B:30:TYR:HE1	1.70	0.57
1:B:24:LEU:O	1:B:27:GLN:HB2	2.06	0.56
1:B:20:THR:HG22	1:B:30:TYR:CE1	2.40	0.55
1:B:144:GLU:HB3	1:B:148:LEU:HD11	1.90	0.53
1:A:75:LEU:HD22	1:A:243:ILE:CD1	2.43	0.48
1:A:172:ILE:HD11	1:A:197:ILE:O	2.15	0.46
1:B:141:ASP:C	1:B:143:MET:N	2.70	0.45
1:B:24:LEU:O	1:B:27:GLN:CB	2.65	0.44
1:A:24:LEU:HD23	1:A:49:TYR:CD2	2.54	0.42
1:B:220:ASP:HB2	1:B:229:ILE:HD12	2.03	0.41
1:A:219:ILE:O	1:A:222:LYS:HB2	2.20	0.41
1:B:112:ARG:HB2	1:B:206:TRP:CE3	2.55	0.41
1:B:111:ARG:NE	1:B:184:ASP:OD2	2.54	0.41
1:B:20:THR:HG23	2:B:306:HOH:O	2.22	0.40
1:B:37:PHE:CD2	1:B:94:LEU:HD11	2.56	0.40

There are no symmetry-related clashes.

## 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	233/250 (93%)	222 (95%)	11 (5%)	0	100	100
1	B	242/250 (97%)	233 (96%)	8 (3%)	1 (0%)	34	46
All	All	475/500 (95%)	455 (96%)	19 (4%)	1 (0%)	47	60

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	142	GLN

### 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	216/227 (95%)	211 (98%)	5 (2%)	50	65
1	B	221/227 (97%)	212 (96%)	9 (4%)	30	41
All	All	437/454 (96%)	423 (97%)	14 (3%)	39	53

All (14) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	27	GLN
1	A	50	LEU
1	A	74	LEU
1	A	120	TRP
1	A	134	GLN
1	B	5	LEU
1	B	115	SER
1	B	120	TRP
1	B	127	GLN
1	B	140	ARG
1	B	142	GLN
1	B	144	GLU

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
1	B	190	ASP
1	B	244	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	100	GLN
1	A	173	GLN
1	A	246	HIS
1	B	127	GLN
1	B	133	HIS

### 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](#)

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

### 6.1 Protein, DNA and RNA chains

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	237/250 (94%)	0.20	16 (6%) 17 20	31, 47, 101, 142	0
1	B	244/250 (97%)	0.04	5 (2%) 65 72	31, 47, 85, 131	0
All	All	481/500 (96%)	0.12	21 (4%) 34 41	31, 47, 93, 142	0

All (21) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	2	SER	5.3
1	A	133	HIS	4.8
1	A	156	ILE	4.7
1	A	138	GLN	4.7
1	B	142	GLN	4.3
1	A	127	GLN	4.0
1	A	154	ILE	3.8
1	A	136	LEU	3.7
1	A	130	MET	3.7
1	B	25	ASP	3.5
1	A	153	VAL	3.5
1	B	46	GLN	3.5
1	A	123	LYS	3.5
1	A	126	LEU	3.2
1	A	155	HIS	3.1
1	A	158	GLN	3.0
1	B	47	GLU	2.9
1	A	159	LEU	2.7
1	A	2	SER	2.7
1	A	134	GLN	2.3
1	A	152	GLU	2.2

## 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.