WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Feb 19, 2024 – 08:45 PM EST

PDB ID	:	4JMQ
Title	:	Crystal structure of pb9: The Dit of bacteriophage T5.
Authors	:	Flayhan, A.; Vellieux, F.M.D.; Girard, E.; Maury, O.; Boulanger, P.; Breyton, C.
Deposited on	:	2013-03-14
Resolution	:	1.90 Å(reported)

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

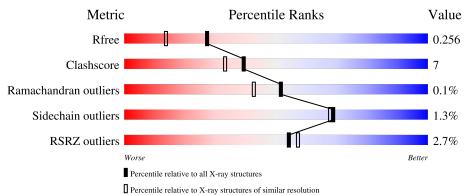|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 1.90 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	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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% 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	217	74%　　13%　　12%
1	B	217	%　72%　　15%　　13%
1	C	217	4%　74%　　17%　·　7%
1	D	217	4%　78%　　9%　　13%

2 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 6935 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 Bacteriophage T5 distal tail protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	190	Total	C	N	O	S	0	1	0
			1505	964	246	293	2			
1	B	189	Total	C	N	O	S	0	1	0
			1493	956	242	293	2			
1	C	201	Total	C	N	O	S	0	1	0
			1608	1029	269	307	3			
1	D	189	Total	C	N	O	S	0	2	0
			1500	961	243	294	2			

There are 52 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	205	GLU	-	expression tag	UNP Q6QGE8
A	206	ASN	-	expression tag	UNP Q6QGE8
A	207	LEU	-	expression tag	UNP Q6QGE8
A	208	TYR	-	expression tag	UNP Q6QGE8
A	209	PHE	-	expression tag	UNP Q6QGE8
A	210	GLN	-	expression tag	UNP Q6QGE8
A	211	GLY	-	expression tag	UNP Q6QGE8
A	212	HIS	-	expression tag	UNP Q6QGE8
A	213	HIS	-	expression tag	UNP Q6QGE8
A	214	HIS	-	expression tag	UNP Q6QGE8
A	215	HIS	-	expression tag	UNP Q6QGE8
A	216	HIS	-	expression tag	UNP Q6QGE8
A	217	HIS	-	expression tag	UNP Q6QGE8
B	205	GLU	-	expression tag	UNP Q6QGE8
B	206	ASN	-	expression tag	UNP Q6QGE8
B	207	LEU	-	expression tag	UNP Q6QGE8
B	208	TYR	-	expression tag	UNP Q6QGE8
B	209	PHE	-	expression tag	UNP Q6QGE8
B	210	GLN	-	expression tag	UNP Q6QGE8
B	211	GLY	-	expression tag	UNP Q6QGE8
B	212	HIS	-	expression tag	UNP Q6QGE8

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	213	HIS	-	expression tag	UNP Q6QGE8
B	214	HIS	-	expression tag	UNP Q6QGE8
B	215	HIS	-	expression tag	UNP Q6QGE8
B	216	HIS	-	expression tag	UNP Q6QGE8
B	217	HIS	-	expression tag	UNP Q6QGE8
C	205	GLU	-	expression tag	UNP Q6QGE8
C	206	ASN	-	expression tag	UNP Q6QGE8
C	207	LEU	-	expression tag	UNP Q6QGE8
C	208	TYR	-	expression tag	UNP Q6QGE8
C	209	PHE	-	expression tag	UNP Q6QGE8
C	210	GLN	-	expression tag	UNP Q6QGE8
C	211	GLY	-	expression tag	UNP Q6QGE8
C	212	HIS	-	expression tag	UNP Q6QGE8
C	213	HIS	-	expression tag	UNP Q6QGE8
C	214	HIS	-	expression tag	UNP Q6QGE8
C	215	HIS	-	expression tag	UNP Q6QGE8
C	216	HIS	-	expression tag	UNP Q6QGE8
C	217	HIS	-	expression tag	UNP Q6QGE8
D	205	GLU	-	expression tag	UNP Q6QGE8
D	206	ASN	-	expression tag	UNP Q6QGE8
D	207	LEU	-	expression tag	UNP Q6QGE8
D	208	TYR	-	expression tag	UNP Q6QGE8
D	209	PHE	-	expression tag	UNP Q6QGE8
D	210	GLN	-	expression tag	UNP Q6QGE8
D	211	GLY	-	expression tag	UNP Q6QGE8
D	212	HIS	-	expression tag	UNP Q6QGE8
D	213	HIS	-	expression tag	UNP Q6QGE8
D	214	HIS	-	expression tag	UNP Q6QGE8
D	215	HIS	-	expression tag	UNP Q6QGE8
D	216	HIS	-	expression tag	UNP Q6QGE8
D	217	HIS	-	expression tag	UNP Q6QGE8

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	215	Total 215	O 215	0	0
2	B	212	Total 212	O 212	0	0
2	C	191	Total 191	O 191	0	0
2	D	211	Total 211	O 211	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 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: Bacteriophage T5 distal tail protein

• Molecule 1: Bacteriophage T5 distal tail protein

• Molecule 1: Bacteriophage T5 distal tail protein

• Molecule 1: Bacteriophage T5 distal tail protein

M1 R2 L3 P4 Y7 S19 D26 PRO MET ILE ARG ASP GLU LEU PRO ASN GLY LYS VAL LYS GLU VAL LYS I43 L57 F58 R68 V81 P84 F89 T99 K105 Q108 N112 P133 S163 S182 F183 G184 S185 T186 N190 G191

I196 L200 L204 E205 ASN LEU TYR PHE GLN GLY HIS HIS HIS HIS HIS HIS

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	55.66Å 70.03Å 71.01Å 91.05° 107.66° 112.03°	Depositor
Resolution (Å)	27.69 – 1.90 66.91 – 1.89	Depositor EDS
% Data completeness (in resolution range)	95.4 (27.69-1.90) 95.4 (66.91-1.89)	Depositor EDS
R_{merge}	0.16	Depositor
R_{sym}	0.19	Depositor
$< I/\sigma(I) >$ [1]	1.36 (at 1.90Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, R_{free}	0.203 , 0.252 0.208 , 0.256	Depositor DCC
R_{free} test set	3549 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	20.8	Xtriage
Anisotropy	0.291	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 56.7	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	6935	wwPDB-VP
Average B, all atoms (Å^2)	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.10% 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	0.37	0/1545	0.55	0/2100
1	B	0.38	0/1532	0.56	0/2082
1	C	0.39	0/1656	0.60	0/2251
1	D	0.35	0/1542	0.55	0/2096
All	All	0.37	0/6275	0.57	0/8529

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	1505	0	1475	23	0
1	B	1493	0	1460	23	0
1	C	1608	0	1549	32	0
1	D	1500	0	1470	14	0
2	A	215	0	0	7	0
2	B	212	0	0	8	0
2	C	191	0	0	5	0
2	D	211	0	0	5	0
All	All	6935	0	5954	89	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 7.

The worst 5 of 89 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:43:ILE:N	2:B:505:HOH:O	1.97	0.97
1:A:43:ILE:N	2:A:464:HOH:O	1.97	0.97
1:A:108:GLN:NE2	2:A:366:HOH:O	1.94	0.94
1:C:60:ASP:OD1	2:C:460:HOH:O	1.91	0.86
1:A:170:GLY:O	2:A:495:HOH:O	2.00	0.80

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	187/217 (86%)	184 (98%)	3 (2%)	0	100	100
1	B	186/217 (86%)	184 (99%)	2 (1%)	0	100	100
1	C	198/217 (91%)	190 (96%)	7 (4%)	1 (0%)	29	18
1	D	187/217 (86%)	183 (98%)	4 (2%)	0	100	100
All	All	758/868 (87%)	741 (98%)	16 (2%)	1 (0%)	51	42

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	27	PRO

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	169/193 (88%)	168 (99%)	1 (1%)	86	87
1	B	168/193 (87%)	165 (98%)	3 (2%)	59	55
1	C	179/193 (93%)	175 (98%)	4 (2%)	52	47
1	D	169/193 (88%)	168 (99%)	1 (1%)	86	87
All	All	685/772 (89%)	676 (99%)	9 (1%)	69	68

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	208	TYR
1	D	204	LEU
1	B	205	GLU
1	C	56	GLU
1	C	60	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 (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(Å²)	Q<0.9
1	A	190/217 (87%)	-0.14	1 (0%)	91	92	21, 32, 60, 97	0
1	B	189/217 (87%)	-0.14	2 (1%)	80	82	19, 31, 57, 83	0
1	C	201/217 (92%)	0.09	9 (4%)	33	36	19, 33, 79, 105	0
1	D	189/217 (87%)	0.15	9 (4%)	30	33	19, 37, 85, 104	0
All	All	769/868 (88%)	-0.01	21 (2%)	54	57	19, 32, 72, 105	0

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	209	PHE	5.8
1	C	208	TYR	5.2
1	C	186	THR	4.6
1	C	185	SER	3.7
1	D	186	THR	3.7

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.