Full wwPDB X-ray Structure Validation Report (i)

Oct 3, 2023 – 04:14 AM EDT

PDB ID	:	6PXU
Title	:	Crystal structure of human GalNAc-T12 bound to a diglycosylated peptide, Mn2+, and UDP
Authors	:	Samara, N.L.; Fernandez, A.J.
Deposited on	:	2019-07-27
Resolution	:	2.01 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 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	:	FAILED
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)	:	2.35.1

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 2.01 Å.

There are no overall percentile quality scores available for this entry.

MolProbity and EDS failed to run properly - the sequence quality summary graphics cannot be shown.

2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 9719 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.

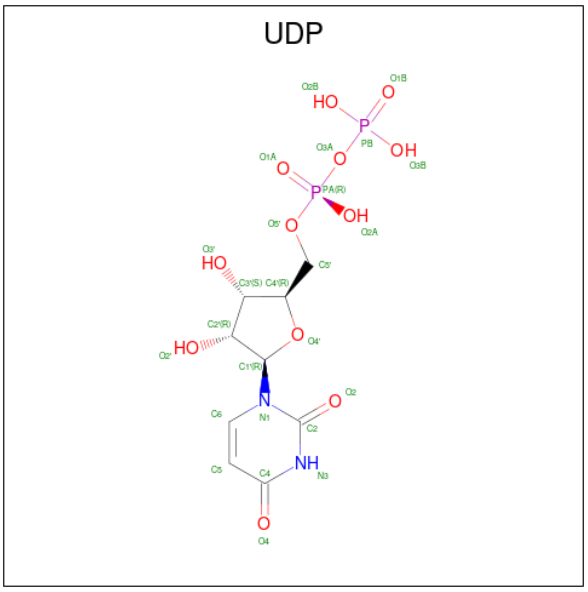
- Molecule 1 is a protein called Polypeptide N-acetylgalactosaminyltransferase 12.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	526	4351	2761	770	793	27	0	10	0
1	B	532	4405	2793	787	801	24	0	8	0

- Molecule 2 is a protein called GAGATGAGAGYYITPRTGAGA.

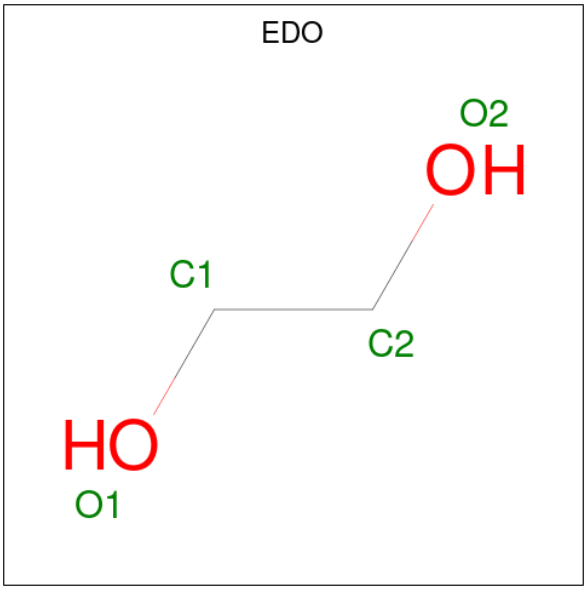
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
2	C	13	93	59	16	18	0	0	0
2	D	15	102	64	18	20	0	0	0

- Molecule 3 is URIDINE-5'-DIPHOSPHATE (three-letter code: UDP) (formula: $C_9H_{14}N_2O_{12}P_2$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	A	1	Total 25	C 9	N 2	O 12	P 2	0	0
3	B	1	Total 25	C 9	N 2	O 12	P 2	0	0

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



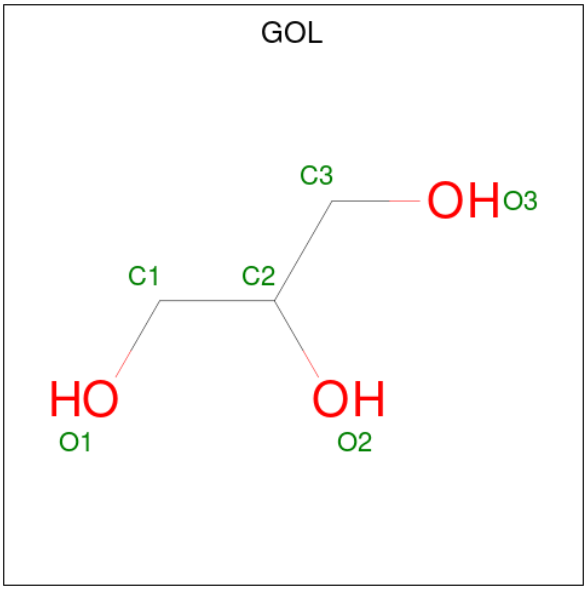
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total 4	C 2	O 2		0	0
4	A	1	Total 4	C 2	O 2		0	0
4	A	1	Total 4	C 2	O 2		0	0
4	A	1	Total 4	C 2	O 2		0	0
4	A	1	Total 4	C 2	O 2		0	0
4	B	1	Total 4	C 2	O 2		0	0
4	B	1	Total 4	C 2	O 2		0	0
4	B	1	Total 4	C 2	O 2		0	0
4	B	1	Total 4	C 2	O 2		0	0
4	B	1	Total 4	C 2	O 2		0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	1	Total C O 4 2 2	0	0
4	B	1	Total C O 4 2 2	0	0

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).

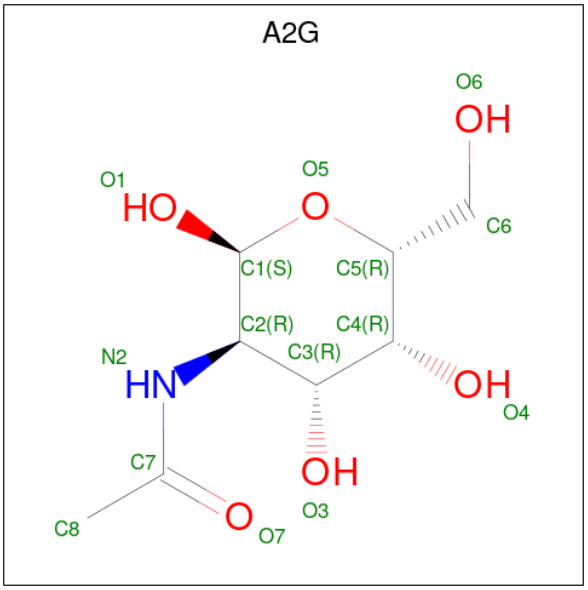


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total C O 6 3 3	0	0
5	A	1	Total C O 6 3 3	0	0
5	A	1	Total C O 6 3 3	0	0
5	B	1	Total C O 6 3 3	0	0
5	B	1	Total C O 6 3 3	0	0
5	B	1	Total C O 6 3 3	0	0
5	B	1	Total C O 6 3 3	0	0
5	B	1	Total C O 6 3 3	0	0

- Molecule 6 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Mn		
6	A	2	2	2	0	0
			Total	Mn		
6	B	2	2	2	0	0

- Molecule 7 is 2-acetamido-2-deoxy-alpha-D-galactopyranose (three-letter code: A2G) (formula: $C_8H_{15}NO_6$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
7	C	1	14	8	1	5	0	0
			Total	C	N	O		
7	C	1	14	8	1	5	0	0
			Total	C	N	O		
7	D	1	14	8	1	5	0	0
			Total	C	N	O		
7	D	1	14	8	1	5	0	0

- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
8	A	271	271	271	0	0
			Total	O		
8	B	280	280	280	0	0
			Total	O		
8	C	6	6	6	0	0

Continued on next page...

Continued from previous page...

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

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

3 Data and refinement statistics (i)

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

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	72.82Å 73.12Å 74.44Å 113.08° 100.50° 108.23°	Depositor
Resolution (Å)	20.03 – 2.01	Depositor
% Data completeness (in resolution range)	97.1 (20.03-2.01)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.67 (at 2.01Å)	Xtriage
Refinement program	PHENIX 1.10.1_2155	Depositor
R, R_{free}	0.171 , 0.219	Depositor
Wilson B-factor (Å^2)	29.3	Xtriage
Anisotropy	0.542	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	9719	wwPDB-VP
Average B, all atoms (Å^2)	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.07% 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.

4 Model quality (i)

4.1 Standard geometry (i)

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

4.2 Too-close contacts (i)

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

4.3 Torsion angles (i)

4.3.1 Protein backbone (i)

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

4.3.2 Protein sidechains (i)

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

4.3.3 RNA (i)

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

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

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

4.5 Carbohydrates (i)

There are no monosaccharides in this entry.

4.6 Ligand geometry (i)

Of 30 ligands modelled in this entry, 4 are monoatomic - leaving 26 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	UDP	A	601	6	24,26,26	1.07	2 (8%)	37,40,40	1.98	10 (27%)
5	GOL	B	607	-	5,5,5	0.38	0	5,5,5	0.35	0
5	GOL	B	610	-	5,5,5	0.38	0	5,5,5	0.21	0
4	EDO	A	608	-	3,3,3	0.49	0	2,2,2	0.37	0
7	A2G	C	102	2	14,14,15	0.51	0	17,19,21	0.83	0
4	EDO	A	609	-	3,3,3	0.46	0	2,2,2	0.37	0
7	A2G	D	101	2	14,14,15	0.45	0	17,19,21	0.93	1 (5%)
4	EDO	A	602	-	3,3,3	0.52	0	2,2,2	0.22	0
5	GOL	B	608	-	5,5,5	0.43	0	5,5,5	0.65	0
4	EDO	B	613	-	3,3,3	0.44	0	2,2,2	0.45	0
4	EDO	A	604	-	3,3,3	0.51	0	2,2,2	0.32	0
5	GOL	B	606	-	5,5,5	0.42	0	5,5,5	0.71	0
7	A2G	C	101	2	14,14,15	0.45	0	17,19,21	0.81	0
5	GOL	B	609	-	5,5,5	0.37	0	5,5,5	0.20	0
4	EDO	A	603	-	3,3,3	0.52	0	2,2,2	0.26	0
7	A2G	D	102	2	14,14,15	0.48	0	17,19,21	2.28	4 (23%)
4	EDO	B	603	-	3,3,3	0.44	0	2,2,2	0.42	0
4	EDO	B	614	-	3,3,3	0.46	0	2,2,2	0.31	0
5	GOL	A	607	-	5,5,5	0.35	0	5,5,5	0.46	0
4	EDO	B	611	-	3,3,3	0.50	0	2,2,2	0.26	0
4	EDO	B	604	-	3,3,3	0.43	0	2,2,2	0.43	0
4	EDO	B	605	-	3,3,3	0.50	0	2,2,2	0.26	0
4	EDO	B	612	-	3,3,3	0.47	0	2,2,2	0.30	0
5	GOL	A	606	-	5,5,5	0.43	0	5,5,5	0.33	0
3	UDP	B	602	6	24,26,26	0.85	1 (4%)	37,40,40	1.55	4 (10%)
5	GOL	A	605	-	5,5,5	0.37	0	5,5,5	0.19	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	UDP	A	601	6	-	8/16/32/32	0/2/2/2
5	GOL	B	607	-	-	0/4/4/4	-
5	GOL	B	610	-	-	2/4/4/4	-
4	EDO	A	608	-	-	0/1/1/1	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
7	A2G	C	102	2	-	2/6/23/26	0/1/1/1
4	EDO	A	609	-	-	1/1/1/1	-
7	A2G	D	101	2	-	2/6/23/26	0/1/1/1
4	EDO	A	602	-	-	0/1/1/1	-
5	GOL	B	608	-	-	1/4/4/4	-
4	EDO	B	613	-	-	0/1/1/1	-
4	EDO	A	604	-	-	0/1/1/1	-
5	GOL	B	606	-	-	2/4/4/4	-
7	A2G	C	101	2	-	0/6/23/26	0/1/1/1
5	GOL	B	609	-	-	2/4/4/4	-
4	EDO	A	603	-	-	1/1/1/1	-
7	A2G	D	102	2	-	5/6/23/26	0/1/1/1
4	EDO	B	603	-	-	0/1/1/1	-
4	EDO	B	614	-	-	0/1/1/1	-
5	GOL	A	607	-	-	0/4/4/4	-
4	EDO	B	611	-	-	1/1/1/1	-
4	EDO	B	604	-	-	0/1/1/1	-
4	EDO	B	605	-	-	0/1/1/1	-
4	EDO	B	612	-	-	0/1/1/1	-
5	GOL	A	606	-	-	0/4/4/4	-
3	UDP	B	602	6	-	3/16/32/32	0/2/2/2
5	GOL	A	605	-	-	4/4/4/4	-

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	601	UDP	O4'-C4'	-2.55	1.39	1.45
3	A	601	UDP	C6-C5	2.13	1.40	1.35
3	B	602	UDP	C6-C5	2.01	1.39	1.35

All (19) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
7	D	102	A2G	C4-C3-C2	-6.36	101.69	111.02
3	A	601	UDP	C4-N3-C2	-6.08	118.56	126.58
3	B	602	UDP	C4-N3-C2	-5.48	119.35	126.58
3	B	602	UDP	N3-C2-N1	4.43	120.77	114.89
3	A	601	UDP	C5-C4-N3	4.35	121.34	114.84
7	D	102	A2G	C2-N2-C7	4.26	128.97	122.90
3	A	601	UDP	N3-C2-N1	3.99	120.18	114.89

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	601	UDP	O5'-PA-O1A	3.94	124.46	109.07
3	B	602	UDP	C5-C4-N3	3.41	119.95	114.84
3	A	601	UDP	O4'-C4'-C5'	-3.38	98.26	109.37
7	D	102	A2G	O5-C1-C2	-3.19	106.25	111.29
3	A	601	UDP	PA-O3A-PB	-2.62	123.83	132.83
7	D	101	A2G	C1-O5-C5	2.55	115.64	112.19
3	A	601	UDP	C5-C6-N1	-2.47	117.67	121.81
3	A	601	UDP	O4-C4-C5	-2.34	121.04	125.16
3	B	602	UDP	O4-C4-C5	-2.26	121.19	125.16
7	D	102	A2G	C8-C7-N2	-2.18	112.41	116.10
3	A	601	UDP	O3B-PB-O3A	2.12	111.73	104.64
3	A	601	UDP	O2-C2-N1	-2.03	120.08	122.79

There are no chirality outliers.

All (34) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	601	UDP	C5'-O5'-PA-O1A
5	A	605	GOL	O1-C1-C2-C3
5	B	606	GOL	O1-C1-C2-C3
5	B	609	GOL	O1-C1-C2-O2
5	B	609	GOL	O1-C1-C2-C3
5	B	610	GOL	O1-C1-C2-O2
5	B	610	GOL	O1-C1-C2-C3
7	D	102	A2G	C3-C2-N2-C7
7	D	102	A2G	C4-C5-C6-O6
7	D	101	A2G	O7-C7-N2-C2
7	D	101	A2G	C8-C7-N2-C2
7	D	102	A2G	O7-C7-N2-C2
7	D	102	A2G	C8-C7-N2-C2
7	C	102	A2G	O5-C5-C6-O6
7	C	102	A2G	C4-C5-C6-O6
7	D	102	A2G	O5-C5-C6-O6
5	A	605	GOL	C1-C2-C3-O3
5	B	608	GOL	C1-C2-C3-O3
5	A	605	GOL	O2-C2-C3-O3
5	B	606	GOL	O1-C1-C2-O2
4	A	603	EDO	O1-C1-C2-O2
4	A	609	EDO	O1-C1-C2-O2
5	A	605	GOL	O1-C1-C2-O2
4	B	611	EDO	O1-C1-C2-O2
3	A	601	UDP	O4'-C4'-C5'-O5'

Continued on next page...

Continued from previous page...

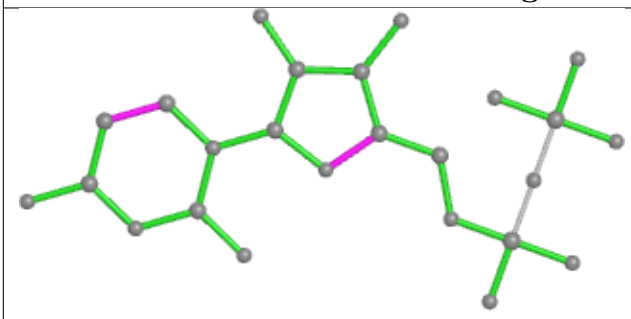
Mol	Chain	Res	Type	Atoms
3	A	601	UDP	C5'-O5'-PA-O3A
3	A	601	UDP	C5'-O5'-PA-O2A
3	A	601	UDP	PB-O3A-PA-O2A
3	A	601	UDP	O4'-C1'-N1-C6
3	B	602	UDP	C4'-C5'-O5'-PA
3	A	601	UDP	C2'-C1'-N1-C6
3	B	602	UDP	C2'-C1'-N1-C6
3	B	602	UDP	O4'-C1'-N1-C6
3	A	601	UDP	O4'-C1'-N1-C2

There are no ring outliers.

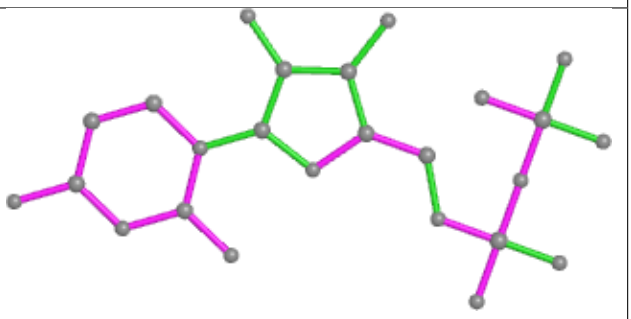
No monomer is involved in short contacts.

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.

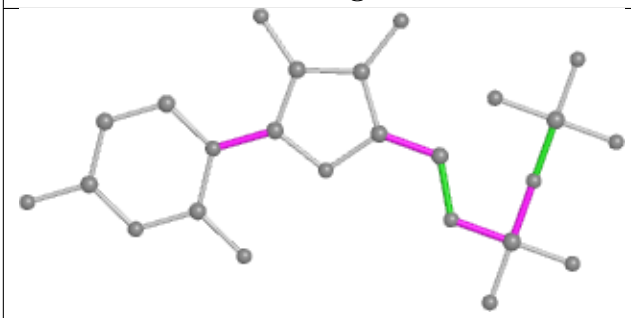
Ligand UDP A 601



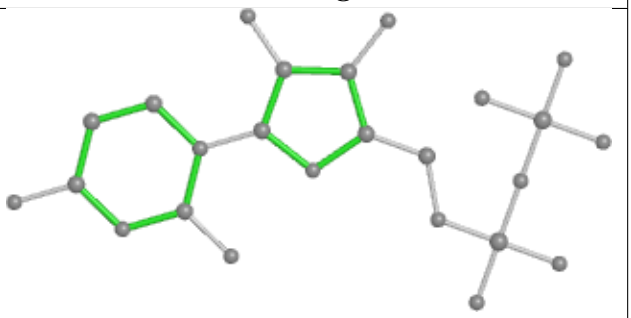
Bond lengths



Bond angles

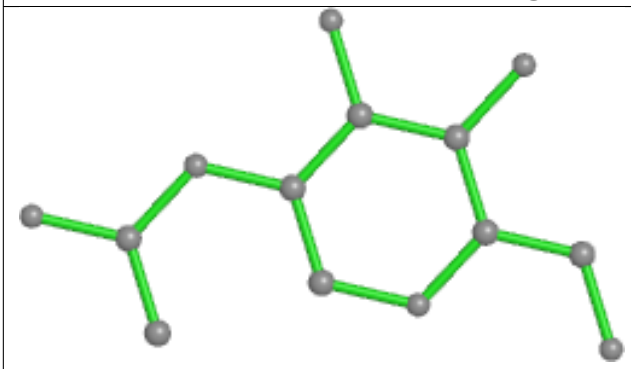


Torsions

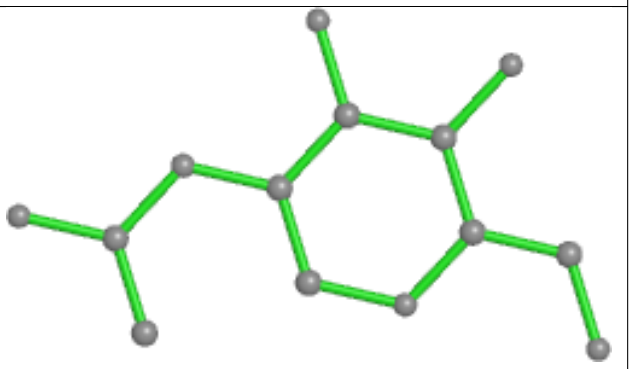


Rings

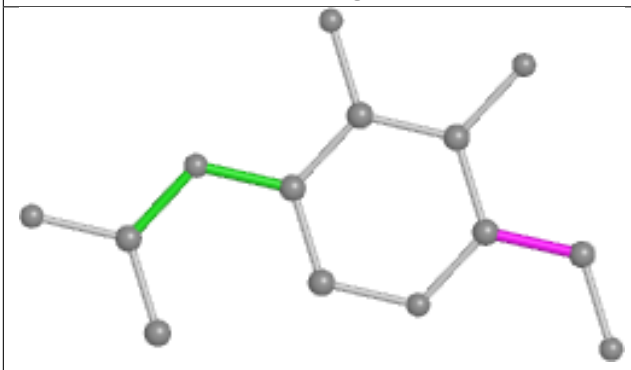
Ligand A2G C 102



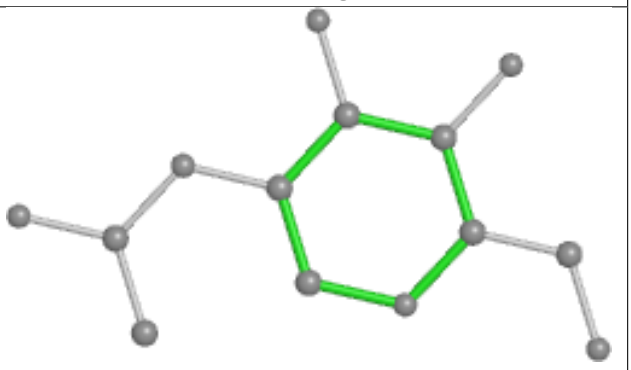
Bond lengths



Bond angles

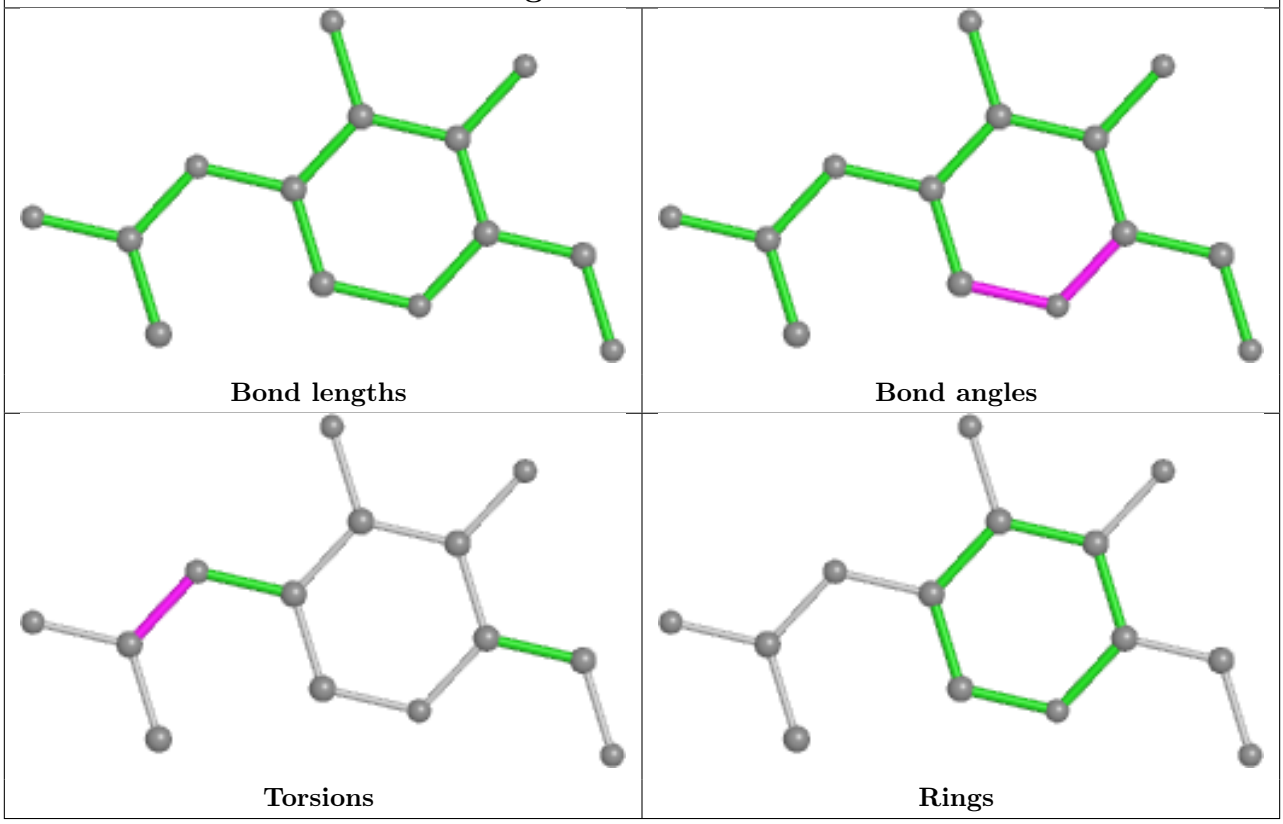


Torsions

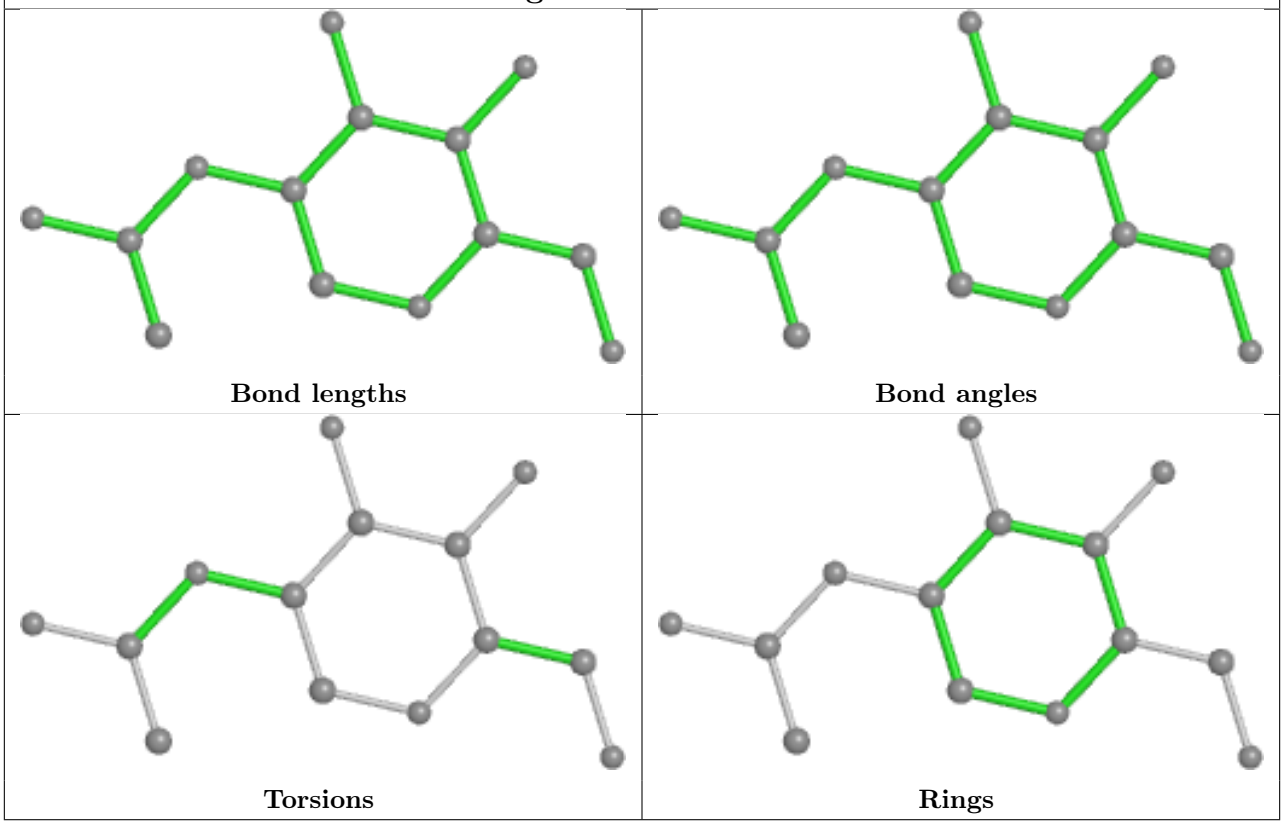


Rings

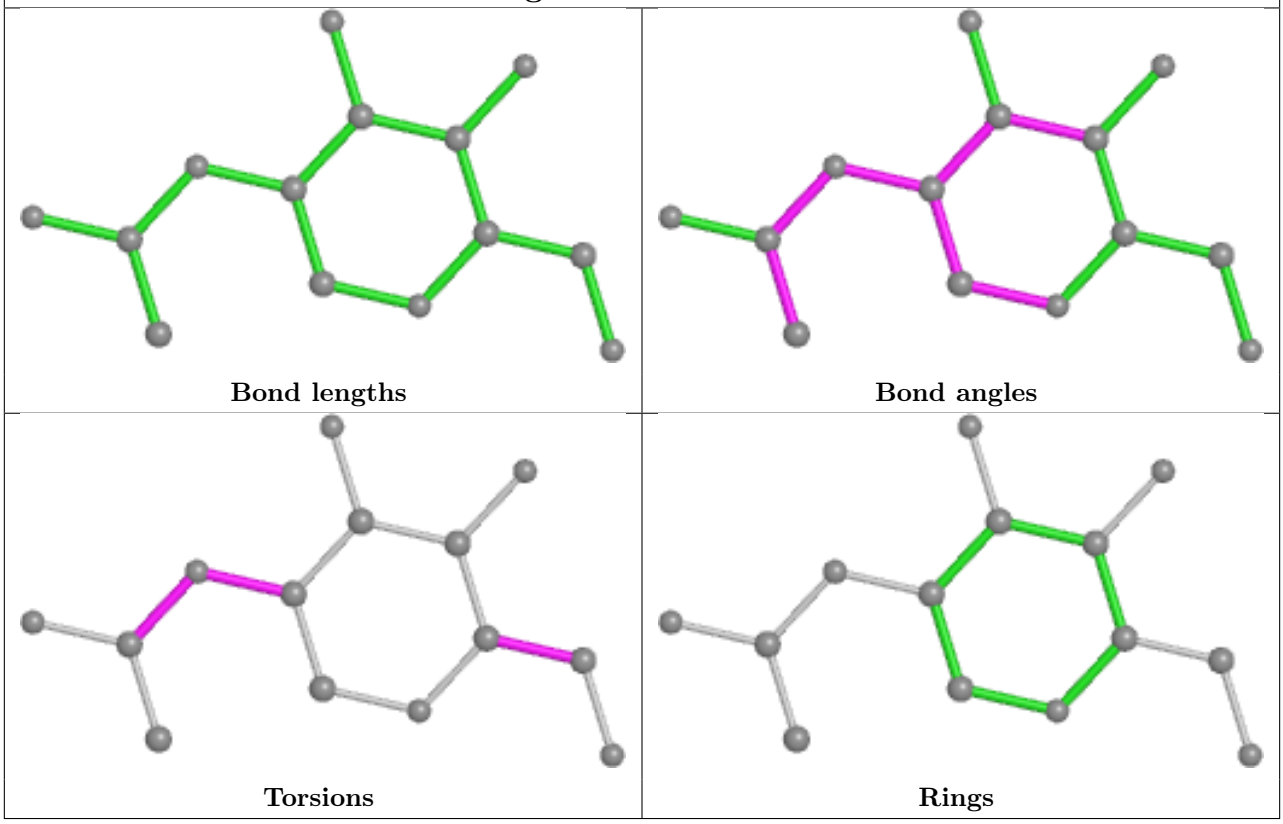
Ligand A2G D 101



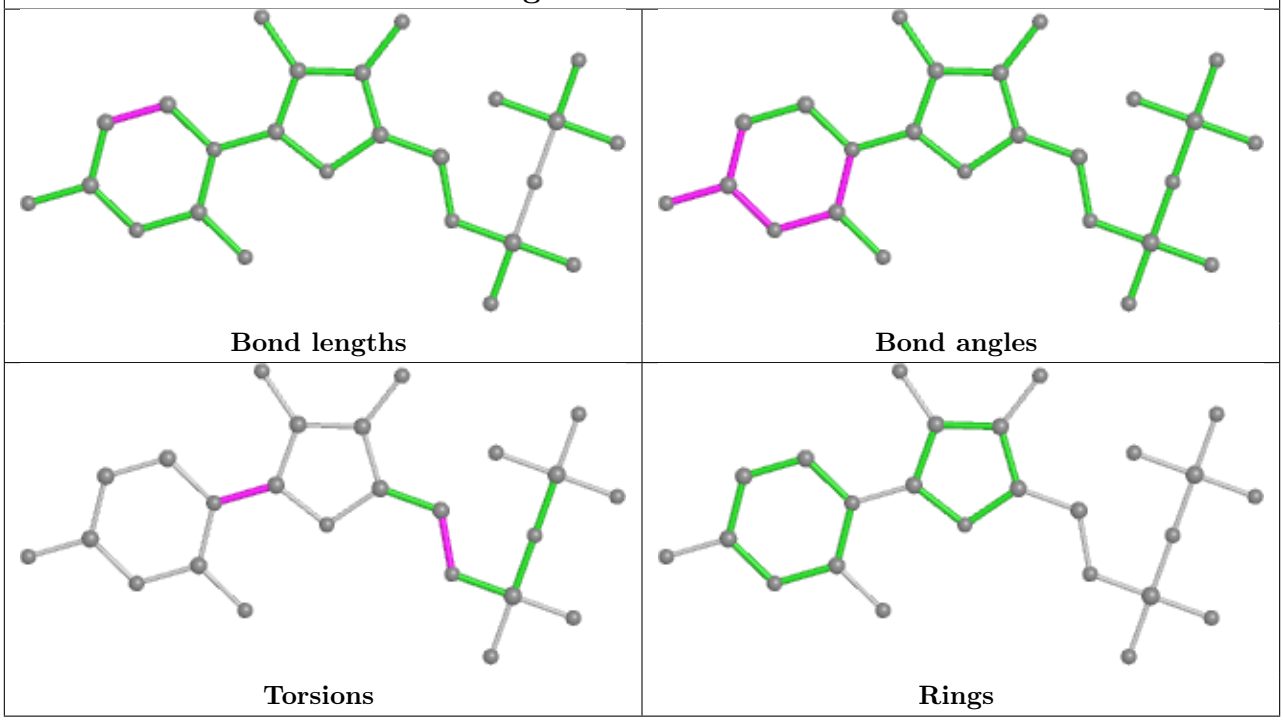
Ligand A2G C 101



Ligand A2G D 102



Ligand UDP B 602



4.7 Other polymers (i)

There are no such residues in this entry.

4.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

5 Fit of model and data (i)

5.1 Protein, DNA and RNA chains (i)

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

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

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

5.3 Carbohydrates (i)

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

5.4 Ligands (i)

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

5.5 Other polymers (i)

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