



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 30, 2020 – 01:52 am BST

PDB ID : 6KON  
Title : Mycobacterium tuberculosis initial transcription complex comprising sigma H  
and 5'-OH RNA of 5 nt  
Authors : Li, L.; Zhang, Y.  
Deposited on : 2019-08-12  
Resolution : 3.00 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

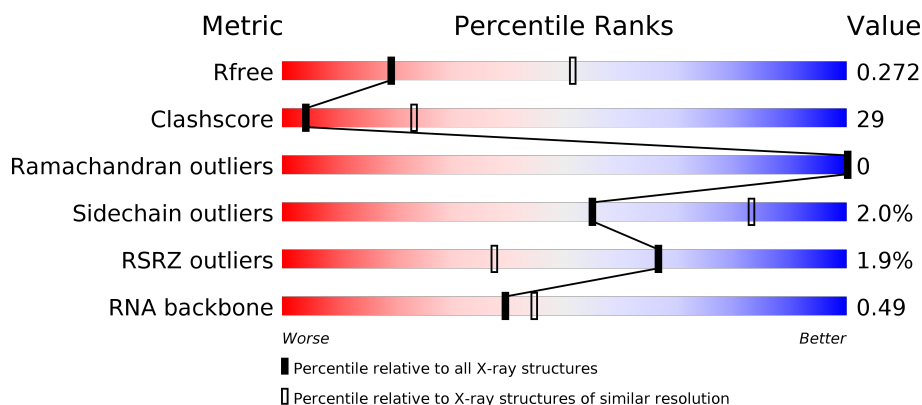
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.00 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.



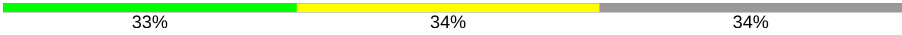


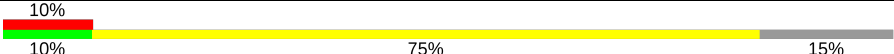

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)
RNA backbone	3102	1173 (3.30-2.70)

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	368	<div> <div></div> <div>36%23%41%</div> </div>
1	B	368	<div> <div>2%</div> <div>29%33%37%</div> </div>
2	C	1174	<div> <div>3%</div> <div>53%40%7%</div> </div>
3	D	1317	<div> <div></div> <div>53%41%5%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
4	E	110	 33% 34% 34%
5	F	218	 2% 44% 37% 17%
6	G	23	 22% 78%
7	H	20	 10% 10% 75% 15%
8	I	5	 40% 60%

## 2 Entry composition

There are 11 unique types of molecules in this entry. The entry contains 24039 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 subunit alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	216	Total	C	N	O	S	0	0	0
			1617	1020	275	320	2			
1	B	233	Total	C	N	O	S	0	0	0
			1700	1074	286	337	3			

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-20	MET	-	initiating methionine	UNP P9WGZ1
A	-19	GLY	-	expression tag	UNP P9WGZ1
A	-18	HIS	-	expression tag	UNP P9WGZ1
A	-17	HIS	-	expression tag	UNP P9WGZ1
A	-16	HIS	-	expression tag	UNP P9WGZ1
A	-15	HIS	-	expression tag	UNP P9WGZ1
A	-14	HIS	-	expression tag	UNP P9WGZ1
A	-13	HIS	-	expression tag	UNP P9WGZ1
A	-12	HIS	-	expression tag	UNP P9WGZ1
A	-11	HIS	-	expression tag	UNP P9WGZ1
A	-10	HIS	-	expression tag	UNP P9WGZ1
A	-9	HIS	-	expression tag	UNP P9WGZ1
A	-8	SER	-	expression tag	UNP P9WGZ1
A	-7	SER	-	expression tag	UNP P9WGZ1
A	-6	GLY	-	expression tag	UNP P9WGZ1
A	-5	HIS	-	expression tag	UNP P9WGZ1
A	-4	ILE	-	expression tag	UNP P9WGZ1
A	-3	GLU	-	expression tag	UNP P9WGZ1
A	-2	GLY	-	expression tag	UNP P9WGZ1
A	-1	ARG	-	expression tag	UNP P9WGZ1
A	0	HIS	-	expression tag	UNP P9WGZ1
B	-20	MET	-	initiating methionine	UNP P9WGZ1
B	-19	GLY	-	expression tag	UNP P9WGZ1
B	-18	HIS	-	expression tag	UNP P9WGZ1
B	-17	HIS	-	expression tag	UNP P9WGZ1

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	-16	HIS	-	expression tag	UNP P9WGZ1
B	-15	HIS	-	expression tag	UNP P9WGZ1
B	-14	HIS	-	expression tag	UNP P9WGZ1
B	-13	HIS	-	expression tag	UNP P9WGZ1
B	-12	HIS	-	expression tag	UNP P9WGZ1
B	-11	HIS	-	expression tag	UNP P9WGZ1
B	-10	HIS	-	expression tag	UNP P9WGZ1
B	-9	HIS	-	expression tag	UNP P9WGZ1
B	-8	SER	-	expression tag	UNP P9WGZ1
B	-7	SER	-	expression tag	UNP P9WGZ1
B	-6	GLY	-	expression tag	UNP P9WGZ1
B	-5	HIS	-	expression tag	UNP P9WGZ1
B	-4	ILE	-	expression tag	UNP P9WGZ1
B	-3	GLU	-	expression tag	UNP P9WGZ1
B	-2	GLY	-	expression tag	UNP P9WGZ1
B	-1	ARG	-	expression tag	UNP P9WGZ1
B	0	HIS	-	expression tag	UNP P9WGZ1

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	1092	Total	C	N	O	S	0	0	0
			8237	5153	1442	1604	38			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-1	MET	-	initiating methionine	UNP P9WGY9
C	0	VAL	-	expression tag	UNP P9WGY9

- Molecule 3 is a protein called DNA-directed RNA polymerase subunit beta'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	1252	Total	C	N	O	S	0	0	0
			9622	6037	1731	1814	40			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	0	MET	-	initiating methionine	UNP P9WGY7
D	1	VAL	-	expression tag	UNP P9WGY7

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

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
4	E	73	Total	C	N	O	0	0	0
			582	373	97	112			

- Molecule 5 is a protein called ECF RNA polymerase sigma factor SigH.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	F	180	Total	C	N	O	S	0	0	0
			1356	845	234	273	4			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	-1	GLY	-	expression tag	UNP P9WGH9
F	0	ALA	-	expression tag	UNP P9WGH9

- Molecule 6 is a DNA chain called DNA (5'-D(P\*GP\*TP\*TP\*GP\*TP\*GP\*GP\*GP\*AP\*GP\*CP\*TP\*GP\*TP\*CP\*AP\*CP\*GP\*GP\*AP\*TP\*GP\*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	G	23	Total	C	N	O	P	0	0	0
			442	207	76	136	23			

- Molecule 7 is a DNA chain called DNA (5'-D(P\*GP\*CP\*AP\*TP\*CP\*CP\*GP\*TP\*GP\*AP\*GP\*TP\*CP\*GP\*AP\*GP\*GP\*GP\*TP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
7	H	17	Total	C	N	O	P	0	0	0
			353	166	68	102	17			

- Molecule 8 is a RNA chain called RNA (5'-R(\*CP\*UP\*CP\*GP\*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
8	I	5	Total	C	N	O	P	0	0	0
			102	47	18	33	4			

- Molecule 9 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	D	2	Total	Zn	0	0
			2	2		

- Molecule 10 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
10	D	1	Total 1	Mg 1	0	0

- Molecule 11 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
11	C	12	Total 12	O 12	0	0
11	D	9	Total 9	O 9	0	0
11	F	1	Total 1	O 1	0	0
11	G	1	Total 1	O 1	0	0
11	H	2	Total 2	O 2	0	0

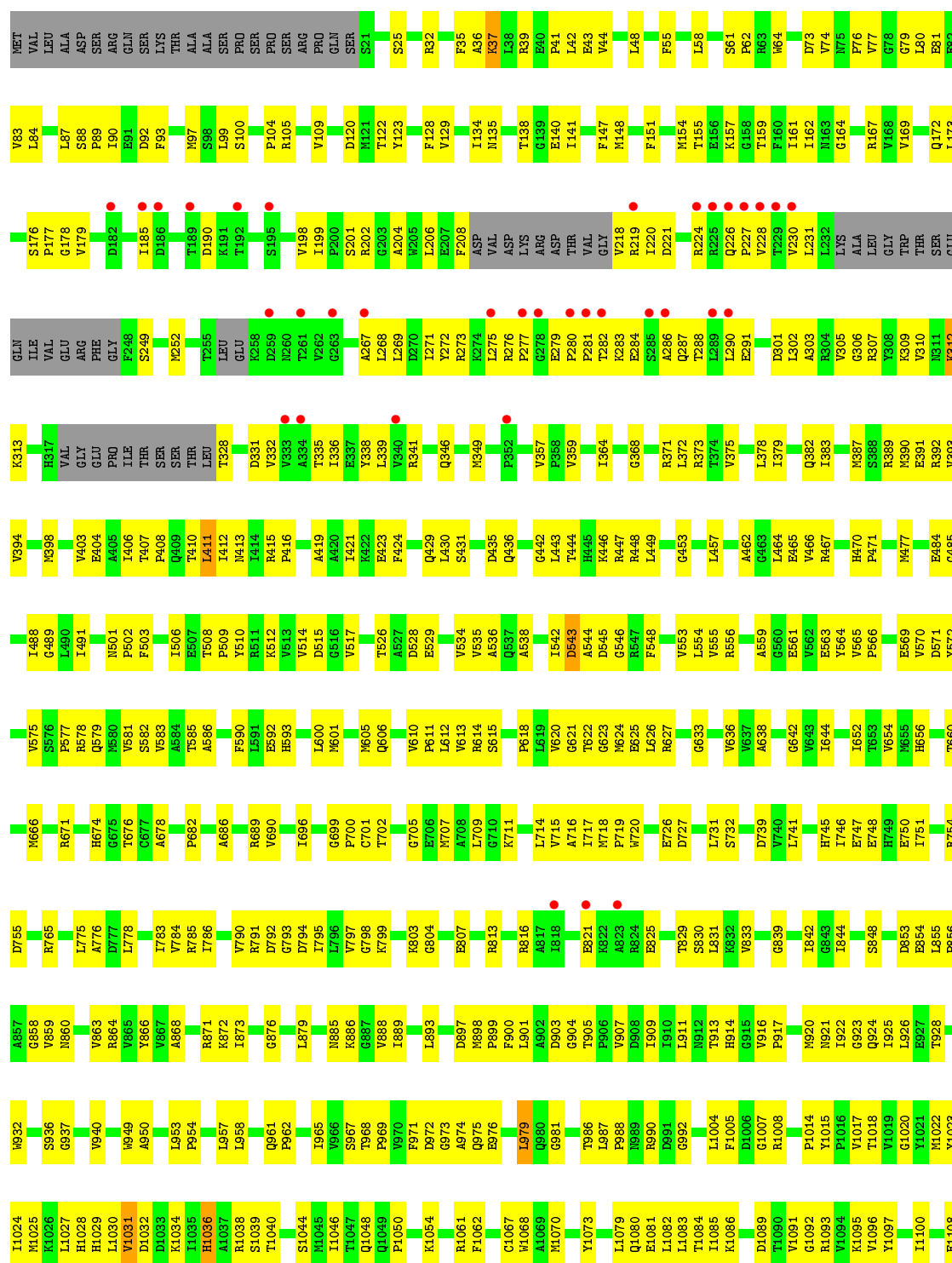




ASP  
GLU  
LEU  
ASP  
GLN  
TYR  
ALA  
GLU  
THR  
GLU  
LEU

• Molecule 2: DNA-directed RNA polymerase subunit beta

Chain C: 







- Molecule 8: RNA (5'-R(\*CP\*UP\*CP\*GP\*A)-3')

Chain I:



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	129.33Å 161.22Å 129.47Å 90.00° 117.78° 90.00°	Depositor
Resolution (Å)	48.65 – 3.00 48.65 – 3.00	Depositor EDS
% Data completeness (in resolution range)	97.2 (48.65-3.00) 97.3 (48.65-3.00)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.75 (at 3.01Å)	Xtriage
Refinement program	PHENIX 1.14 _3260	Depositor
R, $R_{free}$	0.220 , 0.271 0.221 , 0.272	Depositor DCC
$R_{free}$ test set	2008 reflections (2.20%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	84.6	Xtriage
Anisotropy	0.438	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 50.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.000 for -h-l,k,h 0.000 for l,k,-h-l 0.015 for h,-k,-h-l 0.020 for -h-l,-k,l 0.016 for l,-k,h	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	24039	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	95.0	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.

<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

### 5.1 Standard geometry

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	0/1642	0.46	0/2237
1	B	0.24	0/1725	0.46	0/2355
2	C	0.26	0/8380	0.45	0/11387
3	D	0.26	0/9782	0.43	0/13250
4	E	0.23	0/594	0.40	0/809
5	F	0.24	0/1382	0.40	0/1887
6	G	0.59	0/494	0.93	0/761
7	H	0.61	0/396	0.88	0/610
8	I	0.28	0/113	0.92	0/174
All	All	0.28	0/24508	0.47	0/33470

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 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	A	1617	0	1629	94	0
1	B	1700	0	1692	141	0
2	C	8237	0	8007	500	0
3	D	9622	0	9556	532	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	E	582	0	579	56	0
5	F	1356	0	1231	110	0
6	G	442	0	241	41	0
7	H	353	0	191	26	0
8	I	102	0	55	3	0
9	D	2	0	0	0	0
10	D	1	0	0	0	0
11	C	12	0	0	3	0
11	D	9	0	0	0	0
11	F	1	0	0	0	0
11	G	1	0	0	0	0
11	H	2	0	0	0	0
All	All	24039	0	23181	1390	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 29.

The worst 5 of 1390 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)
2:C:201:SER:HB2	2:C:303:ALA:HB2	1.27	1.10
6:G:19:DA:H2''	6:G:20:DT:H5''	1.26	1.09
3:D:1164:ARG:HB2	3:D:1208:MET:HE2	1.31	1.08
7:H:14:DC:H2'	7:H:15:DG:H5''	1.34	1.08
3:D:930:VAL:HG12	3:D:936:VAL:HA	1.35	1.06

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	212/368 (58%)	202 (95%)	10 (5%)	0	100	100
1	B	229/368 (62%)	219 (96%)	10 (4%)	0	100	100
2	C	1080/1174 (92%)	1006 (93%)	74 (7%)	0	100	100
3	D	1244/1317 (94%)	1183 (95%)	61 (5%)	0	100	100
4	E	69/110 (63%)	65 (94%)	4 (6%)	0	100	100
5	F	178/218 (82%)	173 (97%)	5 (3%)	0	100	100
All	All	3012/3555 (85%)	2848 (95%)	164 (5%)	0	100	100

There are no Ramachandran outliers to report.

### 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	179/315 (57%)	175 (98%)	4 (2%)	52	81
1	B	183/315 (58%)	181 (99%)	2 (1%)	73	90
2	C	868/995 (87%)	853 (98%)	15 (2%)	60	85
3	D	998/1096 (91%)	975 (98%)	23 (2%)	50	80
4	E	63/90 (70%)	63 (100%)	0	100	100
5	F	128/175 (73%)	124 (97%)	4 (3%)	40	75
All	All	2419/2986 (81%)	2371 (98%)	48 (2%)	55	83

5 of 48 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	125	LEU
3	D	578	ARG
5	F	183	ARG
3	D	224	SER
3	D	429	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:



Mol	Chain	Res	Type
3	D	465	HIS
5	F	117	ASN
3	D	959	GLN
2	C	579	GLN
3	D	606	HIS

### 5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
8	I	4/5 (80%)	0	0

There are no RNA backbone outliers to report.

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 carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 3 ligands modelled in this entry, 3 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 ⓘ

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	216/368 (58%)	-0.45	0	100	100	53, 76, 116, 128	0
1	B	233/368 (63%)	-0.07	7 (3%)	50	22	85, 118, 151, 158	0
2	C	1092/1174 (93%)	-0.14	36 (3%)	46	20	45, 76, 171, 200	0
3	D	1252/1317 (95%)	-0.29	5 (0%)	92	79	51, 87, 135, 157	0
4	E	73/110 (66%)	-0.17	0	100	100	92, 113, 132, 143	0
5	F	180/218 (82%)	0.11	5 (2%)	53	25	71, 131, 155, 174	0
6	G	23/23 (100%)	1.02	5 (21%)	0	0	119, 156, 181, 185	0
7	H	17/20 (85%)	0.34	2 (11%)	4	1	84, 102, 169, 172	0
8	I	5/5 (100%)	-0.38	0	100	100	75, 77, 84, 96	0
All	All	3091/3603 (85%)	-0.19	60 (1%)	66	37	45, 88, 150, 200	0

The worst 5 of 60 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	225	ARG	7.8
2	C	281	PRO	6.2
2	C	224	ARG	4.7
1	B	155	SER	4.7
2	C	185	ILE	4.4

### 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 [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
10	MG	D	2003	1/1	0.91	0.18	58,58,58,58	0
9	ZN	D	2002	1/1	0.96	0.15	89,89,89,89	0
9	ZN	D	2001	1/1	0.99	0.11	109,109,109,109	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.