wwPDB X-ray Structure Validation Summary Report (i)

Jan 27, 2024 – 09:48 AM EST

PDB ID	:	1AL0
Title	:	PROCAPSID OF BACTERIOPHAGE PHIX174
Authors	:	Rossmann, M.G.; Dokland, T.
Deposited on	:	1997-06-06
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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

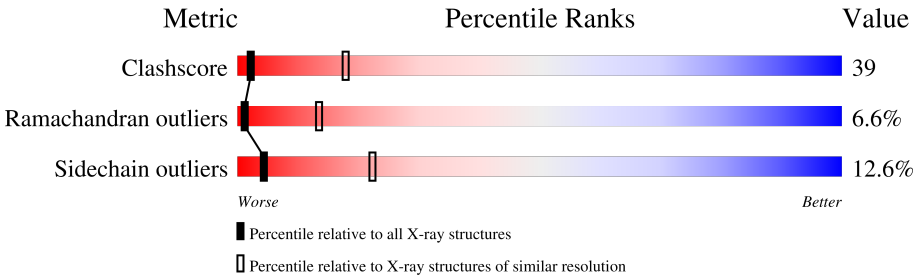
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

1 Overall quality at a glance (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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	141614	1036 (3.58-3.42)
Ramachandran outliers	138981	1005 (3.58-3.42)
Sidechain outliers	138945	1006 (3.58-3.42)

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	1	152	45% 42% 7% • 6%
1	2	152	36% 36% 14% • 12%
1	3	152	40% 39% 12% • 8%
1	4	152	46% 38% 10% • •
2	F	426	35% 46% 14% • •
3	G	175	57% 34% 7% •
4	B	120	8% 18% 9% 5% 59%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 9521 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 SCAFFOLDING PROTEIN GPD.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	1	143	Total	C	N	O	S	0	0	0
			1125	716	194	211	4			
1	2	133	Total	C	N	O	S	0	0	0
			1039	665	175	195	4			
1	3	140	Total	C	N	O	S	0	0	0
			1099	699	187	209	4			
1	4	146	Total	C	N	O	S	0	0	0
			1145	728	197	215	5			

- Molecule 2 is a protein called CAPSID PROTEIN GPF.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	F	418	Total	C	N	O	S	0	0	0
			3358	2139	581	625	13			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	216	ARG	HIS	conflict	UNP P03641

- Molecule 3 is a protein called SPIKE PROTEIN GPG.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	G	175	Total	C	N	O	S	0	0	0
			1340	856	221	255	8			

- Molecule 4 is a protein called SCAFFOLDING PROTEIN GPB.

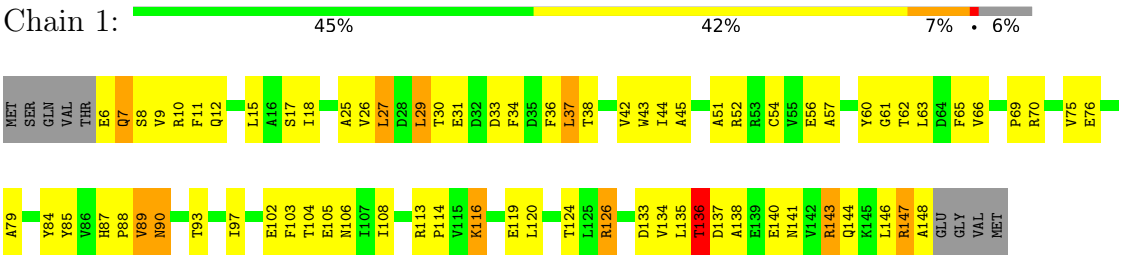
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	B	49	Total	C	N	O	S	0	0	0
			415	262	71	80	2			

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. 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.

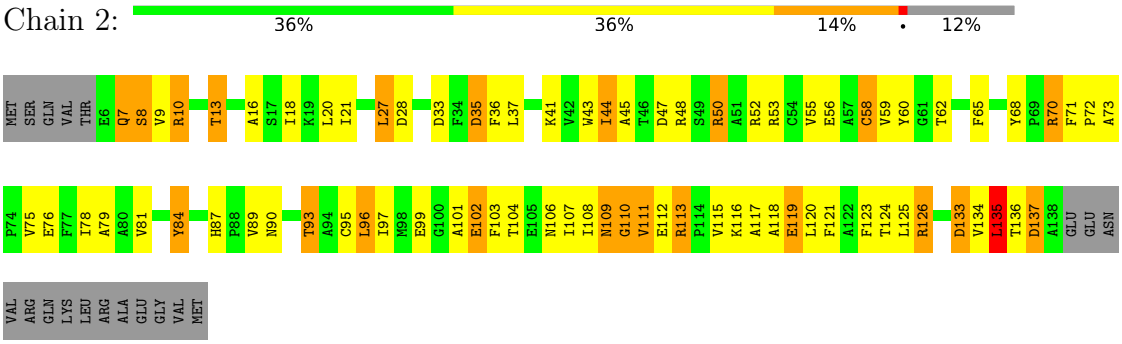
• Molecule 1: SCAFFOLDING PROTEIN GPD

Chain 1:



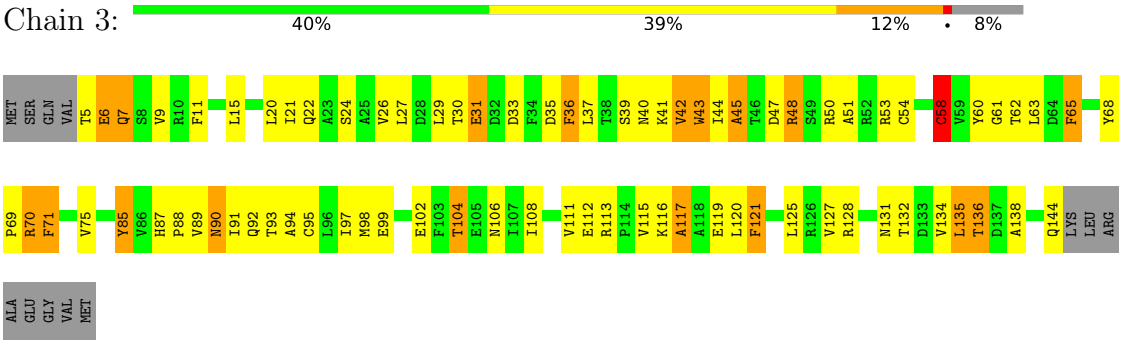
• Molecule 1: SCAFFOLDING PROTEIN GPD

Chain 2:



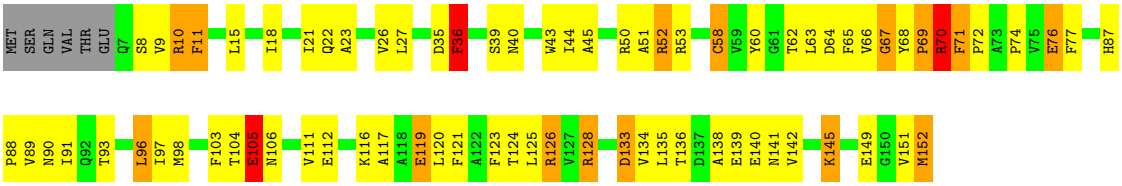
• Molecule 1: SCAFFOLDING PROTEIN GPD

Chain 3:



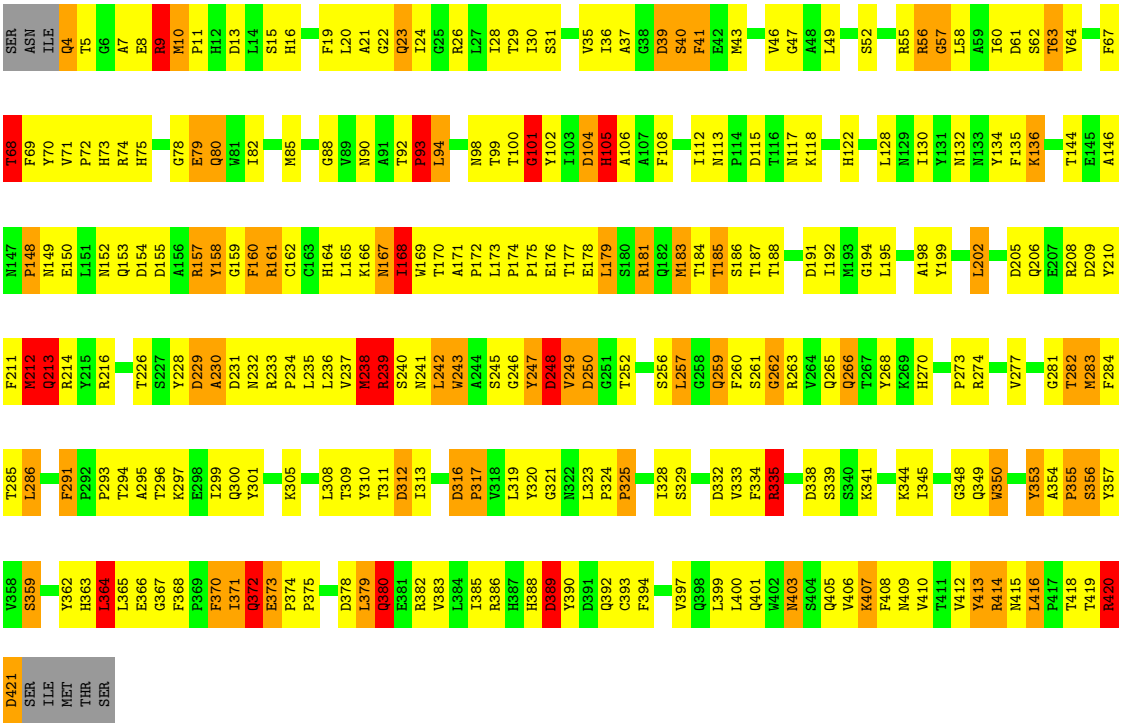
• Molecule 1: SCAFFOLDING PROTEIN GPD

Chain 4:



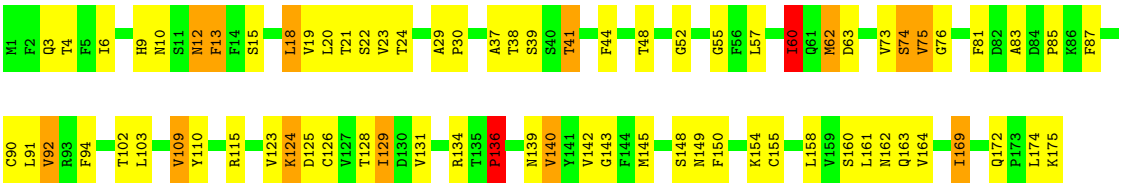
• Molecule 2: CAPSID PROTEIN GPF

Chain F:



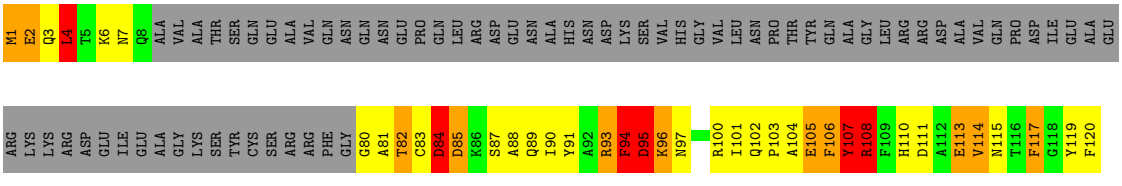
• Molecule 3: SPIKE PROTEIN GPG

Chain G:



• Molecule 4: SCAFFOLDING PROTEIN GPB

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 21 3	Depositor
Cell constants a, b, c, α, β, γ	774.00Å 774.00Å 774.00Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	8.00 – 3.50 44.54 – 3.50	Depositor EDS
% Data completeness (in resolution range)	(Not available) (8.00-3.50) 55.2 (44.54-3.50)	Depositor EDS
R_{merge}	0.25	Depositor
R_{sym}	0.25	Depositor
$< I/\sigma(I) >$ [1]	2.09 (at 3.48Å)	Xtriage
Refinement program	X-PLOR 3.1	Depositor
R, R_{free}	0.316 , (Not available) 0.486 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	70.0	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , -46.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.004 for -l,-k,-h	Xtriage
F_o,F_c correlation	0.21	EDS
Total number of atoms	9521	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 11.75% 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	1	1.15	2/1145 (0.2%)	1.06	5/1557 (0.3%)
1	2	1.06	1/1059 (0.1%)	1.08	5/1443 (0.3%)
1	3	1.18	1/1119 (0.1%)	1.16	6/1524 (0.4%)
1	4	1.11	2/1165 (0.2%)	1.08	6/1582 (0.4%)
2	F	1.13	3/3454 (0.1%)	1.20	36/4702 (0.8%)
3	G	0.91	1/1372 (0.1%)	1.01	10/1872 (0.5%)
4	B	1.30	2/425 (0.5%)	1.41	9/569 (1.6%)
All	All	1.11	12/9739 (0.1%)	1.14	77/13249 (0.6%)

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	3	0	1
1	4	0	1
2	F	0	1
4	B	0	3
All	All	0	6

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	4	112	GLU	CG-CD	6.46	1.61	1.51
2	F	19	PHE	CB-CG	-6.21	1.40	1.51
2	F	162	CYS	CB-SG	-5.90	1.72	1.81
1	1	140	GLU	CG-CD	5.80	1.60	1.51
4	B	80	GLY	N-CA	5.72	1.54	1.46

The worst 5 of 77 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	B	95	ASP	N-CA-C	-9.58	85.14	111.00
2	F	421	ASP	N-CA-CB	8.85	126.53	110.60
1	1	136	THR	N-CA-C	-8.79	87.27	111.00
1	4	70	ARG	NE-CZ-NH2	7.87	124.23	120.30
1	3	48	ARG	NE-CZ-NH2	7.77	124.18	120.30

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	3	68	TYR	Sidechain
1	4	36	PHE	Sidechain
4	B	107	TYR	Sidechain
4	B	82	THR	Mainchain
2	F	160	PHE	Sidechain

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	1	1125	0	1121	75	0
1	2	1039	0	1031	102	0
1	3	1099	0	1086	95	0
1	4	1145	0	1142	94	0
2	F	3358	0	3242	306	0
3	G	1340	0	1323	64	0
4	B	415	0	376	73	0
All	All	9521	0	9321	742	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 39.

The worst 5 of 742 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)
2:F:7:ALA:CB	4:B:100:ARG:CD	1.79	1.58

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:F:7:ALA:HB3	4:B:100:ARG:CD	1.42	1.41
2:F:7:ALA:CB	4:B:100:ARG:HD2	1.44	1.32
1:4:151:VAL:O	1:4:152:MET:HG2	1.13	1.29
2:F:7:ALA:CB	4:B:100:ARG:HD3	1.51	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	1	141/152 (93%)	125 (89%)	14 (10%)	2 (1%)	11 46
1	2	131/152 (86%)	103 (79%)	20 (15%)	8 (6%)	1 15
1	3	138/152 (91%)	114 (83%)	13 (9%)	11 (8%)	1 10
1	4	144/152 (95%)	112 (78%)	26 (18%)	6 (4%)	3 23
2	F	416/426 (98%)	331 (80%)	47 (11%)	38 (9%)	1 8
3	G	173/175 (99%)	152 (88%)	18 (10%)	3 (2%)	9 42
4	B	45/120 (38%)	25 (56%)	9 (20%)	11 (24%)	0 0
All	All	1188/1329 (89%)	962 (81%)	147 (12%)	79 (7%)	1 13

5 of 79 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	2	111	VAL
1	3	9	VAL
1	3	61	GLY
1	3	117	ALA
1	4	8	SER

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	1	119/127 (94%)	108 (91%)	11 (9%)	9	36
1	2	110/127 (87%)	94 (86%)	16 (14%)	3	18
1	3	117/127 (92%)	103 (88%)	14 (12%)	5	24
1	4	121/127 (95%)	107 (88%)	14 (12%)	5	26
2	F	364/372 (98%)	315 (86%)	49 (14%)	4	21
3	G	153/153 (100%)	133 (87%)	20 (13%)	4	21
4	B	42/101 (42%)	37 (88%)	5 (12%)	5	25
All	All	1026/1134 (90%)	897 (87%)	129 (13%)	4	22

5 of 129 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	G	109	VAL
3	G	136	PRO
1	4	105	GLU
1	4	103	PHE
3	G	149	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 28 such sidechains are listed below:

Mol	Chain	Res	Type
2	F	270	HIS
4	B	110	HIS
2	F	415	ASN
3	G	162	ASN
2	F	405	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 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

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.