



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 17, 2026 – 11:27 PM UTC

PDB ID : 7F8I / pdb_00007f8i
Title : Crystal structure of HPV6 L1 pentamer
Authors : Wang, Z.P.; Wang, D.N.; Gu, Y.; Li, S.W.
Deposited on : 2021-07-02
Resolution : 3.37 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
Xtriage (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

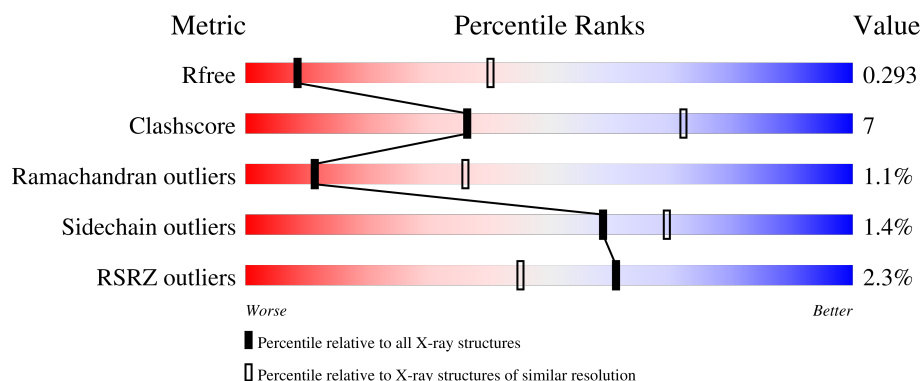
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.37 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}	180053	1099 (3.40-3.32)
Clashscore	190562	1116 (3.40-3.32)
Ramachandran outliers	187476	1101 (3.40-3.32)
Sidechain outliers	187428	1101 (3.40-3.32)
RSRZ outliers	180081	1099 (3.40-3.32)

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	496	0% 67% 16% • 16%
1	B	496	2% 70% 12% • 17%
1	C	496	2% 68% 15% • 16%
1	D	496	2% 68% 15% 17%
1	E	496	3% 67% 16% • 16%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	496	3% 67% 15% • 18%
1	G	496	2% 69% 15% • 16%
1	H	496	% 68% 16% • 16%
1	I	496	2% 65% 16% • 18%
1	J	496	% 67% 15% • 16%

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 32368 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 L1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	416	Total	C	N	O	S	0	0	0
			3258	2069	548	622	19			
1	B	410	Total	C	N	O	S	0	0	0
			3213	2042	540	612	19			
1	C	416	Total	C	N	O	S	0	0	0
			3258	2069	548	622	19			
1	D	410	Total	C	N	O	S	0	0	0
			3216	2044	541	613	18			
1	E	416	Total	C	N	O	S	0	0	0
			3258	2069	548	622	19			
1	F	407	Total	C	N	O	S	0	0	0
			3194	2031	536	609	18			
1	G	418	Total	C	N	O	S	0	0	0
			3274	2080	550	625	19			
1	H	416	Total	C	N	O	S	0	0	0
			3258	2069	548	622	19			
1	I	405	Total	C	N	O	S	0	0	0
			3181	2025	534	604	18			
1	J	416	Total	C	N	O	S	0	0	0
			3258	2069	548	622	19			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	MET	-	initiating methionine	UNP Q9W9C6
A	376	VAL	MET	conflict	UNP Q9W9C6
B	-2	MET	-	initiating methionine	UNP Q9W9C6
B	376	VAL	MET	conflict	UNP Q9W9C6
C	-2	MET	-	initiating methionine	UNP Q9W9C6
C	376	VAL	MET	conflict	UNP Q9W9C6
D	-2	MET	-	initiating methionine	UNP Q9W9C6
D	376	VAL	MET	conflict	UNP Q9W9C6
E	-2	MET	-	initiating methionine	UNP Q9W9C6

Continued on next page...

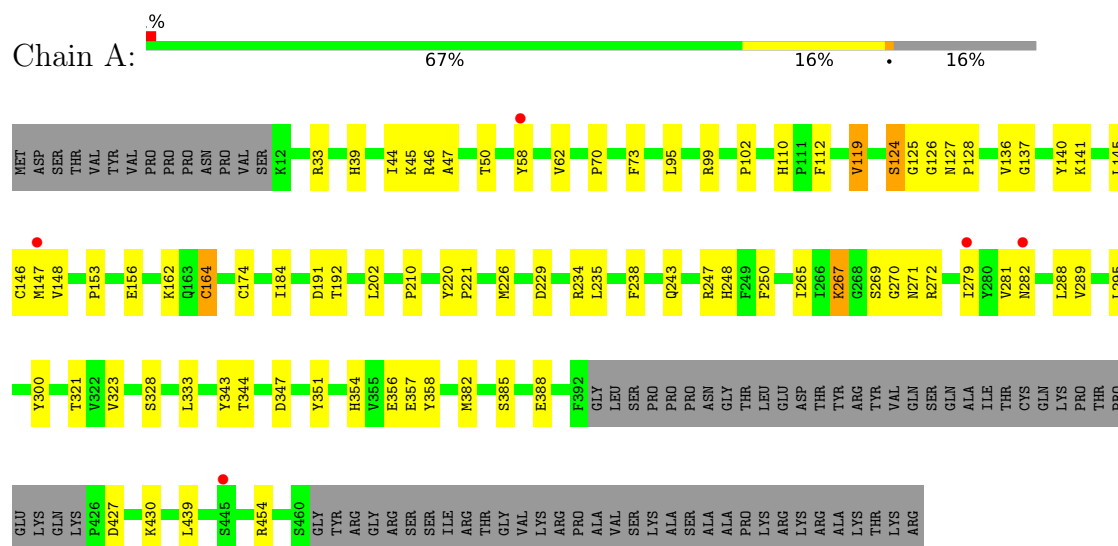
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	376	VAL	MET	conflict	UNP Q9W9C6
F	-2	MET	-	initiating methionine	UNP Q9W9C6
F	376	VAL	MET	conflict	UNP Q9W9C6
G	-2	MET	-	initiating methionine	UNP Q9W9C6
G	376	VAL	MET	conflict	UNP Q9W9C6
H	-2	MET	-	initiating methionine	UNP Q9W9C6
H	376	VAL	MET	conflict	UNP Q9W9C6
I	-2	MET	-	initiating methionine	UNP Q9W9C6
I	376	VAL	MET	conflict	UNP Q9W9C6
J	-2	MET	-	initiating methionine	UNP Q9W9C6
J	376	VAL	MET	conflict	UNP Q9W9C6

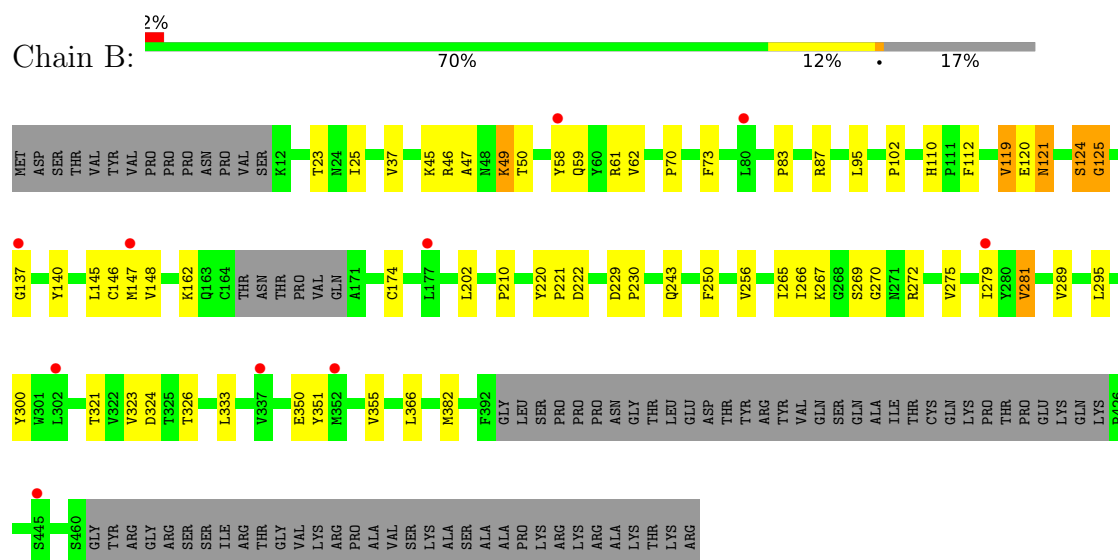
3 Residue-property plots [i](#)

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 L1



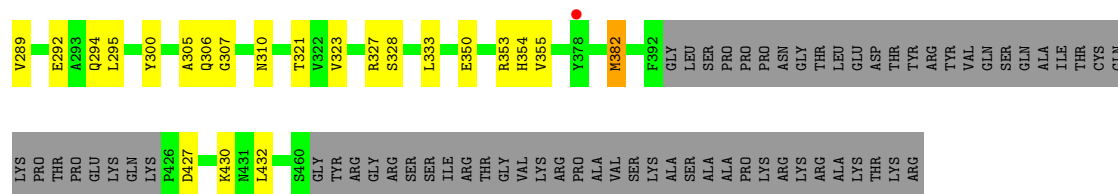
• Molecule 1: Major capsid protein L1



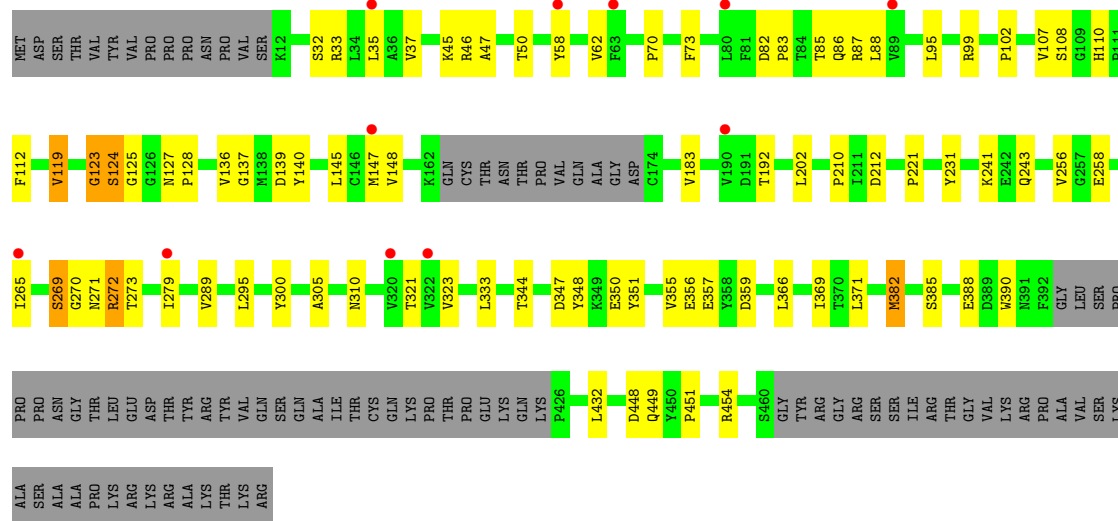
• Molecule 1: Major capsid protein L1



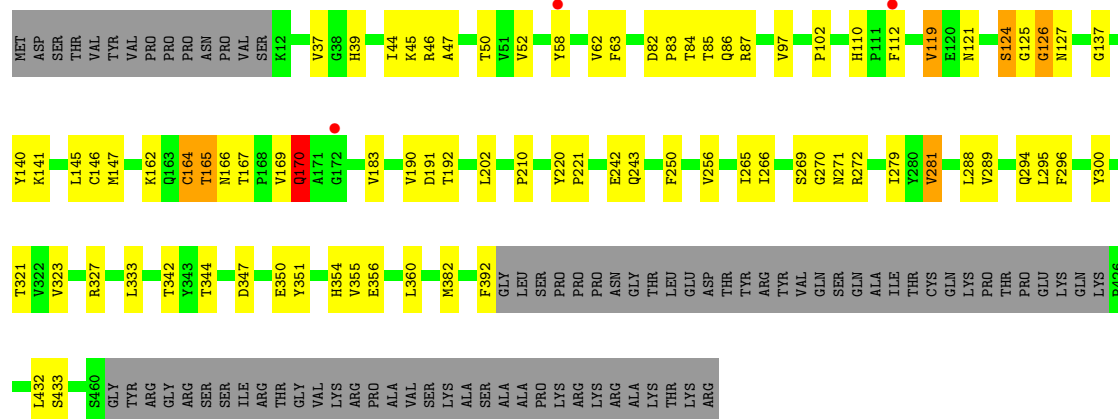
Met	ASP	SER	THR	VAL	TYR	VAL	PRO	PRO	ASN	PRO	VAL	SER	K12	V37	I44	K46	R46	A47	T50	V51	V52	V58	V62	P70	F73	L80	R87	V97	G98	R99	H110	V119	G123	S124	G125	G126	N127	P128	G129	Q130	G137	V140	V441
	L145	C146	M147	K160	G161	K162	Q163	C164	T165	V169	Q170	G174	P175	P176	V183	D191	D201	L202	N205	P210	Y220	P221	Y231	L235	Q243	F250	V256	L265	I266	K267	G270	M271	R272	V275	G276	S277	S278	I279	V281	N282			



• Molecule 1: Major capsid protein L1



• Molecule 1: Major capsid protein L1



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	308.96Å 106.66Å 198.71Å 90.00° 125.96° 90.00°	Depositor
Resolution (Å)	33.88 – 3.37 33.88 – 3.37	Depositor EDS
% Data completeness (in resolution range)	94.4 (33.88-3.37) 94.3 (33.88-3.37)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.11	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.35 (at 3.39Å)	Xtriage
Refinement program	PHENIX 1.15.2_3472	Depositor
R, R_{free}	0.256 , 0.294 0.254 , 0.293	Depositor DCC
R_{free} test set	3592 reflections (4.82%)	wwPDB-VP
Wilson B-factor (Å ²)	95.6	Xtriage
Anisotropy	0.203	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.25 , 40.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.89	EDS
Total number of atoms	32368	wwPDB-VP
Average B, all atoms (Å ²)	105.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.01% 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

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	0.13	0/3342	0.38	0/4544
1	B	0.13	0/3295	0.37	0/4476
1	C	0.11	0/3342	0.35	0/4544
1	D	0.11	0/3298	0.34	0/4480
1	E	0.11	0/3342	0.36	0/4544
1	F	0.11	0/3276	0.34	0/4452
1	G	0.12	0/3359	0.36	0/4567
1	H	0.13	0/3342	0.36	0/4544
1	I	0.13	0/3263	0.36	2/4433 (0.0%)
1	J	0.11	0/3342	0.35	0/4544
All	All	0.12	0/33201	0.36	2/45128 (0.0%)

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	F	0	1
1	H	0	1
All	All	0	2

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	I	124	SER	CA-C-N	5.10	130.87	121.70
1	I	124	SER	C-N-CA	5.10	130.87	121.70

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	F	125	GLY	Peptide
1	H	123	GLY	Peptide

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	3258	0	3150	64	0
1	B	3213	0	3105	52	0
1	C	3258	0	3150	58	0
1	D	3216	0	3108	60	0
1	E	3258	0	3150	59	0
1	F	3194	0	3084	60	0
1	G	3274	0	3162	54	0
1	H	3258	0	3150	53	0
1	I	3181	0	3080	59	0
1	J	3258	0	3150	55	0
All	All	32368	0	31289	461	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 461 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:119:VAL:HG11	1:D:280:TYR:HE1	1.43	0.83
1:D:350:GLU:HB2	1:E:279:ILE:HD13	1.59	0.83
1:D:279:ILE:HD13	1:G:350:GLU:HB2	1.65	0.78
1:H:350:GLU:HB2	1:I:279:ILE:HD13	1.66	0.78
1:I:83:PRO:O	1:I:87:ARG:NH1	2.15	0.78

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	412/496 (83%)	389 (94%)	19 (5%)	4 (1%)	12	38
1	B	404/496 (82%)	385 (95%)	15 (4%)	4 (1%)	12	38
1	C	412/496 (83%)	391 (95%)	15 (4%)	6 (2%)	8	29
1	D	404/496 (82%)	387 (96%)	15 (4%)	2 (0%)	24	52
1	E	412/496 (83%)	390 (95%)	18 (4%)	4 (1%)	12	38
1	F	401/496 (81%)	382 (95%)	15 (4%)	4 (1%)	12	38
1	G	414/496 (84%)	391 (94%)	19 (5%)	4 (1%)	12	38
1	H	412/496 (83%)	393 (95%)	15 (4%)	4 (1%)	12	38
1	I	399/496 (80%)	382 (96%)	12 (3%)	5 (1%)	9	32
1	J	412/496 (83%)	388 (94%)	17 (4%)	7 (2%)	7	27
All	All	4082/4960 (82%)	3878 (95%)	160 (4%)	44 (1%)	11	36

5 of 44 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	382	MET
1	F	126	GLY
1	I	269	SER
1	I	271	ASN
1	J	126	GLY

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	363/432 (84%)	360 (99%)	3 (1%)	73	76
1	B	357/432 (83%)	350 (98%)	7 (2%)	48	65
1	C	363/432 (84%)	358 (99%)	5 (1%)	59	69
1	D	357/432 (83%)	353 (99%)	4 (1%)	65	73
1	E	363/432 (84%)	358 (99%)	5 (1%)	59	69
1	F	355/432 (82%)	353 (99%)	2 (1%)	78	79
1	G	364/432 (84%)	358 (98%)	6 (2%)	55	67
1	H	363/432 (84%)	357 (98%)	6 (2%)	53	67
1	I	354/432 (82%)	350 (99%)	4 (1%)	65	73
1	J	363/432 (84%)	356 (98%)	7 (2%)	50	66
All	All	3602/4320 (83%)	3553 (99%)	49 (1%)	59	69

5 of 49 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	170	GLN
1	H	278	SER
1	G	267	LYS
1	H	119	VAL
1	H	282	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	F	144	GLN
1	F	248	HIS
1	J	142	GLN
1	I	185	GLN
1	I	365	GLN

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 [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 ⓘ

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	416/496 (83%)	-0.06	5 (1%) 76 63	47, 73, 124, 205	0
1	B	410/496 (82%)	0.02	10 (2%) 59 44	56, 87, 150, 194	0
1	C	416/496 (83%)	-0.04	8 (1%) 66 49	67, 99, 168, 259	0
1	D	410/496 (82%)	0.06	12 (2%) 53 39	64, 88, 139, 190	0
1	E	416/496 (83%)	0.05	15 (3%) 46 33	70, 93, 152, 222	0
1	F	407/496 (82%)	0.08	17 (4%) 40 28	67, 117, 175, 226	0
1	G	418/496 (84%)	0.06	9 (2%) 62 46	68, 104, 163, 232	0
1	H	416/496 (83%)	0.07	7 (1%) 69 53	62, 103, 161, 236	0
1	I	405/496 (81%)	0.27	11 (2%) 56 41	73, 126, 190, 225	0
1	J	416/496 (83%)	0.01	3 (0%) 84 73	75, 105, 162, 213	0
All	All	4130/4960 (83%)	0.05	97 (2%) 61 45	47, 99, 164, 259	0

The worst 5 of 97 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	147	MET	5.6
1	H	58	TYR	5.3
1	B	58	TYR	4.6
1	F	58	TYR	4.4
1	E	58	TYR	4.1

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 [i](#)

There are no oligosaccharides in this entry.

6.4 Ligands [i](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.