



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 6, 2026 – 04:27 PM UTC

PDB ID : 7U22 / pdb_00007u22
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

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 : 4-5-2 with Phenix2.0
Mogul : 2022.3.0, CSD as543be (2022)
Xtriage (Phenix) : 2.0
EDS : 3.0
Buster-report : wwPDB partial adaption of 1.1.7 (2018)
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4 : 9.0.010 (Gargrove)
Density-Fitness : 1.0.12
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.49

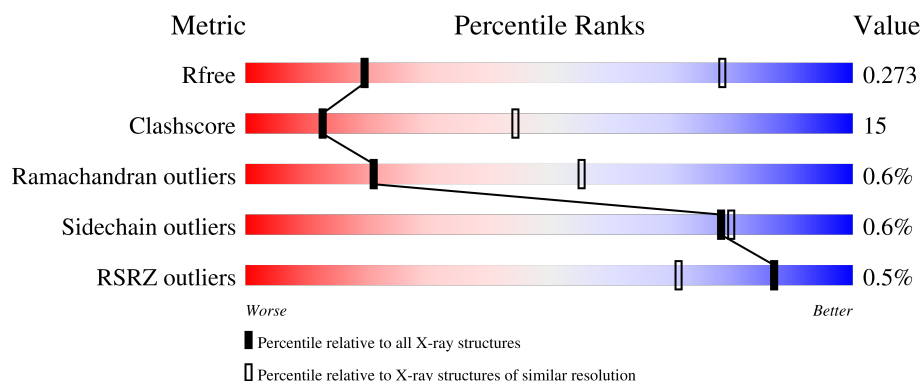
1 Overall quality at a glance

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.



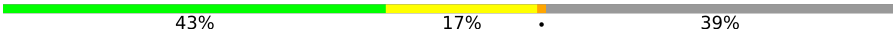
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	180053	1186 (4.06-3.70)
Clashscore	190562	1012 (4.04-3.72)
Ramachandran outliers	187476	1167 (4.06-3.70)
Sidechain outliers	187428	1160 (4.06-3.70)
RSRZ outliers	180081	1185 (4.06-3.70)

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 of the lower bar indicate the fraction of residues that contain outliers for ≥ 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	 43% 17% 39%
6	G	16	 50% 6% 44%
7	H	23	 30% 70%

2 Entry composition

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
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
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
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
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
5	F	320	Total	C	N	O	S	0	0	0
			2543	1583	459	492	9			

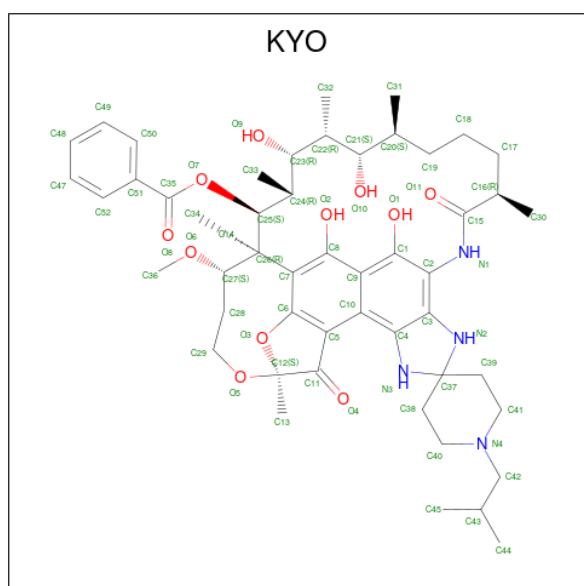
- Molecule 6 is a DNA chain called T DNA.

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

- Molecule 7 is a DNA chain called NT DNA.

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

- 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-tetrahydro spiro[9,4-(epoxypentadecanoimino)furo[2',3':7,8]naphtho[1,2-d]imidazole-2,4'-piperidin]-16-yl benzoate (CCD ID: KYO) (formula: $C_{51}H_{72}N_4O_{11}$) (labeled as "Ligand of Interest" by depositor).



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

- Molecule 9 is ZINC ION (CCD ID: ZN) (formula: Zn).

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

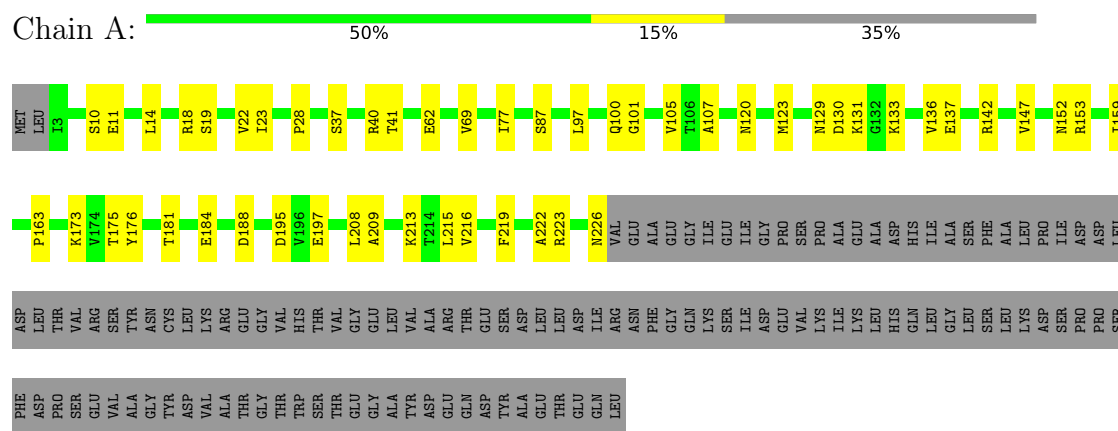
- Molecule 10 is MAGNESIUM ION (CCD ID: MG) (formula: Mg).

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

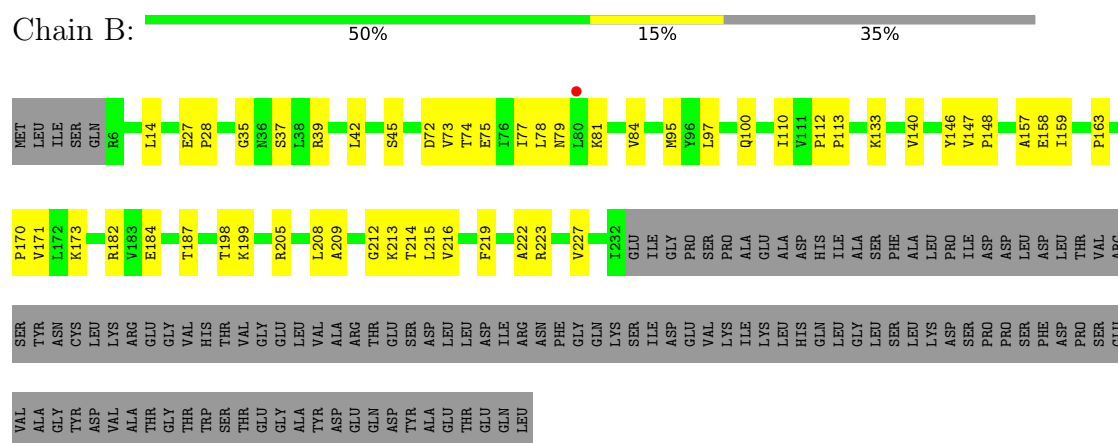
3 Residue-property plots

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

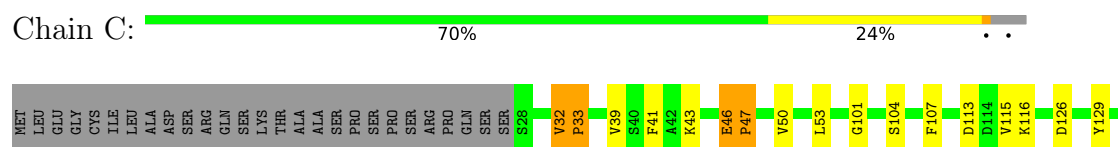
• Molecule 1: DNA-directed RNA polymerase subunit alpha

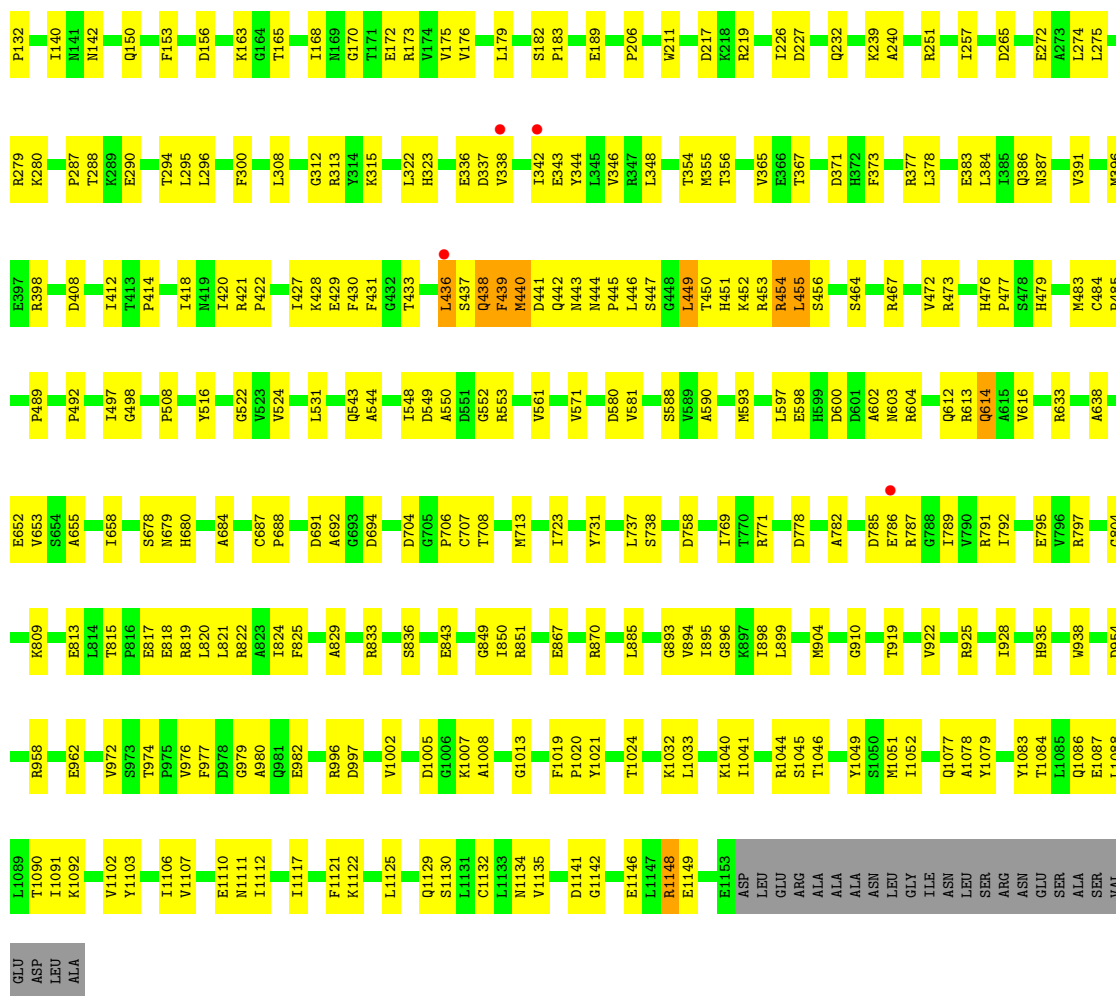


• Molecule 1: DNA-directed RNA polymerase subunit alpha



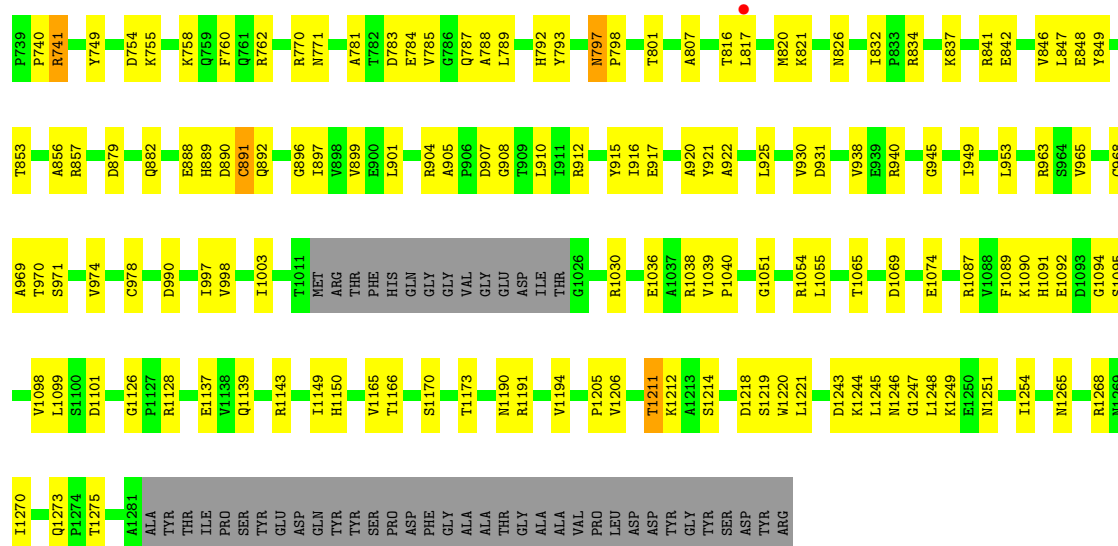
• Molecule 2: DNA-directed RNA polymerase subunit beta





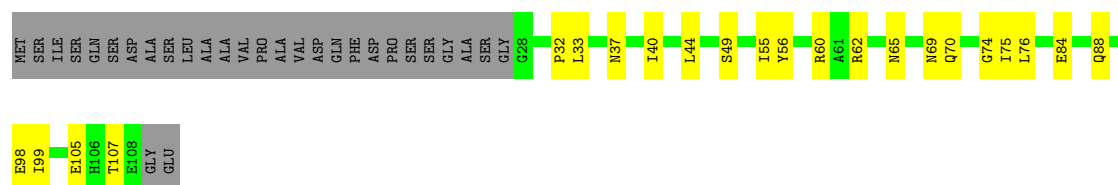
• Molecule 3: DNA-directed RNA polymerase subunit beta'





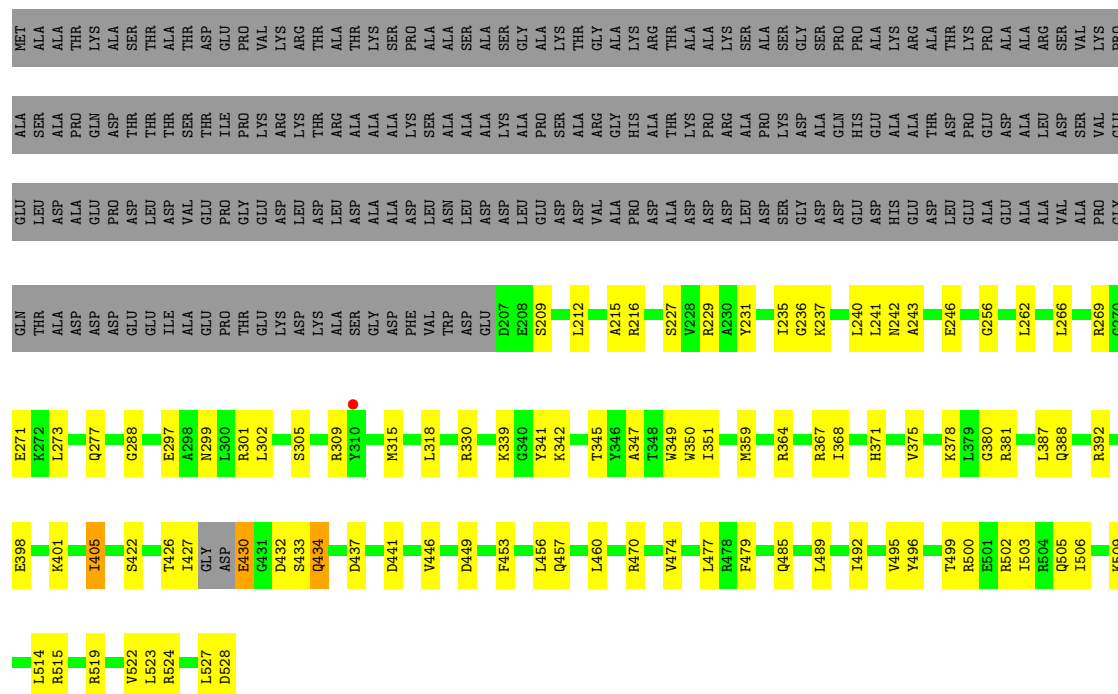
• Molecule 4: DNA-directed RNA polymerase subunit omega

Chain E: 54% 20% 26%



• Molecule 5: RNA polymerase sigma factor SigA

Chain F: 43% 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, α , β , γ	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) 98.9 (48.30-3.87)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.37 (at 3.88Å)	Xtriage
Refinement program	PHENIX 1.19_4092, PHENIX 1.19_4092	Depositor
R, R_{free}	0.240 , 0.278 0.239 , 0.273	Depositor DCC
R_{free} test set	1995 reflections (4.43%)	wwPDB-VP
Wilson B-factor (Å ²)	152.2	Xtriage
Anisotropy	0.310	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 120.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.43$, $\langle L^2 \rangle = 0.26$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	25925	wwPDB-VP
Average B, all atoms (Å ²)	187.0	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.*

¹Intensities estimated from amplitudes.

²Theoretical 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: KYO, 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	0.32	0/1730	0.80	2/2354 (0.1%)
1	B	0.32	0/1741	0.80	2/2371 (0.1%)
2	C	0.31	0/8873	0.77	2/12031 (0.0%)
3	D	0.32	0/10052	0.77	7/13591 (0.1%)
4	E	0.32	0/650	0.76	0/886
5	F	0.32	0/2572	0.77	0/3466
6	G	0.20	0/201	0.34	0/308
7	H	0.20	0/535	0.38	0/826
All	All	0.31	0/26354	0.76	13/35833 (0.0%)

There are no bond length outliers.

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	1126	GLY	CA-C-N	6.16	126.61	119.47
3	D	1126	GLY	C-N-CA	6.16	126.61	119.47
3	D	321	PRO	CA-C-N	5.96	125.66	119.05
3	D	321	PRO	C-N-CA	5.96	125.66	119.05
3	D	797	ASN	CA-C-N	5.56	125.40	119.28

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	50	1
1	B	1715	0	1739	54	0
2	C	8714	0	8635	381	28
3	D	9887	0	9943	331	48
4	E	637	0	635	38	0
5	F	2543	0	2571	134	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	782	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 782 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

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	222/347 (64%)	210 (95%)	11 (5%)	1 (0%)	24	59
1	B	225/347 (65%)	208 (92%)	15 (7%)	2 (1%)	14	47
2	C	1124/1178 (95%)	1049 (93%)	67 (6%)	8 (1%)	18	52
3	D	1261/1316 (96%)	1197 (95%)	58 (5%)	6 (0%)	24	59
4	E	79/110 (72%)	77 (98%)	2 (2%)	0	100	100
5	F	316/528 (60%)	299 (95%)	15 (5%)	2 (1%)	21	55
All	All	3227/3826 (84%)	3040 (94%)	168 (5%)	19 (1%)	21	55

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

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%)	192 (100%)	0	100	100
1	B	192/297 (65%)	192 (100%)	0	100	100
2	C	948/998 (95%)	937 (99%)	11 (1%)	63	72

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
3	D	1048/1095 (96%)	1043 (100%)	5 (0%)	81	81
4	E	68/90 (76%)	68 (100%)	0	100	100
5	F	270/427 (63%)	269 (100%)	1 (0%)	84	83
All	All	2718/3204 (85%)	2701 (99%)	17 (1%)	78	80

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	891	CYS
5	F	430	GLU
2	C	455	LEU
2	C	456	SER
2	C	612	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 27 such sidechains are listed below:

Mol	Chain	Res	Type
3	D	303	GLN
3	D	510	GLN
4	E	70	GLN
3	D	467	GLN
3	D	639	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no oligosaccharides in this entry.

5.6 Ligand geometry

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	-	70,72,72	2.27	22 (31%)	88,109,109	2.46	21 (23%)

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	C15-N1	6.58	1.50	1.35
8	C	1201	KYO	C2-C1	5.83	1.52	1.38
8	C	1201	KYO	C9-C10	5.30	1.55	1.42
8	C	1201	KYO	C8-C9	5.15	1.58	1.43
8	C	1201	KYO	O7-C25	-4.58	1.38	1.44

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
8	C	1201	KYO	C40-N4-C42	-8.84	86.86	111.35
8	C	1201	KYO	C43-C42-N4	8.60	137.02	115.66
8	C	1201	KYO	C41-N4-C42	-8.53	87.71	111.35
8	C	1201	KYO	C29-C28-C27	7.82	125.27	113.22
8	C	1201	KYO	C38-C37-N3	-7.39	106.87	111.23

There are no chirality outliers.

5 of 22 torsion outliers are listed below:

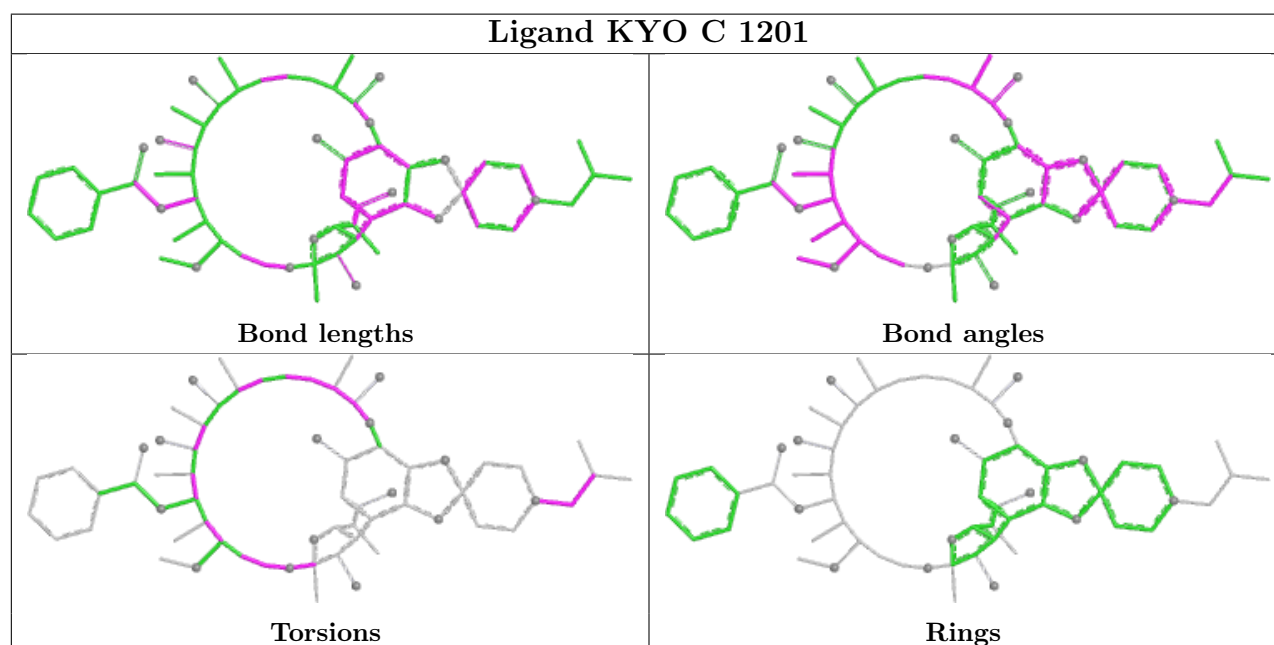
Mol	Chain	Res	Type	Atoms
8	C	1201	KYO	C25-C26-C27-O6
8	C	1201	KYO	C27-C28-C29-O5
8	C	1201	KYO	C15-C16-C17-C18
8	C	1201	KYO	C28-C29-O5-C12
8	C	1201	KYO	N4-C42-C43-C44

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, 95th 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 ²)	Q<0.9
1	A	224/347 (64%)	-0.54	0 100 100	128, 177, 260, 316	0
1	B	227/347 (65%)	-0.53	1 (0%) 88 74	154, 220, 272, 331	0
2	C	1126/1178 (95%)	-0.47	4 (0%) 88 74	101, 162, 263, 329	0
3	D	1265/1316 (96%)	-0.43	9 (0%) 84 66	95, 163, 367, 500	0
4	E	81/110 (73%)	-0.61	0 100 100	119, 171, 218, 238	0
5	F	320/528 (60%)	-0.42	1 (0%) 90 77	112, 185, 324, 402	0
6	G	9/16 (56%)	0.60	0 100 100	269, 283, 289, 312	0
7	H	23/23 (100%)	0.01	0 100 100	138, 233, 298, 341	0
All	All	3275/3865 (84%)	-0.45	15 (0%) 87 71	95, 171, 305, 500	0

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	D	167	ASP	3.7
3	D	197	VAL	3.1
2	C	436	LEU	2.9
1	B	80	LEU	2.6
3	D	193	ALA	2.5

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 oligosaccharides in this entry.

6.4 Ligands [i](#)

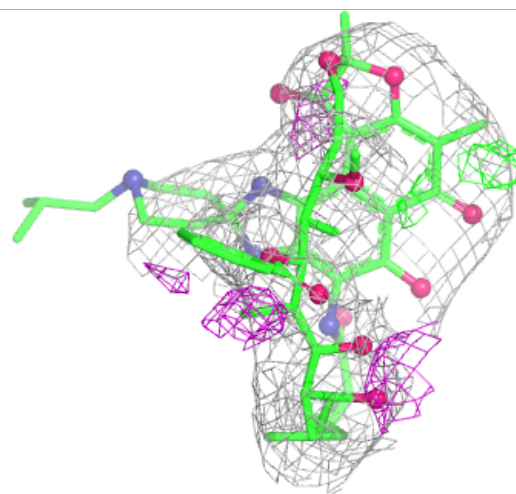
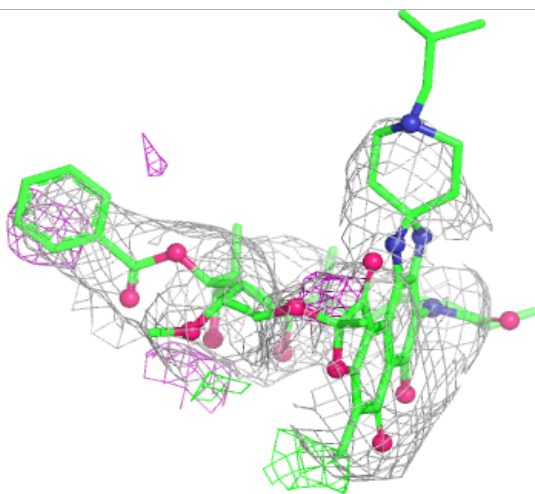
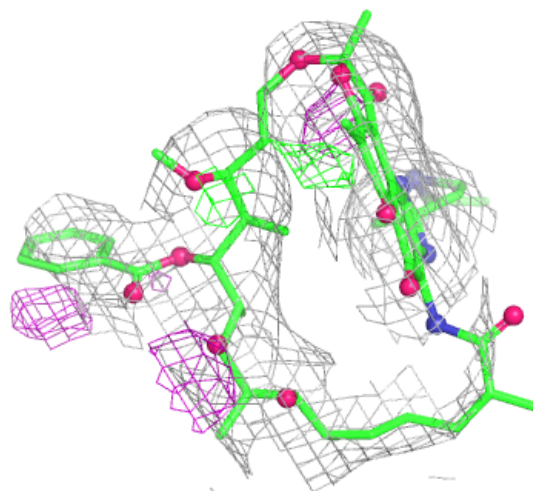
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists 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
8	KYO	C	1201	66/66	0.86	0.10	84,126,203,242	0
9	ZN	D	1401	1/1	-	-	168,168,168,168	1
9	ZN	D	1402	1/1	-	-	154,154,154,154	1
10	MG	D	1403	1/1	-	-	143,143,143,143	1

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around KYO C 1201:

2mF_o-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)



6.5 Other polymers [i](#)

There are no such residues in this entry.