



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

Oct 16, 2017 – 05:48 AM EDT

PDB ID : 5JB1  
EMDB ID: : EMD-8147  
Title : Pseudo-atomic structure of Human Papillomavirus Type 59 L1 Virus-like Particle  
Authors : Li, Z.H.; Yan, X.D.; Yu, H.; Zheng, Q.B.; Gu, Y.; Li, S.W.  
Deposited on : unknown  
Resolution : 6.00 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  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20030345

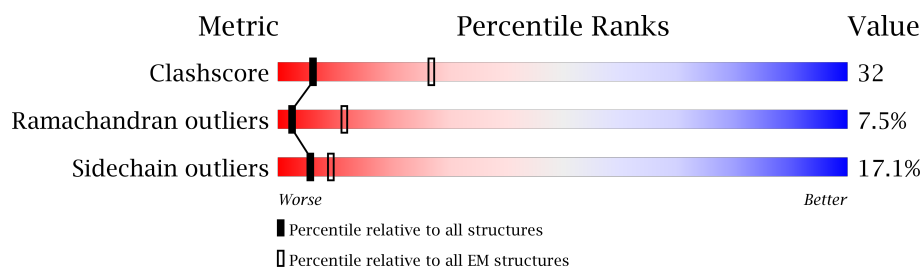
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is 6.00 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	125131	1336
Ramachandran outliers	121729	1120
Sidechain outliers	121581	1026

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	500	
1	B	500	
1	C	500	
1	D	500	
1	E	500	
1	F	500	

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 21696 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.

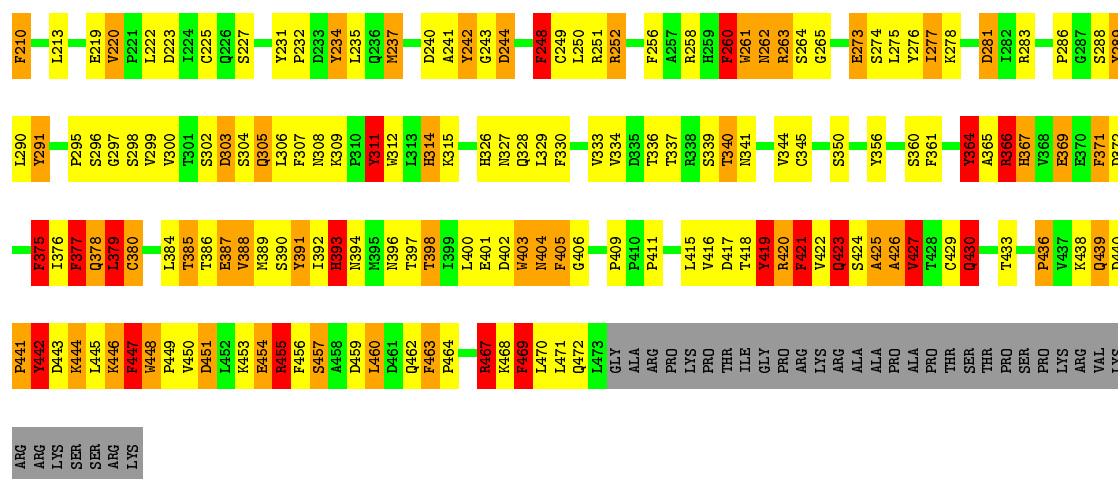
- Molecule 1 is a protein called Major capsid protein L1.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	456	Total	C	N	O	S	0	0
			3598	2289	604	686	19		
1	B	464	Total	C	N	O	S	0	0
			3661	2333	613	696	19		
1	C	457	Total	C	N	O	S	0	0
			3604	2292	605	688	19		
1	D	454	Total	C	N	O	S	0	0
			3586	2281	602	684	19		
1	E	454	Total	C	N	O	S	0	0
			3586	2281	602	684	19		
1	F	464	Total	C	N	O	S	0	0
			3661	2333	613	696	19		

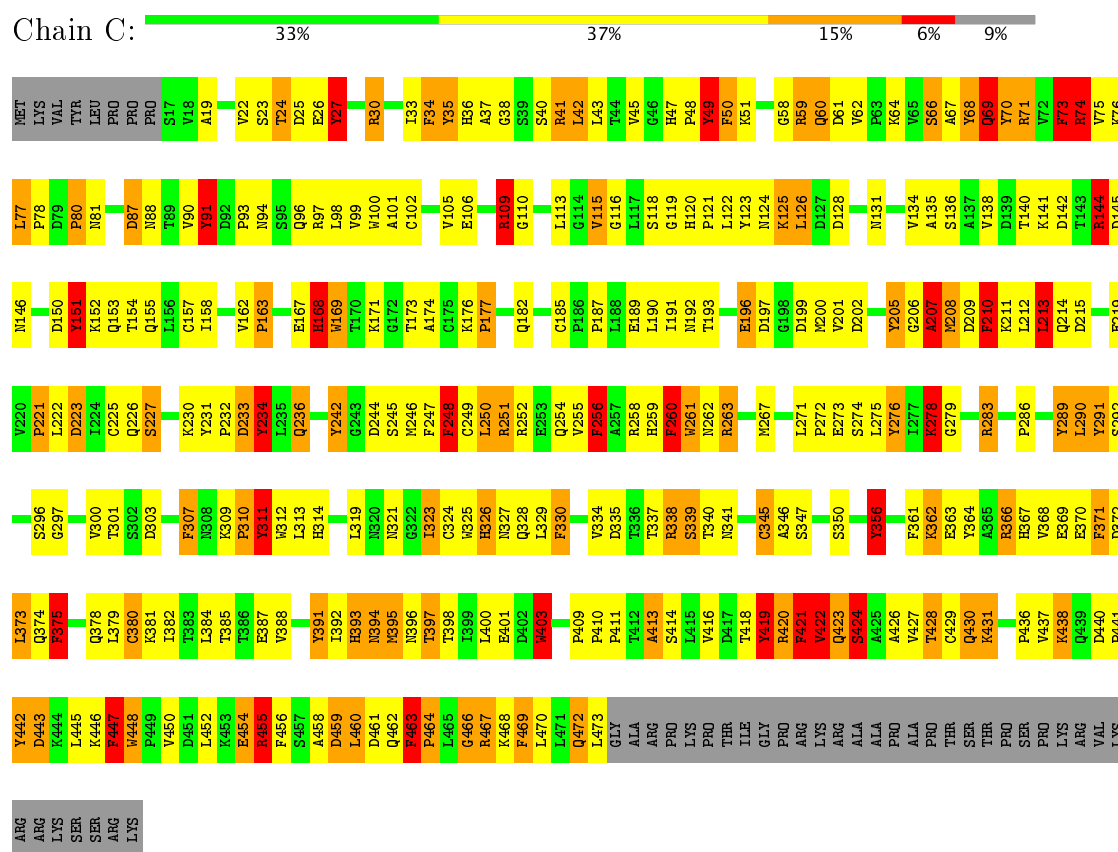
There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	9	MET	-	initiating methionine	UNP Q81971
B	9	MET	-	initiating methionine	UNP Q81971
C	9	MET	-	initiating methionine	UNP Q81971
D	9	MET	-	initiating methionine	UNP Q81971
E	9	MET	-	initiating methionine	UNP Q81971
F	9	MET	-	initiating methionine	UNP Q81971

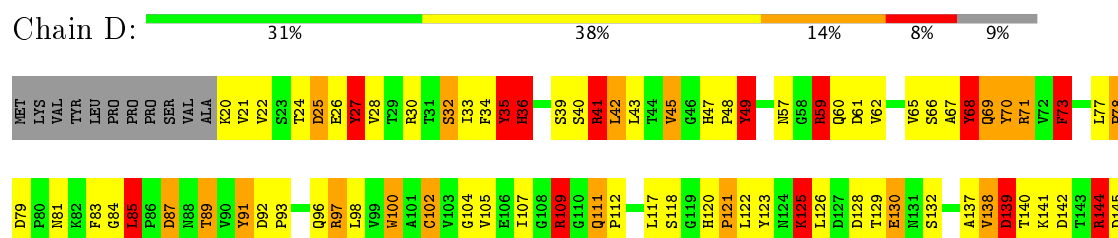


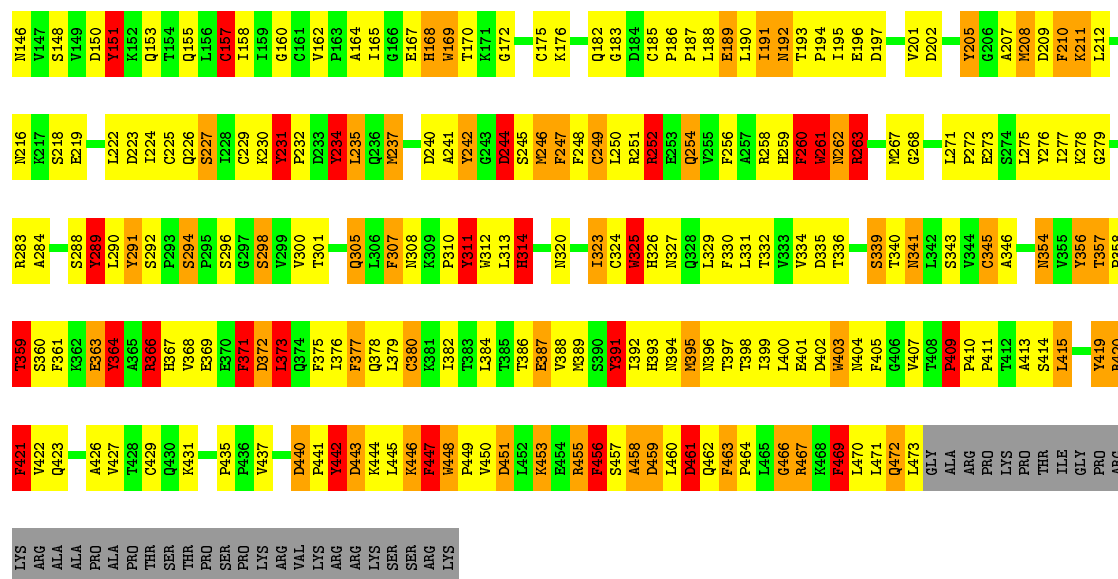


• Molecule 1: Major capsid protein L1

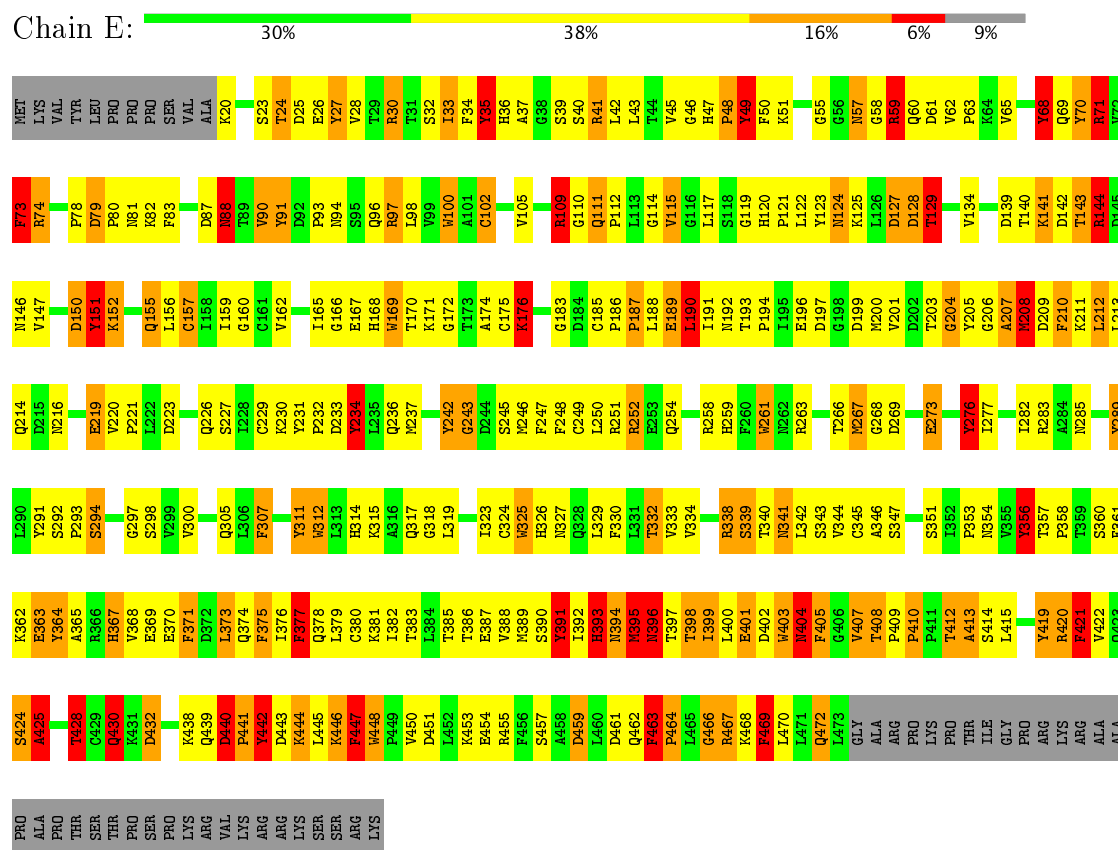


• Molecule 1: Major capsid protein L1

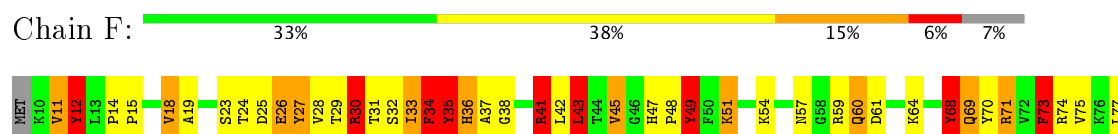




• Molecule 1: Major capsid protein L1



• Molecule 1: Major capsid protein L1



VAL	D440	C380	R307	S227	L156	P78
LYS	P441	K381	N308	I228	C157	D79
ARG	Y442	T382	K309	C229	I158	P80
ARG	D443	T383	P310	K230	I159	N81
LYS	K444	L384	W311	Y231	G160	K82
LYS	L445	T385	W312	P232	G161	K83
SER	K446	T386	I313	D233	G162	G84
SER	F447	R387	H314	Y234	I165	L85
ARG	W448	V388	Q317	L235	G166	P86
LYS	W449	K389	G318	Y236	E167	D87
	V450	S390	H319	K237	H168	N88
	D451	Y391	N320	D240	W169	T89
	L452	K392	I321	A241	T170	V90
	K453	R393	I322	I242	K171	N91
	E454	T394	I323	Y243	G172	D92
	K455	K395	C324	G243	T173	P93
	F456	N396	W325	D244	A174	N94
	S457	T397	H326	S245	C175	S95
	A458	T398	N327	K246	D184	O96
	D459	I399	Q328	F247	C185	A97
	L460	L400	L329	F249	G186	L98
	D461	E401	V333	R251	P187	V99
	F463	N403	W334	R252	L188	W100
	P464	N404	D335	E253	P189	A101
	L465	F405	S339	F256	E189	C102
	G466	G406	T340	A257	L190	V103
	R467	Y407	N341	R258	N191	E106
	K468	T408	H342	H259	T193	I107
	F469	P409	C345	F260	G194	G108
	L470	P410	K346	N261	I195	R109
	Q471	P411	S347	N262	E196	G110
	L472	A413	I352	R263	D199	Q111
	L473	S414	P353	S264	M200	P112
	ALA	L415	N354	N267	V201	H120
	ARG	V416	Y355	Y276	D202	P121
	PRO	D417	Y356	I277	T203	L122
	LYS	T418	S360	K278	G204	Y123
	PRO	Y419	F361	D281	Y205	H124
	THR	R420	K362	E282	M208	K125
	ILE	V421	E363	R283	F209	L126
	GLY	V422	Y364	Y289	P210	D127
	PRO	Q423	A365	L290	K211	D128
	ARG	S424	R366	Y291	L212	S132
	LYS	A425	H367	V299	I213	H133
	ARG	V427	V368	P295	Q214	V134
	ALA	T428	E369	S296	D215	D142
	PRO	Q429	F370	G297	K216	D143
	ALA	K430	D372	C298	R217	T144
	PRO	R431	L373	V299	E219	R144
	THR	D432	Q374	D303	V220	D145
	SER	T433	F375	S304	P221	N146
	SER	A434	I376	Q305	L222	D150
	PRO	P435	F377	L306	D223	Y151
	SER	P436	Q378		L224	Y152
	PRO	V437	L379		C225	K153
	LYS	K438			Q226	
	ARG	Q439				

## 4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, Not provided	Depositor
Number of particles used	3100	Depositor
Resolution determination method	OTHER	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TECNAI F30	Depositor
Voltage (kV)	300	Depositor
Electron dose ( $e^-/\text{\AA}^2$ )	25	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	Not provided	Depositor
Image detector	FEI FALCON I (4k x 4k)	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	1.47	8/3694 (0.2%)	1.88	94/5036 (1.9%)
1	B	1.45	7/3761 (0.2%)	1.89	110/5130 (2.1%)
1	C	1.48	5/3700 (0.1%)	1.91	107/5044 (2.1%)
1	D	1.49	11/3682 (0.3%)	1.87	97/5019 (1.9%)
1	E	1.94	8/3682 (0.2%)	1.89	100/5019 (2.0%)
1	F	1.49	11/3761 (0.3%)	1.88	88/5130 (1.7%)
All	All	1.56	50/22280 (0.2%)	1.89	596/30378 (2.0%)

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	A	0	35
1	B	0	45
1	C	0	46
1	D	0	43
1	E	0	37
1	F	0	47
All	All	0	253

The worst 5 of 50 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	405	PHE	N-CA	77.19	3.00	1.46
1	B	35	TYR	CB-CG	-7.31	1.40	1.51
1	E	469	PHE	CB-CG	-6.46	1.40	1.51
1	F	392	ILE	N-CA	-6.36	1.33	1.46
1	A	447	PHE	CA-C	-6.18	1.36	1.52

The worst 5 of 596 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	C	455	ARG	NE-CZ-NH1	13.91	127.26	120.30
1	B	91	TYR	CB-CG-CD2	-13.52	112.89	121.00
1	C	91	TYR	CB-CG-CD2	-13.48	112.91	121.00
1	C	420	ARG	NE-CZ-NH2	13.26	126.93	120.30
1	A	386	THR	N-CA-C	13.11	146.38	111.00

There are no chirality outliers.

5 of 253 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	35	TYR	Sidechain
1	A	41	ARG	Sidechain
1	A	49	TYR	Sidechain
1	A	68	TYR	Sidechain
1	A	70	TYR	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	3598	0	3503	270	0
1	B	3661	0	3571	243	0
1	C	3604	0	3506	219	0
1	D	3586	0	3488	236	0
1	E	3586	0	3489	244	0
1	F	3661	0	3571	230	0
All	All	21696	0	21128	1384	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 32.

The worst 5 of 1384 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:A:386:THR:O	1:A:386:THR:HG22	1.56	1.06
1:D:169:TRP:CD1	1:D:190:LEU:HD13	2.06	0.90

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:169:TRP:CE2	1:B:190:LEU:HD13	2.05	0.90
1:A:70:TYR:CD1	1:A:201:VAL:HG12	2.07	0.90
1:F:169:TRP:CD1	1:F:190:LEU:HD13	2.11	0.85

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	454/500 (91%)	359 (79%)	59 (13%)	36 (8%)	1	17
1	B	462/500 (92%)	356 (77%)	65 (14%)	41 (9%)	1	15
1	C	455/500 (91%)	363 (80%)	57 (12%)	35 (8%)	1	17
1	D	452/500 (90%)	370 (82%)	53 (12%)	29 (6%)	1	22
1	E	452/500 (90%)	351 (78%)	70 (16%)	31 (7%)	1	20
1	F	462/500 (92%)	367 (79%)	62 (13%)	33 (7%)	1	19
All	All	2737/3000 (91%)	2166 (79%)	366 (13%)	205 (8%)	2	18

5 of 205 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	20	LYS
1	A	50	PHE
1	A	67	ALA
1	A	86	PRO
1	A	404	ASN

### 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 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	406/444 (91%)	341 (84%)	65 (16%)	3	18
1	B	414/444 (93%)	353 (85%)	61 (15%)	3	21
1	C	407/444 (92%)	343 (84%)	64 (16%)	3	18
1	D	405/444 (91%)	328 (81%)	77 (19%)	2	11
1	E	405/444 (91%)	332 (82%)	73 (18%)	2	13
1	F	414/444 (93%)	334 (81%)	80 (19%)	1	11
All	All	2451/2664 (92%)	2031 (83%)	420 (17%)	6	15

5 of 420 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	472	GLN
1	D	305	GLN
1	F	339	SER
1	D	42	LEU
1	D	151	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 43 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	326	HIS
1	D	168	HIS
1	F	168	HIS
1	C	394	ASN
1	C	423	GLN

### 5.3.3 RNA ⓘ

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.