



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

Feb 19, 2018 – 02:50 am GMT

PDB ID : 5G06  
EMDB ID: : EMD-3366  
Title : Cryo-EM structure of yeast cytoplasmic exosome  
Authors : Liu, J.J.; Niu, C.Y.; Wu, Y.; Tan, D.; Wang, Y.; Ye, M.D.; Liu, Y.; Zhao, W.W.; Zhou, K.; Liu, Q.S.; Dai, J.B.; Yang, X.R.; Dong, M.Q.; Huang, N.; Wang, H.W.  
Deposited on : 2016-03-17  
Resolution : 4.20 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

<https://www.wwpdb.org/validation/2017/EMValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

MolProbity : 4.02b-467  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk30686

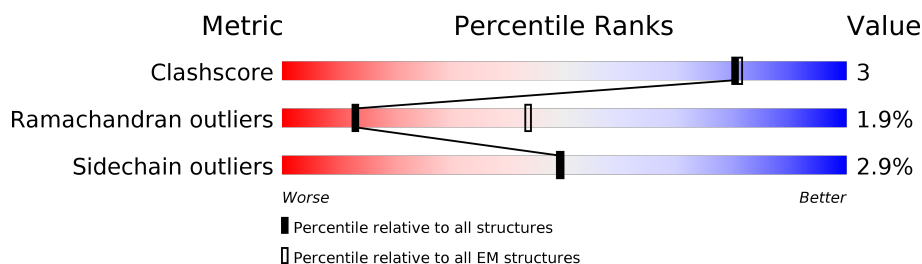
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is 4.20 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	136279	1886
Ramachandran outliers	132675	1663
Sidechain outliers	132484	1531

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	305	79% 19% ..
2	B	246	73% 21% . .
3	C	394	64% 17% . . 15%
4	D	223	81% 18% .
5	E	265	79% 19% .
6	F	250	61% 20% . . 16%
7	G	240	75% 20% . .
8	H	359	58% 21% . 18%
9	I	292	57% 18% . . 22%

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
10	J	1001	<div><div></div><div>74%</div><div>22%</div><div></div></div>
11	P	747	<div><div></div><div>8%</div><div></div><div>87%</div></div>

## 2 Entry composition

There are 11 unique types of molecules in this entry. The entry contains 26794 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 EXOSOME COMPLEX COMPONENT RRP45.

Mol	Chain	Residues	Atoms					AltConf	Trace
1	A	302	Total	C	N	O	S	0	0
			2350	1474	402	457	17		

- Molecule 2 is a protein called EXOSOME COMPLEX COMPONENT SKI6.

Mol	Chain	Residues	Atoms					AltConf	Trace
2	B	240	Total	C	N	O	S	0	0
			1890	1182	338	362	8		

- Molecule 3 is a protein called EXOSOME COMPLEX COMPONENT RRP43.

Mol	Chain	Residues	Atoms					AltConf	Trace
3	C	335	Total	C	N	O	S	0	0
			2622	1653	455	503	11		

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	363	MET	VAL	conflict	UNP P25359

- Molecule 4 is a protein called EXOSOME COMPLEX COMPONENT RRP46.

Mol	Chain	Residues	Atoms					AltConf	Trace
4	D	223	Total	C	N	O	S	0	0
			1707	1073	289	335	10		

- Molecule 5 is a protein called EXOSOME COMPLEX COMPONENT RRP42.

Mol	Chain	Residues	Atoms					AltConf	Trace
5	E	265	Total	C	N	O	S	0	0
			2048	1307	336	400	5		

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	138	ILE	VAL	conflict	UNP Q12277

- Molecule 6 is a protein called EXOSOME COMPLEX COMPONENT MTR3.

Mol	Chain	Residues	Atoms					AltConf	Trace
6	F	211	Total	C	N	O	S	0	0
			1627	1018	276	323	10		

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	75	SER	THR	conflict	UNP P48240

- Molecule 7 is a protein called EXOSOME COMPLEX COMPONENT RRP40.

Mol	Chain	Residues	Atoms					AltConf	Trace
7	G	236	Total	C	N	O	S	0	0
			1831	1169	301	350	11		

- Molecule 8 is a protein called EXOSOME COMPLEX COMPONENT RRP4.

Mol	Chain	Residues	Atoms					AltConf	Trace
8	H	293	Total	C	N	O	S	0	0
			2277	1419	411	435	12		

- Molecule 9 is a protein called EXOSOME COMPLEX COMPONENT CSL4.

Mol	Chain	Residues	Atoms					AltConf	Trace
9	I	227	Total	C	N	O	S	0	0
			1750	1093	310	338	9		

- Molecule 10 is a protein called EXOSOME COMPLEX EXONUCLEASE DIS3.

Mol	Chain	Residues	Atoms					AltConf	Trace
10	J	993	Total	C	N	O	S	0	0
			7942	5002	1395	1509	36		

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
J	171	ASN	ASP	conflict	UNP Q08162
J	551	ASN	ASP	conflict	UNP Q08162

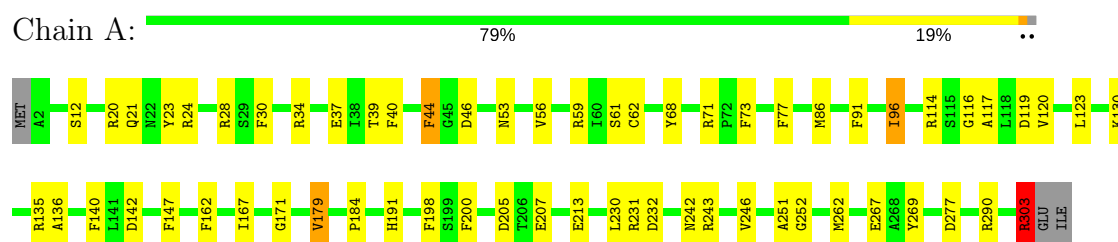
- Molecule 11 is a protein called SUPERKILLER PROTEIN 7.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
11	P	95	750	475	128	146	1	0	0

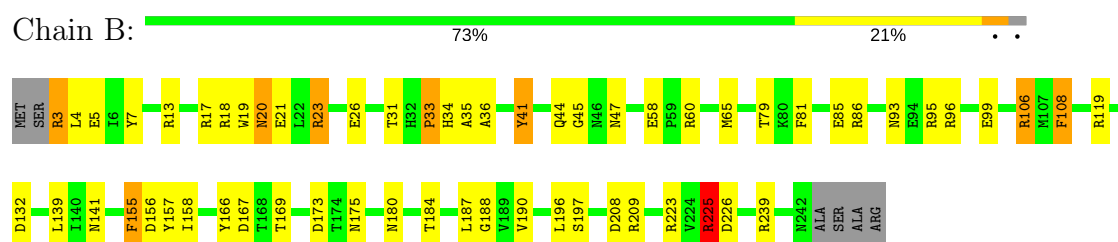
### 3 Residue-property plots

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. 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.

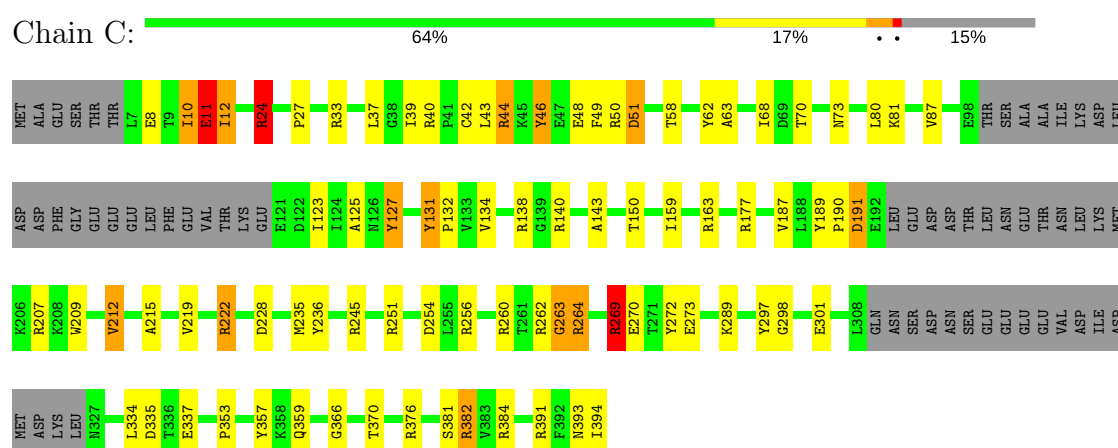
#### • Molecule 1: EXOSOME COMPLEX COMPONENT RRP45



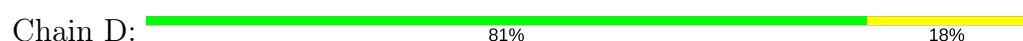
#### • Molecule 2: EXOSOME COMPLEX COMPONENT SKI6



#### • Molecule 3: EXOSOME COMPLEX COMPONENT RRP43



#### • Molecule 4: EXOSOME COMPLEX COMPONENT RRP46



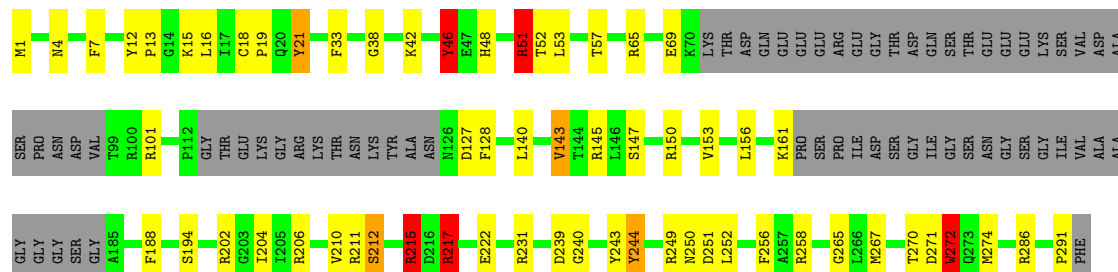






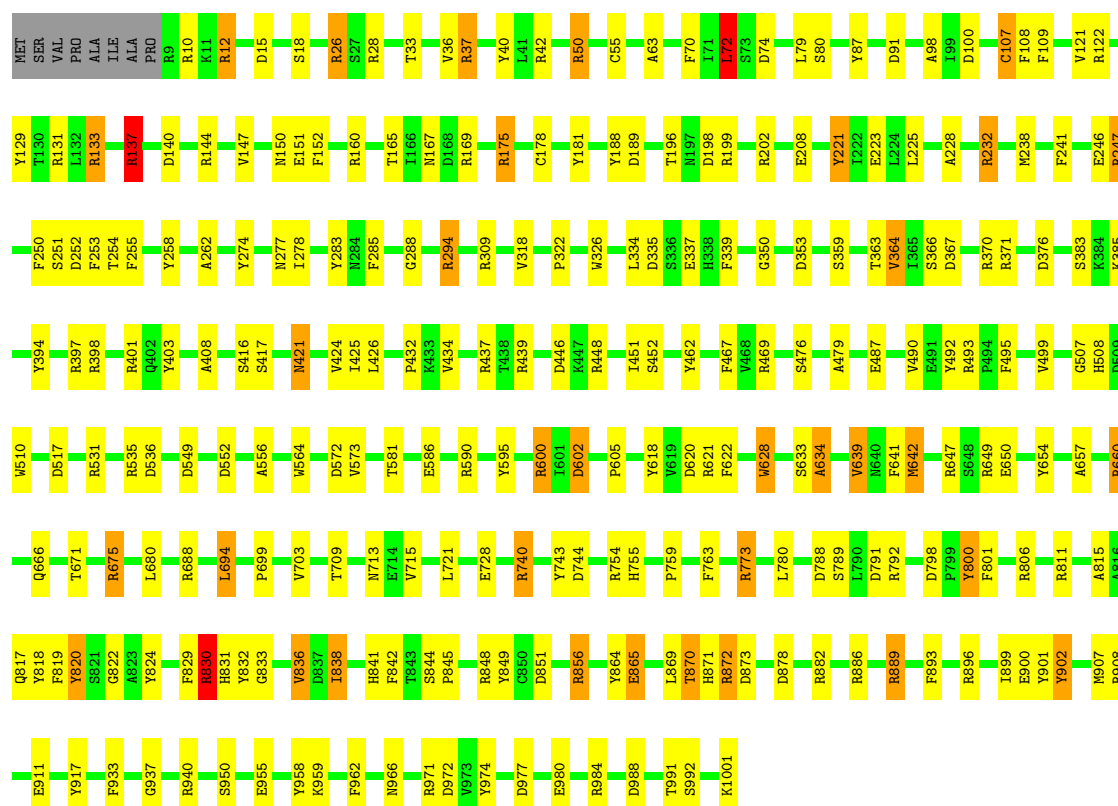
• Molecule 9: EXOSOME COMPLEX COMPONENT CSL4

Chain I: 57% 18% 22%



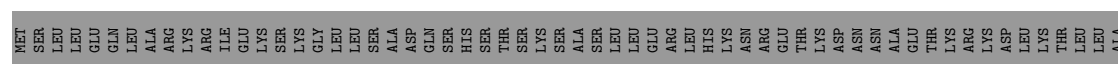
• Molecule 10: EXOSOME COMPLEX EXONUCLEASE DIS3

Chain J: 74% 22%



• Molecule 11: SUPERKILLER PROTEIN 7

Chain P: 8% 87%



[illegible]

## 4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	25000	Depositor
Resolution determination method	Not provided	Depositor
CTF correction method	MICROGRAPHS	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ( $e^-/\text{\AA}^2$ )	21	Depositor
Minimum defocus (nm)	1500	Depositor
Maximum defocus (nm)	3500	Depositor
Magnification	22500	Depositor
Image detector	GATAN K2 SUMMIT (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	1.70	14/2386 (0.6%)	1.91	46/3218 (1.4%)
10	J	1.75	90/8097 (1.1%)	1.98	214/10972 (2.0%)
11	P	1.69	6/763 (0.8%)	1.87	15/1031 (1.5%)
2	B	1.75	25/1914 (1.3%)	1.91	33/2577 (1.3%)
3	C	1.69	23/2659 (0.9%)	2.01	69/3596 (1.9%)
4	D	1.67	9/1725 (0.5%)	1.86	27/2339 (1.2%)
5	E	1.68	11/2087 (0.5%)	1.90	42/2836 (1.5%)
6	F	1.79	16/1649 (1.0%)	2.04	48/2222 (2.2%)
7	G	1.76	17/1868 (0.9%)	1.99	53/2531 (2.1%)
8	H	1.80	30/2311 (1.3%)	2.02	61/3118 (2.0%)
9	I	1.78	26/1774 (1.5%)	1.98	42/2398 (1.8%)
All	All	1.74	267/27233 (1.0%)	1.96	650/36838 (1.8%)

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	3
10	J	0	26
11	P	0	1
2	B	0	7
3	C	0	10
4	D	0	2
5	E	0	4
6	F	0	5
7	G	0	7
8	H	0	5
9	I	0	6
All	All	0	76

The worst 5 of 267 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	G	110	ARG	CZ-NH2	9.83	1.45	1.33
10	J	886	ARG	CZ-NH2	9.40	1.45	1.33
8	H	74	TYR	CZ-OH	8.53	1.52	1.37
10	J	590	ARG	NE-CZ	8.02	1.43	1.33
1	A	34	ARG	CZ-NH2	7.95	1.43	1.33

The worst 5 of 650 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	7	TYR	CB-CG-CD2	20.64	133.39	121.00
8	H	281	TYR	CB-CG-CD2	-16.23	111.26	121.00
3	C	44	ARG	NE-CZ-NH2	-15.84	112.38	120.30
7	G	110	ARG	NE-CZ-NH1	15.73	128.16	120.30
3	C	262	ARG	NE-CZ-NH1	15.64	128.12	120.30

There are no chirality outliers.

5 of 76 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	269	TYR	Sidechain
1	A	303	ARG	Sidechain
1	A	44	PHE	Sidechain
2	B	23	ARG	Sidechain
2	B	3	ARG	Sidechain

## 5.2 Too-close contacts [i](#)

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	2350	0	2345	4	0
2	B	1890	0	1943	5	0
3	C	2622	0	2688	25	0
4	D	1707	0	1766	4	0
5	E	2048	0	2079	4	0
6	F	1627	0	1607	50	0
7	G	1831	0	1829	2	0
8	H	2277	0	2283	9	0
9	I	1750	0	1755	45	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
10	J	7942	0	7947	12	0
11	P	750	0	759	112	0
All	All	26794	0	27001	157	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 3.

The worst 5 of 157 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)
6:F:49:HIS:NE2	11:P:36:LYS:HD2	1.28	1.43
3:C:51:ASP:CB	11:P:47:PRO:HD3	1.49	1.41
9:I:250:ASN:OD1	11:P:86:LEU:CD1	1.71	1.38
3:C:81:LYS:HE2	11:P:35:LYS:CB	1.45	1.38
6:F:52:PHE:CD2	11:P:32:LEU:HD22	1.61	1.34

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	300/305 (98%)	278 (93%)	18 (6%)	4 (1%)	13 54
2	B	238/246 (97%)	223 (94%)	11 (5%)	4 (2%)	10 49
3	C	327/394 (83%)	294 (90%)	22 (7%)	11 (3%)	4 34
4	D	221/223 (99%)	211 (96%)	9 (4%)	1 (0%)	31 72
5	E	263/265 (99%)	248 (94%)	10 (4%)	5 (2%)	9 46
6	F	205/250 (82%)	194 (95%)	8 (4%)	3 (2%)	11 51
7	G	234/240 (98%)	219 (94%)	14 (6%)	1 (0%)	36 76
8	H	287/359 (80%)	263 (92%)	15 (5%)	9 (3%)	4 36

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
9	I	219/292 (75%)	200 (91%)	14 (6%)	5 (2%)	7	42
10	J	991/1001 (99%)	905 (91%)	68 (7%)	18 (2%)	9	47
11	P	93/747 (12%)	82 (88%)	9 (10%)	2 (2%)	7	43
All	All	3378/4322 (78%)	3117 (92%)	198 (6%)	63 (2%)	13	46

5 of 63 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	251	ALA
3	C	10	ILE
3	C	11	GLU
3	C	12	ILE
3	C	143	ALA

### 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	263/266 (99%)	259 (98%)	4 (2%)	67	84
2	B	214/218 (98%)	207 (97%)	7 (3%)	41	68
3	C	294/349 (84%)	286 (97%)	8 (3%)	48	72
4	D	197/197 (100%)	193 (98%)	4 (2%)	58	79
5	E	240/240 (100%)	233 (97%)	7 (3%)	45	71
6	F	181/219 (83%)	174 (96%)	7 (4%)	35	65
7	G	205/209 (98%)	197 (96%)	8 (4%)	35	65
8	H	252/311 (81%)	245 (97%)	7 (3%)	47	71
9	I	190/240 (79%)	184 (97%)	6 (3%)	42	69
10	J	895/901 (99%)	865 (97%)	30 (3%)	40	68
11	P	89/702 (13%)	88 (99%)	1 (1%)	76	87
All	All	3020/3852 (78%)	2931 (97%)	89 (3%)	49	71

5 of 89 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
7	G	96	SER
8	H	195	ARG
10	J	830	ARG
7	G	99	TYR
8	H	68	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 34 such sidechains are listed below:

Mol	Chain	Res	Type
8	H	185	HIS
9	I	126	ASN
11	P	8	ASN
8	H	302	ASN
2	B	131	GLN

### 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 ⓘ

There are no chain breaks in this entry.