Full wwPDB X-ray Structure Validation Report ⓘ

May 24, 2022 – 10:13 am BST

PDB ID	:	7AZ0
Title	:	Structure of the mouse 8-oxoguanine DNA Glycosylase mOGG1 in complex with TH12161
Authors	:	Davies, J.R.; Masuyer, G.; Stenmark, P.
Deposited on	:	2020-11-13
Resolution	:	2.40 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 following versions of software and data (see references ⓘ) were used in the production of this report:

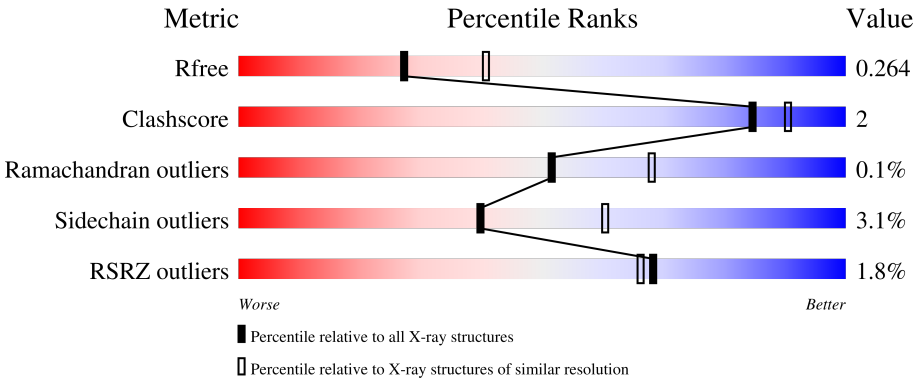
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.28.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.28.1

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 2.40 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	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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	4% 92% 6% ·
1	BBB	318	% 91% 6% ·
1	CCC	318	% 86% 9% · ·

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 7481 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.

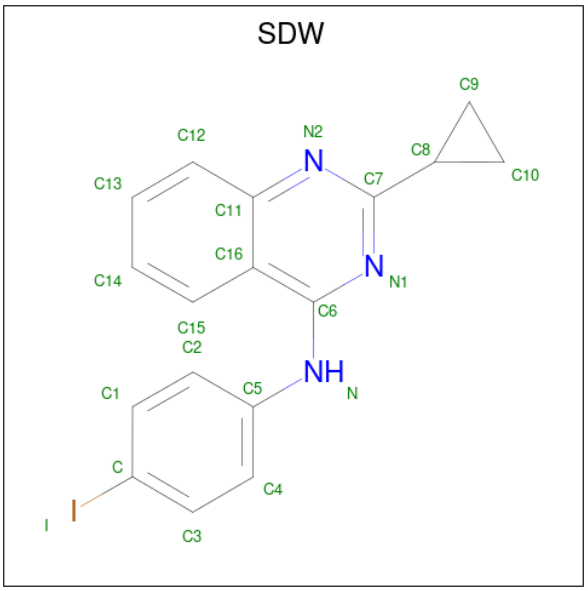
- Molecule 1 is a protein called N-glycosylase/DNA lyase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAA	312	Total	C	N	O	S	0	0	0
			2476	1576	445	444	11			
1	BBB	310	Total	C	N	O	S	0	0	0
			2471	1574	446	440	11			
1	CCC	307	Total	C	N	O	S	0	0	0
			2448	1562	442	433	11			

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 2-cyclopropyl- {N}-(4-iodophenyl)quinazolin-4-amine (three-letter code: SDW) (formula: $C_{17}H_{14}IN_3$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	AAA	1	Total C I N 21 17 1 3	0	0
2	BBB	1	Total C I N 21 17 1 3	0	0

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

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

- Molecule 4 is water.

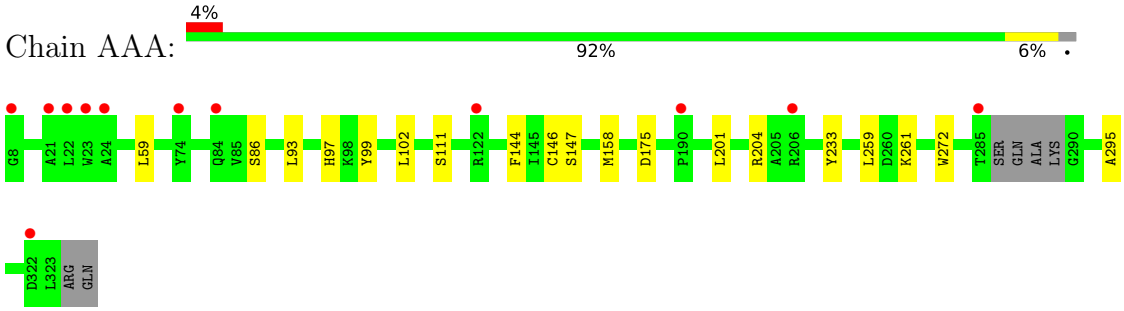
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	AAA	14	Total O 14 14	0	0
4	BBB	18	Total O 18 18	0	0
4	CCC	11	Total O 11 11	0	0

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.

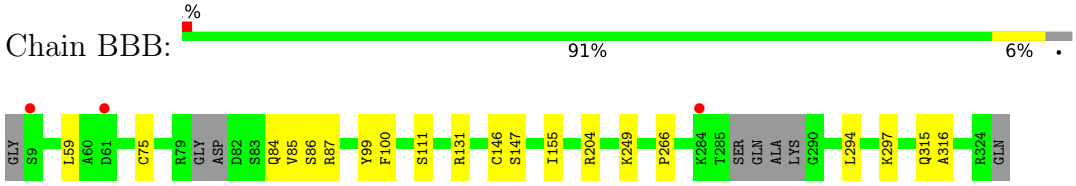
• Molecule 1: N-glycosylase/DNA lyase

Chain AAA:



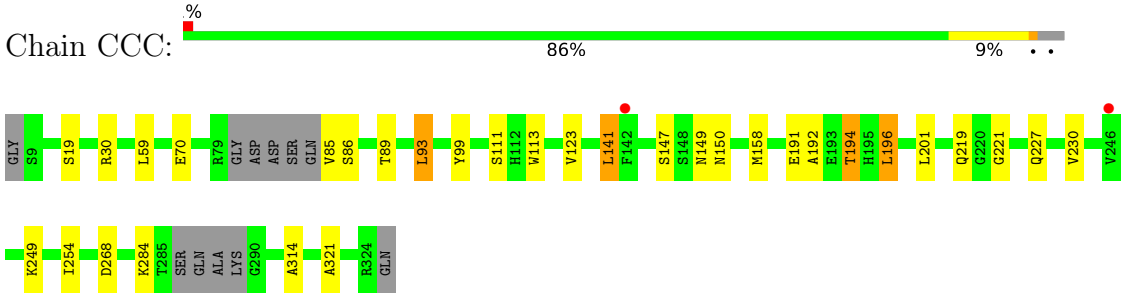
• Molecule 1: N-glycosylase/DNA lyase

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	81.09Å 81.64Å 169.63Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	73.68 – 2.40 73.57 – 2.40	Depositor EDS
% Data completeness (in resolution range)	100.0 (73.68-2.40) 100.0 (73.57-2.40)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.46 (at 2.40Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.230 , 0.257 0.237 , 0.264	Depositor DCC
R_{free} test set	2164 reflections (4.83%)	wwPDB-VP
Wilson B-factor (Å^2)	59.1	Xtriage
Anisotropy	0.267	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.012 for k,h,-l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	7481	wwPDB-VP
Average B, all atoms (Å^2)	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.38% 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: SDW, NI

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	0/2546	0.73	0/3466
1	BBB	0.64	0/2540	0.73	0/3456
1	CCC	0.65	0/2517	0.73	0/3425
All	All	0.64	0/7603	0.73	0/10347

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	2476	0	2415	8	0
1	BBB	2471	0	2417	9	0
1	CCC	2448	0	2400	15	0
2	AAA	21	0	0	1	0
2	BBB	21	0	0	5	0
3	AAA	1	0	0	0	0
4	AAA	14	0	0	0	0
4	BBB	18	0	0	0	0
4	CCC	11	0	0	0	0
All	All	7481	0	7232	34	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 2.

All (34) 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:30:ARG:NH1	1:CCC:70:GLU:OE1	2.30	0.65
1:AAA:146:CYS:HA	1:AAA:204:ARG:HD3	1.88	0.55
1:CCC:227:GLN:O	1:CCC:230:VAL:HG22	2.06	0.55
1:CCC:158:MET:CE	1:CCC:201:LEU:HG	2.37	0.55
1:CCC:191:GLU:O	1:CCC:194:THR:HG23	2.07	0.54
2:BBB:401:SDW:C4	2:BBB:401:SDW:N1	2.72	0.53
1:CCC:141:LEU:HD23	1:CCC:254:ILE:CD1	2.37	0.53
1:BBB:146:CYS:HA	1:BBB:204:ARG:HD3	1.92	0.52
1:AAA:259:LEU:O	1:AAA:261:LYS:HD3	2.09	0.52
2:AAA:401:SDW:C4	2:AAA:401:SDW:N1	2.73	0.50
1:BBB:155:ILE:HD12	2:BBB:401:SDW:C13	2.40	0.50
1:CCC:89:THR:O	1:CCC:93:LEU:HD22	2.12	0.50
1:CCC:147:SER:OG	1:CCC:149:ASN:ND2	2.45	0.49
1:BBB:155:ILE:HD12	2:BBB:401:SDW:C14	2.43	0.48
1:BBB:155:ILE:CD1	2:BBB:401:SDW:C13	2.91	0.48
1:AAA:233:TYR:HB2	1:AAA:261:LYS:HG3	1.98	0.45
1:CCC:192:ALA:O	1:CCC:196:LEU:HD13	2.17	0.45
1:BBB:59:LEU:HD21	1:BBB:99:TYR:CZ	2.53	0.44
1:CCC:196:LEU:HB3	1:CCC:201:LEU:HD22	1.98	0.43
1:CCC:123:VAL:HG13	1:CCC:321:ALA:HB1	2.01	0.43
1:CCC:249:LYS:HD2	1:CCC:268:ASP:HB3	2.00	0.42
1:CCC:141:LEU:HD23	1:CCC:254:ILE:HD13	2.01	0.42
1:CCC:141:LEU:HD23	1:CCC:254:ILE:HD11	2.01	0.42
1:AAA:59:LEU:HD21	1:AAA:99:TYR:CZ	2.55	0.42
1:AAA:272:TRP:HH2	1:AAA:295:ALA:HB1	1.85	0.41
1:AAA:144:PHE:HA	1:AAA:147:SER:HB3	2.02	0.41
1:BBB:316:ALA:HB2	2:BBB:401:SDW:I	2.91	0.41
1:AAA:158:MET:HB3	1:AAA:201:LEU:HD21	2.01	0.41
1:CCC:59:LEU:HD21	1:CCC:99:TYR:CZ	2.55	0.41
1:BBB:147:SER:HA	1:BBB:155:ILE:HD11	2.03	0.40
1:BBB:100:PHE:O	1:BBB:131:ARG:HD3	2.21	0.40
1:AAA:97:HIS:CE1	1:AAA:102:LEU:HB2	2.57	0.40
1:BBB:266:PRO:HD2	1:BBB:315:GLN:OE1	2.20	0.40
1:CCC:113:TRP:CE2	1:CCC:314:ALA:HB2	2.57	0.40

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	308/318 (97%)	297 (96%)	11 (4%)	0	100	100
1	BBB	304/318 (96%)	295 (97%)	9 (3%)	0	100	100
1	CCC	301/318 (95%)	290 (96%)	10 (3%)	1 (0%)	41	55
All	All	913/954 (96%)	882 (97%)	30 (3%)	1 (0%)	51	68

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	CCC	221	GLY

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	261/266 (98%)	257 (98%)	4 (2%)	65	80
1	BBB	261/266 (98%)	252 (97%)	9 (3%)	37	56
1	CCC	258/266 (97%)	247 (96%)	11 (4%)	29	46
All	All	780/798 (98%)	756 (97%)	24 (3%)	40	60

All (24) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	AAA	86	SER
1	AAA	93	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	AAA	111	SER
1	AAA	175	ASP
1	BBB	75	CYS
1	BBB	84	GLN
1	BBB	85	VAL
1	BBB	86	SER
1	BBB	87	ARG
1	BBB	111	SER
1	BBB	249	LYS
1	BBB	294	LEU
1	BBB	297	LYS
1	CCC	19	SER
1	CCC	85	VAL
1	CCC	86	SER
1	CCC	93	LEU
1	CCC	111	SER
1	CCC	141	LEU
1	CCC	150	ASN
1	CCC	194	THR
1	CCC	196	LEU
1	CCC	219	GLN
1	CCC	284	LYS

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 3 ligands modelled in this entry, 1 is monoatomic - leaving 2 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	$\#\|Z\| > 2$
2	SDW	AAA	401	-	23,24,24	0.22	0	32,34,34	0.52	0
2	SDW	BBB	401	-	23,24,24	0.19	0	32,34,34	0.56	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
2	SDW	AAA	401	-	-	1/8/10/10	0/4/4/4
2	SDW	BBB	401	-	-	1/8/10/10	0/4/4/4

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	AAA	401	SDW	N1-C6-N-C5
2	BBB	401	SDW	N1-C6-N-C5

There are no ring outliers.

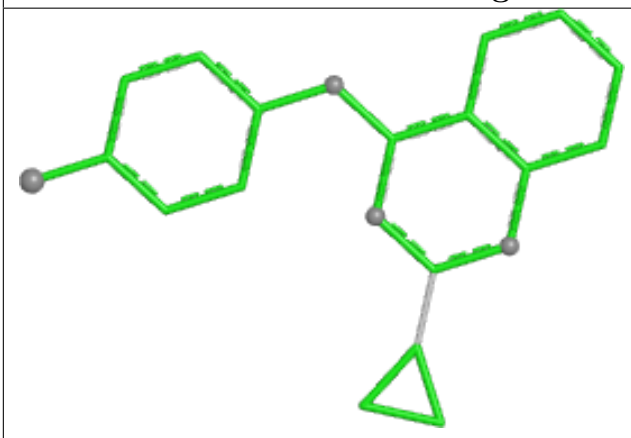
2 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	AAA	401	SDW	1	0
2	BBB	401	SDW	5	0

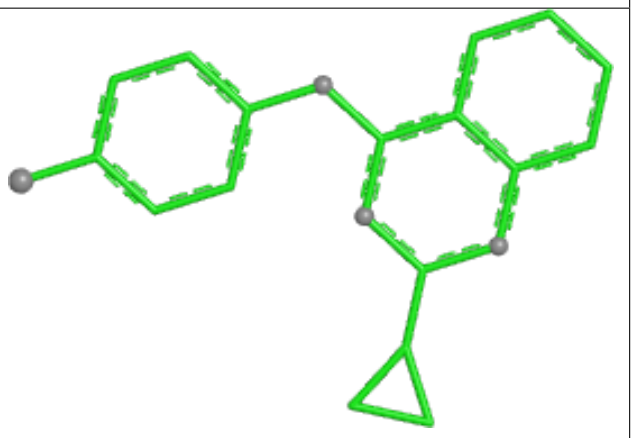
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths,

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 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.

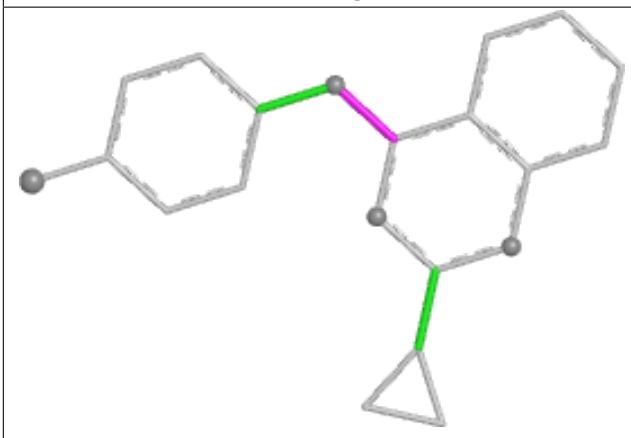
Ligand SDW AAA 401



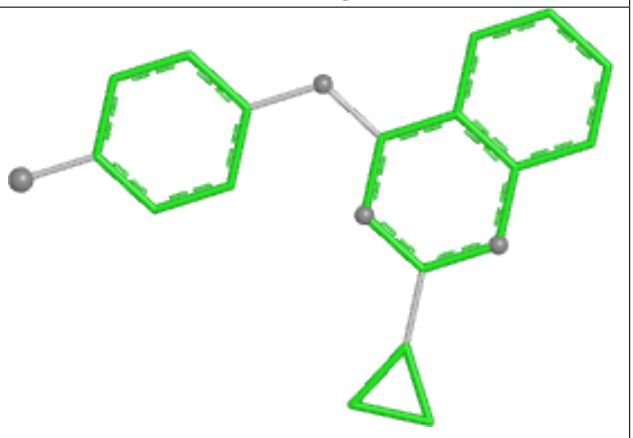
Bond lengths



Bond angles

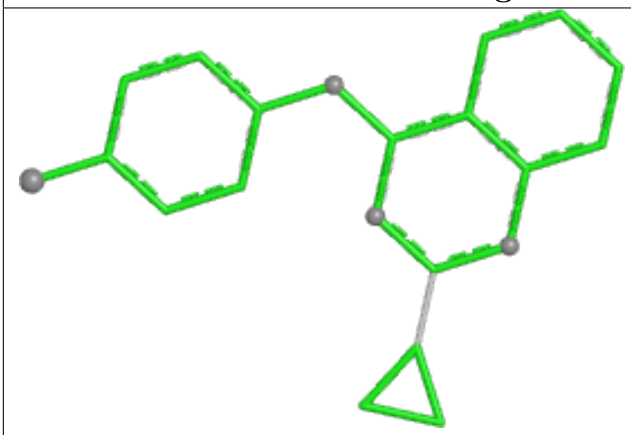


Torsions

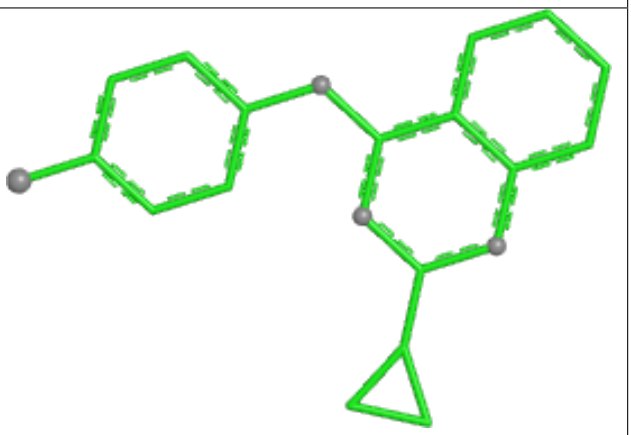


Rings

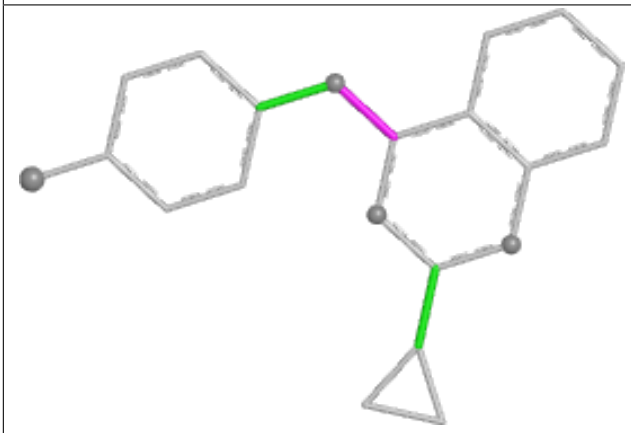
Ligand SDW BBB 401



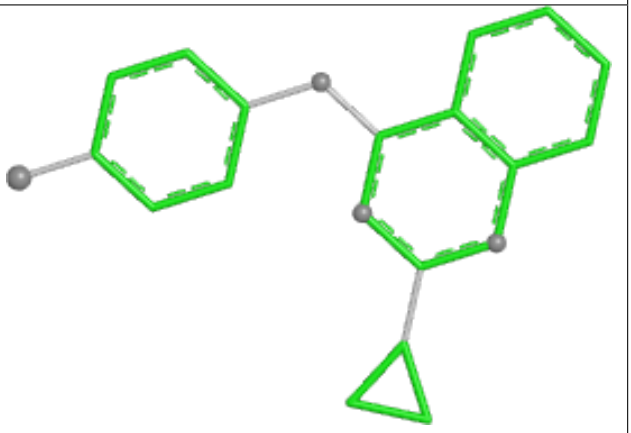
Bond lengths



Bond angles



Torsions



Rings

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

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	312/318 (98%)	0.19	12 (3%)	40	39	47, 72, 109, 123	0
1	BBB	310/318 (97%)	0.07	3 (0%)	82	80	39, 58, 90, 118	0
1	CCC	307/318 (96%)	0.09	2 (0%)	87	86	52, 76, 118, 137	0
All	All	929/954 (97%)	0.12	17 (1%)	68	66	39, 68, 107, 137	0

All (17) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AAA	8	GLY	4.2
1	BBB	61	ASP	3.3
1	AAA	84	GLN	3.1
1	AAA	285	THR	2.9
1	AAA	74	TYR	2.7
1	AAA	23	TRP	2.7
1	AAA	122	ARG	2.6
1	AAA	24	ALA	2.6
1	BBB	9	SER	2.5
1	AAA	22	LEU	2.4
1	CCC	142	PHE	2.3
1	AAA	206	ARG	2.3
1	AAA	21	ALA	2.2
1	AAA	190	PRO	2.1
1	AAA	322	ASP	2.1
1	CCC	246	VAL	2.0
1	BBB	284	LYS	2.0

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

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)

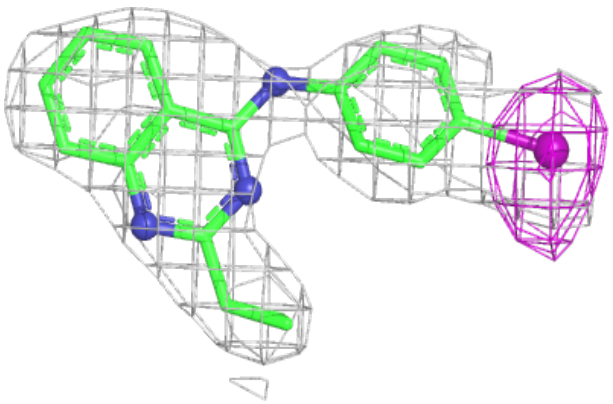
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

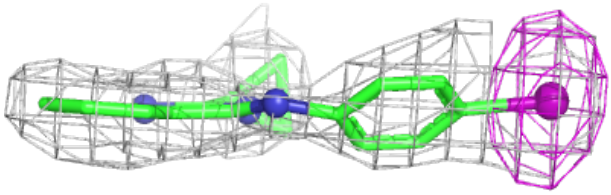
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	SDW	AAA	401	21/21	0.93	0.18	65,70,74,82	21
2	SDW	BBB	401	21/21	0.95	0.22	61,65,69,75	21
3	NI	AAA	402	1/1	0.99	0.12	53,53,53,53	0

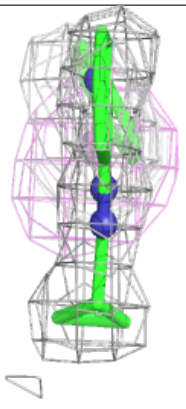
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around SDW 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)

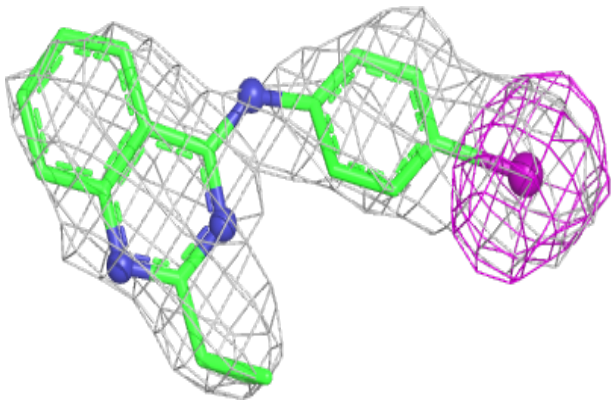


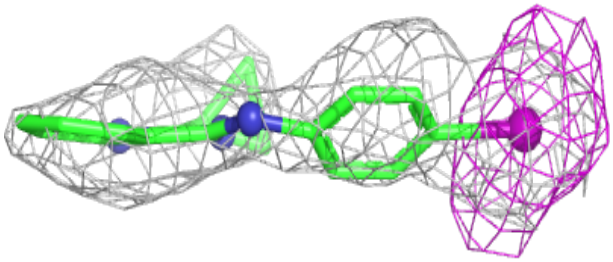


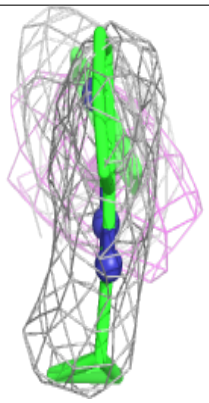


Electron density around SDW BBB 401:

$2mF_o$-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

There are no such residues in this entry.