wwPDB X-ray Structure Validation Summary Report (i)

Jan 2, 2024 – 10:38 pm GMT

PDB ID	:	5IPZ
Title	:	Crystal structure of human carbonic anhydrase isozyme IV with 5-(2-amino-1 ,3-thiazol-4-yl)-2-chlorobenzenesulfonamide
Authors	:	Smirnov, A.; Manakova, E.; Grazulis, S.
Deposited on	:	2016-03-10
Resolution	:	2.10 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:

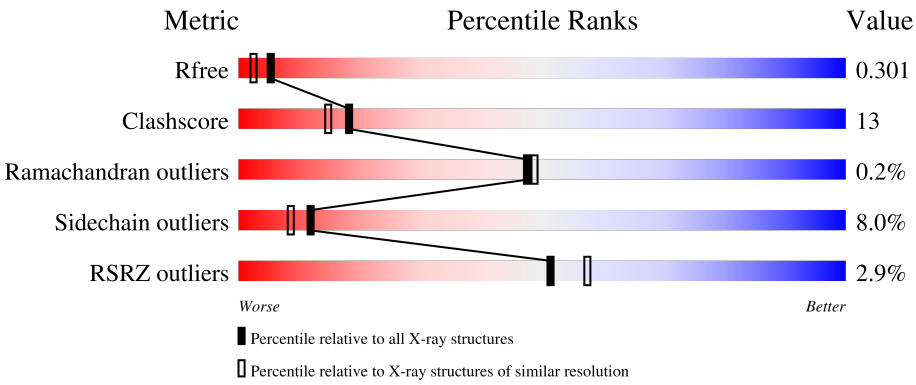
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.10 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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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	A	267	2% 78% 19% • •
1	B	267	2% 78% 16% • •
1	C	267	4% 59% 31% 5% •
1	D	267	3% 60% 30% 5% 5%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 8427 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 4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	261	2115	1339	363	402	11	0	2	0
1	B	256	2062	1310	350	391	11	0	1	0
1	C	255	2043	1297	350	385	11	0	0	0
1	D	253	2032	1294	348	379	11	0	0	0

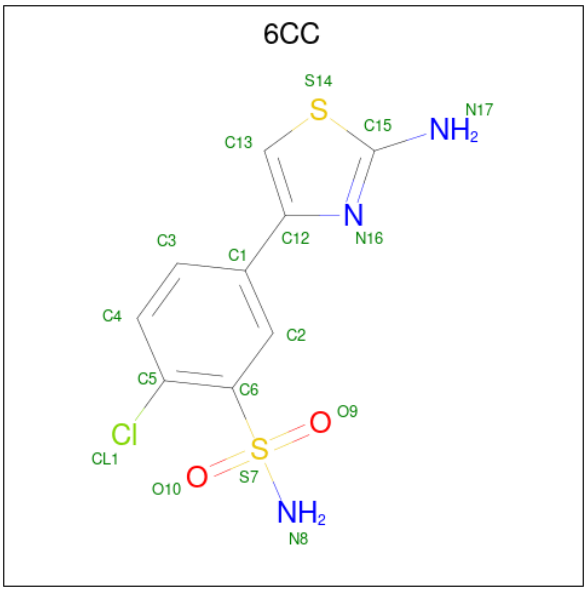
There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP P22748
B	1	MET	-	initiating methionine	UNP P22748
C	1	MET	-	initiating methionine	UNP P22748
D	1	MET	-	initiating methionine	UNP P22748

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Zn		
2	A	1	1	1	0	0
2	B	1	1	1	0	0
2	C	1	1	1	0	0
2	D	1	1	1	0	0

- Molecule 3 is 5-(2-amino-1,3-thiazol-4-yl)-2-chlorobenzene-1-sulfonamide (three-letter code: 6CC) (formula: $C_9H_8ClN_3O_2S_2$).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
			Total	C	Cl	N	O	S		
3	A	1	17	9	1	3	2	2	0	0
			Total	C	Cl	N	O	S		
3	B	1	17	9	1	3	2	2	0	0
			Total	C	Cl	N	O	S		
3	C	1	17	9	1	3	2	2	0	0

- Molecule 4 is water.

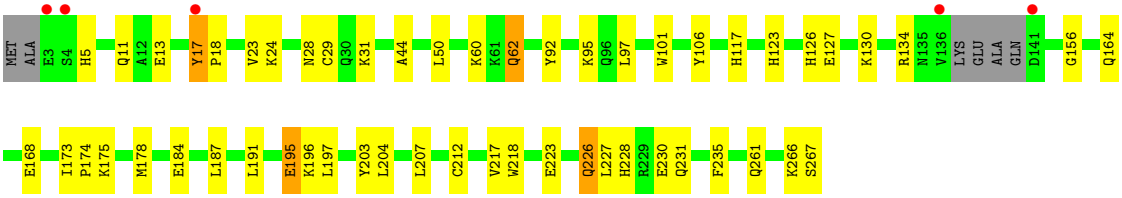
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	51	51	51	0	0
			Total	O		
4	B	45	45	45	0	0
			Total	O		
4	C	13	13	13	0	0
			Total	O		
4	D	11	11	11	0	0

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.

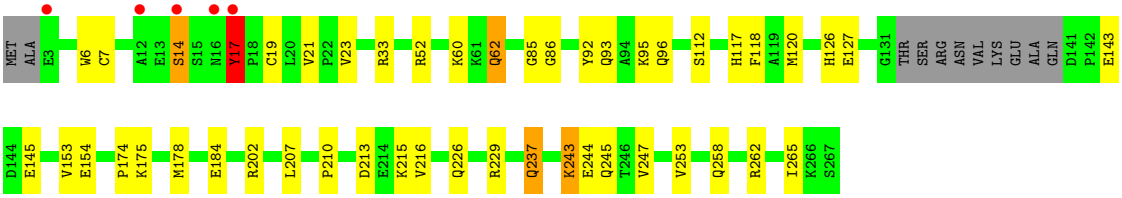
- Molecule 1: Carbonic anhydrase 4

Chain A:



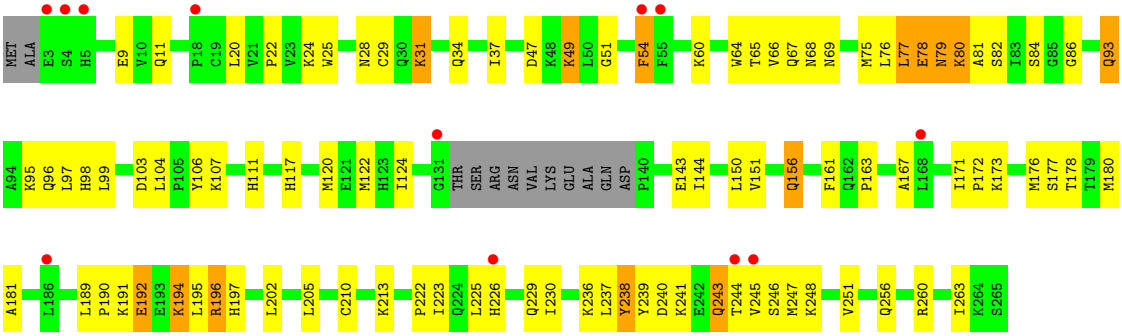
- Molecule 1: Carbonic anhydrase 4

Chain B:



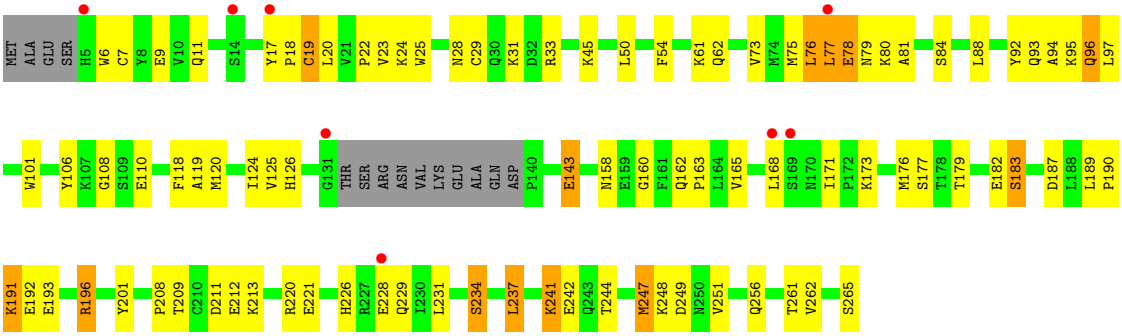
- Molecule 1: Carbonic anhydrase 4

Chain C:



- Molecule 1: Carbonic anhydrase 4

Chain D:


MET
ALA
GLU
SER
H5
W6
C7
Y8
E9
V10
Q11
S14
Y17
P18
C19
L20
V21
P22
V23
K24
W25
N28
C29
Q30
K31
D32
R33
K45
L50
F54
K61
Q62
V73
M74
M75
L76
L77
E78
N79
K80
A81
S84
L88
Y92
Q93
A94
K95
Q96
L97
W101
Y106
K107
G108
S109
E110
F118
A119
M120
I124
V125
H126
G131
THR
SER
ARG
ASN
VAL
LYS
GLU
ALA
GLN
ASP
P140
E143
N158
E159
G160
F161
Q162
P163
L164
V165
L168
S169
N170
I171
P172
K173
M176
S177
T178
T179
E182
S183
D187
L188
L189
P190
K191
E192
E193
R196
Y201
P208
T209
C210
D211
E212
K213
R220
E221
H226
R227
E228
Q229
I230
L231
S234
L237
K241
E242
Q243
T244
M247
K248
D249
N250
V251
Q256
T261
V262
S265

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	48.69Å 256.41Å 48.55Å 90.00° 117.34° 90.00°	Depositor
Resolution (Å)	43.13 – 2.10 43.13 – 2.10	Depositor EDS
% Data completeness (in resolution range)	84.8 (43.13-2.10) 84.8 (43.13-2.10)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.05	Depositor
$< I/\sigma(I) >$ [1]	1.89 (at 2.10Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, R_{free}	0.225 , 0.306 0.222 , 0.301	Depositor DCC
R_{free} test set	5105 reflections (9.83%)	wwPDB-VP
Wilson B-factor (Å^2)	42.3	Xtriage
Anisotropy	0.134	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 44.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.009 for l,k,-h-l 0.009 for -h-l,k,h 0.045 for -h-l,-k,l 0.042 for h,-k,-h-l 0.168 for l,-k,h	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	8427	wwPDB-VP
Average B, all atoms (Å^2)	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.46% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, 6CC

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	0.73	0/2165	0.85	0/2926
1	B	0.71	0/2112	0.82	0/2855
1	C	0.66	0/2092	0.86	3/2826 (0.1%)
1	D	0.67	0/2082	0.79	1/2813 (0.0%)
All	All	0.70	0/8451	0.83	4/11420 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	B	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	C	78	GLU	N-CA-CB	-13.41	86.46	110.60
1	C	77	LEU	N-CA-C	-7.91	89.63	111.00
1	C	81	ALA	N-CA-CB	-5.12	102.93	110.10
1	D	191	LYS	CB-CA-C	5.07	120.54	110.40

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	14	SER	Peptide

5.2 Too-close contacts (i)

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	2115	0	2073	31	0
1	B	2062	0	2019	35	0
1	C	2043	0	2013	87	0
1	D	2032	0	2005	63	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
3	A	17	0	0	3	0
3	B	17	0	0	0	0
3	C	17	0	0	1	0
4	A	51	0	0	0	0
4	B	45	0	0	0	0
4	C	13	0	0	0	0
4	D	11	0	0	1	0
All	All	8427	0	8110	214	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 13.

The worst 5 of 214 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:C:78:GLU:H	1:C:79:ASN:CB	1.32	1.41
1:C:78:GLU:CB	1:C:79:ASN:HB2	1.70	1.20
1:C:78:GLU:N	1:C:79:ASN:HB2	1.57	1.19
1:C:78:GLU:H	1:C:79:ASN:HB3	1.04	1.15
1:C:78:GLU:N	1:C:79:ASN:CB	2.10	1.13

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	A	259/267 (97%)	245 (95%)	14 (5%)	0	100	100
1	B	253/267 (95%)	234 (92%)	18 (7%)	1 (0%)	34	32
1	C	251/267 (94%)	214 (85%)	36 (14%)	1 (0%)	34	32
1	D	249/267 (93%)	215 (86%)	34 (14%)	0	100	100
All	All	1012/1068 (95%)	908 (90%)	102 (10%)	2 (0%)	47	49

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	79	ASN
1	B	17	TYR

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	236/238 (99%)	222 (94%)	14 (6%)	19	17
1	B	229/238 (96%)	216 (94%)	13 (6%)	20	18
1	C	227/238 (95%)	204 (90%)	23 (10%)	7	4
1	D	225/238 (94%)	200 (89%)	25 (11%)	6	3
All	All	917/952 (96%)	842 (92%)	75 (8%)	12	8

5 of 75 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	93	GLN
1	D	241	LYS
1	D	143	GLU
1	D	183	SER
1	B	243	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 19 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	93	GLN
1	D	226	HIS
1	D	256	GLN
1	D	162	GLN
1	B	237	GLN

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, 4 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	6CC	A	302	2	15,18,18	3.09	8 (53%)	22,27,27	3.46	8 (36%)
3	6CC	B	302	-	15,18,18	3.19	6 (40%)	22,27,27	2.84	10 (45%)
3	6CC	C	302	2	15,18,18	3.62	6 (40%)	22,27,27	3.33	8 (36%)

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	6CC	A	302	2	-	1/10/10/10	0/2/2/2
3	6CC	B	302	-	-	0/10/10/10	0/2/2/2
3	6CC	C	302	2	-	6/10/10/10	0/2/2/2

The worst 5 of 20 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	302	6CC	S7-N8	9.23	1.78	1.60
3	B	302	6CC	S7-N8	8.95	1.78	1.60
3	A	302	6CC	O9-S7	7.52	1.57	1.43
3	C	302	6CC	O10-S7	6.57	1.55	1.43
3	C	302	6CC	O9-S7	6.19	1.55	1.43

The worst 5 of 26 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	302	6CC	O10-S7-O9	-10.10	102.15	118.76
3	C	302	6CC	O10-S7-O9	-7.68	106.13	118.76
3	A	302	6CC	C4-C5-C6	-7.16	115.16	120.05
3	B	302	6CC	O10-S7-O9	-7.10	107.09	118.76
3	C	302	6CC	C13-C12-C1	-6.62	120.25	129.44

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	C	302	6CC	C3-C1-C12-C13
3	C	302	6CC	C3-C1-C12-N16
3	C	302	6CC	C2-C1-C12-N16
3	C	302	6CC	C2-C1-C12-C13
3	C	302	6CC	C2-C6-S7-O9

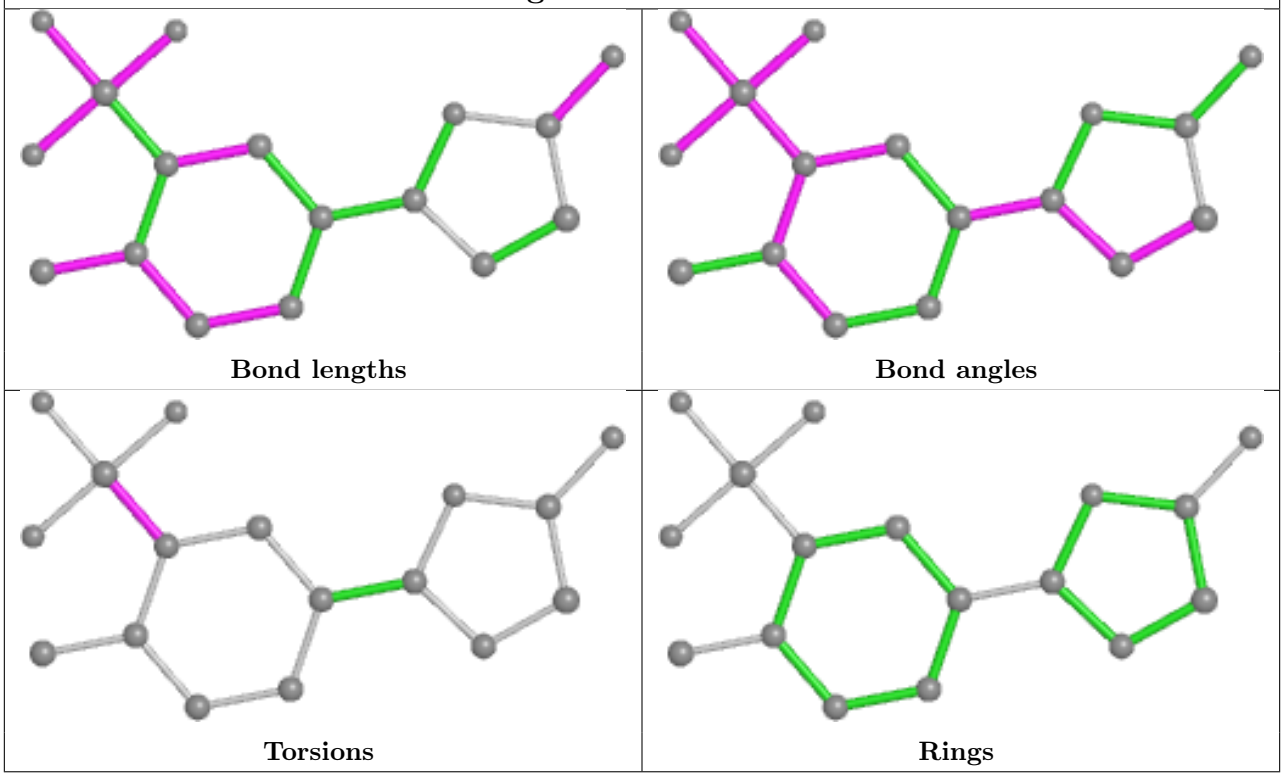
There are no ring outliers.

2 monomers are involved in 4 short contacts:

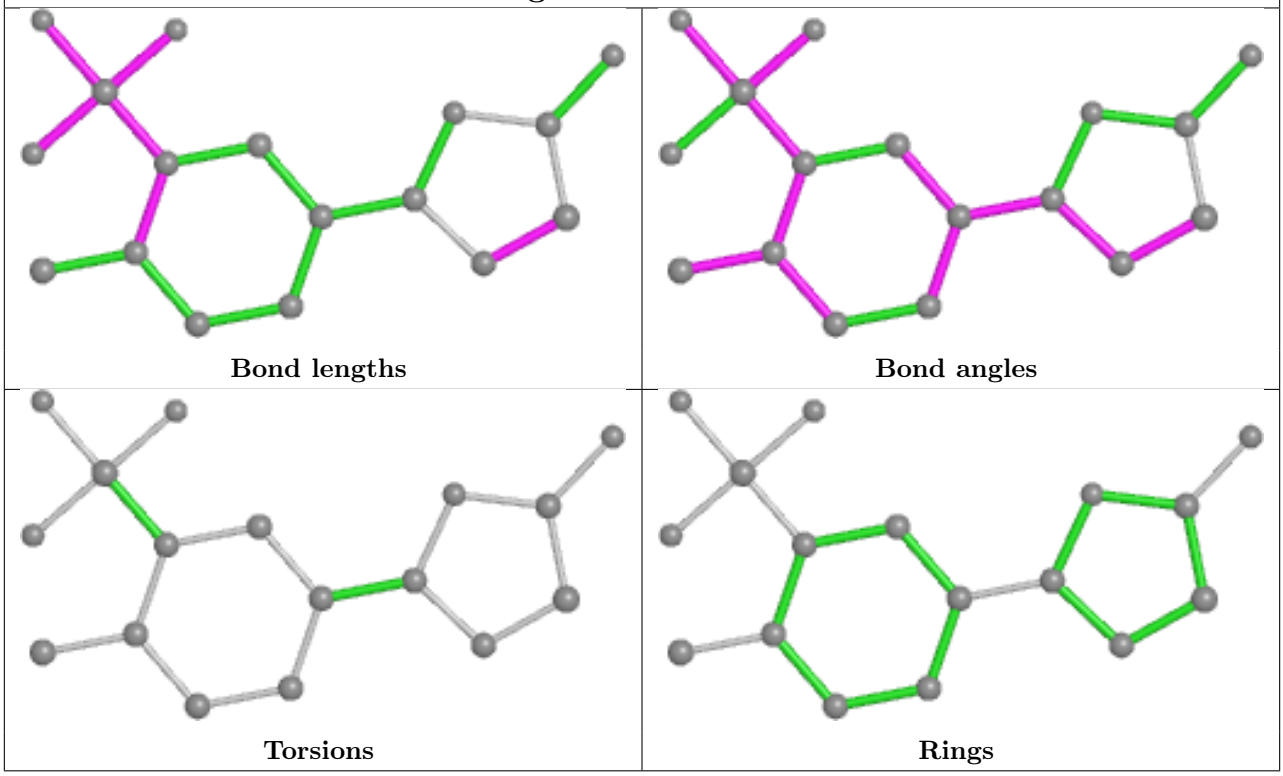
Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	302	6CC	3	0
3	C	302	6CC	1	0

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.

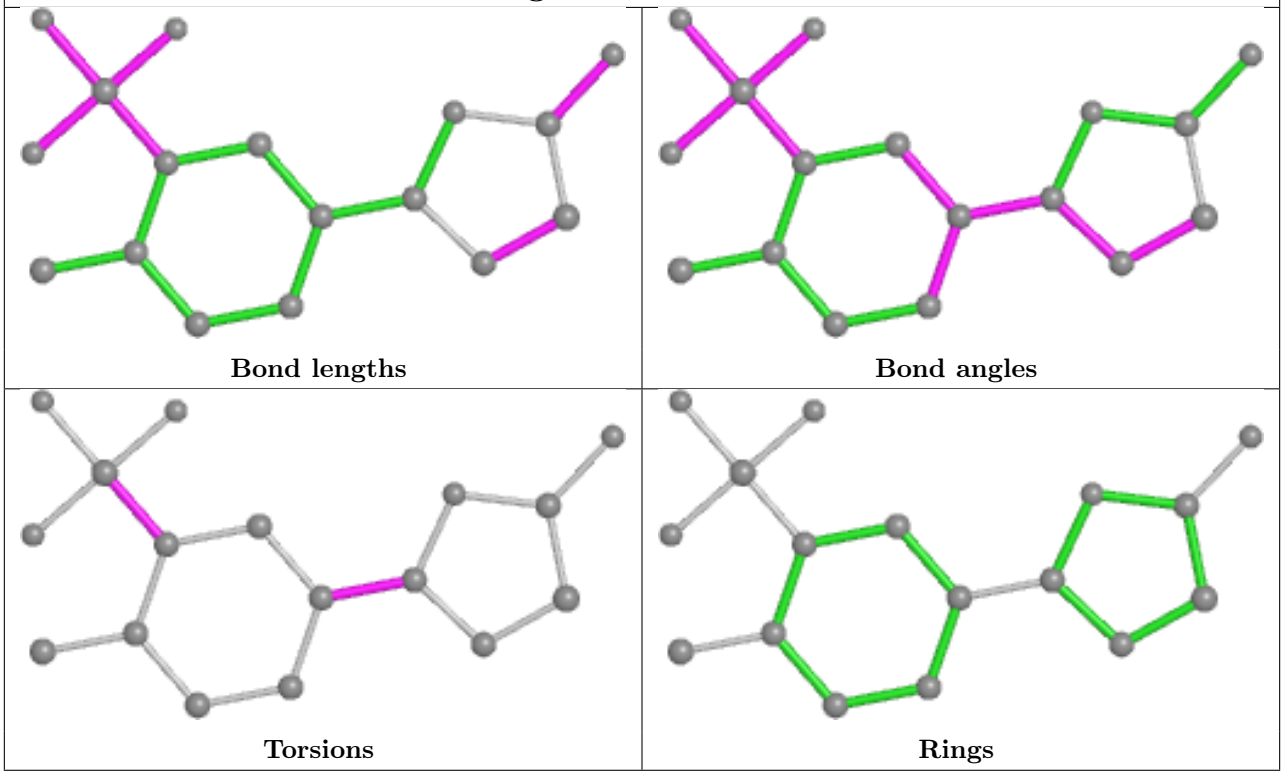
Ligand 6CC A 302



Ligand 6CC B 302



Ligand 6CC C 302



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	261/267 (97%)	-0.19	5 (1%)	66	71	20, 38, 61, 97	0
1	B	256/267 (95%)	-0.16	5 (1%)	65	69	23, 41, 63, 96	0
1	C	255/267 (95%)	0.32	12 (4%)	31	37	30, 61, 93, 120	0
1	D	253/267 (94%)	0.27	8 (3%)	47	54	29, 59, 82, 109	0
All	All	1025/1068 (95%)	0.06	30 (2%)	51	57	20, 47, 82, 120	0

The worst 5 of 30 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	18	PRO	6.9
1	B	17	TYR	4.5
1	B	16	ASN	4.2
1	D	5	HIS	3.9
1	A	3	GLU	3.6

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)

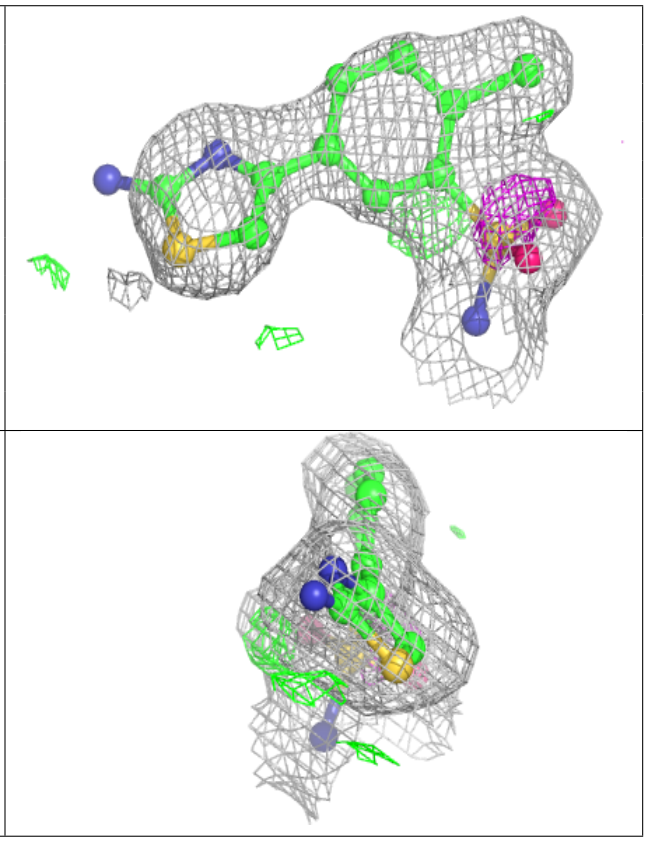
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists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

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	6CC	C	302	17/17	0.94	0.14	33,63,93,98	0
3	6CC	A	302	17/17	0.97	0.12	26,35,52,53	0
3	6CC	B	302	17/17	0.98	0.11	24,31,47,57	0
2	ZN	D	301	1/1	0.98	0.13	36,36,36,36	0
2	ZN	B	301	1/1	1.00	0.12	26,26,26,26	0
2	ZN	C	301	1/1	1.00	0.11	42,42,42,42	0
2	ZN	A	301	1/1	1.00	0.13	23,23,23,23	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.

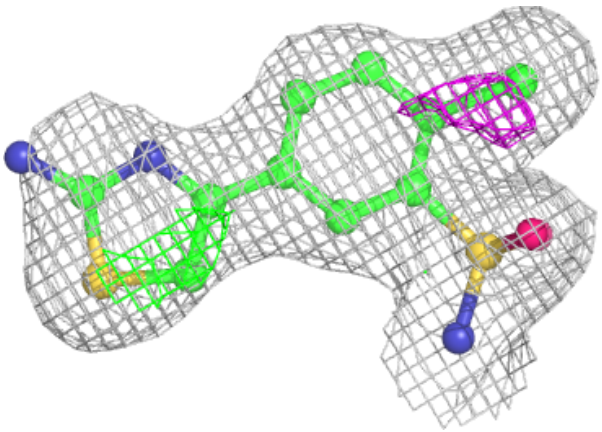
Electron density around 6CC C 302:

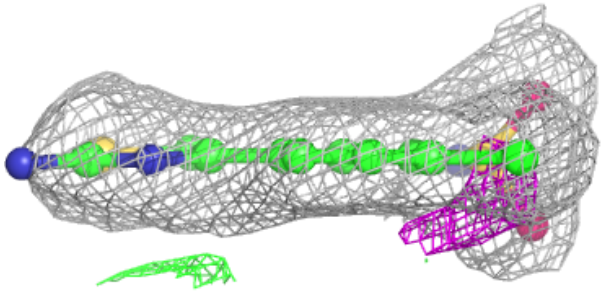
2mF$_o$-DF$_c$ (at 0.7 rmsd) in gray
mF$_o$-DF$_c$ (at 3 rmsd) in purple (negative) and green (positive)

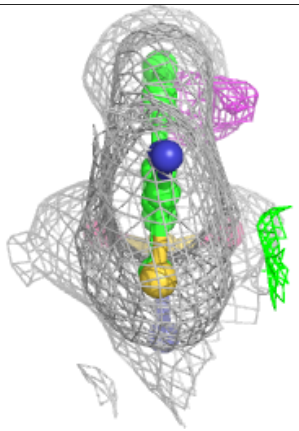


Electron density around 6CC A 302:

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

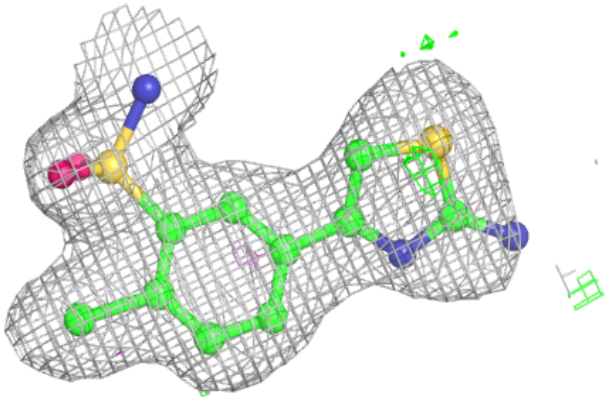


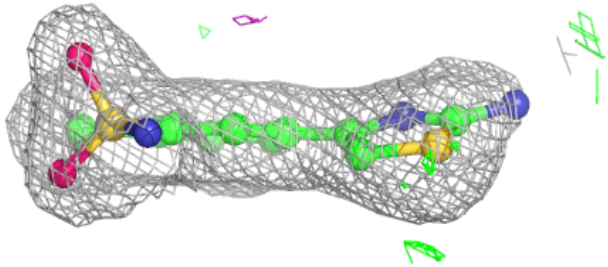


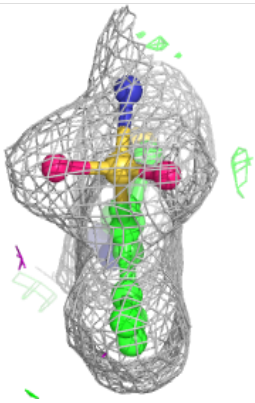


Electron density around 6CC B 302:

$2mF_o$-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.