



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 14, 2023 – 11:59 PM JST

PDB ID : 6IGE
Title : Crystal structure of Human Papillomavirus type 33 pentamer
Authors : Li, Z.H.; Song, S.; He, M.Z.; Gu, Y.; Li, S.W.
Deposited on : 2018-09-25
Resolution : 2.90 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 : 4.02b-467
Xtrriage (Phenix) : 1.13
EDS : 2.36
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) : 2.36

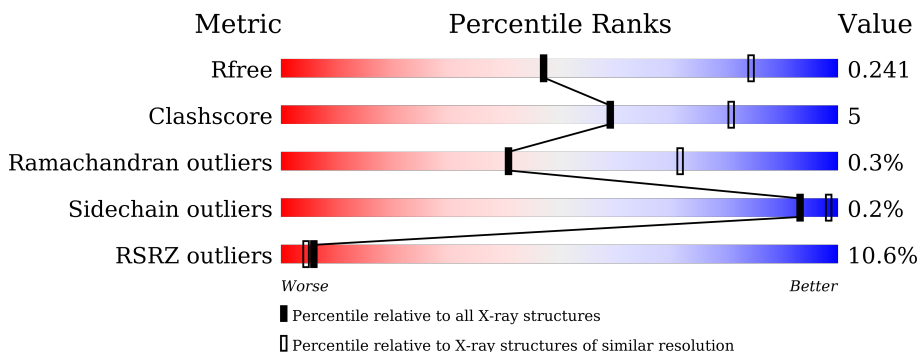
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.90 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.



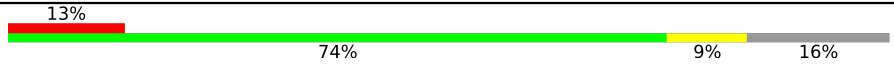
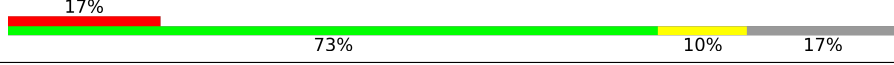
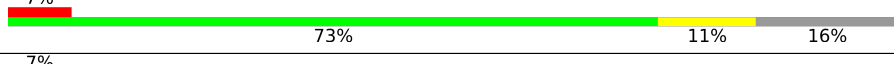

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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	499	 3% 74% 10% 13%
1	B	499	 2% 72% 11% 15%
1	C	499	 3% 73% 10% 14%
1	D	499	 4% 73% 11% 12%
1	E	499	 4% 74% 11% 11%
1	F	499	 28% 73% 11% 17%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	499	
1	H	499	
1	I	499	
1	J	499	

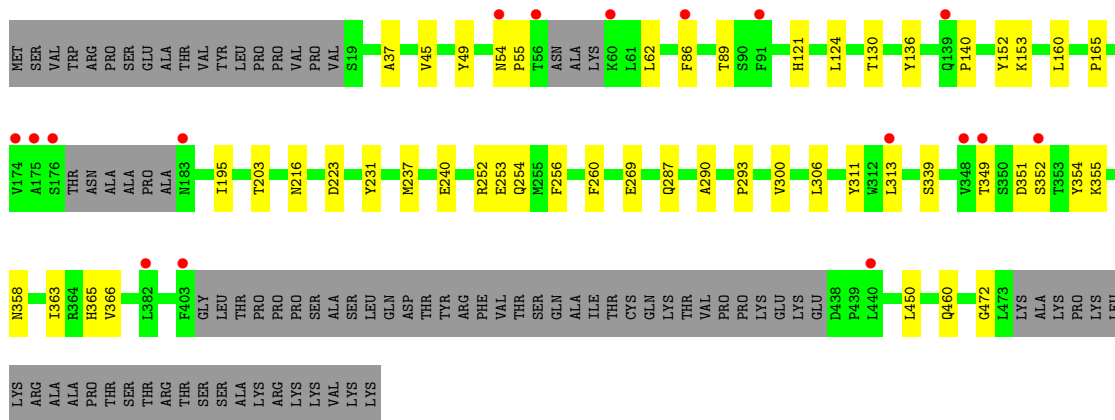
2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 33023 atoms, of which 0 are hydrogens and 0 are deuteriums.

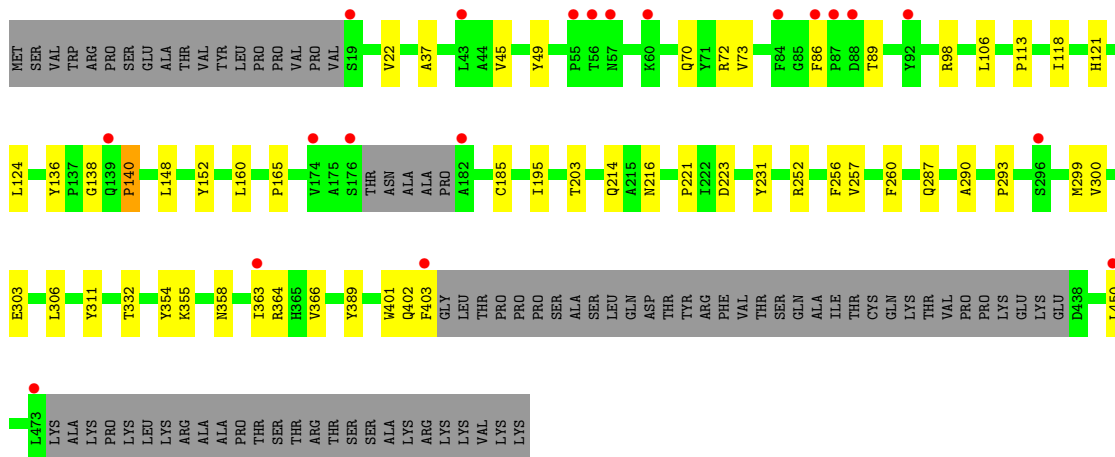
In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Major capsid protein L1.

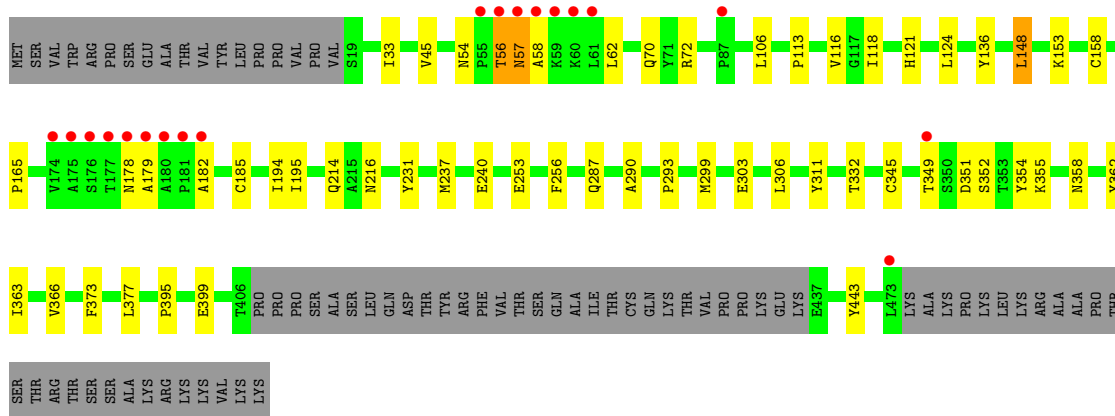
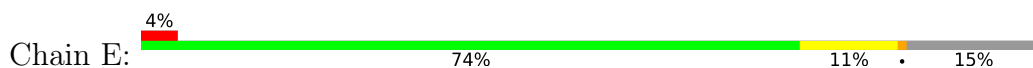
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	417	Total 3309	C 2113	N 548	O 628	S 20	0	0	0
1	B	413	Total 3280	C 2096	N 542	O 622	S 20	0	0	0
1	C	412	Total 3273	C 2092	N 541	O 620	S 20	0	0	0
1	D	416	Total 3300	C 2108	N 547	O 625	S 20	0	0	0
1	E	425	Total 3360	C 2144	N 557	O 639	S 20	0	0	0
1	F	415	Total 3296	C 2106	N 545	O 625	S 20	0	0	0
1	G	417	Total 3310	C 2115	N 548	O 627	S 20	0	0	0
1	H	416	Total 3304	C 2110	N 547	O 627	S 20	0	0	0
1	I	419	Total 3315	C 2117	N 549	O 629	S 20	0	0	0
1	J	412	Total 3276	C 2095	N 542	O 619	S 20	0	0	0



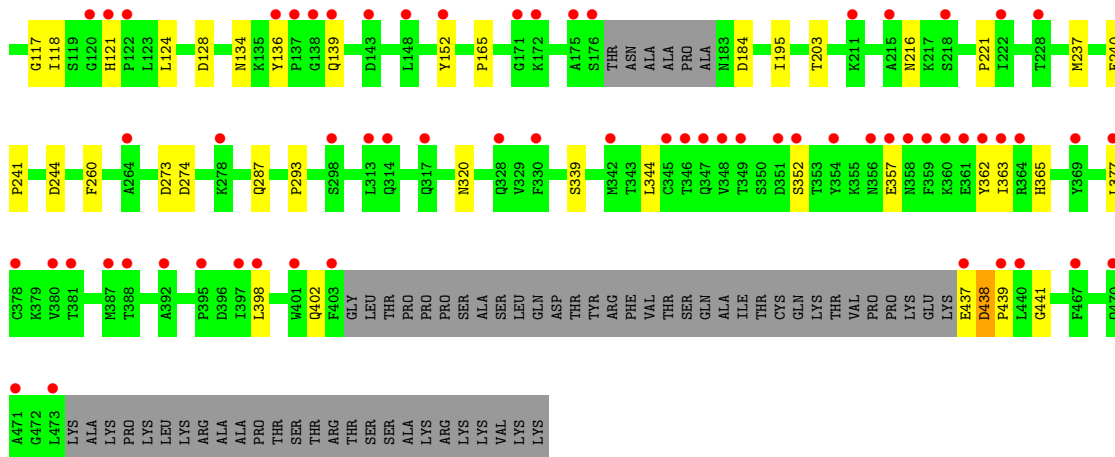
● Molecule 1: Major capsid protein L1



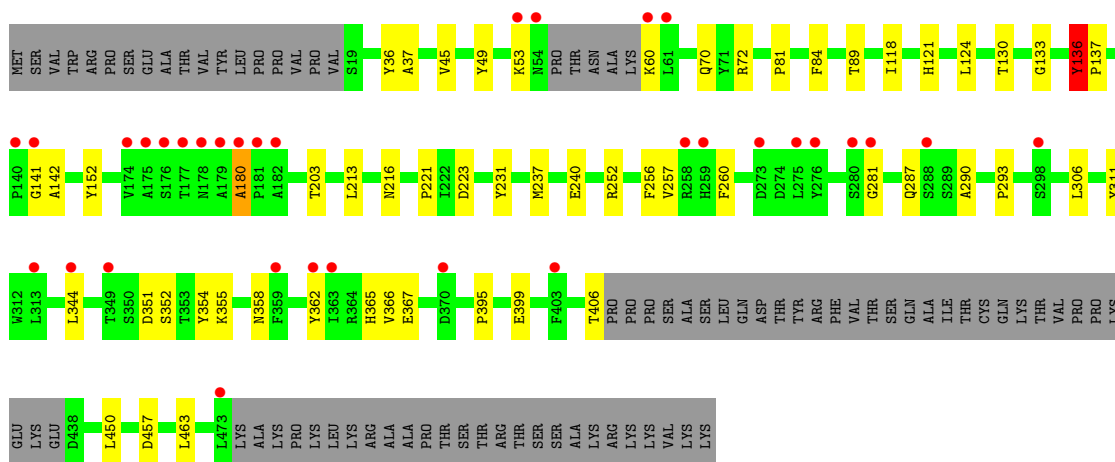
● Molecule 1: Major capsid protein L1



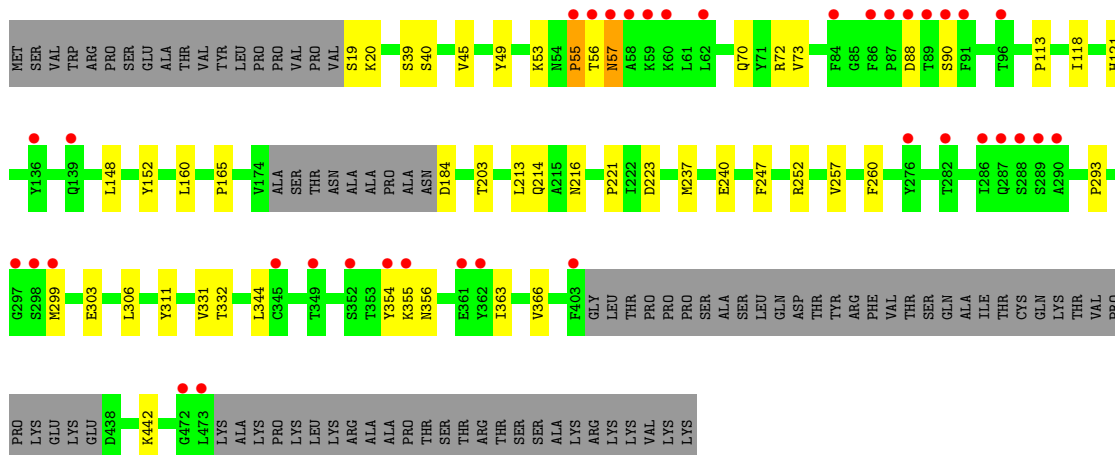
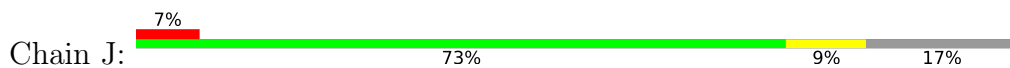
● Molecule 1: Major capsid protein L1



● Molecule 1: Major capsid protein L1



● Molecule 1: Major capsid protein L1



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	98.83Å 171.94Å 145.69Å 90.00° 97.04° 90.00°	Depositor
Resolution (Å)	48.20 – 2.90 48.20 – 2.90	Depositor EDS
% Data completeness (in resolution range)	98.5 (48.20-2.90) 98.5 (48.20-2.90)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.10	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.36 (at 2.91Å)	Xtrriage
Refinement program	PHENIX (1.11.1_2575: ???)	Depositor
R, R_{free}	0.210 , 0.240 0.210 , 0.241	Depositor DCC
R_{free} test set	5243 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å ²)	61.6	Xtrriage
Anisotropy	0.490	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 48.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.35$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	33023	wwPDB-VP
Average B, all atoms (Å ²)	76.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.87% 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	0.27	0/3395	0.48	0/4604
1	B	0.26	0/3365	0.49	0/4562
1	C	0.26	0/3358	0.47	0/4553
1	D	0.26	0/3386	0.49	0/4592
1	E	0.27	0/3448	0.51	1/4680 (0.0%)
1	F	0.26	0/3381	0.48	0/4583
1	G	0.27	0/3396	0.48	0/4605
1	H	0.26	0/3390	0.47	0/4597
1	I	0.27	0/3401	0.50	0/4614
1	J	0.26	0/3362	0.48	0/4559
All	All	0.26	0/33882	0.48	1/45949 (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	I	0	2

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	148	LEU	CA-CB-CG	5.89	128.85	115.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	I	136	TYR	Peptide
1	I	180	ALA	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	3309	0	3202	34	0
1	B	3280	0	3170	41	1
1	C	3273	0	3166	38	0
1	D	3300	0	3196	43	0
1	E	3360	0	3254	43	0
1	F	3296	0	3190	42	1
1	G	3310	0	3206	37	0
1	H	3304	0	3197	38	0
1	I	3315	0	3209	41	1
1	J	3276	0	3175	43	1
All	All	33023	0	31965	306	2

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 5.

The worst 5 of 306 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:I:142:ALA:HB1	1:J:355:LYS:HD2	1.46	0.95
1:J:53:LYS:HG3	1:J:55:PRO:HD3	1.70	0.73
1:J:70:GLN:OE1	1:J:72:ARG:NH2	2.24	0.71
1:F:121:HIS:HB2	1:F:221:PRO:HA	1.74	0.68
1:B:306:LEU:O	1:B:311:TYR:OH	2.07	0.68

All (2) 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:B:357:GLU:OE1	1:I:89:THR:OG1[1_554]	2.14	0.06
1:F:399:GLU:OE2	1:J:442:LYS:NZ[2_546]	2.16	0.04

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	411/499 (82%)	395 (96%)	15 (4%)	1 (0%)	47	78
1	B	405/499 (81%)	392 (97%)	12 (3%)	1 (0%)	47	78
1	C	404/499 (81%)	392 (97%)	11 (3%)	1 (0%)	47	78
1	D	410/499 (82%)	399 (97%)	10 (2%)	1 (0%)	47	78
1	E	421/499 (84%)	404 (96%)	17 (4%)	0	100	100
1	F	407/499 (82%)	394 (97%)	13 (3%)	0	100	100
1	G	411/499 (82%)	395 (96%)	15 (4%)	1 (0%)	47	78
1	H	410/499 (82%)	396 (97%)	12 (3%)	2 (0%)	29	61
1	I	413/499 (83%)	392 (95%)	17 (4%)	4 (1%)	15	45
1	J	406/499 (81%)	390 (96%)	14 (3%)	2 (0%)	29	61
All	All	4098/4990 (82%)	3949 (96%)	136 (3%)	13 (0%)	41	71

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	I	136	TYR
1	I	137	PRO
1	I	180	ALA
1	J	57	ASN
1	I	141	GLY

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	367/439 (84%)	367 (100%)	0	100	100
1	B	364/439 (83%)	363 (100%)	1 (0%)	92	98
1	C	364/439 (83%)	364 (100%)	0	100	100
1	D	366/439 (83%)	366 (100%)	0	100	100
1	E	372/439 (85%)	369 (99%)	3 (1%)	81	94
1	F	366/439 (83%)	366 (100%)	0	100	100
1	G	367/439 (84%)	367 (100%)	0	100	100
1	H	367/439 (84%)	366 (100%)	1 (0%)	92	98
1	I	367/439 (84%)	365 (100%)	2 (0%)	88	96
1	J	364/439 (83%)	364 (100%)	0	100	100
All	All	3664/4390 (84%)	3657 (100%)	7 (0%)	93	98

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	148	LEU
1	H	398	LEU
1	I	406	THR
1	I	136	TYR
1	E	57	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	347	GLN
1	I	178	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 monosaccharides 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	417/499 (83%)	0.26	15 (3%) 42 37	35, 59, 101, 151	0
1	B	413/499 (82%)	0.13	11 (2%) 54 50	31, 49, 88, 121	0
1	C	412/499 (82%)	0.40	17 (4%) 37 32	43, 68, 99, 164	0
1	D	416/499 (83%)	0.33	20 (4%) 30 27	39, 61, 105, 140	0
1	E	425/499 (85%)	0.26	19 (4%) 33 29	31, 47, 95, 178	0
1	F	415/499 (83%)	1.64	138 (33%) 0 0	64, 116, 147, 183	0
1	G	417/499 (83%)	0.91	65 (15%) 2 1	48, 82, 139, 163	0
1	H	416/499 (83%)	1.24	86 (20%) 1 0	64, 97, 141, 170	0
1	I	419/499 (83%)	0.54	33 (7%) 12 10	38, 65, 115, 182	0
1	J	412/499 (82%)	0.59	37 (8%) 9 7	42, 71, 120, 149	0
All	All	4162/4990 (83%)	0.63	441 (10%) 6 4	31, 70, 133, 183	0

The worst 5 of 441 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	179	ALA	14.0
1	I	178	ASN	9.6
1	F	473	LEU	9.5
1	F	176	SER	9.2
1	E	181	PRO	9.1

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 monosaccharides 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.