



wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 14, 2024 – 11:20 am GMT

PDB ID : 6Q8Z
Title : Structure of human galactokinase 1 bound with N-(Cyclobutylmethyl)-1,5-dimethyl-1H-pyrazole-4-carboxamide
Authors : Mackinnon, S.R.; Bezerra, G.A.; Zhang, M.; Foster, W.; Krojer, T.; Brandao-Neto, J.; Douangamath, A.; Arrowsmith, C.; Edwards, A.; Bountra, C.; Brennan, P.; Lai, K.; Yue, W.W.
Deposited on : 2018-12-16
Resolution : 2.40 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

The following versions of software and data (see [references](#) ①) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.8.4, CSD as541be (2020)
Xtrriage (Phenix) : 1.13
EDS : 2.36
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) : 2.36

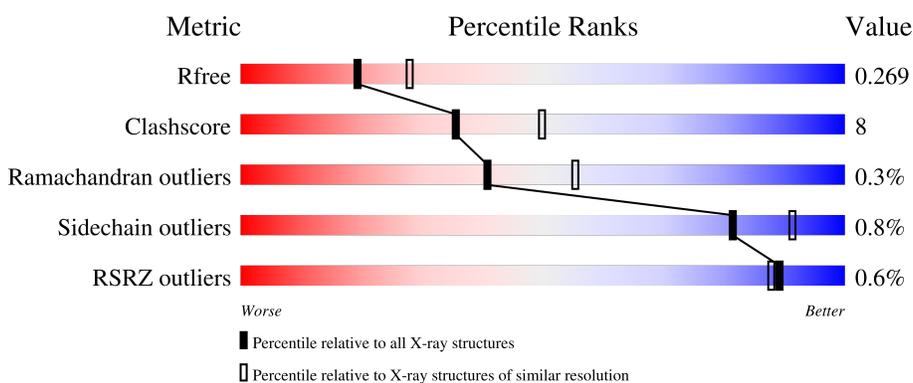
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 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 $\leq 5\%$. 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	392	
1	B	392	
1	C	392	
1	D	392	

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 criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	GAL	A	401	X	-	-	-

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 11861 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.

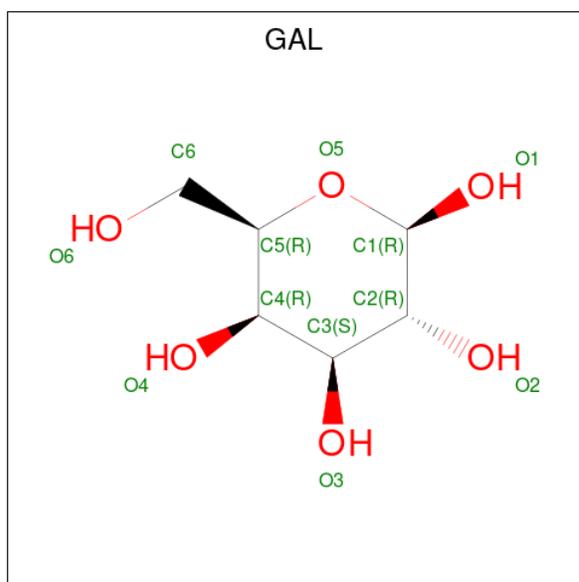
- Molecule 1 is a protein called Galactokinase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	392	2880	1807	508	549	16	0	0	0
1	B	387	2679	1681	471	512	15	0	0	0
1	C	386	2783	1747	489	531	16	0	0	0
1	D	376	2621	1641	456	511	13	0	1	0

There are 4 discrepancies between the modelled and reference sequences:

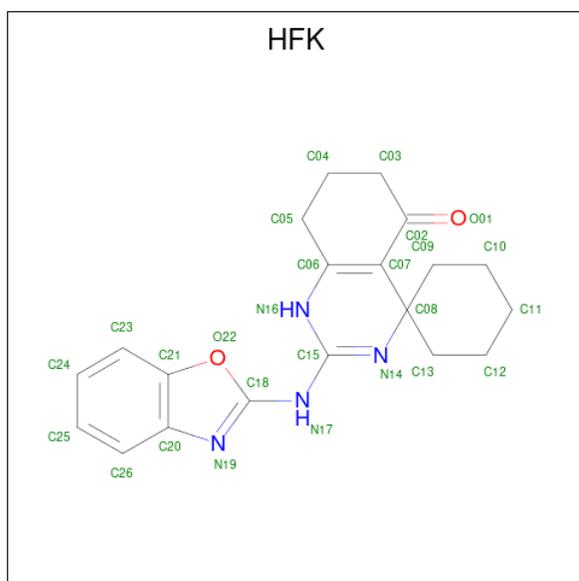
Chain	Residue	Modelled	Actual	Comment	Reference
A	1	HIS	-	expression tag	UNP P51570
B	1	HIS	-	expression tag	UNP P51570
C	1	HIS	-	expression tag	UNP P51570
D	1	HIS	-	expression tag	UNP P51570

- Molecule 2 is beta-D-galactopyranose (three-letter code: GAL) (formula: C₆H₁₂O₆).



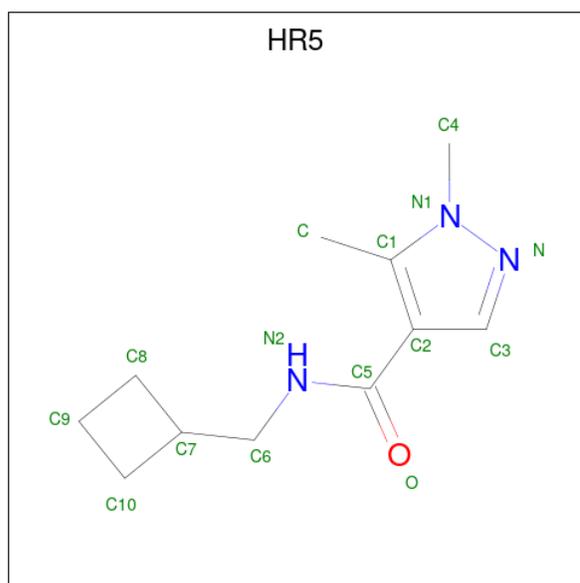
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C O 12 6 6	0	0
2	B	1	Total C O 12 6 6	0	0
2	C	1	Total C O 12 6 6	0	0

- Molecule 3 is 2-(1,3-benzoxazol-2-ylamino)spiro[1,6,7,8-tetrahydroquinazoline-4,1'-cyclohexane]-5-one (three-letter code: HFK) (formula: $C_{20}H_{22}N_4O_2$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			26	20	4	2		
3	B	1	Total	C	N	O	0	0
			26	20	4	2		
3	C	1	Total	C	N	O	0	0
			26	20	4	2		

- Molecule 4 is {N}-(cyclobutylmethyl)-1,5-dimethyl-pyrazole-4-carboxamide (three-letter code: HR5) (formula: C₁₁H₁₇N₃O).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total	C	N	O	0	0
			15	11	3	1		
4	C	1	Total	C	N	O	0	0
			15	11	3	1		
4	C	1	Total	C	N	O	0	0
			15	11	3	1		

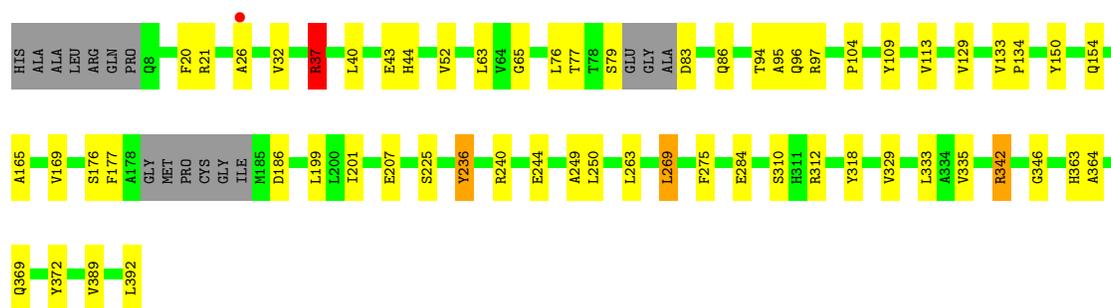
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	222	Total	O	0	0
			222	222		
5	B	216	Total	O	0	0
			216	216		
5	C	159	Total	O	0	0
			159	159		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	D	142	Total 142	O 142	0	0



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	73.62Å 114.39Å 120.85Å 90.00° 100.59° 90.00°	Depositor
Resolution (Å)	52.74 – 2.40 82.40 – 1.81	Depositor EDS
% Data completeness (in resolution range)	99.6 (52.74-2.40) 99.8 (82.40-1.81)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.60 (at 1.81Å)	Xtrriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.241 , 0.283 0.227 , 0.269	Depositor DCC
R_{free} test set	8865 reflections (4.97%)	wwPDB-VP
Wilson B-factor (Å ²)	31.2	Xtrriage
Anisotropy	0.506	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 49.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	11861	wwPDB-VP
Average B, all atoms (Å ²)	47.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.69% 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: GAL, HR5, HFK

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	0.88	2/2934 (0.1%)	0.89	6/3991 (0.2%)
1	B	0.81	5/2726 (0.2%)	0.90	4/3721 (0.1%)
1	C	0.69	1/2834 (0.0%)	0.81	3/3858 (0.1%)
1	D	0.75	2/2667 (0.1%)	0.83	5/3645 (0.1%)
All	All	0.79	10/11161 (0.1%)	0.86	18/15215 (0.1%)

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	37	ARG	CZ-NH1	7.73	1.43	1.33
1	B	37	ARG	CZ-NH2	-6.60	1.24	1.33
1	B	143	ALA	CA-CB	6.21	1.65	1.52
1	B	37	ARG	CZ-NH1	5.93	1.40	1.33
1	D	310	SER	CA-CB	5.87	1.61	1.52

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	37	ARG	NE-CZ-NH1	17.69	129.14	120.30
1	B	37	ARG	NE-CZ-NH1	17.44	129.02	120.30
1	C	37	ARG	NE-CZ-NH1	-14.17	113.21	120.30
1	C	37	ARG	NE-CZ-NH2	12.94	126.77	120.30
1	A	37	ARG	NE-CZ-NH2	-11.25	114.67	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	A	2880	0	2806	32	0
1	B	2679	0	2465	54	0
1	C	2783	0	2660	49	0
1	D	2621	0	2402	40	1
2	A	12	0	9	0	0
2	B	12	0	11	0	0
2	C	12	0	12	0	0
3	A	26	0	0	0	0
3	B	26	0	0	1	0
3	C	26	0	0	0	0
4	A	15	0	0	1	0
4	C	30	0	0	2	0
5	A	222	0	0	6	4
5	B	216	0	0	21	1
5	C	159	0	0	7	0
5	D	142	0	0	6	2
All	All	11861	0	10365	173	4

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 8.

The worst 5 of 173 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:B:37:ARG:NH1	1:B:140:SER:OG	1.62	1.31
1:C:37:ARG:NH1	1:C:140:SER:OG	1.78	1.16
1:A:89:GLN:NE2	5:A:501:HOH:O	1.92	1.01
1:B:37:ARG:NH1	1:B:140:SER:HG	1.54	0.97
1:B:331:ALA:N	5:B:503:HOH:O	2.06	0.88

All (4) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
5:A:630:HOH:O	5:B:713:HOH:O[1_554]	2.08	0.12
5:A:700:HOH:O	5:D:506:HOH:O[1_455]	2.12	0.08
1:D:207:GLU:OE1	5:A:512:HOH:O[1_655]	2.18	0.02
5:A:685:HOH:O	5:D:506:HOH:O[1_455]	2.19	0.01

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	A	390/392 (100%)	385 (99%)	5 (1%)	0	100	100
1	B	379/392 (97%)	368 (97%)	9 (2%)	2 (0%)	29	41
1	C	380/392 (97%)	372 (98%)	7 (2%)	1 (0%)	41	55
1	D	371/392 (95%)	367 (99%)	3 (1%)	1 (0%)	41	55
All	All	1520/1568 (97%)	1492 (98%)	24 (2%)	4 (0%)	41	55

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	7	PRO
1	D	95	ALA
1	B	230	SER
1	C	95	ALA

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	A	290/311 (93%)	289 (100%)	1 (0%)	92	97
1	B	241/311 (78%)	239 (99%)	2 (1%)	81	91
1	C	271/311 (87%)	268 (99%)	3 (1%)	73	87
1	D	241/311 (78%)	239 (99%)	2 (1%)	81	91
All	All	1043/1244 (84%)	1035 (99%)	8 (1%)	81	91

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

Mol	Chain	Res	Type
1	D	363	HIS
1	D	37	ARG
1	C	78	THR
1	C	37	ARG
1	C	188	PHE

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

Mol	Chain	Res	Type
1	B	371	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 monosaccharides in this entry.

5.6 Ligand geometry [i](#)

9 ligands 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
2	GAL	B	401	-	12,12,12	1.60	3 (25%)	17,17,17	1.56	4 (23%)
4	HR5	C	403	-	13,16,16	1.74	3 (23%)	15,22,22	1.18	2 (13%)
2	GAL	A	401	-	12,12,12	2.28	6 (50%)	17,17,17	3.19	6 (35%)
2	GAL	C	401	-	12,12,12	1.35	2 (16%)	17,17,17	1.29	3 (17%)
4	HR5	A	403	-	13,16,16	1.40	2 (15%)	15,22,22	1.77	5 (33%)
3	HFK	A	402	-	26,30,30	4.29	9 (34%)	24,44,44	3.23	9 (37%)
4	HR5	C	404	-	13,16,16	1.62	2 (15%)	15,22,22	1.20	2 (13%)
3	HFK	C	402	-	26,30,30	4.52	10 (38%)	24,44,44	3.85	12 (50%)
3	HFK	B	402	-	26,30,30	4.34	10 (38%)	24,44,44	3.85	9 (37%)

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	GAL	B	401	-	-	1/2/22/22	0/1/1/1
4	HR5	C	403	-	-	0/7/15/15	0/2/2/2
2	GAL	A	401	-	1/1/5/5	1/2/22/22	0/1/1/1
2	GAL	C	401	-	-	1/2/22/22	0/1/1/1
4	HR5	A	403	-	-	6/7/15/15	0/2/2/2
3	HFK	A	402	-	-	2/2/43/43	0/5/5/5
4	HR5	C	404	-	-	1/7/15/15	0/2/2/2
3	HFK	C	402	-	-	2/2/43/43	0/5/5/5
3	HFK	B	402	-	-	2/2/43/43	0/5/5/5

The worst 5 of 47 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	402	HFK	C06-C07	14.37	1.55	1.37
3	B	402	HFK	C06-C07	13.06	1.54	1.37
3	A	402	HFK	C06-C07	12.77	1.53	1.37
3	B	402	HFK	C15-N14	9.45	1.54	1.30
3	C	402	HFK	C15-N16	9.24	1.51	1.36

The worst 5 of 52 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	402	HFk	O01-C02-C03	-9.54	105.23	120.86
3	B	402	HFk	O01-C02-C03	-8.72	106.57	120.86
3	B	402	HFk	C02-C07-C06	-8.29	111.29	119.17
3	A	402	HFk	N16-C15-N14	-7.95	110.58	123.88
3	B	402	HFk	C05-C06-C07	-7.54	113.03	122.97

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	A	401	GAL	C1

5 of 16 torsion outliers are listed below:

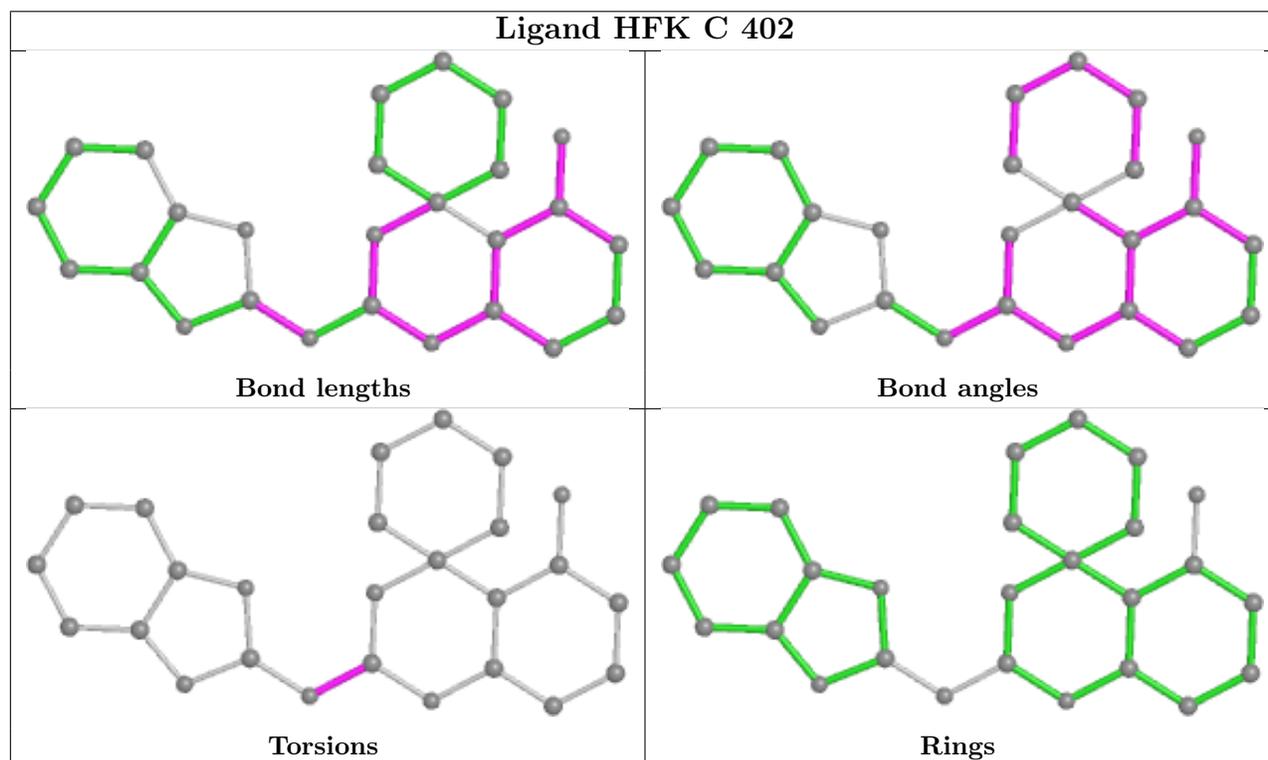
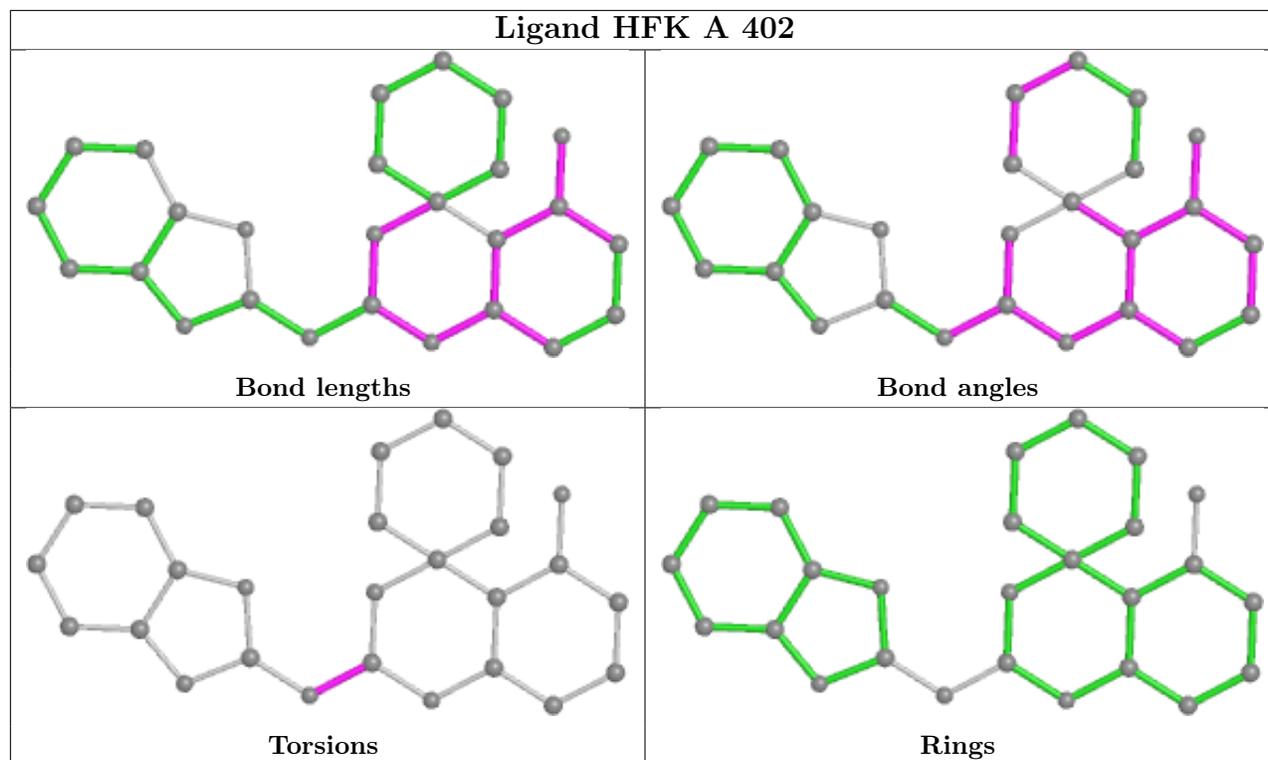
Mol	Chain	Res	Type	Atoms
3	A	402	HFk	N14-C15-N17-C18
3	B	402	HFk	N16-C15-N17-C18
3	C	402	HFk	N14-C15-N17-C18
4	A	403	HR5	C3-C2-C5-O
4	A	403	HR5	C3-C2-C5-N2

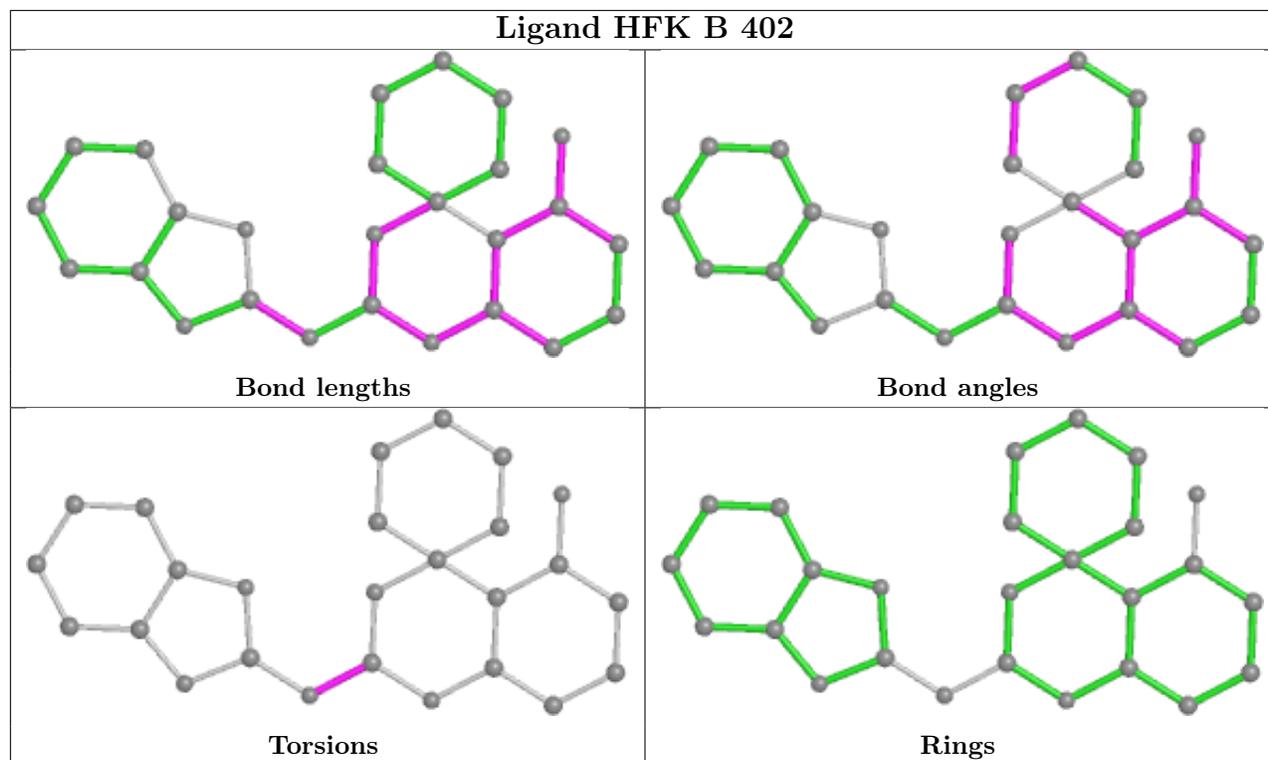
There are no ring outliers.

3 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	403	HR5	1	0
4	C	404	HR5	2	0
3	B	402	HFk	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	392/392 (100%)	-0.89	1 (0%) 94 93	24, 35, 60, 127	0
1	B	387/392 (98%)	-0.46	6 (1%) 72 70	28, 45, 86, 122	0
1	C	386/392 (98%)	-0.62	1 (0%) 94 93	28, 46, 76, 125	0
1	D	376/392 (95%)	-0.41	1 (0%) 94 93	28, 47, 82, 118	0
All	All	1541/1568 (98%)	-0.60	9 (0%) 89 88	24, 43, 81, 127	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	364	ALA	5.3
1	B	6	GLN	4.1
1	B	334	ALA	2.7
1	B	7	PRO	2.6
1	B	371	HIS	2.5

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](#)

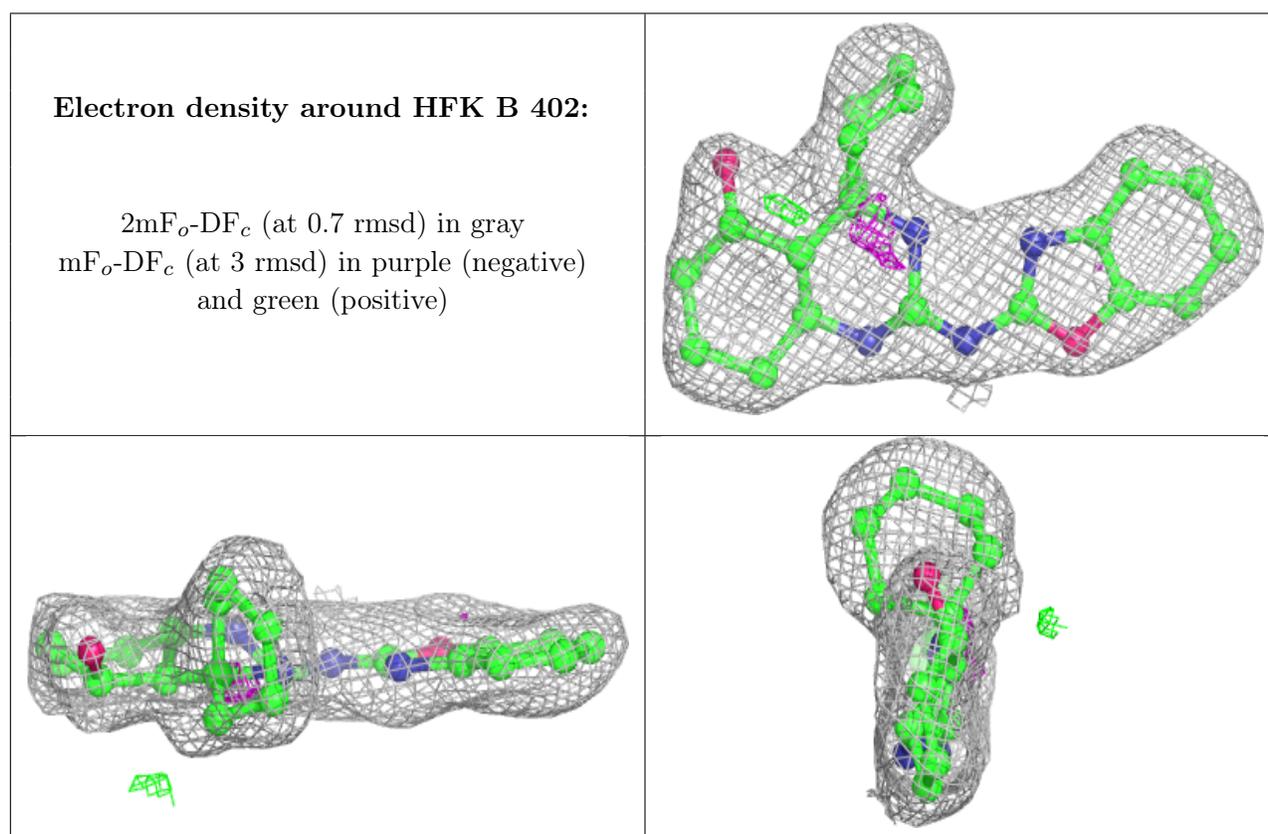
There are no monosaccharides in this entry.

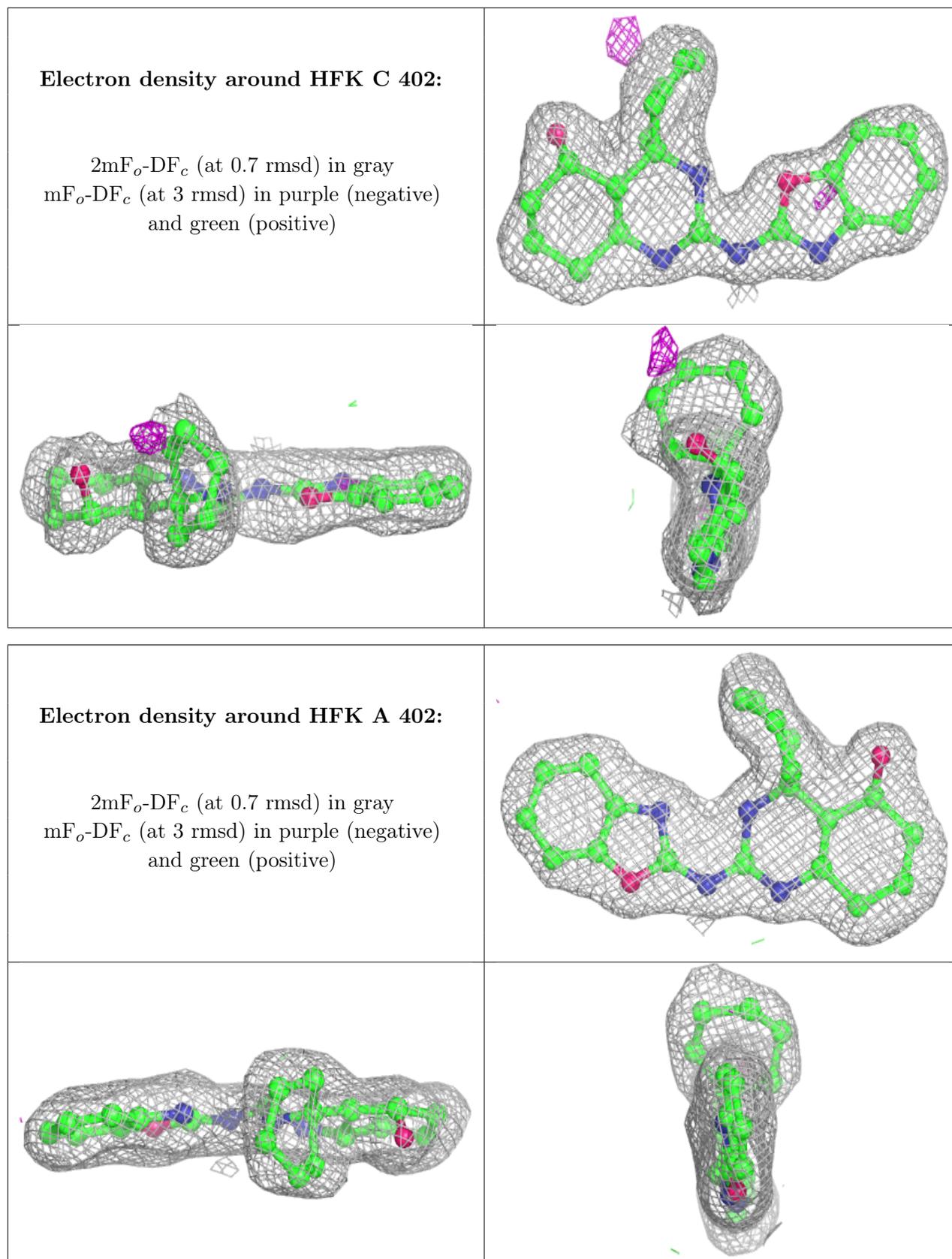
6.4 Ligands [i](#)

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, 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
4	HR5	C	404	15/15	0.70	0.24	53,61,64,66	0
4	HR5	A	403	15/15	0.74	0.26	41,65,75,76	0
3	HFK	B	402	26/26	0.82	0.17	41,46,52,61	0
3	HFK	C	402	26/26	0.83	0.18	45,52,58,60	0
2	GAL	C	401	12/12	0.86	0.15	31,36,43,52	0
4	HR5	C	403	15/15	0.89	0.17	43,50,59,65	0
3	HFK	A	402	26/26	0.94	0.11	28,36,44,47	0
2	GAL	B	401	12/12	0.95	0.10	34,36,39,48	0
2	GAL	A	401	12/12	0.96	0.09	22,25,27,31	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.