



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 16, 2023 – 01:23 AM EDT

PDB ID : 1SFO  
Title : RNA POLYMERASE II STRAND SEPARATED ELONGATION COMPLEX  
Authors : Westover, K.D.; Bushnell, D.A.; Kornberg, R.D.  
Deposited on : 2004-02-20  
Resolution : 3.61 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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 : 2.35  
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) : 2.35

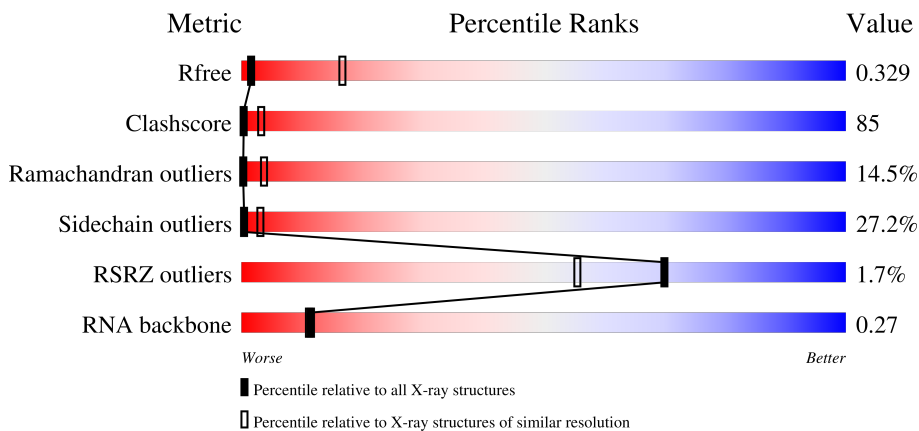
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.61 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	1290 (3.74-3.50)
Clashscore	141614	1387 (3.74-3.50)
Ramachandran outliers	138981	1339 (3.74-3.50)
Sidechain outliers	138945	1339 (3.74-3.50)
RSRZ outliers	127900	1191 (3.74-3.50)
RNA backbone	3102	1018 (4.22-3.00)

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 of 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	R	10	
2	T	14	
3	A	1733	
4	B	1224	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
5	C	318	
6	E	215	
7	F	155	
8	H	146	
9	I	122	
10	J	70	
11	K	120	
12	L	70	

## 2 Entry composition [i](#)

There are 14 unique types of molecules in this entry. The entry contains 28647 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 RNA chain called RNA STRAND.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	R	10	217	98	45	65	9	0	0	0

- Molecule 2 is a DNA chain called DNA STRAND.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	T	14	279	135	48	83	13	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	A	1395	10969	6917	1923	2068	61	0	0	0

- Molecule 4 is a protein called DNA-directed RNA polymerase II 140 kDa polypeptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	B	1106	8793	5568	1538	1632	55	0	0	0

- Molecule 5 is a protein called DNA-directed RNA polymerase II 45 kDa polypeptide.

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

- Molecule 6 is a protein called DNA-directed RNA polymerases I, II, and III 27 kDa polypeptide.

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

- Molecule 7 is a protein called DNA-directed RNA polymerases I, II, and III 23 kDa polypeptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
7	F	84	679	434	115	127	3	0	0	0

- Molecule 8 is a protein called DNA-directed RNA polymerases I, II, and III 14.5 kDa polypeptide.

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

- Molecule 9 is a protein called DNA-directed RNA polymerase II 14.2 kDa polypeptide.

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 8.3 kDa polypeptide.

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 13.6 kDa polypeptide.

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

- Molecule 12 is a protein called DNA-directed RNA polymerases I, II, and III 7.7 kDa polypeptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
12	L	46	364	224	72	64	4	0	0	0

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

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

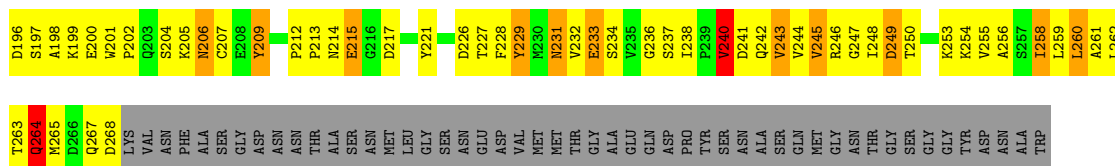




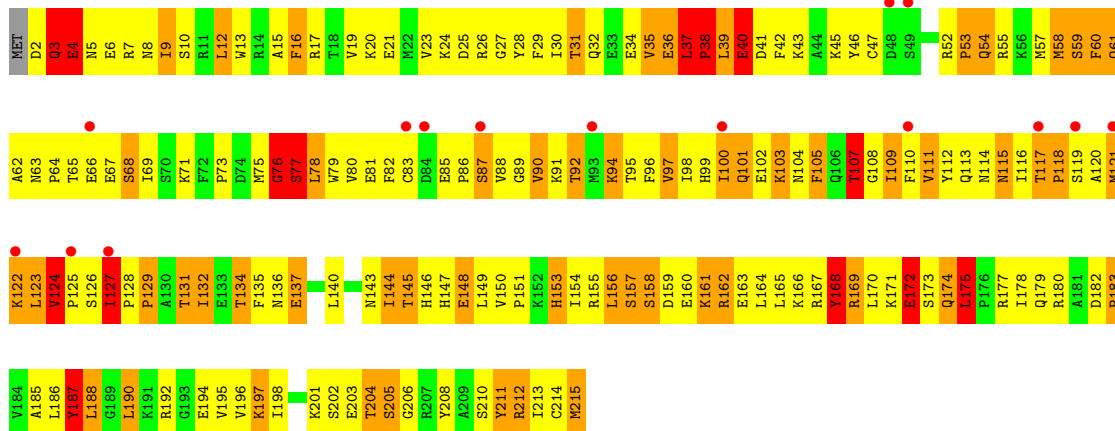
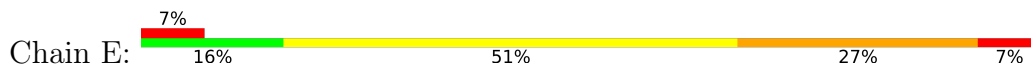




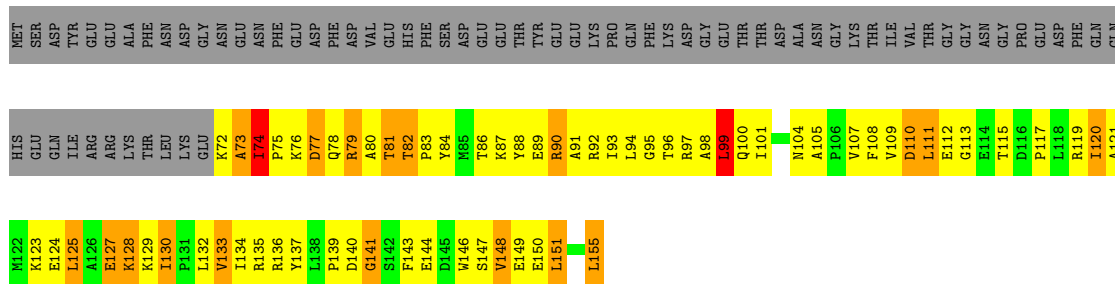




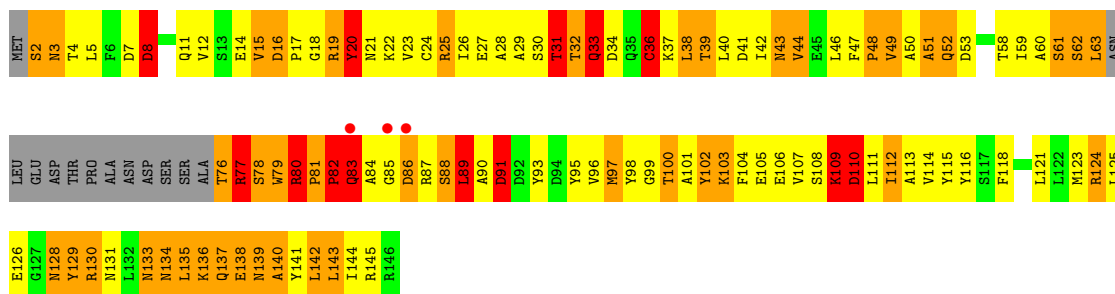
- Molecule 6: DNA-directed RNA polymerases I, II, and III 27 kDa polypeptide



- Molecule 7: DNA-directed RNA polymerases I, II, and III 23 kDa polypeptide

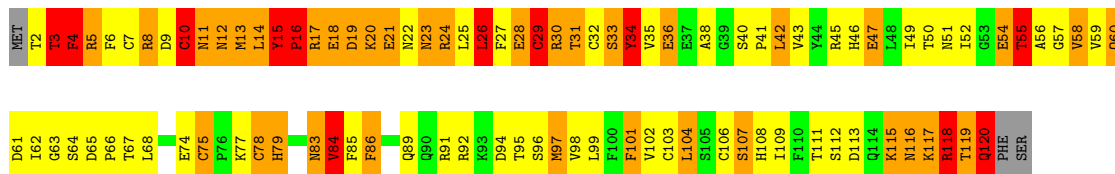


- Molecule 8: DNA-directed RNA polymerases I, II, and III 14.5 kDa polypeptide

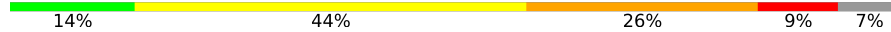


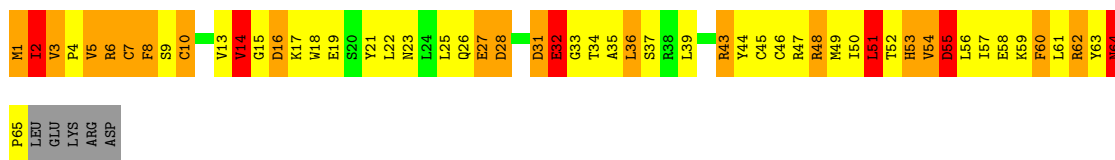
- Molecule 9: DNA-directed RNA polymerase II 14.2 kDa polypeptide

Chain I:  18% 40% 30% 10%

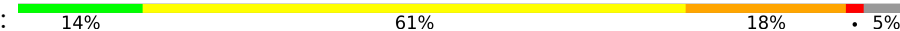


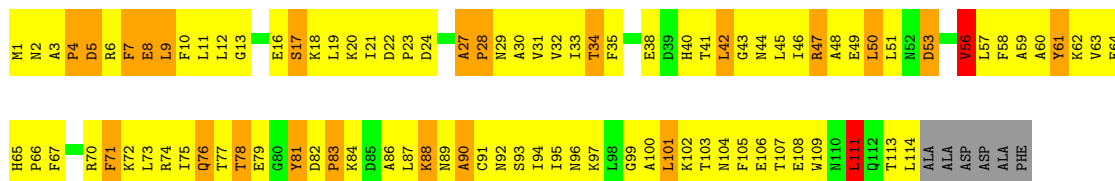
- Molecule 10: DNA-directed RNA polymerases I, II, and III 8.3 kDa polypeptide

Chain J:  14% 44% 26% 9% 7%




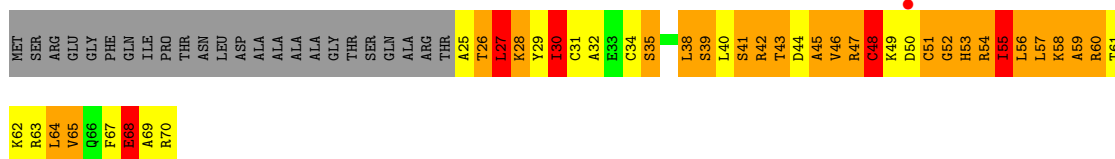
- Molecule 11: DNA-directed RNA polymerase II 13.6 kDa polypeptide

Chain K:  14% 61% 18% 5%



- Molecule 12: DNA-directed RNA polymerases I, II, and III 7.7 kDa polypeptide

Chain L:  % 6% 21% 31% 7% 34%



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	167.08Å 221.26Å 193.69Å 90.00° 100.10° 90.00°	Depositor
Resolution (Å)	39.86 – 3.61 39.86 – 3.60	Depositor EDS
% Data completeness (in resolution range)	92.7 (39.86-3.61) 92.7 (39.86-3.60)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.13 (at 3.57Å)	Xtrriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.315 , 0.343 0.309 , 0.329	Depositor DCC
$R_{free}$ test set	8031 reflections (10.04%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	47.6	Xtrriage
Anisotropy	0.357	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.23 , -1.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.40$ , $\langle L^2 \rangle = 0.22$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.76	EDS
Total number of atoms	28647	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	53.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.72% 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	R	0.67	0/244	0.83	0/380
2	T	0.75	0/311	1.39	3/477 (0.6%)
3	A	0.88	15/11163 (0.1%)	1.45	167/15091 (1.1%)
4	B	0.84	7/8964 (0.1%)	1.38	114/12086 (0.9%)
5	C	0.78	0/2133	1.24	13/2891 (0.4%)
6	E	0.90	2/1788 (0.1%)	1.40	14/2406 (0.6%)
7	F	0.83	0/691	1.28	7/933 (0.8%)
8	H	0.85	1/1086 (0.1%)	1.59	20/1470 (1.4%)
9	I	1.03	2/989 (0.2%)	1.64	23/1331 (1.7%)
10	J	0.78	0/541	1.44	7/727 (1.0%)
11	K	0.74	0/937	1.20	5/1265 (0.4%)
12	L	0.99	1/366 (0.3%)	1.78	12/485 (2.5%)
All	All	0.86	28/29213 (0.1%)	1.41	385/39542 (1.0%)

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
2	T	0	5
3	A	1	6
4	B	0	7
5	C	0	2
6	E	0	1
9	I	0	1
All	All	1	22

The worst 5 of 28 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	319	GLY	C-O	10.97	1.41	1.23
3	A	255	SER	CA-CB	8.59	1.65	1.52
3	A	320	ARG	CA-CB	8.38	1.72	1.53
3	A	320	ARG	CG-CD	7.83	1.71	1.51
4	B	595	ARG	CG-CD	7.49	1.70	1.51

The worst 5 of 385 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	321	PRO	N-CA-C	-18.56	63.84	112.10
3	A	322	VAL	N-CA-C	14.54	150.26	111.00
10	J	10	CYS	CA-CB-SG	12.01	135.61	114.00
3	A	315	LEU	CA-CB-CG	11.82	142.50	115.30
4	B	478	GLY	N-CA-C	-11.76	83.71	113.10

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
3	A	322	VAL	CA

5 of 22 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	T	10	DT	Sidechain
2	T	11	DC	Sidechain
2	T	13	DA	Sidechain
2	T	6	DC	Sidechain
2	T	8	DT	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	R	217	0	110	18	0
2	T	279	0	160	43	0
3	A	10969	0	11070	2106	0
4	B	8793	0	8823	1592	0
5	C	2095	0	2051	337	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	E	1752	0	1776	303	0
7	F	679	0	701	127	0
8	H	1068	0	1040	193	0
9	I	971	0	929	164	0
10	J	532	0	542	125	0
11	K	919	0	929	175	0
12	L	364	0	387	65	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	28647	0	28518	4858	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 85.

The worst 5 of 4858 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)
3:A:853:ASP:OD1	3:A:855:THR:HB	1.32	1.25
3:A:90:VAL:HG12	3:A:297:GLN:NE2	1.49	1.24
4:B:635:ARG:HB2	4:B:636:PRO:CD	1.65	1.21
3:A:321:PRO:O	3:A:322:VAL:HG22	1.41	1.18
3:A:351:THR:HG23	4:B:1103:ILE:HD12	1.23	1.18

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	
3	A	1383/1733 (80%)	851 (62%)	315 (23%)	217 (16%)	0	3
4	B	1088/1224 (89%)	730 (67%)	214 (20%)	144 (13%)	0	4
5	C	264/318 (83%)	187 (71%)	49 (19%)	28 (11%)	0	7
6	E	212/215 (99%)	142 (67%)	41 (19%)	29 (14%)	0	4
7	F	82/155 (53%)	49 (60%)	26 (32%)	7 (8%)	1	10
8	H	129/146 (88%)	87 (67%)	16 (12%)	26 (20%)	0	2
9	I	117/122 (96%)	74 (63%)	24 (20%)	19 (16%)	0	3
10	J	63/70 (90%)	42 (67%)	11 (18%)	10 (16%)	0	3
11	K	112/120 (93%)	81 (72%)	20 (18%)	11 (10%)	0	8
12	L	44/70 (63%)	20 (46%)	10 (23%)	14 (32%)	0	0
All	All	3494/4173 (84%)	2263 (65%)	726 (21%)	505 (14%)	0	4

5 of 505 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	44	THR
3	A	50	ILE
3	A	55	ASP
3	A	56	PRO
3	A	57	ARG

### 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	
3	A	1218/1520 (80%)	859 (70%)	359 (30%)	0	2
4	B	960/1061 (90%)	718 (75%)	242 (25%)	0	4
5	C	234/274 (85%)	175 (75%)	59 (25%)	0	4
6	E	196/197 (100%)	141 (72%)	55 (28%)	0	3
7	F	74/137 (54%)	57 (77%)	17 (23%)	1	5
8	H	117/128 (91%)	80 (68%)	37 (32%)	0	2

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
9	I	113/116 (97%)	84 (74%)	29 (26%)	0	4
10	J	60/65 (92%)	42 (70%)	18 (30%)	0	2
11	K	99/102 (97%)	81 (82%)	18 (18%)	1	10
12	L	40/57 (70%)	27 (68%)	13 (32%)	0	1
All	All	3111/3657 (85%)	2264 (73%)	847 (27%)	0	3

5 of 847 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	B	563	MET
4	B	1119	VAL
10	J	7	CYS
4	B	640	VAL
4	B	556	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 101 such sidechains are listed below:

Mol	Chain	Res	Type
4	B	516	ASN
4	B	1015	HIS
11	K	96	ASN
4	B	572	HIS
4	B	786	ASN

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	R	9/10 (90%)	2 (22%)	0

All (2) RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	R	5	A
1	R	10	A

There are no RNA pucker outliers to report.

## 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 monosaccharides 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	R	10/10 (100%)	1.35	2 (20%) 1 0	30, 82, 200, 200	0
2	T	14/14 (100%)	1.04	4 (28%) 0 0	30, 106, 200, 200	0
3	A	1395/1733 (80%)	-0.15	26 (1%) 66 51	30, 32, 134, 200	0
4	B	1106/1224 (90%)	-0.23	11 (0%) 82 70	30, 30, 118, 198	0
5	C	266/318 (83%)	-0.35	0 100 100	30, 30, 84, 140	0
6	E	214/215 (99%)	0.28	15 (7%) 16 9	30, 67, 153, 200	0
7	F	84/155 (54%)	-0.08	0 100 100	30, 31, 93, 141	0
8	H	133/146 (91%)	-0.08	3 (2%) 60 44	30, 47, 138, 190	0
9	I	119/122 (97%)	-0.09	0 100 100	30, 32, 114, 163	0
10	J	65/70 (92%)	-0.41	0 100 100	30, 30, 96, 141	0
11	K	114/120 (95%)	-0.36	0 100 100	30, 30, 73, 107	0
12	L	46/70 (65%)	-0.24	1 (2%) 62 46	30, 45, 132, 158	0
All	All	3566/4197 (84%)	-0.16	62 (1%) 70 55	30, 31, 130, 200	0

The worst 5 of 62 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	B	866	TYR	5.8
3	A	316	GLN	5.0
2	T	1	DA	4.8
3	A	1175	SER	4.5
3	A	149	GLU	4.4

### 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 monosaccharides 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
14	MG	A	2000	1/1	0.69	0.20	22,22,22,22	0
13	ZN	A	1734	1/1	0.84	0.29	22,22,22,22	0
13	ZN	A	1735	1/1	0.87	0.10	22,22,22,22	0
13	ZN	B	1307	1/1	0.91	0.11	22,22,22,22	0
13	ZN	I	203	1/1	0.92	0.21	22,22,22,22	0
13	ZN	J	101	1/1	0.93	0.11	22,22,22,22	0
13	ZN	I	204	1/1	0.96	0.17	22,22,22,22	0
13	ZN	L	105	1/1	0.98	0.06	22,22,22,22	0
13	ZN	C	319	1/1	0.98	0.06	22,22,22,22	0

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

There are no such residues in this entry.