Full wwPDB X-ray Structure Validation Report ⓘ

Oct 24, 2022 – 02:30 pm BST

PDB ID	:	7PZ1
Title	:	Structure of the mouse 8-oxoguanine DNA Glycosylase mOGG1 in complex with ligand TH8535
Authors	:	Scaletti, E.R.; Helleday, T.; Stenmark, P.
Deposited on	:	2021-10-11
Resolution	:	2.45 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 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:

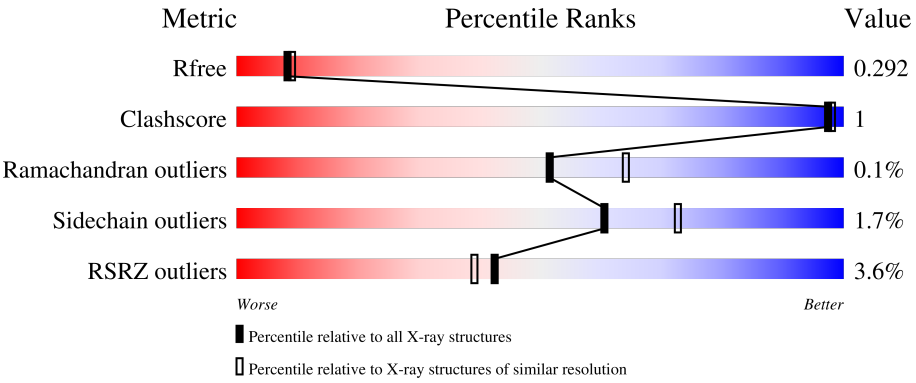
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.31.2
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.2

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 2.45 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	1544 (2.48-2.44)
Clashscore	141614	1613 (2.48-2.44)
Ramachandran outliers	138981	1598 (2.48-2.44)
Sidechain outliers	138945	1598 (2.48-2.44)
RSRZ outliers	127900	1523 (2.48-2.44)

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 <=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	AAA	318	2% 95%
1	BBB	318	4% 93%
1	CCC	318	5% 96%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 7205 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.

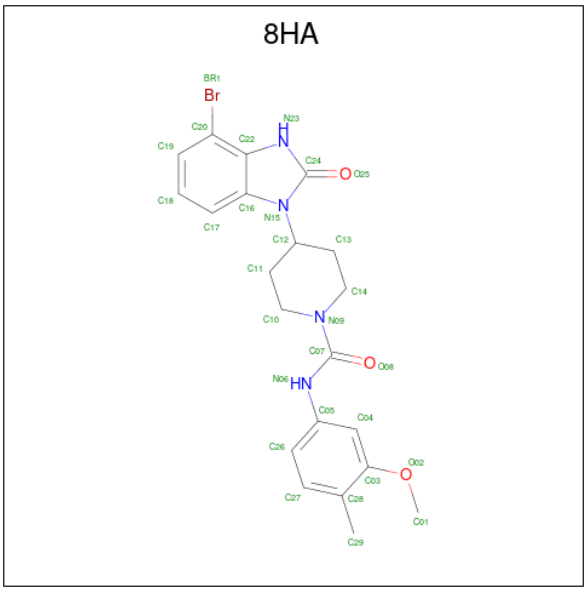
- Molecule 1 is a protein called N-glycosylase/DNA lyase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAA	313	Total	C	N	O	S	0	0	0
			2436	1555	435	435	11			
1	BBB	308	Total	C	N	O	S	0	0	0
			2365	1519	421	414	11			
1	CCC	312	Total	C	N	O	S	0	0	0
			2289	1469	408	402	10			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	8	GLY	-	expression tag	UNP O08760
AAA	10	HIS	SER	conflict	UNP O08760
BBB	8	GLY	-	expression tag	UNP O08760
BBB	10	HIS	SER	conflict	UNP O08760
CCC	8	GLY	-	expression tag	UNP O08760
CCC	10	HIS	SER	conflict	UNP O08760

- Molecule 2 is 4-(4-bromanyl-2-oxidanylidene-3 {H}-benzimidazol-1-yl)- {N}-(3-methoxy-4-methyl-phenyl)piperidine-1-carboxamide (three-letter code: 8HA) (formula: $C_{21}H_{23}BrN_4O_3$) (labeled as "Ligand of Interest" by depositor).

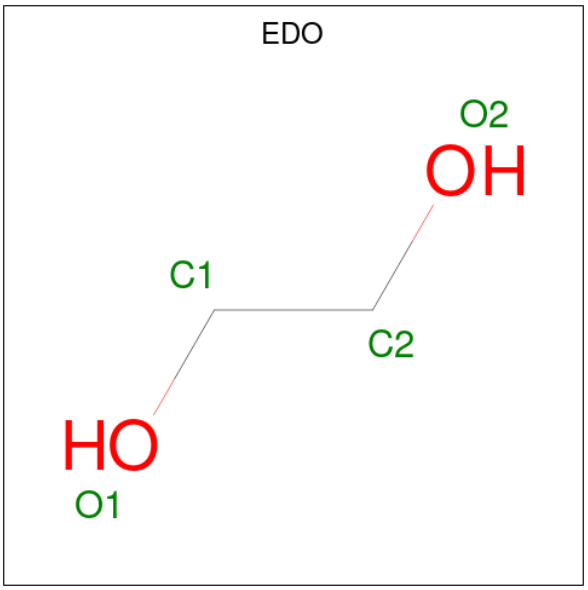


Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	AAA	1	Total	Br	C	N	O	0	0
			29	1	21	4	3		

- Molecule 3 is NICKEL (II) ION (three-letter code: NI) (formula: Ni).

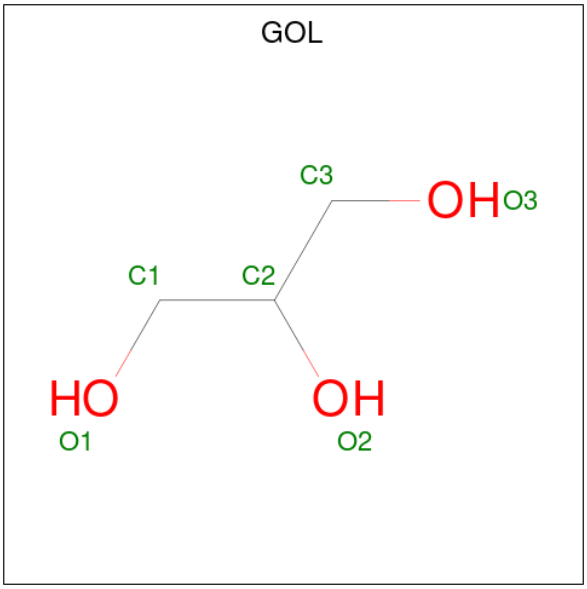
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	AAA	1	Total	Ni	0	0
			1	1		

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	BBB	1	Total C O 4 2 2	0	0

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	CCC	1	Total C O 6 3 3	0	0

- Molecule 6 is water.

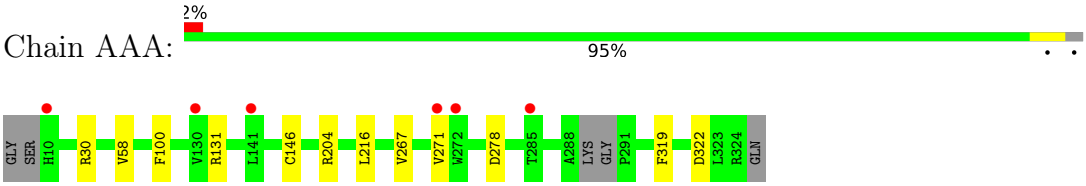
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	AAA	34	Total O 34 34	0	0
6	BBB	31	Total O 31 31	0	0
6	CCC	10	Total O 10 10	0	0

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.

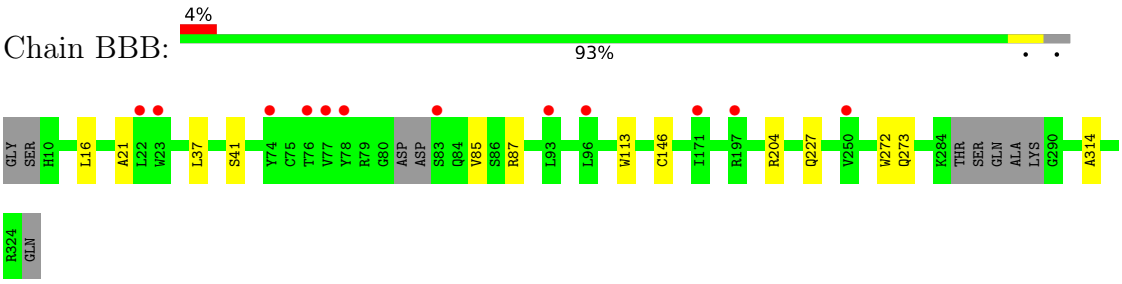
• Molecule 1: N-glycosylase/DNA lyase

Chain AAA:



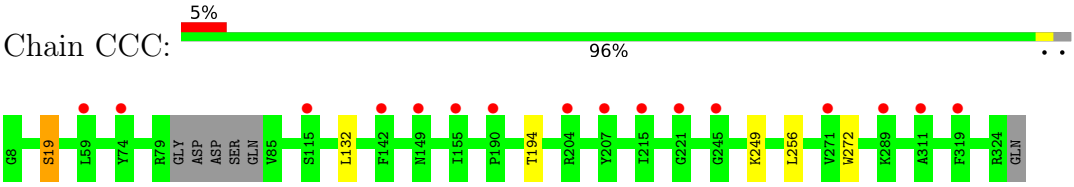
• Molecule 1: N-glycosylase/DNA lyase

Chain BBB:



• Molecule 1: N-glycosylase/DNA lyase

Chain CCC:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	81.19Å 81.83Å 169.56Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	84.78 – 2.45 84.78 – 2.45	Depositor EDS
% Data completeness (in resolution range)	99.9 (84.78-2.45) 99.9 (84.78-2.45)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.34 (at 2.45Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.253 , 0.291 0.253 , 0.292	Depositor DCC
R_{free} test set	2178 reflections (5.15%)	wwPDB-VP
Wilson B-factor (Å^2)	82.6	Xtriage
Anisotropy	0.059	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.001 for k,h,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	7205	wwPDB-VP
Average B, all atoms (Å^2)	92.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.10% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, 8HA, NI, EDO

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	AAA	0.65	0/2505	0.70	0/3419
1	BBB	0.66	0/2433	0.69	0/3325
1	CCC	0.67	0/2356	0.70	0/3233
All	All	0.66	0/7294	0.70	0/9977

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

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	AAA	2436	0	2339	4	0
1	BBB	2365	0	2252	5	0
1	CCC	2289	0	2094	3	0
2	AAA	29	0	0	0	0
3	AAA	1	0	0	0	0
4	BBB	4	0	6	0	0
5	CCC	6	0	8	0	0
6	AAA	34	0	0	0	0
6	BBB	31	0	0	0	0
6	CCC	10	0	0	0	0
All	All	7205	0	6699	11	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 1.

All (11) 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:CCC:132:LEU:CD2	1:CCC:256:LEU:HG	2.37	0.54
1:CCC:19:SER:O	1:CCC:19:SER:OG	2.21	0.53
1:BBB:21:ALA:O	1:BBB:87:ARG:NH1	2.44	0.51
1:AAA:100:PHE:O	1:AAA:131:ARG:HD3	2.15	0.46
1:AAA:278:ASP:HA	1:BBB:273:GLN:HE22	1.81	0.46
1:BBB:113:TRP:CE2	1:BBB:314:ALA:HB2	2.52	0.45
1:CCC:132:LEU:HD21	1:CCC:256:LEU:HG	1.98	0.44
1:BBB:37:LEU:O	1:BBB:41:SER:HB3	2.18	0.43
1:AAA:146:CYS:HA	1:AAA:204:ARG:HD3	2.00	0.43
1:BBB:146:CYS:HA	1:BBB:204:ARG:HD3	2.01	0.42
1:AAA:267:VAL:HA	1:AAA:271:VAL:HG21	2.01	0.42

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	AAA	309/318 (97%)	293 (95%)	16 (5%)	0	100 100
1	BBB	302/318 (95%)	285 (94%)	16 (5%)	1 (0%)	41 49
1	CCC	308/318 (97%)	288 (94%)	20 (6%)	0	100 100
All	All	919/954 (96%)	866 (94%)	52 (6%)	1 (0%)	51 64

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	BBB	85	VAL

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	AAA	249/266 (94%)	244 (98%)	5 (2%)	55	67
1	BBB	235/266 (88%)	232 (99%)	3 (1%)	69	79
1	CCC	211/266 (79%)	207 (98%)	4 (2%)	57	69
All	All	695/798 (87%)	683 (98%)	12 (2%)	60	73

All (12) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	AAA	30	ARG
1	AAA	58	VAL
1	AAA	216	LEU
1	AAA	319	PHE
1	AAA	322	ASP
1	BBB	16	LEU
1	BBB	227	GLN
1	BBB	272	TRP
1	CCC	19	SER
1	CCC	194	THR
1	CCC	249	LYS
1	CCC	272	TRP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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)

Of 4 ligands modelled in this entry, 1 is monoatomic - leaving 3 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	$\#\|Z\| > 2$
5	GOL	CCC	401	-	5,5,5	0.10	0	5,5,5	0.30	0
2	8HA	AAA	401	-	32,32,32	2.31	11 (34%)	46,46,46	2.18	14 (30%)
4	EDO	BBB	401	-	3,3,3	0.07	0	2,2,2	0.23	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
5	GOL	CCC	401	-	-	1/4/4/4	-
2	8HA	AAA	401	-	-	2/14/24/24	0/4/4/4
4	EDO	BBB	401	-	-	0/1/1/1	-

All (11) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	AAA	401	8HA	C07-N09	5.99	1.47	1.36
2	AAA	401	8HA	C07-N06	5.91	1.47	1.37
2	AAA	401	8HA	O02-C03	4.75	1.44	1.37
2	AAA	401	8HA	C05-N06	3.05	1.47	1.41
2	AAA	401	8HA	C17-C16	3.01	1.44	1.39
2	AAA	401	8HA	C10-N09	2.81	1.52	1.47
2	AAA	401	8HA	C22-C16	-2.81	1.37	1.46
2	AAA	401	8HA	C24-N23	-2.62	1.34	1.37
2	AAA	401	8HA	C22-C20	2.41	1.43	1.39

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	AAA	401	8HA	C12-N15	2.29	1.52	1.48
2	AAA	401	8HA	O25-C24	-2.10	1.19	1.23

All (14) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	AAA	401	8HA	C11-C12-N15	-7.75	91.84	112.27
2	AAA	401	8HA	C13-C12-N15	5.28	126.20	112.27
2	AAA	401	8HA	C13-C14-N09	-5.07	103.12	110.82
2	AAA	401	8HA	N06-C07-N09	3.13	119.47	115.89
2	AAA	401	8HA	O25-C24-N23	-3.01	124.81	127.50
2	AAA	401	8HA	C11-C12-C13	2.86	117.78	111.19
2	AAA	401	8HA	C14-N09-C10	-2.84	107.16	112.62
2	AAA	401	8HA	C11-C10-N09	-2.57	106.91	110.82
2	AAA	401	8HA	O02-C03-C28	2.53	119.64	114.99
2	AAA	401	8HA	C10-C11-C12	2.51	115.59	110.81
2	AAA	401	8HA	C16-N15-C24	-2.26	107.61	109.25
2	AAA	401	8HA	C20-C22-N23	2.16	132.62	129.39
2	AAA	401	8HA	O08-C07-N09	-2.04	118.92	121.78
2	AAA	401	8HA	C05-N06-C07	-2.01	122.09	126.12

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	AAA	401	8HA	C04-C03-O02-C01
5	CCC	401	GOL	O2-C2-C3-O3
2	AAA	401	8HA	C28-C03-O02-C01

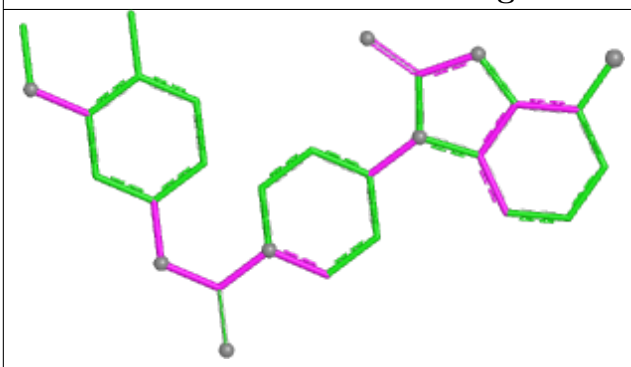
There are no ring outliers.

No monomer is involved in short contacts.

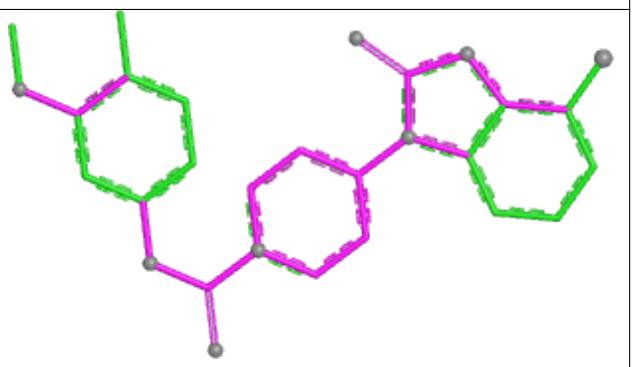
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 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.

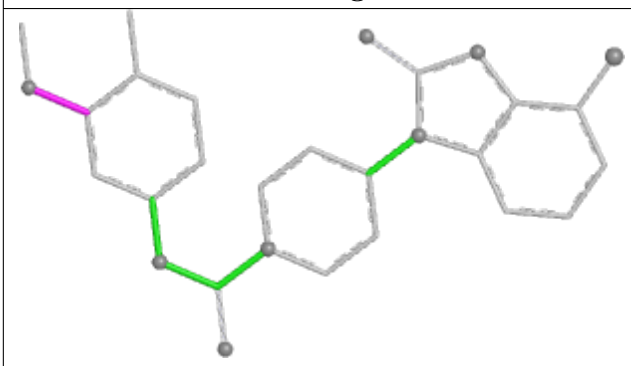
Ligand 8HA AAA 401



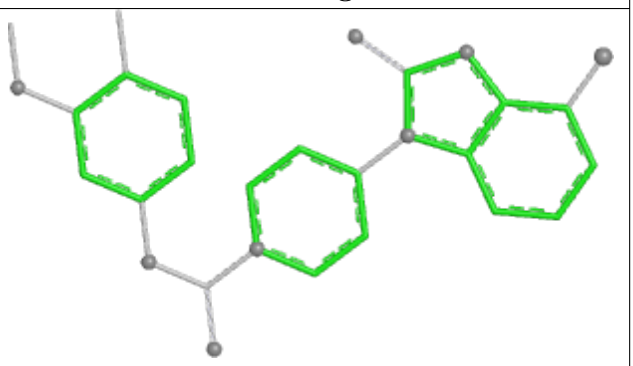
Bond lengths



Bond angles



Torsions



Rings

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

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^{th} 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	AAA	313/318 (98%)	0.43	6 (1%)	66	64	59, 80, 111, 137	0
1	BBB	308/318 (96%)	0.57	12 (3%)	39	36	64, 91, 121, 137	0
1	CCC	312/318 (98%)	0.56	16 (5%)	28	25	70, 103, 140, 156	1 (0%)
All	All	933/954 (97%)	0.52	34 (3%)	42	39	59, 92, 129, 156	1 (0%)

All (34) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	CCC	142	PHE	5.0
1	BBB	74	TYR	4.4
1	BBB	93	LEU	3.6
1	BBB	76	THR	3.5
1	BBB	96	LEU	3.4
1	BBB	77	VAL	3.4
1	CCC	115	SER	3.4
1	AAA	285	THR	3.3
1	CCC	271	VAL	3.1
1	CCC	311	ALA	3.0
1	CCC	245	GLY	2.8
1	BBB	250	VAL	2.7
1	CCC	74	TYR	2.7
1	CCC	221	GLY	2.6
1	BBB	23	TRP	2.6
1	AAA	130	VAL	2.5
1	CCC	319	PHE	2.5
1	CCC	204	ARG	2.5
1	CCC	155	ILE	2.4
1	CCC	59	LEU	2.4
1	AAA	141	LEU	2.4
1	BBB	171	ILE	2.3
1	AAA	271	VAL	2.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	BBB	83	SER	2.2
1	AAA	10	HIS	2.2
1	CCC	289	LYS	2.2
1	CCC	190	PRO	2.2
1	CCC	215	ILE	2.1
1	CCC	207	TYR	2.1
1	AAA	272	TRP	2.1
1	BBB	22	LEU	2.0
1	CCC	149	ASN	2.0
1	BBB	78	TYR	2.0
1	BBB	197	ARG	2.0

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)

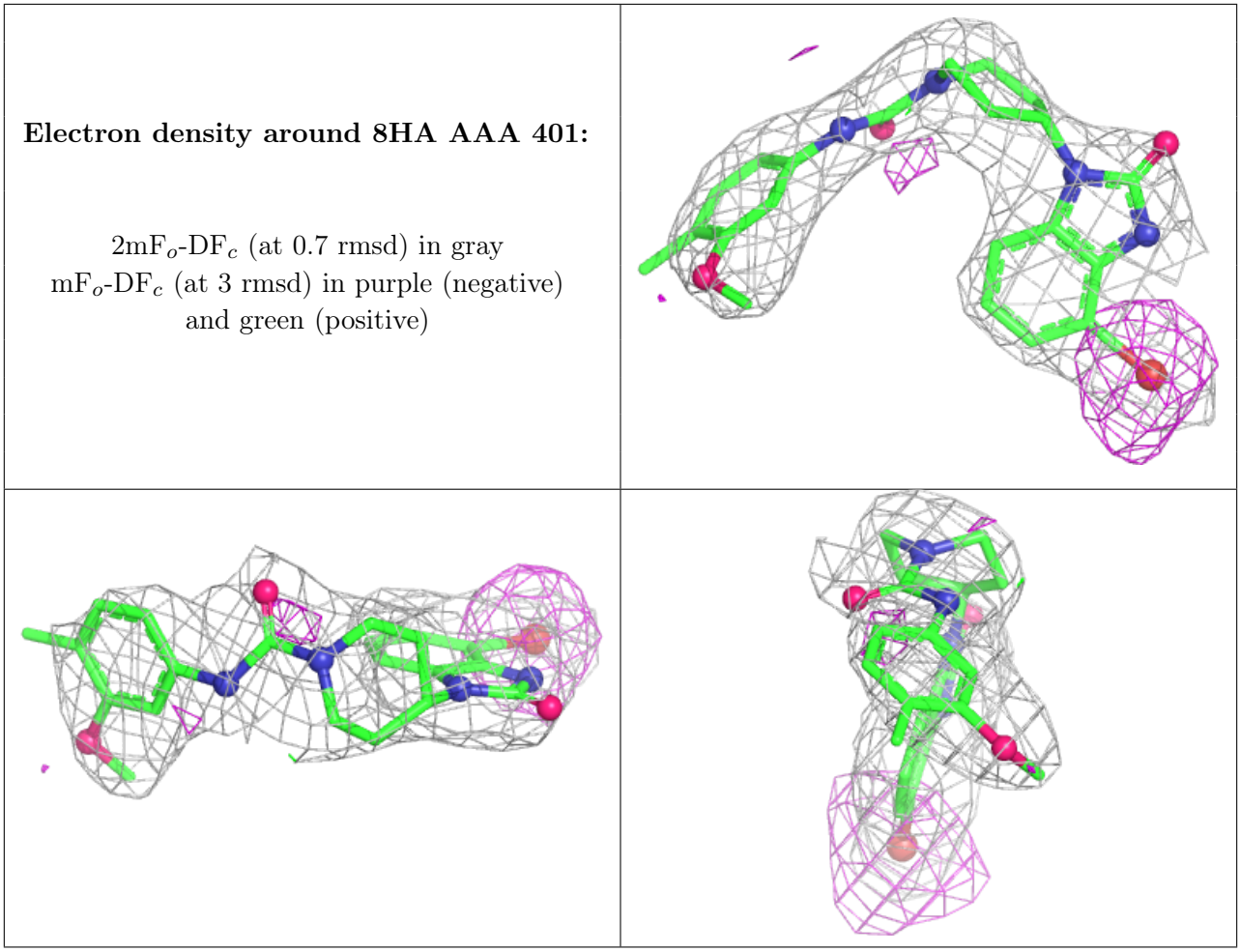
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' 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	GOL	CCC	401	6/6	0.59	0.19	92,94,95,95	0
4	EDO	BBB	401	4/4	0.72	0.23	111,112,112,113	0
2	8HA	AAA	401	29/29	0.82	0.26	82,97,116,131	0
3	NI	AAA	402	1/1	0.97	0.22	71,71,71,71	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.

Electron density around 8HA AAA 401:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)



6.5 Other polymers (i)

There are no such residues in this entry.