



Full wwPDB X-ray Structure Validation Report ⓘ

May 14, 2020 – 11:19 pm BST

PDB ID : 3GQH
Title : Crystal Structure of the Bacteriophage phi29 gene product 12 C-terminal fragment
Authors : Xiang, Y.; Rossmann, M.G.
Deposited on : 2009-03-24
Resolution : 1.80 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

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 : 2.11
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) : 2.11

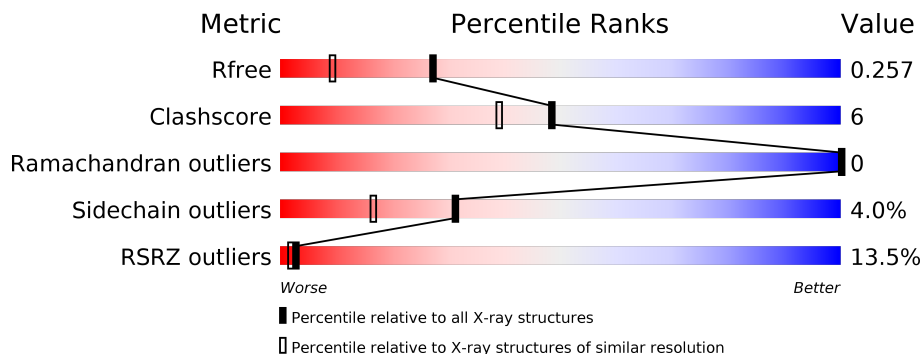
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.80 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	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)
RSRZ outliers	127900	5850 (1.80-1.80)

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 $\leq 5\%$. 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	163	 15% 89% 9% •
1	B	163	 14% 85% 13% •
1	C	163	 12% 85% 15%

2 Entry composition i

There are 2 unique types of molecules in this entry. The entry contains 4074 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 Preneck appendage protein.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace	
			Total	C	N				O
1	A	163	1310	840	216	254	0	5	0
1	B	163	1300	835	215	250	0	4	0
1	C	163	1295	833	215	247	0	4	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	701	GLY	ASP	SEE REMARK 999	UNP B3VMP8
A	817	SER	GLY	SEE REMARK 999	UNP B3VMP8
B	701	GLY	ASP	SEE REMARK 999	UNP B3VMP8
B	817	SER	GLY	SEE REMARK 999	UNP B3VMP8
C	701	GLY	ASP	SEE REMARK 999	UNP B3VMP8
C	817	SER	GLY	SEE REMARK 999	UNP B3VMP8

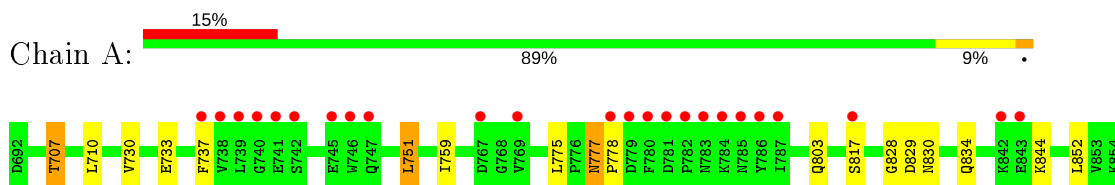
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	51	Total	O	0	0
			51	51		
2	B	49	Total	O	0	0
			49	49		
2	C	69	Total	O	0	0
			69	69		

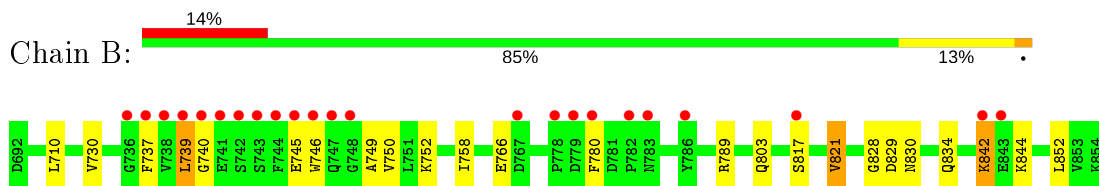
3 Residue-property plots [i](#)

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.

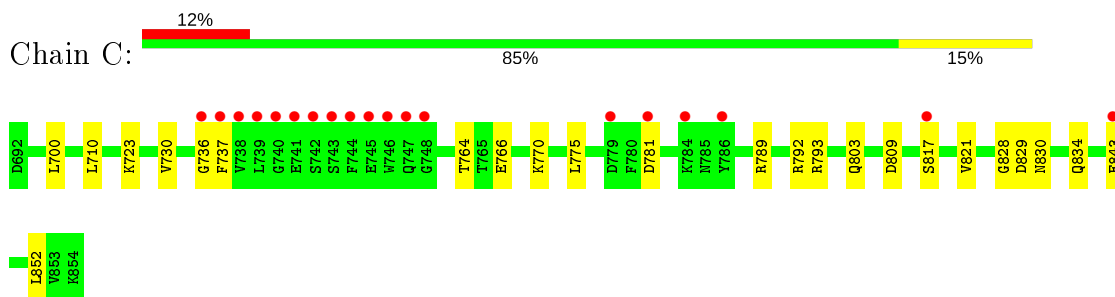
- Molecule 1: Preneck appendage protein



- Molecule 1: Preneck appendage protein



- Molecule 1: Preneck appendage protein



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	81.14Å 84.16Å 86.23Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	42.08 – 1.80 42.08 – 1.80	Depositor EDS
% Data completeness (in resolution range)	99.4 (42.08-1.80) 99.4 (42.08-1.80)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	0.09	Depositor
$\langle I/\sigma(I) \rangle$ ¹	7.28 (at 1.81Å)	Xtrriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.228 , 0.256 0.227 , 0.257	Depositor DCC
R_{free} test set	2795 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å ²)	19.8	Xtrriage
Anisotropy	0.050	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.42 , 57.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.015 for -h,l,k	Xtrriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	4074	wwPDB-VP
Average B, all atoms (Å ²)	26.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.96% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² 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	0.48	0/1336	0.61	0/1804
1	B	0.46	0/1326	0.55	0/1791
1	C	0.51	0/1321	0.65	1/1785 (0.1%)
All	All	0.48	0/3983	0.60	1/5380 (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	C	809	ASP	CB-CG-OD2	5.79	123.51	118.30

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	1310	0	1296	15	0
1	B	1300	0	1288	21	0
1	C	1295	0	1281	14	1
2	A	51	0	0	0	0
2	B	49	0	0	2	1
2	C	69	0	0	3	0
All	All	4074	0	3865	47	1

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 6.

All (47) 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:707:THR:HG21	1:A:733:GLU:OE2	1.59	1.00
1:B:803:GLN:HG2	1:B:834:GLN:HE22	1.32	0.93
1:A:777:ASN:HD22	1:A:778:PRO:HD2	1.50	0.75
1:A:803:GLN:HG2	1:A:834:GLN:HE22	1.49	0.75
1:C:803:GLN:HG2	1:C:834:GLN:HE22	1.49	0.75
1:B:739:LEU:HD21	1:C:736[B]:GLY:HA3	1.69	0.74
1:A:777:ASN:HD22	1:A:778:PRO:CD	2.04	0.70
1:C:792:ARG:HA	2:C:117:HOH:O	1.95	0.66
1:C:766:GLU:HG2	2:C:61:HOH:O	1.94	0.66
1:B:821:VAL:HG13	2:B:78:HOH:O	1.98	0.64
1:A:737[A]:PHE:CD1	1:B:737[A]:PHE:CZ	2.86	0.63
1:B:842:LYS:HD2	1:B:842:LYS:H	1.63	0.62
1:B:739:LEU:HD23	1:C:737[B]:PHE:CZ	2.35	0.61
1:A:834:GLN:HE21	1:A:852:LEU:HB2	1.66	0.61
1:B:803:GLN:HG2	1:B:834:GLN:NE2	2.11	0.60
1:A:817:SER:HB2	1:A:828:GLY:O	2.01	0.59
1:C:817:SER:HB3	1:C:830:ASN:N	2.21	0.56
1:C:817:SER:HB2	1:C:828:GLY:O	2.07	0.55
1:A:817:SER:HB3	1:A:830:ASN:N	2.22	0.54
1:A:751:LEU:HD12	1:A:775:LEU:HD23	1.89	0.53
1:B:842:LYS:HD2	1:B:842:LYS:N	2.22	0.53
1:B:817:SER:HB2	1:B:828:GLY:O	2.09	0.53
1:C:764:THR:HG23	2:C:72:HOH:O	2.09	0.53
1:A:710:LEU:HD23	1:A:730:VAL:HG22	1.91	0.53
1:B:817:SER:HB2	1:B:829:ASP:HA	1.90	0.53
1:B:710:LEU:HD23	1:B:730:VAL:HG22	1.90	0.52
1:C:764:THR:HG22	1:C:770:LYS:HG2	1.91	0.52
1:B:739:LEU:HD13	1:B:740:GLY:N	2.26	0.51
1:A:817:SER:HB3	1:A:829:ASP:C	2.32	0.50
1:B:750:VAL:HG11	1:B:758:ILE:HG23	1.93	0.49
1:A:707:THR:CG2	1:A:733:GLU:OE2	2.48	0.49
1:B:817:SER:HB3	1:B:830:ASN:N	2.29	0.48
1:C:834:GLN:HE21	1:C:852:LEU:HB2	1.78	0.48
1:B:834:GLN:HE21	1:B:852:LEU:HB2	1.79	0.47
1:A:803:GLN:HG2	1:A:834:GLN:NE2	2.23	0.46
1:B:746:TRP:HD1	1:B:752:LYS:HB2	1.81	0.45
1:C:817:SER:HB3	1:C:829:ASP:C	2.37	0.45
1:A:777:ASN:ND2	1:A:778:PRO:HD2	2.25	0.45

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:739:LEU:HD13	1:B:740:GLY:H	1.82	0.45
1:B:842:LYS:CD	1:B:842:LYS:H	2.30	0.44
1:C:710:LEU:HD23	1:C:730:VAL:HG22	1.99	0.43
1:C:736[B]:GLY:O	1:C:737[B]:PHE:CD1	2.71	0.43
1:C:723:LYS:HG3	1:C:821:VAL:HG22	2.02	0.41
1:A:751:LEU:HD13	1:A:759:ILE:HD12	2.02	0.41
1:B:817:SER:HB3	1:B:829:ASP:C	2.41	0.41
1:B:821:VAL:CG1	2:B:78:HOH:O	2.64	0.40
1:B:749:ALA:HA	1:B:780:PHE:CG	2.56	0.40

All (1) 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:C:793:ARG:NH1	2:B:124:HOH:O[3_545]	2.17	0.03

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	166/163 (102%)	164 (99%)	2 (1%)	0	100	100
1	B	165/163 (101%)	160 (97%)	5 (3%)	0	100	100
1	C	165/163 (101%)	160 (97%)	5 (3%)	0	100	100
All	All	496/489 (101%)	484 (98%)	12 (2%)	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	138/135 (102%)	134 (97%)	4 (3%)	42	29
1	B	136/135 (101%)	129 (95%)	7 (5%)	24	10
1	C	134/135 (99%)	129 (96%)	5 (4%)	34	19
All	All	408/405 (101%)	392 (96%)	16 (4%)	31	17

All (16) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	707	THR
1	A	751	LEU
1	A	777	ASN
1	A	844	LYS
1	B	739	LEU
1	B	745	GLU
1	B	766	GLU
1	B	789	ARG
1	B	821	VAL
1	B	842	LYS
1	B	844	LYS
1	C	700	LEU
1	C	775	LEU
1	C	781	ASP
1	C	789	ARG
1	C	843	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (9) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	703	GLN
1	A	777	ASN
1	A	785	ASN
1	A	834	GLN
1	B	791	GLN
1	B	834	GLN
1	C	747	GLN
1	C	783	ASN
1	C	834	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

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, 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	163/163 (100%)	0.64	24 (14%) 2 1	12, 23, 47, 55	0
1	B	163/163 (100%)	0.77	23 (14%) 2 1	10, 22, 49, 61	0
1	C	163/163 (100%)	0.72	19 (11%) 4 3	10, 21, 46, 49	0
All	All	489/489 (100%)	0.71	66 (13%) 3 2	10, 22, 47, 61	0

All (66) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	737[A]	PHE	12.9
1	B	737[A]	PHE	11.3
1	B	746	TRP	9.9
1	A	737[A]	PHE	9.9
1	C	736[A]	GLY	7.5
1	B	739	LEU	7.3
1	B	744	PHE	6.8
1	A	782	PRO	6.7
1	A	779	ASP	5.8
1	C	747	GLN	5.8
1	B	747	GLN	5.7
1	C	786	TYR	5.6
1	B	740	GLY	5.5
1	A	786	TYR	5.0
1	C	740	GLY	4.9
1	B	736[A]	GLY	4.8
1	B	742	SER	4.8
1	A	783	ASN	4.7
1	B	786	TYR	4.6
1	B	745	GLU	4.5
1	C	742	SER	4.5
1	C	746	TRP	4.4
1	B	748	GLY	4.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	843	GLU	4.2
1	A	778	PRO	4.1
1	B	780	PHE	3.9
1	C	739	LEU	3.9
1	B	743	SER	3.7
1	A	781	ASP	3.7
1	A	784	LYS	3.5
1	A	747	GLN	3.5
1	C	817	SER	3.5
1	C	781	ASP	3.1
1	A	740	GLY	3.1
1	C	779	ASP	3.1
1	C	744	PHE	3.1
1	C	741	GLU	3.0
1	B	842	LYS	3.0
1	A	746	TRP	3.0
1	A	817	SER	3.0
1	B	779	ASP	3.0
1	B	843	GLU	2.9
1	A	742	SER	2.9
1	B	782	PRO	2.8
1	C	745	GLU	2.8
1	A	745	GLU	2.8
1	C	738[A]	VAL	2.8
1	A	741	GLU	2.8
1	B	778	PRO	2.7
1	A	767	ASP	2.6
1	A	739	LEU	2.5
1	C	743	SER	2.5
1	B	738[A]	VAL	2.4
1	B	817	SER	2.4
1	A	785	ASN	2.4
1	A	787	ILE	2.4
1	A	842	LYS	2.4
1	B	767	ASP	2.3
1	C	843	GLU	2.3
1	A	738[A]	VAL	2.3
1	B	741	GLU	2.2
1	B	783	ASN	2.1
1	C	784	LYS	2.1
1	A	780	PHE	2.1
1	C	748	GLY	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	769	VAL	2.0

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.