



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 13, 2020 – 04:14 am BST

PDB ID : 2FRP  
Title : Bacteriophage HK97 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 : 2006-01-19  
Resolution : 7.50 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](mailto: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

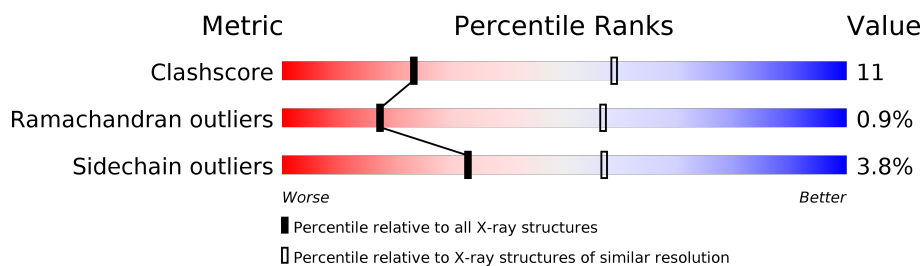
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 7.50 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	1069 (10.00-3.90)
Ramachandran outliers	138981	1002 (10.00-3.90)
Sidechain outliers	138945	1002 (10.00-3.86)

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  $\geq 3$ , 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  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	A	282	70% 28% ..
1	B	282	74% 23% ..
1	C	282	78% 21% .
1	D	282	76% 22% ..
1	E	282	74% 25% ..
1	F	282	63% 25% . 9%
1	G	282	60% 23% . 14%

## 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.

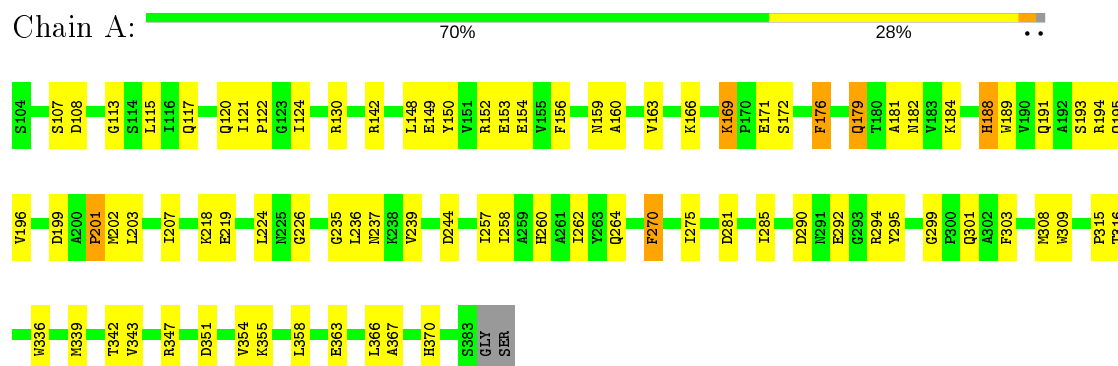
- Molecule 1 is a protein called Major capsid protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
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	256	Total	C	N	O	S	0	0	0
			1986	1241	349	388	8			
1	G	243	Total	C	N	O	S	0	0	0
			1890	1181	333	368	8			

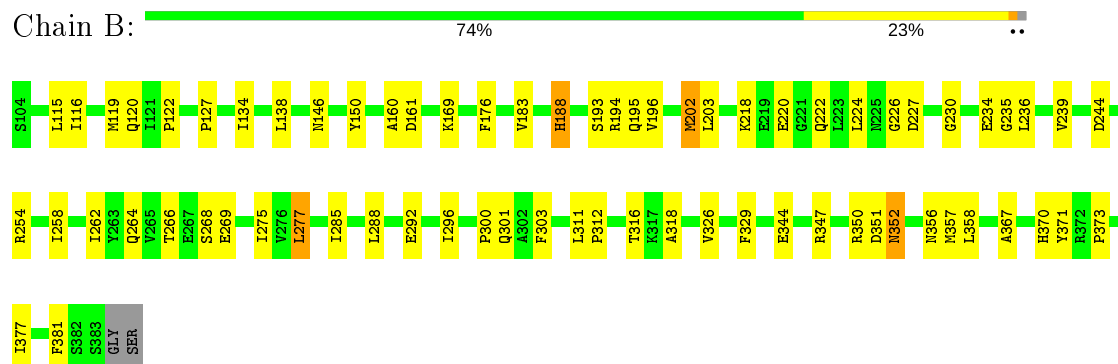
### 3 Residue-property plots

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. 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.

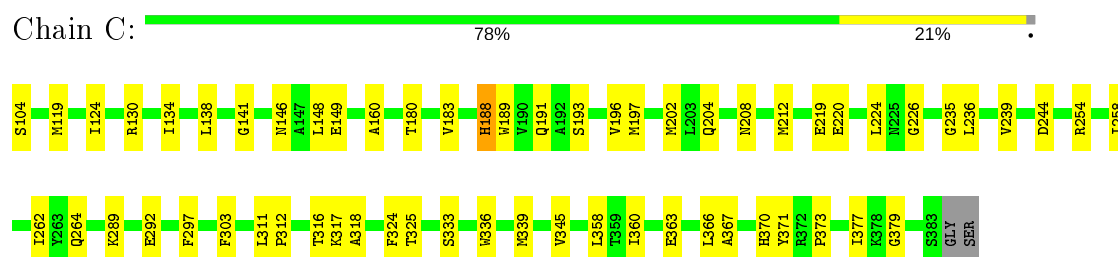
- Molecule 1: Major capsid protein



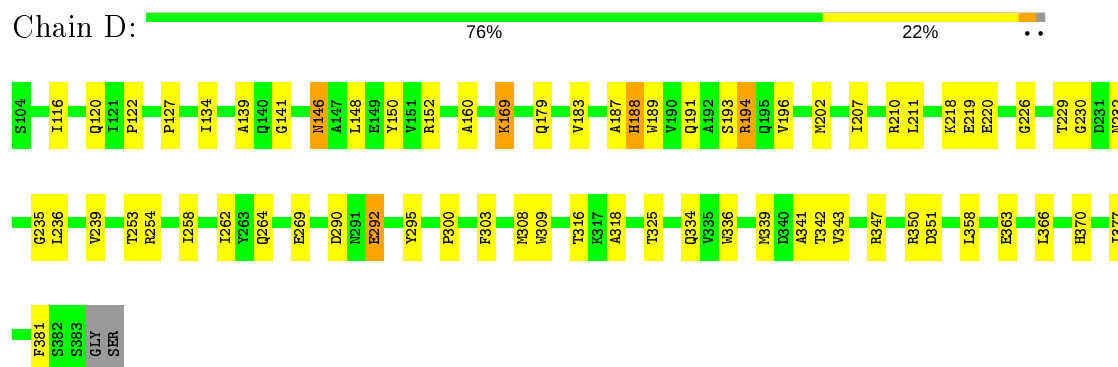
- Molecule 1: Major capsid protein



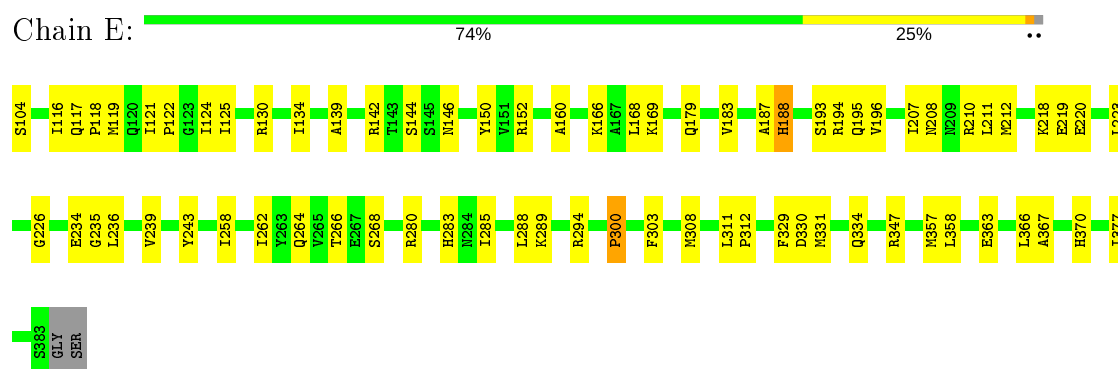
- Molecule 1: Major capsid protein



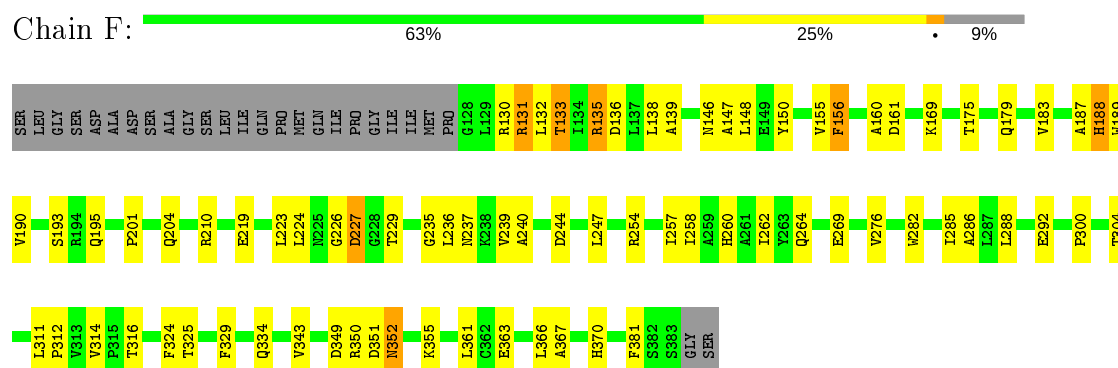
- Molecule 1: Major capsid protein



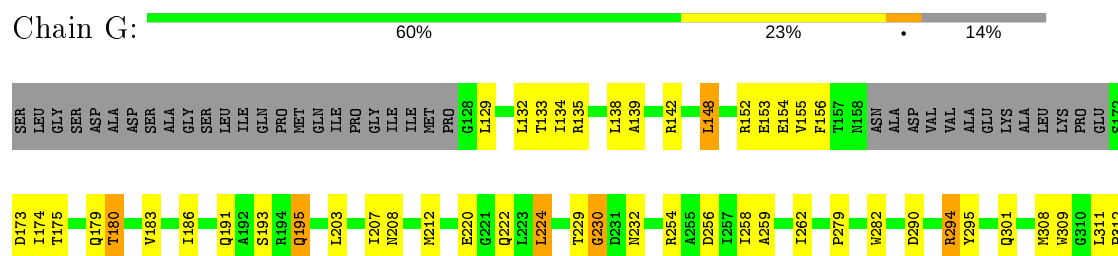
- Molecule 1: Major capsid protein



- 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, $\alpha$ , $\beta$ , $\gamma$	1009.80 Å   1009.80 Å   732.87 Å 90.00°   90.00°   90.00°	Depositor
Resolution (Å)	40.00 – 7.50 40.16 – 7.51	Depositor EDS
% Data completeness (in resolution range)	64.7 (40.00-7.50) 64.9 (40.16-7.51)	Depositor EDS
$R_{merge}$	0.20	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.21 (at 7.33 Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.421 , (Not available) 0.363 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	152.8	Xtriage
Anisotropy	0.763	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 176.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.40$ , $\langle L^2 \rangle = 0.22$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.13	EDS
Total number of atoms	14631	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	60.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.61% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>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	0.30	0/2188	0.50	0/2969
1	B	0.32	1/2188 (0.0%)	0.49	0/2969
1	C	0.28	0/2188	0.48	0/2969
1	D	0.34	1/2188 (0.0%)	0.49	0/2969
1	E	0.28	0/2188	0.51	0/2969
1	F	0.28	0/2020	0.50	0/2740
1	G	0.49	1/1922 (0.1%)	0.63	4/2605 (0.2%)
All	All	0.33	3/14882 (0.0%)	0.51	4/20190 (0.0%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	G	180	THR	C-N	-17.02	0.94	1.34
1	D	169	LYS	CE-NZ	8.56	1.70	1.49
1	B	356	ASN	CG-ND2	8.00	1.52	1.32

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	G	148	LEU	O-C-N	12.69	143.00	122.70
1	G	148	LEU	CA-C-N	-9.85	95.54	117.20
1	G	148	LEU	C-N-CA	-7.34	103.35	121.70
1	G	180	THR	O-C-N	-5.36	114.13	122.70

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	57	0
1	B	2151	0	2119	54	0
1	C	2151	0	2119	48	0
1	D	2151	0	2119	52	0
1	E	2151	0	2119	51	0
1	F	1986	0	1951	60	0
1	G	1890	0	1850	51	0
All	All	14631	0	14396	327	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 327 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	1.54
1:F:224:LEU:HD11	1:F:276:VAL:HG11	1.52	0.89
1:G:138:LEU:HD23	1:G:329:PHE:HB3	1.56	0.88
1:G:345:VAL:HB	1:G:358:LEU:HD21	1.56	0.87
1:A:218:LYS:HE2	1:B:161:ASP:OD1	1.77	0.84

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%)	22 63
1	B	278/282 (99%)	252 (91%)	22 (8%)	4 (1%)	11 46
1	C	278/282 (99%)	258 (93%)	20 (7%)	0	100 100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	D	278/282 (99%)	256 (92%)	22 (8%)	0	100	100
1	E	278/282 (99%)	246 (88%)	29 (10%)	3 (1%)	14	52
1	F	254/282 (90%)	229 (90%)	21 (8%)	4 (2%)	9	44
1	G	239/282 (85%)	208 (87%)	27 (11%)	4 (2%)	9	42
All	All	1883/1974 (95%)	1698 (90%)	168 (9%)	17 (1%)	17	57

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	G	224	LEU

### 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	40
1	B	230/231 (100%)	221 (96%)	9 (4%)	32	56
1	C	230/231 (100%)	226 (98%)	4 (2%)	60	78
1	D	230/231 (100%)	223 (97%)	7 (3%)	41	63
1	E	230/231 (100%)	223 (97%)	7 (3%)	41	63
1	F	211/231 (91%)	205 (97%)	6 (3%)	43	65
1	G	201/231 (87%)	190 (94%)	11 (6%)	21	47
All	All	1562/1617 (97%)	1502 (96%)	60 (4%)	33	57

5 of 60 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	224	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	D	202	MET
1	G	294	ARG
1	D	152	ARG
1	D	303	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 35 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	352	ASN
1	D	334	GLN
1	G	237	ASN
1	D	191	GLN
1	D	222	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](#)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	G	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	G	180:THR	C	181:ALA	N	0.94

## 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.