wwPDB X-ray Structure Validation Summary Report (i)

Jun 23, 2024 – 01:15 AM EDT

PDB ID	:	6S5V
Title	:	Crystal structure of the Cap-Midlink region of the H5N1 Influenza A virus polymerase in complex with a Cap-domain binding analogue
Authors	:	Keown, J.R.; Fodor, E.; Grimes, J.M.
Deposited on	:	2019-07-02
Resolution	:	1.35 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:

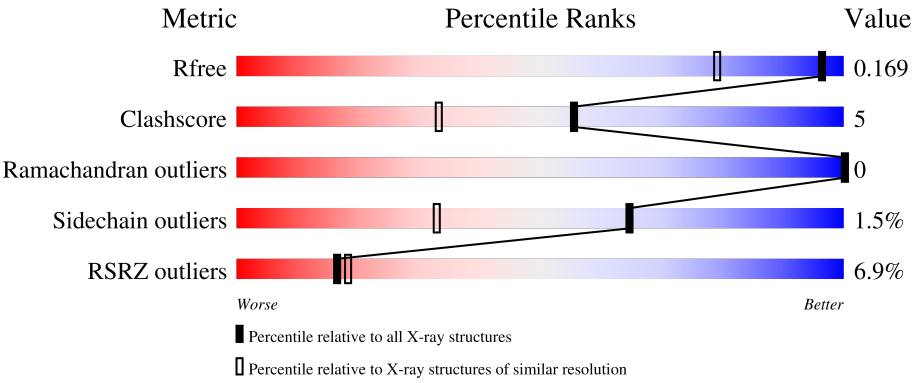
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 1.35 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	1509 (1.38-1.34)
Clashscore	141614	1551 (1.38-1.34)
Ramachandran outliers	138981	1530 (1.38-1.34)
Sidechain outliers	138945	1530 (1.38-1.34)
RSRZ outliers	127900	1487 (1.38-1.34)

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	AAAA	304	5% 89% · 10%
1	ABAA	304	8% 88% · 10%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 9719 atoms, of which 4713 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 Polymerase basic protein 2.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	AAAA	275	Total	C	H	N	O	S	0	17	0
			4572	1393	2327	413	419	20			
1	ABAA	275	Total	C	H	N	O	S	0	17	0
			4540	1388	2309	408	418	17			

There are 28 discrepancies between the modelled and reference sequences:

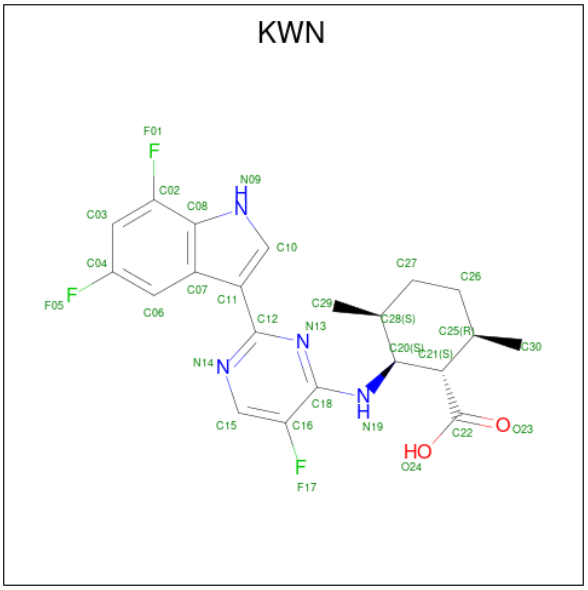
Chain	Residue	Modelled	Actual	Comment	Reference
AAAA	233	HIS	-	expression tag	UNP Q6E3N3
AAAA	234	HIS	-	expression tag	UNP Q6E3N3
AAAA	235	HIS	-	expression tag	UNP Q6E3N3
AAAA	236	HIS	-	expression tag	UNP Q6E3N3
AAAA	237	HIS	-	expression tag	UNP Q6E3N3
AAAA	238	HIS	-	expression tag	UNP Q6E3N3
AAAA	239	GLY	-	expression tag	UNP Q6E3N3
AAAA	240	GLU	-	expression tag	UNP Q6E3N3
AAAA	241	ASN	-	expression tag	UNP Q6E3N3
AAAA	242	LEU	-	expression tag	UNP Q6E3N3
AAAA	243	TYR	-	expression tag	UNP Q6E3N3
AAAA	244	PHE	-	expression tag	UNP Q6E3N3
AAAA	245	GLN	-	expression tag	UNP Q6E3N3
AAAA	246	GLY	-	expression tag	UNP Q6E3N3
ABAA	233	HIS	-	expression tag	UNP Q6E3N3
ABAA	234	HIS	-	expression tag	UNP Q6E3N3
ABAA	235	HIS	-	expression tag	UNP Q6E3N3
ABAA	236	HIS	-	expression tag	UNP Q6E3N3
ABAA	237	HIS	-	expression tag	UNP Q6E3N3
ABAA	238	HIS	-	expression tag	UNP Q6E3N3
ABAA	239	GLY	-	expression tag	UNP Q6E3N3
ABAA	240	GLU	-	expression tag	UNP Q6E3N3
ABAA	241	ASN	-	expression tag	UNP Q6E3N3
ABAA	242	LEU	-	expression tag	UNP Q6E3N3
ABAA	243	TYR	-	expression tag	UNP Q6E3N3

Continued on next page...

Continued from previous page...

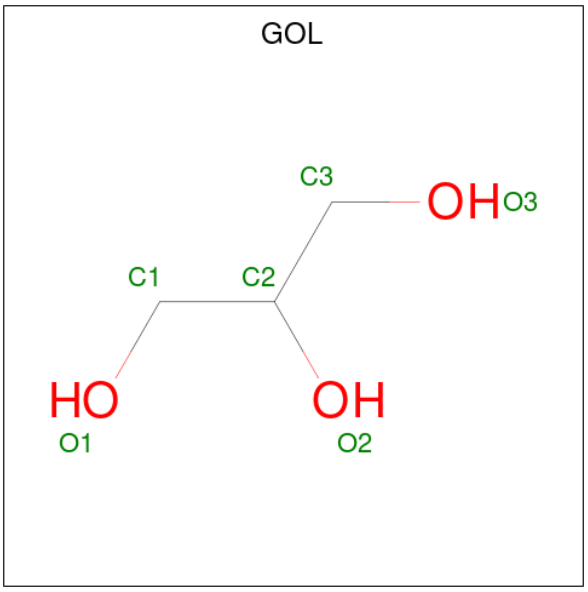
Chain	Residue	Modelled	Actual	Comment	Reference
ABAA	244	PHE	-	expression tag	UNP Q6E3N3
ABAA	245	GLN	-	expression tag	UNP Q6E3N3
ABAA	246	GLY	-	expression tag	UNP Q6E3N3

- Molecule 2 is (1 {S},2 {S},3 {S},6 {R})-2-[[2-[5,7-bis(fluoranyl)-1 {H}-indol-3-yl]-5-fluoranyl-pyrimidin-4-yl]amino]-3,6-dimethyl-cyclohexane-1-carboxylic acid (three-letter code: KWN) (formula: $C_{21}H_{21}F_3N_4O_2$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
2	AAAA	1	Total	C	F	H	N	O	0	0
			48	21	3	18	4	2		
2	ABAA	1	Total	C	F	H	N	O	0	0
			48	21	3	18	4	2		
2	ABAA	1	Total	C	F	H	N	O	0	0
			48	21	3	18	4	2		

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	AAAA	1	Total C H O 14 3 8 3	0	0
3	AAAA	1	Total C H O 14 3 8 3	0	0
3	ABAA	1	Total C H O 13 3 7 3	0	0

- Molecule 4 is POTASSIUM ION (three-letter code: K) (formula: K).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	ABAA	1	Total K 1 1	0	0

- Molecule 5 is water.

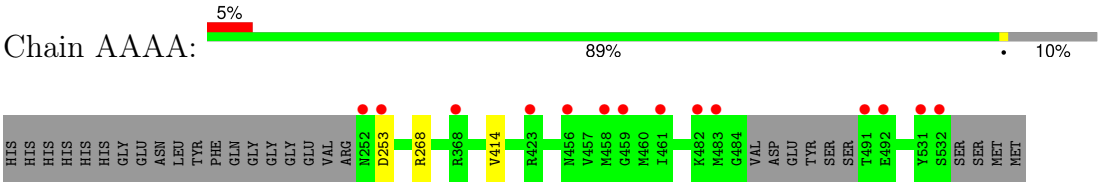
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	AAAA	207	Total O 207 207	0	0
5	ABAA	214	Total O 214 214	0	0

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.

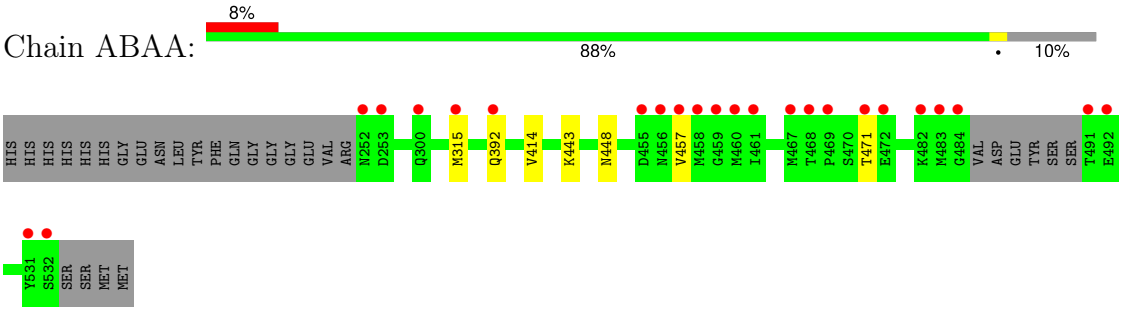
• Molecule 1: Polymerase basic protein 2

Chain AAAA:



• Molecule 1: Polymerase basic protein 2

Chain ABAA:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	105.70Å 105.70Å 140.66Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	84.50 – 1.35 84.50 – 1.35	Depositor EDS
% Data completeness (in resolution range)	100.0 (84.50-1.35) 100.0 (84.50-1.35)	Depositor EDS
R_{merge}	0.02	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.30 (at 1.35Å)	Xtriage
Refinement program	PHENIX 1.15.2_3472	Depositor
R, R_{free}	0.148 , 0.165 0.153 , 0.169	Depositor DCC
R_{free} test set	8936 reflections (5.14%)	wwPDB-VP
Wilson B-factor (Å^2)	18.2	Xtriage
Anisotropy	0.026	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.45 , 49.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.98	EDS
Total number of atoms	9719	wwPDB-VP
Average B, all atoms (Å^2)	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 66.22 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 6.2621e-06. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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: K, GOL, KWN

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	AAAA	0.43	0/2311	0.69	3/3104 (0.1%)
1	ABAA	0.43	0/2298	0.72	1/3090 (0.0%)
All	All	0.43	0/4609	0.70	4/6194 (0.1%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	ABAA	414	VAL	CG1-CB-CG2	7.39	122.73	110.90
1	AAAA	414	VAL	CG1-CB-CG2	7.29	122.56	110.90
1	AAAA	268	ARG	NE-CZ-NH2	-6.00	117.30	120.30
1	AAAA	268	ARG	NE-CZ-NH1	5.88	123.24	120.30

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	AAAA	2245	2327	0	0	0
1	ABAA	2231	2309	0	0	0
2	AAAA	30	18	0	0	0
2	ABAA	60	36	0	0	0
3	AAAA	12	16	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	ABAA	6	7	0	0	0
4	ABAA	1	0	0	0	0
5	AAAA	207	0	0	0	0
5	ABAA	214	0	0	0	0
All	All	5006	4713	0	0	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 5.

There are no clashes within the asymmetric unit.

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	AAAA	288/304 (95%)	284 (99%)	4 (1%)	0	100	100
1	ABAA	288/304 (95%)	283 (98%)	5 (2%)	0	100	100
All	All	576/608 (95%)	567 (98%)	9 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	AAAA	257/266 (97%)	256 (100%)	1 (0%)	91	81

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	ABAA	254/266 (96%)	247 (97%)	7 (3%)	43 10
All	All	511/532 (96%)	503 (98%)	8 (2%)	65 30

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	ABAA	471	THR
1	ABAA	457	VAL
1	ABAA	443	LYS
1	ABAA	392	GLN
1	ABAA	448	ASN

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 7 ligands modelled in this entry, 1 is monoatomic - leaving 6 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	KWN	ABAA	601	-	30,33,33	3.40	13 (43%)	32,49,49	1.80	7 (21%)
2	KWN	ABAA	603	-	30,33,33	4.36	13 (43%)	32,49,49	1.84	7 (21%)
3	GOL	AAAA	602	-	5,5,5	0.93	0	5,5,5	0.94	0
3	GOL	ABAA	602	-	5,5,5	1.04	0	5,5,5	1.24	1 (20%)
2	KWN	AAAA	601	-	30,33,33	3.12	12 (40%)	32,49,49	1.64	7 (21%)
3	GOL	AAAA	603	-	5,5,5	0.83	0	5,5,5	1.09	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	KWN	ABAA	601	-	-	4/8/29/29	0/4/4/4
2	KWN	ABAA	603	-	-	4/8/29/29	0/4/4/4
3	GOL	AAAA	602	-	-	3/4/4/4	-
3	GOL	ABAA	602	-	-	4/4/4/4	-
2	KWN	AAAA	601	-	-	0/8/29/29	0/4/4/4
3	GOL	AAAA	603	-	-	0/4/4/4	-

The worst 5 of 38 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	ABAA	603	KWN	C02-C08	12.83	1.53	1.41
2	ABAA	601	KWN	C02-C08	9.61	1.50	1.41
2	AAAA	601	KWN	C02-C08	8.68	1.49	1.41
2	ABAA	603	KWN	C06-C04	8.64	1.49	1.36
2	ABAA	603	KWN	C12-N14	7.27	1.46	1.34

The worst 5 of 22 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	ABAA	603	KWN	C16-C15-N14	-6.41	117.27	122.68
2	ABAA	601	KWN	C16-C15-N14	-5.74	117.84	122.68
2	AAAA	601	KWN	C16-C15-N14	-4.51	118.87	122.68
2	AAAA	601	KWN	C30-C25-C21	-3.35	107.03	112.51
2	ABAA	603	KWN	C30-C25-C21	-3.33	107.07	112.51

There are no chirality outliers.

5 of 15 torsion outliers are listed below:

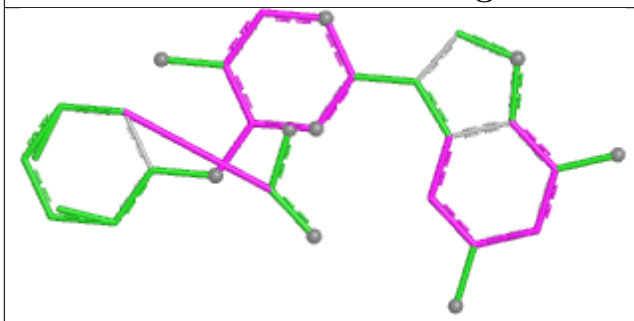
Mol	Chain	Res	Type	Atoms
3	AAAA	602	GOL	O1-C1-C2-O2
3	AAAA	602	GOL	O1-C1-C2-C3
3	ABAA	602	GOL	O1-C1-C2-O2
3	ABAA	602	GOL	O1-C1-C2-C3
3	ABAA	602	GOL	C1-C2-C3-O3

There are no ring outliers.

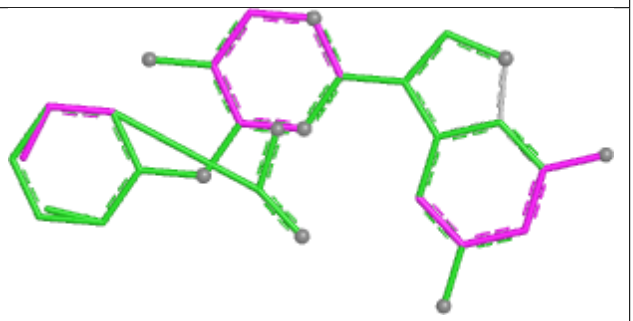
No monomer is involved in short contacts.

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.

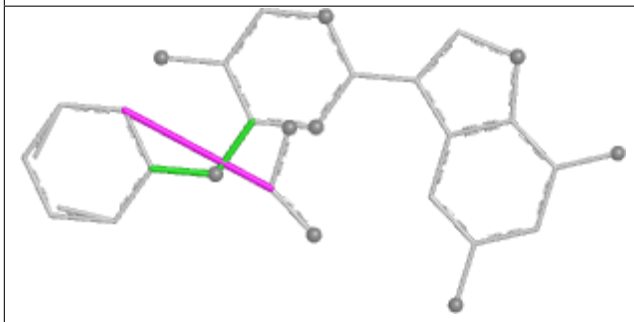
Ligand KWN ABAA 601



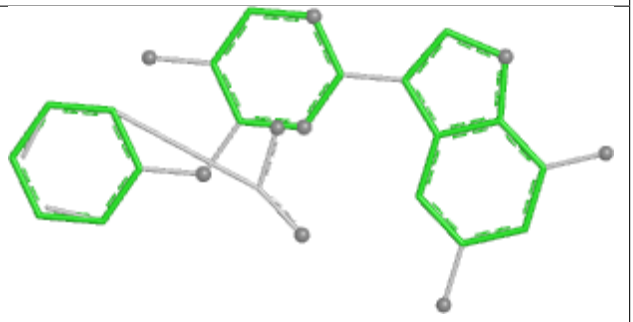
Bond lengths



Bond angles

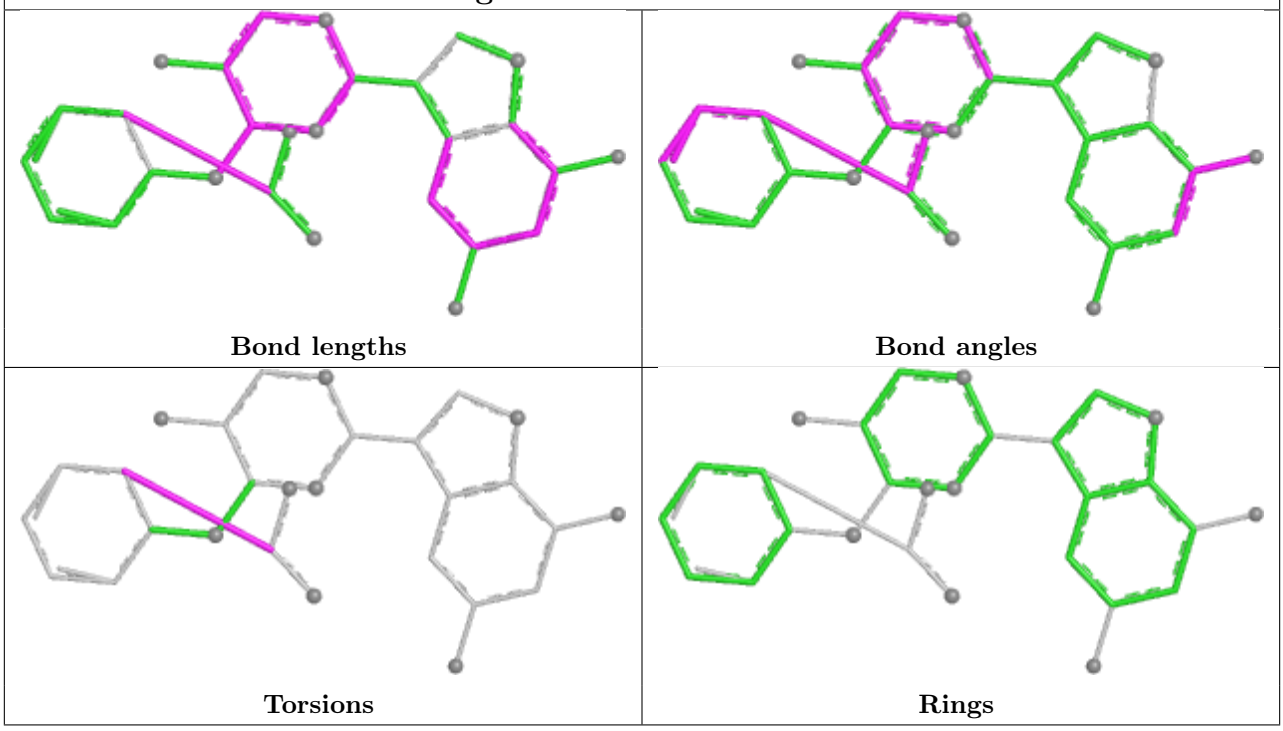


Torsions

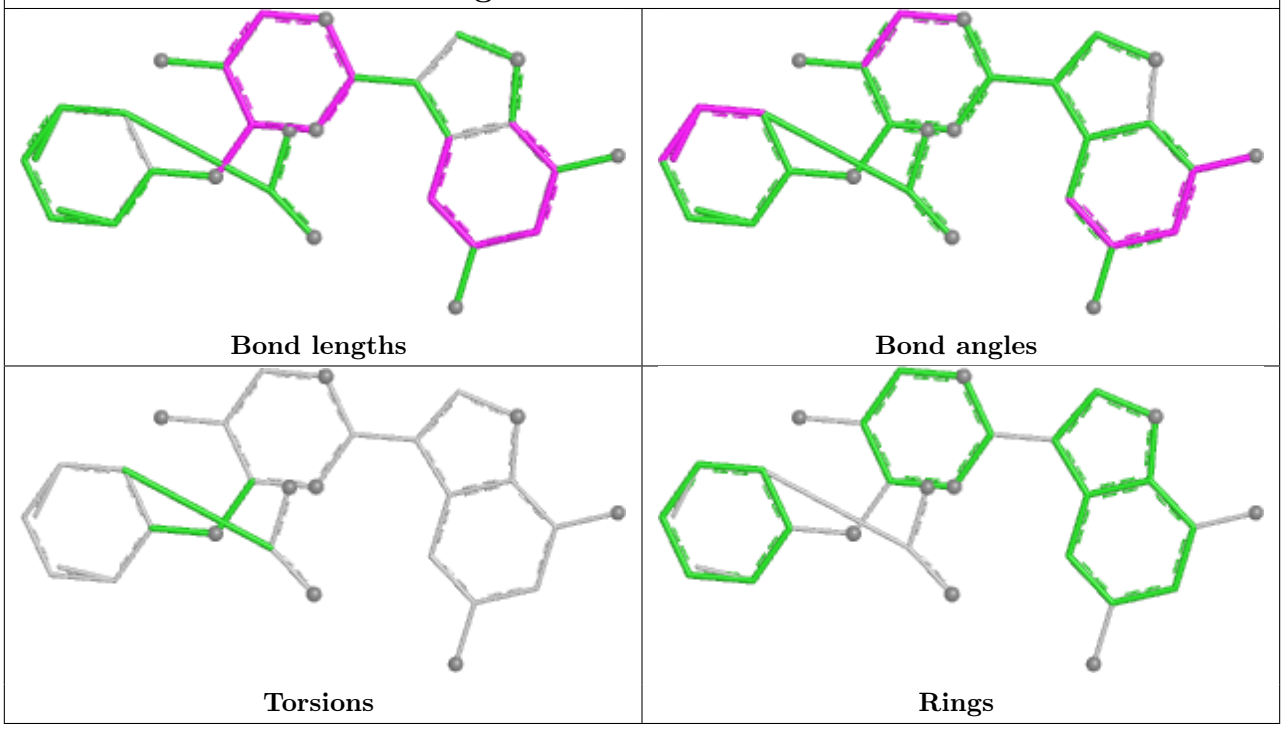


Rings

Ligand KWN ABAA 603



Ligand KWN AAAA 601



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	AAAA	275/304 (90%)	0.33	14 (5%)	28	31	13, 20, 44, 71	0
1	ABAA	275/304 (90%)	0.42	24 (8%)	10	12	13, 20, 45, 82	0
All	All	550/608 (90%)	0.38	38 (6%)	16	18	13, 20, 45, 82	0

The worst 5 of 38 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	ABAA	483	MET	7.3
1	AAAA	252	ASN	6.8
1	ABAA	456	ASN	6.0
1	AAAA	253	ASP	5.5
1	ABAA	459	GLY	5.4

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

There are no monosaccharides in this entry.

6.4 Ligands

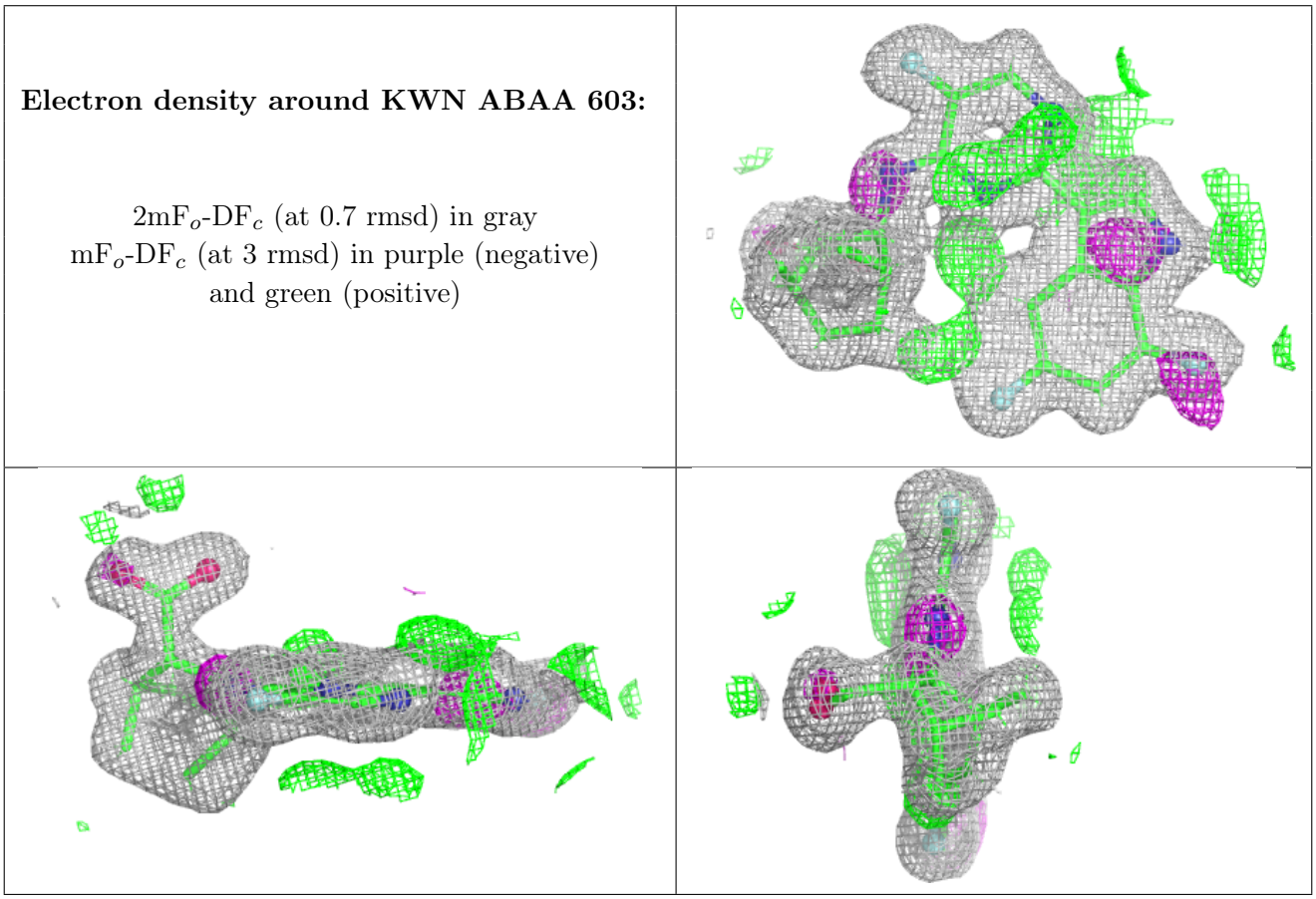
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' 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
3	GOL	AAAA	602	6/6	0.55	0.39	84,101,102,103	0
3	GOL	ABAA	602	6/6	0.73	0.23	66,70,83,84	0
3	GOL	AAAA	603	6/6	0.75	0.29	68,82,84,85	0
2	KWN	ABAA	603	30/30	0.92	0.12	17,25,31,32	0
2	KWN	AAAA	601	30/30	0.98	0.08	12,17,23,23	0
2	KWN	ABAA	601	30/30	0.98	0.08	13,18,24,24	0
4	K	ABAA	604	1/1	0.98	0.05	27,27,27,27	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.

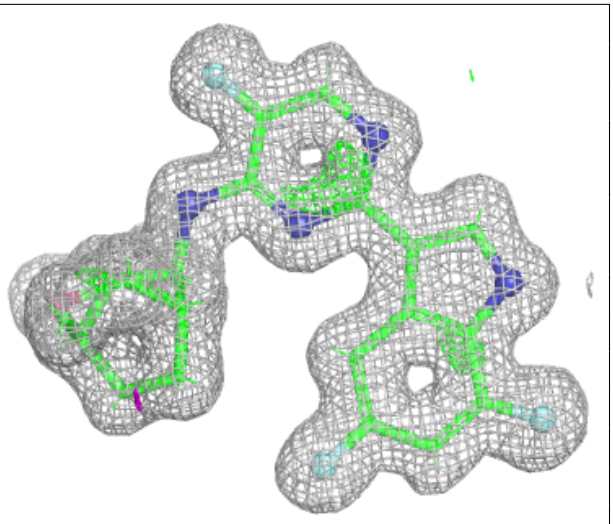
Electron density around KWN ABAA 603:

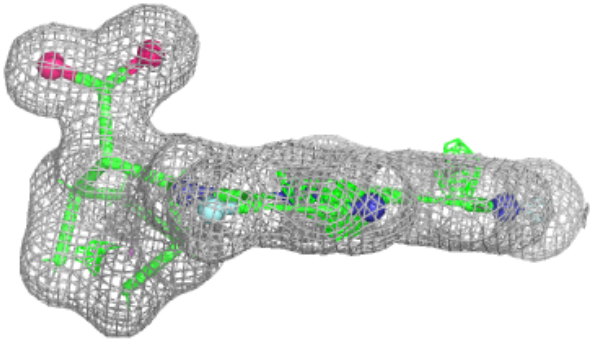
$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

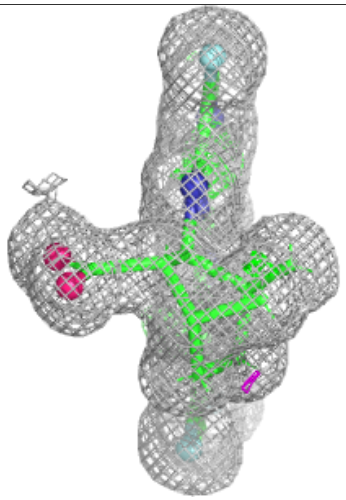


Electron density around KWN AAAA 601:

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

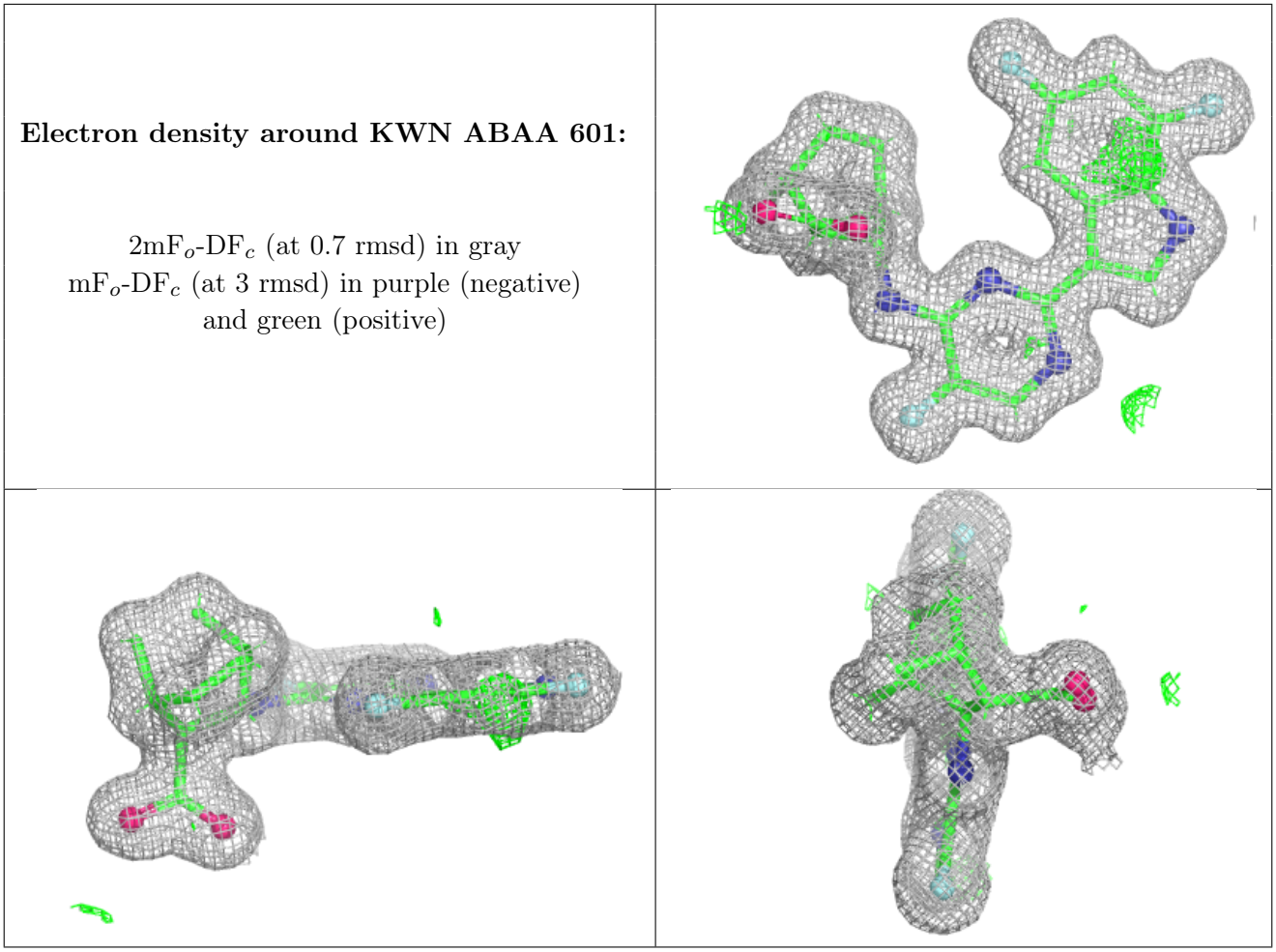






Electron density around KWN ABAA 601:

$2mF_o$-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.