



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 08:23 am BST

PDB ID : 5U5Q  
Title : 12 Subunit RNA Polymerase II at Room Temperature collected using SFX  
Authors : Bushnell, D.A.; Oberthur, D.; Mariani, V.; Yefanov, O.; Tolstikova, A.; Barty, A.  
Deposited on : 2016-12-07  
Resolution : 3.80 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](mailto: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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Xtriage (Phenix) : 1.13  
EDS : 2.11  
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) : 2.11

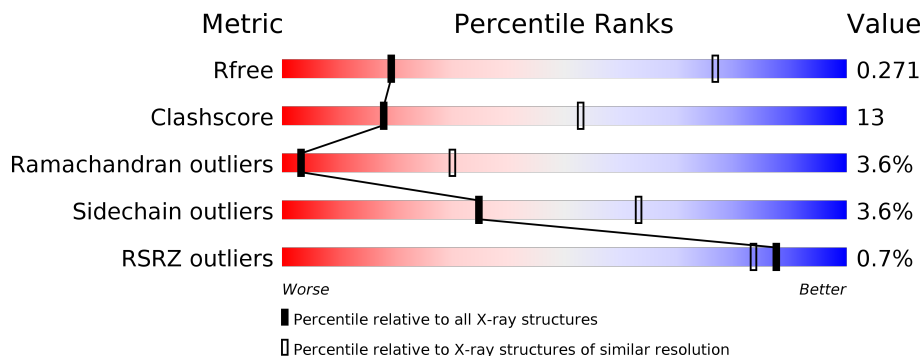
# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.80 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	1212 (4.00-3.60)
Clashscore	141614	1288 (4.00-3.60)
Ramachandran outliers	138981	1243 (4.00-3.60)
Sidechain outliers	138945	1237 (4.00-3.60)
RSRZ outliers	127900	1121 (4.00-3.60)

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 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	1733	
2	B	1224	
3	C	318	
4	D	221	
5	E	215	
6	F	155	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
7	G	171	 67% 32% .
8	H	146	 % 55% 36% . 6%
9	I	122	 % 73% 24% ..
10	J	70	 49% 36% 9% 7%
11	K	120	 % 76% 18% . .
12	L	70	 40% 23% . 34%

## 2 Entry composition [i](#)

There are 14 unique types of molecules in this entry. The entry contains 31634 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 DNA-directed RNA polymerase II subunit RPB1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	1434	11285	7106	1968	2149	62	38	0	0

- Molecule 2 is a protein called DNA-directed RNA polymerase II subunit RPB2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	1150	9138	5778	1601	1703	56	50	0	0

- Molecule 3 is a protein called DNA-directed RNA polymerase II subunit RPB3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	266	2095	1317	348	417	13	0	0	0

- Molecule 4 is a protein called DNA-directed RNA polymerase II subunit RPB4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	184	1415	877	250	285	3	0	0	0

- Molecule 5 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	E	215	1760	1116	310	322	12	0	0	0

- Molecule 6 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
6	F	87	705	451	119	132	3	0	0	0

- Molecule 7 is a protein called DNA-directed RNA polymerase II subunit RPB7.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
7	G	171	1340	861	222	249	8	0	0	0

- Molecule 8 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
8	H	137	1101	692	185	220	4	0	0	0

- Molecule 9 is a protein called DNA-directed RNA polymerase II subunit RPB9.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
9	I	119	971	596	179	186	10	0	0	0

- Molecule 10 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC5.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
10	J	65	532	339	93	94	6	0	0	0

- Molecule 11 is a protein called DNA-directed RNA polymerase II subunit RPB11.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
11	K	115	920	590	157	171	2	0	0	1

- Molecule 12 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
12	L	46	363	224	72	63	4	0	0	0

- Molecule 13 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
13	J	1	Total	Zn	0	0
			1	1		
13	B	1	Total	Zn	0	0
			1	1		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
13	I	2	Total 2	Zn 2	0	0
13	C	1	Total 1	Zn 1	0	0
13	A	2	Total 2	Zn 2	0	0
13	L	1	Total 1	Zn 1	0	0

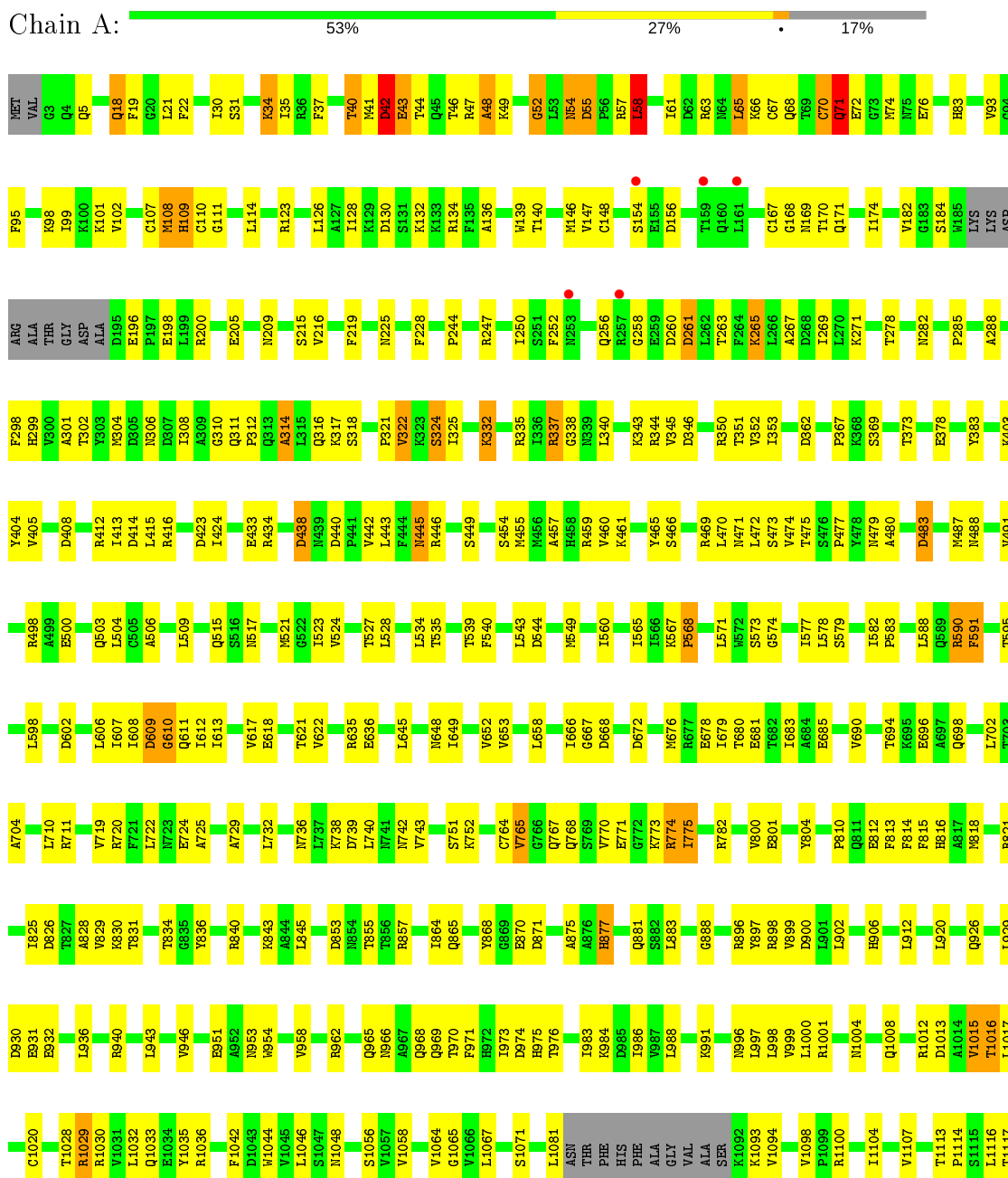
- Molecule 14 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
14	A	1	Total 1	Mg 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 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: DNA-directed RNA polymerase II subunit RPB1












Chain J:  49% 36% 9% 7%



- Molecule 11: DNA-directed RNA polymerase II subunit RPB11

Chain K:  % 76% 18% . .



- Molecule 12: DNA-directed RNA polymerases I, II, and III subunit RPABC4

Chain L:  40% 23% . 34%



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	223.40Å 396.80Å 287.90Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.61 – 3.80 40.61 – 3.80	Depositor EDS
% Data completeness (in resolution range)	98.9 (40.61-3.80) 98.9 (40.61-3.80)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.17 (at 3.76Å)	Xtriage
Refinement program	PHENIX (dev_2356: ???)	Depositor
R, $R_{free}$	0.218 , 0.271 0.218 , 0.271	Depositor DCC
$R_{free}$ test set	2493 reflections (2.01%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	141.0	Xtriage
Anisotropy	0.340	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.24 , 109.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.37$ , $\langle L^2 \rangle = 0.19$	Xtriage
Estimated twinning fraction	0.105 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.118 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	31634	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	169.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.91% of the height of the origin peak. No significant pseudotranslation is detected.*

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

<sup>2</sup> Theoretical 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](#)

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

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	0.25	0/11488	0.44	0/15539
2	B	0.25	0/9317	0.46	1/12566 (0.0%)
3	C	0.24	0/2133	0.43	0/2891
4	D	0.24	0/1426	0.41	0/1919
5	E	0.38	1/1796 (0.1%)	0.41	0/2416
6	F	0.24	0/717	0.41	0/967
7	G	0.25	0/1368	0.45	0/1844
8	H	0.24	0/1119	0.46	0/1515
9	I	0.25	0/989	0.46	0/1331
10	J	0.25	0/541	0.47	0/727
11	K	0.25	0/938	0.42	0/1267
12	L	0.23	0/365	0.46	0/485
All	All	0.26	1/32197 (0.0%)	0.44	1/43467 (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	A	0	1
2	B	0	4
All	All	0	5

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	E	124	VAL	C-N	12.32	1.57	1.34

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
2	B	758	PHE	CB-CG-CD1	-5.04	117.27	120.80

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	55	ASP	Peptide
2	B	755	ILE	Peptide
2	B	756	ILE	Peptide
2	B	757	PRO	Peptide
2	B	758	PHE	Peptide

## 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	11285	0	11338	331	0
2	B	9138	0	9158	285	0
3	C	2095	0	2052	70	0
4	D	1415	0	1368	45	0
5	E	1760	0	1788	36	0
6	F	705	0	731	22	0
7	G	1340	0	1357	33	0
8	H	1101	0	1067	36	0
9	I	971	0	929	19	0
10	J	532	0	544	33	0
11	K	920	0	929	17	0
12	L	363	0	386	11	0
13	A	2	0	0	0	0
13	B	1	0	0	0	0
13	C	1	0	0	0	0
13	I	2	0	0	0	0
13	J	1	0	0	0	0
13	L	1	0	0	0	0
14	A	1	0	0	0	0
All	All	31634	0	31647	851	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 13.

The worst 5 of 851 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)
4:D:66:ARG:O	4:D:70:PHE:HB2	1.52	1.10
2:B:475:SER:H	2:B:476:ARG:HA	1.35	0.91
7:G:34:VAL:HG12	7:G:45:ILE:HG21	1.62	0.82
3:C:66:ARG:NH2	10:J:3:VAL:O	2.13	0.81
2:B:983:ARG:HD2	2:B:1091:TYR:HB3	1.63	0.80

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 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	1428/1733 (82%)	1222 (86%)	149 (10%)	57 (4%)	3	28
2	B	1142/1224 (93%)	943 (83%)	155 (14%)	44 (4%)	3	28
3	C	264/318 (83%)	221 (84%)	35 (13%)	8 (3%)	4	33
4	D	180/221 (81%)	155 (86%)	19 (11%)	6 (3%)	4	32
5	E	213/215 (99%)	188 (88%)	21 (10%)	4 (2%)	8	42
6	F	85/155 (55%)	77 (91%)	7 (8%)	1 (1%)	13	50
7	G	169/171 (99%)	156 (92%)	11 (6%)	2 (1%)	13	50
8	H	133/146 (91%)	103 (77%)	22 (16%)	8 (6%)	1	20
9	I	117/122 (96%)	98 (84%)	15 (13%)	4 (3%)	3	31
10	J	63/70 (90%)	51 (81%)	9 (14%)	3 (5%)	2	24
11	K	113/120 (94%)	105 (93%)	8 (7%)	0	100	100
12	L	44/70 (63%)	25 (57%)	14 (32%)	5 (11%)	0	7
All	All	3951/4565 (86%)	3344 (85%)	465 (12%)	142 (4%)	3	30

5 of 142 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	47	ARG
1	A	54	ASN
1	A	1114	PRO
1	A	1416	ALA
1	A	1437	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	1255/1520 (83%)	1205 (96%)	50 (4%)	31	59
2	B	995/1061 (94%)	957 (96%)	38 (4%)	33	61
3	C	234/274 (85%)	230 (98%)	4 (2%)	60	78
4	D	147/200 (74%)	144 (98%)	3 (2%)	55	75
5	E	197/197 (100%)	193 (98%)	4 (2%)	55	75
6	F	77/137 (56%)	74 (96%)	3 (4%)	32	60
7	G	152/152 (100%)	144 (95%)	8 (5%)	22	54
8	H	121/128 (94%)	118 (98%)	3 (2%)	47	70
9	I	113/116 (97%)	110 (97%)	3 (3%)	44	69
10	J	60/65 (92%)	56 (93%)	4 (7%)	16	47
11	K	99/102 (97%)	96 (97%)	3 (3%)	41	66
12	L	40/57 (70%)	38 (95%)	2 (5%)	24	55
All	All	3490/4009 (87%)	3365 (96%)	125 (4%)	35	63

5 of 125 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	199	MET
2	B	670	GLU
9	I	103	CYS
2	B	261	ARG

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
2	B	401	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	449	ASN
4	D	115	HIS
3	C	31	ASN
1	A	1140	HIS
2	B	770	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](#)

Of 9 ligands modelled in this entry, 9 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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, 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	1429/1733 (82%)	-0.50	8 (0%) 89 85	57, 151, 244, 348	0
2	B	1144/1224 (93%)	-0.38	16 (1%) 75 68	78, 157, 266, 369	0
3	C	266/318 (83%)	-0.56	0 100 100	83, 158, 209, 278	0
4	D	184/221 (83%)	-0.56	0 100 100	135, 220, 290, 320	0
5	E	215/215 (100%)	-0.51	1 (0%) 91 87	102, 201, 277, 336	0
6	F	87/155 (56%)	-0.69	0 100 100	71, 124, 184, 213	0
7	G	171/171 (100%)	-0.44	0 100 100	121, 185, 244, 270	0
8	H	137/146 (93%)	-0.47	2 (1%) 73 66	140, 201, 276, 327	0
9	I	119/122 (97%)	-0.65	1 (0%) 86 81	121, 172, 245, 277	0
10	J	65/70 (92%)	-0.51	0 100 100	110, 154, 200, 238	0
11	K	115/120 (95%)	-0.54	1 (0%) 84 79	85, 141, 206, 267	0
12	L	46/70 (65%)	-0.48	0 100 100	123, 198, 252, 288	0
All	All	3978/4565 (87%)	-0.48	29 (0%) 87 83	57, 161, 260, 369	0

The worst 5 of 29 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	441	ASP	4.9
2	B	509	ALA	3.7
8	H	2	SER	3.6
2	B	467	GLY	3.6
1	A	253	ASN	3.4

### 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 carbohydrates in this entry.

### 6.4 Ligands [i](#)

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. 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	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
13	ZN	L	101	1/1	0.96	0.10	135,135,135,135	0
13	ZN	A	1801	1/1	0.96	0.06	233,233,233,233	0
13	ZN	I	201	1/1	0.97	0.08	100,100,100,100	0
13	ZN	A	1802	1/1	0.98	0.06	107,107,107,107	0
14	MG	A	1803	1/1	0.98	0.18	67,67,67,67	0
13	ZN	C	401	1/1	0.99	0.16	171,171,171,171	0
13	ZN	B	1301	1/1	0.99	0.09	122,122,122,122	0
13	ZN	I	202	1/1	0.99	0.03	193,193,193,193	0
13	ZN	J	101	1/1	0.99	0.21	131,131,131,131	0

### 6.5 Other polymers [i](#)

There are no such residues in this entry.