



wwPDB X-ray Structure Validation Summary Report ⓘ

May 19, 2020 – 07:29 am BST

PDB ID : 4IKF
Title : PFV intasome with inhibitor MB-76
Authors : Taltynov, O.; Demeulemeester, J.; Desimmie, B.A.; Suchaud, V.; Billamboz, M.; Lion, C.; Bailly, F.; Debyser, Z.; Cotellet, P.; Christ, F.; Strelkov, S.V.
Deposited on : 2012-12-26
Resolution : 3.40 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

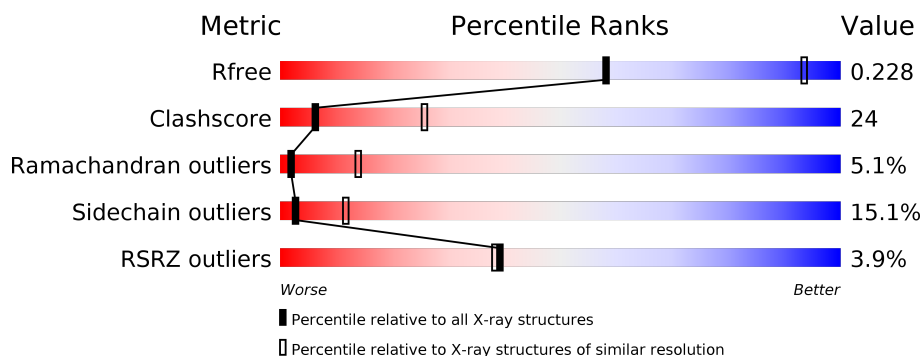
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.40 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 (3.48-3.32)
Clashscore	141614	1055 (3.48-3.32)
Ramachandran outliers	138981	1038 (3.48-3.32)
Sidechain outliers	138945	1038 (3.48-3.32)
RSRZ outliers	127900	2173 (3.50-3.30)

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 $\leq 5\%$. 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	395	<div> <div>3%</div> <div>43%</div> <div>43%</div> <div>8%</div> <div>7%</div> </div>
1	B	395	<div> <div>3%</div> <div>22%</div> <div>18%</div> <div>5%</div> <div>53%</div> </div>
2	C	19	<div> <div>5%</div> <div>32%</div> <div>63%</div> </div>
3	D	17	<div> <div>6%</div> <div>53%</div> <div>41%</div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
6	NH4	A	403	-	-	-	X

2 Entry composition

There are 10 unique types of molecules in this entry. The entry contains 5177 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 Integrase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	368	Total	C	N	O	S	0	1	0
			2934	1881	517	532	4			
1	B	184	Total	C	N	O	S	0	0	0
			1441	935	234	271	1			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	EXPRESSION TAG	UNP P14350
A	-1	PRO	-	EXPRESSION TAG	UNP P14350
A	0	GLY	-	EXPRESSION TAG	UNP P14350
A	217	SER	GLY	CONFLICT	UNP P14350
A	218	GLY	SER	CONFLICT	UNP P14350
B	-2	GLY	-	EXPRESSION TAG	UNP P14350
B	-1	PRO	-	EXPRESSION TAG	UNP P14350
B	0	GLY	-	EXPRESSION TAG	UNP P14350
B	217	SER	GLY	CONFLICT	UNP P14350
B	218	GLY	SER	CONFLICT	UNP P14350

- Molecule 2 is a DNA chain called 5'-D(*AP*TP*TP*GP*TP*CP*AP*TP*GP*GP*AP*A P*TP*TP*TP*CP*GP*CP*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	19	Total	C	N	O	P	0	0	0
			387	187	68	114	18			

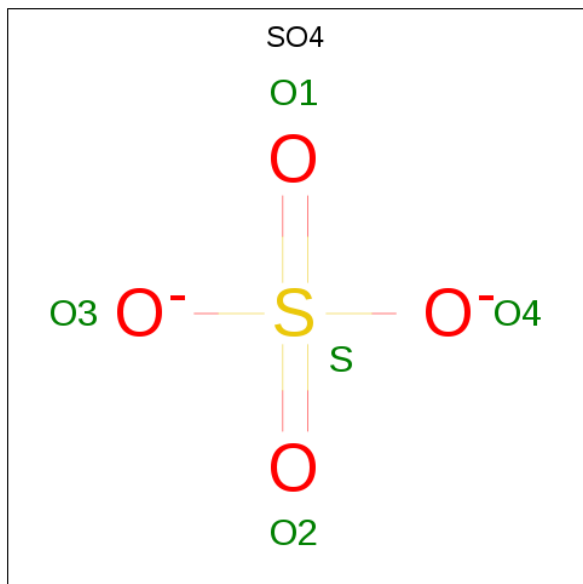
- Molecule 3 is a DNA chain called 5'-D(*TP*GP*CP*GP*AP*AP*AP*TP*TP*CP*CP*A P*TP*GP*AP*CP*A)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	17	Total	C	N	O	P	0	0	0
			345	166	65	98	16			

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

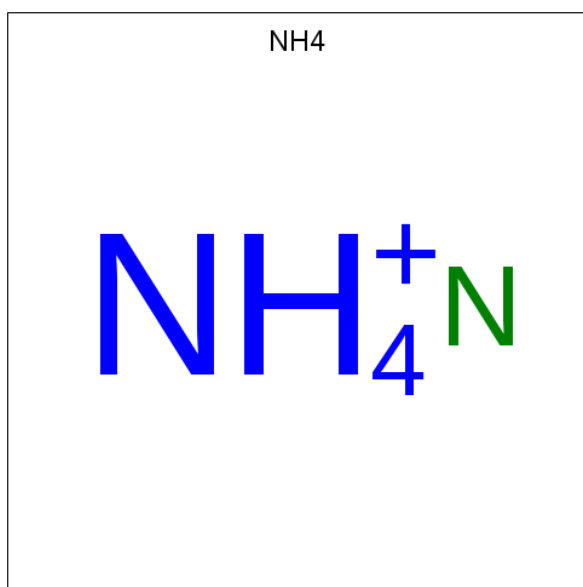
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Zn	0	0
			1	1		

- Molecule 5 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	O	S	0	0
			5	4	1		
5	B	1	Total	O	S	0	0
			5	4	1		

- Molecule 6 is AMMONIUM ION (three-letter code: NH4) (formula: H₄N).

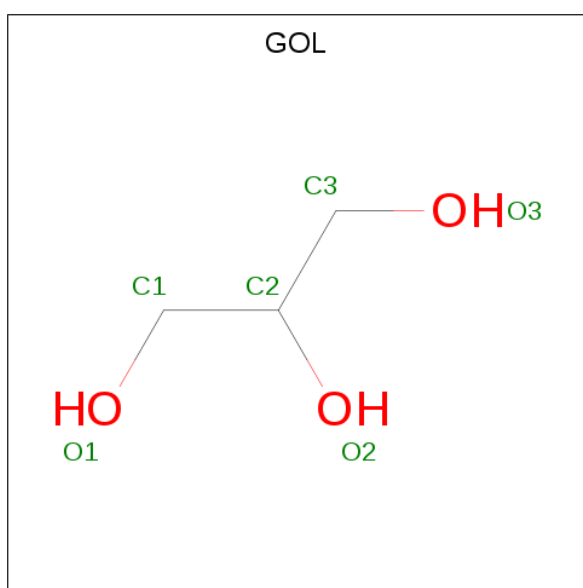


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	1	Total N 1 1	0	0

- Molecule 7 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

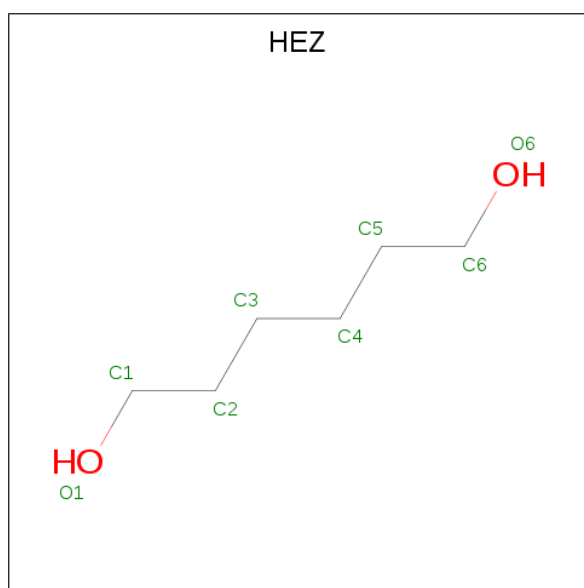
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	A	2	Total Mg 2 2	0	0

- Molecule 8 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).



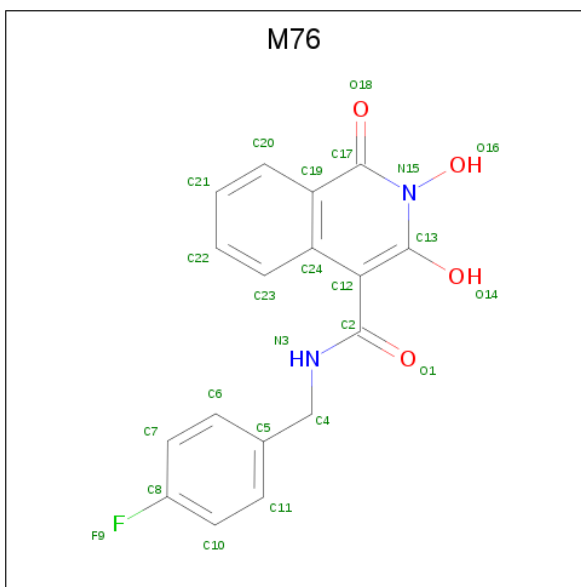
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
8	A	1	Total	C	O	0	0
			6	3	3		
8	A	1	Total	C	O	0	0
			6	3	3		
8	A	1	Total	C	O	0	0
			6	3	3		
8	B	1	Total	C	O	0	0
			6	3	3		

- Molecule 9 is HEXANE-1,6-DIOL (three-letter code: HEZ) (formula: $C_6H_{14}O_2$).

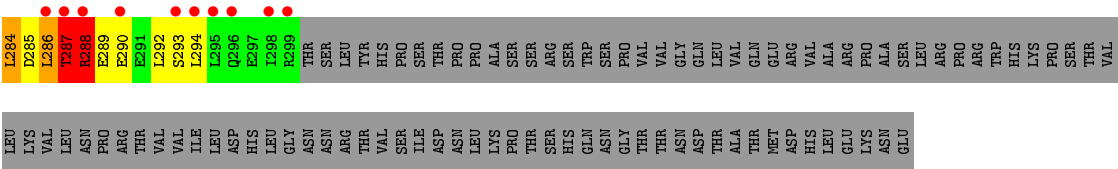


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
9	B	1	Total	C	O	0	0
			8	6	2		

- Molecule 10 is N-(4-fluorobenzyl)-2,3-dihydroxy-1-oxo-1,2-dihydroisoquinoline-4-carboxamide (three-letter code: M76) (formula: $C_{17}H_{13}FN_2O_4$).



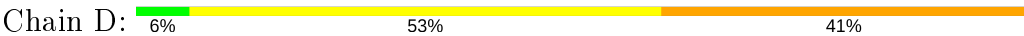
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	F	N	O		
10	D	1	24	17	1	2	4	0	0



● Molecule 2: 5'-D(*AP*TP*TP*GP*TP*CP*AP*TP*GP*GP*AP*AP*TP*TP*TP*CP*GP*CP*A)-3'



● Molecule 3: 5'-D(*TP*GP*CP*GP*AP*AP*AP*TP*TP*CP*CP*AP*TP*GP*AP*CP*A)-3',



4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α , β , γ	159.07Å 159.07Å 123.82Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.85 – 3.40 48.85 – 3.40	Depositor EDS
% Data completeness (in resolution range)	93.8 (48.85-3.40) 93.8 (48.85-3.40)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.41 (at 3.40Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.2_869), REFMAC (CCP4)	Depositor
R, R_{free}	0.180 , 0.230 0.179 , 0.228	Depositor DCC
R_{free} test set	1080 reflections (5.13%)	wwPDB-VP
Wilson B-factor (Å ²)	114.1	Xtriage
Anisotropy	0.002	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 94.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	5177	wwPDB-VP
Average B, all atoms (Å ²)	115.0	wwPDB-VP

5 Model quality ⓘ

5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, MG, M76, ZN, HEZ, NH4, SO4

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	0.82	1/3013 (0.0%)	0.95	2/4109 (0.0%)
1	B	0.74	0/1481	0.84	0/2025
2	C	1.59	3/433 (0.7%)	2.73	46/667 (6.9%)
3	D	1.32	1/387 (0.3%)	2.24	19/595 (3.2%)
All	All	0.93	5/5314 (0.1%)	1.33	67/7396 (0.9%)

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	6	DC	C3'-O3'	-8.32	1.33	1.44
2	C	7	DA				

5.2 Too-close contacts ⓘ

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	2934	0	2978	157	1
1	B	1441	0	1421	70	0
2	C	387	0	218	12	0
3	D	345	0	193	14	1
4	A	1	0	0	0	0
5	A	5	0	0	0	0
5	B	5	0	0	0	0
6	A	1	0	0	0	0
7	A	2	0	0	0	0
8	A	18	0	24	1	0
8	B	6	0	8	1	0
9	B	8	0	14	0	0
10	D	24	0	11	6	0
All	All	5177	0	4867	243	1

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 24.

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	366/395 (93%)	301 (82%)	50 (14%)	15 (4%)	3	18
1	B	182/395 (46%)	140 (77%)	29 (16%)	13 (7%)	1	7
All	All	548/790 (69%)	441 (80%)	79 (14%)	28 (5%)	2	14

5 of 28 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	282	ASP
1	A	328	ALA
1	B	216	SER
1	B	286	LEU
1	B	288	ARG

5.3.2 Protein sidechains

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score

Continued from previous page...

Mol	Chain	Res	Type
1	A	360	ASN
1	B	274	SER
1	A	312	SER
1	A	341	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	357	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 12 ligands modelled in this entry, 3 are monoatomic and 1 is modelled with single atom - leaving 8 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

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
8	GOL	A	408	-	5,5,5	0.47	0	5,5,5	0.59	0
10	M76	D	101	7	24,26,26	1.60	4 (16%)	27,37,37	1.09	3 (11%)
8	GOL	A	407	-	5,5,5	0.50	0	5,5,5	1.02	0
5	SO4	A	402	-	4,4,4	0.22	0	6,6,6	0.29	0
9	HEZ	B	403	-	7,7,7	0.34	0	6,6,6	1.15	0

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	GOL	B	401	-	-	2/4/4/4	-

Continued from previous page...

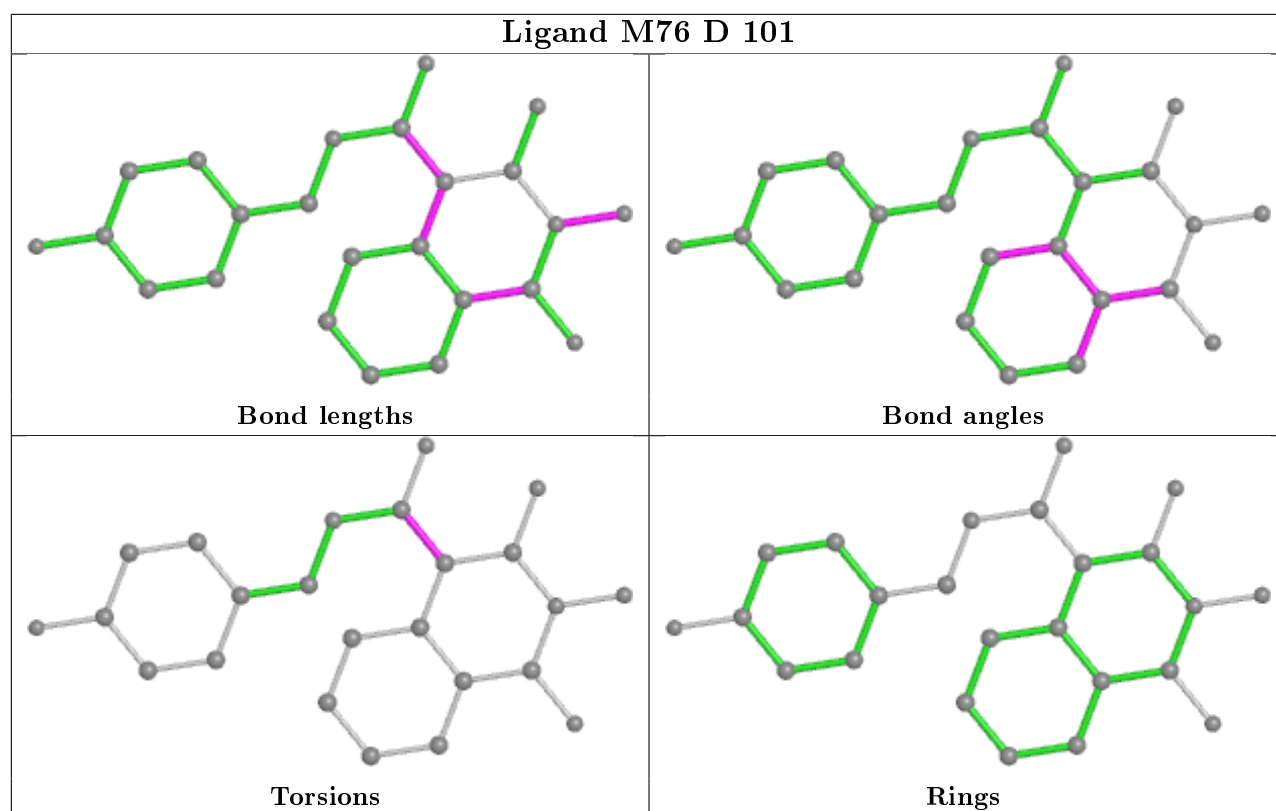
Mol	Chain	Res	Type	Atoms
8	A	406	GOL	O2-C2-C3-O3
10	D	101	M76	C24-C12-C2-O1
10	D	101	M76	C24-C12-C2-N3

There are no ring outliers.

3 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
8	B	401	GOL	1	0
10	D	101	M76	6	0
8	A	407	GOL	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 [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	368/395 (93%)	0.14	11 (2%) 50 49	80, 104, 154, 194	1 (0%)
1	B	184/395 (46%)	0.10	11 (5%) 21 23	90, 120, 207, 230	0
2	C	19/19 (100%)	-0.12	1 (5%) 26 27	85, 107, 136, 158	0
3	D	17/17 (100%)	-0.14	0 100 100	91, 103, 142, 166	0
All	All	588/826 (71%)	0.11	23 (3%) 39 38	80, 107, 178, 230	1 (0%)

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	8	LEU	7.5
1	B	298	ILE	4.4
1	B	299	ARG	4.2
1	A	37	VAL	3.7
1	A	36	LYS	3.4

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](#)

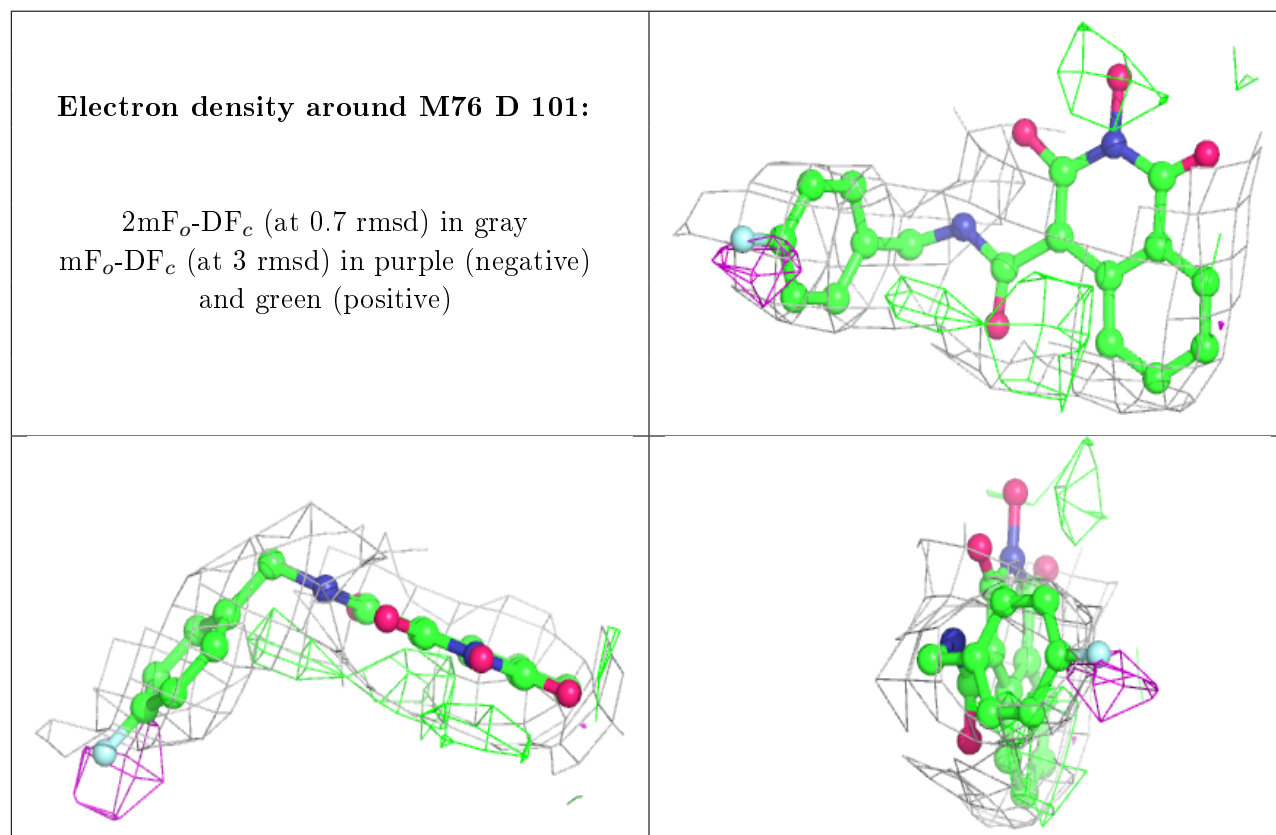
There are no carbohydrates in this entry.

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(\AA^2)	Q<0.9
6	NH4	A	403	1/1	0.49	0.46	57,57,57,57	0
8	GOL	A	408	6/6	0.70	0.36	124,132,143,151	0
9	HEZ	B	403	8/8	0.72	0.40	124,132,141,142	0
8	GOL	B	401	6/6	0.82	0.98	116,124,135,142	0
8	GOL	A	406	6/6	0.89	0.15	109,124,132,133	0
5	SO4	B	402	5/5	0.91	0.12	141,147,160,169	0
5	SO4	A	402	5/5	0.93	0.17	135,136,148,156	0
8	GOL	A	407	6/6	0.93	0.15	109,116,126,151	0
10	M76	D	101	24/24	0.95	0.22	64,81,90,100	0
7	MG	A	404	1/1	0.96	0.24	95,95,95,95	0
4	ZN	A	401	1/1	0.99	0.17	94,94,94,94	0
7	MG	A	405	1/1	0.99	0.16	77,77,77,77	0

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.