



# Full wwPDB X-ray Structure Validation Report

Feb 28, 2014 – 11:12 PM GMT

PDB ID : 3S2D  
Title : RNA Polymerase II Initiation Complex with a 5-nt RNA containing a 5Br-U  
Authors : Liu, X.; Bushnell, D.A.; Silva, D.A.; Huang, X.; Kornberg, R.D.  
Deposited on : 2011-05-16  
Resolution : 3.20 Å(reported)

This is a full wwPDB validation 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>

---

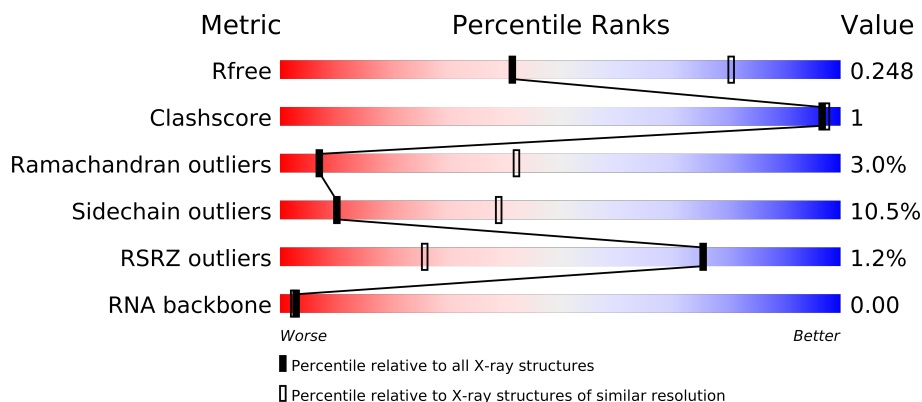
The following versions of software and data (see [references](#)) were used in the production of this report:

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 3.20 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)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	66092	1824 (3.30-3.10)
Clashscore	79885	1078 (3.26-3.14)
Ramachandran outliers	78287	1059 (3.26-3.14)
Sidechain outliers	78261	1058 (3.26-3.14)
RSRZ outliers	66119	1825 (3.30-3.10)
RNA backbone	1838	1002 (3.72-2.68)

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	1733	
2	B	1224	
3	C	318	
4	E	215	
5	F	155	
6	H	146	
7	I	122	
8	J	70	
9	K	120	
10	L	70	
11	R	5	
12	T	29	

## 2 Entry composition

There are 14 unique types of molecules in this entry. The entry contains 28672 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 DNA-directed RNA polymerase II subunit RPB1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1405	Total	C	N	O	S	0	0	0
			11043	6965	1936	2081	61			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	1114	Total	C	N	O	S	0	0	0
			8861	5610	1549	1647	55			

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

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

- Molecule 4 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	E	214	Total	C	N	O	S	0	0	0
			1752	1111	309	321	11			

- Molecule 5 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	85	Total	C	N	O	S	0	0	0
			688	439	116	130	3			

- Molecule 6 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	H	133	Total	C	N	O	S	0	0	0
			1068	673	180	211	4			

- Molecule 7 is a protein called DNA-directed RNA polymerase II subunit RPB9.

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

- Molecule 8 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC5.

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

- Molecule 9 is a protein called DNA-directed RNA polymerase II subunit RPB11.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
9	K	114	Total	C	N	O	S	0	0	0
			919	590	156	171	2			

- Molecule 10 is a protein called DNA-directed RNA polymerases I, II, and III subunit RPABC4.

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

- Molecule 11 is a RNA chain called RNA (5'-R(\*AP\*GP\*GP\*(5BU)P\*G)-3').

Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
11	R	5	Total	Br	C	N	O	P		0	0	0
			109	1	49	22	33	4				

- Molecule 12 is a DNA chain called DNA (5'-D(\*CP\*TP\*AP\*CP\*CP\*GP\*AP\*TP\*AP\*AP\*GP\*CP\*AP\*GP\*AP\*CP\*GP\*AP\*TP\*CP\*AP\*CP\*CP\*TP\*CP\*GP\*AP\*TP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
12	T	13	Total	C	N	O	P	0	0	0
			262	125	46	78	13			

- 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		

*Continued on next page...*

*Continued from previous page...*

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

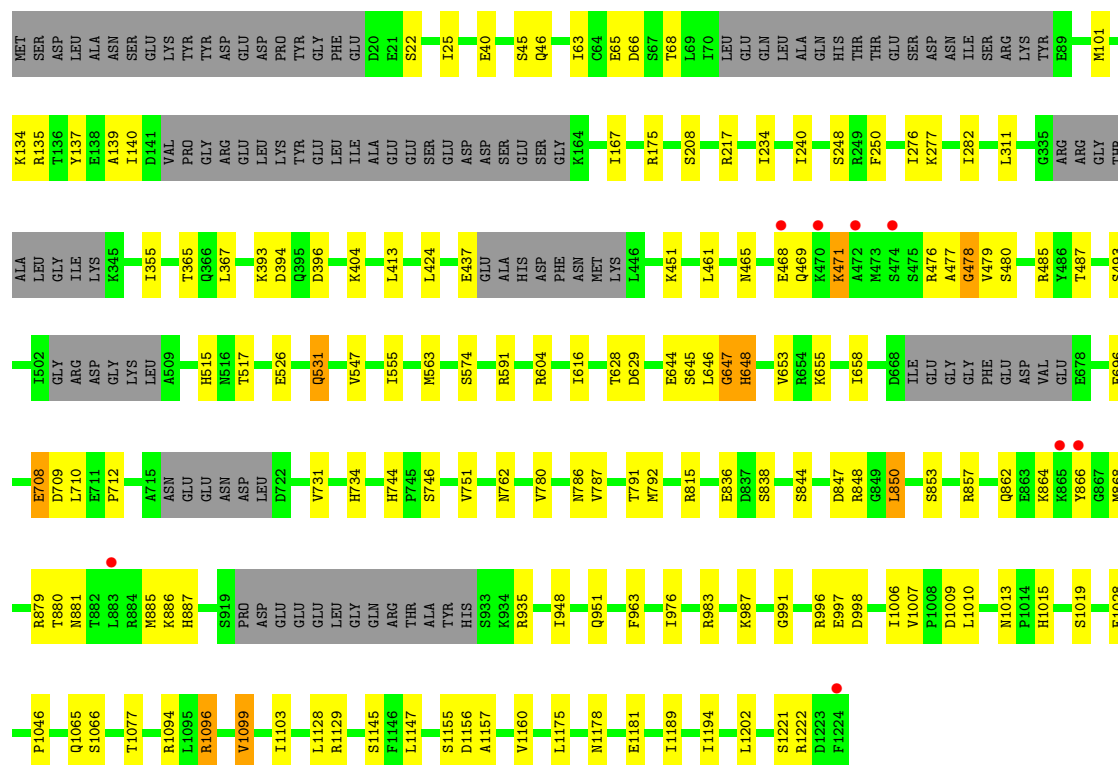
- Molecule 14 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
14	A	1	Total 1	Mg 1	0	0



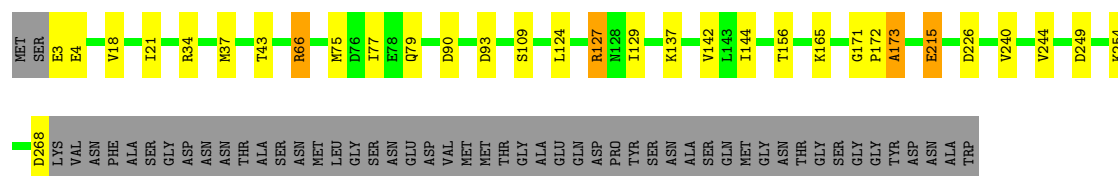
- Molecule 2: DNA-directed RNA polymerase II subunit RPB2

Chain B:



- Molecule 3: DNA-directed RNA polymerase II subunit RPB3

Chain C:



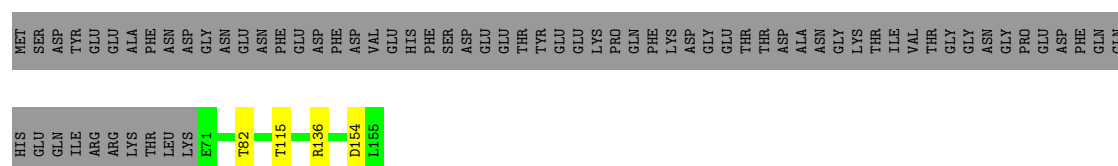
- Molecule 4: DNA-directed RNA polymerases I, II, and III subunit RPABC1

Chain E:



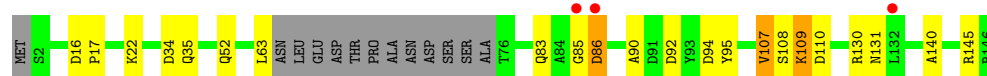
- Molecule 5: DNA-directed RNA polymerases I, II, and III subunit RPABC2

Chain F:



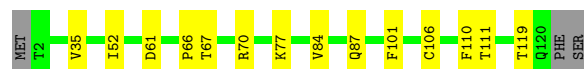
- Molecule 6: DNA-directed RNA polymerases I, II, and III subunit RPABC3

Chain H: 



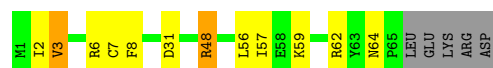
- Molecule 7: DNA-directed RNA polymerase II subunit RPB9

Chain I: 



- Molecule 8: DNA-directed RNA polymerases I, II, and III subunit RPABC5

Chain J: 



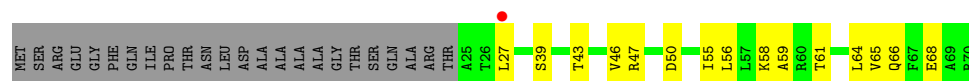
- Molecule 9: DNA-directed RNA polymerase II subunit RPB11

Chain K: 



- Molecule 10: DNA-directed RNA polymerases I, II, and III subunit RPABC4

Chain L: 



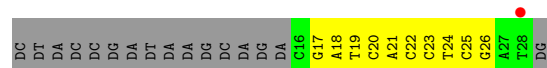
- Molecule 11: RNA (5'-R(\*AP\*GP\*GP\*(5BU)P\*G)-3')

Chain R: 



- Molecule 12: DNA (5'-D(\*CP\*TP\*AP\*CP\*CP\*GP\*AP\*TP\*AP\*AP\*GP\*CP\*AP\*GP\*AP\*C  
P\*GP\*AP\*TP\*CP\*AP\*CP\*CP\*TP\*CP\*GP\*AP\*TP\*G)-3')

Chain T: 





## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	158.29Å 220.77Å 192.24Å 90.00° 97.69° 90.00°	Depositor
Resolution (Å)	47.76 – 3.20 47.76 – 3.19	Depositor EDS
% Data completeness (in resolution range)	(Not available) (47.76-3.20) 98.5 (47.76-3.19)	Depositor EDS
$R_{merge}$	0.15	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.40 (at 3.19Å)	Xtriage
Refinement program	BUSTER 2.8.0	Depositor
R, $R_{free}$	0.183 , 0.231 0.202 , 0.248	Depositor DCC
$R_{free}$ test set	5353 reflections (4.99%)	DCC
Wilson B-factor (Å <sup>2</sup> )	67.2	Xtriage
Anisotropy	0.685	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 56.7	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.45$ , $\langle L^2 \rangle = 0.28$	Xtriage
Outliers	2 of 107358 reflections (0.002%)	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	28672	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	94.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.36% of the height of the origin peak. No significant pseudotranslation is detected.*

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

## 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: 5BU, 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	0.44	0/11241	0.72	1/15199 (0.0%)
2	B	0.46	0/9033	0.74	4/12181 (0.0%)
3	C	0.42	0/2133	0.76	1/2891 (0.0%)
4	E	0.42	0/1788	0.66	0/2406
5	F	0.44	0/700	0.66	0/945
6	H	0.43	0/1086	0.74	0/1470
7	I	0.41	0/989	0.72	0/1331
8	J	0.48	0/541	0.81	0/727
9	K	0.40	0/937	0.65	0/1265
10	L	0.49	0/365	0.84	0/485
11	R	0.99	0/99	1.65	2/154 (1.3%)
12	T	1.08	0/292	1.86	11/447 (2.5%)
All	All	0.46	0/29204	0.76	19/39501 (0.0%)

There are no bond length outliers.

All (19) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	647	GLY	C-N-CA	7.61	140.72	121.70
12	T	17	DG	O4'-C1'-N9	7.56	113.29	108.00
12	T	24	DT	O4'-C1'-N1	7.48	113.24	108.00
12	T	26	DG	P-O3'-C3'	7.27	128.43	119.70
12	T	22	DC	O4'-C4'-C3'	-7.10	101.66	104.50
3	C	172	PRO	C-N-CA	6.83	138.77	121.70
12	T	17	DG	C1'-O4'-C4'	-6.67	103.43	110.10
2	B	628	THR	C-N-CA	6.03	136.77	121.70
11	R	8	G	C4'-C3'-C2'	-6.00	96.60	102.60
11	R	10	G	O4'-C1'-N9	5.85	112.88	108.20
12	T	23	DC	O4'-C1'-N1	5.85	112.09	108.00
12	T	20	DC	O4'-C1'-N1	5.81	112.07	108.00

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
12	T	19	DT	C4-C5-C7	5.54	122.33	119.00
12	T	22	DC	C4'-C3'-C2'	-5.40	98.24	103.10
2	B	648	HIS	N-CA-CB	5.33	120.18	110.60
12	T	18	DA	O4'-C1'-N9	5.27	111.69	108.00
2	B	140	ILE	C-N-CA	5.23	134.76	121.70
1	A	1123	GLY	C-N-CA	5.15	134.58	121.70
12	T	21	DA	C4'-C3'-C2'	-5.02	98.58	103.10

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	11043	0	23	10	0
2	B	8861	0	0	11	0
3	C	2095	0	0	4	0
4	E	1752	0	0	0	0
5	F	688	0	5	1	0
6	H	1068	0	0	1	0
7	I	971	0	0	2	0
8	J	532	0	0	3	0
9	K	919	0	0	3	0
10	L	363	0	0	0	0
11	R	109	0	0	0	0
12	T	262	0	0	1	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	28672	0	28	26	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 1.

All (26) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:469:ARG:NH2	2:B:991:GLY:O	2.26	0.68
1:A:871:ASP:OD1	1:A:1366:ARG:NH2	2.28	0.66
1:A:399:HIS:O	1:A:401:GLY:N	2.29	0.66
3:C:66:ARG:NH2	8:J:3:VAL:O	2.35	0.59
2:B:1013:ASN:ND2	2:B:1015:HIS:CD2	2.77	0.53
9:K:38:GLU:O	9:K:69:ALA:O	2.30	0.49
3:C:165:LYS:O	9:K:6:ARG:NH1	2.46	0.49
2:B:1009:ASP:OD2	8:J:48:ARG:NH2	2.46	0.49
1:A:839:ARG:NH2	1:A:1402:PHE:O	2.44	0.49
1:A:55:ASP:O	1:A:57:ARG:N	2.46	0.48
2:B:515:HIS:CD2	2:B:517:THR:OG1	2.67	0.47
1:A:567:LYS:NZ	6:H:95:TYR:CZ	2.85	0.44
1:A:789:LYS:N	7:I:67:THR:O	2.51	0.44
5:F:82:THR:O	5:F:136:ARG:NH1	2.51	0.43
1:A:567:LYS:CB	1:A:568:PRO:CD	2.96	0.43
2:B:708:GLU:O	2:B:710:LEU:N	2.52	0.43
2:B:848:ARG:NH1	8:J:8:PHE:O	2.52	0.43
7:I:101:PHE:N	7:I:110:PHE:O	2.51	0.43
3:C:171:GLY:C	3:C:173:ALA:N	2.70	0.43
2:B:857:ARG:NH2	12:T:25:DC:OP1	2.53	0.42
3:C:93:ASP:O	3:C:127:ARG:NH2	2.52	0.41
2:B:744:HIS:CD2	2:B:746:SER:OG	2.72	0.41
2:B:847:ASP:OD2	9:K:6:ARG:NH2	2.53	0.41
1:A:840:ARG:NH2	1:A:1106:ASN:OD1	2.54	0.41
1:A:482:PHE:N	2:B:836:GLU:O	2.54	0.41
2:B:476:ARG:O	2:B:478:GLY:N	2.54	0.41

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	1395/1733 (80%)	1234 (88%)	124 (9%)	37 (3%)	8	46
2	B	1096/1224 (90%)	962 (88%)	96 (9%)	38 (4%)	6	37
3	C	264/318 (83%)	239 (90%)	21 (8%)	4 (2%)	15	64
4	E	212/215 (99%)	203 (96%)	7 (3%)	2 (1%)	25	76
5	F	83/155 (54%)	75 (90%)	7 (8%)	1 (1%)	19	70
6	H	129/146 (88%)	103 (80%)	16 (12%)	10 (8%)	1	11
7	I	117/122 (96%)	105 (90%)	11 (9%)	1 (1%)	25	76
8	J	63/70 (90%)	58 (92%)	2 (3%)	3 (5%)	4	27
9	K	112/120 (93%)	107 (96%)	4 (4%)	1 (1%)	25	76
10	L	44/70 (63%)	30 (68%)	7 (16%)	7 (16%)	0	1
All	All	3515/4173 (84%)	3116 (89%)	295 (8%)	104 (3%)	7	42

All (104) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	40	THR
1	A	54	ASN
1	A	55	ASP
1	A	250	ILE
1	A	399	HIS
1	A	418	SER
1	A	424	ILE
1	A	567	LYS
1	A	1223	ASP
2	B	250	PHE
2	B	477	ALA
2	B	648	HIS
2	B	708	GLU
2	B	709	ASP
2	B	712	PRO
2	B	731	VAL
2	B	734	HIS
2	B	751	VAL
2	B	850	LEU
2	B	1046	PRO
2	B	1156	ASP
2	B	1181	GLU
3	C	173	ALA
8	J	2	ILE

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
8	J	6	ARG
10	L	50	ASP
10	L	56	LEU
1	A	118	HIS
1	A	286	HIS
1	A	312	PRO
1	A	324	SER
1	A	332	LYS
1	A	775	ILE
1	A	1123	GLY
1	A	1234	GLU
1	A	1437	GLY
2	B	367	LEU
2	B	465	ASN
2	B	469	GLN
2	B	526	GLU
2	B	629	ASP
2	B	887	HIS
2	B	976	ILE
2	B	1155	SER
3	C	142	VAL
4	E	126	SER
6	H	85	GLY
6	H	131	ASN
10	L	64	LEU
1	A	56	PRO
1	A	68	GLN
1	A	308	ILE
1	A	423	ASP
1	A	1221	LYS
2	B	137	TYR
2	B	139	ALA
2	B	471	LYS
2	B	646	LEU
2	B	879	ARG
2	B	1221	SER
5	F	154	ASP
6	H	90	ALA
6	H	109	LYS
7	I	77	LYS
9	K	64	GLU
10	L	46	VAL

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
10	L	59	ALA
1	A	214	ILE
1	A	257	ARG
1	A	610	GLY
1	A	672	ASP
1	A	852	TYR
1	A	1083	THR
2	B	248	SER
2	B	478	GLY
2	B	1096	ARG
2	B	1157	ALA
2	B	1178	ASN
3	C	90	ASP
3	C	215	GLU
4	E	86	PRO
6	H	86	ASP
10	L	47	ARG
1	A	35	ILE
1	A	958	VAL
2	B	531	GLN
2	B	881	ASN
6	H	83	GLN
6	H	140	ALA
8	J	64	ASN
10	L	39	SER
1	A	310	GLY
1	A	404	TYR
2	B	563	MET
2	B	792	MET
2	B	1099	VAL
6	H	17	PRO
6	H	108	SER
1	A	78	PRO
1	A	331	GLY
1	A	400	PRO
1	A	1384	VAL
6	H	107	VAL
2	B	647	GLY

### 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	1225/1520 (81%)	1083 (88%)	142 (12%)	8	34
2	B	967/1061 (91%)	866 (90%)	101 (10%)	10	39
3	C	234/274 (85%)	209 (89%)	25 (11%)	10	38
4	E	196/197 (100%)	186 (95%)	10 (5%)	33	77
5	F	75/137 (55%)	74 (99%)	1 (1%)	80	95
6	H	117/128 (91%)	103 (88%)	14 (12%)	7	32
7	I	113/116 (97%)	103 (91%)	10 (9%)	14	50
8	J	60/65 (92%)	52 (87%)	8 (13%)	6	27
9	K	99/102 (97%)	91 (92%)	8 (8%)	17	56
10	L	40/57 (70%)	32 (80%)	8 (20%)	2	9
All	All	3126/3657 (86%)	2799 (90%)	327 (10%)	10	39

All (327) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	18	GLN
1	A	30	ILE
1	A	40	THR
1	A	41	MET
1	A	43	GLU
1	A	58	LEU
1	A	61	ILE
1	A	65	LEU
1	A	69	THR
1	A	70	CYS
1	A	88	LYS
1	A	90	VAL
1	A	93	VAL
1	A	126	LEU
1	A	130	ASP
1	A	133	LYS
1	A	143	LYS
1	A	144	THR

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
1	A	171	GLN
1	A	204	THR
1	A	222	LEU
1	A	225	ASN
1	A	252	PHE
1	A	254	GLU
1	A	270	LEU
1	A	282	ASN
1	A	298	PHE
1	A	303	TYR
1	A	308	ILE
1	A	320	ARG
1	A	323	LYS
1	A	325	ILE
1	A	330	LYS
1	A	335	ARG
1	A	344	ARG
1	A	351	THR
1	A	354	SER
1	A	368	LYS
1	A	389	THR
1	A	407	ARG
1	A	408	ASP
1	A	434	ARG
1	A	443	LEU
1	A	445	ASN
1	A	450	LEU
1	A	451	HIS
1	A	455	MET
1	A	470	LEU
1	A	475	THR
1	A	476	SER
1	A	496	GLU
1	A	498	ARG
1	A	527	THR
1	A	532	ARG
1	A	535	THR
1	A	541	ILE
1	A	544	ASP
1	A	571	LEU
1	A	590	ARG
1	A	595	THR

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	598	LEU
1	A	603	ASN
1	A	612	ILE
1	A	618	GLU
1	A	622	VAL
1	A	625	SER
1	A	629	LEU
1	A	634	THR
1	A	637	LYS
1	A	640	GLN
1	A	666	ILE
1	A	672	ASP
1	A	688	LYS
1	A	695	LYS
1	A	702	LEU
1	A	705	LYS
1	A	710	LEU
1	A	716	ASP
1	A	722	LEU
1	A	731	ARG
1	A	740	LEU
1	A	756	ILE
1	A	771	GLU
1	A	788	SER
1	A	801	GLU
1	A	806	ARG
1	A	821	ARG
1	A	826	ASP
1	A	847	ASP
1	A	855	THR
1	A	885	THR
1	A	908	LEU
1	A	920	LEU
1	A	925	LEU
1	A	932	GLU
1	A	938	LYS
1	A	953	ASN
1	A	969	GLN
1	A	988	LEU
1	A	1005	GLU
1	A	1015	VAL
1	A	1022	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	1025	ARG
1	A	1029	ARG
1	A	1064	VAL
1	A	1067	LEU
1	A	1081	LEU
1	A	1084	PHE
1	A	1110	ASN
1	A	1113	THR
1	A	1161	THR
1	A	1168	GLU
1	A	1176	LEU
1	A	1187	GLN
1	A	1195	LEU
1	A	1230	GLU
1	A	1233	ASP
1	A	1237	ILE
1	A	1257	ASP
1	A	1261	LYS
1	A	1264	GLU
1	A	1267	MET
1	A	1269	GLU
1	A	1277	GLU
1	A	1280	GLU
1	A	1288	ASP
1	A	1297	GLU
1	A	1299	VAL
1	A	1303	GLU
1	A	1309	ASP
1	A	1333	ILE
1	A	1334	ASP
1	A	1351	GLU
1	A	1354	ASN
1	A	1364	ASN
1	A	1366	ARG
1	A	1382	THR
1	A	1391	ARG
1	A	1393	ASN
1	A	1398	MET
1	A	1403	GLU
1	A	1426	GLU
2	B	22	SER
2	B	25	ILE

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
2	B	40	GLU
2	B	45	SER
2	B	46	GLN
2	B	63	ILE
2	B	65	GLU
2	B	66	ASP
2	B	68	THR
2	B	101	MET
2	B	134	LYS
2	B	135	ARG
2	B	167	ILE
2	B	175	ARG
2	B	208	SER
2	B	217	ARG
2	B	234	ILE
2	B	240	ILE
2	B	276	ILE
2	B	277	LYS
2	B	282	ILE
2	B	311	LEU
2	B	355	ILE
2	B	365	THR
2	B	393	LYS
2	B	394	ASP
2	B	396	ASP
2	B	404	LYS
2	B	413	LEU
2	B	424	LEU
2	B	437	GLU
2	B	451	LYS
2	B	461	LEU
2	B	468	GLU
2	B	471	LYS
2	B	479	VAL
2	B	480	SER
2	B	485	ARG
2	B	487	THR
2	B	493	SER
2	B	531	GLN
2	B	547	VAL
2	B	555	ILE
2	B	574	SER

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
2	B	591	ARG
2	B	604	ARG
2	B	616	ILE
2	B	644	GLU
2	B	645	SER
2	B	653	VAL
2	B	655	LYS
2	B	658	ILE
2	B	696	GLU
2	B	762	ASN
2	B	780	VAL
2	B	786	ASN
2	B	787	VAL
2	B	791	THR
2	B	815	ARG
2	B	838	SER
2	B	844	SER
2	B	850	LEU
2	B	853	SER
2	B	862	GLN
2	B	864	LYS
2	B	866	TYR
2	B	868	MET
2	B	880	THR
2	B	885	MET
2	B	886	LYS
2	B	935	ARG
2	B	948	ILE
2	B	951	GLN
2	B	963	PHE
2	B	983	ARG
2	B	987	LYS
2	B	996	ARG
2	B	997	GLU
2	B	998	ASP
2	B	1006	ILE
2	B	1007	VAL
2	B	1010	LEU
2	B	1019	SER
2	B	1028	GLU
2	B	1065	GLN
2	B	1066	SER

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
2	B	1077	THR
2	B	1094	ARG
2	B	1096	ARG
2	B	1099	VAL
2	B	1103	ILE
2	B	1128	LEU
2	B	1129	ARG
2	B	1145	SER
2	B	1147	LEU
2	B	1160	VAL
2	B	1175	LEU
2	B	1189	ILE
2	B	1194	ILE
2	B	1202	LEU
2	B	1222	ARG
3	C	3	GLU
3	C	4	GLU
3	C	18	VAL
3	C	21	ILE
3	C	34	ARG
3	C	37	MET
3	C	43	THR
3	C	66	ARG
3	C	75	MET
3	C	77	ILE
3	C	79	GLN
3	C	109	SER
3	C	124	LEU
3	C	127	ARG
3	C	129	ILE
3	C	137	LYS
3	C	144	ILE
3	C	156	THR
3	C	215	GLU
3	C	226	ASP
3	C	240	VAL
3	C	244	VAL
3	C	249	ASP
3	C	254	LYS
3	C	268	ASP
4	E	7	ARG
4	E	65	THR

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
4	E	78	LEU
4	E	92	THR
4	E	122	LYS
4	E	127	ILE
4	E	156	LEU
4	E	169	ARG
4	E	191	LYS
4	E	204	THR
5	F	115	THR
6	H	16	ASP
6	H	22	LYS
6	H	34	ASP
6	H	35	GLN
6	H	52	GLN
6	H	63	LEU
6	H	86	ASP
6	H	92	ASP
6	H	94	ASP
6	H	107	VAL
6	H	109	LYS
6	H	110	ASP
6	H	130	ARG
6	H	145	ARG
7	I	35	VAL
7	I	52	ILE
7	I	61	ASP
7	I	66	PRO
7	I	70	ARG
7	I	84	VAL
7	I	87	GLN
7	I	106	CYS
7	I	111	THR
7	I	119	THR
8	J	3	VAL
8	J	7	CYS
8	J	31	ASP
8	J	48	ARG
8	J	56	LEU
8	J	57	ILE
8	J	59	LYS
8	J	62	ARG
9	K	14	GLU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
9	K	20	LYS
9	K	22	ASP
9	K	31	VAL
9	K	49	GLU
9	K	51	LEU
9	K	78	THR
9	K	114	LEU
10	L	27	LEU
10	L	43	THR
10	L	55	ILE
10	L	58	LYS
10	L	61	THR
10	L	65	VAL
10	L	66	GLN
10	L	68	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
11	R	5/5 (100%)	4 (80%)	1 (20%)

All (4) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
11	R	7	G
11	R	8	G
11	R	9	5BU
11	R	10	G

All (1) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
11	R	6	A

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

1 non-standard protein/DNA/RNA residue is modelled in this entry.



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	$\# Z  > 2$
11	5BU	R	9	11,12	20,22,23	1.85	3 (15%)	25,32,35	2.97	8 (32%)

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
11	5BU	R	9	11,12	-	0/6/25/26	0/2/2/2

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
11	R	9	5BU	C4-C5	5.71	1.49	1.39
11	R	9	5BU	C2-N1	3.85	1.42	1.38
11	R	9	5BU	C6-N1	2.28	1.39	1.34

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
11	R	9	5BU	C6-N1-C2	-8.57	119.97	122.41
11	R	9	5BU	C5-C6-N1	7.14	124.38	119.67
11	R	9	5BU	P-O5'-C5'	6.82	149.83	123.19
11	R	9	5BU	C2-N1-C1'	4.18	120.83	118.21
11	R	9	5BU	O3'-C3'-C4'	-2.37	104.10	111.08
11	R	9	5BU	O4'-C4'-C3'	-2.25	100.62	105.17
11	R	9	5BU	C3'-C2'-C1'	-2.18	97.50	100.91
11	R	9	5BU	C2'-C3'-C4'	-2.00	98.66	102.65

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

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.

## 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	1405/1733 (81%)	-0.02	29 (2%) 60 15	42, 85, 175, 219	0
2	B	1114/1224 (91%)	-0.11	8 (0%) 84 38	37, 78, 147, 189	0
3	C	266/318 (83%)	-0.24	0 100 100	49, 75, 114, 155	0
4	E	214/215 (99%)	-0.09	0 100 100	61, 115, 162, 175	0
5	F	85/155 (54%)	-0.25	0 100 100	66, 96, 133, 152	0
6	H	133/146 (91%)	0.10	3 (2%) 57 13	80, 119, 154, 166	0
7	I	119/122 (97%)	-0.23	0 100 100	55, 93, 129, 142	0
8	J	65/70 (92%)	-0.23	0 100 100	49, 67, 102, 117	0
9	K	114/120 (95%)	-0.23	0 100 100	51, 77, 104, 120	0
10	L	46/70 (65%)	-0.03	1 (2%) 59 14	61, 116, 143, 158	0
11	R	5/5 (100%)	-0.61	0 100 100	78, 84, 109, 113	0
12	T	13/29 (44%)	0.36	1 (7%) 13 3	84, 115, 174, 181	0
All	All	3579/4207 (85%)	-0.09	42 (1%) 75 26	37, 84, 162, 219	0

All (42) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
12	T	28	DT	5.2
1	A	1087	ALA	5.2
1	A	1176	LEU	5.1
1	A	1085	HIS	3.7
2	B	883	LEU	3.6
1	A	141	LEU	3.4
1	A	317	LYS	3.3
1	A	45	GLN	3.3
6	H	85	GLY	3.2
1	A	49	LYS	3.0
1	A	186	LYS	2.9

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	69	THR	2.7
2	B	865	LYS	2.7
2	B	474	SER	2.6
1	A	1086	PHE	2.6
1	A	1088	GLY	2.5
6	H	132	LEU	2.5
1	A	146	MET	2.5
1	A	103	CYS	2.5
6	H	86	ASP	2.5
1	A	176	LYS	2.4
2	B	468	GLU	2.4
10	L	27	LEU	2.4
2	B	1224	PHE	2.4
1	A	257	ARG	2.3
1	A	1083	THR	2.3
1	A	173	THR	2.3
1	A	87	ALA	2.3
1	A	174	ILE	2.3
1	A	118	HIS	2.3
1	A	255	SER	2.3
1	A	183	GLY	2.2
1	A	182	VAL	2.2
2	B	470	LYS	2.2
1	A	144	THR	2.2
1	A	1091	SER	2.1
1	A	1123	GLY	2.1
1	A	116	ASP	2.1
2	B	472	ALA	2.1
1	A	254	GLU	2.0
2	B	866	TYR	2.0
1	A	143	LYS	2.0

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

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( $\text{\AA}^2$ )	Q<0.9
11	5BU	R	9	21/22	0.15	-0.68	77,80,90,102	1

### 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( $\text{\AA}^2$ )	Q<0.9
14	MG	A	2001	1/1	0.15	0.02	21,21,21,21	0
13	ZN	I	203	1/1	0.11	-0.64	87,87,87,87	0
13	ZN	I	204	1/1	0.11	-0.86	66,66,66,66	0
13	ZN	C	319	1/1	0.12	-0.87	67,67,67,67	0
13	ZN	B	1307	1/1	0.10	-0.89	149,149,149,149	0
13	ZN	L	105	1/1	0.08	-1.14	102,102,102,102	0
13	ZN	J	101	1/1	0.19	-1.28	58,58,58,58	0
13	ZN	A	1734	1/1	0.13	-1.32	300,300,300,300	0
13	ZN	A	1735	1/1	0.09	-1.74	135,135,135,135	0

### 6.5 Other polymers

There are no such residues in this entry.