



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 06:39 am GMT

PDB ID : 5J6R
Title : Crystal structure of Human Papillomavirus Type 59 L1 pentamer
Authors : Li, Z.H.; Yan, X.D.; Yu, H.; Gu, Y.; Li, S.W.
Deposited on : 2016-04-05
Resolution : 4.01 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

<http://wwpdb.org/validation/2016/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)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

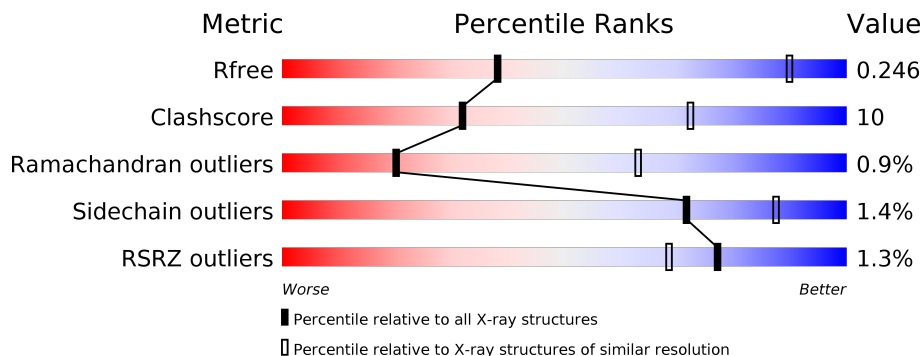
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 4.01 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}	100719	1089 (4.42-3.60)
Clashscore	112137	1189 (4.44-3.60)
Ramachandran outliers	110173	1140 (4.42-3.60)
Sidechain outliers	110143	1127 (4.42-3.60)
RSRZ outliers	101464	1100 (4.42-3.60)

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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. 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	500	
1	B	500	
1	C	500	
1	D	500	
1	E	500	
1	F	500	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	500	<div><div>%</div><div><div></div><div>62%</div><div>21%</div><div>•</div><div>16%</div></div></div>
1	H	500	<div><div>%</div><div><div></div><div>61%</div><div>21%</div><div>•</div><div>17%</div></div></div>
1	I	500	<div><div>%</div><div><div></div><div>63%</div><div>20%</div><div>•</div><div>16%</div></div></div>
1	J	500	<div><div>%</div><div><div></div><div>63%</div><div>20%</div><div>•</div><div>16%</div></div></div>

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 33270 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	420	Total	C	N	O	S	0	0	0
			3334	2119	561	637	17			
1	B	416	Total	C	N	O	S	0	0	0
			3305	2101	556	631	17			
1	C	420	Total	C	N	O	S	0	0	0
			3334	2119	561	637	17			
1	D	420	Total	C	N	O	S	0	0	0
			3334	2119	561	637	17			
1	E	420	Total	C	N	O	S	0	0	0
			3334	2119	561	637	17			
1	F	420	Total	C	N	O	S	0	0	0
			3334	2119	561	637	17			
1	G	420	Total	C	N	O	S	0	0	0
			3334	2119	561	637	17			
1	H	414	Total	C	N	O	S	0	0	0
			3293	2094	554	628	17			
1	I	420	Total	C	N	O	S	0	0	0
			3334	2119	561	637	17			
1	J	420	Total	C	N	O	S	0	0	0
			3334	2119	561	637	17			

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	9	MET	-	initiating methionine	UNP Q81971
A	175	SER	CYS	engineered mutation	UNP Q81971
B	9	MET	-	initiating methionine	UNP Q81971
B	175	SER	CYS	engineered mutation	UNP Q81971
C	9	MET	-	initiating methionine	UNP Q81971
C	175	SER	CYS	engineered mutation	UNP Q81971
D	9	MET	-	initiating methionine	UNP Q81971
D	175	SER	CYS	engineered mutation	UNP Q81971
E	9	MET	-	initiating methionine	UNP Q81971

Continued on next page...

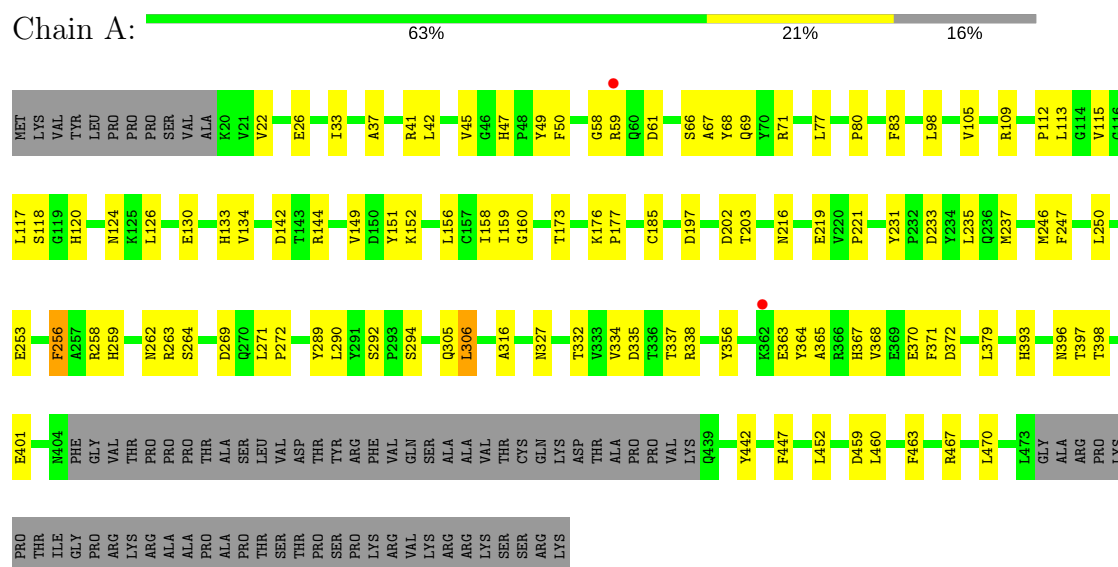
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	175	SER	CYS	engineered mutation	UNP Q81971
F	9	MET	-	initiating methionine	UNP Q81971
F	175	SER	CYS	engineered mutation	UNP Q81971
G	9	MET	-	initiating methionine	UNP Q81971
G	175	SER	CYS	engineered mutation	UNP Q81971
H	9	MET	-	initiating methionine	UNP Q81971
H	175	SER	CYS	engineered mutation	UNP Q81971
I	9	MET	-	initiating methionine	UNP Q81971
I	175	SER	CYS	engineered mutation	UNP Q81971
J	9	MET	-	initiating methionine	UNP Q81971
J	175	SER	CYS	engineered mutation	UNP Q81971

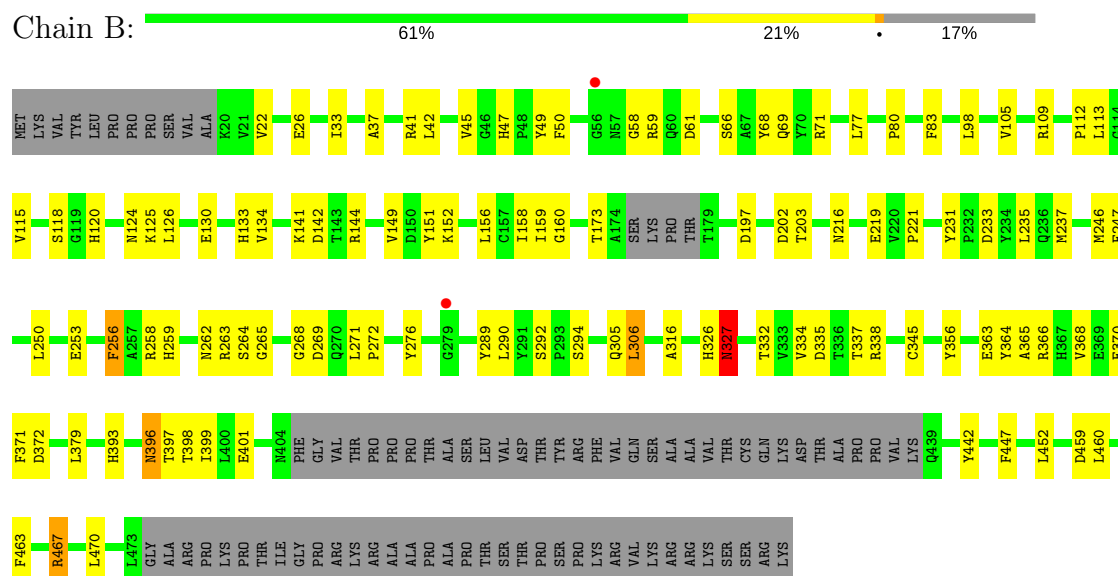
3 Residue-property plots

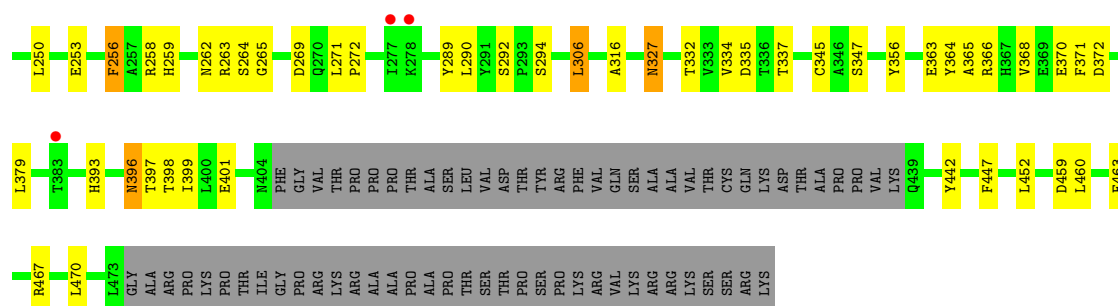
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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

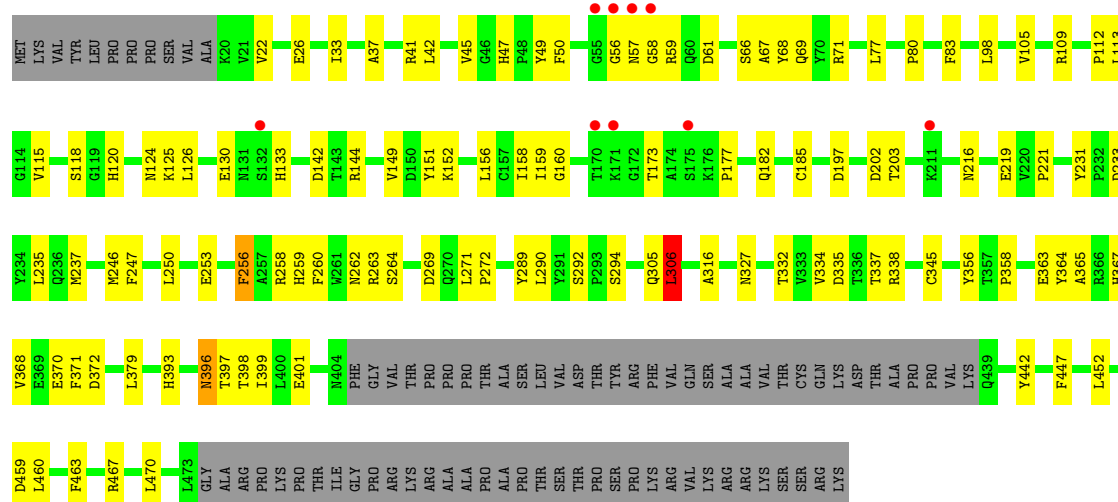


- Molecule 1: Major capsid protein L1

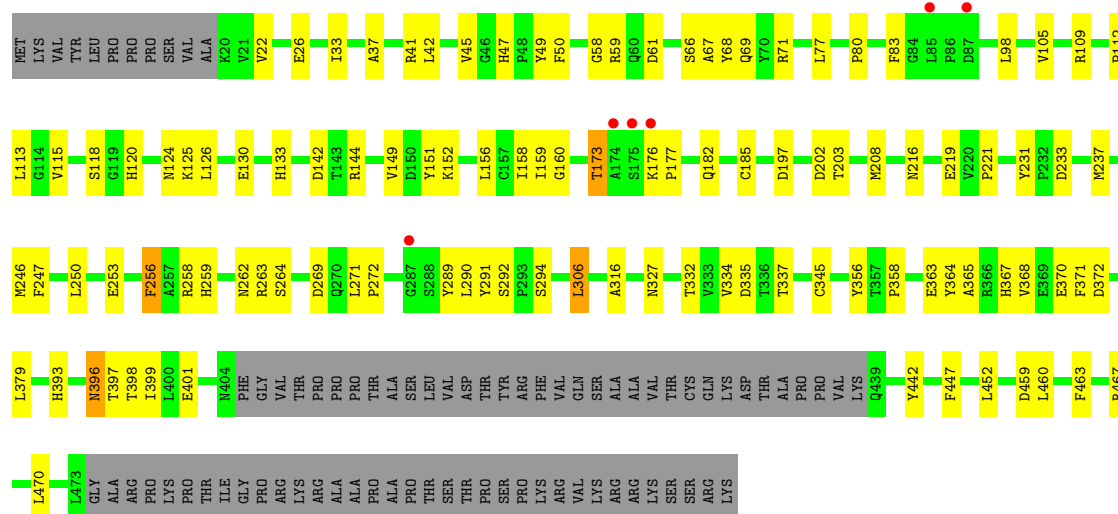




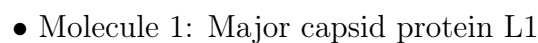
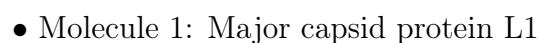
• Molecule 1: Major capsid protein L1



• Molecule 1: Major capsid protein L1



• Molecule 1: Major capsid protein L1



[illegible]

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	114.24Å 161.74Å 154.73Å 90.00° 110.34° 90.00°	Depositor
Resolution (Å)	49.70 – 4.01 49.70 – 4.01	Depositor EDS
% Data completeness (in resolution range)	99.4 (49.70-4.01) 99.3 (49.70-4.01)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.28	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.11 (at 4.00Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.3_928)	Depositor
R, R_{free}	0.243 , 0.267 0.223 , 0.246	Depositor DCC
R_{free} test set	2231 reflections (5.07%)	DCC
Wilson B-factor (Å ²)	85.1	Xtriage
Anisotropy	1.019	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.27 , 63.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	0.011 for h,-k,-h-l	Xtriage
F_o, F_c correlation	0.84	EDS
Total number of atoms	33270	wwPDB-VP
Average B, all atoms (Å ²)	121.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.85% 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	0.25	0/3421	0.43	0/4656
1	B	0.48	5/3390 (0.1%)	0.62	3/4612 (0.1%)
1	C	0.26	1/3421 (0.0%)	0.43	0/4656
1	D	0.29	0/3421	0.64	5/4656 (0.1%)
1	E	0.30	2/3421 (0.1%)	0.43	0/4656
1	F	0.27	1/3421 (0.0%)	0.45	1/4656 (0.0%)
1	G	0.23	0/3421	0.42	0/4656
1	H	0.23	0/3378	0.42	1/4595 (0.0%)
1	I	0.27	1/3421 (0.0%)	0.43	0/4656
1	J	0.23	0/3421	0.42	0/4656
All	All	0.29	10/34136 (0.0%)	0.48	10/46455 (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	D	0	1

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	276	TYR	CG-CD2	-12.48	1.23	1.39
1	B	276	TYR	CG-CD1	-12.19	1.23	1.39
1	B	276	TYR	CE1-CZ	-10.87	1.24	1.38
1	B	276	TYR	CE2-CZ	-9.84	1.25	1.38
1	E	327	ASN	CG-OD1	-7.29	1.07	1.24

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	467	ARG	NE-CZ-NH1	-19.69	110.46	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	467	ARG	NE-CZ-NH1	-19.51	110.54	120.30
1	B	467	ARG	NE-CZ-NH2	19.44	130.02	120.30
1	D	467	ARG	NE-CZ-NH2	19.32	129.96	120.30
1	D	467	ARG	CD-NE-CZ	9.71	137.19	123.60

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	461	ASP	Sidechain

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	3334	0	3234	75	0
1	B	3305	0	3201	82	0
1	C	3334	0	3234	84	0
1	D	3334	0	3234	76	1
1	E	3334	0	3234	77	0
1	F	3334	0	3234	81	2
1	G	3334	0	3234	78	1
1	H	3293	0	3189	81	1
1	I	3334	0	3234	81	2
1	J	3334	0	3234	78	1
All	All	33270	0	32262	662	4

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 10.

The worst 5 of 662 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:C:41:ARG:NH2	1:H:233:ASP:OD2	2.18	0.76
1:B:290:LEU:HD22	1:D:363:GLU:HB3	1.68	0.74

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:126:LEU:HB3	1:B:262:ASN:HB3	1.72	0.72
1:F:233:ASP:OD2	1:J:41:ARG:NH2	2.23	0.72
1:J:126:LEU:HB3	1:J:262:ASN:HB3	1.72	0.72

All (4) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:G:173:THR:OG1	1:J:57:ASN:O[2_547]	1.85	0.35
1:D:280:THR:OG1	1:I:141:LYS:NZ[1_554]	1.95	0.25
1:F:57:ASN:ND2	1:I:386:THR:OG1[2_657]	2.15	0.05
1:F:57:ASN:ND2	1:H:88:ASN:OD1[2_657]	2.16	0.04

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	416/500 (83%)	376 (90%)	36 (9%)	4 (1%)	18	61
1	B	410/500 (82%)	372 (91%)	35 (8%)	3 (1%)	25	68
1	C	416/500 (83%)	376 (90%)	36 (9%)	4 (1%)	18	61
1	D	416/500 (83%)	376 (90%)	36 (9%)	4 (1%)	18	61
1	E	416/500 (83%)	376 (90%)	36 (9%)	4 (1%)	18	61
1	F	416/500 (83%)	374 (90%)	37 (9%)	5 (1%)	15	58
1	G	416/500 (83%)	376 (90%)	36 (9%)	4 (1%)	18	61
1	H	408/500 (82%)	370 (91%)	35 (9%)	3 (1%)	25	68
1	I	416/500 (83%)	376 (90%)	36 (9%)	4 (1%)	18	61
1	J	416/500 (83%)	376 (90%)	36 (9%)	4 (1%)	18	61
All	All	4146/5000 (83%)	3748 (90%)	359 (9%)	39 (1%)	20	63

5 of 39 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	130	GLU
1	B	130	GLU
1	C	130	GLU
1	D	130	GLU
1	E	130	GLU

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	376/444 (85%)	370 (98%)	6 (2%)	68	86
1	B	372/444 (84%)	365 (98%)	7 (2%)	62	83
1	C	376/444 (85%)	370 (98%)	6 (2%)	68	86
1	D	376/444 (85%)	371 (99%)	5 (1%)	73	88
1	E	376/444 (85%)	371 (99%)	5 (1%)	73	88
1	F	376/444 (85%)	371 (99%)	5 (1%)	73	88
1	G	376/444 (85%)	371 (99%)	5 (1%)	73	88
1	H	371/444 (84%)	367 (99%)	4 (1%)	78	89
1	I	376/444 (85%)	370 (98%)	6 (2%)	68	86
1	J	376/444 (85%)	371 (99%)	5 (1%)	73	88
All	All	3751/4440 (84%)	3697 (99%)	54 (1%)	71	87

5 of 54 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	173	THR
1	F	256	PHE
1	J	173	THR
1	E	256	PHE
1	E	306	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such

sidechains are listed below:

Mol	Chain	Res	Type
1	E	47	HIS
1	F	216	ASN
1	I	47	HIS
1	D	216	ASN
1	I	216	ASN

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](#)

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	420/500 (84%)	-0.10	2 (0%) 90 86	77, 114, 203, 264	0
1	B	416/500 (83%)	-0.08	2 (0%) 90 86	78, 113, 187, 265	0
1	C	420/500 (84%)	-0.08	7 (1%) 70 61	78, 113, 202, 265	0
1	D	420/500 (84%)	-0.02	2 (0%) 90 86	78, 114, 203, 264	0
1	E	420/500 (84%)	-0.09	8 (1%) 67 58	78, 114, 202, 265	0
1	F	420/500 (84%)	0.06	9 (2%) 64 54	81, 114, 205, 265	0
1	G	420/500 (84%)	-0.08	6 (1%) 75 66	77, 114, 206, 264	0
1	H	414/500 (82%)	-0.01	6 (1%) 75 66	80, 114, 187, 265	0
1	I	420/500 (84%)	0.01	5 (1%) 79 71	79, 114, 202, 265	0
1	J	420/500 (84%)	-0.06	7 (1%) 70 61	78, 114, 202, 269	0
All	All	4190/5000 (83%)	-0.05	54 (1%) 77 68	77, 114, 202, 269	0

The worst 5 of 54 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	175	SER	6.6
1	J	175	SER	5.0
1	F	57	ASN	4.7
1	E	57	ASN	4.6
1	C	175	SER	4.5

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 carbohydrates 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.