



# wwPDB X-ray Structure Validation Summary Report

Feb 28, 2014 – 12:24 AM GMT

PDB ID : 4IFD  
Title : Crystal structure of an 11-subunit eukaryotic exosome complex bound to RNA  
Authors : Makino, D.L.; Conti, E.  
Deposited on : 2012-12-14  
Resolution : 2.81 Å(reported)

This is a wwPDB validation summary report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

---

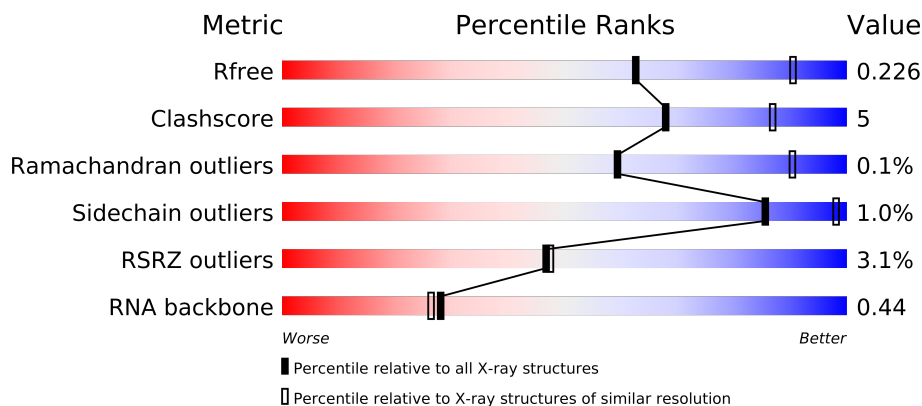
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.15 2013  
Xtriage (Phenix) : dev-1323  
EDS : stable22639  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable22683

# 1 Overall quality at a glance

The reported resolution of this entry is 2.81 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))
$R_{free}$	66092	1799 (2.80-2.80)
Clashscore	79885	2295 (2.80-2.80)
Ramachandran outliers	78287	2252 (2.80-2.80)
Sidechain outliers	78261	2254 (2.80-2.80)
RSRZ outliers	66119	1802 (2.80-2.80)
RNA backbone	1838	1076 (3.30-2.30)

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  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	304	
2	B	248	
3	C	393	
4	D	245	
5	E	267	
6	F	250	
7	G	242	
8	H	361	
9	I	301	
10	J	1003	
11	K	179	
12	R	45	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
13	BR	J	1104	-	X
14	GOL	H	402	-	X
14	GOL	H	403	-	X
14	GOL	H	404	-	X
15	MES	G	302	-	X

## 2 Entry composition

There are 18 unique types of molecules in this entry. The entry contains 26848 atoms, of which 0 are hydrogen and 0 are deuterium.

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 Exosome complex component RRP45.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	300	Total	C	N	O	S	0	0	0
			2301	1443	394	448	16			

- Molecule 2 is a protein called Exosome complex component SKI6.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	242	Total	C	N	O	S	0	0	0
			1882	1177	336	361	8			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-1	GLY	-	EXPRESSION TAG	UNP P46948
B	0	HIS	-	EXPRESSION TAG	UNP P46948

- Molecule 3 is a protein called Exosome complex component RRP43.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	338	Total	C	N	O	S	0	3	0
			2627	1660	453	503	11			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	102	SER	ALA	ENGINEERED MUTATION	UNP P25359
C	363	MET	VAL	ENGINEERED MUTATION	UNP P25359

- Molecule 4 is a protein called Exosome complex component RRP46.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	223	Total	C	N	O	S	0	0	0
			1702	1070	288	334	10			

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	-21	GLY	-	EXPRESSION TAG	UNP P53256
D	-20	HIS	-	EXPRESSION TAG	UNP P53256
D	-19	GLY	-	EXPRESSION TAG	UNP P53256
D	-18	ASN	-	EXPRESSION TAG	UNP P53256
D	-17	ASN	-	EXPRESSION TAG	UNP P53256
D	-16	LYS	-	EXPRESSION TAG	UNP P53256
D	-15	GLU	-	EXPRESSION TAG	UNP P53256
D	-14	PRO	-	EXPRESSION TAG	UNP P53256
D	-13	ASN	-	EXPRESSION TAG	UNP P53256
D	-12	THR	-	EXPRESSION TAG	UNP P53256
D	-11	LYS	-	EXPRESSION TAG	UNP P53256
D	-10	ASN	-	EXPRESSION TAG	UNP P53256
D	-9	ARG	-	EXPRESSION TAG	UNP P53256
D	-8	LEU	-	EXPRESSION TAG	UNP P53256
D	-7	ASP	-	EXPRESSION TAG	UNP P53256
D	-6	SER	-	EXPRESSION TAG	UNP P53256
D	-5	ALA	-	EXPRESSION TAG	UNP P53256
D	-4	GLU	-	EXPRESSION TAG	UNP P53256
D	-3	LYS	-	EXPRESSION TAG	UNP P53256
D	-2	LYS	-	EXPRESSION TAG	UNP P53256
D	-1	LYS	-	EXPRESSION TAG	UNP P53256
D	0	LYS	-	EXPRESSION TAG	UNP P53256

- Molecule 5 is a protein called Exosome complex component RRP42.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	E	267	Total	C	N	O	S	0	0	0
			2042	1304	338	395	5			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	-1	GLY	-	EXPRESSION TAG	UNP Q12277
E	0	HIS	-	EXPRESSION TAG	UNP Q12277
E	138	ILE	VAL	ENGINEERED MUTATION	UNP Q12277

- Molecule 6 is a protein called Exosome complex component MTR3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	F	213	Total	C	N	O	S	0	0	0
			1640	1023	280	327	10			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	75	SER	THR	ENGINEERED MUTATION	UNP P48240
F	161	THR	MET	ENGINEERED MUTATION	UNP P48240

- Molecule 7 is a protein called Exosome complex component RRP40.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	G	236	Total	C	N	O	S	0	0	0
			1804	1150	296	348	10			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	-1	GLY	-	EXPRESSION TAG	UNP Q08285
G	0	HIS	-	EXPRESSION TAG	UNP Q08285

- Molecule 8 is a protein called Exosome complex component RRP4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	H	295	Total	C	N	O	S	0	0	0
			2248	1400	405	431	12			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
H	-1	ARG	-	EXPRESSION TAG	UNP P38792
H	0	SER	-	EXPRESSION TAG	UNP P38792

- Molecule 9 is a protein called Exosome complex component CSL4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
9	I	230	Total	C	N	O	S	0	0	0
			1711	1071	303	329	8			

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
I	-8	MET	-	EXPRESSION TAG	UNP P53859
I	-7	LYS	-	EXPRESSION TAG	UNP P53859
I	-6	HIS	-	EXPRESSION TAG	UNP P53859
I	-5	HIS	-	EXPRESSION TAG	UNP P53859

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
I	-4	HIS	-	EXPRESSION TAG	UNP P53859
I	-3	HIS	-	EXPRESSION TAG	UNP P53859
I	-2	HIS	-	EXPRESSION TAG	UNP P53859
I	-1	HIS	-	EXPRESSION TAG	UNP P53859
I	0	PRO	-	EXPRESSION TAG	UNP P53859

- Molecule 10 is a protein called Exosome complex exonuclease DIS3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
10	J	948	Total	C	N	O	S	0	0	0
			7430	4693	1310	1392	35			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
J	-1	GLY	-	EXPRESSION TAG	UNP Q08162
J	0	ALA	-	EXPRESSION TAG	UNP Q08162
J	171	ASN	ASP	ENGINEERED MUTATION	UNP Q08162
J	551	ASN	ASP	ENGINEERED MUTATION	UNP Q08162

- Molecule 11 is a protein called Exosome complex exonuclease RRP6.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
11	K	81	Total	C	N	O	S	0	0	0
			623	393	110	118	2			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
K	515	ARG	-	EXPRESSION TAG	UNP Q12149
K	516	SER	-	EXPRESSION TAG	UNP Q12149
K	517	MET	-	EXPRESSION TAG	UNP Q12149

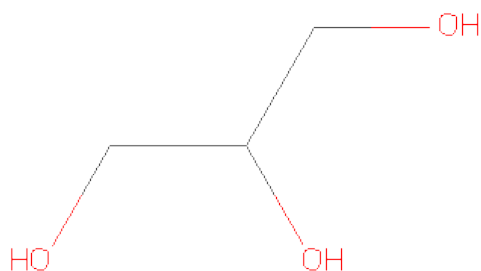
- Molecule 12 is a RNA chain called RNA (45-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
12	R	30	Total	C	N	O	P	0	0	0
			540	237	65	208	30			

- Molecule 13 is BROMIDE ION (three-letter code: BR) (formula: Br).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
13	J	4	Total 4	Br 4	0	0
13	E	1	Total 1	Br 1	0	0
13	H	1	Total 1	Br 1	0	0
13	I	1	Total 1	Br 1	0	0
13	A	2	Total 2	Br 2	0	0
13	F	1	Total 1	Br 1	0	0

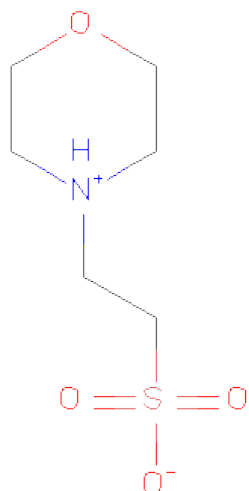
- Molecule 14 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
14	C	1	Total 6	C 3	O 3	0	0
14	G	1	Total 6	C 3	O 3	0	0
14	H	1	Total 6	C 3	O 3	0	0
14	H	1	Total 6	C 3	O 3	0	0
14	H	1	Total 6	C 3	O 3	0	0

- Molecule 15 is 2-(N-MORPHOLINO)-ETHANESULFONICACID (three-letter code: MES)

(formula: C<sub>6</sub>H<sub>13</sub>NO<sub>4</sub>S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
15	G	1	Total	C	N	O	S	0	0
			12	6	1	4	1		

- Molecule 16 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
16	J	1	Total	Mg	0	0
			1	1		

- Molecule 17 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
17	J	1	Total	Zn	0	0
			1	1		

- Molecule 18 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
18	A	35	Total	O	0	0
			35	35		
18	B	35	Total	O	0	0
			35	35		
18	C	20	Total	O	0	0
			20	20		
18	D	38	Total	O	0	0
			38	38		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
18	E	23	Total 23	O 23	0	0
18	F	15	Total 15	O 15	0	0
18	G	19	Total 19	O 19	0	0
18	H	23	Total 23	O 23	0	0
18	I	6	Total 6	O 6	0	0
18	J	27	Total 27	O 27	0	0
18	K	2	Total 2	O 2	0	0
18	R	1	Total 1	O 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 errors 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: Exosome complex component RRP45

Chain A: 



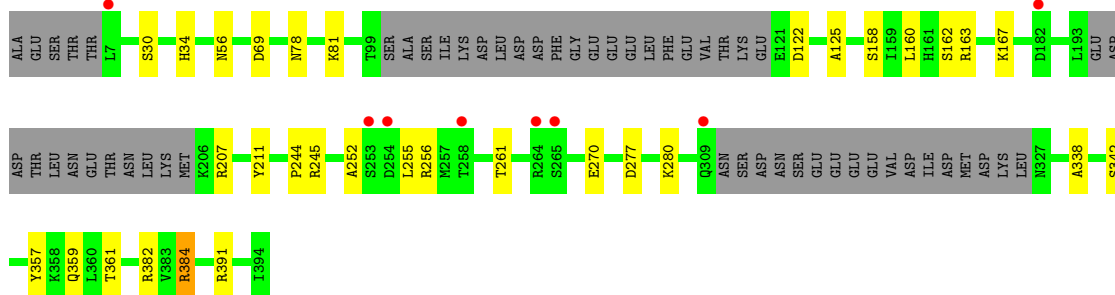
- Molecule 2: Exosome complex component SKI6

Chain B: 



- Molecule 3: Exosome complex component RRP43

Chain C: 



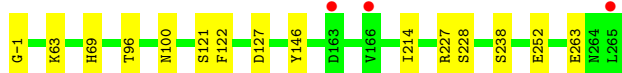
- Molecule 4: Exosome complex component RRP46

Chain D: 

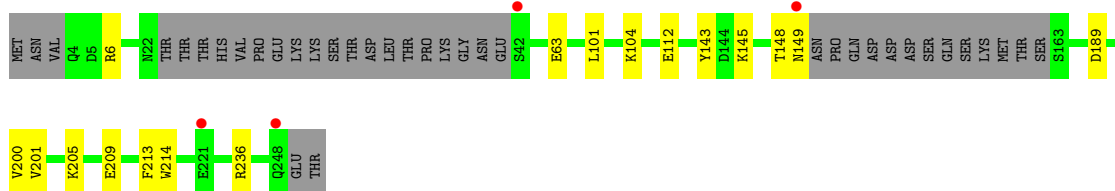


- Molecule 5: Exosome complex component RRP42

Chain E: 



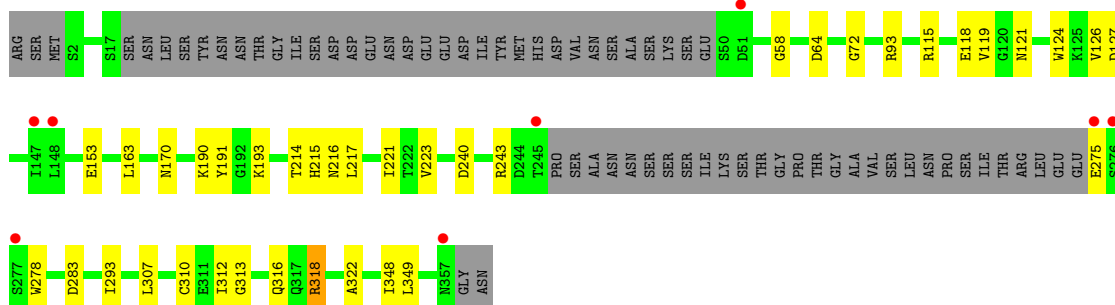
- Chain F:



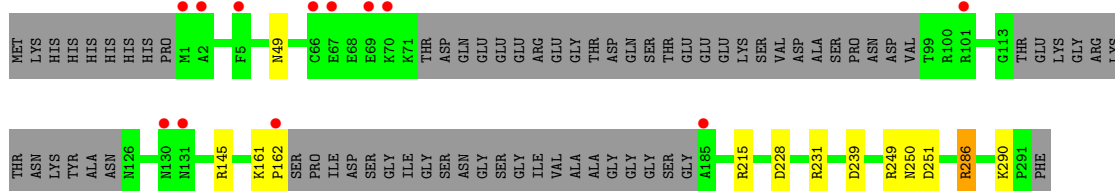
- Chain G: 



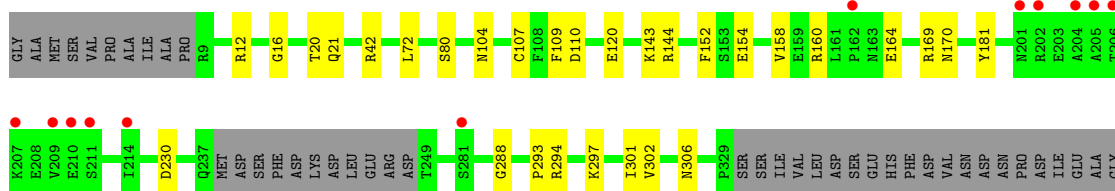
- Chain H:

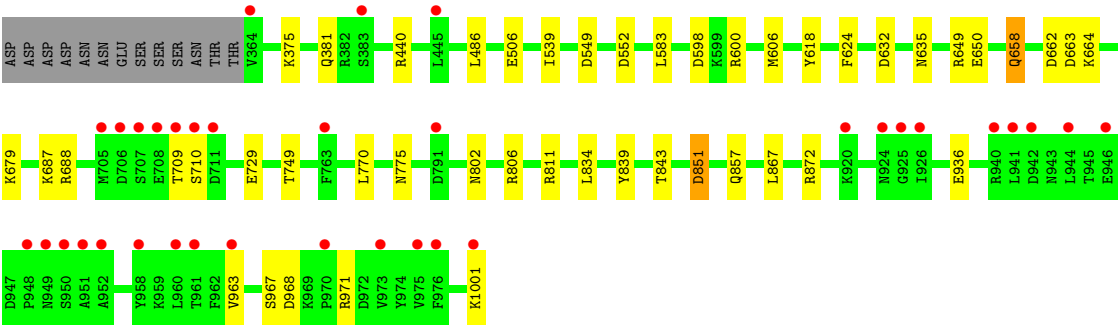


- Chain I: 



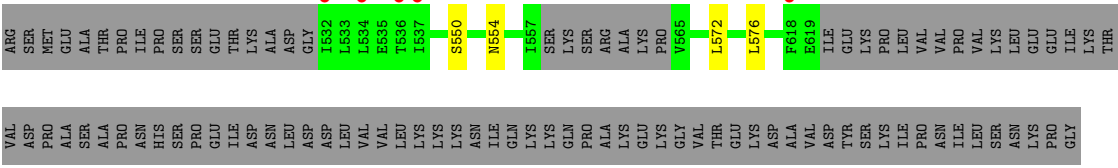
- Chain J: 





• Molecule 11: Exosome complex exonuclease RRP6

Chain K:



• Molecule 12: RNA (45-MER)

Chain R:



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	154.09Å 107.44Å 150.46Å 90.00° 110.63° 90.00°	Depositor
Resolution (Å)	49.64 – 2.81 49.64 – 2.80	Depositor EDS
% Data completeness (in resolution range)	87.9 (49.64-2.81) 87.9 (49.64-2.80)	Depositor EDS
$R_{merge}$	0.18	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.01 (at 2.81Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.1_1168)	Depositor
R, $R_{free}$	0.182 , 0.224 0.187 , 0.226	Depositor DCC
$R_{free}$ test set	4954 reflections (5.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	54.9	Xtriage
Anisotropy	0.333	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 36.4	EDS
Estimated twinning fraction	0.020 for l,-k,h	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 99113 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	26848	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.48% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, MG, ZN, MES, BR

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	0.21	0/2337	0.35	0/3158
2	B	0.20	0/1906	0.36	0/2570
3	C	0.20	0/2670	0.38	0/3619
4	D	0.20	0/1720	0.39	0/2335
5	E	0.21	0/2082	0.37	0/2834
6	F	0.21	0/1662	0.37	0/2240
7	G	0.21	0/1841	0.37	0/2503
8	H	0.21	0/2282	0.38	0/3086
9	I	0.20	0/1735	0.38	0/2354
10	J	0.20	0/7575	0.37	0/10290
11	K	0.21	0/629	0.37	0/850
12	R	0.15	0/595	0.75	0/919
All	All	0.20	0/27034	0.39	0/36758

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2301	0	0	11	0
2	B	1882	0	0	9	0
3	C	2627	0	0	19	0
4	D	1702	0	0	10	0
5	E	2042	0	0	11	0
6	F	1640	0	0	14	0
7	G	1804	0	0	4	0
8	H	2248	0	1370	22	0
9	I	1711	0	0	8	0
10	J	7430	0	0	41	0
11	K	623	0	0	3	0
12	R	540	0	0	2	0
13	A	2	0	0	2	0
13	E	1	0	0	0	0
13	F	1	0	0	0	0
13	H	1	0	0	0	0
13	I	1	0	0	0	0
13	J	4	0	0	5	0
14	C	6	0	8	1	0
14	G	6	0	8	0	0
14	H	18	0	24	1	0
15	G	12	0	12	1	0
16	J	1	0	0	0	0
17	J	1	0	0	0	0
18	A	35	0	0	1	0
18	B	35	0	0	0	0
18	C	20	0	0	1	0
18	D	38	0	0	1	0
18	E	23	0	0	2	0
18	F	15	0	0	1	0
18	G	19	0	0	0	0
18	H	23	0	0	0	0
18	I	6	0	0	1	0
18	J	27	0	0	0	0
18	K	2	0	0	0	0
18	R	1	0	0	0	0
All	All	26848	0	1422	136	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 5.

The worst 5 of 136 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
10:J:811:ARG:NH2	10:J:936:GLU:OE2	2.22	0.71
3:C:357:TYR:OH	3:C:382:ARG:NH1	2.23	0.71
8:H:216:ASN:HD21	8:H:240:ASP:HB3	1.54	0.71
8:H:313:GLY:O	8:H:318:ARG:NH1	2.25	0.69
9:I:249:ARG:NH1	9:I:251:ASP:OD1	2.27	0.68

There are no symmetry-related clashes.

## 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	298/304 (98%)	288 (97%)	10 (3%)	0	100	100
2	B	240/248 (97%)	235 (98%)	4 (2%)	1 (0%)	43	80
3	C	333/393 (85%)	313 (94%)	20 (6%)	0	100	100
4	D	221/245 (90%)	215 (97%)	5 (2%)	1 (0%)	38	76
5	E	265/267 (99%)	249 (94%)	16 (6%)	0	100	100
6	F	207/250 (83%)	192 (93%)	15 (7%)	0	100	100
7	G	234/242 (97%)	227 (97%)	7 (3%)	0	100	100
8	H	289/361 (80%)	281 (97%)	7 (2%)	1 (0%)	50	84
9	I	222/301 (74%)	216 (97%)	6 (3%)	0	100	100
10	J	942/1003 (94%)	901 (96%)	40 (4%)	1 (0%)	59	90
11	K	77/179 (43%)	71 (92%)	6 (8%)	0	100	100
All	All	3328/3793 (88%)	3188 (96%)	136 (4%)	4 (0%)	59	90

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	167	ASP
4	D	140	LYS
8	H	121	ASN
10	J	598	ASP

### 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	253/265 (96%)	251 (99%)	2 (1%)	89	98
2	B	210/219 (96%)	209 (100%)	1 (0%)	94	99
3	C	290/349 (83%)	285 (98%)	5 (2%)	73	95
4	D	196/216 (91%)	196 (100%)	0	100	100
5	E	236/241 (98%)	235 (100%)	1 (0%)	95	99
6	F	182/219 (83%)	181 (100%)	1 (0%)	94	99
7	G	198/210 (94%)	195 (98%)	3 (2%)	76	96
8	H	242/313 (77%)	239 (99%)	3 (1%)	82	97
9	I	177/249 (71%)	175 (99%)	2 (1%)	84	98
10	J	812/901 (90%)	801 (99%)	11 (1%)	78	96
11	K	69/163 (42%)	69 (100%)	0	100	100
All	All	2865/3345 (86%)	2836 (99%)	29 (1%)	85	98

5 of 29 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
8	H	119	VAL
9	I	49	ASN
10	J	834	LEU
8	H	126	VAL
9	I	286	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
8	H	174	GLN

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
12	R	28/45 (62%)	9 (32%)	0

5 of 9 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
12	R	-39	A
12	R	-37	A
12	R	-33	G
12	R	-31	U
12	R	-14	U

There are no RNA pucker outliers to report.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 18 ligands modelled in this entry, 12 are monoatomic - leaving 6 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
14	GOL	C	401	-	5,5,5	0.33	0	5,5,5	0.26	0
14	GOL	G	301	-	5,5,5	0.33	0	5,5,5	0.26	0
15	MES	G	302	-	12,12,12	2.13	1 (8%)	16,16,16	2.08	4 (25%)
14	GOL	H	402	-	5,5,5	0.34	0	5,5,5	0.23	0
14	GOL	H	403	-	5,5,5	0.33	0	5,5,5	0.28	0
14	GOL	H	404	-	5,5,5	0.33	0	5,5,5	0.24	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
14	GOL	C	401	-	-	0/4/4/4	0/0/0/0
14	GOL	G	301	-	-	0/4/4/4	0/0/0/0
15	MES	G	302	-	-	0/6/14/14	0/1/1/1
14	GOL	H	402	-	-	0/4/4/4	0/0/0/0
14	GOL	H	403	-	-	0/4/4/4	0/0/0/0
14	GOL	H	404	-	-	0/4/4/4	0/0/0/0

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
15	G	302	MES	C8-S	-6.47	1.66	1.78

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
15	G	302	MES	C8-C7-N4	-4.09	105.44	112.44
15	G	302	MES	C5-N4-C3	4.00	117.68	109.75
15	G	302	MES	C6-C5-N4	-3.53	105.44	109.96
15	G	302	MES	C7-N4-C5	2.12	117.20	111.66

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 5.7 Other polymers ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	300/304 (98%)	-0.37	6 (2%) 62 63	31, 49, 110, 147	0
2	B	242/248 (97%)	-0.53	1 (0%) 90 91	29, 44, 88, 132	0
3	C	338/393 (86%)	-0.16	8 (2%) 56 57	35, 57, 127, 162	0
4	D	223/245 (91%)	-0.44	2 (0%) 81 81	32, 48, 76, 120	0
5	E	267/267 (100%)	-0.31	3 (1%) 77 78	35, 58, 106, 131	0
6	F	213/250 (85%)	-0.28	4 (1%) 64 64	36, 58, 115, 141	0
7	G	236/242 (97%)	-0.31	5 (2%) 60 61	32, 58, 122, 159	0
8	H	295/361 (81%)	-0.27	8 (2%) 52 52	38, 59, 114, 140	0
9	I	230/301 (76%)	0.13	12 (5%) 26 26	51, 81, 128, 140	0
10	J	948/1003 (94%)	-0.02	47 (4%) 28 28	47, 81, 136, 169	0
11	K	81/179 (45%)	0.12	5 (6%) 20 19	48, 90, 131, 143	0
12	R	30/45 (66%)	1.22	6 (20%) 2 1	72, 163, 201, 210	0
All	All	3403/3838 (88%)	-0.18	107 (3%) 47 47	29, 64, 128, 210	0

The worst 5 of 107 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
10	J	209	VAL	6.9
8	H	276	SER	6.8
10	J	206	THR	6.6
11	K	532	ILE	6.5
10	J	710	SER	6.3

### 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 ⓘ

There are no carbohydrates in this entry.

### 6.4 Ligands ⓘ

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
14	GOL	H	403	6/6	0.26	11.86	72,74,77,81	0
14	GOL	H	404	6/6	0.24	4.33	58,64,66,67	0
13	BR	J	1104	1/1	0.29	2.78	137,137,137,137	0
15	MES	G	302	12/12	0.25	2.54	77,107,110,111	0
14	GOL	H	402	6/6	0.21	2.24	66,72,77,77	0
14	GOL	G	301	6/6	0.19	1.58	81,85,91,92	0
13	BR	A	402	1/1	0.20	1.04	136,136,136,136	0
13	BR	J	1103	1/1	0.13	-0.08	110,110,110,110	0
14	GOL	C	401	6/6	0.17	-0.32	79,83,84,84	0
16	MG	J	1105	1/1	0.14	-0.60	72,72,72,72	0
17	ZN	J	1106	1/1	0.12	-0.97	62,62,62,62	0
13	BR	H	401	1/1	0.10	-1.20	75,75,75,75	0
13	BR	J	1101	1/1	0.09	-2.39	76,76,76,76	0
13	BR	I	301	1/1	0.06	-2.99	90,90,90,90	0
13	BR	J	1102	1/1	0.12	-3.56	106,106,106,106	0
13	BR	A	401	1/1	0.08	-3.94	94,94,94,94	0
13	BR	E	301	1/1	0.08	-3.96	101,101,101,101	0
13	BR	F	301	1/1	0.03	-4.12	119,119,119,119	0

### 6.5 Other polymers ⓘ

There are no such residues in this entry.