



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 9, 2023 – 02:03 PM EDT

PDB ID : 7U22  
Title : Mycobacterium tuberculosis RNA polymerase sigma A holoenzyme open promoter complex containing UMN-7  
Authors : Molodtsov, V.; Ebright, R.H.  
Deposited on : 2022-02-22  
Resolution : 3.87 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  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtrriage (Phenix) : 1.13  
EDS : 2.35.1  
buster-report : 1.1.7 (2018)  
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

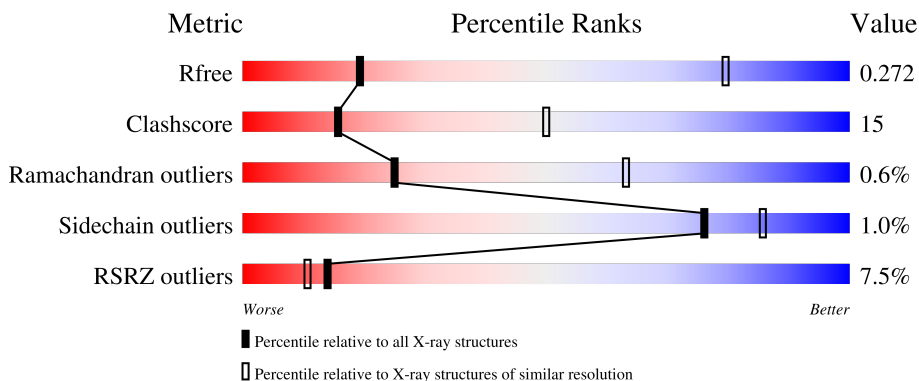
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.87 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.



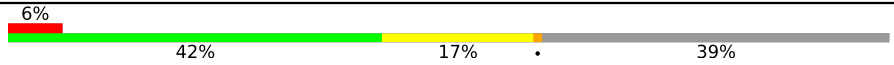
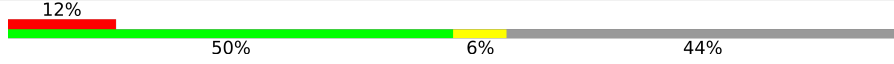
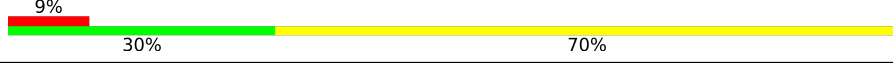
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1026 (4.12-3.64)
Clashscore	141614	1045 (4.10-3.66)
Ramachandran outliers	138981	1008 (4.10-3.66)
Sidechain outliers	138945	1001 (4.10-3.66)
RSRZ outliers	127900	1213 (4.16-3.60)

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 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	A	347	
1	B	347	
2	C	1178	
3	D	1316	
4	E	110	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
5	F	528	
6	G	16	
7	H	23	

## 2 Entry composition [i](#)

There are 10 unique types of molecules in this entry. The entry contains 25925 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
			Total	C	N	O	S			
1	A	224	Total	C	N	O	S	0	0	0
			1704	1072	295	335	2			
1	B	227	Total	C	N	O	S	0	0	0
			1715	1080	291	342	2			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	C	1126	Total	C	N	O	S	0	0	0
			8714	5454	1528	1693	39			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	D	1265	Total	C	N	O	S	0	0	0
			9887	6188	1793	1866	40			

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

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
4	E	81	Total	C	N	O	0	0	0
			637	408	106	123			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	F	320	Total	C	N	O	S	0	0	0
			2543	1583	459	492	9			

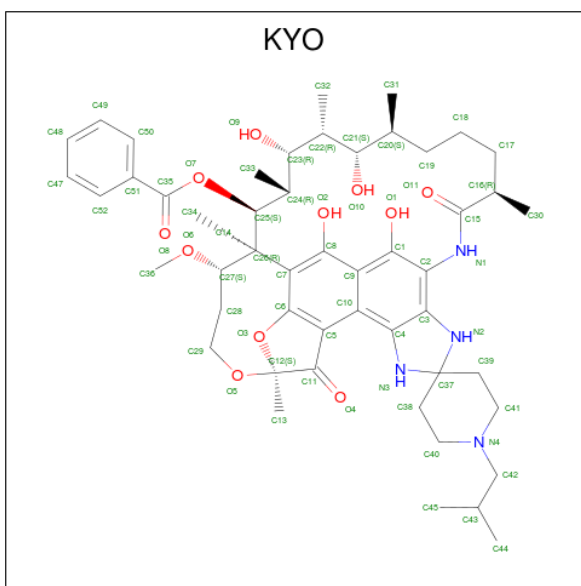
- Molecule 6 is a DNA chain called T DNA.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace	
			Total	C	N	O				P
6	G	9	180	87	33	52	8	0	0	0

- Molecule 7 is a DNA chain called NT DNA.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace	
			Total	C	N	O				P
7	H	23	476	227	91	136	22	0	0	0

- Molecule 8 is (9S,14S,15R,16S,17R,18R,19R,20S,21S,25R)-5,6,18,20-tetrahydroxy-14-methoxy-7,9,15,17,19,21,25-heptamethyl-1'-(2-methylpropyl)-10,26-dioxo-1,3,9,10-tetrahydrospiro[9,4-(epoxypentadecanoimino)furo[2',3':7,8]naphtho[1,2-d]imidazole-2,4'-piperidin]-16-yl benzoate (three-letter code: KYO) (formula: C<sub>51</sub>H<sub>72</sub>N<sub>4</sub>O<sub>11</sub>) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
8	C	1	66	51	4	11	0	0

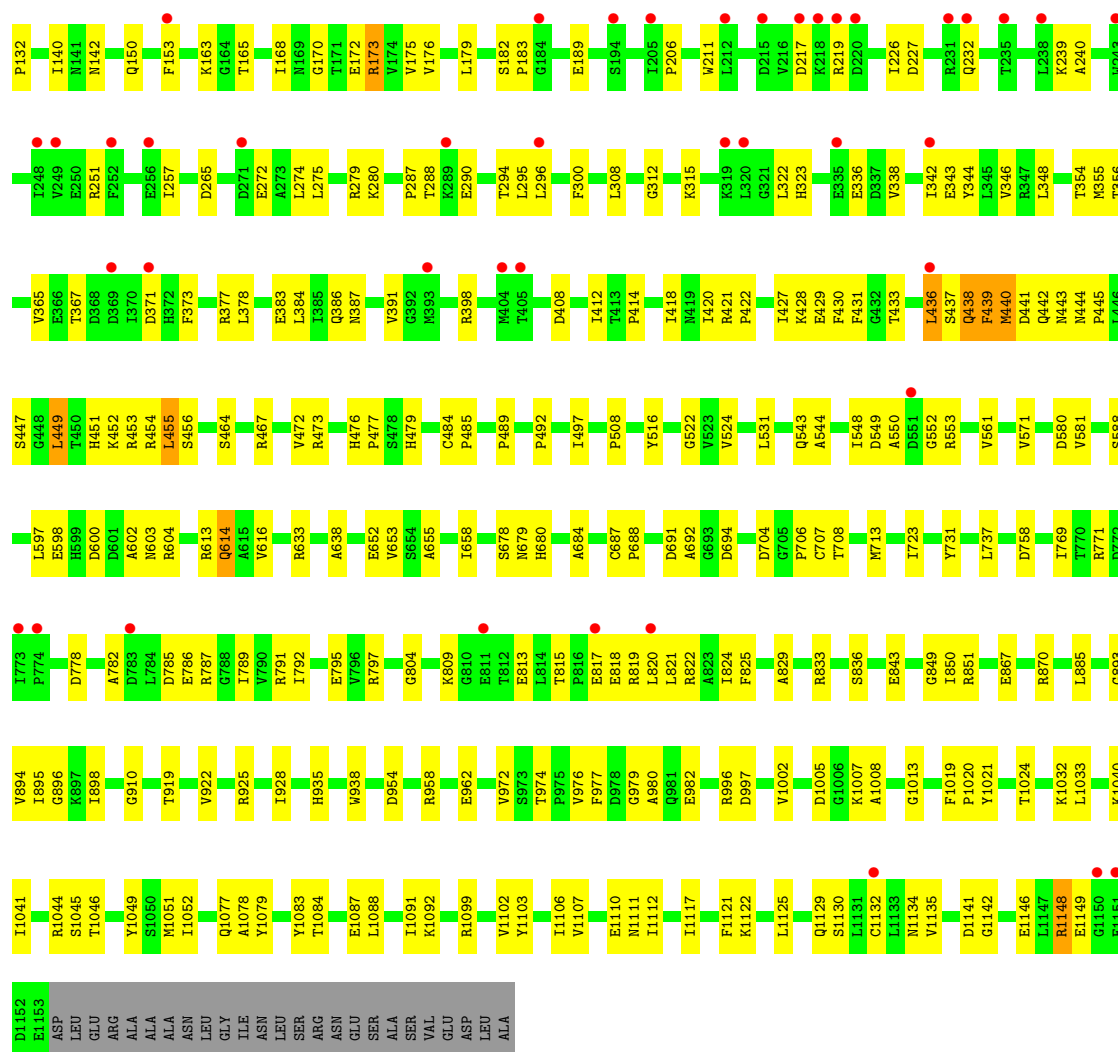
- Molecule 9 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

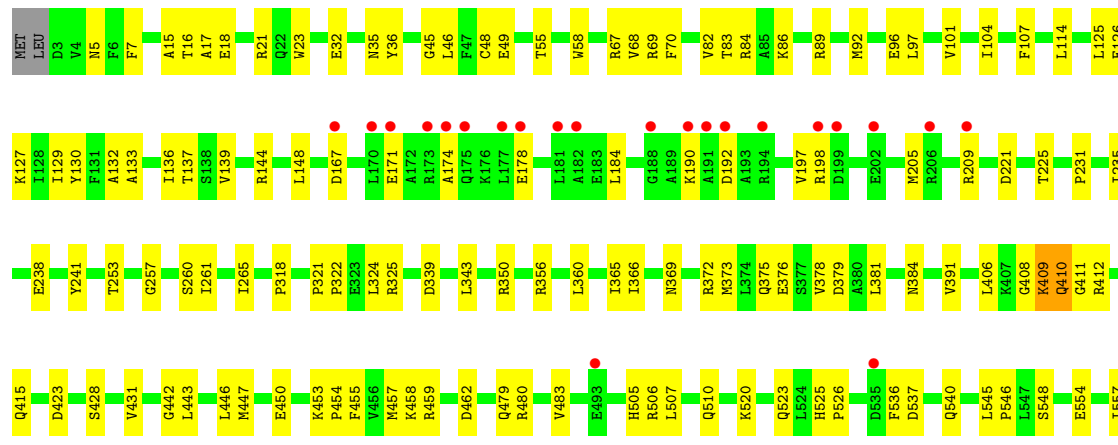
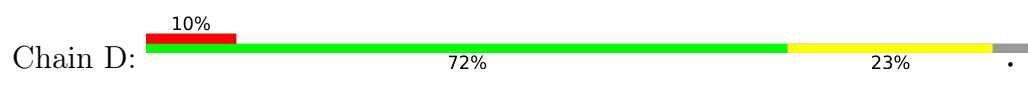
- Molecule 10 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
10	D	1	Total	Mg	1	0
			1	1		

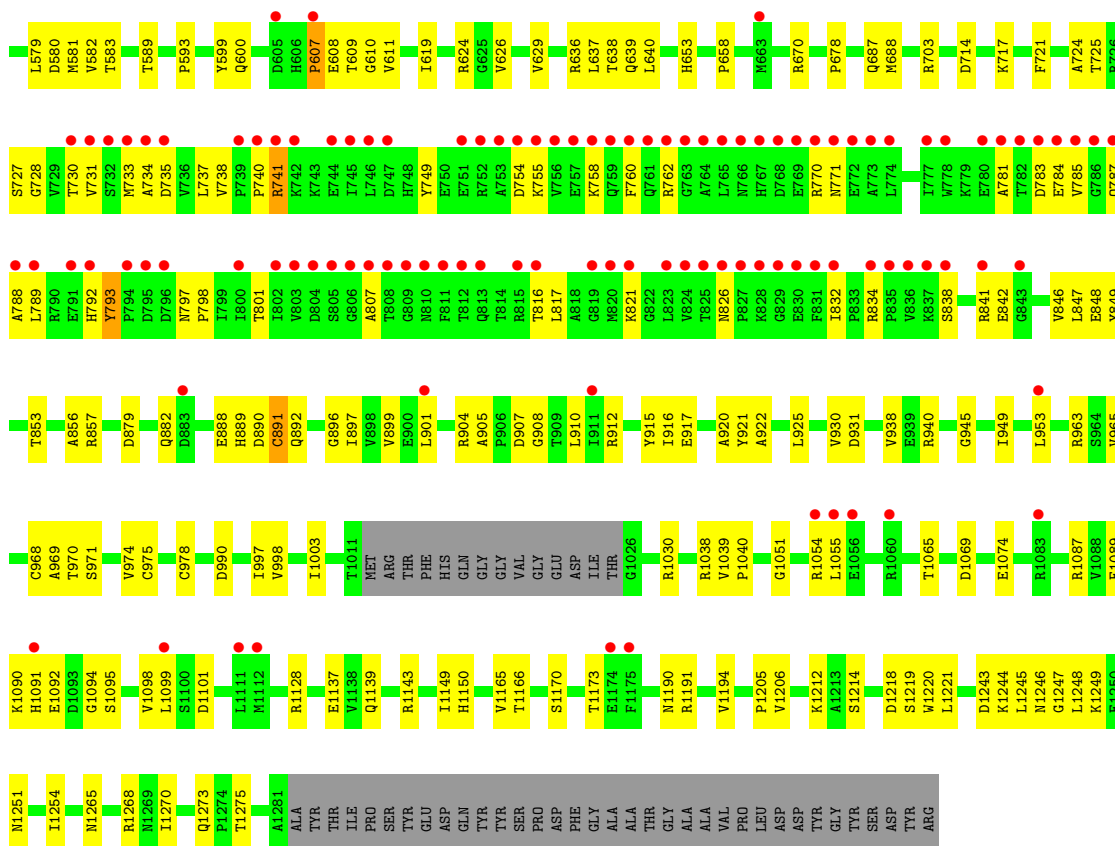




● Molecule 3: DNA-directed RNA polymerase subunit beta'

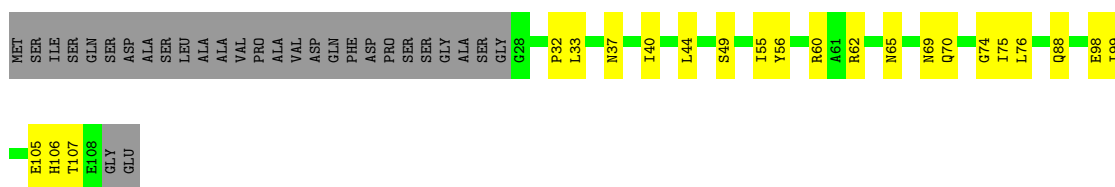






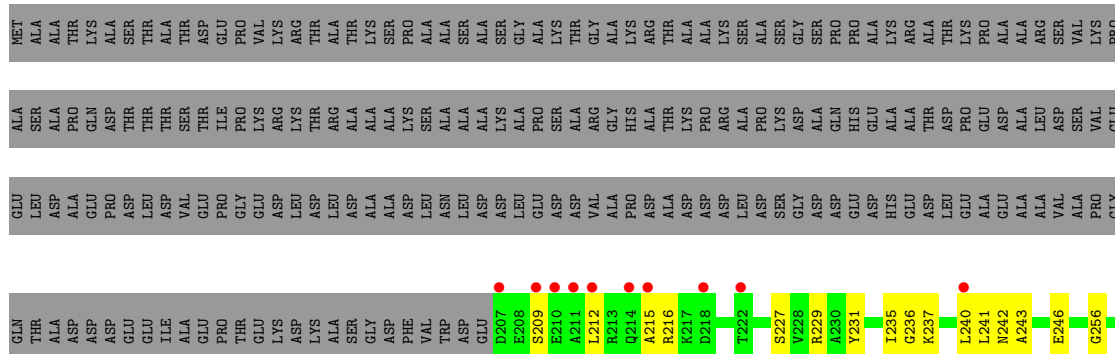
- Molecule 4: DNA-directed RNA polymerase subunit omega

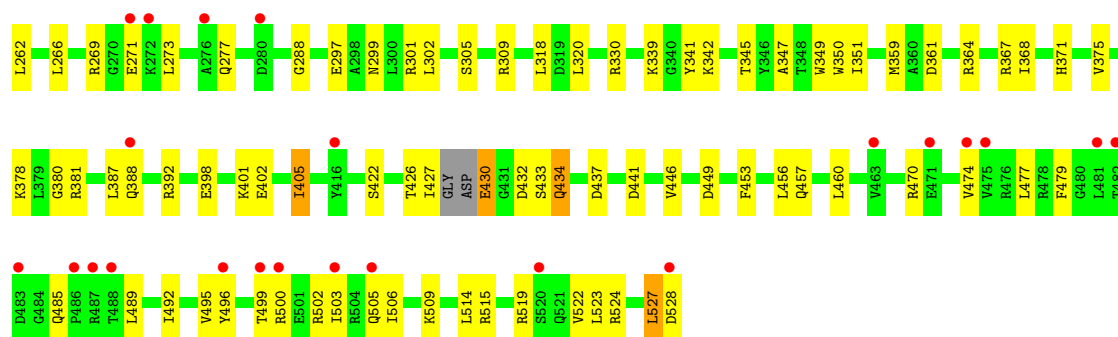
Chain E: 54% 20% 26%



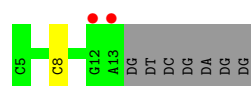
- Molecule 5: RNA polymerase sigma factor SigA

Chain F: 6% 42% 17% 39%

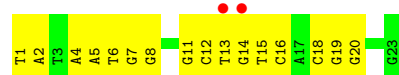




- Molecule 6: T DNA



- Molecule 7: NT DNA



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	149.48Å 161.44Å 196.60Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.30 – 3.87 48.30 – 3.87	Depositor EDS
% Data completeness (in resolution range)	99.0 (48.30-3.87) 99.0 (48.30-3.87)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.37 (at 3.88Å)	Xtrriage
Refinement program	PHENIX 1.19_4092, PHENIX 1.19_4092	Depositor
R, $R_{free}$	0.240 , 0.278 0.238 , 0.272	Depositor DCC
$R_{free}$ test set	1985 reflections (4.45%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	152.2	Xtrriage
Anisotropy	0.310	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 99.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.43$ , $\langle L^2 \rangle = 0.26$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	25925	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	187.0	wwPDB-VP

Xtrriage'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: MG, ZN, KYO

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	0.23	0/1730	0.44	0/2354
1	B	0.23	0/1741	0.44	0/2371
2	C	0.24	0/8873	0.42	1/12031 (0.0%)
3	D	0.24	0/10052	0.42	0/13591
4	E	0.25	0/650	0.42	0/886
5	F	0.25	0/2572	0.42	0/3466
6	G	0.42	0/201	0.84	0/308
7	H	0.55	0/535	0.89	0/826
All	All	0.25	0/26354	0.45	1/35833 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	48	LEU	CA-CB-CG	5.41	127.73	115.30

There are no chirality outliers.

There are no planarity outliers.

### 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	1704	0	1741	48	1
1	B	1715	0	1739	52	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	C	8714	0	8635	366	28
3	D	9887	0	9943	324	48
4	E	637	0	635	36	0
5	F	2543	0	2571	135	7
6	G	180	0	103	1	0
7	H	476	0	261	37	0
8	C	66	0	0	15	0
9	D	2	0	0	0	0
10	D	1	0	0	0	0
All	All	25925	0	25628	754	48

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 15.

The worst 5 of 754 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:D:916:ILE:HG22	3:D:921:TYR:CE2	1.30	1.67
2:C:275:LEU:HD11	5:F:212:LEU:CD1	1.29	1.59
2:C:275:LEU:HD13	5:F:212:LEU:CG	1.40	1.51
2:C:275:LEU:CD1	5:F:212:LEU:CD1	2.03	1.37
2:C:275:LEU:CD1	5:F:212:LEU:HG	1.57	1.35

The worst 5 of 48 symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:C:778:ASP:O	3:D:178:GLU:OE2[3_745]	1.03	1.17
2:C:771:ARG:NH1	3:D:198:ARG:NH1[3_745]	1.12	1.08
1:A:159:ILE:CD1	3:D:171:GLU:OE1[3_745]	1.21	0.99
3:D:755:LYS:CE	3:D:888:GLU:OE2[4_575]	1.22	0.98
2:C:787:ARG:NH2	3:D:209:ARG:NH2[3_745]	1.35	0.85

## 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	222/347 (64%)	210 (95%)	11 (5%)	1 (0%)	29	67
1	B	225/347 (65%)	208 (92%)	15 (7%)	2 (1%)	17	54
2	C	1124/1178 (95%)	1049 (93%)	67 (6%)	8 (1%)	22	60
3	D	1261/1316 (96%)	1197 (95%)	58 (5%)	6 (0%)	29	67
4	E	79/110 (72%)	77 (98%)	2 (2%)	0	100	100
5	F	316/528 (60%)	299 (95%)	15 (5%)	2 (1%)	25	63
All	All	3227/3826 (84%)	3040 (94%)	168 (5%)	19 (1%)	25	63

5 of 19 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	D	678	PRO
3	D	412	ARG
1	A	184	GLU
2	C	1148	ARG
5	F	405	ILE

### 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	192/297 (65%)	191 (100%)	1 (0%)	88	93
1	B	192/297 (65%)	192 (100%)	0	100	100
2	C	948/998 (95%)	938 (99%)	10 (1%)	73	84
3	D	1048/1095 (96%)	1036 (99%)	12 (1%)	73	84
4	E	68/90 (76%)	67 (98%)	1 (2%)	65	80
5	F	270/427 (63%)	266 (98%)	4 (2%)	65	80
All	All	2718/3204 (85%)	2690 (99%)	28 (1%)	76	86

5 of 28 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	410	GLN
5	F	527	LEU
3	D	741	ARG
5	F	361	ASP
3	D	714	ASP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	465	HIS
4	E	65	ASN
5	F	388	GLN
4	E	69	ASN
2	C	612	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 monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 4 ligands modelled in this entry, 3 are monoatomic - leaving 1 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
8	KYO	C	1201	-	71,72,72	2.54	22 (30%)	86,109,109	2.44	18 (20%)

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
8	KYO	C	1201	-	-	22/63/100/100	0/6/7/7

The worst 5 of 22 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
8	C	1201	KYO	C37-N3	-7.69	1.35	1.46
8	C	1201	KYO	C37-N2	-7.03	1.36	1.46
8	C	1201	KYO	C15-N1	6.62	1.50	1.35
8	C	1201	KYO	C2-C1	5.62	1.52	1.38
8	C	1201	KYO	C9-C10	5.06	1.55	1.42

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
8	C	1201	KYO	C43-C42-N4	10.05	137.02	115.75
8	C	1201	KYO	C40-N4-C42	-8.38	86.86	111.20
8	C	1201	KYO	C41-N4-C42	-8.09	87.71	111.20
8	C	1201	KYO	C29-C28-C27	7.49	125.27	113.24
8	C	1201	KYO	C16-C15-N1	6.75	121.44	114.45

There are no chirality outliers.

5 of 22 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
8	C	1201	KYO	N1-C15-C16-C17
8	C	1201	KYO	O11-C15-C16-C17
8	C	1201	KYO	C25-C26-C27-O6
8	C	1201	KYO	C27-C28-C29-O5
8	C	1201	KYO	C15-C16-C17-C18

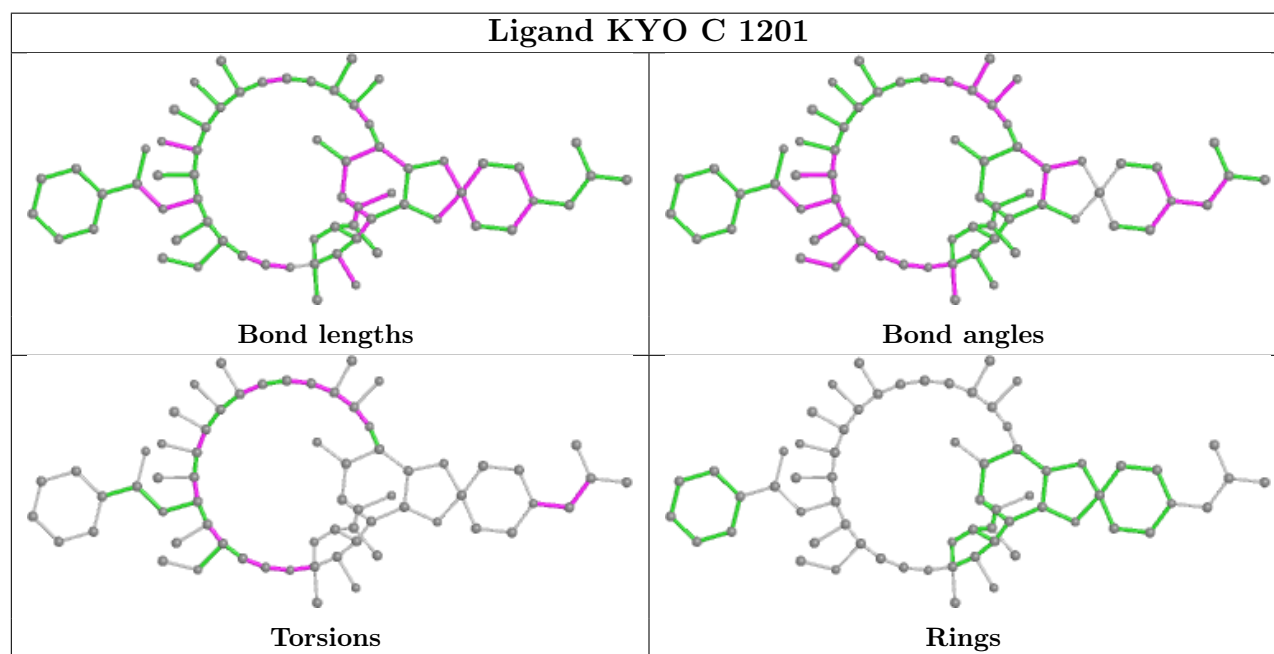
There are no ring outliers.

1 monomer is involved in 15 short contacts:



Mol	Chain	Res	Type	Clashes	Symm-Clashes
8	C	1201	KYO	15	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less than 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.



## 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	224/347 (64%)	0.05	9 (4%) 38 31	128, 177, 260, 316	0
1	B	227/347 (65%)	0.48	27 (11%) 4 4	154, 220, 272, 331	0
2	C	1126/1178 (95%)	0.12	42 (3%) 41 33	101, 162, 263, 329	0
3	D	1265/1316 (96%)	0.48	130 (10%) 6 6	95, 163, 367, 500	0
4	E	81/110 (73%)	-0.09	0 100 100	119, 171, 218, 238	0
5	F	320/528 (60%)	0.45	33 (10%) 6 6	112, 185, 324, 402	0
6	G	9/16 (56%)	1.30	2 (22%) 0 0	269, 283, 298, 312	0
7	H	23/23 (100%)	0.47	2 (8%) 10 8	138, 233, 298, 341	0
All	All	3275/3865 (84%)	0.31	245 (7%) 14 11	95, 171, 305, 500	0

The worst 5 of 245 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	D	756	VAL	14.1
3	D	835	PRO	13.7
5	F	211	ALA	12.7
3	D	826	ASN	11.9
5	F	210	GLU	11.8

### 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](#)

LIGAND-RSR INFOmissingINFO

## 6.5 Other polymers [i](#)

There are no such residues in this entry.