



# wwPDB/EMDataBank EM Map/Model Validation Summary Report ⓘ

Feb 13, 2017 – 11:58 AM EST

PDB ID : 5KEQ  
EMDB ID: : EMDB-6619  
Title : High resolution cryo-EM maps of Human papillomavirus 16 reveal L2 location and heparin-induced conformational changes  
Authors : Guan, J.; Bywaters, S.M.; Brendle, S.A.; Ashley, R.E.; Makhov, A.M.; Conway, J.F.; Christensen, N.D.; Hafenstein, S.  
Deposited on : 2016-06-10  
Resolution : 4.30 Å(reported)

This is a wwPDB/EMDataBank EM Map/Model Validation Summary Report for a publicly released PDB/EMDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/EMValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

MolProbity : 4.02b-467  
Mogul : unknown  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20028442

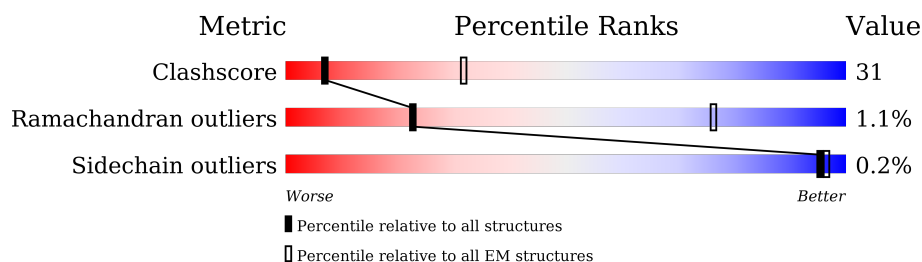
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is 4.30 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)	EM structures (#Entries)
Clashscore	114402	924
Ramachandran outliers	111179	726
Sidechain outliers	111093	686

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	A	483	
1	B	483	
1	C	483	
1	D	483	
1	E	483	
1	F	483	

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 22267 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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.

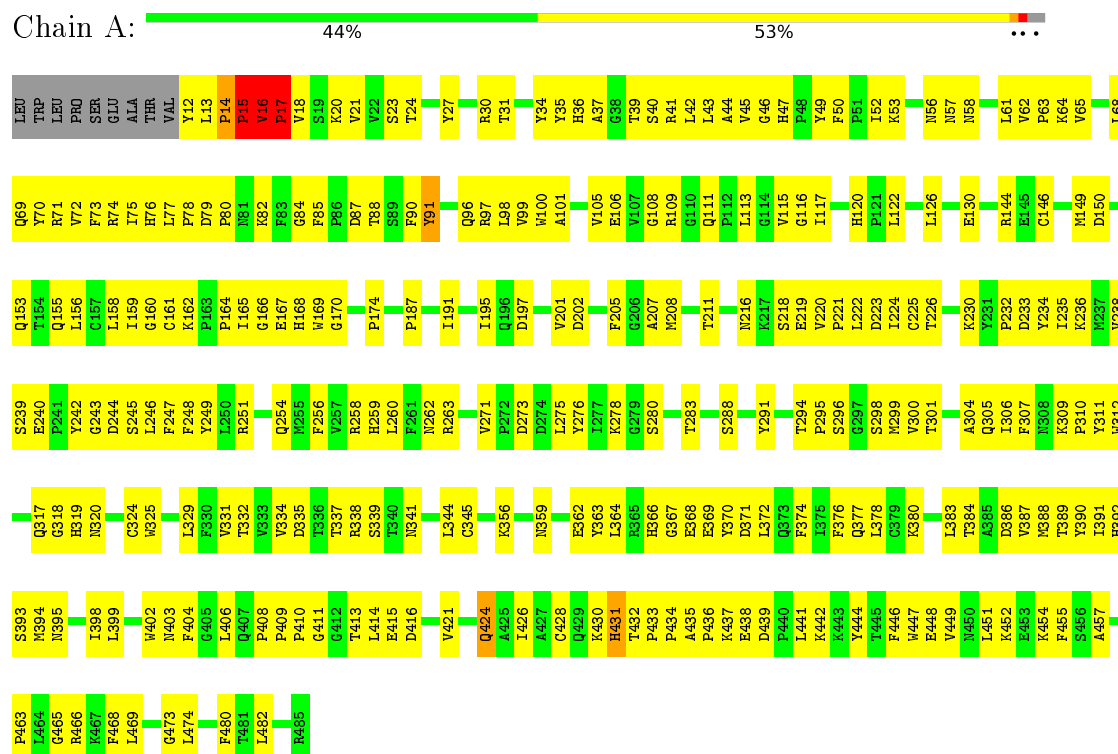
- Molecule 1 is a protein called Major capsid protein L1.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	474	Total	C	N	O	S	0	0
			3725	2380	627	697	21		
1	B	463	Total	C	N	O	S	0	0
			3638	2321	612	684	21		
1	C	472	Total	C	N	O	S	0	0
			3705	2368	621	695	21		
1	D	471	Total	C	N	O	S	0	0
			3700	2365	620	694	21		
1	E	482	Total	C	N	O	S	0	0
			3785	2422	633	709	21		
1	F	473	Total	C	N	O	S	0	0
			3714	2373	622	698	21		

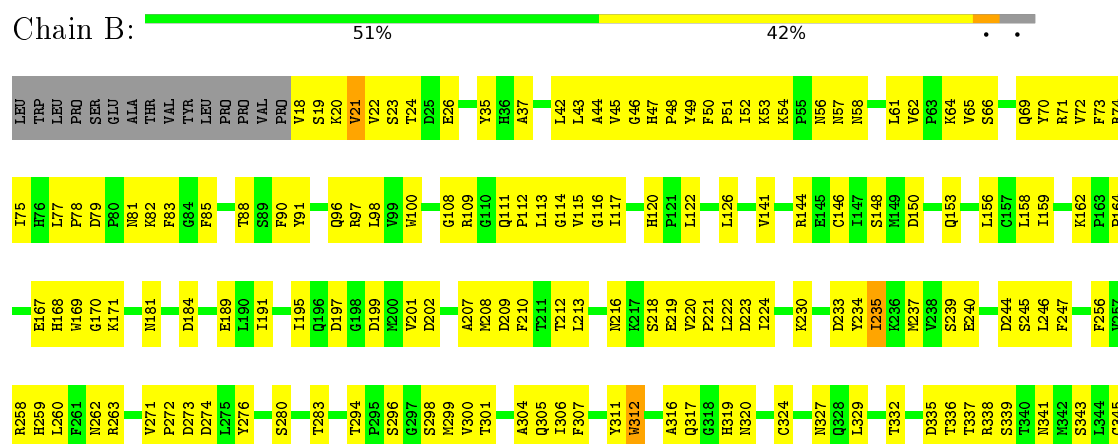
### 3 Residue-property plots

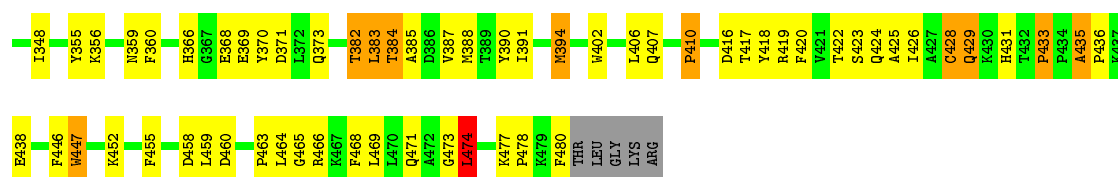
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according to the number of geometric 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 L1



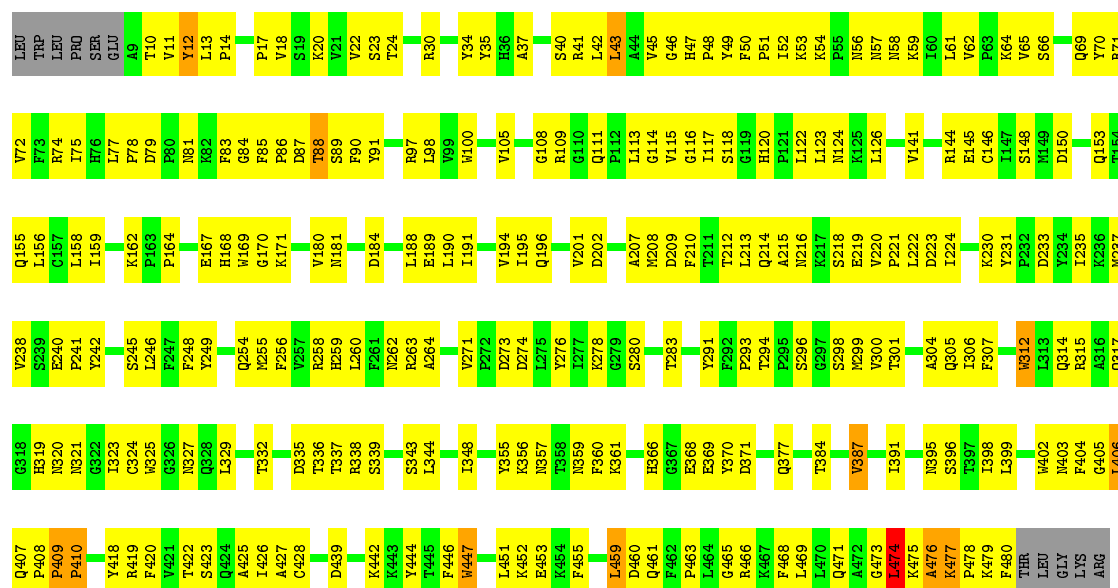
- Molecule 1: Major capsid protein L1





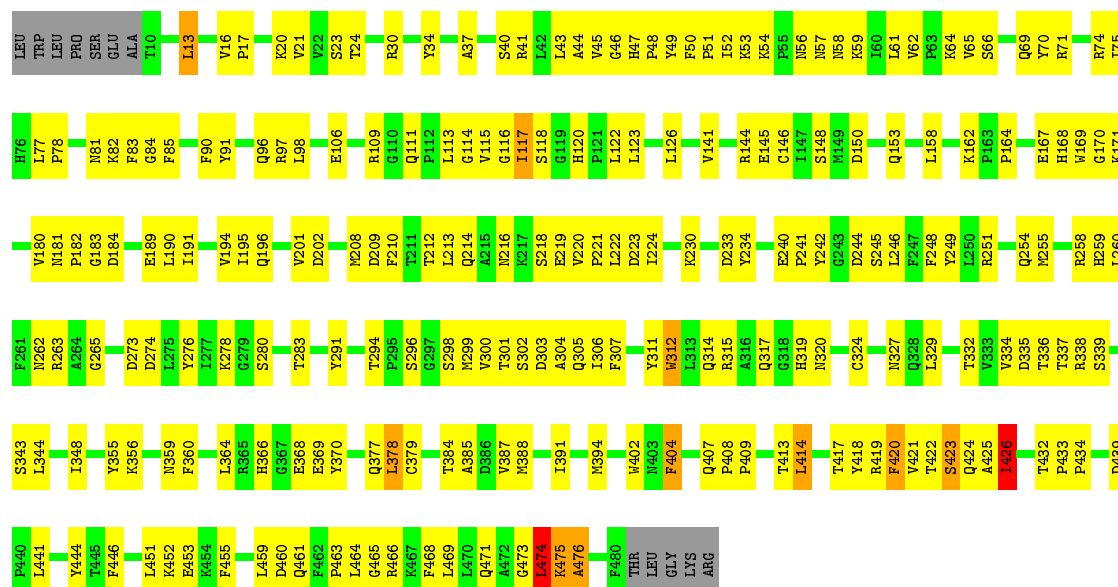
• Molecule 1: Major capsid protein L1

Chain C: 47% 48% ..

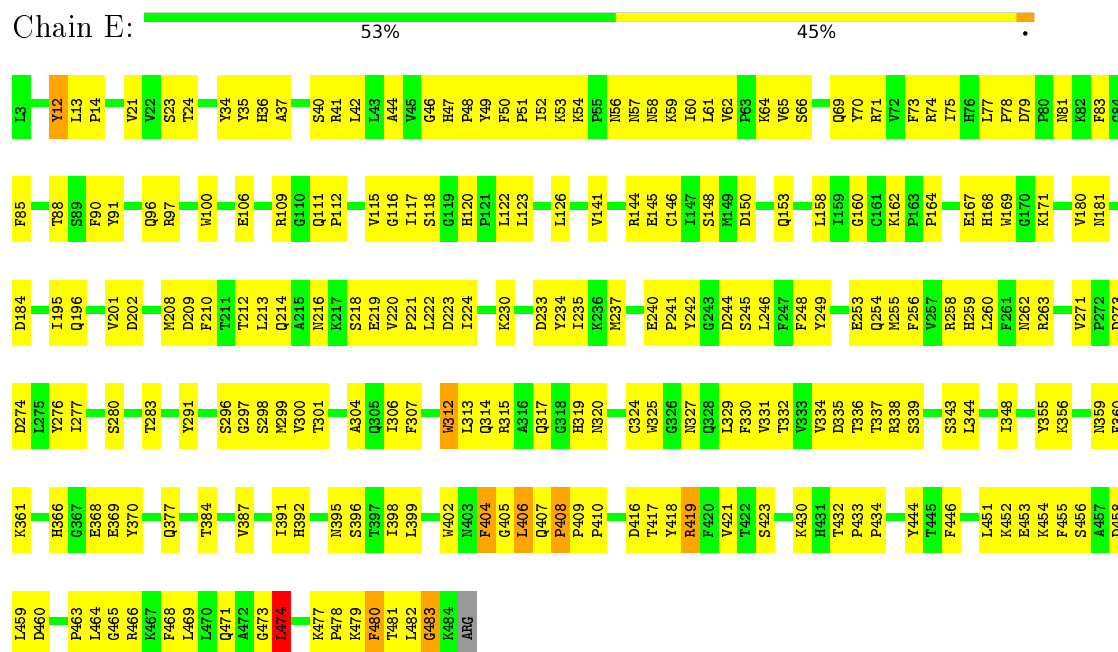


• Molecule 1: Major capsid protein L1

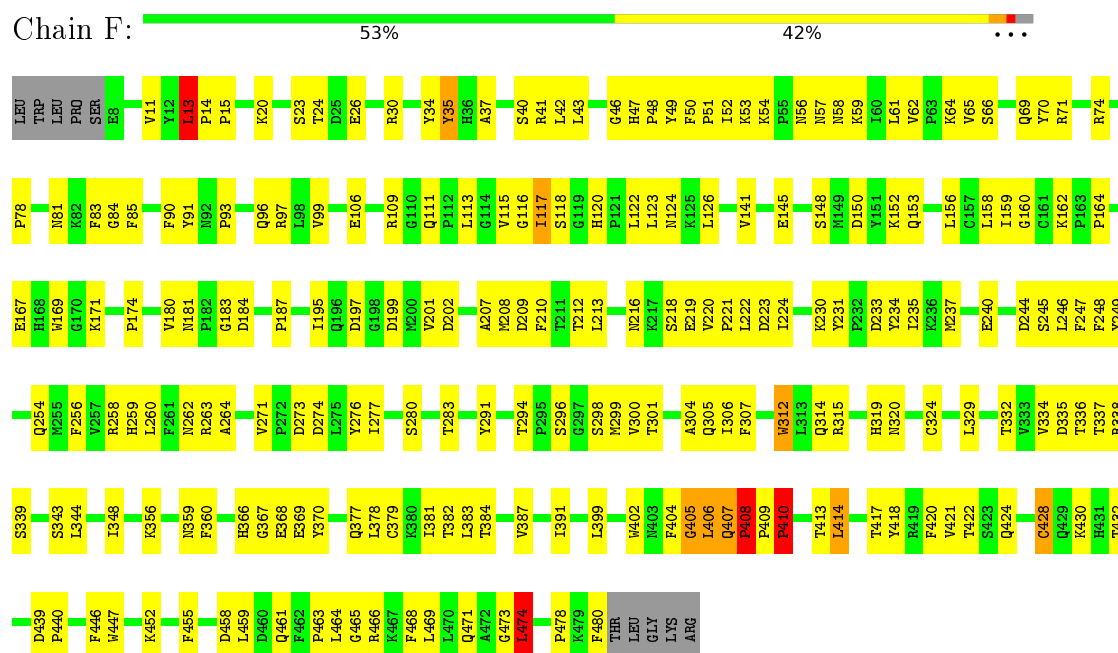
Chain D: 51% 44% ..



• Molecule 1: Major capsid protein L1



• Molecule 1: Major capsid protein L1



## 4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, I	Depositor
Number of particles used	51422	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	Not provided	Depositor
Microscope	FEI POLARA 300	Depositor
Voltage (kV)	300	Depositor
Electron dose ( $e^-/\text{\AA}^2$ )	Not provided	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	Not provided	Depositor
Image detector	Not provided	Depositor

## 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	$\# Z  > 2$	RMSZ	$\# Z  > 2$
1	A	0.50	0/3827	0.68	8/5209 (0.2%)
1	B	0.57	2/3736 (0.1%)	0.67	6/5083 (0.1%)
1	C	0.64	4/3807 (0.1%)	0.74	7/5185 (0.1%)
1	D	0.60	3/3802 (0.1%)	0.78	11/5178 (0.2%)
1	E	0.60	4/3890 (0.1%)	0.67	6/5299 (0.1%)
1	F	0.69	8/3816 (0.2%)	0.78	16/5197 (0.3%)
All	All	0.60	21/22878 (0.1%)	0.72	54/31151 (0.2%)

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	A	0	6
1	B	0	7
1	C	0	4
1	D	0	3
1	E	0	5
1	F	0	6
All	All	0	31

The worst 5 of 21 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	F	408	PRO	N-CA	15.98	1.74	1.47
1	F	409	PRO	N-CA	11.28	1.66	1.47
1	F	409	PRO	C-N	10.84	1.54	1.34
1	C	409	PRO	C-N	-8.11	1.18	1.34
1	E	404	PHE	CB-CG	-7.85	1.38	1.51

The worst 5 of 54 bond angle outliers are listed below:



Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	426	ILE	CG1-CB-CG2	-16.52	75.05	111.40
1	F	13	LEU	CB-CG-CD1	-12.63	89.52	111.00
1	D	426	ILE	CB-CG1-CD1	-12.57	78.71	113.90
1	D	378	LEU	CB-CG-CD2	-9.98	94.03	111.00
1	D	474	LEU	CA-C-O	8.64	138.24	120.10

There are no chirality outliers.

5 of 31 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	17	PRO	Peptide
1	A	415	GLU	Peptide
1	A	424	GLN	Peptide
1	A	430	LYS	Peptide
1	A	91	TYR	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	3725	0	3648	256	0
1	B	3638	0	3551	233	0
1	C	3705	0	3622	278	0
1	D	3700	0	3617	250	0
1	E	3785	0	3706	236	0
1	F	3714	0	3628	227	0
All	All	22267	0	21772	1373	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 31.

The worst 5 of 1373 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:469:LEU:HA	1:D:474:LEU:CD2	1.44	1.48
1:D:469:LEU:CA	1:D:474:LEU:HD21	1.48	1.41

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:F:408:PRO:CA	1:F:408:PRO:N	1.74	1.31
1:E:455:PHE:HB2	1:F:13:LEU:HD11	1.15	1.10
1:C:475:LYS:O	1:C:477:LYS:N	1.89	1.03

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 PDB entries followed by that with respect to all EM entries.

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	472/483 (98%)	434 (92%)	33 (7%)	5 (1%)	17	64
1	B	461/483 (95%)	428 (93%)	26 (6%)	7 (2%)	13	58
1	C	470/483 (97%)	428 (91%)	37 (8%)	5 (1%)	17	64
1	D	469/483 (97%)	426 (91%)	36 (8%)	7 (2%)	13	58
1	E	480/483 (99%)	439 (92%)	39 (8%)	2 (0%)	39	80
1	F	471/483 (98%)	438 (93%)	29 (6%)	4 (1%)	24	70
All	All	2823/2898 (97%)	2593 (92%)	200 (7%)	30 (1%)	23	64

5 of 30 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	384	THR
1	C	476	ALA
1	C	477	LYS
1	D	474	LEU
1	E	474	LEU

### 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 PDB entries followed by that with respect to all EM entries.

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	412/420 (98%)	411 (100%)	1 (0%)	95	97
1	B	402/420 (96%)	401 (100%)	1 (0%)	95	97
1	C	410/420 (98%)	409 (100%)	1 (0%)	95	97
1	D	410/420 (98%)	408 (100%)	2 (0%)	92	96
1	E	419/420 (100%)	419 (100%)	0	100	100
1	F	411/420 (98%)	410 (100%)	1 (0%)	95	97
All	All	2464/2520 (98%)	2458 (100%)	6 (0%)	95	97

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	474	LEU
1	F	474	LEU
1	D	474	LEU
1	B	474	LEU
1	D	475	LYS

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

Mol	Chain	Res	Type
1	A	319	HIS
1	B	366	HIS

### 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	C	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	C	409:PRO	C	410:PRO	N	1.18