



wwPDB X-ray Structure Validation Summary Report ⓘ

Apr 25, 2026 – 07:35 AM EDT

PDB ID : 2FSY / pdb_00002fsy
Title : Bacteriophage HK97 Pepsin-treated Expansion Intermediate IV
Authors : Gan, L.; Speir, J.A.; Conway, J.F.; Lander, G.; Cheng, N.; Firek, B.A.; Hendrix, R.W.; Duda, R.L.; Liljas, L.; Johnson, J.E.
Deposited on : 2006-01-23
Resolution : 3.80 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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity : 4-5-2 with Phenix2.0
Xtrriage (Phenix) : 2.0
EDS : 3.0
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4 : 9.0.010 (Gargrove)
Density-Fitness : 1.0.12
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.49

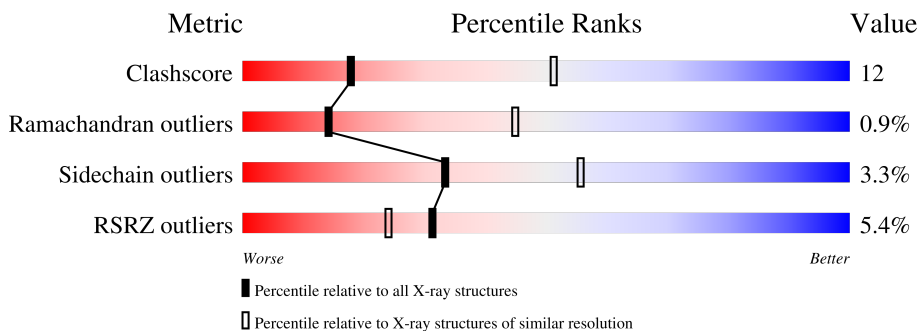
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.80 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.



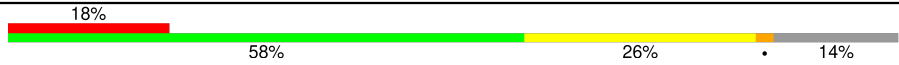
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	190562	1012 (3.94-3.66)
Ramachandran outliers	187476	1048 (3.96-3.64)
Sidechain outliers	187428	1043 (3.96-3.64)
RSRZ outliers	180081	1064 (3.96-3.64)

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 $\leq 5\%$. 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	282	 4% 68% 30% ..
1	B	282	 5% 72% 26% ..
1	C	282	 3% 76% 23% .
1	D	282	 2% 72% 25% ..
1	E	282	 2% 72% 26% ..
1	F	282	 4% 60% 28% . 9%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	282	 <p>A horizontal bar chart representing the quality of chain. The bar is divided into four segments: a red segment (18%), a green segment (58%), a yellow segment (26%), and a grey segment (14%). The percentages are labeled below each segment.</p>

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 14631 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.

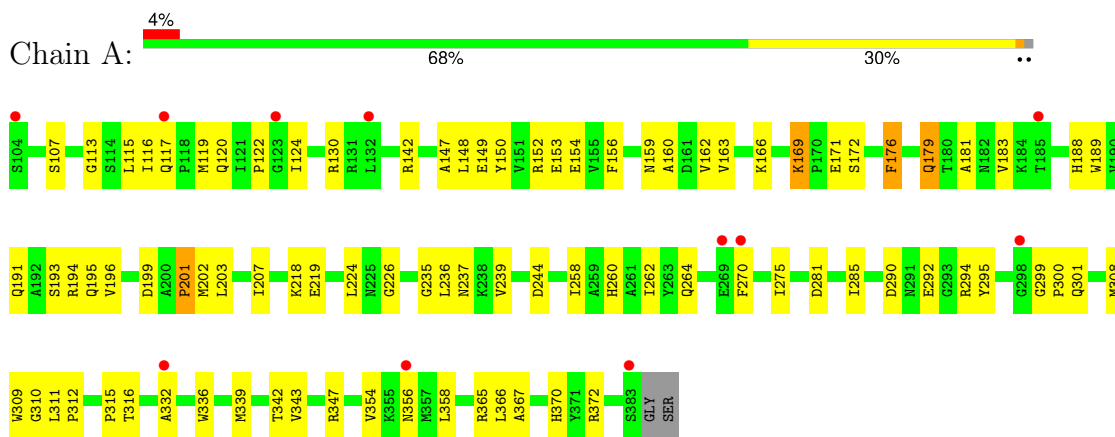
- Molecule 1 is a protein called major capsid protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	280	2151	1344	375	422	10	0	0	0
1	B	280	2151	1344	375	422	10	0	0	0
1	C	280	2151	1344	375	422	10	0	0	0
1	D	280	2151	1344	375	422	10	0	0	0
1	E	280	2151	1344	375	422	10	0	0	0
1	F	256	1986	1241	349	388	8	0	0	0
1	G	243	1890	1181	333	368	8	0	0	0

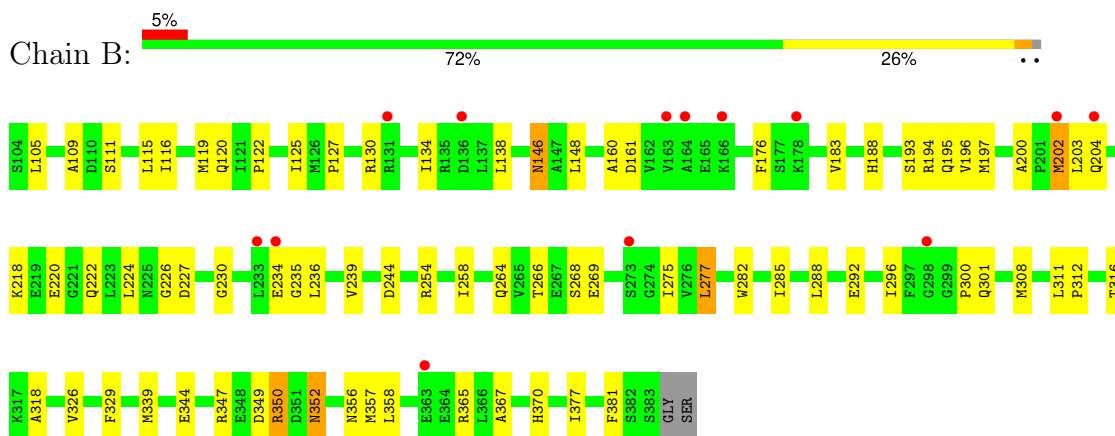
3 Residue-property plots [i](#)

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.

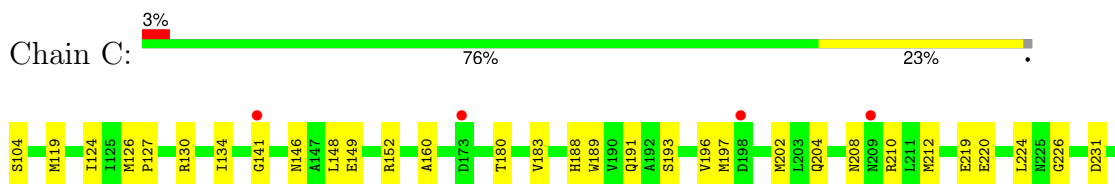
- Molecule 1: major capsid protein



- Molecule 1: major capsid protein



- Molecule 1: major capsid protein



4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α , β , γ	1006.39Å 1006.39Å 728.69Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 3.80 50.00 – 3.80	Depositor EDS
% Data completeness (in resolution range)	59.9 (50.00-3.80) 60.1 (50.00-3.80)	Depositor EDS
R_{merge}	0.17	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.26 (at 3.77Å)	Xtrriage
Refinement program	CNS	Depositor
R, R_{free}	0.337 , (Not available) 0.327 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å ²)	73.2	Xtrriage
Anisotropy	0.772	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 61.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.28	EDS
Total number of atoms	14631	wwPDB-VP
Average B, all atoms (Å ²)	98.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.42% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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	0.32	1/2188 (0.0%)	0.83	3/2969 (0.1%)
1	B	0.36	1/2188 (0.0%)	0.83	3/2969 (0.1%)
1	C	0.33	1/2188 (0.0%)	0.82	3/2969 (0.1%)
1	D	0.34	1/2188 (0.0%)	0.83	3/2969 (0.1%)
1	E	0.31	0/2188	0.86	6/2969 (0.2%)
1	F	0.31	0/2020	0.86	4/2740 (0.1%)
1	G	0.32	0/1922	0.84	2/2605 (0.1%)
All	All	0.33	4/14882 (0.0%)	0.84	24/20190 (0.1%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	356	ASN	CG-ND2	9.38	1.52	1.33
1	D	169	LYS	CE-NZ	7.00	1.70	1.49
1	A	356	ASN	CG-ND2	5.53	1.44	1.33
1	C	356	ASN	CG-ND2	5.16	1.44	1.33

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	299	GLY	CA-C-N	8.77	130.81	119.84
1	E	299	GLY	C-N-CA	8.77	130.81	119.84
1	F	200	ALA	CA-C-N	8.19	127.76	119.24
1	F	200	ALA	C-N-CA	8.19	127.76	119.24
1	D	350	ARG	CB-CA-C	-6.74	108.78	116.54

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	2151	0	2119	60	0
1	B	2151	0	2119	65	0
1	C	2151	0	2119	53	0
1	D	2151	0	2119	58	0
1	E	2151	0	2119	53	0
1	F	1986	0	1951	68	0
1	G	1890	0	1851	54	0
All	All	14631	0	14397	357	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 12.

The worst 5 of 357 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:D:169:LYS:CE	1:D:169:LYS:NZ	1.70	1.50
1:F:224:LEU:HD11	1:F:276:VAL:HG11	1.53	0.89
1:G:345:VAL:HB	1:G:358:LEU:HD21	1.56	0.88
1:B:349:ASP:OD2	1:B:350:ARG:HG2	1.75	0.85
1:G:316:THR:HG22	1:G:318:ALA:H	1.41	0.85

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	A	278/282 (99%)	249 (90%)	27 (10%)	2 (1%)	18	51
1	B	278/282 (99%)	252 (91%)	22 (8%)	4 (1%)	9	37
1	C	278/282 (99%)	258 (93%)	20 (7%)	0	100	100
1	D	278/282 (99%)	256 (92%)	22 (8%)	0	100	100
1	E	278/282 (99%)	246 (88%)	29 (10%)	3 (1%)	11	41
1	F	254/282 (90%)	229 (90%)	21 (8%)	4 (2%)	7	35
1	G	239/282 (85%)	209 (87%)	26 (11%)	4 (2%)	7	34
All	All	1883/1974 (95%)	1699 (90%)	167 (9%)	17 (1%)	14	45

5 of 17 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	300	PRO
1	F	352	ASN
1	G	352	ASN
1	E	330	ASP
1	F	131	ARG

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	230/231 (100%)	218 (95%)	12 (5%)	21	46
1	B	230/231 (100%)	222 (96%)	8 (4%)	32	54
1	C	230/231 (100%)	228 (99%)	2 (1%)	70	74
1	D	230/231 (100%)	224 (97%)	6 (3%)	40	60
1	E	230/231 (100%)	220 (96%)	10 (4%)	26	50
1	F	211/231 (91%)	205 (97%)	6 (3%)	38	58
1	G	201/231 (87%)	194 (96%)	7 (4%)	32	54
All	All	1562/1617 (97%)	1511 (97%)	51 (3%)	33	56

5 of 51 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	115	LEU
1	E	300	PRO
1	G	350	ARG
1	E	116	ILE
1	E	195	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 45 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	291	ASN
1	F	260	HIS
1	E	306	ASN
1	F	158	ASN
1	F	334	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 oligosaccharides 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	280/282 (99%)	0.24	11 (3%) 43 31	43, 89, 136, 236	0
1	B	280/282 (99%)	0.37	13 (4%) 37 27	39, 79, 127, 147	0
1	C	280/282 (99%)	0.26	8 (2%) 53 37	39, 80, 128, 182	0
1	D	280/282 (99%)	0.26	6 (2%) 63 44	44, 80, 123, 162	0
1	E	280/282 (99%)	0.24	5 (1%) 67 46	40, 79, 121, 169	0
1	F	256/282 (90%)	0.18	10 (3%) 43 31	47, 87, 147, 208	0
1	G	243/282 (86%)	1.14	50 (20%) 2 4	81, 167, 255, 321	0
All	All	1899/1974 (96%)	0.37	103 (5%) 31 24	39, 86, 182, 321	0

The worst 5 of 103 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	382	SER	8.1
1	E	383	SER	7.0
1	F	128	GLY	5.2
1	A	269	GLU	5.0
1	G	378	LYS	4.8

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 oligosaccharides 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.