



wwPDB X-ray Structure Validation Summary Report ⓘ

May 15, 2020 – 06:57 pm BST

PDB ID : 5UHB
Title : Crystal structure of Mycobacterium tuberculosis transcription initiation complex in complex with Rifampin
Authors : Lin, W.; Das, K.; Feng, Y.; Ebright, R.H.
Deposited on : 2017-01-11
Resolution : 4.29 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 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 : 2.11
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) : 2.11

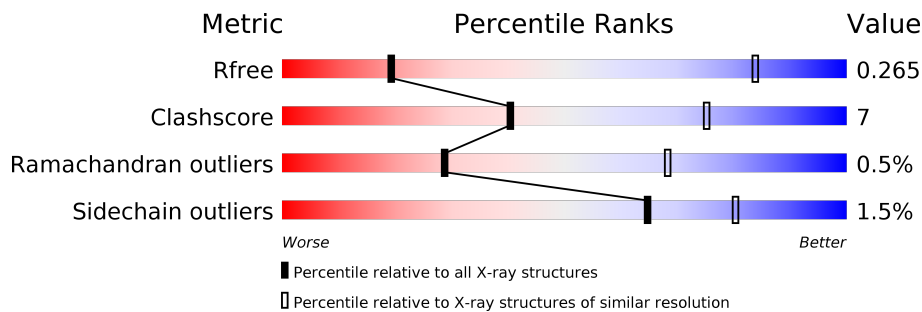
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 4.29 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	1014 (4.80-3.80)
Clashscore	141614	1077 (4.80-3.80)
Ramachandran outliers	138981	1029 (4.80-3.80)
Sidechain outliers	138945	1012 (4.80-3.80)

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 ≥ 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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	347	
1	B	347	
2	C	1178	
3	D	1316	
4	E	110	
5	F	528	
6	H	23	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
7	G	16	 44% 25% 6% 25%

2 Entry composition [i](#)

There are 10 unique types of molecules in this entry. The entry contains 25979 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			

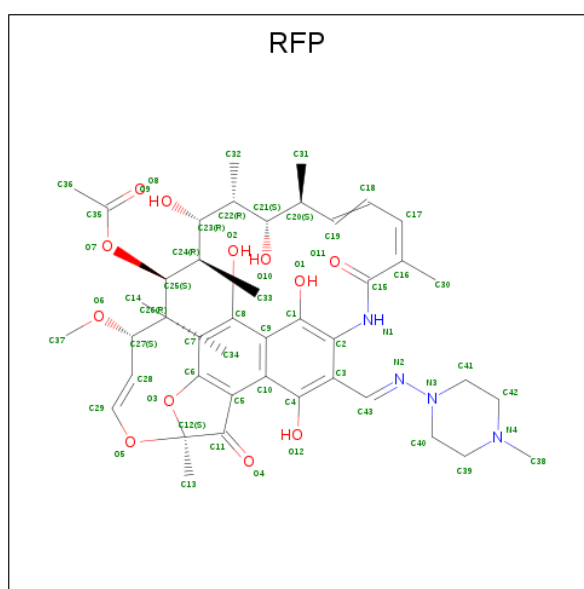
- Molecule 6 is a DNA chain called DNA (5'-D(*TP*AP*TP*AP*AP*TP*GP*GP*GP*AP*GP*CP*TP*GP*TP*CP*AP*CP*GP*GP*AP*TP*G)-3').

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

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
7	G	12	241	116	43	71	11	0	0	0

- Molecule 8 is RIFAMPICIN (three-letter code: RFP) (formula: C₄₃H₅₈N₄O₁₂).



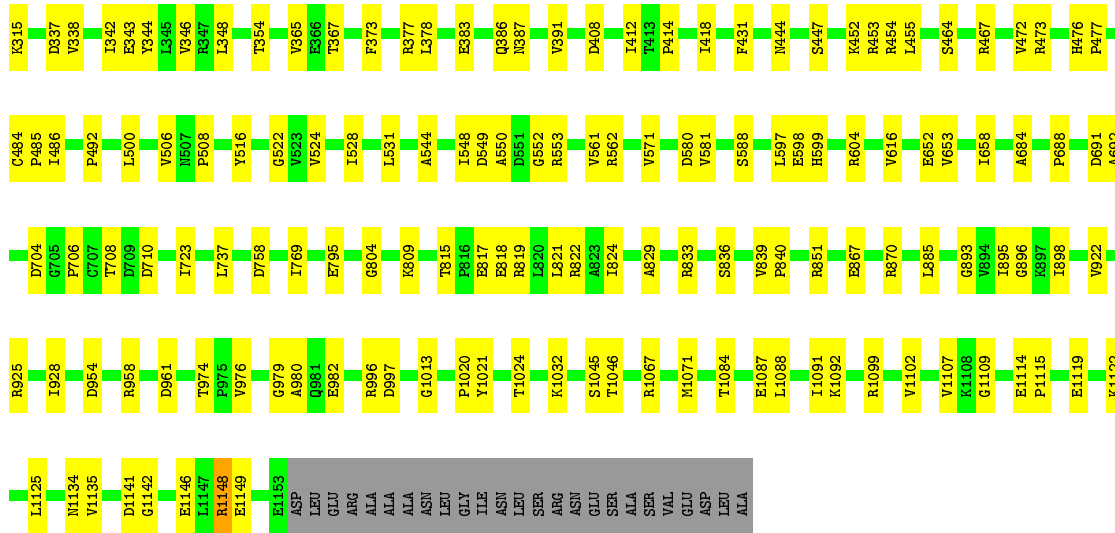
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
8	C	1	59	43	4	12	0	0

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

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

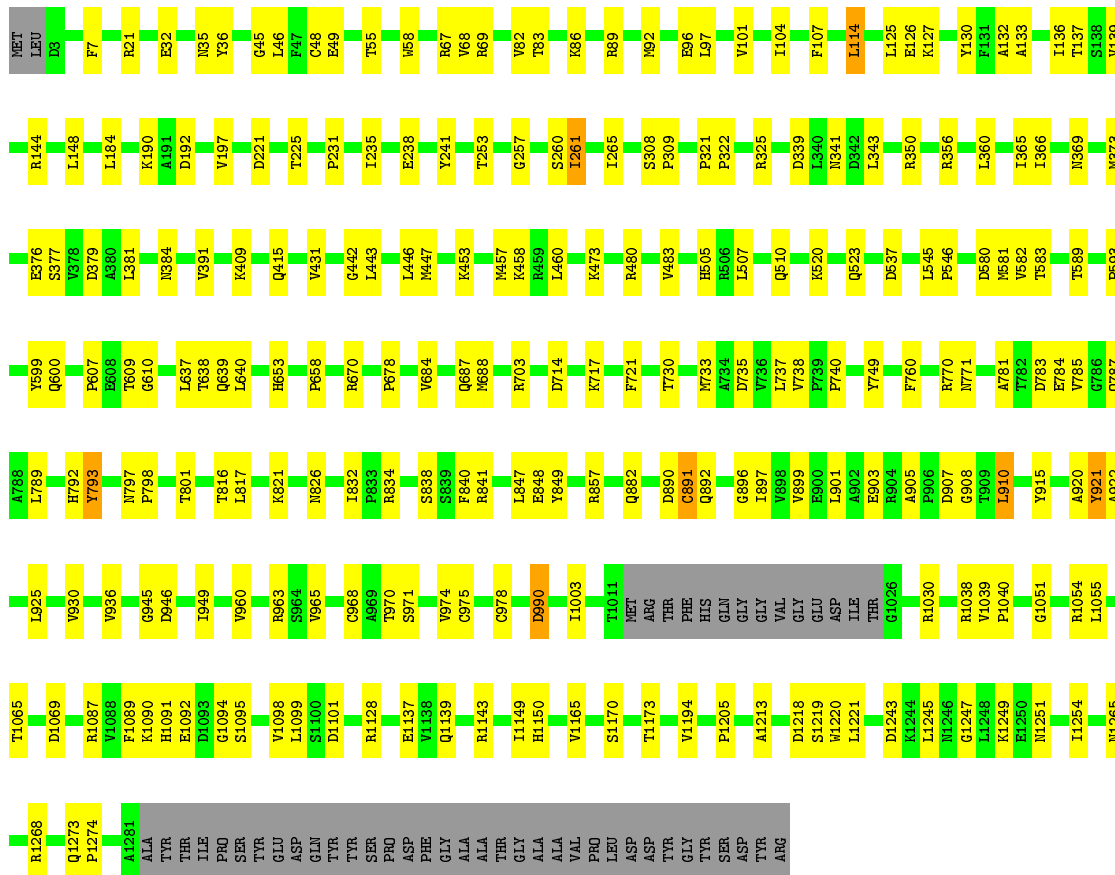
- Molecule 10 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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



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

Chain D: 78% 17%



• Molecule 4: DNA-directed RNA polymerase subunit omega

Chain E: 58% 15% 26%

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	154.19Å 164.23Å 200.20Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.78 – 4.29 49.78 – 4.29	Depositor EDS
% Data completeness (in resolution range)	83.4 (49.78-4.29) 83.4 (49.78-4.29)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.67 (at 4.29Å)	Xtrriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.208 , 0.267 0.211 , 0.265	Depositor DCC
R_{free} test set	2006 reflections (6.78%)	wwPDB-VP
Wilson B-factor (Å ²)	31.8	Xtrriage
Anisotropy	0.545	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.27 , 16.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.41$, $\langle L^2 \rangle = 0.24$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.79	EDS
Total number of atoms	25979	wwPDB-VP
Average B, all atoms (Å ²)	62.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.62% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² 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, RFP, 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	0.23	0/1730	0.44	0/2354
1	B	0.23	0/1741	0.44	0/2371
2	C	0.23	0/8873	0.42	1/12031 (0.0%)
3	D	0.24	0/10052	0.45	5/13591 (0.0%)
4	E	0.25	0/650	0.42	0/886
5	F	0.25	0/2572	0.41	0/3466
6	H	0.55	0/535	0.89	0/826
7	G	0.55	0/269	1.11	1/413 (0.2%)
All	All	0.25	0/26422	0.46	7/35938 (0.0%)

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
7	G	12	DG	P-O3'-C3'	11.67	133.70	119.70
3	D	910	LEU	O-C-N	11.09	140.44	122.70
3	D	910	LEU	CA-C-N	-8.90	97.62	117.20
3	D	910	LEU	C-N-CA	-7.78	102.25	121.70
3	D	1213	ALA	N-CA-CB	6.34	118.97	110.10

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	30	0
1	B	1715	0	1739	22	0
2	C	8714	0	8636	129	0
3	D	9887	0	9943	159	0
4	E	637	0	635	9	0
5	F	2543	0	2571	55	0
6	H	476	0	261	10	0
7	G	241	0	137	8	0
8	C	59	0	58	6	0
9	D	2	0	0	0	0
10	D	1	0	0	0	0
All	All	25979	0	25721	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 7.

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:562:ARG:NH2	3:D:847:LEU:CD1	2.05	1.19
2:C:562:ARG:NH2	3:D:847:LEU:HD11	1.62	1.11
3:D:891:CYS:SG	3:D:970:THR:OG1	2.09	1.09
2:C:562:ARG:HH22	3:D:847:LEU:HD11	1.23	0.93
7:G:11:DT:H2''	7:G:12:DG:H8	1.32	0.92

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	222/347 (64%)	210 (95%)	11 (5%)	1 (0%)	29 68
1	B	225/347 (65%)	208 (92%)	15 (7%)	2 (1%)	17 56

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	C	1124/1178 (95%)	1053 (94%)	63 (6%)	8 (1%)	22	62
3	D	1261/1316 (96%)	1197 (95%)	59 (5%)	5 (0%)	34	72
4	E	79/110 (72%)	77 (98%)	2 (2%)	0	100	100
5	F	316/528 (60%)	302 (96%)	13 (4%)	1 (0%)	41	76
All	All	3227/3826 (84%)	3047 (94%)	163 (5%)	17 (0%)	29	68

5 of 17 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	D	678	PRO
1	A	184	GLU
2	C	1148	ARG
5	F	405	ILE
1	B	158	GLU

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%)	190 (99%)	2 (1%)	76	86
1	B	192/297 (65%)	192 (100%)	0	100	100
2	C	948/998 (95%)	939 (99%)	9 (1%)	78	88
3	D	1048/1095 (96%)	1025 (98%)	23 (2%)	52	71
4	E	68/90 (76%)	65 (96%)	3 (4%)	28	54
5	F	270/427 (63%)	265 (98%)	5 (2%)	57	75
All	All	2718/3204 (85%)	2676 (98%)	42 (2%)	65	80

5 of 42 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	261	ILE
3	D	714	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
5	F	361	ASP
3	D	582	VAL
3	D	653	HIS

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

Mol	Chain	Res	Type
2	C	969	ASN
3	D	465	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 carbohydrates 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	RFP	C	1201	-	63,63,63	2.62	12 (19%)	94,94,94	2.24	20 (21%)

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	RFP	C	1201	-	-	21/60/85/85	0/5/5/5

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
8	C	1201	RFP	O4-C11	11.49	1.40	1.21
8	C	1201	RFP	C43-N2	8.14	1.50	1.27
8	C	1201	RFP	C17-C16	7.26	1.55	1.34
8	C	1201	RFP	C3-C43	5.40	1.56	1.46
8	C	1201	RFP	C15-N1	5.06	1.46	1.35

The worst 5 of 20 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
8	C	1201	RFP	C43-N2-N3	-8.54	108.44	120.43
8	C	1201	RFP	C38-N4-C42	8.44	123.28	110.66
8	C	1201	RFP	C30-C16-C17	-7.35	105.63	123.42
8	C	1201	RFP	C2-C3-C43	-6.73	117.33	124.17
8	C	1201	RFP	O3-C6-C7	5.17	130.03	121.14

There are no chirality outliers.

5 of 21 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
8	C	1201	RFP	C4-C3-C43-N2
8	C	1201	RFP	C15-C16-C17-C18
8	C	1201	RFP	C26-C27-C28-C29
8	C	1201	RFP	O6-C27-C28-C29
8	C	1201	RFP	C43-N2-N3-C40

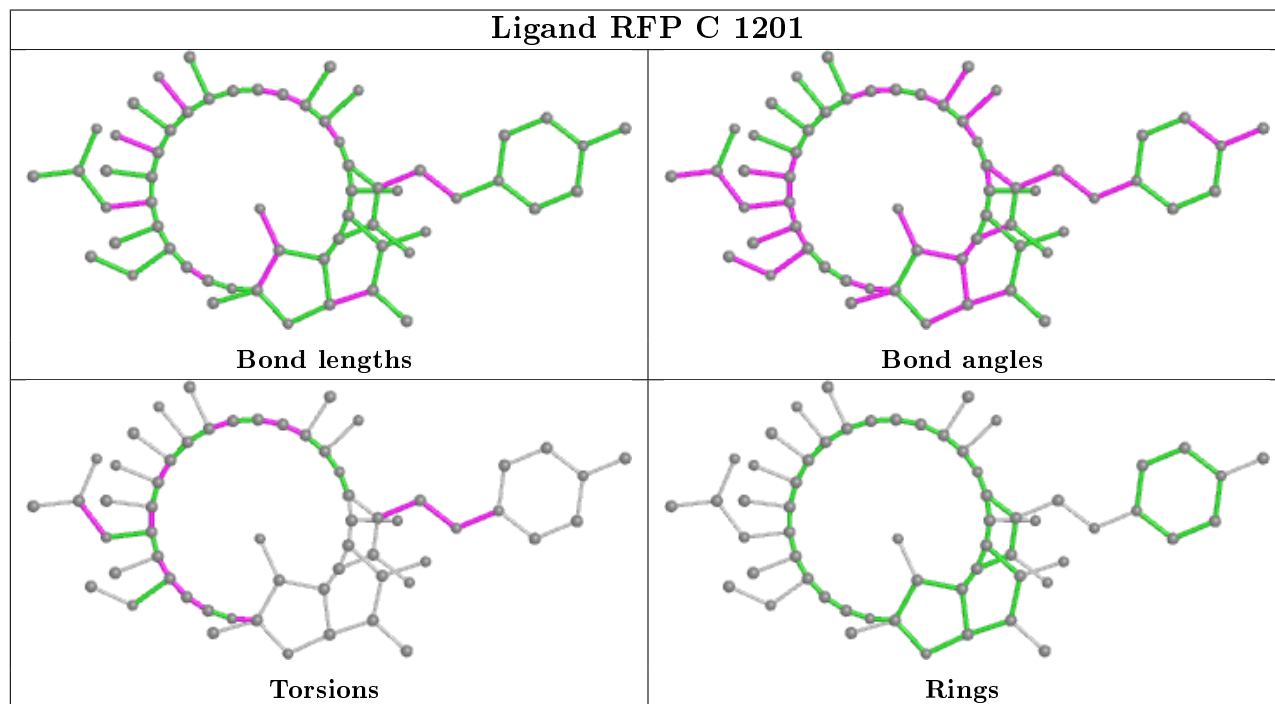
There are no ring outliers.

1 monomer is involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
8	C	1201	RFP	6	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

6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

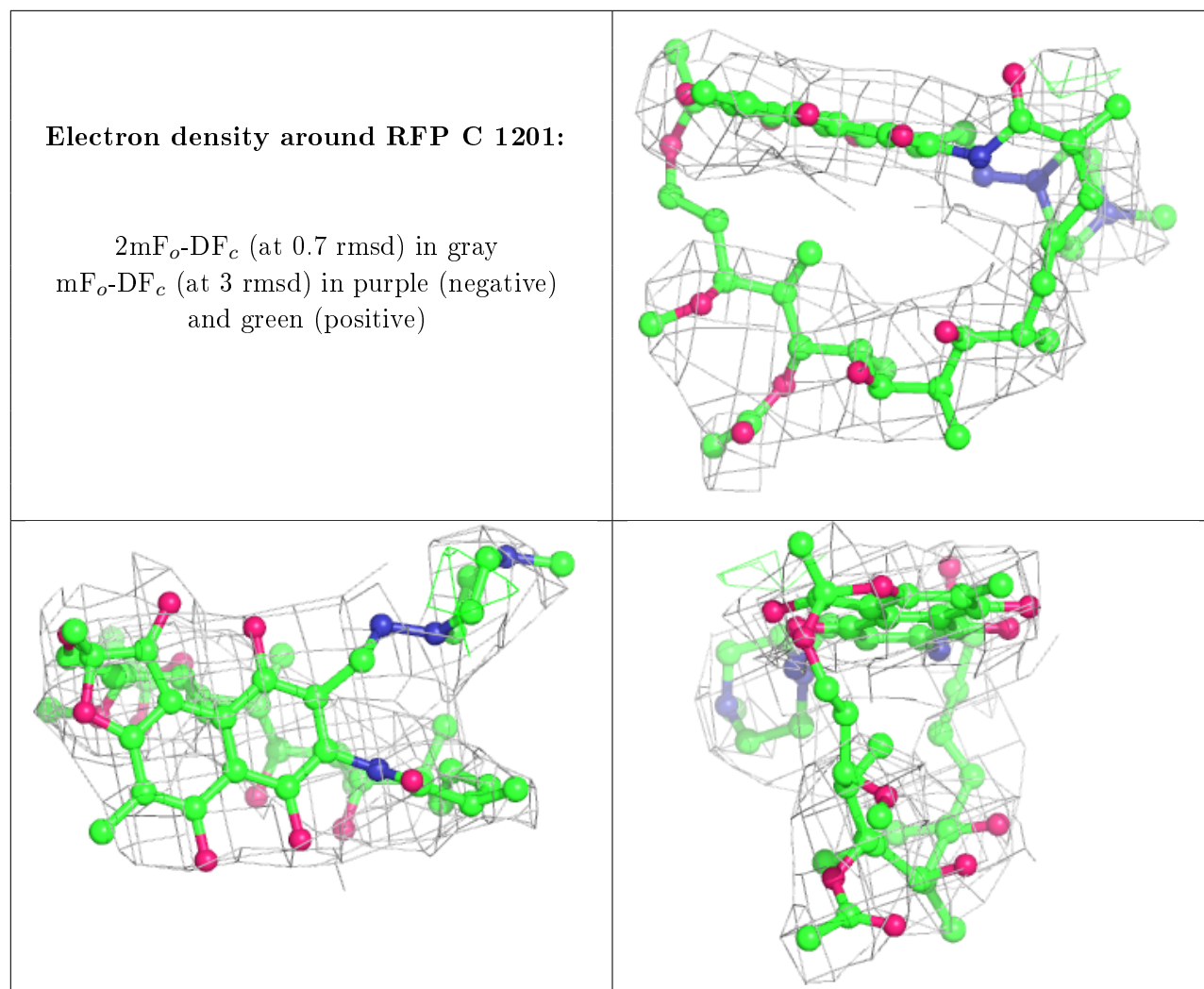
6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

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 [\(i\)](#)

Unable to reproduce the depositors R factor - this section is therefore empty.