



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 11, 2023 – 11:26 AM EDT

PDB ID : 7RWI  
Title : Mycobacterium tuberculosis RNA polymerase sigma L holoenzyme open promoter complex containing TNP-2198  
Authors : Molodtsov, V.; Ebright, R.H.  
Deposited on : 2021-08-19  
Resolution : 3.70 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)  
Xtriage (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

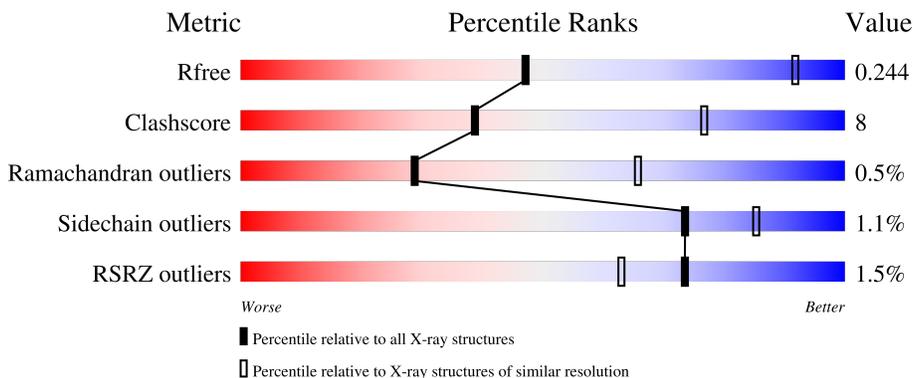
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.70 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	1049 (3.88-3.52)
Clashscore	141614	1027 (3.86-3.54)
Ramachandran outliers	138981	1069 (3.88-3.52)
Sidechain outliers	138945	1065 (3.88-3.52)
RSRZ outliers	127900	1578 (3.90-3.50)

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	177	
6	G	23	
7	H	27	

## 2 Entry composition [i](#)

There are 8 unique types of molecules in this entry. The entry contains 24903 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	225	Total	C	N	O	S	0	0	0
			1716	1080	296	338	2			
1	B	232	Total	C	N	O	S	0	0	0
			1732	1093	296	341	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
			8724	5459	1531	1695	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	1264	Total	C	N	O	S	0	0	0
			9884	6189	1790	1865	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
			630	403	106	121			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	F	174	Total	C	N	O	S	0	0	0
			1352	840	256	254	2			

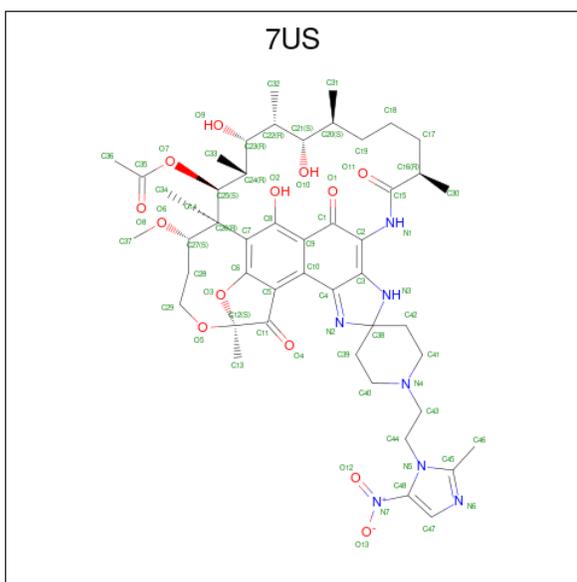
- Molecule 6 is a DNA chain called T DNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
6	G	15	306	146	58	88	14	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	24	491	234	90	144	23	0	0	0

- Molecule 8 is (3aM,9S,10bP,14S,15R,16S,17R,18R,19R,20S,21S,25R)-6,18,20-trihydroxy-14-methoxy-7,9,15,17,19,21,25-heptamethyl-1'-[2-(2-methyl-5-nitro-1H-imidazol-1-yl)ethyl]-5,10,26-trioxo-3,5,9,10-tetrahydrospiro[9,4-(epoxypentadecanoimino)furo[2',3':7,8]naphtho[1,2-d]imidazole-2,4'-piperidin]-16-yl acetate (three-letter code: 7US) (formula: C<sub>48</sub>H<sub>67</sub>N<sub>7</sub>O<sub>13</sub>).

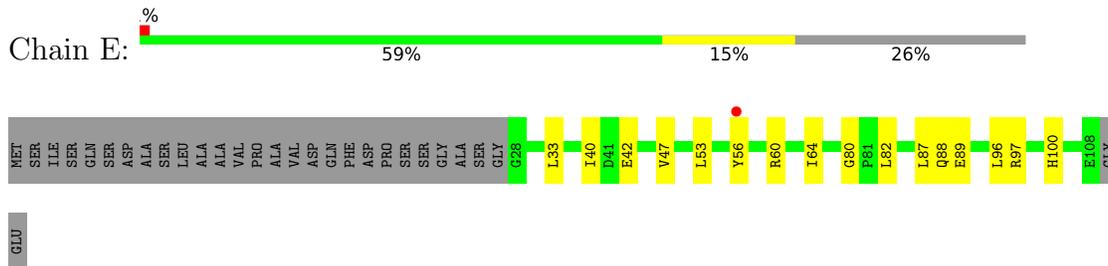


Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
8	C	1	68	48	7	13	38	0

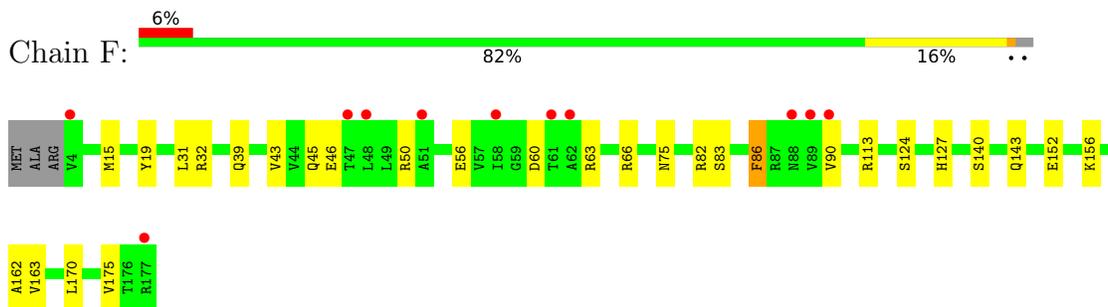




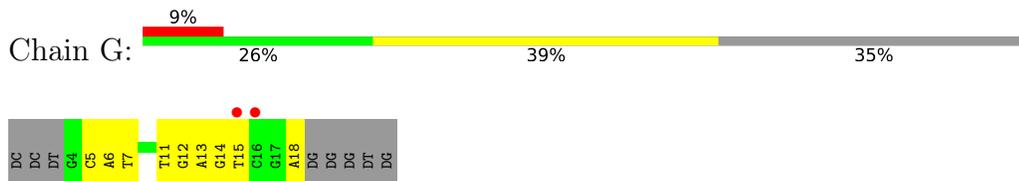
● Molecule 4: DNA-directed RNA polymerase subunit omega



● Molecule 5: RNA polymerase sigma factor



● 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, $\alpha$ , $\beta$ , $\gamma$	142.77Å 160.61Å 238.65Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.71 – 3.70 48.71 – 3.70	Depositor EDS
% Data completeness (in resolution range)	97.1 (48.71-3.70) 97.1 (48.71-3.70)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.53 (at 3.67Å)	Xtriage
Refinement program	PHENIX 1.19_4092, PHENIX 1.19_4092	Depositor
R, $R_{free}$	0.208 , 0.245 0.207 , 0.244	Depositor DCC
$R_{free}$ test set	1997 reflections (3.48%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	132.1	Xtriage
Anisotropy	0.612	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 97.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	24903	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	167.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.51% 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: 7US

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	0.26	0/1742	0.56	0/2370
1	B	0.26	0/1758	0.55	0/2397
2	C	0.25	0/8883	0.51	1/12043 (0.0%)
3	D	0.25	0/10049	0.50	0/13583
4	E	0.24	0/643	0.45	0/877
5	F	0.25	0/1374	0.52	0/1869
6	G	0.44	0/343	0.84	0/528
7	H	0.54	0/550	0.93	0/848
All	All	0.26	0/25342	0.53	1/34515 (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	677	ARG	O-C-N	-5.96	113.16	122.70

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	1716	0	1756	42	0
1	B	1732	0	1754	37	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	C	8724	0	8651	160	0
3	D	9884	0	9947	124	0
4	E	630	0	622	11	0
5	F	1352	0	1346	20	0
6	G	306	0	170	10	0
7	H	491	0	272	18	0
8	C	68	0	0	10	0
All	All	24903	0	24518	369	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 8.

The worst 5 of 369 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:442:GLN:HG3	2:C:679:ASN:ND2	1.37	1.36
2:C:442:GLN:CG	2:C:679:ASN:HD22	1.43	1.30
2:C:430:PHE:O	2:C:434:SER:HB2	1.26	1.25
2:C:454:ARG:HH22	8:C:1201:7US:C19	1.53	1.22
2:C:442:GLN:CB	2:C:679:ASN:HD22	1.60	1.13

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	223/347 (64%)	213 (96%)	9 (4%)	1 (0%)	34 69
1	B	230/347 (66%)	217 (94%)	10 (4%)	3 (1%)	12 47
2	C	1124/1178 (95%)	1061 (94%)	54 (5%)	9 (1%)	19 56
3	D	1258/1316 (96%)	1206 (96%)	50 (4%)	2 (0%)	47 78

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
4	E	79/110 (72%)	76 (96%)	3 (4%)	0	100	100
5	F	172/177 (97%)	167 (97%)	5 (3%)	0	100	100
All	All	3086/3475 (89%)	2940 (95%)	131 (4%)	15 (0%)	29	66

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	184	GLU
1	B	159	ILE
2	C	370	ILE
2	C	677	ARG
2	C	680	HIS

### 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	194/297 (65%)	194 (100%)	0	100	100
1	B	191/297 (64%)	189 (99%)	2 (1%)	76	86
2	C	950/998 (95%)	939 (99%)	11 (1%)	71	84
3	D	1049/1095 (96%)	1037 (99%)	12 (1%)	73	85
4	E	66/90 (73%)	66 (100%)	0	100	100
5	F	134/136 (98%)	130 (97%)	4 (3%)	41	66
All	All	2584/2913 (89%)	2555 (99%)	29 (1%)	73	85

5 of 29 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	279	ASP
5	F	66	ARG
3	D	443	LEU
3	D	804	ASP
3	D	307	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
2	C	679	ASN
2	C	680	HIS

### 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](#)

1 ligand 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
8	7US	C	1201	-	65,74,74	4.15	18 (27%)	80,113,113	3.54	26 (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
8	7US	C	1201	-	-	20/60/121/121	1/6/7/7

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
8	C	1201	7US	C4-N2	23.52	1.46	1.29
8	C	1201	7US	C43-N4	-11.73	1.20	1.47
8	C	1201	7US	O3-C6	9.85	1.56	1.37
8	C	1201	7US	C10-C4	9.39	1.58	1.46
8	C	1201	7US	C15-N1	7.26	1.49	1.36

The worst 5 of 26 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
8	C	1201	7US	C39-C38-N2	21.37	130.87	111.42
8	C	1201	7US	O3-C6-C5	-9.96	107.56	114.36
8	C	1201	7US	C29-C28-C27	7.50	125.28	113.24
8	C	1201	7US	C16-C15-N1	6.73	121.42	114.45
8	C	1201	7US	C5-C6-C7	-6.46	120.25	125.33

There are no chirality outliers.

5 of 20 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
8	C	1201	7US	N1-C15-C16-C17
8	C	1201	7US	O11-C15-C16-C17
8	C	1201	7US	C15-C16-C17-C18
8	C	1201	7US	C25-C26-C27-O6
8	C	1201	7US	C27-C28-C29-O5

All (1) ring outliers are listed below:

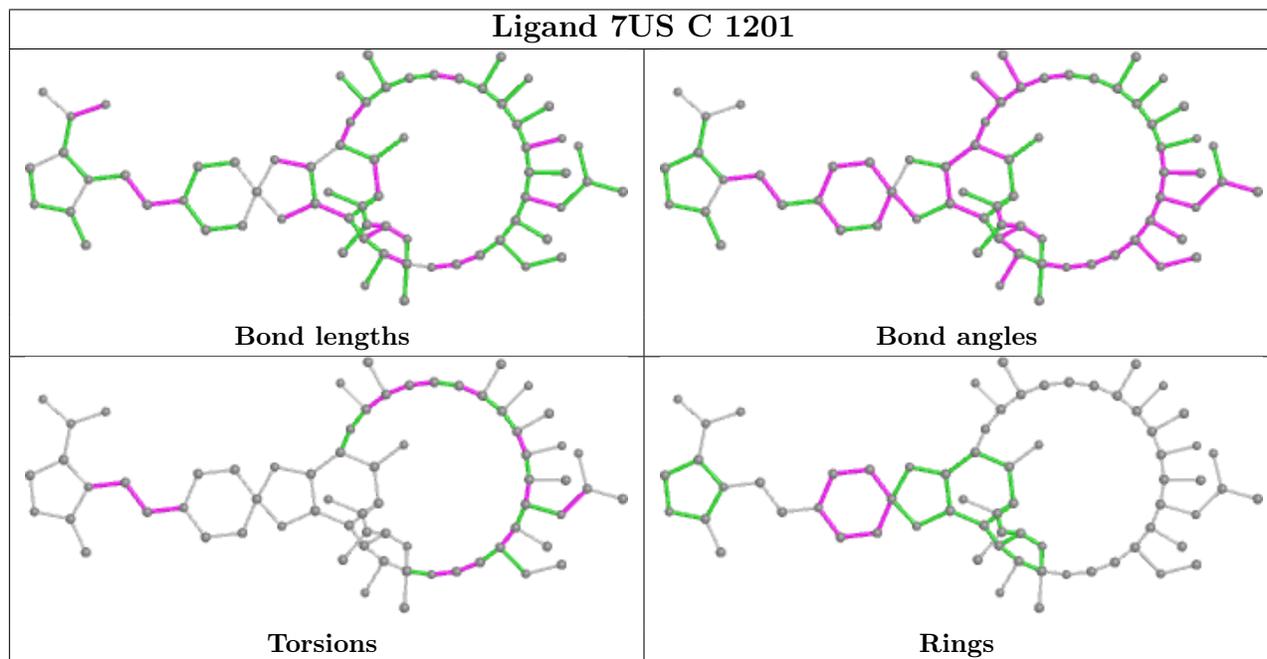
Mol	Chain	Res	Type	Atoms
8	C	1201	7US	C38-C39-C40-C41-C42-N4

1 monomer is involved in 10 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
8	C	1201	7US	10	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	225/347 (64%)	-0.30	3 (1%) 77 67	88, 125, 198, 262	0
1	B	232/347 (66%)	-0.24	5 (2%) 62 50	102, 152, 228, 274	0
2	C	1126/1178 (95%)	-0.15	13 (1%) 79 69	86, 145, 242, 291	0
3	D	1264/1316 (96%)	-0.13	12 (0%) 84 76	83, 163, 270, 312	0
4	E	81/110 (73%)	0.03	1 (1%) 79 69	122, 161, 232, 285	0
5	F	174/177 (98%)	-0.01	11 (6%) 20 13	111, 173, 251, 294	0
6	G	15/23 (65%)	0.07	2 (13%) 3 3	212, 245, 280, 283	0
7	H	24/27 (88%)	0.04	1 (4%) 36 27	224, 263, 294, 310	0
All	All	3141/3525 (89%)	-0.15	48 (1%) 73 63	83, 155, 256, 312	0

The worst 5 of 48 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	157	ALA	6.6
1	B	156	GLY	5.7
5	F	89	VAL	4.5
2	C	192	ASP	4.1
1	B	158	GLU	4.0

### 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](#)

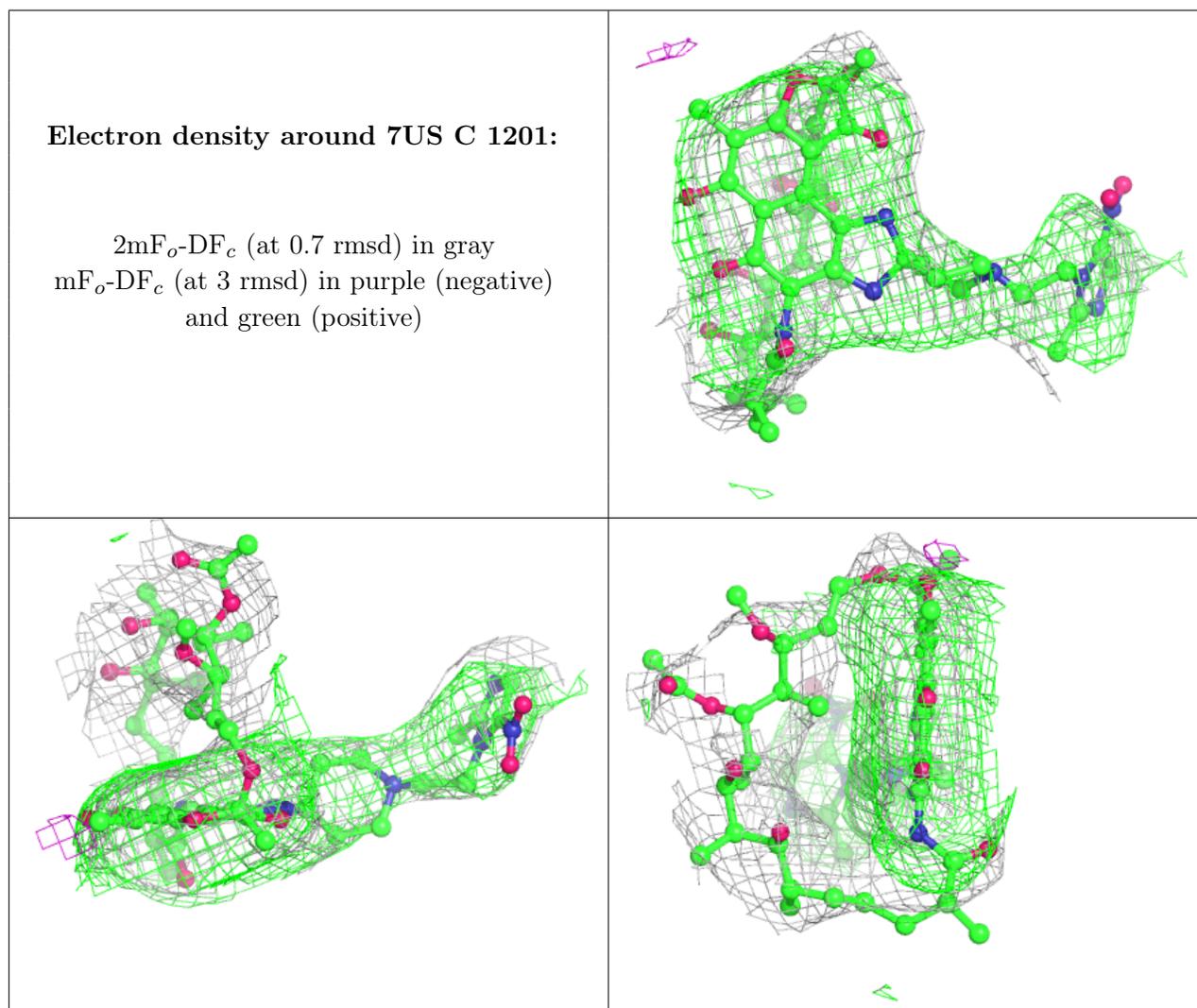
There are no monosaccharides in this entry.

## 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
8	7US	C	1201	68/68	0.82	0.32	80,161,189,211	38

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.



## 6.5 Other polymers [i](#)

There are no such residues in this entry.