wwPDB X-ray Structure Validation Summary Report (i)

Nov 21, 2023 – 07:29 AM JST

PDB ID	:	7F7R
Title	:	Enterococcus faecalis GH31 alpha-N-acetylgalactosaminidase D455N in complex with Tn antigen
Authors	:	Miyazaki, T.
Deposited on	:	2021-06-30
Resolution	:	1.63 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:

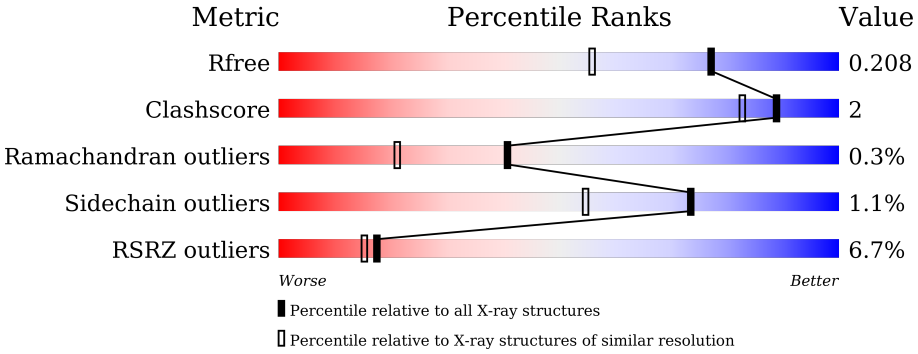
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (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

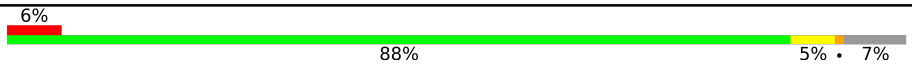
The reported resolution of this entry is 1.63 Å.

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	3122 (1.66-1.62)
Clashscore	141614	3268 (1.66-1.62)
Ramachandran outliers	138981	3215 (1.66-1.62)
Sidechain outliers	138945	3215 (1.66-1.62)
RSRZ outliers	127900	3079 (1.66-1.62)

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	963	

2 Entry composition (i)

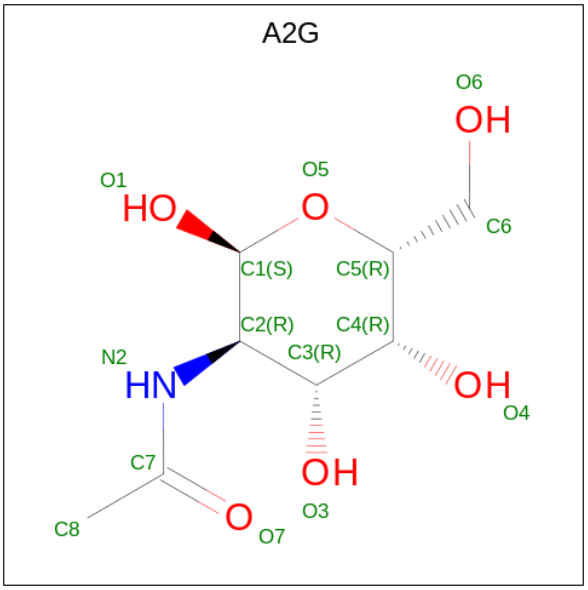
There are 7 unique types of molecules in this entry. The entry contains 7682 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 GH31 alpha-N-acetylgalactosaminidase.

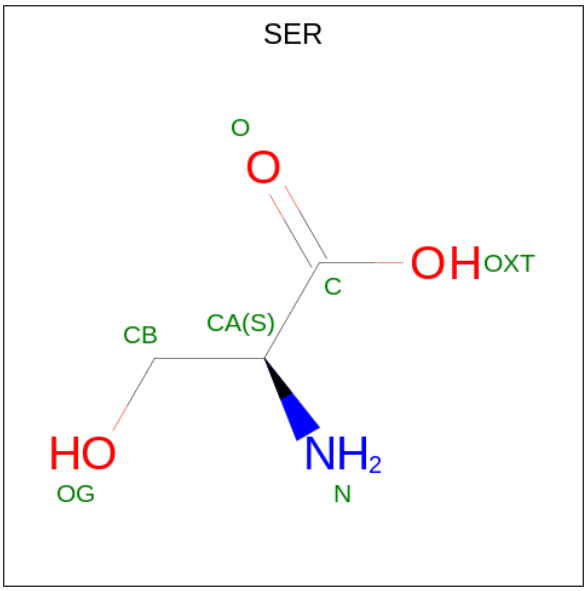
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	899	7044	4453	1174	1401	16	0	1	0

- Molecule 2 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).



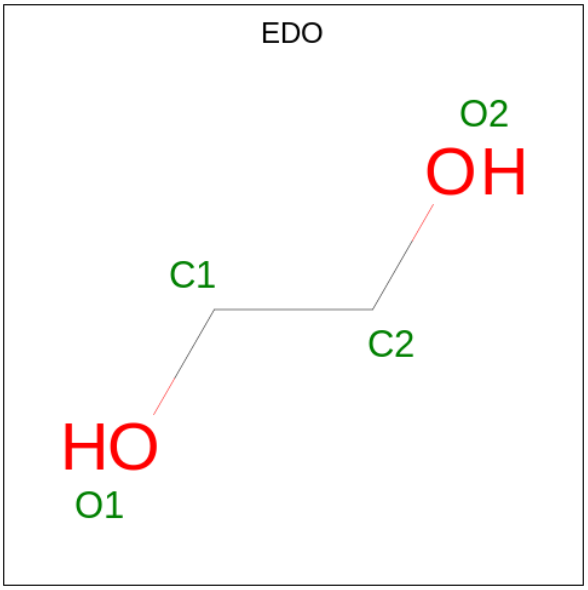
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	A	1	14	8	1	5	0	0

- Molecule 3 is SERINE (three-letter code: SER) (formula: $C_3H_7NO_3$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
3	A	1	7	3	1	3	0	0

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



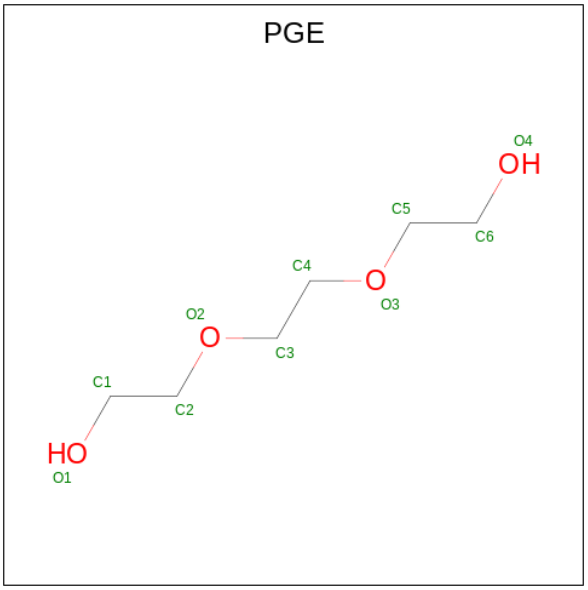
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	O			
4	A	1	4	2	2		0	0
			Total	C	O			
4	A	1	4	2	2		0	0
			Total	C	O			
4	A	1	4	2	2		0	0

Continued on next page...

Continued from previous page...

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

- Molecule 5 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula: $C_6H_{14}O_4$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
5	A	1	10	6	4	0	0

- Molecule 6 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Ca		
6	A	1	1	1	0	0

- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
7	A	558	558	558	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: GH31 alpha-N-acetylgalactosaminidase

Chain A: 6% 88% 5% • 7%

MET GLY SER SER HIS HIS HIS HIS HIS HIS SER SER GLY LEU VAL PRO ARG GLY SER HIS MET GLN GLU GLN THR ALA LYS GLU ASP VAL ALA ASP SER ALA THR S57 V58 V62 I73 T74 I83 L95 D96 P97 T98 E103 Q127 S137 F141 K149

K154 L162 R243 T268 L307 N308 A309 Y310 N311 W315 G343 W379 Q392 E416 W420 T421 Q422 N455 S491 L492 I504 W505 H520 P562 N573 P600 E656 K682 V688 V699 P710 E751 D772 G812

K818 L832 L862 A869 I874 K877 T890 T891 V892 D893 D894 Q895 L896 T897 I898 PRO ALA ASN ILE ALA I904 N905 E906 E907 K908 T909 T910 PRO SER SER L914 T915 L916 Q917 TRP ASP GLN VAL THR GLU ALA THR S926 Y927 E930 R931 D932 G933 F936 I939 Q940

T941 N942 T943 A944 T945 F946 D947 G948 F949 S950 F951 L952 S953 E954 H955 T956 F957 R958 V959 R960 A961 V962 G966 V967 S968 E969 W970 S971 GLU PRO ILE LYS GLY LYS THR GLN ASP ASP PRO TYR LYS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	82.71Å 82.97Å 148.58Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.04 – 1.63 46.00 – 1.63	Depositor EDS
% Data completeness (in resolution range)	100.0 (46.04-1.63) 100.0 (46.00-1.63)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.28 (at 1.63Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.177 , 0.200 0.187 , 0.208	Depositor DCC
R_{free} test set	6325 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å^2)	21.2	Xtriage
Anisotropy	0.317	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 40.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.024 for k,h,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	7682	wwPDB-VP
Average B, all atoms (Å^2)	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.58% 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: EDO, PGE, A2G, CA

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	0.77	3/7208 (0.0%)	0.84	2/9776 (0.0%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	751	GLU	CD-OE2	-6.26	1.18	1.25
1	A	416	GLU	CD-OE2	-5.33	1.19	1.25
1	A	103	GLU	CD-OE2	5.23	1.31	1.25

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	243	ARG	NE-CZ-NH2	-6.48	117.06	120.30
1	A	243	ARG	NE-CZ-NH1	5.06	122.83	120.30

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	7044	0	6735	27	0
2	A	14	0	12	0	0
3	A	7	0	3	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	48	0	72	2	0
5	A	10	0	14	0	0
6	A	1	0	0	0	0
7	A	558	0	0	2	0
All	All	7682	0	6836	27	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 2.

The worst 5 of 27 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:A:898:ILE:HG22	1:A:898:ILE:O	1.86	0.76
1:A:898:ILE:O	1:A:898:ILE:CG2	2.42	0.66
1:A:315:TRP:CZ2	4:A:1011:EDO:H21	2.43	0.53
1:A:127:GLN:NE2	7:A:1104:HOH:O	2.41	0.53
1:A:772:ASP:OD2	1:A:877:LYS:HE3	2.10	0.51

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	892/963 (93%)	862 (97%)	27 (3%)	3 (0%)	41 21

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	897	THR
1	A	970	TRP
1	A	562	PRO

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	752/805 (93%)	744 (99%)	8 (1%)	73 55

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	898	ILE
1	A	877	LYS
1	A	573	ASN
1	A	520	HIS
1	A	656	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	311	ASN
1	A	468	ASN
1	A	761	ASN
1	A	715	ASN
1	A	250	ASN

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 16 ligands modelled in this entry, 1 is monoatomic - leaving 15 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	$\#\|Z\| > 2$
4	EDO	A	1012	-	3,3,3	0.24	0	2,2,2	0.54	0
3	SER	A	1002	2	5,6,6	0.72	0	5,7,7	1.33	0
4	EDO	A	1008	-	3,3,3	0.34	0	2,2,2	0.37	0
4	EDO	A	1009	-	3,3,3	0.19	0	2,2,2	0.44	0
4	EDO	A	1004	-	3,3,3	0.26	0	2,2,2	0.57	0
4	EDO	A	1011	-	3,3,3	0.27	0	2,2,2	0.23	0
4	EDO	A	1014	-	3,3,3	0.09	0	2,2,2	0.28	0
2	A2G	A	1001	3	14,14,15	0.53	0	17,19,21	1.10	2 (11%)
4	EDO	A	1003	-	3,3,3	0.12	0	2,2,2	0.35	0
4	EDO	A	1007	-	3,3,3	0.17	0	2,2,2	0.07	0
4	EDO	A	1005	-	3,3,3	0.21	0	2,2,2	0.38	0
4	EDO	A	1015	-	3,3,3	0.17	0	2,2,2	0.35	0
4	EDO	A	1010	-	3,3,3	0.17	0	2,2,2	0.31	0
5	PGE	A	1013	-	9,9,9	0.18	0	8,8,8	0.18	0
4	EDO	A	1006	-	3,3,3	0.36	0	2,2,2	0.15	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
4	EDO	A	1012	-	-	1/1/1/1	-
3	SER	A	1002	2	-	4/6/6/6	-
4	EDO	A	1008	-	-	0/1/1/1	-
4	EDO	A	1009	-	-	0/1/1/1	-
4	EDO	A	1004	-	-	0/1/1/1	-
4	EDO	A	1011	-	-	1/1/1/1	-
4	EDO	A	1014	-	-	0/1/1/1	-
2	A2G	A	1001	3	-	1/6/23/26	0/1/1/1
4	EDO	A	1003	-	-	0/1/1/1	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	EDO	A	1007	-	-	0/1/1/1	-
4	EDO	A	1005	-	-	0/1/1/1	-
4	EDO	A	1015	-	-	1/1/1/1	-
4	EDO	A	1010	-	-	0/1/1/1	-
5	PGE	A	1013	-	-	0/7/7/7	-
4	EDO	A	1006	-	-	0/1/1/1	-

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	1001	A2G	O5-C1-C2	-2.09	107.98	111.29
2	A	1001	A2G	O4-C4-C3	-2.06	105.58	110.35

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	1002	SER	O-C-CA-CB
3	A	1002	SER	OXT-C-CA-CB
4	A	1011	EDO	O1-C1-C2-O2
4	A	1012	EDO	O1-C1-C2-O2
2	A	1001	A2G	O5-C5-C6-O6

There are no ring outliers.

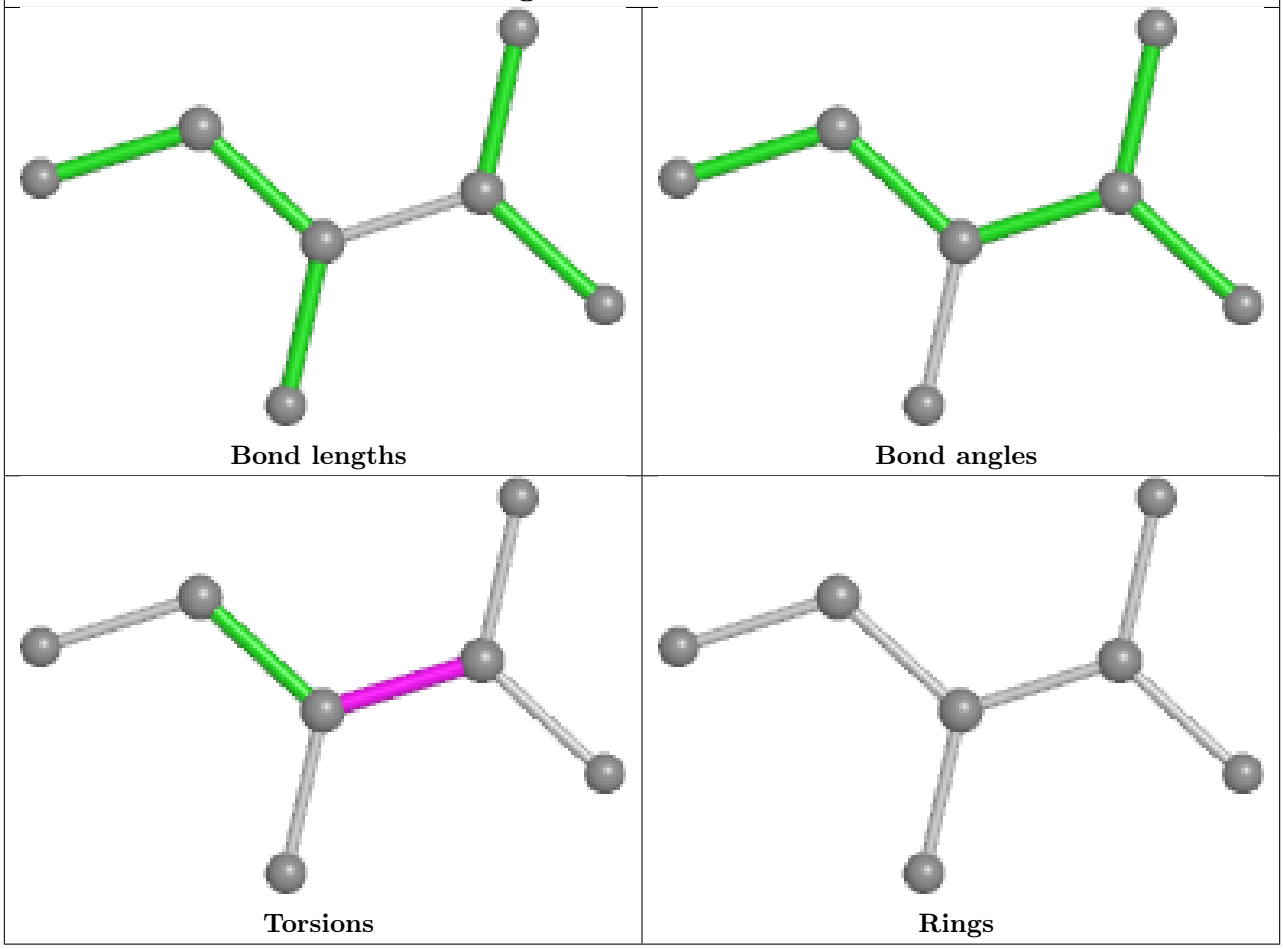
2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1011	EDO	1	0
4	A	1015	EDO	1	0

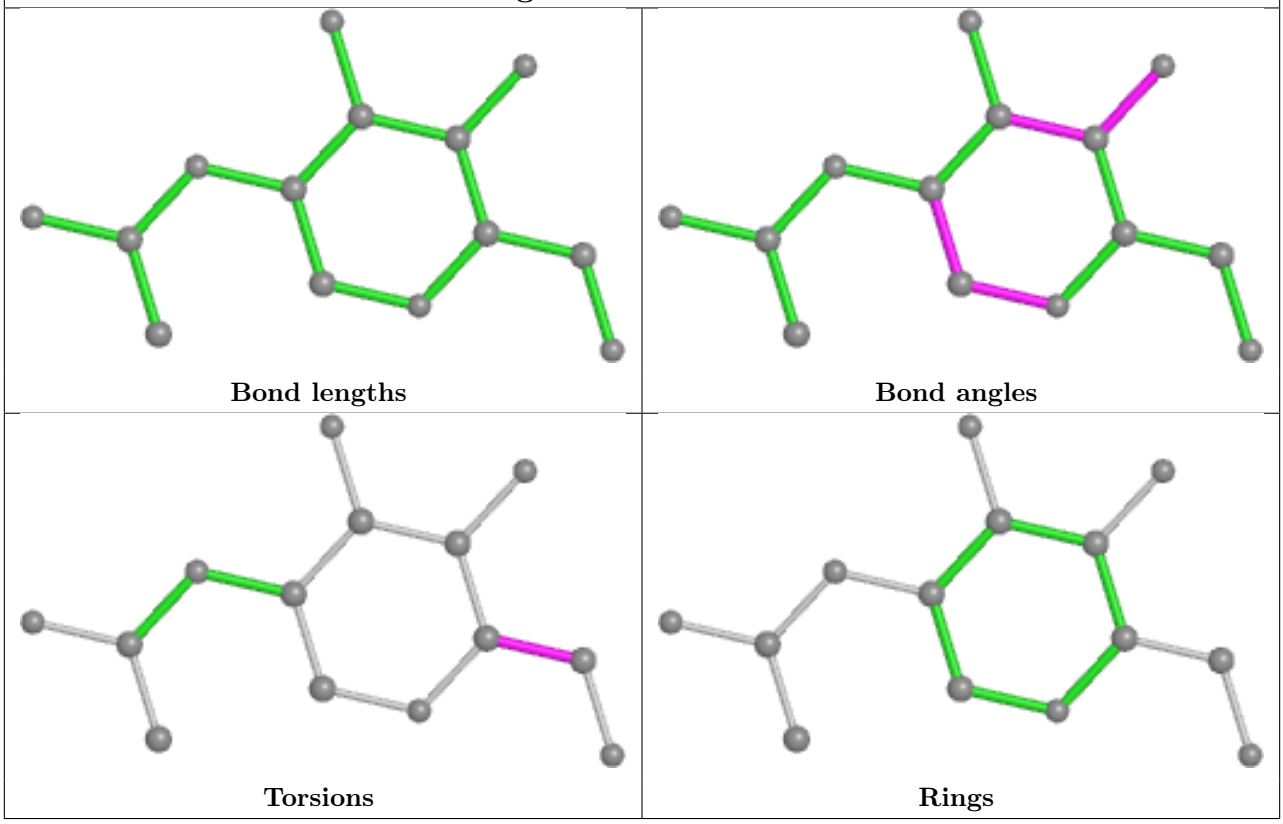
The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. 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 SER A 1002



Ligand A2G A 1001



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	899/963 (93%)	0.11	60 (6%) 17 15	16, 25, 55, 85	0

The worst 5 of 60 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	957	PHE	9.7
1	A	916	LEU	8.6
1	A	904	ILE	8.1
1	A	951	PHE	7.0
1	A	962	VAL	6.6

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
4	EDO	A	1012	4/4	0.81	0.19	45,48,48,50	0
4	EDO	A	1009	4/4	0.84	0.16	35,36,38,41	0

Continued on next page...

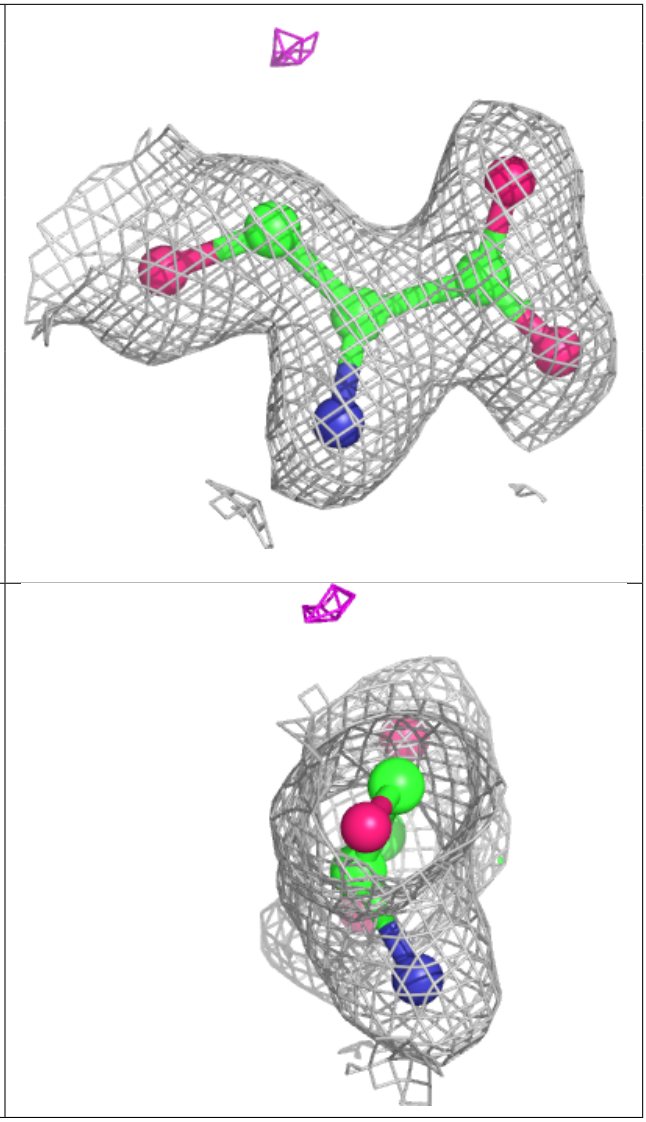
Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
5	PGE	A	1013	10/10	0.84	0.13	46,49,54,56	0
3	SER	A	1002	7/7	0.88	0.11	24,30,34,38	0
4	EDO	A	1008	4/4	0.89	0.09	30,35,36,36	0
4	EDO	A	1003	4/4	0.90	0.12	22,23,25,26	0
4	EDO	A	1007	4/4	0.91	0.14	34,35,37,38	0
4	EDO	A	1015	4/4	0.92	0.15	36,38,40,44	0
4	EDO	A	1011	4/4	0.93	0.11	30,35,35,38	0
4	EDO	A	1006	4/4	0.93	0.11	35,40,41,42	0
4	EDO	A	1005	4/4	0.95	0.07	31,31,32,32	0
4	EDO	A	1010	4/4	0.95	0.14	34,35,35,36	0
6	CA	A	1016	1/1	0.95	0.34	46,46,46,46	0
4	EDO	A	1004	4/4	0.96	0.08	24,25,25,27	0
2	A2G	A	1001	14/15	0.96	0.06	19,20,21,22	0
4	EDO	A	1014	4/4	0.97	0.07	32,41,44,45	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.

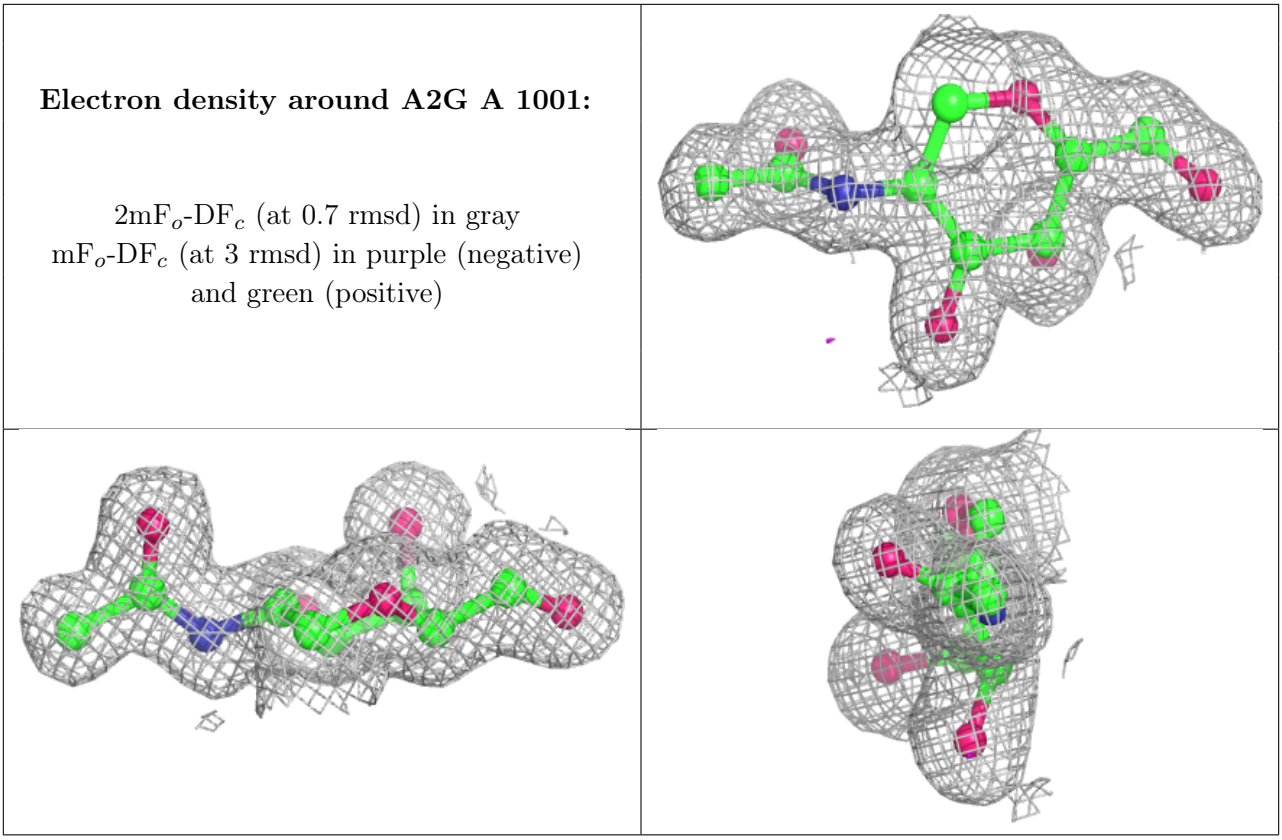
Electron density around SER A 1002:

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



Electron density around A2G A 1001:

$2mF_o$-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.