Full wwPDB X-ray Structure Validation Report (i)

Sep 7, 2020 – 02:36 PM BST

PDB ID	:	6SZ7
Title	:	Crystal structure of YTHDC1 with fragment 5 (DHU_DC1_066)
Authors	:	Bedi, R.K.; Huang, D.; Sledz, P.; Caflisch, A.
Deposited on	:	2019-10-02
Resolution	:	2.31 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 following versions of software and data (see references (1)) were used in the production of this report:

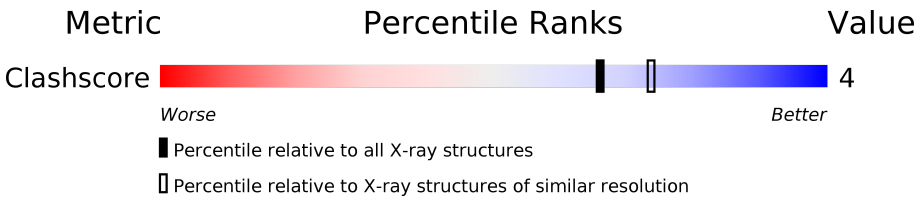
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	FAILED
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.14.2

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 2.31 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))
Clashscore	141614	6604 (2.34-2.30)

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 on 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%

Note EDS failed to run properly.

Mol	Chain	Length	Quality of chain
1	A	183	83% 6% 11%
1	B	183	83% 7% 10%

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	M1Q	A	601	X	-	-	-
2	M1Q	B	601	X	-	-	-
3	SO4	B	604	-	-	X	-

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2842 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 YTH domain-containing protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	163	1289	832	228	223	6	0	3	0
1	B	164	1297	839	224	228	6	0	3	0

There are 36 discrepancies between the modelled and reference sequences:

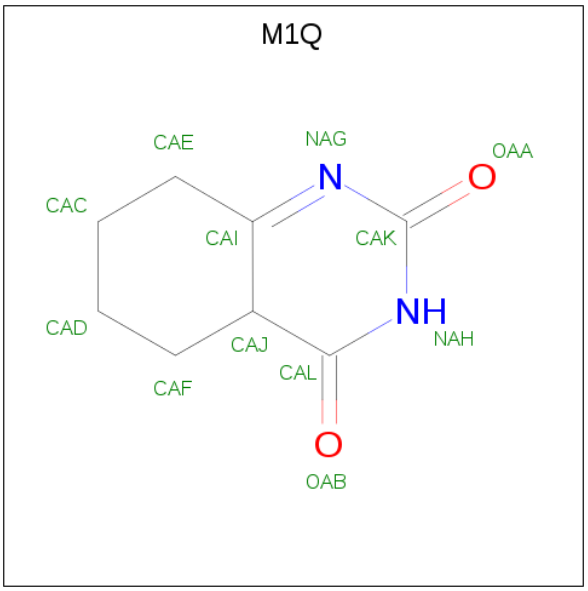
Chain	Residue	Modelled	Actual	Comment	Reference
A	327	MET	-	initiating methionine	UNP Q96MU7
A	328	HIS	-	expression tag	UNP Q96MU7
A	329	HIS	-	expression tag	UNP Q96MU7
A	330	HIS	-	expression tag	UNP Q96MU7
A	331	HIS	-	expression tag	UNP Q96MU7
A	332	HIS	-	expression tag	UNP Q96MU7
A	333	HIS	-	expression tag	UNP Q96MU7
A	334	SER	-	expression tag	UNP Q96MU7
A	335	SER	-	expression tag	UNP Q96MU7
A	336	GLY	-	expression tag	UNP Q96MU7
A	337	ARG	-	expression tag	UNP Q96MU7
A	338	GLU	-	expression tag	UNP Q96MU7
A	339	ASN	-	expression tag	UNP Q96MU7
A	340	LEU	-	expression tag	UNP Q96MU7
A	341	TYR	-	expression tag	UNP Q96MU7
A	342	PHE	-	expression tag	UNP Q96MU7
A	343	GLN	-	expression tag	UNP Q96MU7
A	344	GLY	-	expression tag	UNP Q96MU7
B	327	MET	-	initiating methionine	UNP Q96MU7
B	328	HIS	-	expression tag	UNP Q96MU7
B	329	HIS	-	expression tag	UNP Q96MU7
B	330	HIS	-	expression tag	UNP Q96MU7
B	331	HIS	-	expression tag	UNP Q96MU7
B	332	HIS	-	expression tag	UNP Q96MU7
B	333	HIS	-	expression tag	UNP Q96MU7

Continued on next page...

Continued from previous page...

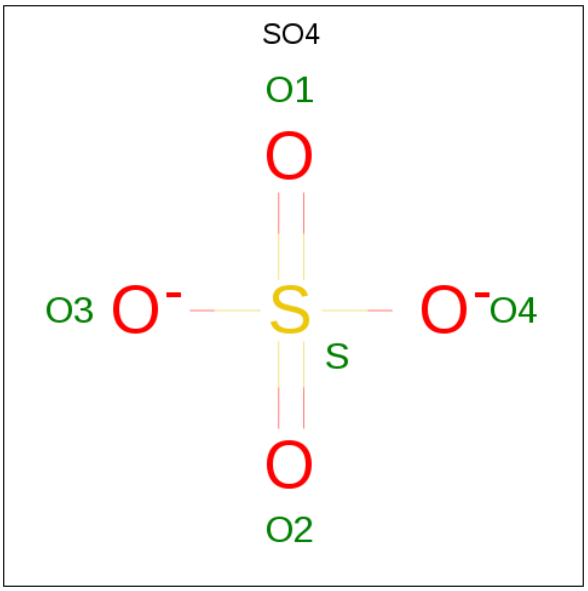
Chain	Residue	Modelled	Actual	Comment	Reference
B	334	SER	-	expression tag	UNP Q96MU7
B	335	SER	-	expression tag	UNP Q96MU7
B	336	GLY	-	expression tag	UNP Q96MU7
B	337	ARG	-	expression tag	UNP Q96MU7
B	338	GLU	-	expression tag	UNP Q96MU7
B	339	ASN	-	expression tag	UNP Q96MU7
B	340	LEU	-	expression tag	UNP Q96MU7
B	341	TYR	-	expression tag	UNP Q96MU7
B	342	PHE	-	expression tag	UNP Q96MU7
B	343	GLN	-	expression tag	UNP Q96MU7
B	344	GLY	-	expression tag	UNP Q96MU7

- Molecule 2 is 5,6,7,8-tetrahydro-4 {a} {H}-quinazoline-2,4-dione (three-letter code: M1Q) (formula: $C_8H_{10}N_2O_2$) (labeled as "Ligand of Interest" by author).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	A	1	12	8	2	2	0	0
			Total	C	N	O		
2	B	1	12	8	2	2	0	0

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total O S 5 4 1	0	0
3	A	1	Total O S 5 4 1	0	0
3	B	1	Total O S 5 4 1	0	0
3	B	1	Total O S 5 4 1	0	0
3	B	1	Total O S 5 4 1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	83	Total O 83 83	0	0
4	B	124	Total O 124 124	0	0

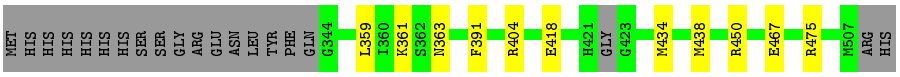
3 Residue-property plots (i)

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. 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. 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.

Note EDS failed to run properly.

- Molecule 1: YTH domain-containing protein 1

Chain A:



- Molecule 1: YTH domain-containing protein 1

Chain B:

4 Data and refinement statistics (i)

EDS failed to run properly - this section is therefore incomplete.

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	39.82Å 103.57Å 42.52Å 90.00° 105.08° 90.00°	Depositor
Resolution (Å)	41.06 – 2.31	Depositor
% Data completeness (in resolution range)	96.2 (41.06-2.31)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.14 (at 2.32Å)	Xtriage
Refinement program	PHENIX 1.16_3549	Depositor
R, R_{free}	0.197 , 0.265	Depositor
Wilson B-factor (Å^2)	30.8	Xtriage
Anisotropy	0.102	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.53$, $< L^2 > = 0.36$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	2842	wwPDB-VP
Average B, all atoms (Å^2)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.34% 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: M1Q, 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	A	0.42	0/1330	0.58	0/1797
1	B	0.46	0/1341	0.61	1/1813 (0.1%)
All	All	0.44	0/2671	0.60	1/3610 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	489	CYS	CA-CB-SG	-5.30	104.46	114.00

There are no chirality outliers.

There are no planarity outliers.

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	1289	0	1289	9	0
1	B	1297	0	1304	11	0
2	A	12	0	0	0	0
2	B	12	0	0	0	0
3	A	10	0	0	0	0
3	B	15	0	0	2	0
4	A	83	0	0	1	0
4	B	124	0	0	6	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	2842	0	2593	21	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.

All (21) 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)
3:B:604:SO4:S	4:B:703:HOH:O	2.38	0.80
1:B:446:ASP:OD1	4:B:701:HOH:O	2.15	0.65
1:B:496:GLU:OE2	4:B:702:HOH:O	2.16	0.61
1:B:404:ARG:NH2	4:B:703:HOH:O	2.30	0.60
1:A:404:ARG:HH22	1:A:475:ARG:HH11	1.50	0.58
1:B:363:ASN:OD1	1:B:404:ARG:HG3	2.06	0.56
1:A:359:LEU:HD21	1:A:361:LYS:HE3	1.88	0.55
1:A:363:ASN:N	1:A:363:ASN:OD1	2.43	0.48
1:A:404:ARG:HH22	1:A:475:ARG:NH1	2.12	0.47
1:A:467:GLU:O	1:B:366:GLU:HG2	2.14	0.47
1:A:475:ARG:HA	1:A:475:ARG:HD2	1.75	0.47
1:A:450:ARG:HD2	4:A:718:HOH:O	2.14	0.46
1:B:496:GLU:CD	4:B:702:HOH:O	2.55	0.45
1:B:419:SER:HB3	1:B:443:PHE:CZ	2.53	0.43
3:B:604:SO4:O4	4:B:703:HOH:O	2.20	0.43
1:A:391:PHE:CD2	1:A:418:GLU:HG3	2.54	0.42
1:B:372:LYS:HA	1:B:447:TRP:CE2	2.55	0.42
1:A:434:MET:HA	1:A:438:MET:SD	2.61	0.41
1:B:368:VAL:O	1:B:372:LYS:HG3	2.21	0.40
1:B:500:LEU:HD23	1:B:500:LEU:HA	1.87	0.40
1:B:372:LYS:HA	1:B:447:TRP:NE1	2.36	0.40

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

There are no protein backbone outliers to report in this entry.

5.3.2 Protein sidechains (i)

There are no protein residues with a non-rotameric sidechain to report in this entry.

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)

7 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	$\#\|Z\| > 2$
3	SO4	B	602	-	4,4,4	0.26	0	6,6,6	0.24	0
3	SO4	A	603	-	4,4,4	0.15	0	6,6,6	0.30	0
2	M1Q	A	601	-	12,13,13	5.46	7 (58%)	14,18,18	1.77	4 (28%)
3	SO4	B	604	-	4,4,4	0.19	0	6,6,6	0.23	0
3	SO4	A	602	-	4,4,4	0.20	0	6,6,6	0.39	0
2	M1Q	B	601	-	12,13,13	5.43	7 (58%)	14,18,18	2.09	6 (42%)
3	SO4	B	603	-	4,4,4	0.12	0	6,6,6	0.29	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	M1Q	A	601	-	1/1/4/5	-	0/2/2/2
2	M1Q	B	601	-	1/1/4/5	-	0/2/2/2

All (14) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	601	M1Q	CAJ-CAL	-12.96	1.41	1.52
2	A	601	M1Q	CAJ-CAL	-12.29	1.42	1.52
2	A	601	M1Q	OAA-CAK	9.60	1.42	1.24
2	B	601	M1Q	OAA-CAK	8.96	1.40	1.24
2	A	601	M1Q	OAB-CAL	8.38	1.40	1.23
2	B	601	M1Q	OAB-CAL	8.06	1.39	1.23
2	B	601	M1Q	CAI-NAG	4.02	1.35	1.28
2	A	601	M1Q	CAI-NAG	3.89	1.34	1.28
2	A	601	M1Q	CAE-CAI	3.23	1.55	1.50
2	A	601	M1Q	CAL-NAH	-3.02	1.32	1.37
2	B	601	M1Q	CAE-CAI	2.80	1.54	1.50
2	B	601	M1Q	CAL-NAH	-2.79	1.32	1.37
2	A	601	M1Q	CAK-NAH	-2.38	1.33	1.39
2	B	601	M1Q	CAK-NAH	-2.30	1.33	1.39

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	601	M1Q	CAL-NAH-CAK	-4.66	118.40	125.42
2	A	601	M1Q	CAF-CAJ-CAL	3.65	120.81	112.28
2	B	601	M1Q	CAJ-CAL-NAH	3.35	120.38	116.25
2	A	601	M1Q	CAL-NAH-CAK	-3.17	120.64	125.42
2	B	601	M1Q	OAB-CAL-CAJ	-2.62	116.19	121.95
2	B	601	M1Q	OAA-CAK-NAG	-2.33	117.96	121.83
2	B	601	M1Q	CAF-CAJ-CAL	2.32	117.70	112.28
2	A	601	M1Q	CAD-CAF-CAJ	-2.16	108.58	111.28
2	B	601	M1Q	NAH-CAK-NAG	2.13	123.56	119.38
2	A	601	M1Q	CAJ-CAL-NAH	2.05	118.78	116.25

All (2) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	A	601	M1Q	CAJ
2	B	601	M1Q	CAJ

There are no torsion outliers.

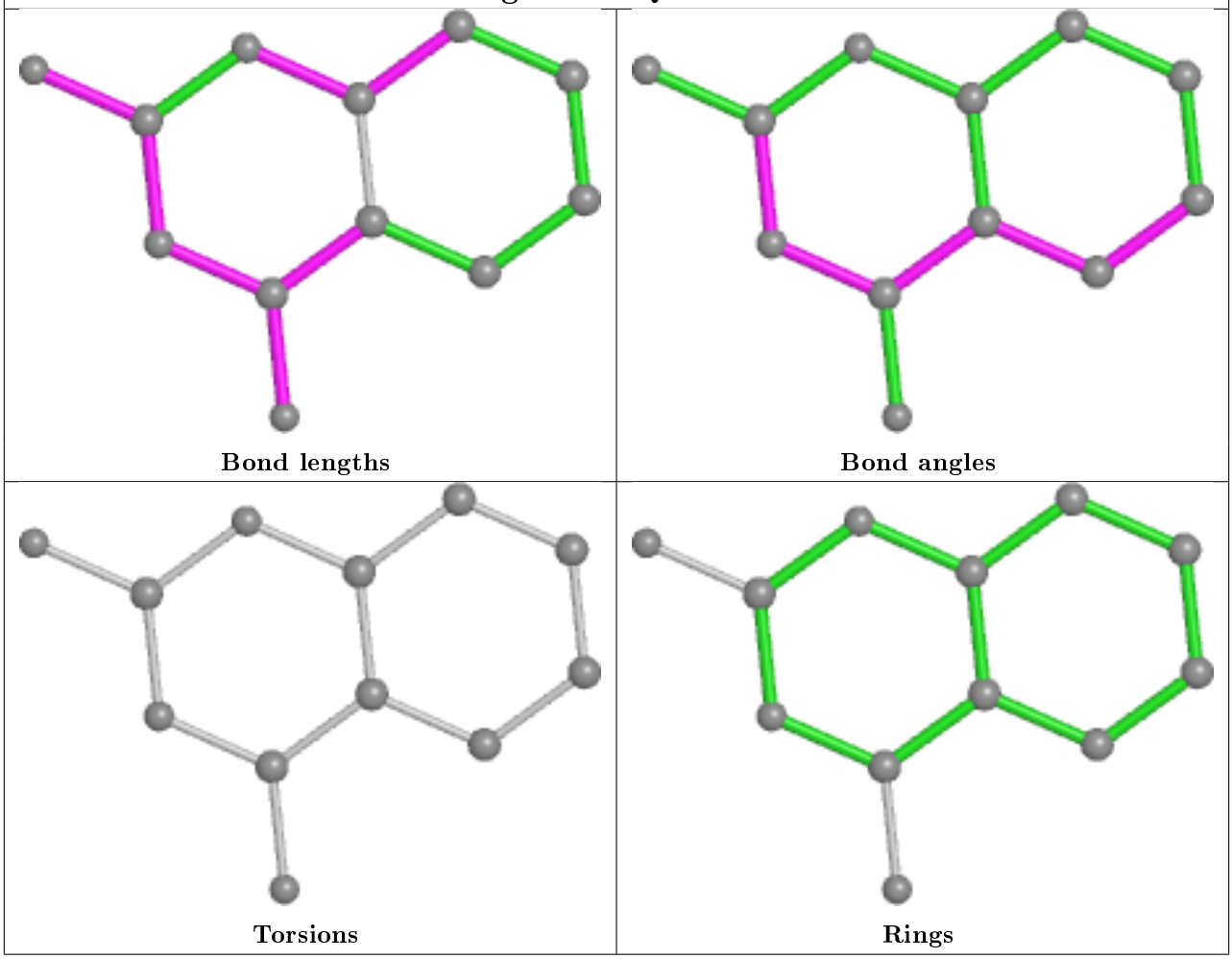
There are no ring outliers.

1 monomer is involved in 2 short contacts:

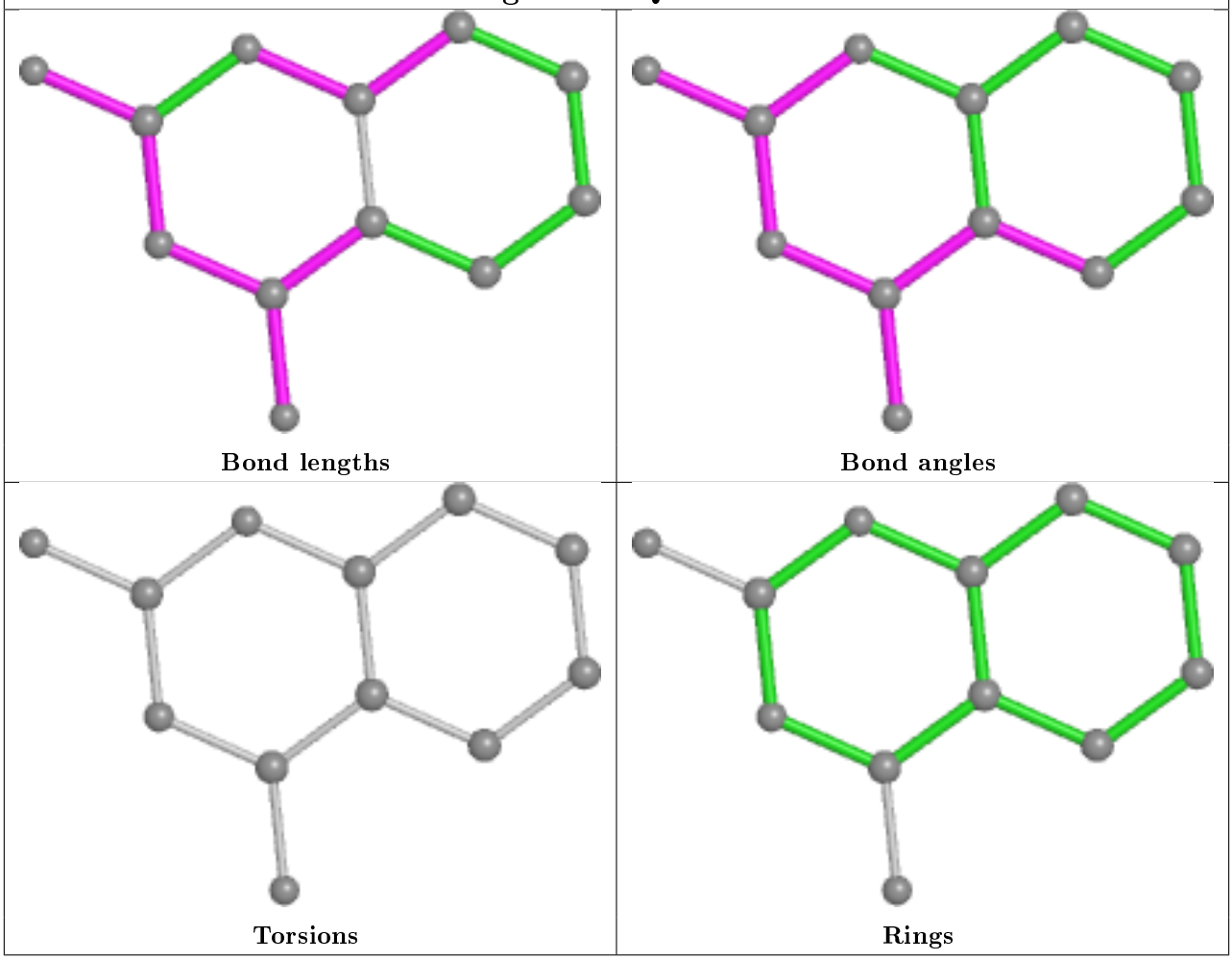
Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	604	SO4	2	0

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 M1Q A 601



Ligand M1Q B 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 (i)

6.1 Protein, DNA and RNA chains (i)

EDS failed to run properly - this section is therefore empty.

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

EDS failed to run properly - this section is therefore empty.

6.3 Carbohydrates (i)

EDS failed to run properly - this section is therefore empty.

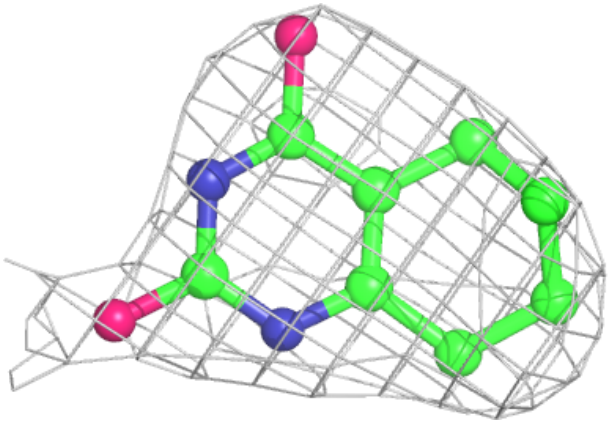
6.4 Ligands (i)

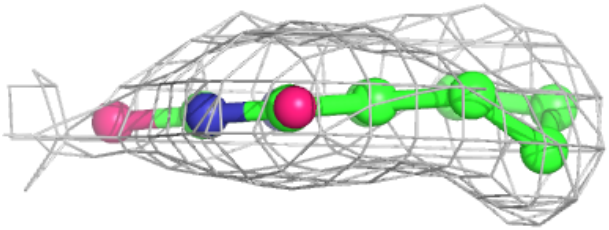
EDS failed to run properly - this section is therefore empty.

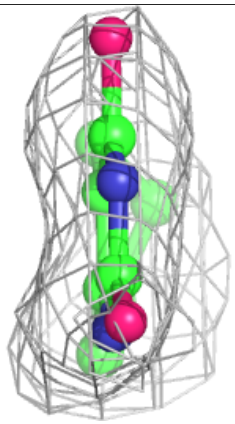
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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

Electron density around M1Q A 601:

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

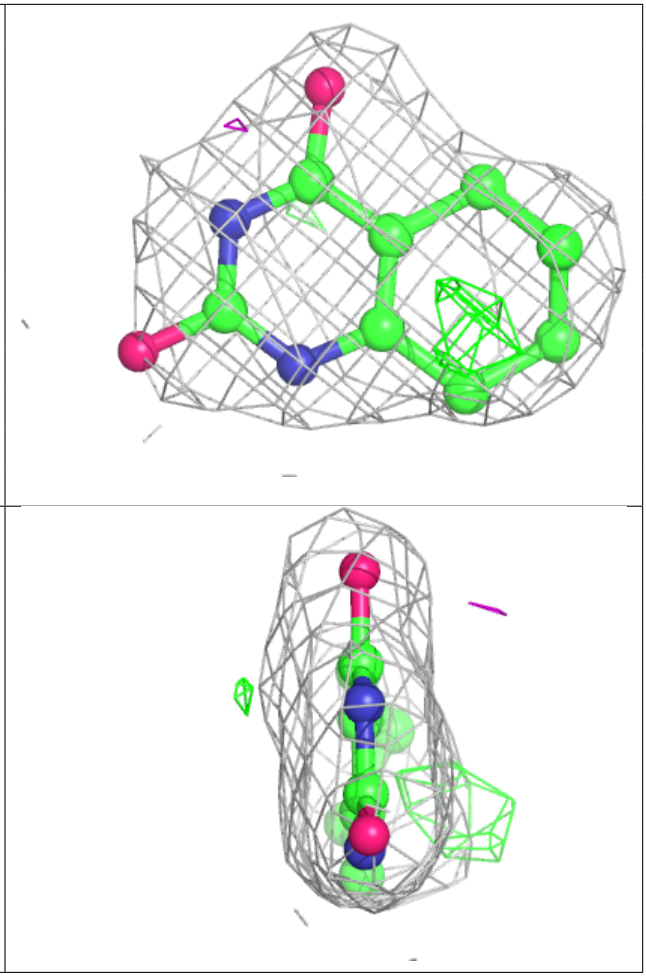






Electron density around M1Q B 601:

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)

EDS failed to run properly - this section is therefore empty.