



wwPDB X-ray Structure Validation Summary Report i

Oct 27, 2025 – 12:11 PM EDT

PDB ID : 9E20 / pdb_00009e20
Title : [iU:Hg2+:F] Mercury base pair with 2-thiothymidine and iodouracil
Authors : Vecchioni, S.; Lu, B.; Woloszyn, K.; Ohayon, Y.P.; Sha, R.
Deposited on : 2024-10-21
Resolution : 4.72 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 i 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](#) i) 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
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)
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) : 2.46

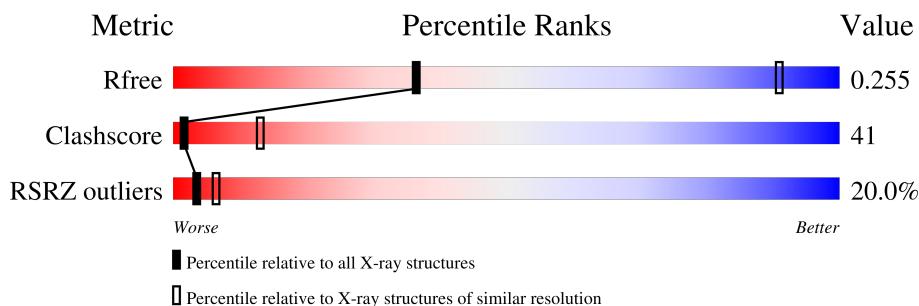
1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

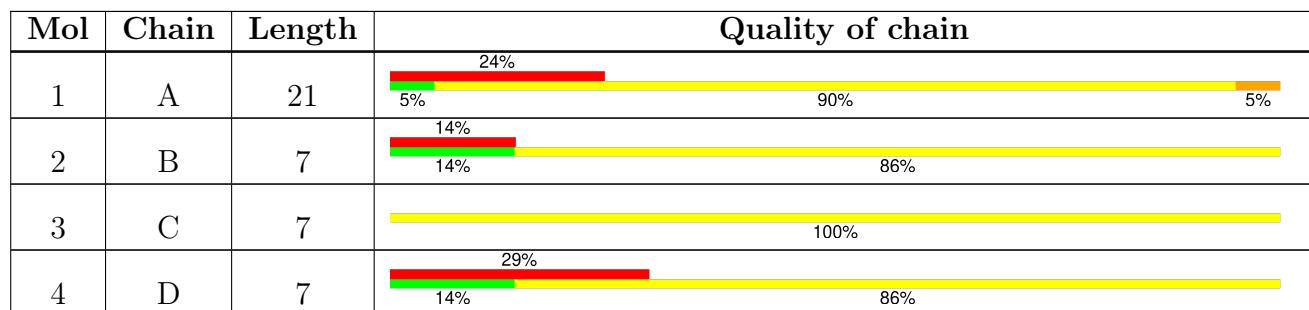
The reported resolution of this entry is 4.72 Å.

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}	164625	1101 (5.54-3.90)
Clashscore	180529	1023 (5.52-3.92)
RSRZ outliers	164620	1097 (5.54-3.90)

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.



2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 860 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 DNA chain called DNA (5'-D(*GP*AP*GP*CP*AP*GP*CP*CP*TP*GP*TP*(5IU)P*TP*GP*GP*AP*CP*AP*TP*CP*A)-3').

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
1	A	21	Total C I N O P 429 204 1 80 124 20	0	0	0

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

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
2	B	7	Total C N O P S 140 67 26 39 7 1	0	0	0

- Molecule 3 is a DNA chain called DNA (5'-D(P*GP*GP*CP*TP*GP*CP*T)-3').

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
3	C	7	Total C N O P 144 68 25 44 7	0	0	0

- Molecule 4 is a DNA chain called DNA (5'-D(P*CP*TP*GP*AP*TP*GP*T)-3').

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
4	D	7	Total C N O P 145 69 24 45 7	0	0	0

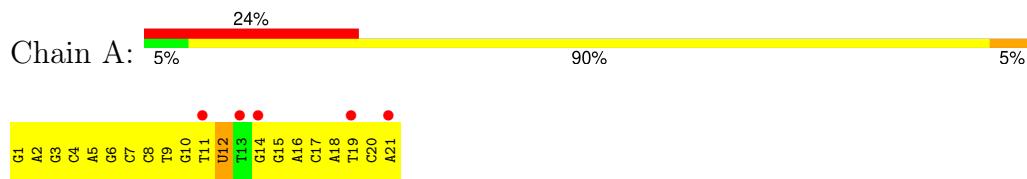
- Molecule 5 is MERCURY (II) ION (CCD ID: HG) (formula: Hg) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total Hg 1 1	0	0
5	B	1	Total Hg 1 1	0	0

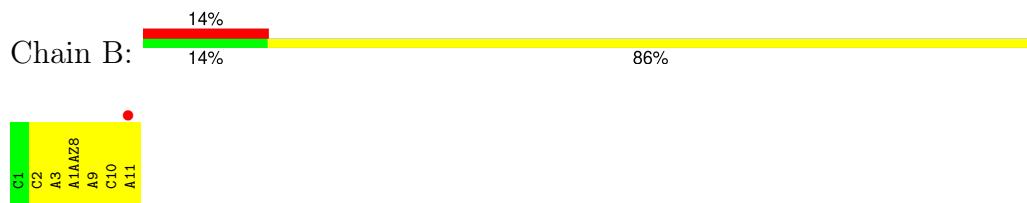
3 Residue-property plots [\(i\)](#)

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 (5'-D(*GP*AP*GP*CP*AP*GP*CP*CP*TP*GP*TP*(5IU)P*TP*GP*GP*AP*CP*AP*TP*CP*A)-3')



- Molecule 2: DNA (5'-D(P*CP*CP*AP*(DF)P*AP*CP*A)-3')



- Molecule 3: DNA (5'-D(P*GP*GP*CP*TP*GP*CP*T)-3')



- Molecule 4: DNA (5'-D(P*CP*TP*GP*AP*TP*GP*T)-3')



4 Data and refinement statistics (i)

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, α , β , γ	105.00 Å 105.00 Å 96.01 Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	42.45 – 4.72 42.45 – 4.72	Depositor EDS
% Data completeness (in resolution range)	69.3 (42.45-4.72) 62.5 (42.45-4.72)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	0.19 (at 4.65 Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R , R_{free}	0.189, 0.253 0.194, 0.255	Depositor DCC
R_{free} test set	84 reflections (2.44%)	wwPDB-VP
Wilson B-factor (Å ²)	191.5	Xtriage
Anisotropy	0.356	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.14, 999.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.44$, $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	0.074 for h,-h-k,-l	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	860	wwPDB-VP
Average B, all atoms (Å ²)	236.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.74% 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: 5IU, HG, A1AAZ

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	0.67	0/458	0.83	0/703
2	B	0.62	0/133	0.71	0/199
3	C	0.68	0/160	0.78	0/245
4	D	0.63	0/161	0.86	0/245
All	All	0.66	0/912	0.81	0/1392

There are no bond length outliers.

There are no bond angle outliers.

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	429	0	235	35	0
2	B	140	0	67	7	0
3	C	144	0	80	9	0
4	D	145	0	81	6	0
5	A	1	0	0	0	0
5	B	1	0	0	0	0
All	All	860	0	463	54	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 41.

The worst 5 of 54 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)
1:A:12:5IU:C4'	1:A:12:5IU:O4'	1.63	1.22
1:A:15:DG:H2'	1:A:16:DA:C8	2.15	0.81
4:D:2:DT:H2'	4:D:3:DG:C8	2.27	0.69
1:A:14:DG:H8	1:A:14:DG:H5"	1.62	0.64
2:B:2:DC:H2"	2:B:3:DA:N7	2.14	0.62

There are no symmetry-related clashes.

5.3 Torsion angles [\(i\)](#)

5.3.1 Protein backbone [\(i\)](#)

There are no protein molecules in this entry.

5.3.2 Protein sidechains [\(i\)](#)

There are no protein molecules in this entry.

5.3.3 RNA [\(i\)](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [\(i\)](#)

2 non-standard protein/DNA/RNA residues are 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
1	5IU	A	12	5,1	18,21,22	4.42	12 (66%)	25,30,33	2.38	11 (44%)
2	A1AAZ	B	8	2	17,21,22	4.59	11 (64%)	20,30,33	2.37	6 (30%)

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
1	5IU	A	12	5,1	-	3/7/21/22	0/2/2/2
2	A1AAZ	B	8	2	-	6/7/21/22	0/2/2/2

The worst 5 of 23 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	8	A1AAZ	C2-S1	8.75	1.82	1.67
1	A	12	5IU	O4'-C4'	8.38	1.63	1.45
2	B	8	A1AAZ	O4'-C4'	7.98	1.62	1.45
2	B	8	A1AAZ	O4-C4	7.76	1.38	1.23
2	B	8	A1AAZ	C3'-C4'	-6.97	1.34	1.53

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	8	A1AAZ	O4-C4-C5	-6.10	117.94	124.92
2	B	8	A1AAZ	C5-C4-N3	5.01	119.68	115.32
1	A	12	5IU	O4'-C1'-N1	4.40	115.67	107.86
1	A	12	5IU	C4-N3-C2	-4.37	121.61	127.34
2	B	8	A1AAZ	C5-C6-N1	-4.28	118.67	123.31

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	12	5IU	C4'-C5'-O5'-P
1	A	12	5IU	C3'-C4'-C5'-O5'
2	B	8	A1AAZ	C3'-C4'-C5'-O5'
2	B	8	A1AAZ	O4'-C4'-C5'-O5'
2	B	8	A1AAZ	C2'-C1'-N1-C6

There are no ring outliers.

1 monomer is involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	A	12	5IU	4	0

5.5 Carbohydrates [\(i\)](#)

There are no oligosaccharides in this entry.

5.6 Ligand geometry [\(i\)](#)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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	20/21 (95%)	1.49	5 (25%) 2 4	213, 240, 264, 268	0
2	B	6/7 (85%)	1.29	1 (16%) 5 8	199, 205, 212, 217	0
3	C	7/7 (100%)	0.99	0 100 100	223, 254, 264, 265	0
4	D	7/7 (100%)	1.54	2 (28%) 1 4	195, 223, 250, 260	0
All	All	40/42 (95%)	1.38	8 (20%) 3 6	195, 239, 264, 268	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	13	DT	5.2
1	A	14	DG	4.5
1	A	11	DT	4.2
4	D	6	DG	3.1
4	D	7	DT	2.5

6.2 Non-standard residues in protein, DNA, RNA chains [\(i\)](#)

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(Å ²)	Q<0.9
1	5IU	A	12	20/21	0.80	0.40	198,221,233,237	0
2	A1AAZ	B	8	20/21	0.89	0.18	207,227,246,251	0

6.3 Carbohydrates [\(i\)](#)

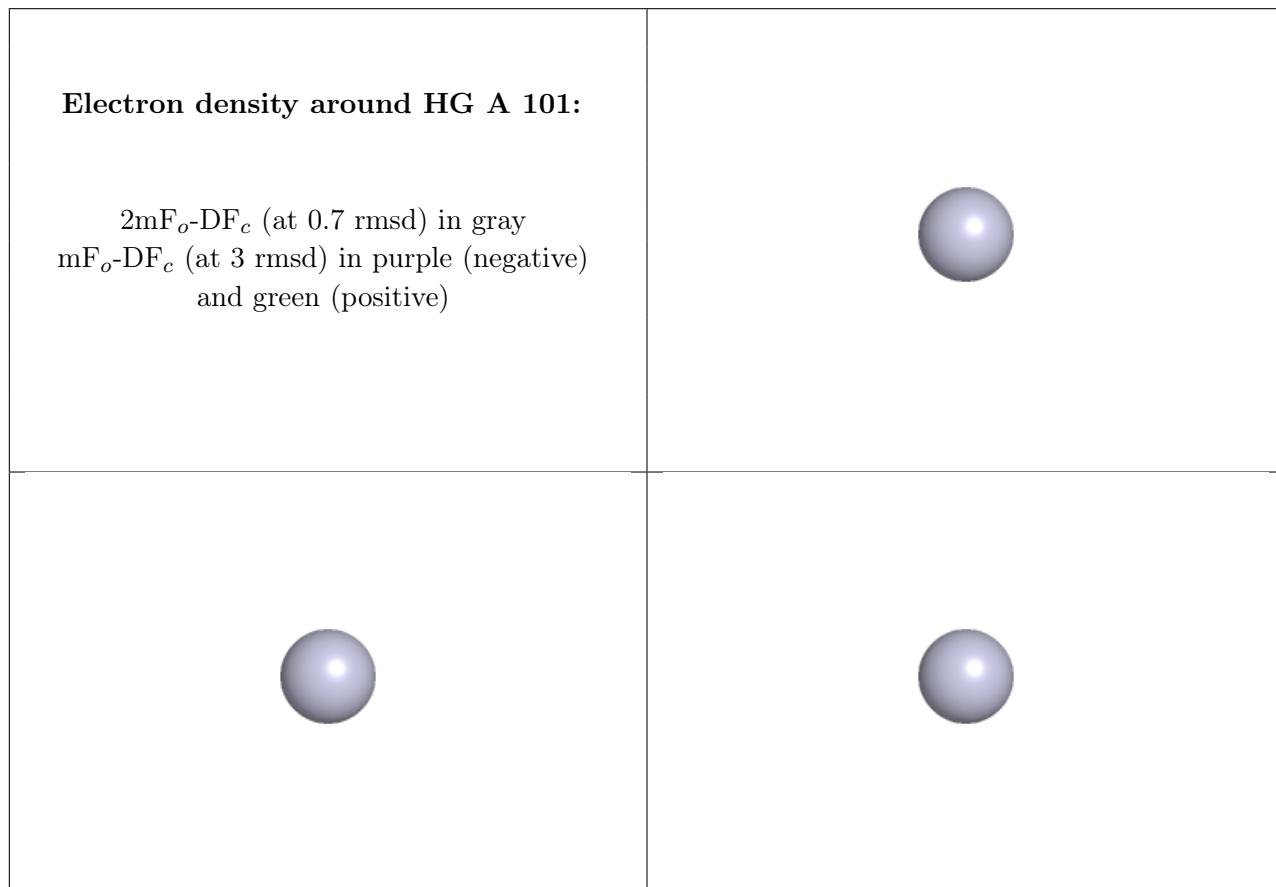
There are no oligosaccharides in this entry.

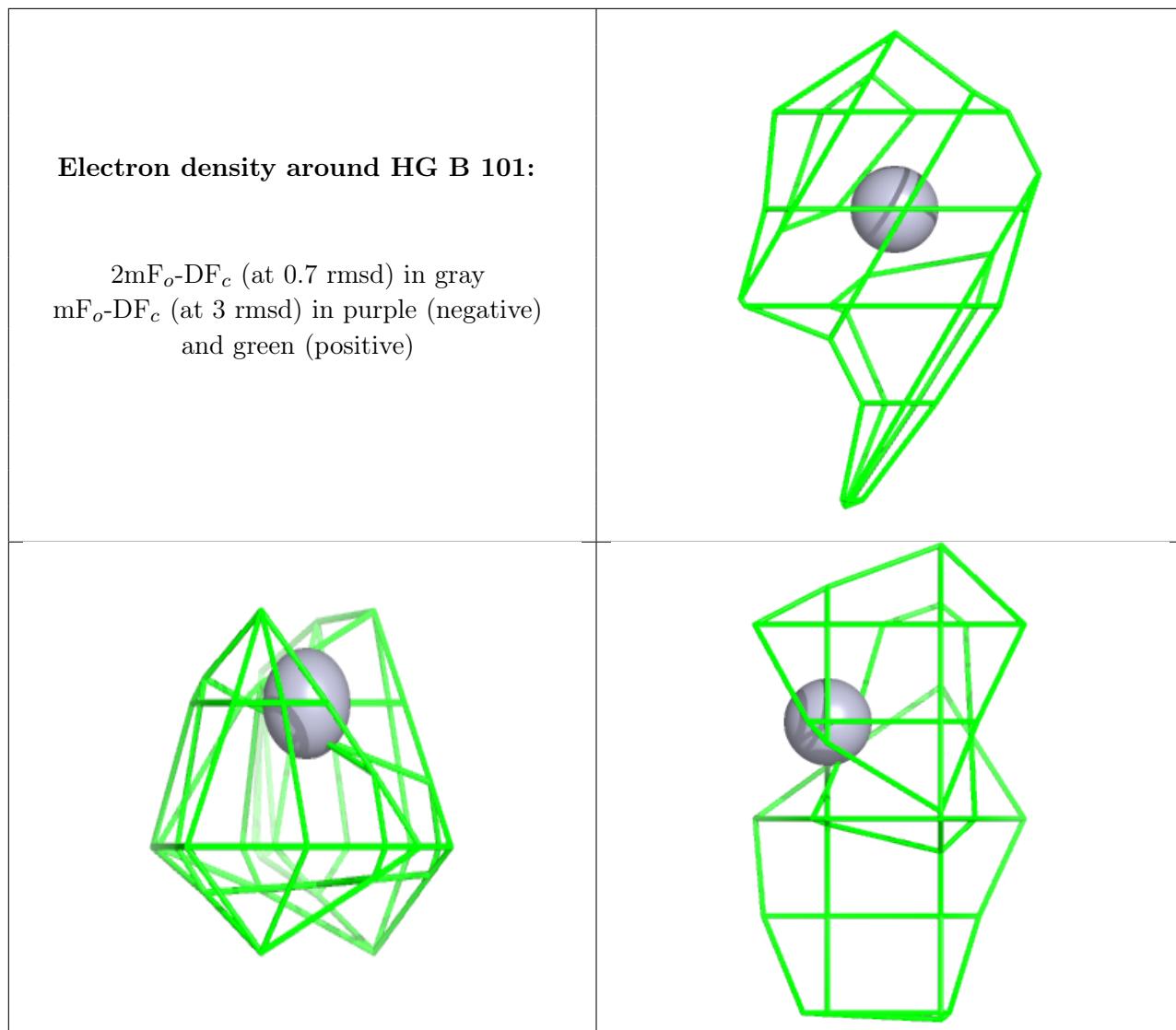
6.4 Ligands [\(i\)](#)

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(Å ²)	Q<0.9
5	HG	A	101	1/1	0.96	0.15	316,316,316,316	0
5	HG	B	101	1/1	0.98	0.16	239,239,239,239	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.