



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 24, 2020 – 04:35 am BST

PDB ID : 5UHF  
Title : Crystal structure of Mycobacterium tuberculosis transcription initiation complex in complex with D-IX336  
Authors : Lin, W.; Das, K.; Feng, Y.; Ebright, R.H.  
Deposited on : 2017-01-11  
Resolution : 4.34 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  
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

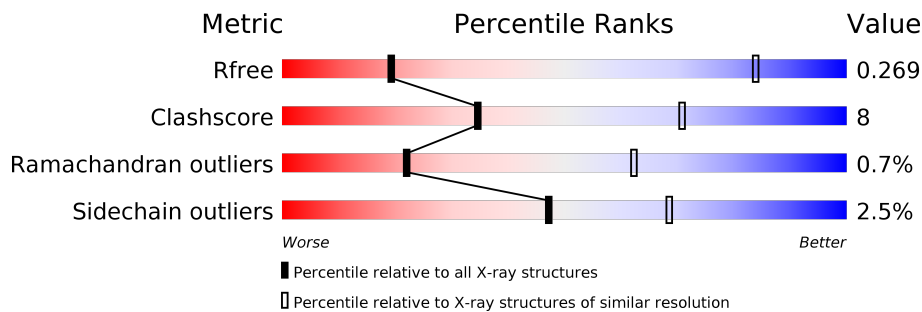
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 4.34 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.



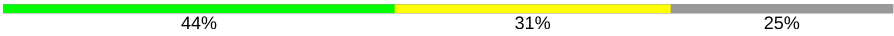
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1018 (4.84-3.80)
Clashscore	141614	1081 (4.84-3.80)
Ramachandran outliers	138981	1033 (4.84-3.80)
Sidechain outliers	138945	1016 (4.84-3.80)

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  $\geq 3$ , 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  $\leq 5\%$

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% 31% 25%

## 2 Entry composition [i](#)

There are 10 unique types of molecules in this entry. The entry contains 25949 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	225	Total	C	N	O	S	0	0	0
			1706	1075	289	340	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	322	Total	C	N	O	S	0	0	0
			2555	1589	461	496	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\*CP\*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

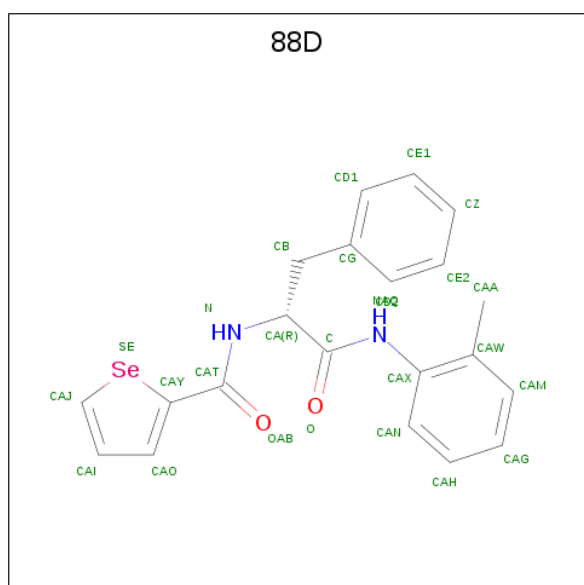
- Molecule 8 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

- Molecule 9 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

- Molecule 10 is N-(2-methylphenyl)-N-alpha-(selenophene-2-carbonyl)-D-phenylalaninamide (three-letter code: 88D) (formula: C<sub>21</sub>H<sub>20</sub>N<sub>2</sub>O<sub>2</sub>Se).

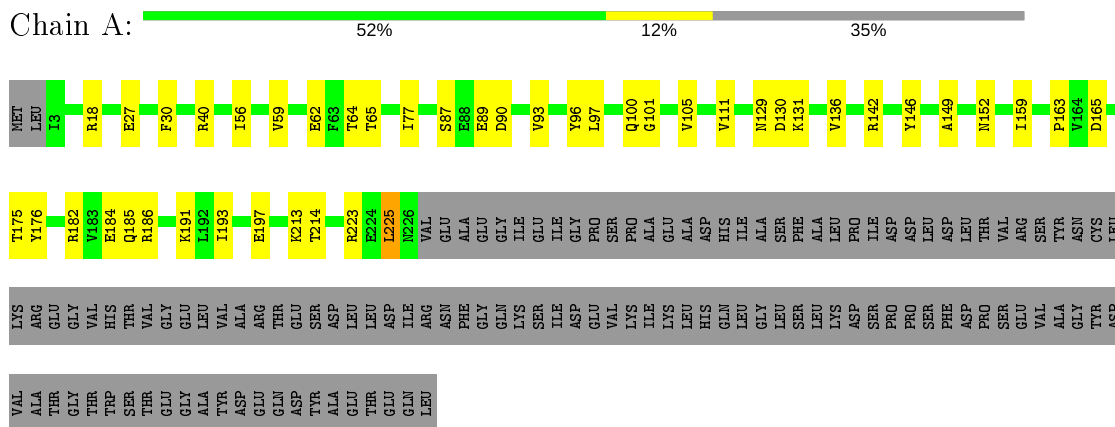


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	Se		
10	D	1	26	21	2	2	1	0	0

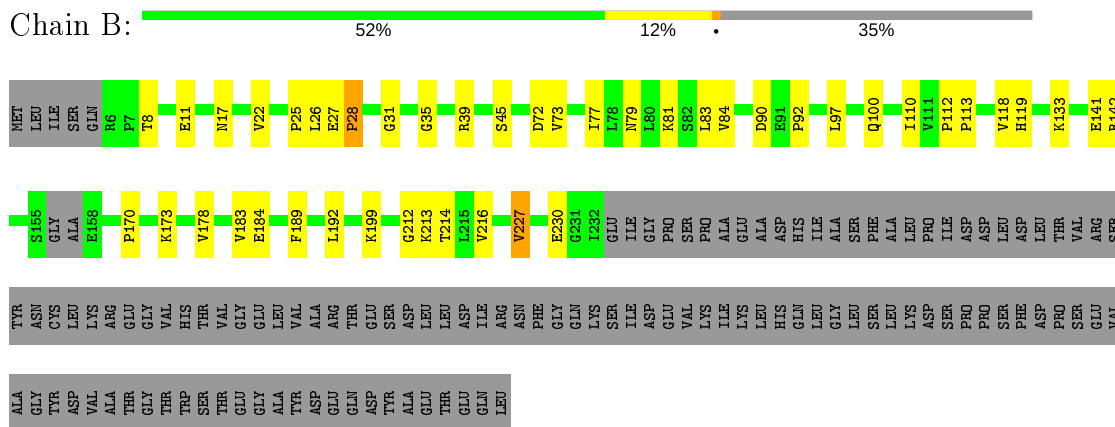
### 3 Residue-property plots [i](#)

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

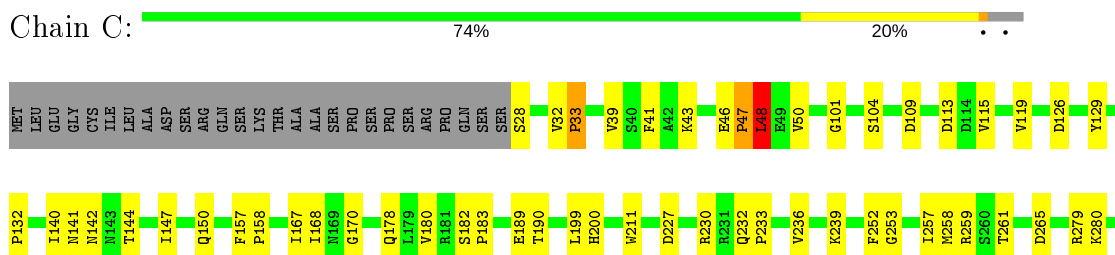
- Molecule 1: DNA-directed RNA polymerase subunit alpha

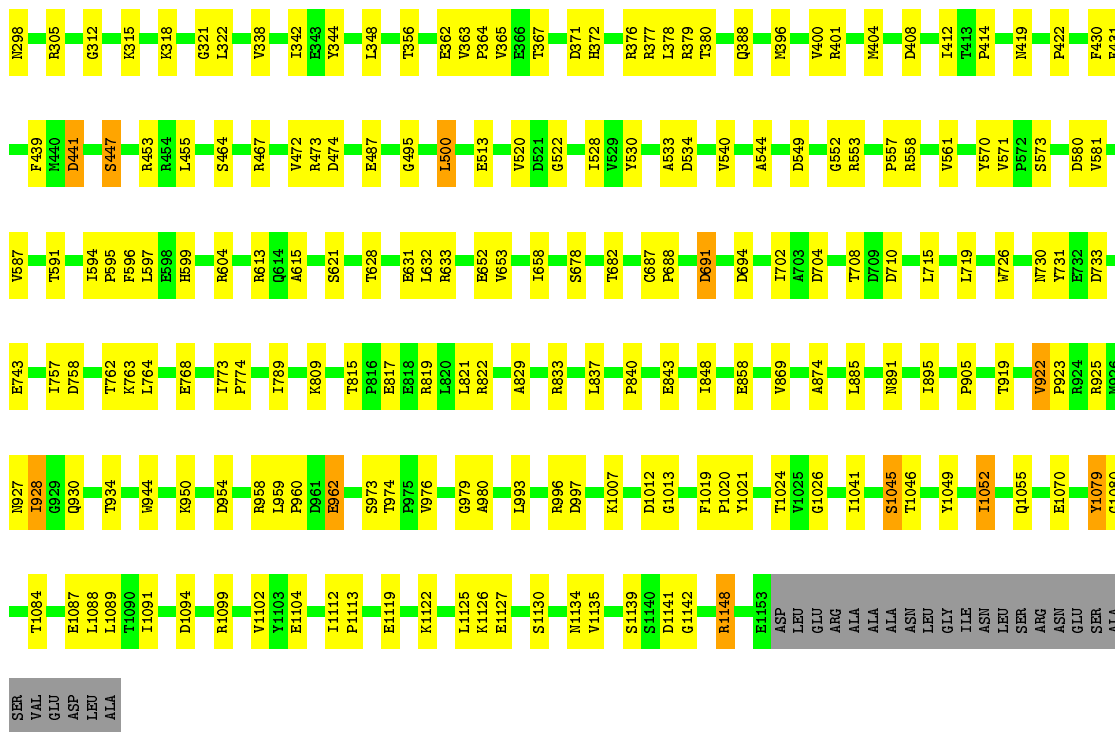


- Molecule 1: DNA-directed RNA polymerase subunit alpha

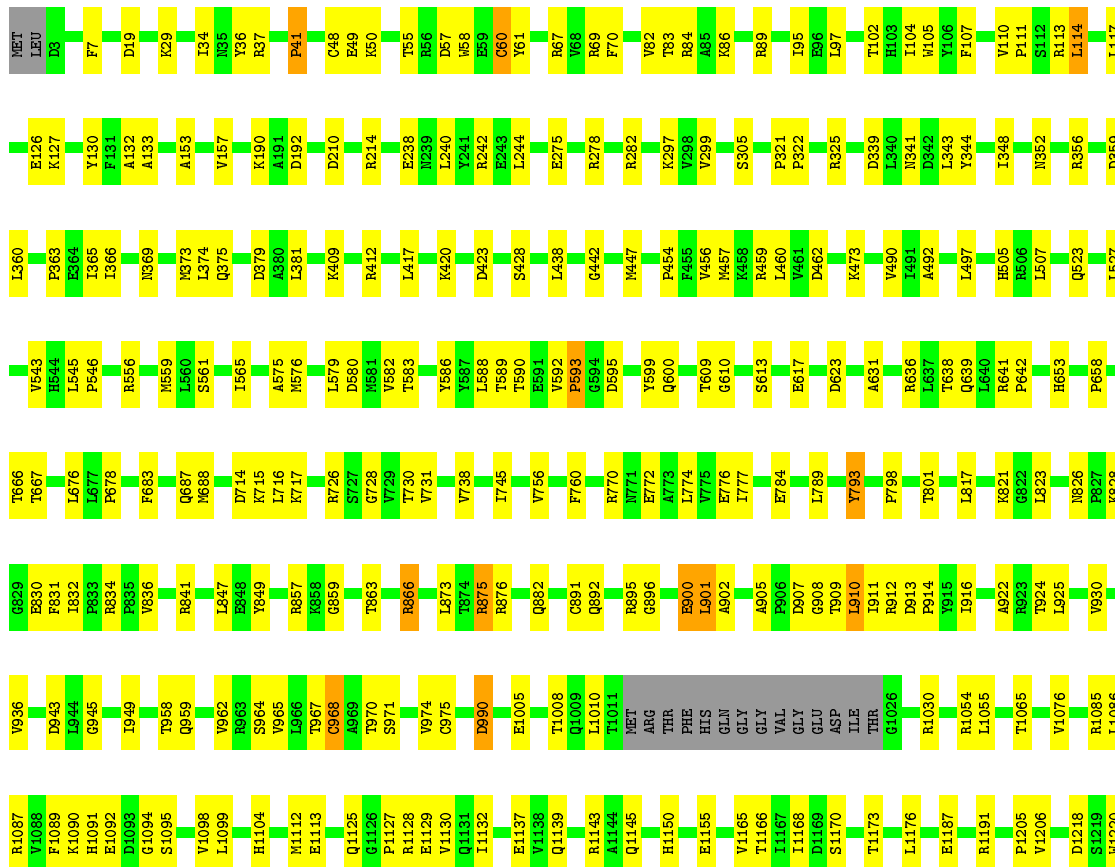


- Molecule 2: DNA-directed RNA polymerase subunit beta





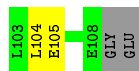
• Molecule 3: DNA-directed RNA polymerase subunit beta'







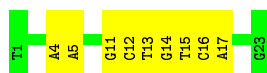
- Molecule 4: DNA-directed RNA polymerase subunit omega



- Molecule 5: RNA polymerase sigma factor SigA



- Molecule 6: 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')



- Molecule 7: DNA (5'-D(\*CP\*AP\*TP\*CP\*CP\*GP\*TP\*GP\*AP\*GP\*TP\*CP\*CP\*AP\*GP\*G)-3')



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	152.37Å 163.46Å 197.38Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.50 – 4.34 49.80 – 4.34	Depositor EDS
% Data completeness (in resolution range)	71.6 (48.50-4.34) 86.3 (49.80-4.34)	Depositor EDS
$R_{merge}$	0.20	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.58 (at 4.29Å)	Xtrriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.198 , 0.267 0.204 , 0.269	Depositor DCC
$R_{free}$ test set	2004 reflections (7.01%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	52.4	Xtrriage
Anisotropy	0.747	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.24 , 32.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.81	EDS
Total number of atoms	25949	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	68.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.34% 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: ZN, MG, 88D

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	0.22	0/1730	0.44	0/2354
1	B	0.22	0/1731	0.44	0/2356
2	C	0.23	0/8873	0.42	1/12031 (0.0%)
3	D	0.24	0/10052	0.42	0/13591
4	E	0.23	0/650	0.41	0/886
5	F	0.24	0/2585	0.41	0/3485
6	H	0.56	0/535	0.89	0/826
7	G	0.58	0/269	0.90	0/413
All	All	0.25	0/26425	0.45	1/35942 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
2	C	48	LEU	CA-CB-CG	5.30	127.49	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	25	0
1	B	1706	0	1730	28	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	C	8714	0	8636	160	0
3	D	9887	0	9943	181	0
4	E	637	0	635	14	0
5	F	2555	0	2579	56	0
6	H	476	0	261	14	0
7	G	241	0	137	4	0
8	D	2	0	0	0	0
9	D	1	0	0	0	0
10	D	26	0	0	1	0
All	All	25949	0	25662	428	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 428 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:891:CYS:SG	3:D:975:CYS:HB3	2.07	0.95
3:D:902:ALA:HB3	3:D:958:THR:HG22	1.53	0.90
6:H:16:DC:O2	7:G:12:DG:N2	2.06	0.89
3:D:900:GLU:OE1	3:D:959:GLN:NE2	2.06	0.87
3:D:901:LEU:O	3:D:901:LEU:HD22	1.85	0.77

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%)	205 (92%)	16 (7%)	1 (0%)	29 68
1	B	221/347 (64%)	201 (91%)	17 (8%)	3 (1%)	11 47

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	C	1124/1178 (95%)	1042 (93%)	70 (6%)	12 (1%)	14	52
3	D	1261/1316 (96%)	1187 (94%)	67 (5%)	7 (1%)	25	65
4	E	79/110 (72%)	75 (95%)	4 (5%)	0	100	100
5	F	320/528 (61%)	301 (94%)	18 (6%)	1 (0%)	41	76
All	All	3227/3826 (84%)	3011 (93%)	192 (6%)	24 (1%)	22	62

5 of 24 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	227	VAL
3	D	593	PRO
3	D	678	PRO
3	D	971	SER
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%)	187 (97%)	5 (3%)	46	67
1	B	192/297 (65%)	191 (100%)	1 (0%)	88	93
2	C	948/998 (95%)	928 (98%)	20 (2%)	53	72
3	D	1048/1095 (96%)	1018 (97%)	30 (3%)	42	64
4	E	68/90 (76%)	66 (97%)	2 (3%)	42	64
5	F	271/427 (64%)	261 (96%)	10 (4%)	34	59
All	All	2719/3204 (85%)	2651 (98%)	68 (2%)	47	68

5 of 68 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	D	105	TRP
3	D	580	ASP

Continued on next page...

*Continued from previous page...*

Mol	Chain	Res	Type
5	F	367	ARG
3	D	110	VAL
3	D	297	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
2	C	141	ASN
3	D	375	GLN
3	D	510	GLN
5	F	415	GLN
5	F	505	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 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	$\# Z  > 2$
10	88D	D	1404	-	25,28,28	2.88	5 (20%)	31,37,37	0.87	1 (3%)

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
10	88D	D	1404	-	-	0/17/20/20	0/3/3/3

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
10	D	1404	88D	CAY-CAT	-10.60	1.33	1.49
10	D	1404	88D	CAA-CAW	-5.85	1.39	1.51
10	D	1404	88D	CB-CG	-4.67	1.40	1.51
10	D	1404	88D	CAO-CAY	-4.29	1.34	1.38
10	D	1404	88D	CAX-NAQ	-3.92	1.34	1.41

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
10	D	1404	88D	CAX-NAQ-C	-2.67	119.61	126.92

There are no chirality outliers.

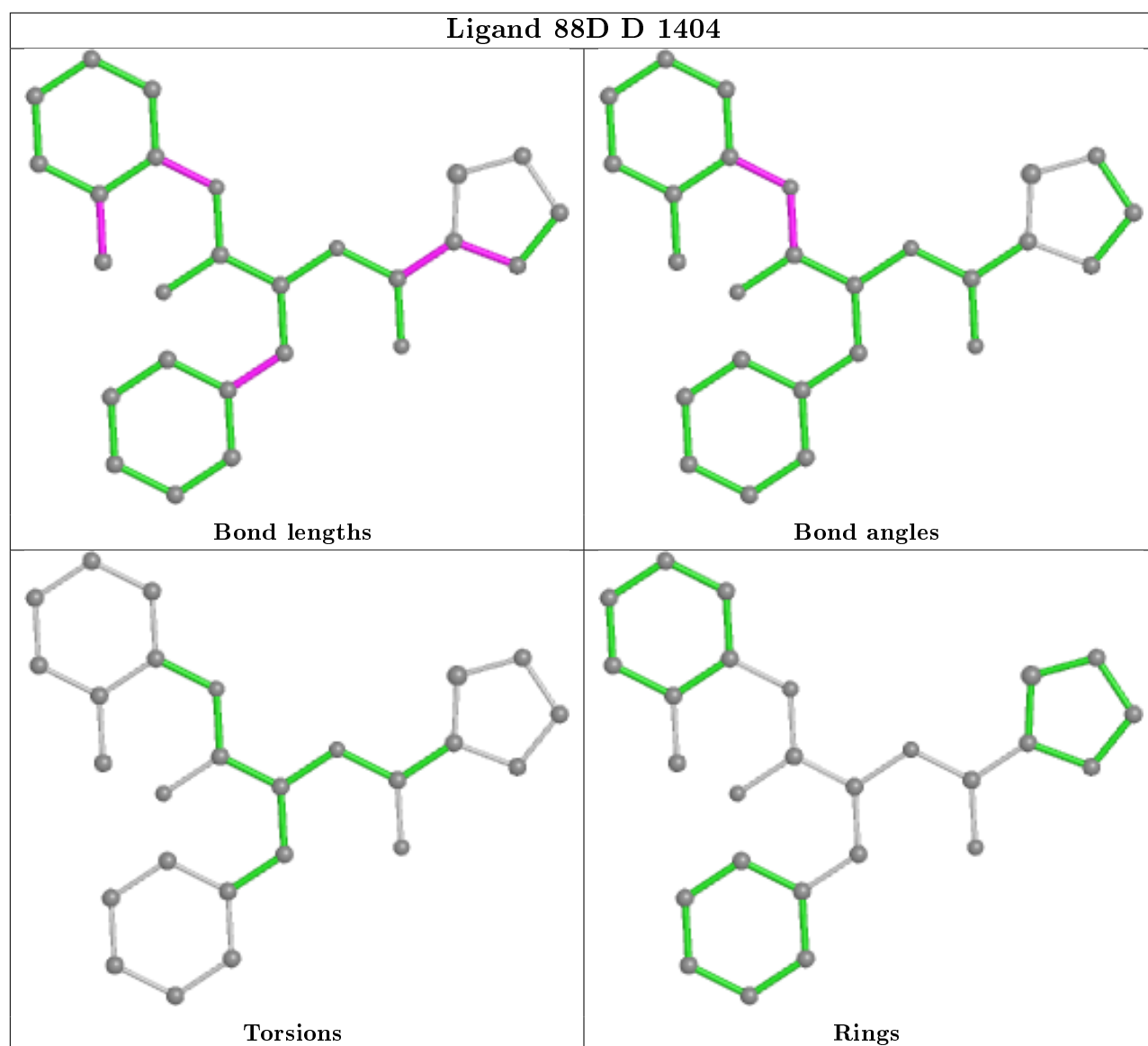
There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
10	D	1404	88D	1	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.

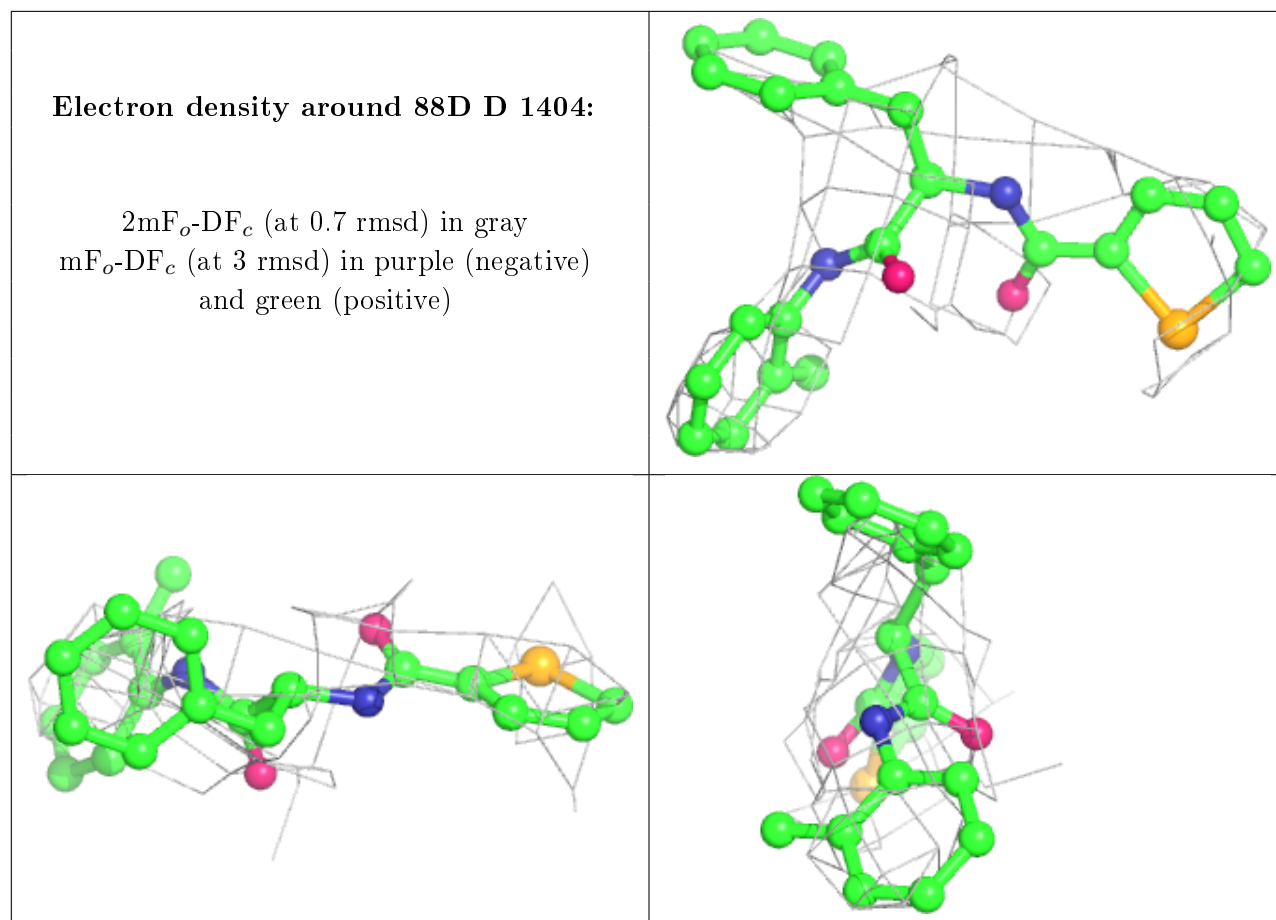
### 6.3 Carbohydrates

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 depositor's R factor - this section is therefore empty.