wwPDB X-ray Structure Validation Summary Report (i)

Nov 15, 2022 – 12:52 pm GMT

PDB ID	:	6YYN
Title	:	Structure of Cathepsin S in complex with Compound 14
Authors	:	Wagener, M.; Schade, M.; Merla, B.; Hars, U.; Kueckelhaus, S.Q.
Deposited on	:	2020-05-05
Resolution	:	2.22 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:

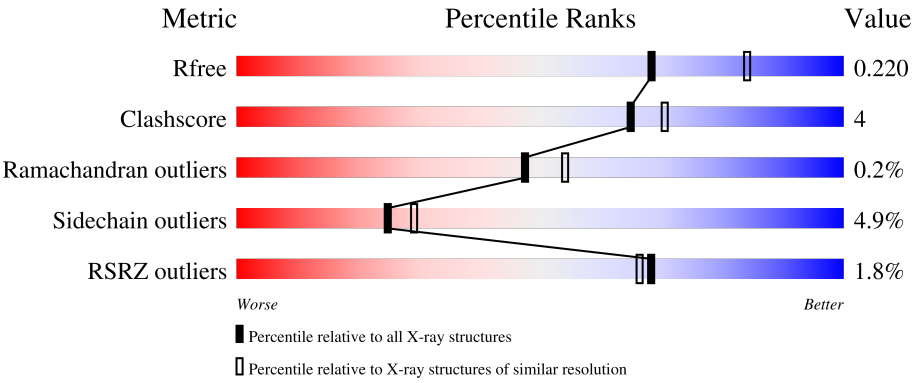
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.31.2
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0267
CCP4	:	7.1.010 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.2

1 Overall quality at a glance (i)

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

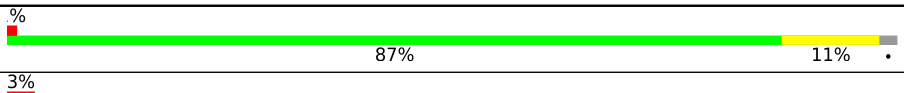
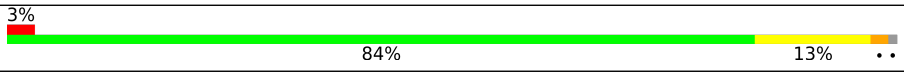
The reported resolution of this entry is 2.22 Å.

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	5912 (2.24-2.20)
Clashscore	141614	6646 (2.24-2.20)
Ramachandran outliers	138981	6543 (2.24-2.20)
Sidechain outliers	138945	6544 (2.24-2.20)
RSRZ outliers	127900	5797 (2.24-2.20)

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	225	
1	BBB	225	

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
4	SO4	BBB	1001	-	-	-	X
4	SO4	BBB	1006	-	-	-	X

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 4029 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.

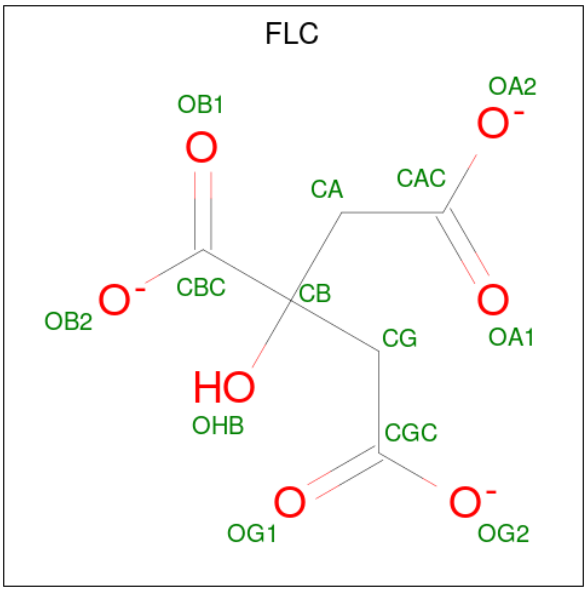
- Molecule 1 is a protein called Cathepsin S.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAA	220	Total	C	N	O	S	0	0	0
			1713	1082	293	326	12			
1	BBB	222	Total	C	N	O	S	0	0	0
			1734	1094	300	328	12			

There are 12 discrepancies between the modelled and reference sequences:

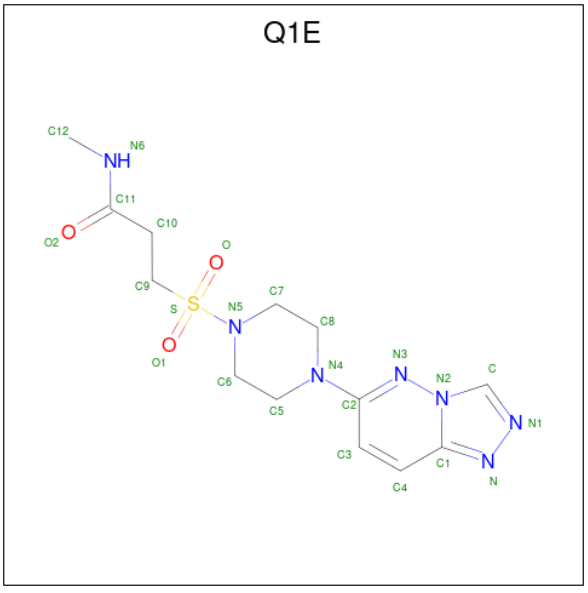
Chain	Residue	Modelled	Actual	Comment	Reference
AAA	218	HIS	-	expression tag	UNP P25774
AAA	219	HIS	-	expression tag	UNP P25774
AAA	220	HIS	-	expression tag	UNP P25774
AAA	221	HIS	-	expression tag	UNP P25774
AAA	222	HIS	-	expression tag	UNP P25774
AAA	223	HIS	-	expression tag	UNP P25774
BBB	218	HIS	-	expression tag	UNP P25774
BBB	219	HIS	-	expression tag	UNP P25774
BBB	220	HIS	-	expression tag	UNP P25774
BBB	221	HIS	-	expression tag	UNP P25774
BBB	222	HIS	-	expression tag	UNP P25774
BBB	223	HIS	-	expression tag	UNP P25774

- Molecule 2 is CITRATE ANION (three-letter code: FLC) (formula: $C_6H_5O_7$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	AAA	1	Total C O 13 6 7	0	0
2	AAA	1	Total C O 13 6 7	0	0

- Molecule 3 is {N}-methyl-3-[4-([1,2,4]triazolo[4,3-b]pyridazin-6-yl)piperazin-1-yl]sulfonyl-propanamide (three-letter code: Q1E) (formula: $C_{13}H_{19}N_7O_3S$) (labeled as "Ligand of Interest" by depositor).



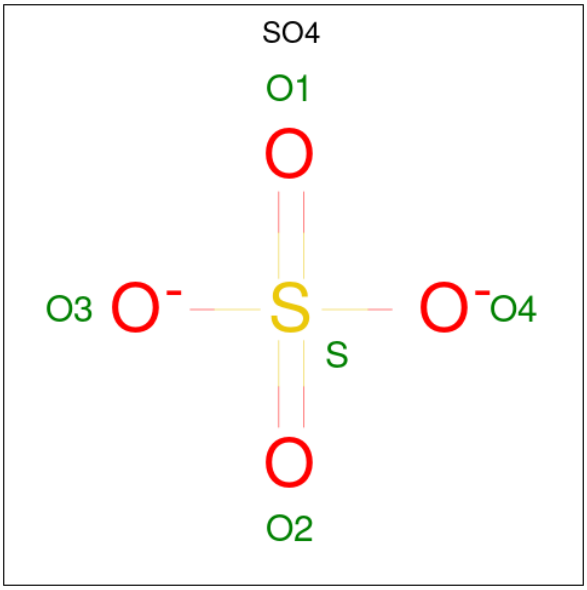
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	AAA	1	Total C N O S 24 13 7 3 1	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	BBB	1	Total	C	N	O	S	0	0
			24	13	7	3	1		

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	BBB	1	Total	O	S	0	0
			5	4	1		
4	BBB	1	Total	O	S	0	0
			5	4	1		
4	BBB	1	Total	O	S	0	0
			5	4	1		
4	BBB	1	Total	O	S	0	0
			5	4	1		
4	BBB	1	Total	O	S	0	0
			5	4	1		
4	BBB	1	Total	O	S	0	0
			5	4	1		

- Molecule 5 is water.

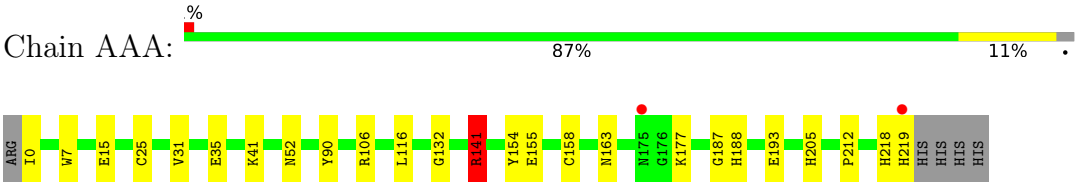
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	AAA	248	Total	O	0	0
			248	248		
5	BBB	230	Total	O	0	0
			230	230		

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.

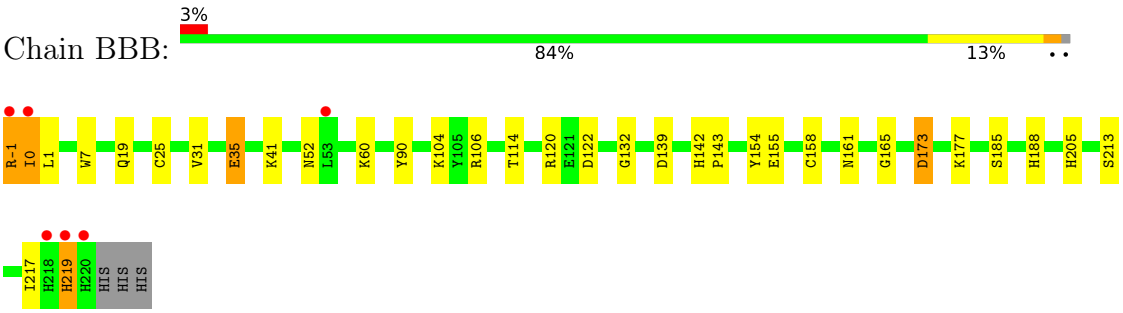
• Molecule 1: Cathepsin S

Chain AAA:



• Molecule 1: Cathepsin S

Chain BBB:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 2 2	Depositor
Cell constants a, b, c, α, β, γ	85.67Å 85.67Å 150.92Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.51 – 2.22 19.51 – 2.22	Depositor EDS
% Data completeness (in resolution range)	99.0 (19.51-2.22) 99.2 (19.51-2.22)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.11 (at 2.21Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.158 , 0.217 0.166 , 0.220	Depositor DCC
R_{free} test set	1436 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å^2)	31.0	Xtriage
Anisotropy	0.258	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	(Not available) , (Not available)	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	4029	wwPDB-VP
Average B, all atoms (Å^2)	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.25% 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: FLC, Q1E, SO4

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	0.84	1/1757 (0.1%)	0.92	2/2376 (0.1%)
1	BBB	0.85	2/1779 (0.1%)	0.91	1/2405 (0.0%)
All	All	0.84	3/3536 (0.1%)	0.91	3/4781 (0.1%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	BBB	35	GLU	CD-OE1	6.03	1.32	1.25
1	BBB	-1	ARG	N-CA	5.94	1.58	1.46
1	AAA	15	GLU	CD-OE2	5.16	1.31	1.25

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	AAA	141	ARG	CB-CA-C	5.72	121.85	110.40
1	AAA	106	ARG	NE-CZ-NH1	5.71	123.16	120.30
1	BBB	106	ARG	NE-CZ-NH2	-5.07	117.77	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	AAA	1713	0	1632	11	0
1	BBB	1734	0	1652	17	0
2	AAA	26	0	10	0	0
3	AAA	24	0	0	0	0
3	BBB	24	0	0	0	0
4	BBB	30	0	0	2	0
5	AAA	248	0	0	2	0
5	BBB	230	0	0	2	0
All	All	4029	0	3294	28	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 28 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:139:ASP:OD2	1:BBB:142:HIS:HE1	1.74	0.71
1:AAA:155:GLU:H	1:AAA:205:HIS:CE1	2.20	0.59
1:AAA:218:HIS:O	1:AAA:219:HIS:HB2	2.02	0.58
1:AAA:31:VAL:O	1:AAA:35:GLU:HG3	2.04	0.58
1:AAA:116:LEU:HD22	1:AAA:212:PRO:HB2	1.84	0.58

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	218/225 (97%)	211 (97%)	7 (3%)	0	100 100
1	BBB	220/225 (98%)	213 (97%)	6 (3%)	1 (0%)	29 30
All	All	438/450 (97%)	424 (97%)	13 (3%)	1 (0%)	47 54

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	BBB	0	ILE

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	181/186 (97%)	174 (96%)	7 (4%)	32	40
1	BBB	183/186 (98%)	172 (94%)	11 (6%)	19	21
All	All	364/372 (98%)	346 (95%)	18 (5%)	25	29

5 of 18 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	BBB	158	CYS
1	BBB	219	HIS
1	BBB	173	ASP
1	BBB	1	LEU
1	BBB	143	PRO

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)

10 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	$\#\|Z\| > 2$
4	SO4	BBB	1005	-	4,4,4	0.32	0	6,6,6	0.25	0
3	Q1E	BBB	1007	-	21,26,26	2.54	6 (28%)	26,37,37	2.80	10 (38%)
2	FLC	AAA	1001	-	12,12,12	1.49	1 (8%)	17,17,17	1.23	2 (11%)
4	SO4	BBB	1002	-	4,4,4	0.24	0	6,6,6	0.07	0
2	FLC	AAA	1002	-	12,12,12	1.69	1 (8%)	17,17,17	2.51	6 (35%)
4	SO4	BBB	1003	-	4,4,4	0.51	0	6,6,6	0.18	0
4	SO4	BBB	1006	-	4,4,4	0.37	0	6,6,6	0.37	0
3	Q1E	AAA	1003	-	21,26,26	2.16	6 (28%)	26,37,37	2.51	13 (50%)
4	SO4	BBB	1004	-	4,4,4	0.31	0	6,6,6	0.13	0
4	SO4	BBB	1001	-	4,4,4	0.37	0	6,6,6	0.17	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	FLC	AAA	1001	-	-	6/16/16/16	-
3	Q1E	BBB	1007	-	-	2/18/28/28	0/3/3/3
3	Q1E	AAA	1003	-	-	0/18/28/28	0/3/3/3
2	FLC	AAA	1002	-	-	4/16/16/16	-

The worst 5 of 14 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	BBB	1007	Q1E	C2-N3	7.04	1.42	1.32
3	BBB	1007	Q1E	N1-N	-6.09	1.25	1.37
3	AAA	1003	Q1E	C2-N3	5.79	1.40	1.32
3	AAA	1003	Q1E	N1-N	-4.76	1.28	1.37
2	AAA	1002	FLC	CB-CBC	4.24	1.57	1.53

The worst 5 of 31 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	BBB	1007	Q1E	C7-N5-C6	-7.46	103.91	112.17
2	AAA	1002	FLC	OB1-CBC-CB	-6.05	113.69	122.25
3	BBB	1007	Q1E	O1-S-C9	5.87	118.00	107.97
3	BBB	1007	Q1E	C3-C2-N3	-5.23	118.08	124.15
2	AAA	1002	FLC	OB2-CBC-CB	5.11	121.92	113.05

There are no chirality outliers.

5 of 12 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	AAA	1001	FLC	CAC-CA-CB-CBC
2	AAA	1001	FLC	CAC-CA-CB-CG
2	AAA	1002	FLC	CA-CB-CG-CGC
2	AAA	1002	FLC	CBC-CB-CG-CGC
2	AAA	1002	FLC	OHB-CB-CG-CGC

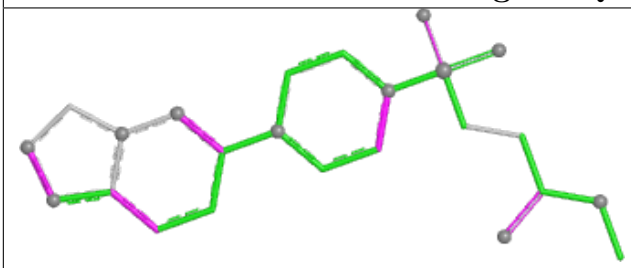
There are no ring outliers.

2 monomers are involved in 2 short contacts:

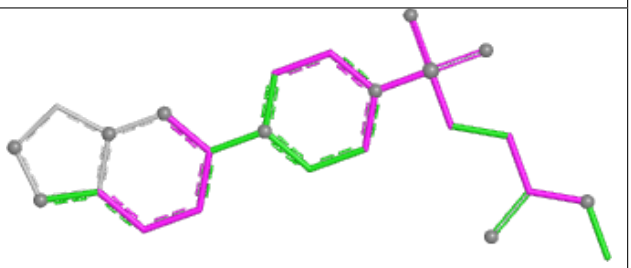
Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	BBB	1003	SO4	1	0
4	BBB	1004	SO4	1	0

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.

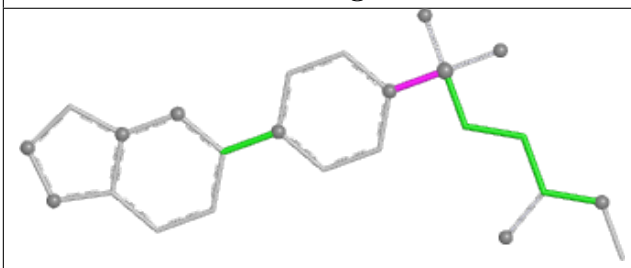
Ligand Q1E BBB 1007



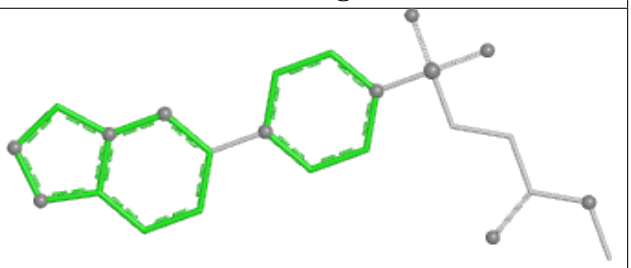
Bond lengths



Bond angles

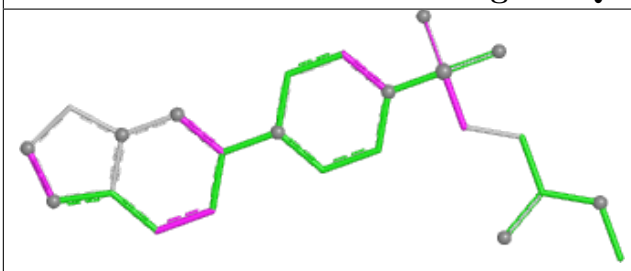


Torsions

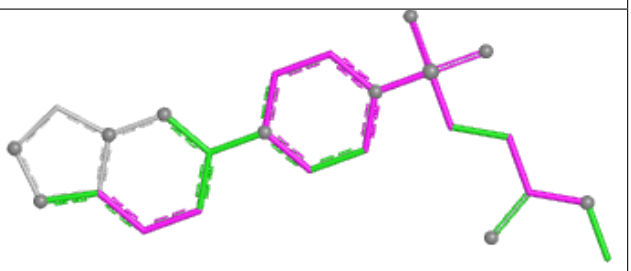


Rings

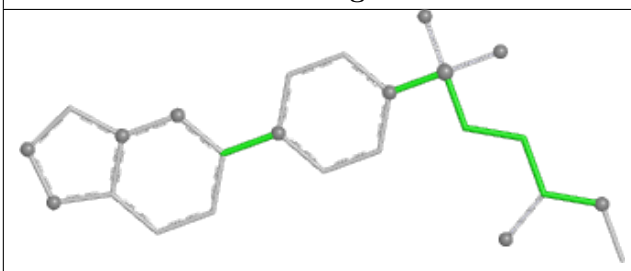
Ligand Q1E AAA 1003



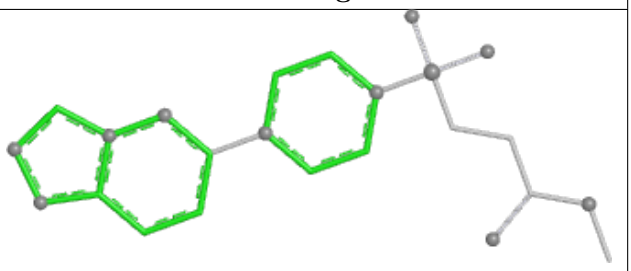
Bond lengths



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	220/225 (97%)	-0.49	2 (0%)	84	83	23, 32, 46, 93	0
1	BBB	222/225 (98%)	-0.43	6 (2%)	54	52	23, 32, 47, 117	0
All	All	442/450 (98%)	-0.46	8 (1%)	68	66	23, 32, 47, 117	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	BBB	220	HIS	8.1
1	BBB	0	ILE	7.4
1	AAA	219	HIS	5.9
1	BBB	219	HIS	4.8
1	BBB	-1	ARG	4.4

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

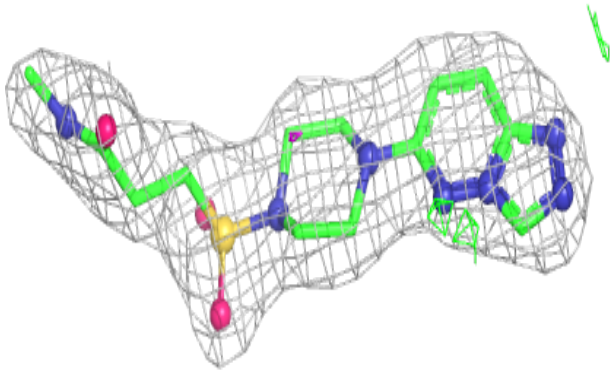
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9'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

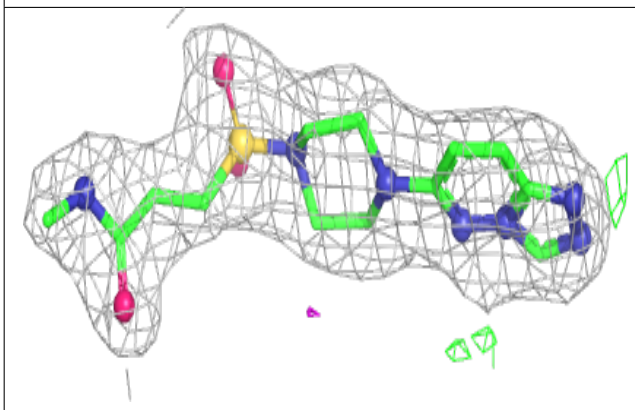
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	SO4	BBB	1001	5/5	0.56	0.59	147,150,156,160	0
4	SO4	BBB	1006	5/5	0.73	0.45	64,86,104,113	0
2	FLC	AAA	1002	13/13	0.77	0.22	58,71,81,97	0
2	FLC	AAA	1001	13/13	0.79	0.24	45,48,49,51	13
4	SO4	BBB	1002	5/5	0.82	0.21	87,92,103,105	0
4	SO4	BBB	1004	5/5	0.86	0.41	94,103,103,108	0
4	SO4	BBB	1003	5/5	0.94	0.18	47,48,49,50	5
4	SO4	BBB	1005	5/5	0.97	0.14	71,71,73,76	5
3	Q1E	AAA	1003	24/24	0.98	0.06	32,36,47,50	0
3	Q1E	BBB	1007	24/24	0.98	0.08	29,34,45,48	0

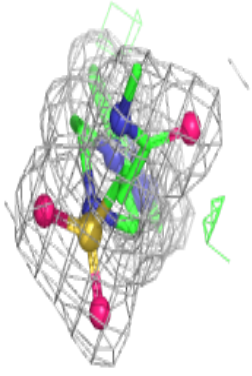
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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

Electron density around Q1E AAA 1003:

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

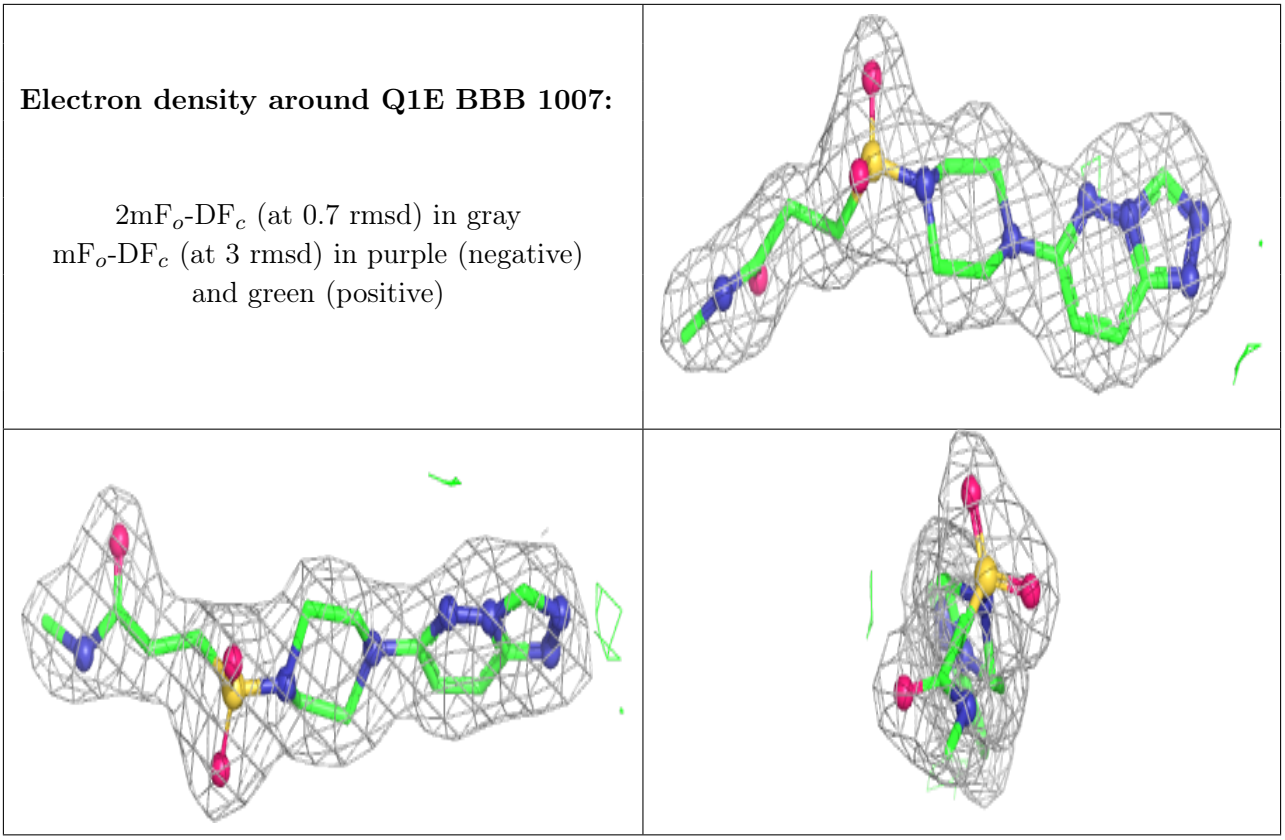






Electron density around Q1E BBB 1007:

$2mF_o$-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.