W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 15, 2024 – 03:43 pm BST

PDB ID	:	8QUN
Title	:	human carbonic anhydrase I in complex with 2-((5-cyano-4-hydroxy-6-(p-tolyl)pyrimidin-2-yl)thio)-N-(4-sulfamoylphenyl)acetamide
Authors	:	Angeli, A.; Ferraroni, M.
Deposited on	:	2023-10-16
Resolution	:	1.61 Å(reported)

The following versions of software and data (see references ⓘ) were used in the production of this report:

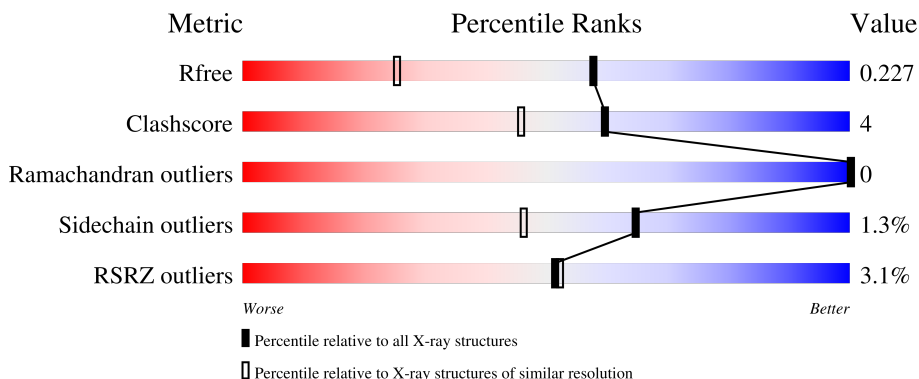|		
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	3.0
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

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 1.61 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	Percentile Ranks	Value
Rfree		0.227
Clashscore		4
Ramachandran outliers		0
Sidechain outliers		1.3%
RSRZ outliers		3.1%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	4274 (1.60-1.60)
Clashscore	180529	4682 (1.60-1.60)
Ramachandran outliers	177936	4583 (1.60-1.60)
Sidechain outliers	177891	4582 (1.60-1.60)
RSRZ outliers	164620	4272 (1.60-1.60)

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	261	3% 90% 8% ••
1	BBB	261	3% 92% 6% ••

WORLDWIDE
wwPDB
PROTEIN DATA BANK

2 Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 4451 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 Carbonic anhydrase 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAA	257	Total 2025	C 1280	N 350	O 392	S 3	0	3	0
1	BBB	257	Total 2022	C 1277	N 350	O 392	S 3	0	3	0

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	AAA	1	Total 1	Zn 1	0	0
2	BBB	1	Total 1	Zn 1	0	0

- Molecule 3 is 2-[5-cyano-4-(4-methylphenyl)-6-oxidanyl-pyrimidin-2-yl]sulfanyl-N-(4-sulfamoylphenyl)ethanamide (three-letter code: WZH) (formula: $C_{20}H_{17}N_5O_4S_2$) (labeled as "Ligand of Interest" by depositor).
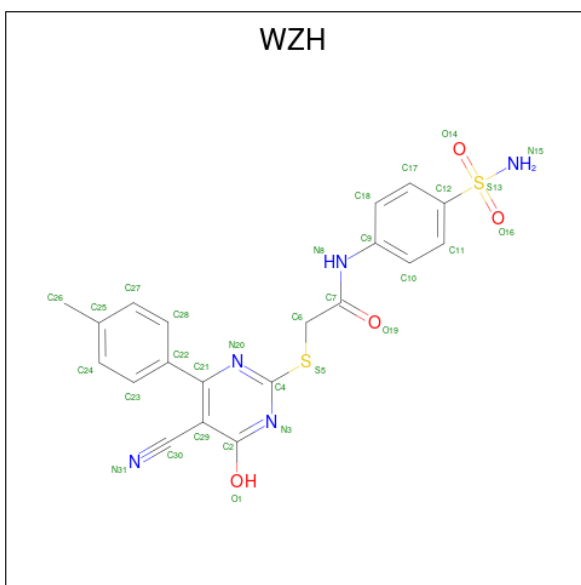
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	AAA	1	Total 31	C 20	N 5	O 4	S 2	0	0
3	BBB	1	Total 31	C 20	N 5	O 4	S 2	0	0

- Molecule 4 is DIMETHYL SULFOXIDE (three-letter code: DMS) (formula: C_2H_6OS).
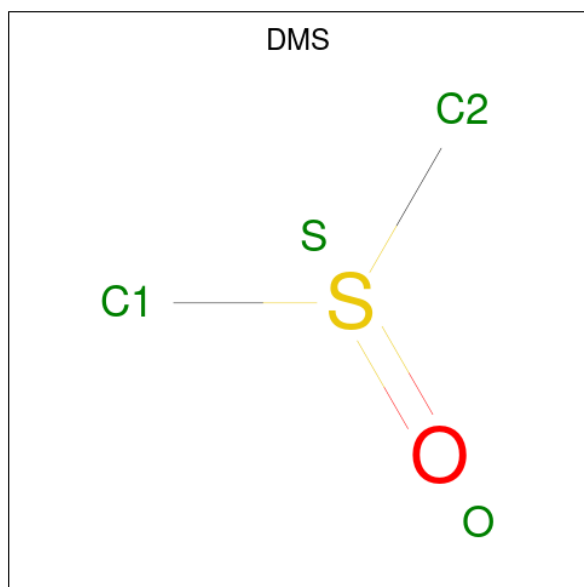
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	BBB	1	Total 4	C 2	O 1	S 1	0	0

- Molecule 5 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	BBB	1	Total 1	Na 1	0	0

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	AAA	165	Total 165	O 165	0	0
6	BBB	170	Total 170	O 170	0	0

3 Residue-property plots ⓘ

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.
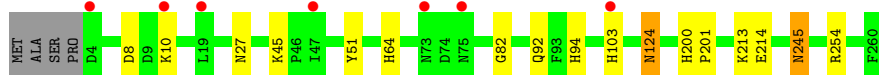
• Molecule 1: Carbonic anhydrase 1

Chain AAA:

• Molecule 1: Carbonic anhydrase 1

Chain BBB:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	62.78Å 71.39Å 121.08Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	60.61 – 1.61 60.61 – 1.61	Depositor EDS
% Data completeness (in resolution range)	99.9 (60.61-1.61) 99.9 (60.61-1.61)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.46 (at 1.61Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.184 , 0.215 0.197 , 0.227	Depositor DCC
R_{free} test set	3477 reflections (4.85%)	wwPDB-VP
Wilson B-factor (Å^2)	22.2	Xtriage
Anisotropy	0.143	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 27.7	EDS
L-test for twinning[2]	$< |L| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	4451	wwPDB-VP
Average B, all atoms (Å^2)	26.0	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 21.48 % 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 6.9492e-03. 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: DMS, ZN, WZH, NA

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	0.77	2/2091 (0.1%)	0.86	0/2843
1	BBB	0.76	0/2088	0.88	1/2839 (0.0%)
All	All	0.76	2/4179 (0.0%)	0.87	1/5682 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	AAA	106	GLU	CD-OE1	-5.22	1.20	1.25
1	AAA	214	GLU	CD-OE1	-5.05	1.20	1.25

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	BBB	51	TYR	CB-CG-CD2	-5.08	117.95	121.00

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	2025	0	1954	16	0
1	BBB	2022	0	1945	15	0
2	AAA	1	0	0	0	0

Continued on next page...

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	BBB	1	0	0	0	0
3	AAA	31	0	0	1	0
3	BBB	31	0	0	1	0
4	BBB	4	0	6	1	0
5	BBB	1	0	0	0	0
6	AAA	165	0	0	3	0
6	BBB	170	0	0	5	0
All	All	4451	0	3905	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 4.

The worst 5 of 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:BBB:103:HIS:O	1:BBB:103:HIS:CG	2.22	0.91
1:AAA:92:GLN:HE21	1:AAA:94:HIS:HD1	1.34	0.75
1:AAA:245:ASN:HD22	1:AAA:245:ASN:H	1.41	0.69
1:BBB:245:ASN:H	1:BBB:245:ASN:HD22	1.43	0.67
1:BBB:213:LYS:CE	1:BBB:214:GLU:OE2	2.44	0.66

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	258/261 (99%)	251 (97%)	7 (3%)	0	100	100
1	BBB	258/261 (99%)	249 (96%)	9 (4%)	0	100	100
All	All	516/522 (99%)	500 (97%)	16 (3%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains ⓘ

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	226/226 (100%)	222 (98%)	4 (2%)	54	31
1	BBB	225/226 (100%)	223 (99%)	2 (1%)	75	62
All	All	451/452 (100%)	445 (99%)	6 (1%)	65	46

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	AAA	245	ASN
1	BBB	124	ASN
1	BBB	245	ASN
1	AAA	14	GLU
1	AAA	4	ASP

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

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 6 ligands modelled in this entry, 3 are 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	#$|Z| > 2$
3	WZH	BBB	304	2	33,33,33	0.70	1 (3%)	45,47,47	0.99	3 (6%)
3	WZH	AAA	302	2	33,33,33	0.81	2 (6%)	45,47,47	0.69	0
4	DMS	BBB	301	-	3,3,3	0.23	0	3,3,3	0.31	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
3	WZH	BBB	304	2	-	6/20/21/21	0/3/3/3
3	WZH	AAA	302	2	-	7/20/21/21	0/3/3/3

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	AAA	302	WZH	C6-S5	2.76	1.87	1.80
3	AAA	302	WZH	S13-N15	-2.24	1.56	1.60
3	BBB	304	WZH	C6-S5	2.23	1.85	1.80

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	BBB	304	WZH	O16-S13-O14	-3.37	113.22	118.76
3	BBB	304	WZH	C6-S5-C4	3.17	105.72	101.63
3	BBB	304	WZH	C12-S13-N15	2.32	111.67	108.38

There are no chirality outliers.

5 of 13 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	AAA	302	WZH	C21-C29-C30-N31
3	AAA	302	WZH	C17-C12-S13-O16

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
3	AAA	302	WZH	C11-C12-S13-O16
3	BBB	304	WZH	C11-C12-S13-O16
3	AAA	302	WZH	C11-C12-S13-N15

There are no ring outliers.

3 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	BBB	304	WZH	1	0
3	AAA	302	WZH	1	0
4	BBB	301	DMS	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 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 WZH BBB 304

Bond lengths   Bond angles

Torsions   Rings

Ligand WZH AAA 302

Bond lengths

Bond angles

Torsions

Rings
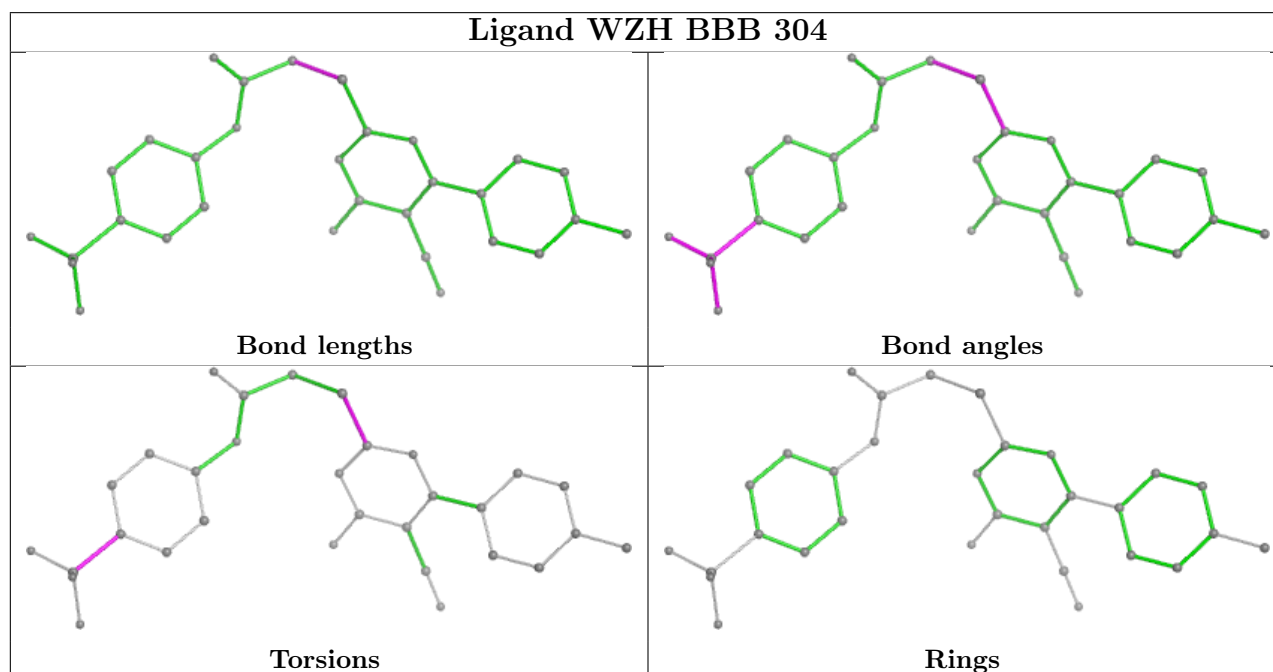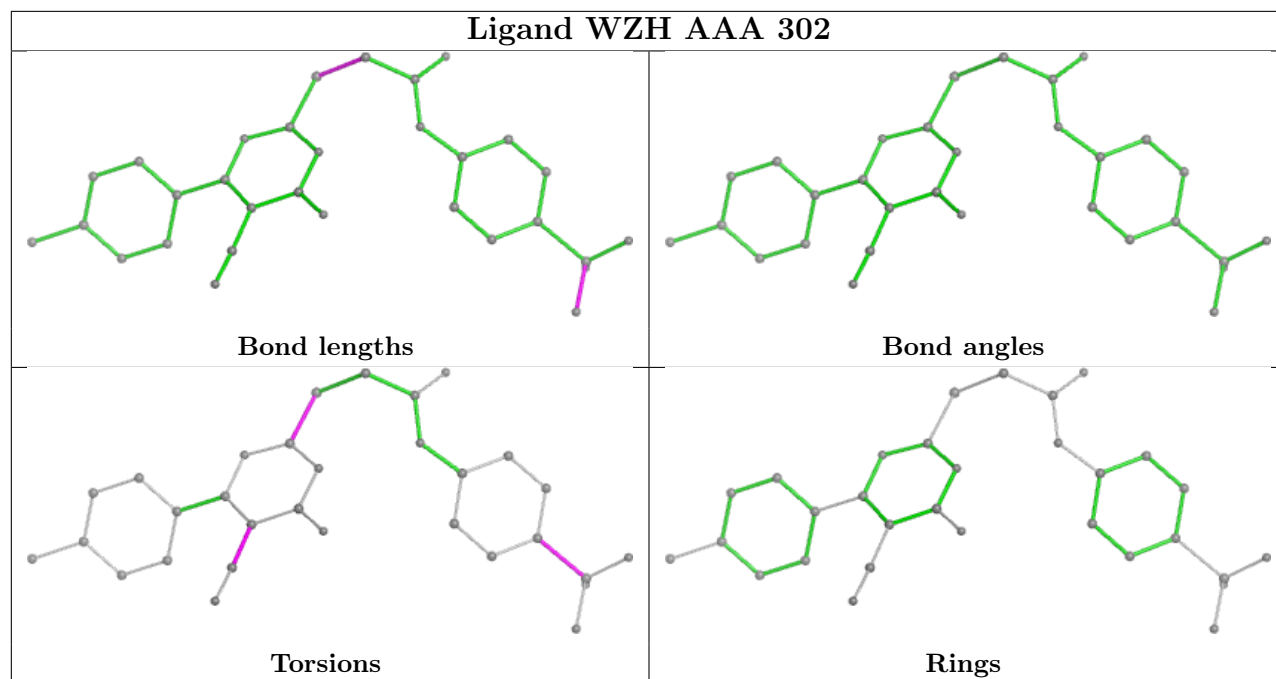
5.7 Other polymers ⓘ

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, 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(Å²)	Q<0.9
1	AAA	257/261 (98%)	0.15	9 (3%)	47	48	13, 25, 42, 55	3 (1%)
1	BBB	257/261 (98%)	0.09	7 (2%)	56	57	13, 24, 39, 58	3 (1%)
All	All	514/522 (98%)	0.12	16 (3%)	51	52	13, 25, 41, 58	6 (1%)

The worst 5 of 16 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	BBB	103	HIS	4.3
1	AAA	19	LEU	3.2
1	BBB	73	ASN	2.8
1	AAA	103	HIS	2.5
1	AAA	15	GLN	2.3

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

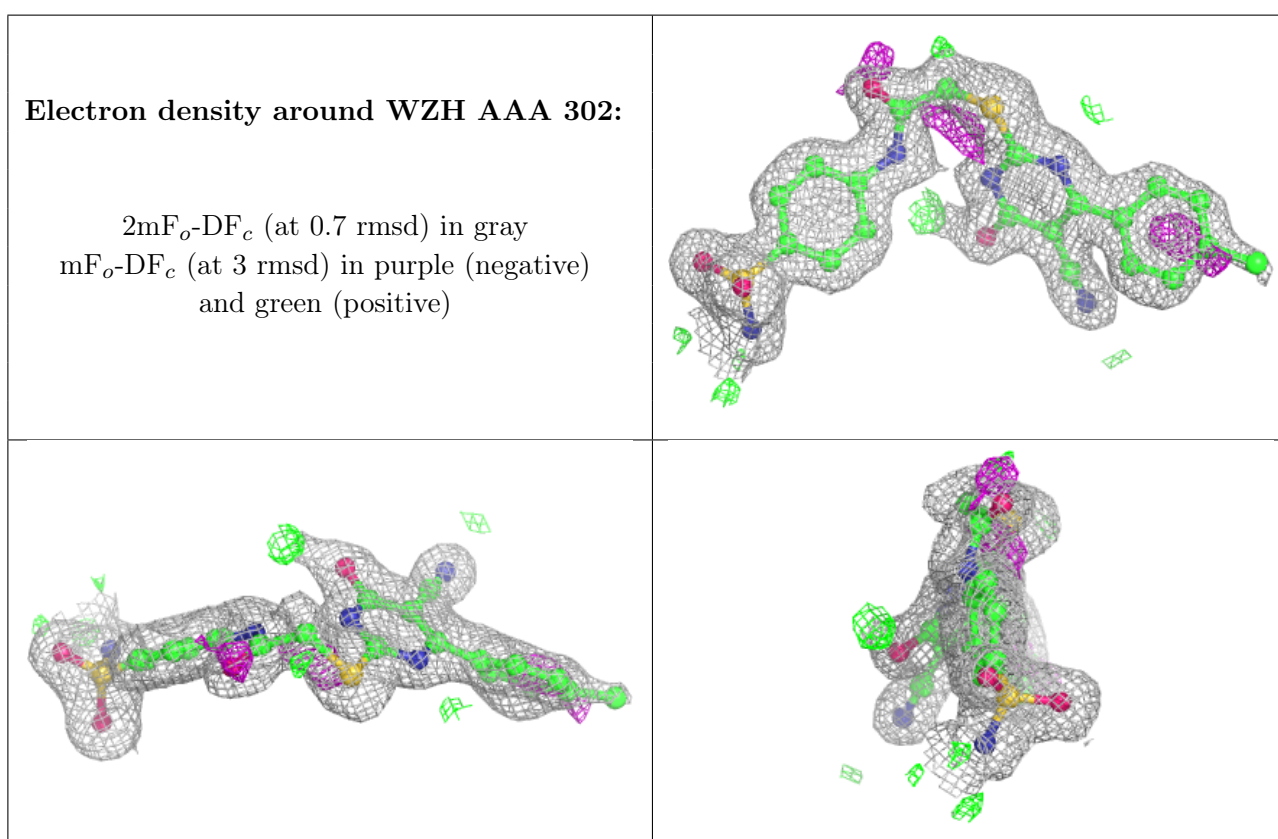There are no non-standard protein/DNA/RNA residues in this entry.

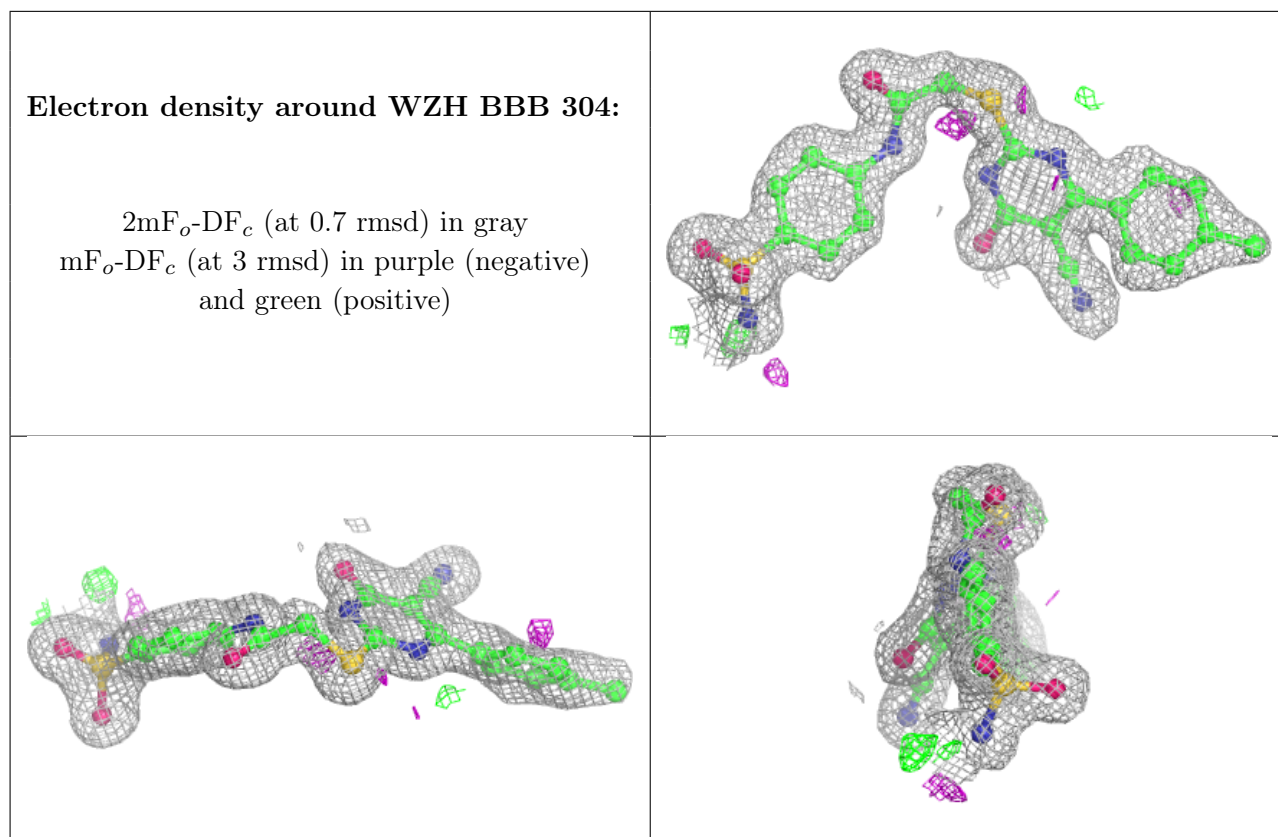## 6.3 Carbohydrates ⓘ

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, 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($\mathring{A}^2$)	Q<0.9
4	DMS	BBB	301	4/4	0.80	0.24	67,71,78,78	0
5	NA	BBB	303	1/1	0.94	0.23	34,34,34,34	0
3	WZH	AAA	302	31/31	0.97	0.08	18,28,36,38	0
3	WZH	BBB	304	31/31	0.97	0.08	18,31,37,40	0
2	ZN	BBB	302	1/1	0.99	0.02	15,15,15,15	0
2	ZN	AAA	301	1/1	1.00	0.01	17,17,17,17	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 WZH AAA 302:

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

Electron density around WZH BBB 304:

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

6.5 Other polymers ⓘ

There are no such residues in this entry.