WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Sep 14, 2023 – 10:23 PM EDT

PDB ID	:	2FT1
Title	:	Bacteriophage HK97 Head II
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	:	3.90 Å(reported)

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

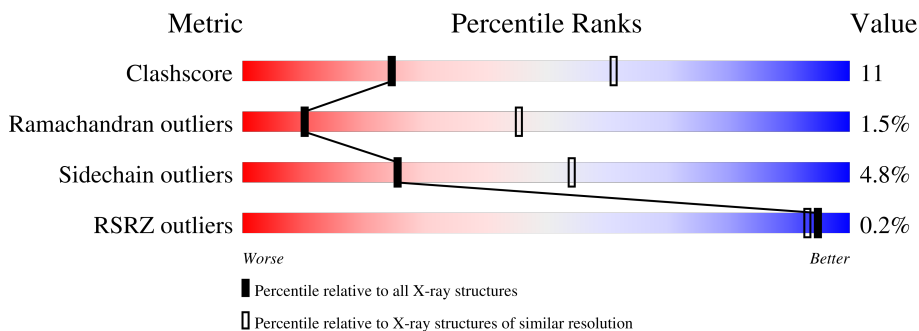|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 3.90 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.
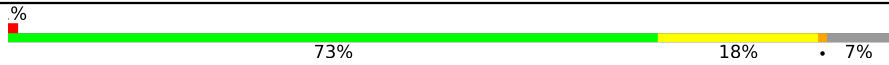
Metric	Percentile Ranks	Value
Clashscore		11
Ramachandran outliers		1.5%
Sidechain outliers		4.8%
RSRZ outliers		0.2%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	141614	1004 (4.12-3.68)
Ramachandran outliers	138981	1021 (4.14-3.66)
Sidechain outliers	138945	1014 (4.14-3.66)
RSRZ outliers	127900	1275 (4.20-3.60)

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	282	71% 26% · ·
1	B	282	76% 21% · ·
1	C	282	74% 22% · ·
1	D	282	72% 26% · ·
1	E	282	75% 22% · ·
1	F	282	62% 26% · 9%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	282	

2 Entry composition ⓘ

There is only 1 type of molecule in this entry. The entry contains 14777 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
1	A	280	Total	C	N	O	S	0	0	0
			2151	1344	375	422	10			
1	B	280	Total	C	N	O	S	0	0	0
			2151	1344	375	422	10			
1	C	280	Total	C	N	O	S	0	0	0
			2151	1344	375	422	10			
1	D	280	Total	C	N	O	S	0	0	0
			2151	1344	375	422	10			
1	E	280	Total	C	N	O	S	0	0	0
			2151	1344	375	422	10			
1	F	258	Total	C	N	O	S	0	0	0
			2001	1251	351	390	9			
1	G	261	Total	C	N	O	S	0	0	0
			2021	1265	354	393	9			

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.
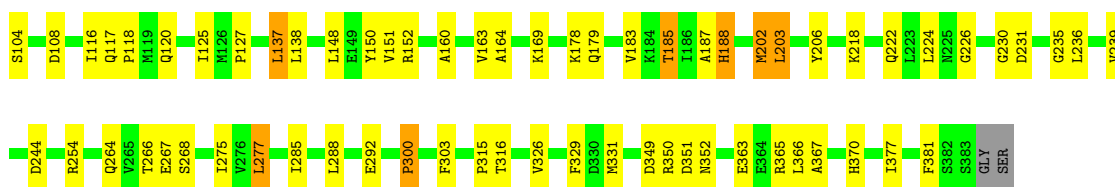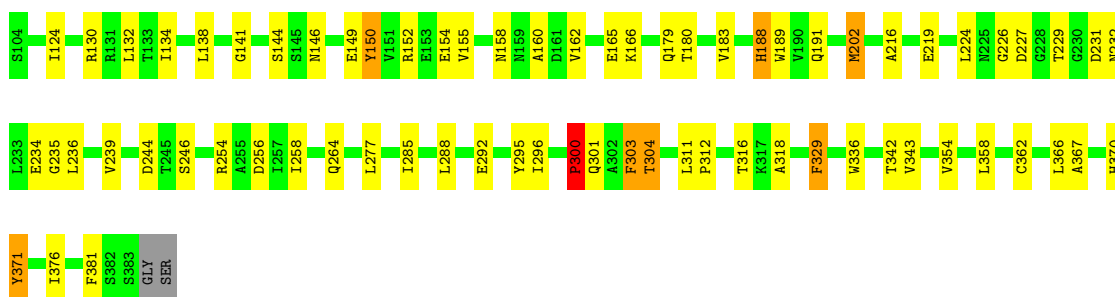
• Molecule 1: major capsid protein
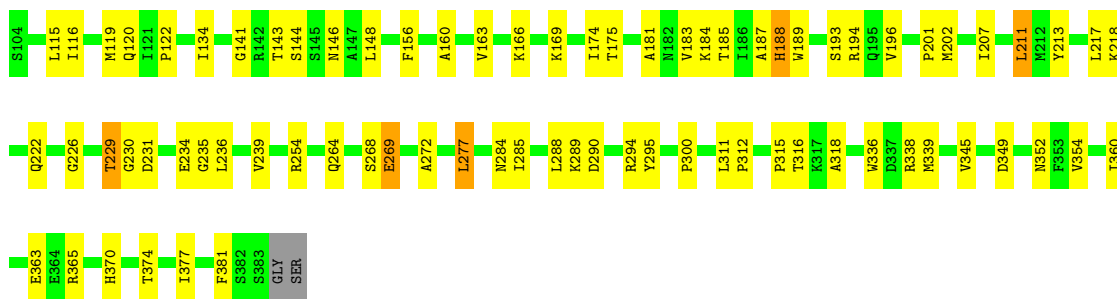
Chain A:

• Molecule 1: major capsid protein

Chain B:

• Molecule 1: major capsid protein

Chain C:

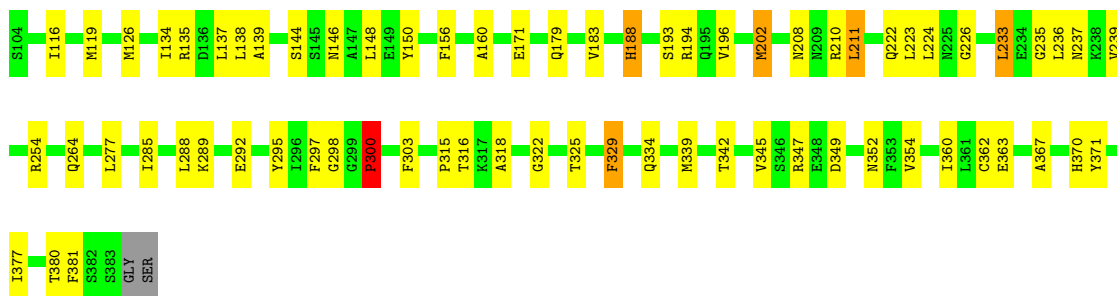• Molecule 1: major capsid protein

Chain D:

72%   26%

S104 L115 I116 M119 Q120 I121 P122 I134 G141 R142 T143 S144 S145 N146 A147 L148 F156 A160 V163 K166 K169 I174 T175 A181 N182 V183 K184 T185 I186 A187 H188 W189 S193 R194 Q195 V196 P201 M202 I207 L211 M212 Y213 L217 K218

Q222 G226 T229 G230 D231 E234 G235 L236 V239 R254 Q264 S268 E269 A272 L277 N284 I285 L288 K289 D290 R294 Y295 P300 L311 P312 P315 T316 K317 A318 W336 D337 R338 M339 V345 D349 N352 F353 V354 I360

E363 E364 R365 H370 T374 I377 F381 S382 S383 GLY SER

• Molecule 1: major capsid protein

Chain E:

75%   22%

S104 I116 M119 M126 I134 R135 D136 L137 L138 A139 S144 S145 N146 A147 L148 E149 Y150 F156 A160 E171 Q179 V183 H188 S193 R194 Q195 V196 M202 N208 N209 R210 L211 Q222 L223 L224 N225 G226 L233 E234 G235 L236 N237 K238 V239

R254 Q264 L277 I285 L288 K289 E292 Y295 I296 F297 G298 G299 P300 F303 P315 T316 K317 A318 G322 T325 F329 Q334 M339 T342 V345 S346 R347 E348 D349 N352 F353 V354 I360 L361 C362 E363 A367 H370 Y371

I377 T380 F381 S382 S383 GLY SER

• Molecule 1: major capsid protein

Chain F:

62%   26%   •   9%

SER LEU GLY SER ASP ALA ASP SER ALA GLY SER LEU ILE GLN PRO MET GLN ILE PRO GLY ILE ILE M126 P127 R130 R131 L132 T133 I134 R135 D136 L137 L138 A139 Q140 G141 R142 N146 A147 L148 E149 Y150 F156 T157 A160 Q179 V183 H188 S193 R194 Q195 V196

P201 M202 R210 L215 A216 L224 N225 G226 D227 G228 T229 G230 D231 G235 L236 N237 V239 D244 D252 T253 R254 A255 D256 Q264 V265 T266 E267 S268 E269 L277 D281 W282 H283 N284 I285 L288 K289 D290 N291 E292 G293 R294 G299 P300 Q301 A302 F303

N306 V314 P315 T316 K317 A318 F324 F329 Q334 W335 W336 T342 V343 E344 V345 D349 R350 D351 N352 F353 I360 L361 A367 H370 Y371 I376 I377 K378 F381 S382 S383 GLY SER

• Molecule 1: major capsid protein

Chain G:

%

73%   18%   •   7%

SER LEU GLY SER ASP ALA ASP SER ALA GLY SER LEU ILE GLN PRO MET GLN ILE PRO G123 I124 I125 M126 L129 A139 R142 R152 V155 T175 Q179 V183 H188 W189 V190 Q191 R194 Q195 L203 Y206 I207 L215 A216 G226 D227 G228 T229

G230 D231 G235 L236 D252 T253 R254 A255 D256 I257 I258 I262 Y263 Q264 S273 I285 L288 K289 Y295 Q301 L311 P312 T316 K317 A318 Q334 M339 T342 V343 D349 N352 F353 V354 R365 L366 A367 H370 Y371 R372 I377 F381

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	1009.80Å 1009.80Å 732.87Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 3.90 30.00 – 3.90	Depositor EDS
% Data completeness (in resolution range)	52.1 (30.00-3.90) 52.1 (30.00-3.90)	Depositor EDS
R_{merge}	0.20	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.47 (at 3.86Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.297 , (Not available) 0.279 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	48.5	Xtriage
Anisotropy	0.201	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.26 , 45.1	EDS
L-test for twinning[2]	$<|L|> = 0.40$, $<L^2> = 0.22$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.76	EDS
Total number of atoms	14777	wwPDB-VP
Average B, all atoms (Å^2)	47.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.44% 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 (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	0.26	0/2188	0.52	0/2969
1	B	0.25	0/2188	0.51	0/2969
1	C	0.26	0/2188	0.52	0/2969
1	D	0.27	0/2188	0.51	0/2969
1	E	0.25	0/2188	0.53	0/2969
1	F	0.26	0/2036	0.52	0/2762
1	G	0.28	0/2056	0.51	0/2789
All	All	0.26	0/15032	0.52	0/20396

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 (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	48	0
1	C	2151	0	2119	61	0
1	D	2151	0	2119	45	0
1	E	2151	0	2119	50	0
1	F	2001	0	1967	64	0
1	G	2021	0	1992	34	0
All	All	14777	0	14554	320	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 11.

The worst 5 of 320 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:229:THR:HG23	1:D:230:GLY:H	1.35	0.91
1:F:227:ASP:HB3	1:F:229:THR:HG22	1.54	0.90
1:E:210:ARG:HH21	1:F:156:PHE:HD2	1.20	0.90
1:E:116:ILE:HG21	1:F:146:ASN:HB2	1.59	0.84
1:D:207:ILE:HG23	1:D:211:LEU:HD23	1.63	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	278/282 (99%)	243 (87%)	32 (12%)	3 (1%)	14	51
1	B	278/282 (99%)	243 (87%)	31 (11%)	4 (1%)	11	46
1	C	278/282 (99%)	246 (88%)	26 (9%)	6 (2%)	6	38
1	D	278/282 (99%)	248 (89%)	25 (9%)	5 (2%)	8	42
1	E	278/282 (99%)	252 (91%)	21 (8%)	5 (2%)	8	42
1	F	256/282 (91%)	229 (90%)	22 (9%)	5 (2%)	7	40
1	G	259/282 (92%)	230 (89%)	29 (11%)	0	100	100
All	All	1905/1974 (96%)	1691 (89%)	186 (10%)	28 (2%)	10	45

5 of 28 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	105	LEU
1	C	134	ILE

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	304	THR
1	D	229	THR
1	E	134	ILE

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	230/231 (100%)	214 (93%)	16 (7%)	15	44
1	B	230/231 (100%)	218 (95%)	12 (5%)	23	53
1	C	230/231 (100%)	220 (96%)	10 (4%)	29	57
1	D	230/231 (100%)	220 (96%)	10 (4%)	29	57
1	E	230/231 (100%)	220 (96%)	10 (4%)	29	57
1	F	213/231 (92%)	202 (95%)	11 (5%)	23	53
1	G	215/231 (93%)	208 (97%)	7 (3%)	38	63
All	All	1578/1617 (98%)	1502 (95%)	76 (5%)	25	54

5 of 76 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	354	VAL
1	G	188	HIS
1	F	150	TYR
1	F	202	MET
1	G	372	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 29 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	191	GLN
1	G	284	ASN
1	E	334	GLN
1	G	191	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	E	237	ASN

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 ⓘ

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	280/282 (99%)	-0.67	0	100	100	1, 42, 106, 140	0
1	B	280/282 (99%)	-0.74	0	100	100	1, 36, 96, 150	0
1	C	280/282 (99%)	-0.71	0	100	100	1, 37, 100, 154	0
1	D	280/282 (99%)	-0.71	0	100	100	1, 37, 94, 133	0
1	E	280/282 (99%)	-0.76	0	100	100	1, 35, 93, 193	0
1	F	258/282 (91%)	-0.68	0	100	100	2, 38, 106, 201	0
1	G	261/282 (92%)	-0.60	4 (1%)	73	64	2, 53, 136, 214	0
All	All	1919/1974 (97%)	-0.70	4 (0%)	95	93	1, 39, 105, 214	0

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	123	GLY	6.0
1	G	124	ILE	4.3
1	G	125	ILE	2.5
1	G	383	SER	2.0

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.