Full wwPDB X-ray Structure Validation Report (i)

Jan 4, 2024 – 03:49 am GMT

PDB ID	:	5G2T
Title	:	BT1596 in complex with its substrate 4,5 unsaturated uronic acid alpha 1,4 D-Glucosamine-2-N, 6-O-disulfate
Authors	:	Cartmell, A.; Lowe, E.C.; Basle, A.; Crouch, L.I.; Czjzek, M.; Turnbull, J.; Henrissat, B.; Terrapon, N.; Thomas, S.; Murray, H.; Firbank, S.J.; Bolam, D.N.
Deposited on	:	2016-04-13
Resolution	:	1.90 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 (i)) were used in the production of this report:

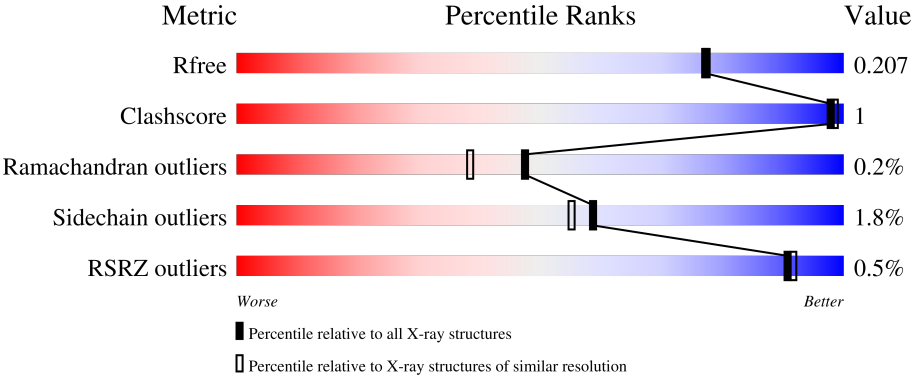
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
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 1.90 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	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (1.90-1.90)

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	489	93% • •
1	B	489	92% • 5%
1	C	489	% 91% • 5%
1	D	489	% 91% • 5%
2	E	2	100%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	2	100%

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	EDO	B	507	-	-	-	X
4	EDO	D	508	-	-	-	X

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 15889 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 2-O GLYCOSAMINOGLYCAN SULFATASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	467	3623	2290	633	679	21	0	3	0
1	B	465	3642	2299	636	686	21	0	6	0
1	C	465	3605	2278	627	679	21	0	1	0
1	D	464	3605	2276	630	678	21	0	1	0

There are 80 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	expression tag	UNP Q8A7C8
A	2	GLY	-	expression tag	UNP Q8A7C8
A	3	SER	-	expression tag	UNP Q8A7C8
A	4	SER	-	expression tag	UNP Q8A7C8
A	5	HIS	-	expression tag	UNP Q8A7C8
A	6	HIS	-	expression tag	UNP Q8A7C8
A	7	HIS	-	expression tag	UNP Q8A7C8
A	8	HIS	-	expression tag	UNP Q8A7C8
A	9	HIS	-	expression tag	UNP Q8A7C8
A	10	HIS	-	expression tag	UNP Q8A7C8
A	11	SER	-	expression tag	UNP Q8A7C8
A	12	SER	-	expression tag	UNP Q8A7C8
A	13	GLY	-	expression tag	UNP Q8A7C8
A	14	PRO	-	expression tag	UNP Q8A7C8
A	15	GLN	-	expression tag	UNP Q8A7C8
A	16	GLN	-	expression tag	UNP Q8A7C8
A	17	GLY	-	expression tag	UNP Q8A7C8
A	18	LEU	-	expression tag	UNP Q8A7C8
A	19	ARG	-	expression tag	UNP Q8A7C8
A	20	GLN	-	expression tag	UNP Q8A7C8
B	1	MET	-	expression tag	UNP Q8A7C8

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	2	GLY	-	expression tag	UNP Q8A7C8
B	3	SER	-	expression tag	UNP Q8A7C8
B	4	SER	-	expression tag	UNP Q8A7C8
B	5	HIS	-	expression tag	UNP Q8A7C8
B	6	HIS	-	expression tag	UNP Q8A7C8
B	7	HIS	-	expression tag	UNP Q8A7C8
B	8	HIS	-	expression tag	UNP Q8A7C8
B	9	HIS	-	expression tag	UNP Q8A7C8
B	10	HIS	-	expression tag	UNP Q8A7C8
B	11	SER	-	expression tag	UNP Q8A7C8
B	12	SER	-	expression tag	UNP Q8A7C8
B	13	GLY	-	expression tag	UNP Q8A7C8
B	14	PRO	-	expression tag	UNP Q8A7C8
B	15	GLN	-	expression tag	UNP Q8A7C8
B	16	GLN	-	expression tag	UNP Q8A7C8
B	17	GLY	-	expression tag	UNP Q8A7C8
B	18	LEU	-	expression tag	UNP Q8A7C8
B	19	ARG	-	expression tag	UNP Q8A7C8
B	20	GLN	-	expression tag	UNP Q8A7C8
C	1	MET	-	expression tag	UNP Q8A7C8
C	2	GLY	-	expression tag	UNP Q8A7C8
C	3	SER	-	expression tag	UNP Q8A7C8
C	4	SER	-	expression tag	UNP Q8A7C8
C	5	HIS	-	expression tag	UNP Q8A7C8
C	6	HIS	-	expression tag	UNP Q8A7C8
C	7	HIS	-	expression tag	UNP Q8A7C8
C	8	HIS	-	expression tag	UNP Q8A7C8
C	9	HIS	-	expression tag	UNP Q8A7C8
C	10	HIS	-	expression tag	UNP Q8A7C8
C	11	SER	-	expression tag	UNP Q8A7C8
C	12	SER	-	expression tag	UNP Q8A7C8
C	13	GLY	-	expression tag	UNP Q8A7C8
C	14	PRO	-	expression tag	UNP Q8A7C8
C	15	GLN	-	expression tag	UNP Q8A7C8
C	16	GLN	-	expression tag	UNP Q8A7C8
C	17	GLY	-	expression tag	UNP Q8A7C8
C	18	LEU	-	expression tag	UNP Q8A7C8
C	19	ARG	-	expression tag	UNP Q8A7C8
C	20	GLN	-	expression tag	UNP Q8A7C8
D	1	MET	-	expression tag	UNP Q8A7C8
D	2	GLY	-	expression tag	UNP Q8A7C8
D	3	SER	-	expression tag	UNP Q8A7C8

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	4	SER	-	expression tag	UNP Q8A7C8
D	5	HIS	-	expression tag	UNP Q8A7C8
D	6	HIS	-	expression tag	UNP Q8A7C8
D	7	HIS	-	expression tag	UNP Q8A7C8
D	8	HIS	-	expression tag	UNP Q8A7C8
D	9	HIS	-	expression tag	UNP Q8A7C8
D	10	HIS	-	expression tag	UNP Q8A7C8
D	11	SER	-	expression tag	UNP Q8A7C8
D	12	SER	-	expression tag	UNP Q8A7C8
D	13	GLY	-	expression tag	UNP Q8A7C8
D	14	PRO	-	expression tag	UNP Q8A7C8
D	15	GLN	-	expression tag	UNP Q8A7C8
D	16	GLN	-	expression tag	UNP Q8A7C8
D	17	GLY	-	expression tag	UNP Q8A7C8
D	18	LEU	-	expression tag	UNP Q8A7C8
D	19	ARG	-	expression tag	UNP Q8A7C8
D	20	GLN	-	expression tag	UNP Q8A7C8

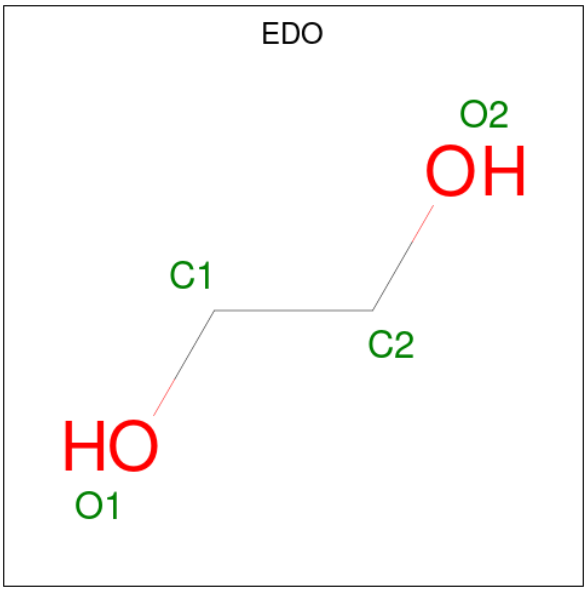
- Molecule 2 is an oligosaccharide called 4-deoxy-2-O-sulfo-alpha-L-threo-hex-4-enopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo-2-(sulfoamino)-alpha-D-glucopyranose.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	E	2	35	12	1	19	3	0	0	0
			Total	C	N	O	S			
2	F	2	35	12	1	19	3	0	0	0

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Zn		
3	A	1	1	1	0	0
			Total	Zn		
3	B	1	1	1	0	0
			Total	Zn		
3	C	1	1	1	0	0
			Total	Zn		
3	D	1	1	1	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
4	A	1	4	2	2	0	0
4	A	1	4	2	2	0	0
4	A	1	4	2	2	0	0
4	A	1	4	2	2	0	0
4	A	1	4	2	2	0	0
4	A	1	4	2	2	0	0
4	A	1	4	2	2	0	0
4	B	1	4	2	2	0	0
4	B	1	4	2	2	0	0
4	B	1	4	2	2	0	0
4	B	1	4	2	2	0	0
4	B	1	4	2	2	0	0
4	B	1	4	2	2	0	0

Continued on next page...

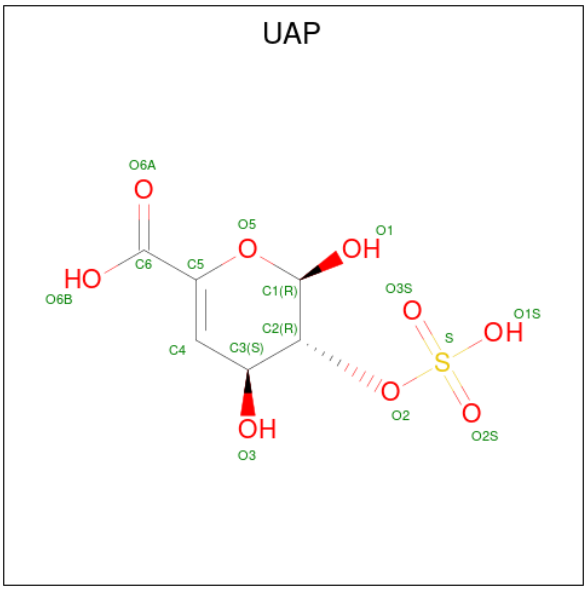
Continued from previous page...

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

- Molecule 5 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total Cl 1 1	0	0
5	B	1	Total Cl 1 1	0	0
5	C	1	Total Cl 1 1	0	0
5	D	1	Total Cl 1 1	0	0

- Molecule 6 is 4-deoxy-2-O-sulfo-alpha-L-threo-hex-4-enopyranuronic acid (three-letter code: UAP) (formula: $C_6H_8O_9S$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	C	1	Total C O S 16 6 9 1	0	0
6	D	1	Total C O S 16 6 9 1	0	0

- Molecule 7 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	A	326	Total O 326 326	0	0
7	B	301	Total O 301 301	0	0
7	C	279	Total O 279 279	0	0

Continued on next page...

Continued from previous page...

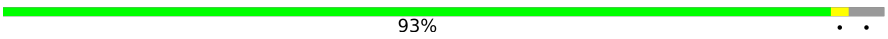
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
7	D	266	266	266	0	0

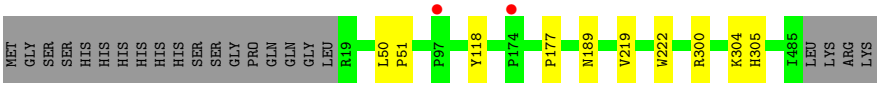
3 Residue-property plots (i)

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: 2-O GLYCOSAMINOGLYCAN SULFATASE

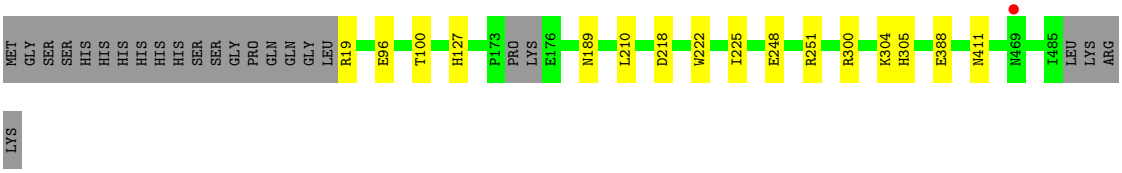
Chain A:





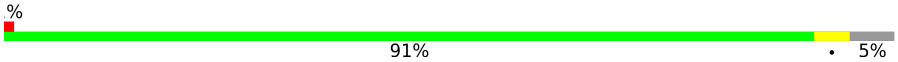
• Molecule 1: 2-O GLYCOSAMINOGLYCAN SULFATASE

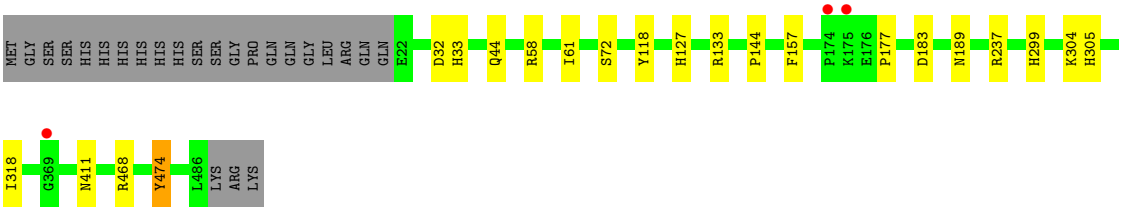
Chain B:



• Molecule 1: 2-O GLYCOSAMINOGLYCAN SULFATASE

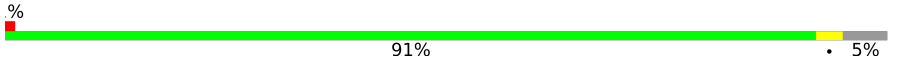
Chain C:

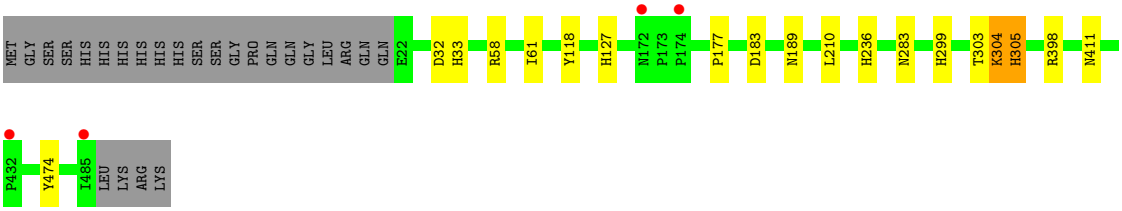




• Molecule 1: 2-O GLYCOSAMINOGLYCAN SULFATASE

Chain D:





• Molecule 2: 4-deoxy-2-O-sulfo-alpha-L-threo-hex-4-enopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo -2-(sulfoamino)-alpha-D-glucopyranose

Chain E: 100%

SGN1
UAP2

• Molecule 2: 4-deoxy-2-O-sulfo-alpha-L-threo-hex-4-enopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo -2-(sulfoamino)-alpha-D-glucopyranose

Chain F: 100%

SGN1
UAP2

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	69.21Å 114.53Å 127.29Å 90.00° 100.04° 90.00°	Depositor
Resolution (Å)	125.34 – 1.90 46.40 – 1.90	Depositor EDS
% Data completeness (in resolution range)	100.0 (125.34-1.90) 100.0 (46.40-1.90)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.80 (at 1.90Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
R, R_{free}	0.160 , 0.198 0.170 , 0.207	Depositor DCC
R_{free} test set	7624 reflections (4.97%)	wwPDB-VP
Wilson B-factor (Å^2)	16.8	Xtriage
Anisotropy	0.635	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 50.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	15889	wwPDB-VP
Average B, all atoms (Å^2)	21.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 62.82 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.0628e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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: SGN, ZN, UAP, CL, EDO

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	0.59	0/3736	0.74	2/5096 (0.0%)
1	B	0.60	0/3762	0.72	2/5125 (0.0%)
1	C	0.59	0/3712	0.73	3/5061 (0.1%)
1	D	0.60	0/3712	0.73	2/5060 (0.0%)
All	All	0.60	0/14922	0.73	9/20342 (0.0%)

There are no bond length outliers.

All (9) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	133	ARG	NE-CZ-NH2	-6.42	117.09	120.30
1	A	300	ARG	NE-CZ-NH1	5.84	123.22	120.30
1	B	300	ARG	NE-CZ-NH1	5.67	123.14	120.30
1	B	300	ARG	NE-CZ-NH2	-5.41	117.59	120.30
1	C	58	ARG	NE-CZ-NH1	5.24	122.92	120.30
1	D	58	ARG	NE-CZ-NH1	5.21	122.90	120.30
1	A	300	ARG	NE-CZ-NH2	-5.16	117.72	120.30
1	D	398	ARG	NE-CZ-NH1	5.08	122.84	120.30
1	C	468	ARG	NE-CZ-NH2	-5.06	117.77	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	3623	0	3442	3	0
1	B	3642	0	3478	3	0
1	C	3605	0	3420	8	0
1	D	3605	0	3424	5	0
2	E	35	0	13	0	0
2	F	35	0	13	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
3	D	1	0	0	0	0
4	A	32	0	48	0	1
4	B	36	0	54	0	1
4	C	36	0	54	0	0
4	D	28	0	42	0	1
5	A	1	0	0	0	0
5	B	1	0	0	0	0
5	C	1	0	0	0	0
5	D	1	0	0	0	0
6	C	16	0	7	0	0
6	D	16	0	7	0	0
7	A	326	0	0	0	0
7	B	301	0	0	2	1
7	C	279	0	0	2	2
7	D	266	0	0	1	0
All	All	15889	0	14002	19	3

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 1.

All (19) 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:B:251:ARG:NH1	7:B:2157:HOH:O	1.92	0.74
1:B:19:ARG:CB	7:B:2001:HOH:O	2.56	0.53
1:D:32:ASP:O	1:D:33:HIS:HB2	2.09	0.53
1:A:118:TYR:CE2	1:A:177:PRO:HB2	2.45	0.52
1:C:72:SER:OG	1:C:127:HIS:HD2	1.93	0.50
1:C:61:ILE:HG22	1:C:318:ILE:HG12	1.92	0.50
1:C:32:ASP:O	1:C:33:HIS:HB2	2.13	0.49
1:D:299:HIS:HE1	7:D:2145:HOH:O	1.98	0.45
1:C:44:GLN:HG2	7:C:2012:HOH:O	2.15	0.45
1:C:118:TYR:CE2	1:C:177:PRO:HB2	2.51	0.45

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:118:TYR:CE2	1:D:177:PRO:HB2	2.53	0.44
1:C:299:HIS:HE1	7:C:2151:HOH:O	1.99	0.44
1:C:144:PRO:HA	1:C:157:PHE:CD1	2.54	0.43
1:C:474:TYR:CD1	1:C:474:TYR:N	2.86	0.42
1:B:222:TRP:CE3	1:B:225:ILE:HD11	2.54	0.42
1:D:236:HIS:CE1	1:D:303:THR:OG1	2.73	0.42
1:A:219:VAL:HG21	1:A:222:TRP:CH2	2.55	0.42
1:A:50[B]:LEU:HB3	1:A:51:PRO:HD3	2.03	0.40
1:D:304:LYS:O	1:D:305:HIS:HB2	2.22	0.40

All (3) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:B:507:EDO:O2	7:C:2188:HOH:O[2_453]	2.00	0.20
4:A:508:EDO:O2	4:D:508:EDO:O2[2_444]	2.08	0.12
7:B:2277:HOH:O	7:C:2156:HOH:O[1_655]	2.11	0.09

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	468/489 (96%)	458 (98%)	9 (2%)	1 (0%)	47	38
1	B	467/489 (96%)	458 (98%)	8 (2%)	1 (0%)	47	38
1	C	464/489 (95%)	450 (97%)	13 (3%)	1 (0%)	47	38
1	D	463/489 (95%)	453 (98%)	9 (2%)	1 (0%)	47	38
All	All	1862/1956 (95%)	1819 (98%)	39 (2%)	4 (0%)	47	38

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	305	HIS
1	C	305	HIS
1	D	305	HIS
1	A	305	HIS

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	387/424 (91%)	385 (100%)	2 (0%)	88	89
1	B	394/424 (93%)	384 (98%)	10 (2%)	47	41
1	C	385/424 (91%)	379 (98%)	6 (2%)	62	60
1	D	386/424 (91%)	377 (98%)	9 (2%)	50	45
All	All	1552/1696 (92%)	1525 (98%)	27 (2%)	59	57

All (27) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	189	ASN
1	A	304	LYS
1	B	96	GLU
1	B	100	THR
1	B	127	HIS
1	B	189	ASN
1	B	210	LEU
1	B	218	ASP
1	B	248	GLU
1	B	304	LYS
1	B	388	GLU
1	B	411	ASN
1	C	183	ASP
1	C	189	ASN
1	C	237	ARG
1	C	304	LYS
1	C	411	ASN
1	C	474	TYR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	D	61	ILE
1	D	127	HIS
1	D	183	ASP
1	D	189	ASN
1	D	210	LEU
1	D	283	ASN
1	D	304	LYS
1	D	411	ASN
1	D	474	TYR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (9) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	172	ASN
1	A	299	HIS
1	B	299	HIS
1	C	127	HIS
1	C	299	HIS
1	D	89	ASN
1	D	236	HIS
1	D	299	HIS
1	D	384	HIS

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)

4 monosaccharides 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	$\#\|Z\| > 2$
2	SGN	E	1	2	18,19,20	1.51	2 (11%)	22,29,31	1.89	4 (18%)
2	UAP	E	2	2	16,16,16	2.06	5 (31%)	19,24,24	1.68	4 (21%)
2	SGN	F	1	2	18,19,20	4.78	4 (22%)	22,29,31	1.69	4 (18%)
2	UAP	F	2	2	16,16,16	1.88	2 (12%)	19,24,24	1.68	4 (21%)

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	SGN	E	1	2	-	1/11/27/31	0/1/1/1
2	UAP	E	2	2	-	0/9/25/25	0/1/1/1
2	SGN	F	1	2	-	0/11/27/31	0/1/1/1
2	UAP	F	2	2	-	0/9/25/25	0/1/1/1

All (13) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	F	1	SGN	O2S-S1	19.30	1.64	1.42
2	E	2	UAP	C5-C6	-5.83	1.34	1.48
2	F	2	UAP	C5-C6	-5.52	1.35	1.48
2	F	1	SGN	O4S-S2	4.54	1.64	1.45
2	E	1	SGN	O4S-S2	4.53	1.64	1.45
2	E	2	UAP	C4-C5	2.44	1.36	1.33
2	E	1	SGN	O2S-S1	2.23	1.44	1.42
2	F	2	UAP	O5-C5	2.20	1.46	1.37
2	F	1	SGN	C1-C2	2.12	1.55	1.52
2	E	2	UAP	C3-C4	2.12	1.53	1.50
2	E	2	UAP	O6B-C6	-2.11	1.24	1.30
2	E	2	UAP	O5-C5	2.06	1.45	1.37
2	F	1	SGN	O1S-S1	2.06	1.44	1.42

All (16) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	E	1	SGN	O5-C1-C2	5.25	114.79	109.52
2	F	1	SGN	O1S-S1-O2S	-4.98	108.39	120.16

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	E	1	SGN	O1S-S1-O2S	-4.47	109.59	120.16
2	E	2	UAP	O1-C1-O5	-4.31	104.58	110.18
2	F	1	SGN	O5-C1-C2	4.07	113.61	109.52
2	E	2	UAP	O2-C2-C1	3.81	112.70	107.58
2	E	1	SGN	C4-C3-C2	-3.65	107.05	111.46
2	F	2	UAP	O2-C2-C1	3.64	112.47	107.58
2	F	2	UAP	O1-C1-O5	-3.52	105.61	110.18
2	F	2	UAP	O6B-C6-C5	2.73	121.02	114.20
2	F	1	SGN	C4-C3-C2	-2.44	108.51	111.46
2	F	2	UAP	O5-C5-C4	-2.40	121.81	124.61
2	E	2	UAP	O6B-C6-C5	2.32	120.00	114.20
2	E	1	SGN	C1-O5-C5	2.26	117.88	113.66
2	E	2	UAP	O5-C5-C4	-2.14	122.11	124.61
2	F	1	SGN	C1-C2-C3	2.02	113.30	110.54

There are no chirality outliers.

All (1) torsion outliers are listed below:

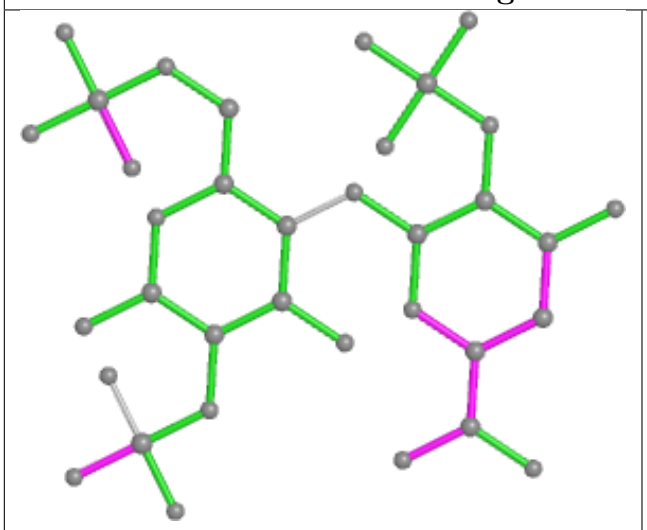
Mol	Chain	Res	Type	Atoms
2	E	1	SGN	C6-O6-S2-O5S

There are no ring outliers.

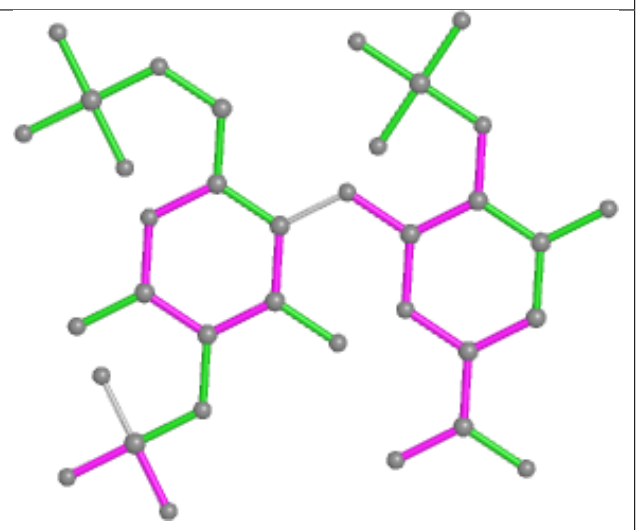
No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for oligosaccharide.

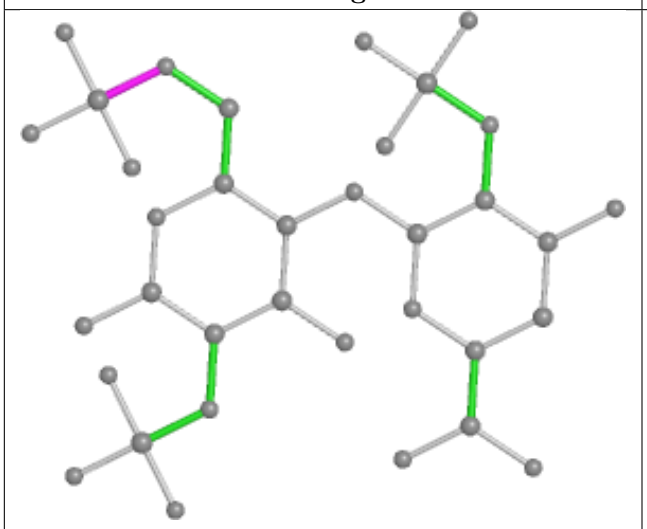
Oligosaccharide Chain E



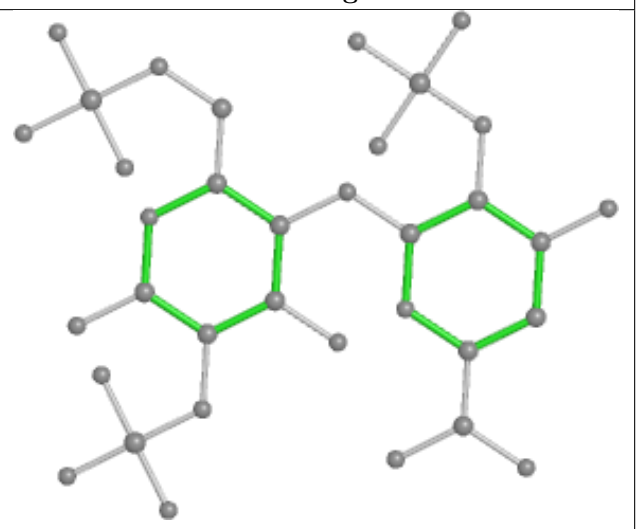
Bond lengths



Bond angles

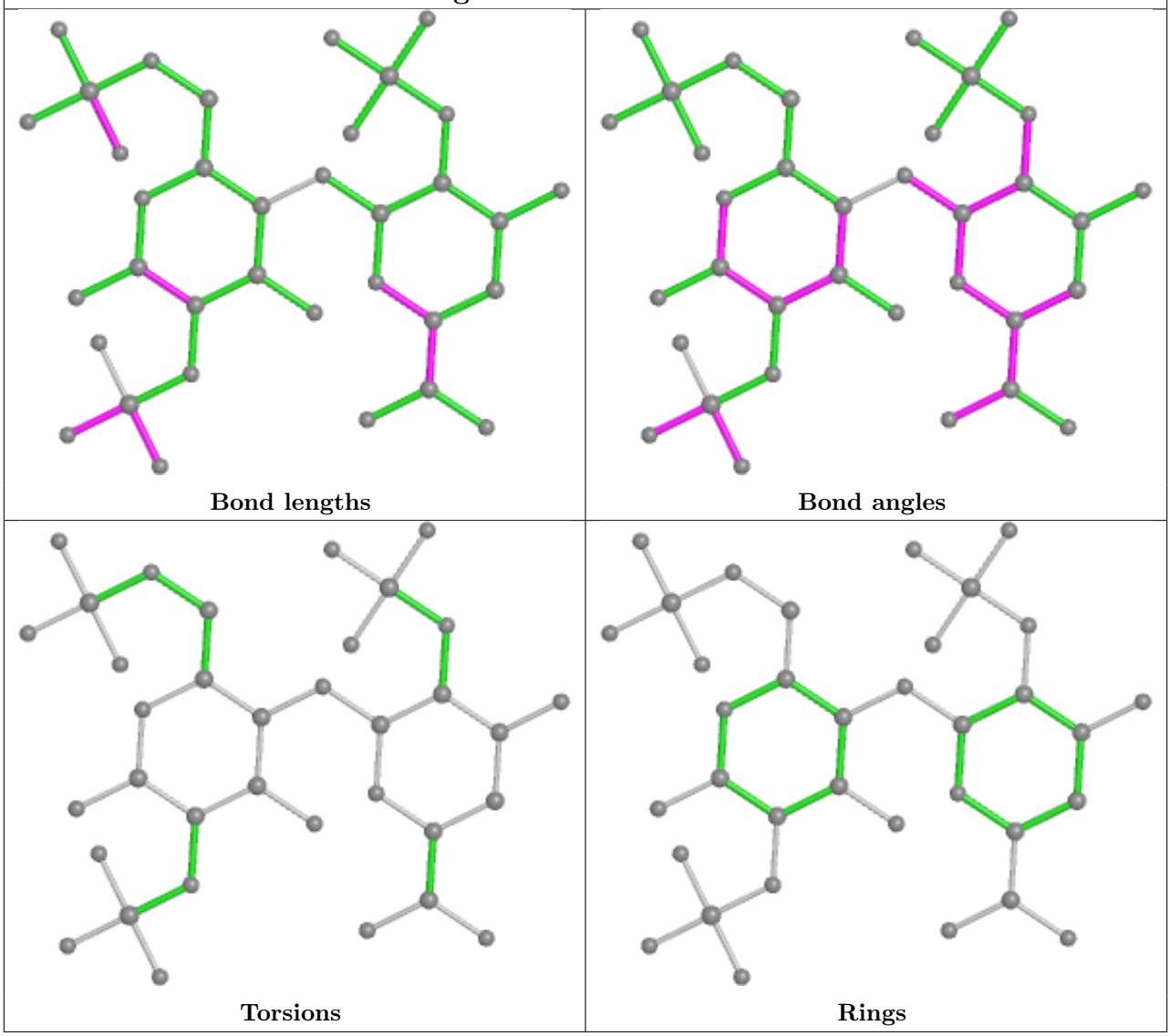


Torsions



Rings

Oligosaccharide Chain F



5.6 Ligand geometry (i)

Of 43 ligands modelled in this entry, 8 are monoatomic - leaving 35 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	505	-	3,3,3	0.56	0	2,2,2	0.13	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
4	EDO	A	506	-	3,3,3	0.45	0	2,2,2	0.36	0
4	EDO	C	502	-	3,3,3	0.42	0	2,2,2	0.13	0
4	EDO	C	510	-	3,3,3	0.51	0	2,2,2	0.28	0
4	EDO	B	504	-	3,3,3	0.51	0	2,2,2	0.30	0
4	EDO	B	508	-	3,3,3	0.45	0	2,2,2	0.30	0
4	EDO	B	510	-	3,3,3	0.53	0	2,2,2	0.62	0
4	EDO	C	504	-	3,3,3	0.25	0	2,2,2	0.38	0
4	EDO	B	502	-	3,3,3	0.34	0	2,2,2	0.54	0
6	UAP	D	510	-	16,16,16	2.31	4 (25%)	19,24,24	2.04	4 (21%)
4	EDO	C	506	-	3,3,3	0.66	0	2,2,2	0.15	0
4	EDO	D	502	-	3,3,3	0.49	0	2,2,2	0.30	0
4	EDO	C	508	-	3,3,3	0.52	0	2,2,2	0.25	0
4	EDO	A	503	-	3,3,3	0.58	0	2,2,2	0.49	0
4	EDO	A	504	-	3,3,3	0.50	0	2,2,2	0.76	0
4	EDO	B	503	-	3,3,3	0.38	0	2,2,2	0.55	0
4	EDO	B	505	-	3,3,3	0.41	0	2,2,2	0.57	0
4	EDO	B	507	-	3,3,3	0.40	0	2,2,2	0.09	0
4	EDO	A	509	-	3,3,3	0.46	0	2,2,2	0.35	0
4	EDO	D	507	-	3,3,3	0.39	0	2,2,2	0.54	0
4	EDO	C	507	-	3,3,3	0.46	0	2,2,2	0.41	0
4	EDO	A	502	-	3,3,3	0.45	0	2,2,2	0.13	0
4	EDO	D	505	-	3,3,3	0.58	0	2,2,2	0.13	0
4	EDO	D	506	-	3,3,3	0.50	0	2,2,2	0.18	0
4	EDO	B	506	-	3,3,3	0.51	0	2,2,2	0.29	0
4	EDO	C	505	-	3,3,3	0.47	0	2,2,2	0.01	0
4	EDO	B	509	-	3,3,3	0.41	0	2,2,2	0.26	0
4	EDO	A	508	-	3,3,3	0.27	0	2,2,2	0.24	0
4	EDO	D	503	-	3,3,3	0.63	0	2,2,2	0.19	0
4	EDO	C	503	-	3,3,3	0.46	0	2,2,2	0.22	0
4	EDO	D	504	-	3,3,3	0.50	0	2,2,2	0.70	0
4	EDO	D	508	-	3,3,3	0.54	0	2,2,2	0.13	0
4	EDO	C	509	-	3,3,3	0.41	0	2,2,2	0.53	0
4	EDO	A	507	-	3,3,3	0.36	0	2,2,2	0.47	0
6	UAP	C	511	-	16,16,16	2.20	5 (31%)	19,24,24	2.30	6 (31%)

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	505	-	-	0/1/1/1	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	EDO	A	506	-	-	0/1/1/1	-
4	EDO	C	502	-	-	1/1/1/1	-
4	EDO	C	510	-	-	1/1/1/1	-
4	EDO	B	504	-	-	0/1/1/1	-
4	EDO	B	508	-	-	0/1/1/1	-
4	EDO	B	510	-	-	1/1/1/1	-
4	EDO	C	504	-	-	0/1/1/1	-
4	EDO	B	502	-	-	0/1/1/1	-
6	UAP	D	510	-	-	0/9/25/25	0/1/1/1
4	EDO	C	506	-	-	0/1/1/1	-
4	EDO	D	502	-	-	0/1/1/1	-
4	EDO	C	508	-	-	0/1/1/1	-
4	EDO	A	503	-	-	1/1/1/1	-
4	EDO	A	504	-	-	0/1/1/1	-
4	EDO	B	503	-	-	0/1/1/1	-
4	EDO	B	505	-	-	0/1/1/1	-
4	EDO	B	507	-	-	1/1/1/1	-
4	EDO	A	509	-	-	1/1/1/1	-
4	EDO	D	507	-	-	0/1/1/1	-
4	EDO	C	507	-	-	0/1/1/1	-
4	EDO	A	502	-	-	0/1/1/1	-
4	EDO	D	505	-	-	0/1/1/1	-
4	EDO	D	506	-	-	1/1/1/1	-
4	EDO	B	506	-	-	1/1/1/1	-
4	EDO	C	505	-	-	0/1/1/1	-
4	EDO	B	509	-	-	1/1/1/1	-
4	EDO	A	508	-	-	1/1/1/1	-
4	EDO	D	503	-	-	0/1/1/1	-
4	EDO	C	503	-	-	0/1/1/1	-
4	EDO	D	504	-	-	0/1/1/1	-
4	EDO	D	508	-	-	1/1/1/1	-
4	EDO	C	509	-	-	0/1/1/1	-
4	EDO	A	507	-	-	0/1/1/1	-
6	UAP	C	511	-	-	0/9/25/25	0/1/1/1

All (9) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	D	510	UAP	C5-C6	-5.52	1.35	1.48
6	C	511	UAP	C5-C6	-5.43	1.35	1.48
6	D	510	UAP	C1-C2	4.74	1.56	1.52
6	C	511	UAP	C1-C2	3.53	1.55	1.52

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	C	511	UAP	C3-C4	2.93	1.54	1.50
6	D	510	UAP	C4-C5	2.79	1.37	1.33
6	D	510	UAP	O5-C5	2.39	1.47	1.37
6	C	511	UAP	O5-C5	2.32	1.46	1.37
6	C	511	UAP	C4-C5	2.21	1.36	1.33

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
6	C	511	UAP	O2-C2-C1	5.38	114.81	107.58
6	D	510	UAP	O1-C1-O5	-5.20	103.42	110.18
6	C	511	UAP	O1-C1-O5	-4.95	103.74	110.18
6	D	510	UAP	O2-C2-C1	4.83	114.08	107.58
6	C	511	UAP	O6B-C6-C5	3.80	123.68	114.20
6	C	511	UAP	O3-C3-C4	2.32	114.53	109.31
6	C	511	UAP	O5-C5-C4	-2.22	122.02	124.61
6	D	510	UAP	C1-C2-C3	2.21	114.18	110.36
6	D	510	UAP	C1-O5-C5	2.19	118.67	114.43
6	C	511	UAP	O6B-C6-O6A	-2.14	118.72	123.61

There are no chirality outliers.

All (11) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	C	502	EDO	O1-C1-C2-O2
4	D	508	EDO	O1-C1-C2-O2
4	A	508	EDO	O1-C1-C2-O2
4	D	506	EDO	O1-C1-C2-O2
4	A	503	EDO	O1-C1-C2-O2
4	B	510	EDO	O1-C1-C2-O2
4	B	506	EDO	O1-C1-C2-O2
4	C	510	EDO	O1-C1-C2-O2
4	A	509	EDO	O1-C1-C2-O2
4	B	509	EDO	O1-C1-C2-O2
4	B	507	EDO	O1-C1-C2-O2

There are no ring outliers.

3 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	B	507	EDO	0	1

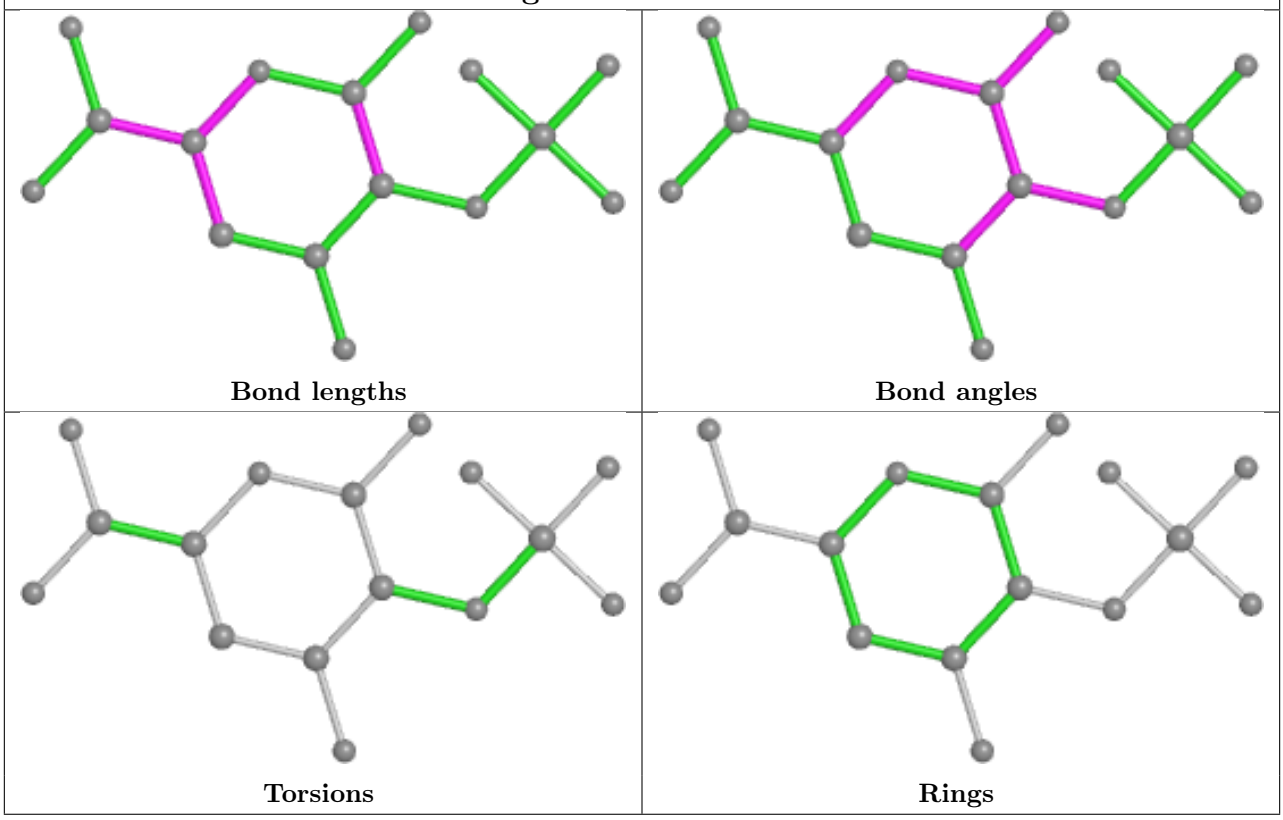
Continued on next page...

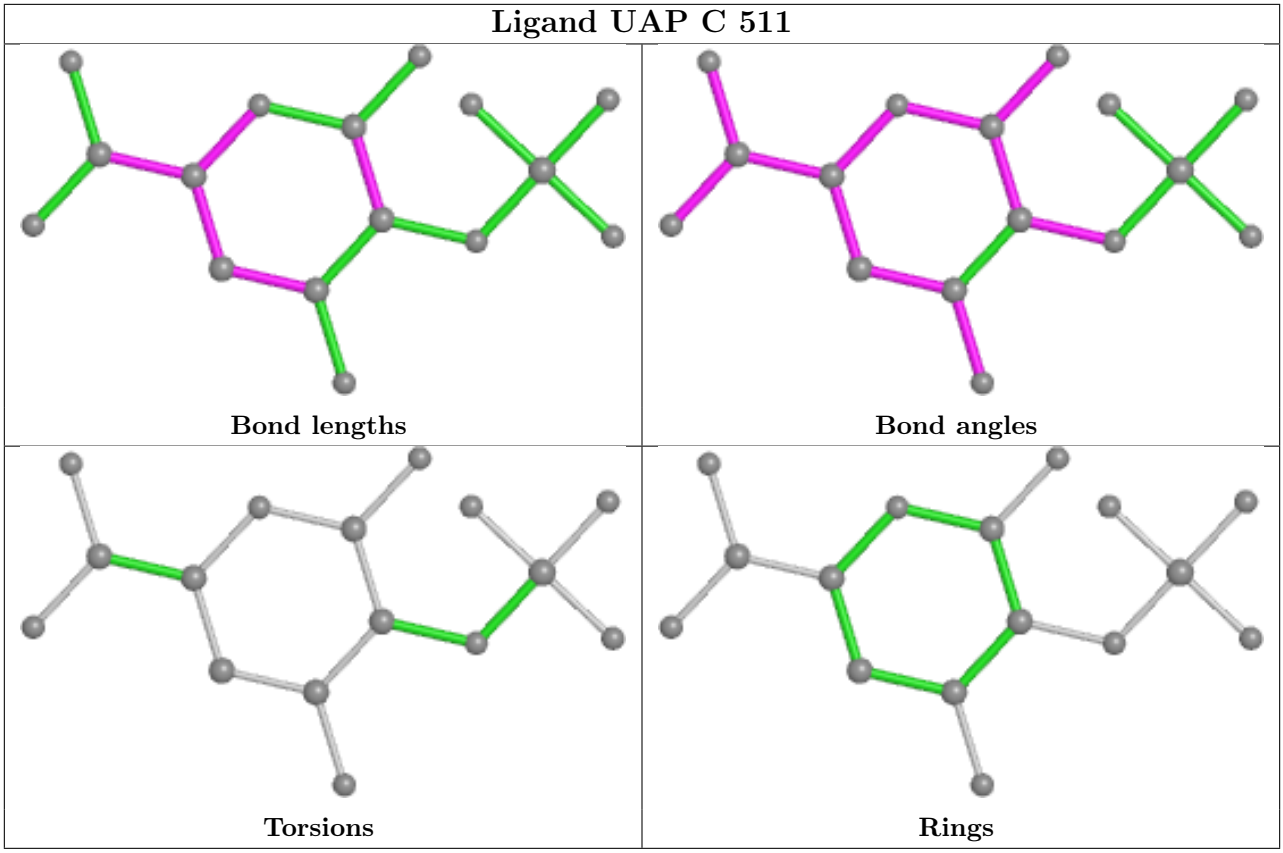
Continued from previous page...

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	508	EDO	0	1
4	D	508	EDO	0	1

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 UAP D 510





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	467/489 (95%)	-0.48	2 (0%)	92	93	9, 16, 34, 51	0
1	B	465/489 (95%)	-0.43	1 (0%)	95	95	10, 18, 34, 50	0
1	C	465/489 (95%)	-0.31	3 (0%)	89	90	9, 19, 38, 63	0
1	D	464/489 (94%)	-0.33	4 (0%)	84	85	10, 19, 36, 57	0
All	All	1861/1956 (95%)	-0.39	10 (0%)	91	92	9, 18, 36, 63	0

All (10) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	174	PRO	8.4
1	D	174	PRO	7.2
1	A	174	PRO	3.7
1	C	369	GLY	3.2
1	C	175	LYS	2.7
1	B	469	ASN	2.7
1	A	97	PRO	2.3
1	D	485	ILE	2.1
1	D	432	PRO	2.1
1	D	172	ASN	2.0

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)

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,

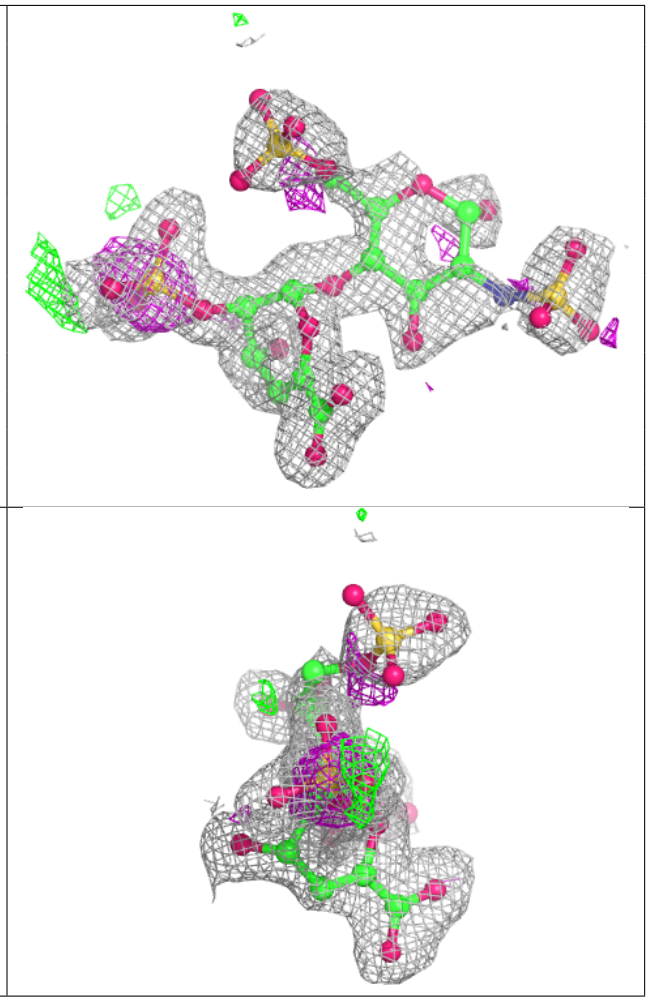
median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	SGN	F	1	19/20	0.76	0.24	57,70,87,90	0
2	SGN	E	1	19/20	0.80	0.24	48,60,80,82	0
2	UAP	F	2	16/16	0.91	0.12	25,34,37,40	0
2	UAP	E	2	16/16	0.94	0.11	18,26,29,31	0

The following is a graphical depiction of the model fit to experimental electron density for oligosaccharide. Each fit is shown from different orientation to approximate a three-dimensional view.

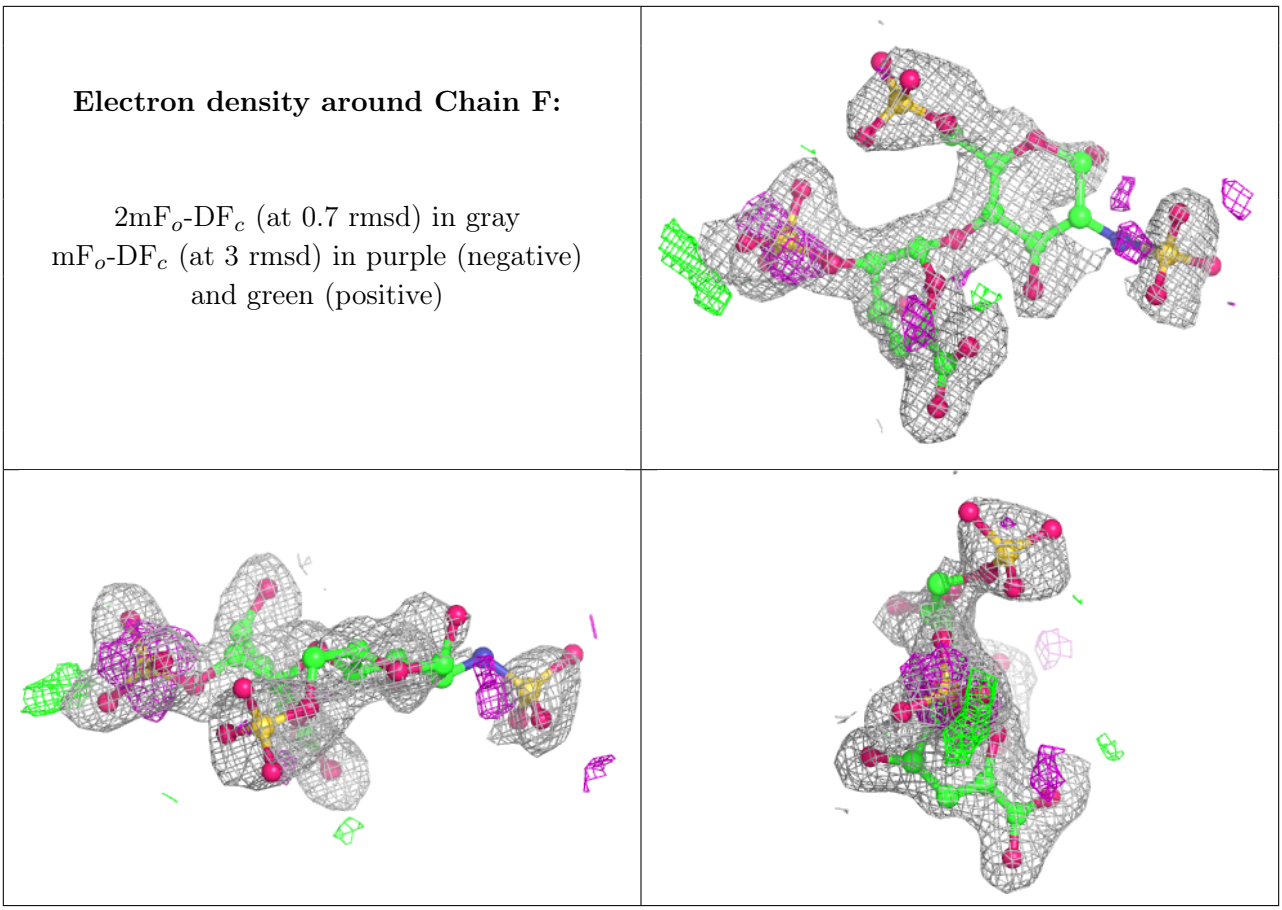
Electron density around Chain E:

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 Chain F:

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



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	D	508	4/4	0.14	1.46	128,134,134,137	0
4	EDO	B	507	4/4	0.55	0.99	188,189,195,199	0
4	EDO	C	510	4/4	0.74	0.22	48,49,50,52	0
4	EDO	C	509	4/4	0.77	0.18	49,50,51,52	0
4	EDO	D	507	4/4	0.81	0.15	50,51,53,54	0
4	EDO	B	506	4/4	0.83	0.26	37,39,41,42	0
4	EDO	B	509	4/4	0.84	0.24	41,41,42,42	0
4	EDO	C	505	4/4	0.85	0.19	33,34,36,37	0
4	EDO	C	508	4/4	0.86	0.15	33,35,35,36	0
4	EDO	D	503	4/4	0.86	0.16	26,26,27,29	0
4	EDO	A	503	4/4	0.87	0.19	22,30,32,39	0
4	EDO	B	502	4/4	0.88	0.16	33,34,34,38	0

Continued on next page...

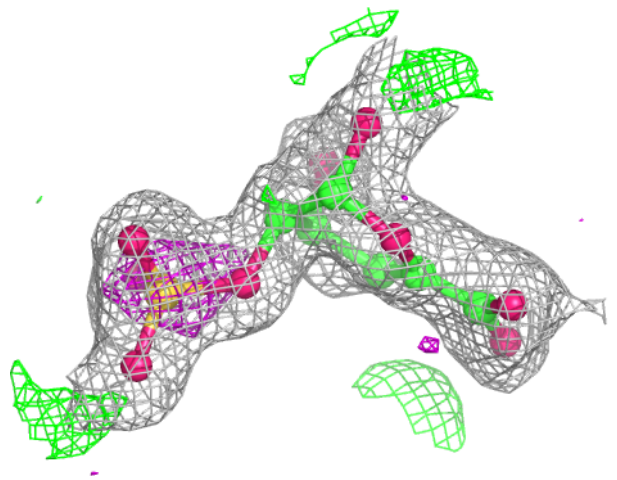
Continued from previous page...

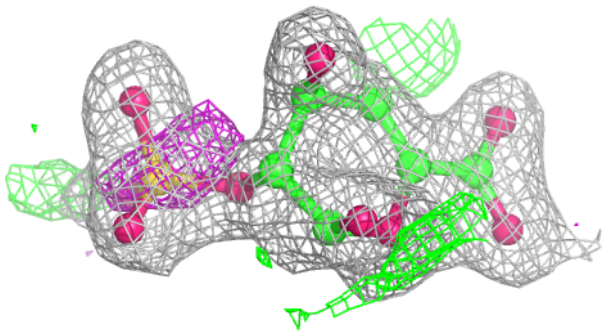
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	EDO	B	503	4/4	0.88	0.24	41,45,47,52	0
4	EDO	B	510	4/4	0.89	0.20	21,27,29,33	0
4	EDO	A	509	4/4	0.89	0.22	36,38,39,43	0
4	EDO	B	508	4/4	0.90	0.15	48,48,49,50	0
4	EDO	D	504	4/4	0.90	0.12	23,23,24,25	0
6	UAP	D	510	16/16	0.90	0.11	25,32,34,36	0
4	EDO	A	506	4/4	0.91	0.17	32,37,38,41	0
6	UAP	C	511	16/16	0.92	0.11	22,29,32,37	0
4	EDO	A	505	4/4	0.92	0.11	20,23,23,25	0
4	EDO	B	504	4/4	0.93	0.09	22,23,24,25	0
4	EDO	D	505	4/4	0.93	0.12	29,32,32,34	0
4	EDO	C	506	4/4	0.94	0.11	24,28,28,29	0
4	EDO	C	507	4/4	0.94	0.15	23,24,25,28	0
4	EDO	A	507	4/4	0.94	0.12	32,33,33,35	0
4	EDO	C	502	4/4	0.95	0.08	27,29,30,35	0
4	EDO	D	506	4/4	0.95	0.07	29,30,31,35	0
4	EDO	C	503	4/4	0.95	0.11	20,22,22,23	0
5	CL	D	509	1/1	0.96	0.07	22,22,22,22	0
4	EDO	A	508	4/4	0.96	0.15	24,24,25,26	0
4	EDO	D	502	4/4	0.96	0.08	29,31,32,32	0
4	EDO	A	502	4/4	0.97	0.12	22,24,25,25	0
4	EDO	C	504	4/4	0.97	0.12	19,20,22,23	0
4	EDO	B	505	4/4	0.97	0.08	16,16,17,17	0
5	CL	C	512	1/1	0.98	0.07	19,19,19,19	0
3	ZN	D	501	1/1	0.98	0.07	18,18,18,18	0
4	EDO	A	504	4/4	0.98	0.08	15,17,18,18	0
5	CL	B	513	1/1	0.98	0.07	22,22,22,22	0
5	CL	A	511	1/1	0.99	0.06	19,19,19,19	0
3	ZN	A	501	1/1	0.99	0.04	20,20,20,20	0
3	ZN	B	501	1/1	0.99	0.04	20,20,20,20	0
3	ZN	C	501	1/1	1.00	0.07	16,16,16,16	0

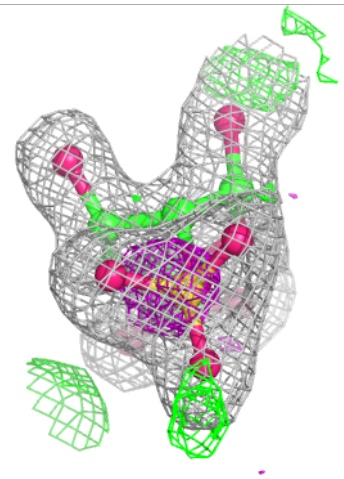
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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

Electron density around UAP D 510:

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

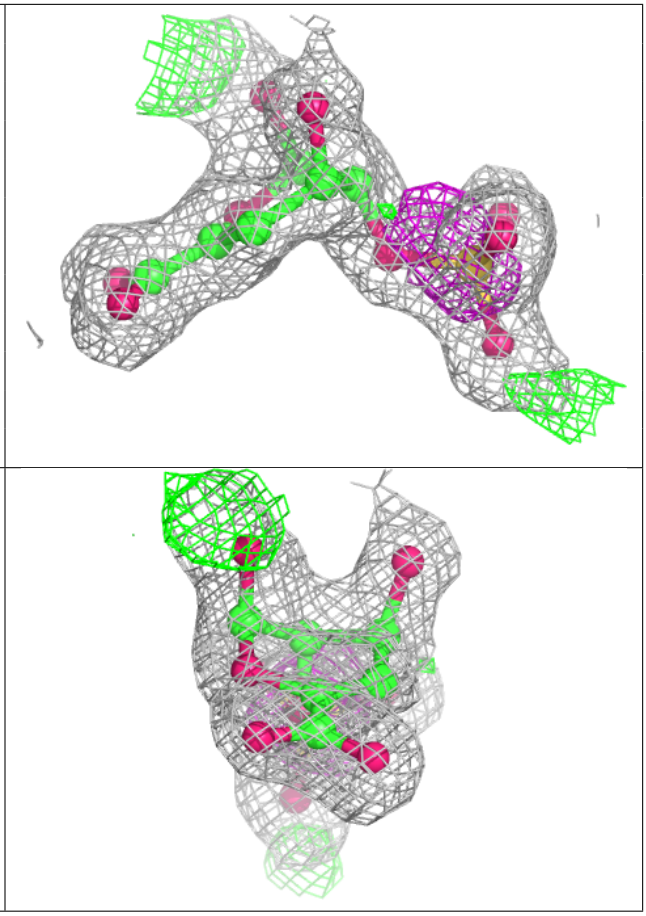






Electron density around UAP C 511:

$2mF_o$-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.