W O R L D W I D E

wwPDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Oct 12, 2024 – 10:18 AM EDT

PDB ID	:	5T03
Title	:	Crystal structure of heparan sulfate 6-O-sulfotransferase with bound PAP and glucuronic acid containing hexasaccharide substrate
Authors	:	Pedersen, L.C.; Moon, A.F.; Krahn, J.M.; Liu, J.
Deposited on	:	2016-08-15
Resolution	:	2.10 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

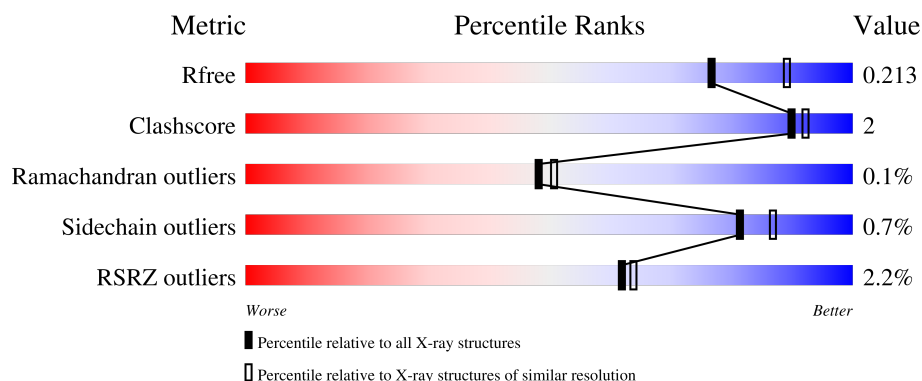|		
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	3.0
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

1   Overall quality at a glance   ⓘ

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

The reported resolution of this entry is 2.10 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	Percentile Ranks	Value
Rfree		0.213
Clashscore		2
Ramachandran outliers		0.1%
Sidechain outliers		0.7%
RSRZ outliers		2.2%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	6234 (2.10-2.10)
Clashscore	180529	6893 (2.10-2.10)
Ramachandran outliers	177936	6839 (2.10-2.10)
Sidechain outliers	177891	6840 (2.10-2.10)
RSRZ outliers	164620	6234 (2.10-2.10)

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	692	% — 93% · ·
1	B	692	4% — 92% 5% ·
2	C	6	33% 50% 17%
2	E	6	17% 67% 17%
3	D	4	25% 75%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	4	50% 50%

2 Entry composition (i)

There are 10 unique types of molecules in this entry. The entry contains 11901 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 maltose binding protein - heparan sulfate 6-O-sulfotransferase isoform 3 fusion protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	670	Total	C	N	O	S	0	2	0
			5243	3338	891	988	26			
1	B	675	Total	C	N	O	S	0	6	0
			5285	3369	897	992	27			

There are 26 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	MET	-	initiating methionine	UNP P0AEY0
A	82	ALA	ASP	engineered mutation	UNP P0AEY0
A	83	ALA	LYS	engineered mutation	UNP P0AEY0
A	172	ALA	GLU	engineered mutation	UNP P0AEY0
A	173	ALA	ASN	engineered mutation	UNP P0AEY0
A	239	ALA	LYS	engineered mutation	UNP P0AEY0
A	359	ALA	GLU	engineered mutation	UNP P0AEY0
A	362	ALA	LYS	engineered mutation	UNP P0AEY0
A	363	ALA	ASP	engineered mutation	UNP P0AEY0
A	367	ASN	-	linker	UNP P0AEY0
A	368	ALA	-	linker	UNP P0AEY0
A	369	ALA	-	linker	UNP P0AEY0
A	370	ALA	-	linker	UNP P0AEY0
B	0	MET	-	initiating methionine	UNP P0AEY0
B	82	ALA	ASP	engineered mutation	UNP P0AEY0
B	83	ALA	LYS	engineered mutation	UNP P0AEY0
B	172	ALA	GLU	engineered mutation	UNP P0AEY0
B	173	ALA	ASN	engineered mutation	UNP P0AEY0
B	239	ALA	LYS	engineered mutation	UNP P0AEY0
B	359	ALA	GLU	engineered mutation	UNP P0AEY0
B	362	ALA	LYS	engineered mutation	UNP P0AEY0
B	363	ALA	ASP	engineered mutation	UNP P0AEY0
B	367	ASN	-	linker	UNP P0AEY0
B	368	ALA	-	linker	UNP P0AEY0

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	369	ALA	-	linker	UNP P0AEY0
B	370	ALA	-	linker	UNP P0AEY0

- Molecule 2 is an oligosaccharide called 2-deoxy-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)
  -beta-D-glucopyranuronic acid-(1-4)-2-deoxy-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-b
 eta-D-glucopyranuronic acid-(1-4)-2-deoxy-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-bet
  a-D-glucopyranuronic acid.
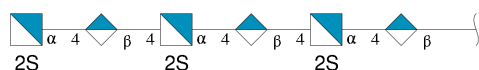
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	6	Total	C	N	O	S	0	0	0
			81	36	3	39	3			
2	E	6	Total	C	N	O	S	0	4	0
			96	42	4	46	4			

- Molecule 3 is an oligosaccharide called alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose-
  (1-4)-alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace
3	D	4	Total	C	O	0	0	0
			45	24	21			
3	F	4	Total	C	O	0	0	0
			45	24	21			

- Molecule 4 is ADENOSINE-3'-5'-DIPHOSPHATE (three-letter code: A3P) (formula: $C_{10}H_{15}N_5O_{10}P_2$).

A3P
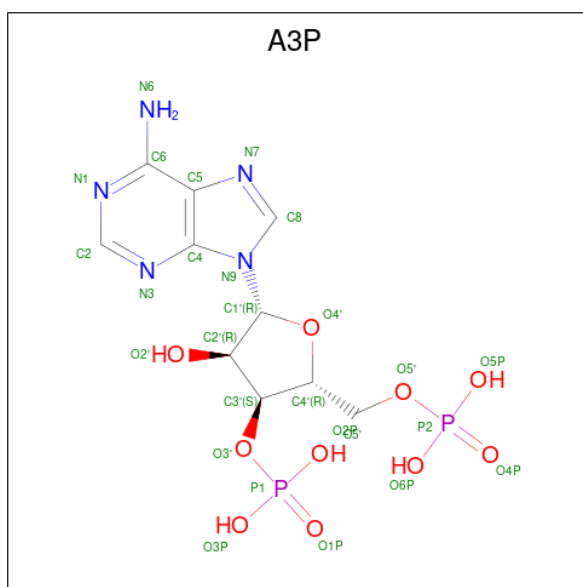
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	A	1	Total	C	N	O	P	0	0
			27	10	5	10	2		
4	B	1	Total	C	N	O	P	0	0
			27	10	5	10	2		

- Molecule 5 is P-NITROPHENOL (three-letter code: NPO) (formula: $C_6H_5NO_3$).

NPO
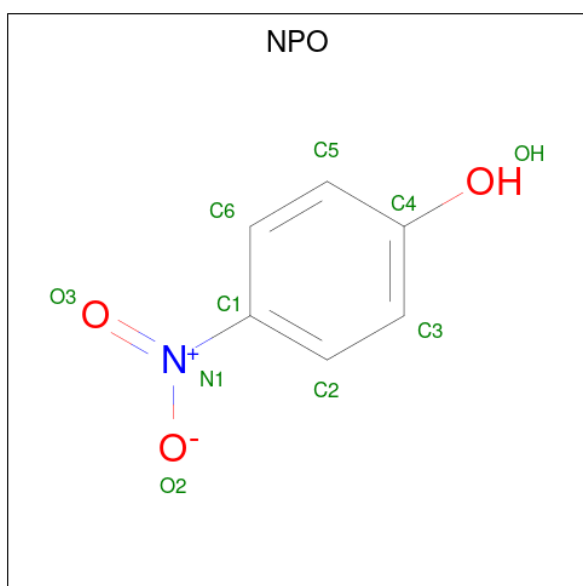
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	1	Total	C	N	O	0	0
			10	6	1	3		
5	B	1	Total	C	N	O	0	1
			10	6	1	3		

- Molecule 6 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).

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

Continued on next page...

Continued from previous page...

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

- Molecule 7 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	3	Total 3	Cl 3	0	0
7	B	2	Total 2	Cl 2	0	0

- Molecule 8 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	4	Total 4	Na 4	0	0
8	B	2	Total 2	Na 2	0	1

- Molecule 9 is L(+)-TARTARIC ACID (three-letter code: TLA) (formula: $C_4H_6O_6$).
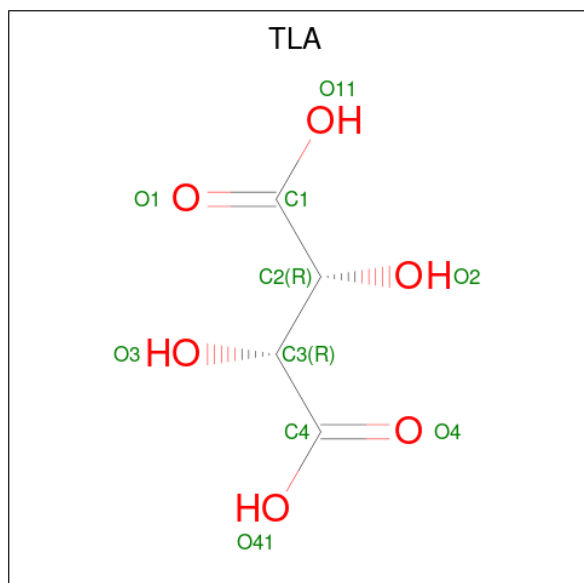
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
9	A	1	Total 10	C 4	O 6	0	0
9	A	1	Total 10	C 4	O 6	0	0
9	B	1	Total 10	C 4	O 6	0	0
9	B	1	Total 10	C 4	O 6	0	0

• Molecule 10 is water.

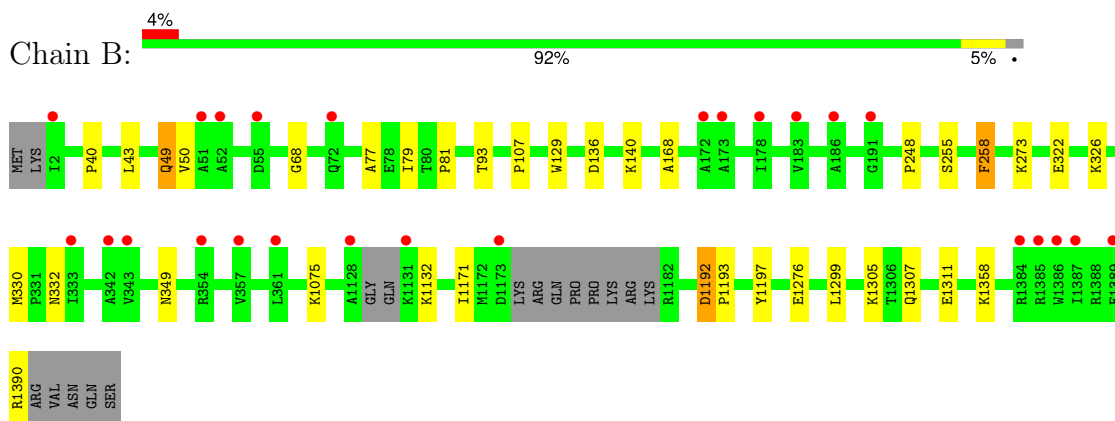| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
10	A	490	Total 491	O 491	0	1
10	B	425	Total 426	O 426	0	3

3   Residue-property plots ⓘ

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: maltose binding protein - heparan sulfate 6-O-sulfotransferase isoform 3 fusion protein

Chain A:

• Molecule 1: maltose binding protein - heparan sulfate 6-O-sulfotransferase isoform 3 fusion protein

Chain B:

• Molecule 2: 2-deoxy-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-beta-D-glucopyranuronic acid-(1-4)-2-deoxy-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-beta-D-glucopyranuronic acid-(1-4)-2-deoxy-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-beta-D-glucopyranuronic acid
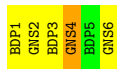

Chain C:

• Molecule 2: 2-deoxy-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-beta-D-glucopyranuronic acid-(1-4)-2-deoxy-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-beta-D-glucopyranuronic acid-(1-4)-2-

deoxy-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-beta-D-glucopyranuronic acid

Chain E:

• Molecule 3: alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose

Chain D:

• Molecule 3: alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose

Chain F:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	75.78Å 128.40Å 179.72Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	38.65 – 2.10 38.65 – 2.10	Depositor EDS
% Data completeness (in resolution range)	99.5 (38.65-2.10) 95.4 (38.65-2.10)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.14 (at 2.10Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.173 , 0.212 0.174 , 0.213	Depositor DCC
R_{free} test set	2274 reflections (2.24%)	wwPDB-VP
Wilson B-factor (Å^2)	29.6	Xtriage
Anisotropy	0.279	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 50.8	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	11901	wwPDB-VP
Average B, all atoms (Å^2)	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.34% 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 ⓘ

5.1   Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: GNS, A3P, CL, BDP, NA, EDO, GLC, TLA, NPO

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	0.32	0/5366	0.49	0/7286
1	B	0.33	0/5409	0.49	0/7349
All	All	0.32	0/10775	0.49	0/14635

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2   Too-close contacts ⓘ

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	5243	0	5021	14	0
1	B	5285	0	5020	23	0
2	C	81	0	40	1	0
2	E	96	0	47	2	0
3	D	45	0	39	0	0
3	F	45	0	39	0	0
4	A	27	0	11	0	0
4	B	27	0	11	0	0
5	A	10	0	4	0	0
5	B	10	0	4	0	0
6	A	44	0	66	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	B	20	0	30	0	0
7	A	3	0	0	0	0
7	B	2	0	0	1	0
8	A	4	0	0	0	0
8	B	2	0	0	0	0
9	A	20	0	8	0	0
9	B	20	0	7	0	0
10	A	491	0	0	3	1
10	B	426	0	0	4	1
All	All	11901	0	10347	37	1

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 37 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:1132:LYS:HB2	2:E:4[A]:GNS:O3S	2.01	0.61
1:A:219:LYS:NZ	10:A:1510:HOH:O	2.37	0.56
1:B:258:PHE:CD1	1:B:330:MET:HG2	2.41	0.56
1:B:349:ASN:OD1	10:B:1501:HOH:O	2.17	0.55
1:B:49[A]:GLN:HG2	1:B:50:VAL:HG13	1.89	0.54

All (1) 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)
10:A:1823:HOH:O	10:B:1876:HOH:O[2_454]	2.17	0.03

5.3　Torsion angles　ⓘ

5.3.1　Protein backbone　ⓘ

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	668/692 (96%)	653 (98%)	15 (2%)	0	100	100
1	B	675/692 (98%)	660 (98%)	14 (2%)	1 (0%)	48	51
All	All	1343/1384 (97%)	1313 (98%)	29 (2%)	1 (0%)	48	51

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	168	ALA

5.3.2 Protein sidechains ⓘ

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	535/582 (92%)	532 (99%)	3 (1%)	84	89
1	B	532/582 (91%)	527 (99%)	5 (1%)	75	82
All	All	1067/1164 (92%)	1059 (99%)	8 (1%)	81	87

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

Mol	Chain	Res	Type
1	B	1192	ASP
1	B	1171	ILE
1	B	49[B]	GLN
1	B	49[A]	GLN
1	B	258	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA ⓘ

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)

21 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	BDP	C	1	2,5	12,12,13	0.84	0	14,17,19	1.24	1 (7%)
2	GNS	C	2	2	15,15,16	0.74	0	17,22,24	1.50	2 (11%)
2	BDP	C	3	2	12,12,13	0.81	0	14,17,19	0.71	0
2	GNS	C	4	2	15,15,16	0.61	0	17,22,24	1.53	2 (11%)
2	BDP	C	5	8,2	12,12,13	0.83	0	14,17,19	0.71	0
2	GNS	C	6	2	15,15,16	0.62	0	17,22,24	1.72	3 (17%)
3	GLC	D	1	3	12,12,12	0.53	0	17,17,17	0.57	0
3	GLC	D	2	3	11,11,12	0.66	0	15,15,17	0.93	1 (6%)
3	GLC	D	3	3	11,11,12	0.62	0	15,15,17	0.97	1 (6%)
3	GLC	D	4	3	11,11,12	0.50	0	15,15,17	1.04	1 (6%)
2	BDP	E	1[A]	2,5	12,12,13	0.83	0	14,17,19	1.16	1 (7%)
2	GNS	E	2[A]	2	15,15,16	0.54	0	17,22,24	1.96	4 (23%)
2	BDP	E	3[A]	2	12,12,13	0.84	0	14,17,19	1.09	1 (7%)
2	GNS	E	4[A]	2	15,15,16	0.55	0	17,22,24	1.88	5 (29%)
2	GNS	E	4[B]	2	15,15,16	0.67	0	17,22,24	1.47	1 (5%)
2	BDP	E	5	2	12,12,13	0.86	0	14,17,19	0.89	0
2	GNS	E	6	2	15,15,16	0.65	0	17,22,24	1.71	2 (11%)
3	GLC	F	1	3	12,12,12	0.48	0	17,17,17	0.48	0
3	GLC	F	2	3	11,11,12	0.58	0	15,15,17	0.96	1 (6%)
3	GLC	F	3	3	11,11,12	0.59	0	15,15,17	0.79	0
3	GLC	F	4	3	11,11,12	0.49	0	15,15,17	1.36	1 (6%)

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	BDP	C	1	2,5	-	0/4/21/24	0/1/1/1
2	GNS	C	2	2	-	0/7/24/27	0/1/1/1
2	BDP	C	3	2	-	0/4/21/24	0/1/1/1
2	GNS	C	4	2	-	1/7/24/27	0/1/1/1
2	BDP	C	5	8,2	-	1/4/21/24	0/1/1/1
2	GNS	C	6	2	-	1/7/24/27	0/1/1/1
3	GLC	D	1	3	-	0/2/22/22	0/1/1/1
3	GLC	D	2	3	-	0/2/19/22	0/1/1/1
3	GLC	D	3	3	-	0/2/19/22	0/1/1/1
3	GLC	D	4	3	-	0/2/19/22	0/1/1/1
2	BDP	E	1[A]	2,5	-	0/4/21/24	0/1/1/1
2	GNS	E	2[A]	2	-	2/7/24/27	0/1/1/1
2	BDP	E	3[A]	2	-	0/4/21/24	0/1/1/1
2	GNS	E	4[A]	2	-	0/7/24/27	0/1/1/1
2	GNS	E	4[B]	2	-	0/7/24/27	0/1/1/1
2	BDP	E	5	2	-	0/4/21/24	0/1/1/1
2	GNS	E	6	2	-	0/7/24/27	0/1/1/1
3	GLC	F	1	3	-	1/2/22/22	0/1/1/1
3	GLC	F	2	3	-	0/2/19/22	0/1/1/1
3	GLC	F	3	3	-	0/2/19/22	0/1/1/1
3	GLC	F	4	3	-	0/2/19/22	0/1/1/1

There are no bond length outliers.

The worst 5 of 27 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	6	GNS	O2S-S1-O3S	-5.26	108.67	120.36
2	E	4[B]	GNS	O2S-S1-O3S	-5.12	108.97	120.36
2	E	6	GNS	O2S-S1-O3S	-5.06	109.11	120.36
2	E	2[A]	GNS	O2S-S1-O3S	-4.93	109.39	120.36
2	C	2	GNS	O2S-S1-O3S	-4.92	109.42	120.36

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	E	2[A]	GNS	O5-C5-C6-O6
2	E	2[A]	GNS	C4-C5-C6-O6
3	F	1	GLC	C4-C5-C6-O6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
2	C	5	BDP	O5-C5-C6-O6A
2	C	4	GNS	C2-N2-S1-O2S

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	C	2	GNS	1	0
2	E	4[A]	GNS	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 oligosaccharide.

Oligosaccharide Chain C

Bond lengths

Bond angles

Torsions

Rings

Oligosaccharide Chain E

Bond lengths

Bond angles

Torsions

Rings

Oligosaccharide Chain D

Bond lengths

Bond angles

Torsions

Rings

Oligosaccharide Chain F

Bond lengths

Bond angles

Torsions

Rings
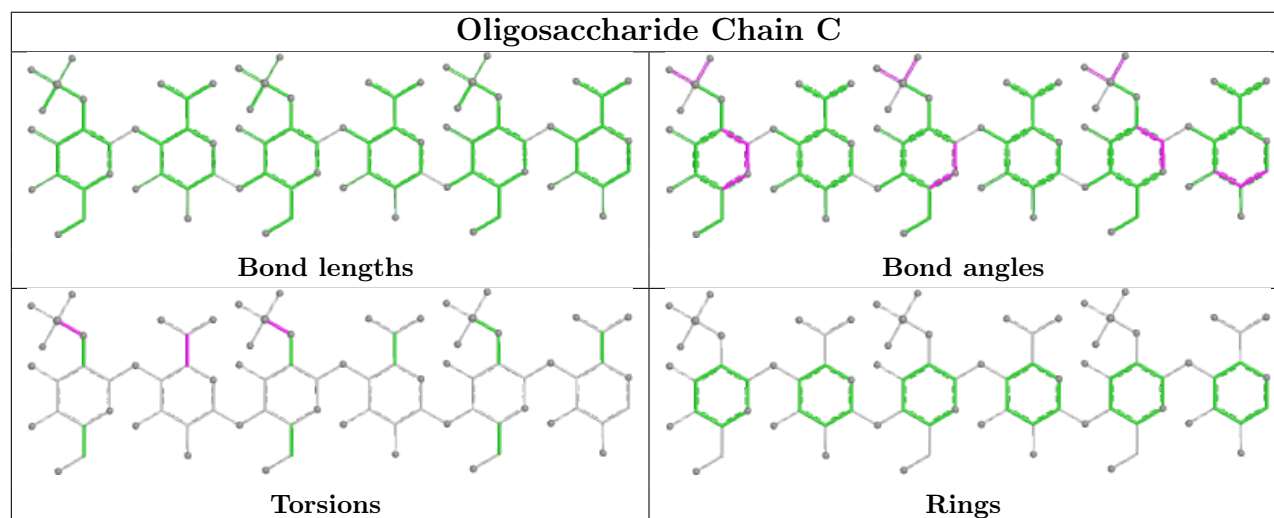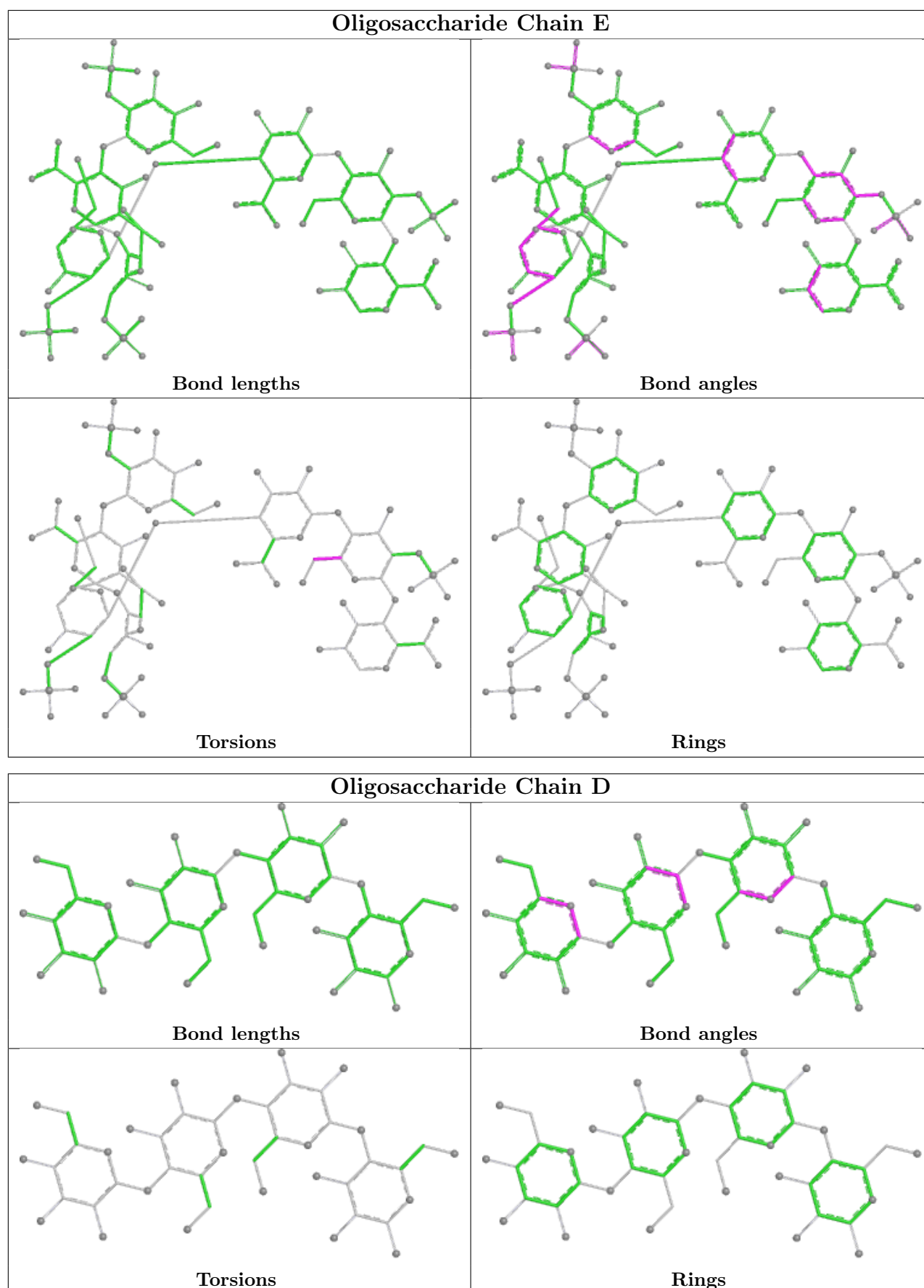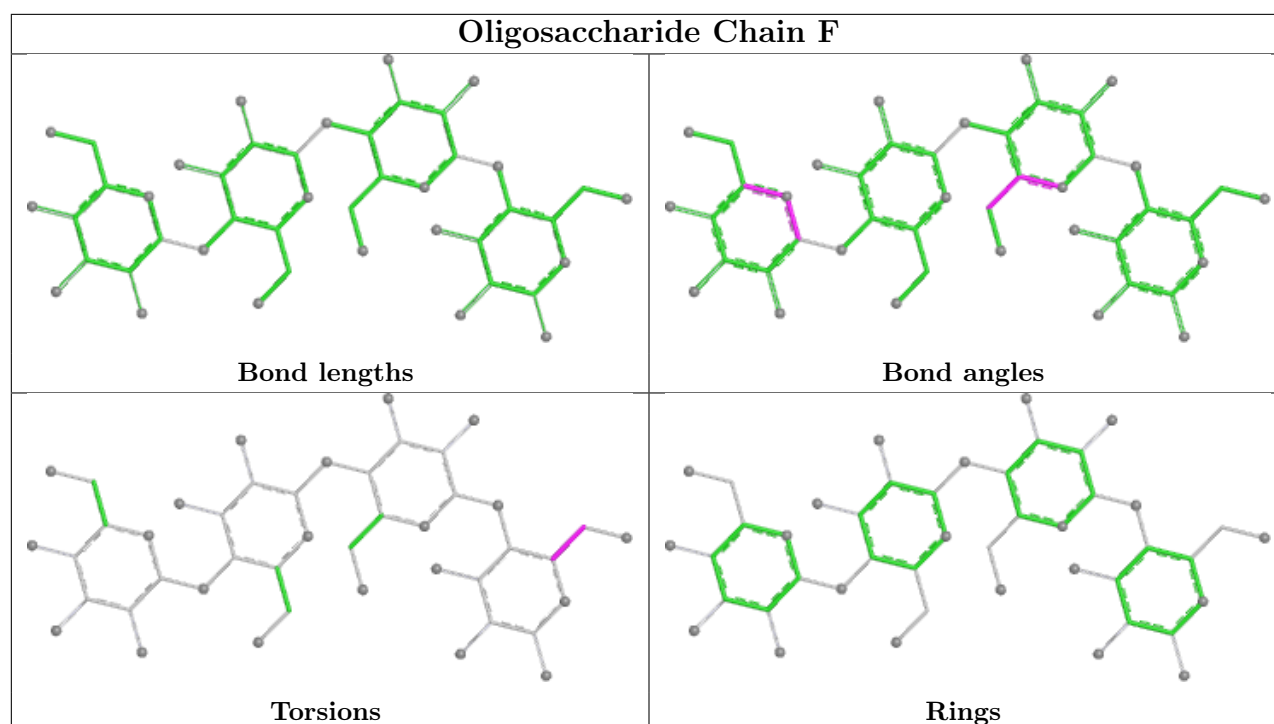
5.6 Ligand geometry ⓘ

Of 35 ligands modelled in this entry, 11 are monoatomic - leaving 24 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	A3P	B	1401	-	25,29,29	0.72	0	31,45,45	0.79	1 (3%)
5	NPO	A	1408	2	10,10,10	0.94	0	11,13,13	0.95	0
6	EDO	A	1421	-	3,3,3	0.44	0	2,2,2	0.58	0
6	EDO	A	1418	-	3,3,3	0.42	0	2,2,2	0.34	0
6	EDO	B	1414	-	3,3,3	0.44	0	2,2,2	0.61	0
6	EDO	A	1415	-	3,3,3	0.39	0	2,2,2	0.46	0
6	EDO	B	1413	-	3,3,3	0.57	0	2,2,2	0.36	0
6	EDO	A	1414	-	3,3,3	0.44	0	2,2,2	0.39	0
9	TLA	B	1422	8	9,9,9	1.00	0	12,12,12	1.02	0
6	EDO	B	1416	-	3,3,3	0.49	0	2,2,2	0.48	0
9	TLA	B	1423	-	9,9,9	1.03	0	12,12,12	1.46	3 (25%)
6	EDO	B	1415	-	3,3,3	0.39	0	2,2,2	0.55	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	#$\|Z\|$ > 2	Counts	RMSZ	#$\|Z\|$ > 2
6	EDO	A	1417	-	3,3,3	0.42	0	2,2,2	0.52	0
5	NPO	B	1408[A]	2	10,10,10	0.96	0	11,13,13	1.39	2 (18%)
6	EDO	A	1420	-	3,3,3	0.53	0	2,2,2	0.08	0
6	EDO	A	1422	-	3,3,3	0.46	0	2,2,2	0.48	0
6	EDO	A	1413	-	3,3,3	0.36	0	2,2,2	0.80	0
6	EDO	A	1419	-	3,3,3	0.46	0	2,2,2	0.42	0
6	EDO	A	1423	-	3,3,3	0.53	0	2,2,2	0.23	0
9	TLA	A	1431	8	9,9,9	0.99	0	12,12,12	0.99	0
6	EDO	B	1417	-	3,3,3	0.51	0	2,2,2	0.51	0
6	EDO	A	1416	-	3,3,3	0.49	0	2,2,2	0.43	0
9	TLA	A	1432	-	9,9,9	1.01	0	12,12,12	1.21	1 (8%)
4	A3P	A	1401	-	25,29,29	0.73	0	31,45,45	0.67	1 (3%)

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	A3P	B	1401	-	-	0/11/31/31	0/3/3/3
5	NPO	A	1408	2	-	0/2/4/4	0/1/1/1
6	EDO	A	1421	-	-	1/1/1/1	-
6	EDO	A	1418	-	-	0/1/1/1	-
6	EDO	B	1414	-	-	0/1/1/1	-
6	EDO	A	1415	-	-	1/1/1/1	-
6	EDO	B	1413	-	-	0/1/1/1	-
6	EDO	A	1414	-	-	0/1/1/1	-
9	TLA	B	1422	8	-	0/12/12/12	-
6	EDO	B	1416	-	-	0/1/1/1	-
9	TLA	B	1423	-	-	3/12/12/12	-
6	EDO	B	1415	-	-	0/1/1/1	-
6	EDO	A	1417	-	-	0/1/1/1	-
5	NPO	B	1408[A]	2	-	0/2/4/4	0/1/1/1
6	EDO	A	1420	-	-	0/1/1/1	-
6	EDO	A	1422	-	-	1/1/1/1	-
6	EDO	A	1413	-	-	1/1/1/1	-
6	EDO	A	1419	-	-	0/1/1/1	-
6	EDO	A	1423	-	-	1/1/1/1	-
9	TLA	A	1431	8	-	0/12/12/12	-
6	EDO	B	1417	-	-	1/1/1/1	-
6	EDO	A	1416	-	-	0/1/1/1	-
9	TLA	A	1432	-	-	0/12/12/12	-

Continued on next page...

WORLDWIDE PDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	A3P	A	1401	-	-	0/11/31/31	0/3/3/3

There are no bond length outliers.

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	B	1408[A]	NPO	C6-C1-N1	-2.69	116.98	119.34
9	B	1423	TLA	O11-C1-C2	2.57	120.45	113.31
5	B	1408[A]	NPO	C2-C1-N1	2.57	121.58	119.34
4	A	1401	A3P	C5-C6-N6	2.23	123.71	120.31
9	B	1423	TLA	O41-C4-C3	2.17	119.35	113.31

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
6	A	1423	EDO	O1-C1-C2-O2
9	B	1423	TLA	C2-C3-C4-O4
6	B	1417	EDO	O1-C1-C2-O2
9	B	1423	TLA	C2-C3-C4-O41
6	A	1413	EDO	O1-C1-C2-O2

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 A3P B 1401

Bond lengths

Bond angles

Torsions

Rings

Ligand A3P A 1401

Bond lengths

Bond angles

Torsions

Rings
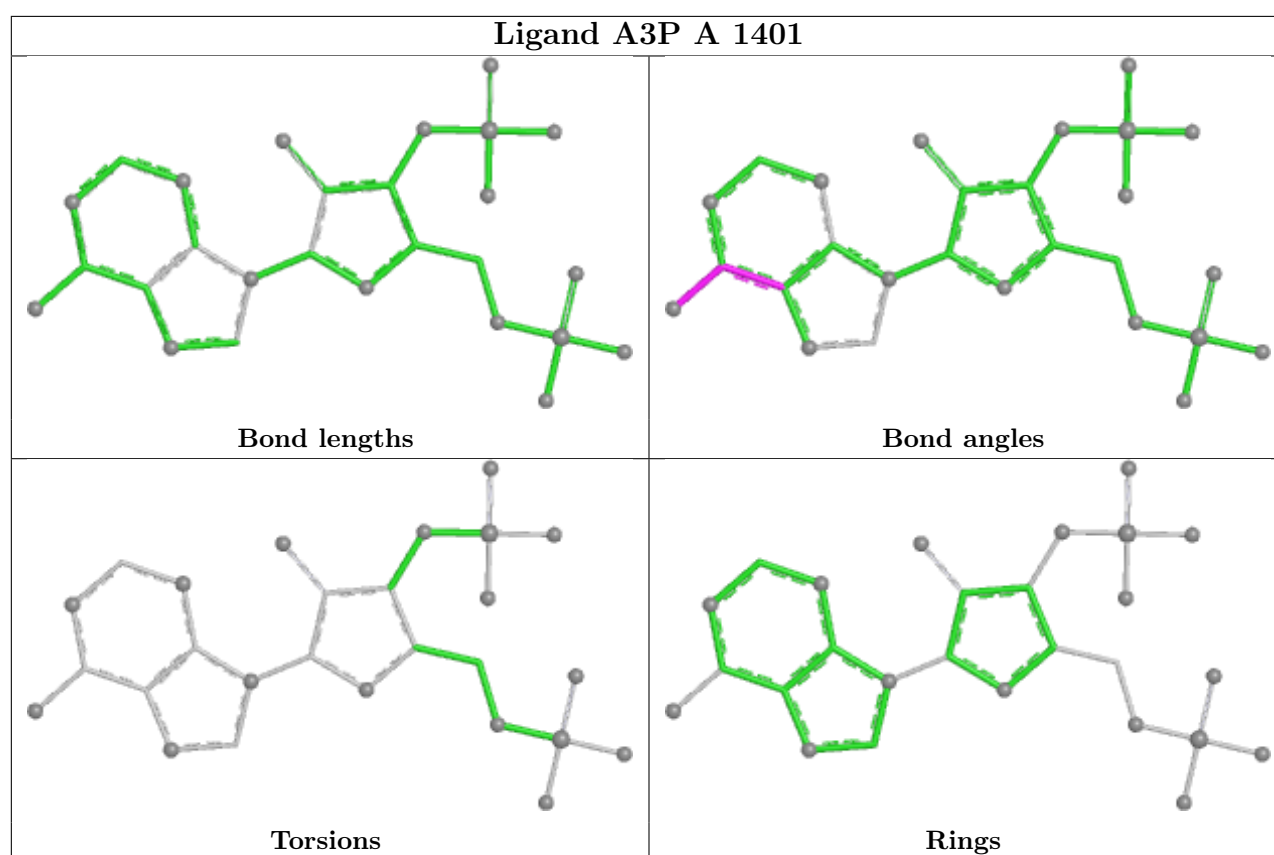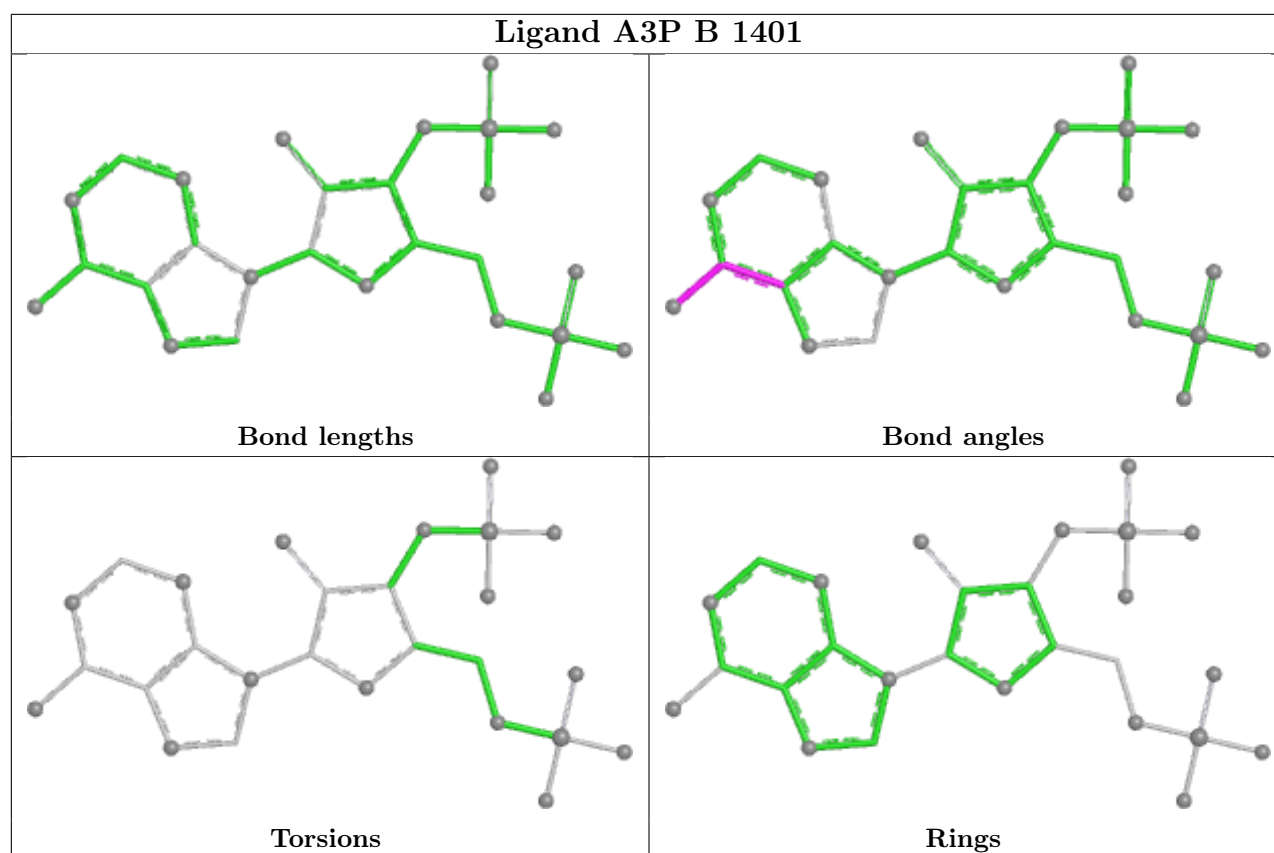
5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

6   Fit of model and data ⓘ

6.1   Protein, DNA and RNA chains ⓘ

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 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(Å²)	Q<0.9
1	A	670/692 (96%)	-0.44	5 (0%)	84	85	17, 31, 57, 98	2 (0%)
1	B	675/692 (97%)	-0.03	25 (3%)	45	48	11, 38, 70, 108	7 (1%)
All	All	1345/1384 (97%)	-0.24	30 (2%)	62	64	11, 33, 65, 108	9 (0%)

The worst 5 of 30 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	1387	ILE	5.1
1	B	1386	TRP	4.9
1	B	1128	ALA	4.4
1	B	55	ASP	3.7
1	A	92	PHE	3.6

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

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

6.3   Carbohydrates ⓘ

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.
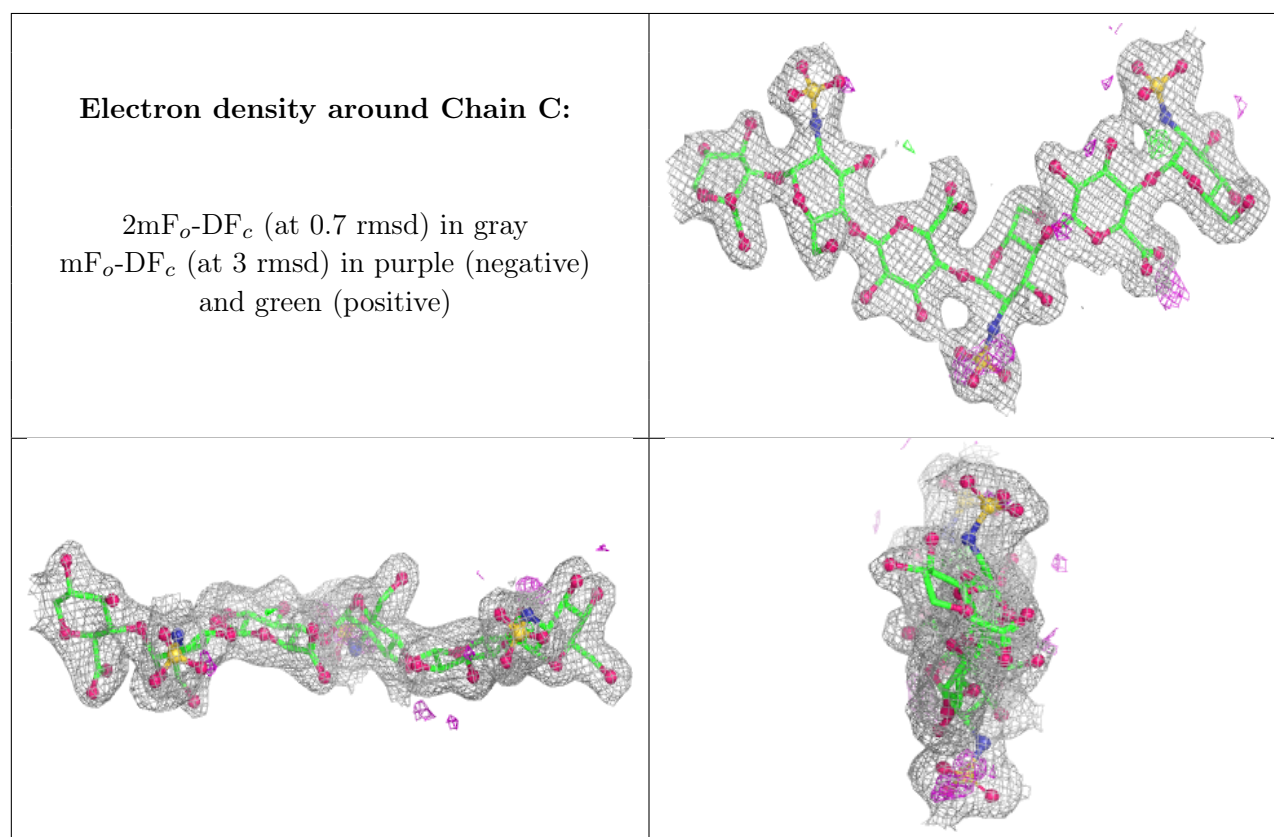
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å²)	Q<0.9
3	GLC	F	4	11/12	0.71	0.13	59,63,66,69	0
2	GNS	E	2[A]	15/16	0.72	0.14	55,59,66,69	15
2	BDP	E	1[A]	12/13	0.79	0.11	52,61,67,68	12
3	GLC	D	4	11/12	0.82	0.13	34,43,51,53	11
2	GNS	C	2	15/16	0.84	0.09	45,55,74,83	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors($Å^2$)	Q<0.9
2	BDP	C	3	12/13	0.85	0.09	38,44,53,59	0
3	GLC	F	2	11/12	0.86	0.10	35,36,41,44	0
2	BDP	C	1	12/13	0.88	0.07	45,50,55,56	0
3	GLC	F	1	12/12	0.88	0.10	25,32,36,38	0
2	BDP	E	3[A]	12/13	0.90	0.09	34,42,53,53	12
2	GNS	E	4[A]	15/16	0.91	0.11	27,34,44,49	15
3	GLC	F	3	11/12	0.91	0.08	35,38,45,47	0
2	GNS	E	4[B]	15/16	0.91	0.11	25,30,34,36	15
2	GNS	C	4	15/16	0.93	0.08	29,37,44,47	0
3	GLC	D	3	11/12	0.96	0.06	23,29,35,35	0
2	BDP	C	5	12/13	0.96	0.06	21,25,30,32	0
2	BDP	E	5	12/13	0.96	0.06	18,23,25,26	0
3	GLC	D	1	12/12	0.97	0.05	22,26,30,30	0
3	GLC	D	2	11/12	0.97	0.04	18,22,24,25	0
2	GNS	C	6	15/16	0.97	0.06	18,23,28,28	0
2	GNS	E	6	15/16	0.97	0.05	16,20,28,33	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 C:

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

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)
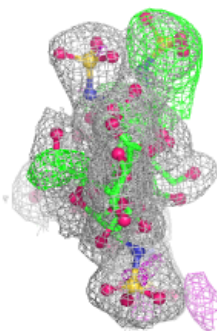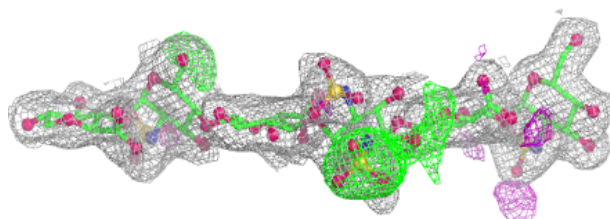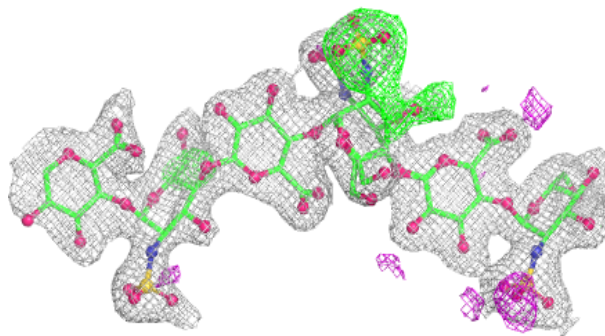
Electron density around Chain D:

$2mF_o$-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)
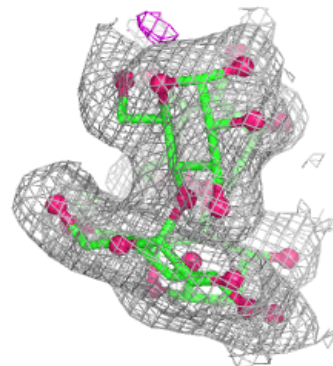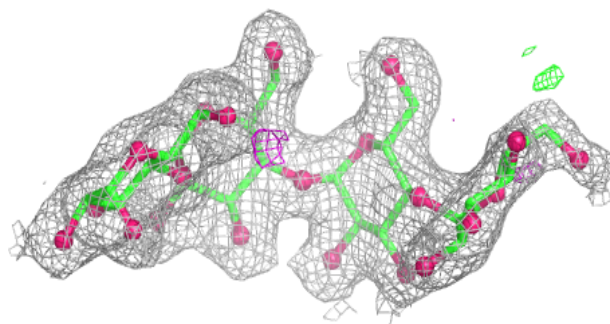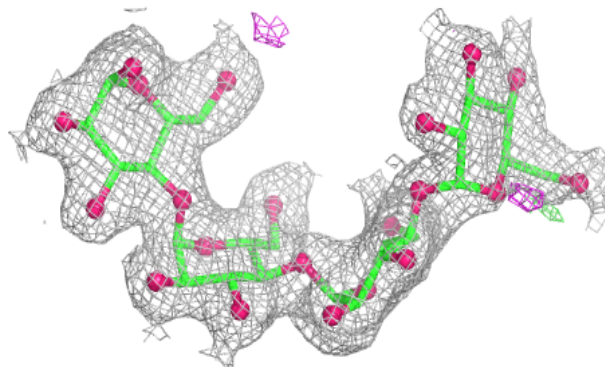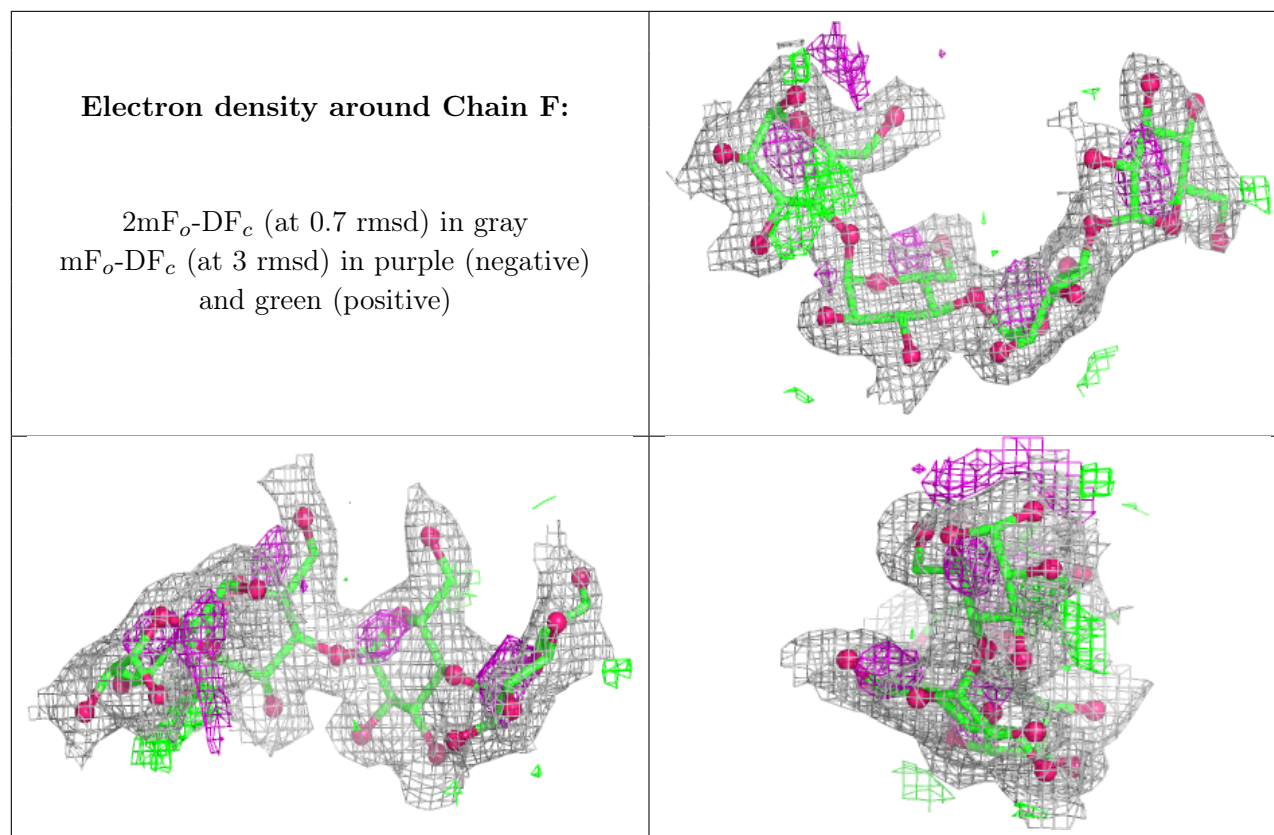
6.4 Ligands ⓘ

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
6	EDO	B	1416	4/4	0.75	0.22	44,45,46,48	0
8	NA	A	1427	1/1	0.76	0.13	54,54,54,54	0
9	TLA	A	1432	10/10	0.78	0.10	52,57,62,70	0
6	EDO	B	1417	4/4	0.80	0.15	37,41,49,56	0
8	NA	B	1421[A]	1/1	0.82	0.14	56,56,56,56	1
6	EDO	A	1421	4/4	0.83	0.15	33,38,43,57	0
6	EDO	A	1420	4/4	0.83	0.19	30,32,39,41	0
5	NPO	B	1408[A]	10/10	0.84	0.15	60,62,68,72	10
8	NA	A	1430	1/1	0.85	0.11	69,69,69,69	0
6	EDO	A	1418	4/4	0.87	0.13	36,38,43,46	0
6	EDO	B	1413	4/4	0.88	0.11	22,29,30,30	0
6	EDO	A	1423	4/4	0.89	0.12	31,33,42,46	0
6	EDO	A	1422	4/4	0.89	0.13	47,49,57,64	0
6	EDO	B	1414	4/4	0.89	0.13	38,43,49,50	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors($Å^2$)	Q<0.9
9	TLA	A	1431	10/10	0.90	0.07	33,40,44,50	0
6	EDO	A	1416	4/4	0.91	0.10	25,28,29,38	0
8	NA	A	1429	1/1	0.91	0.13	40,40,40,40	0
8	NA	A	1428	1/1	0.92	0.11	38,38,38,38	1
6	EDO	A	1414	4/4	0.92	0.11	37,38,39,39	0
6	EDO	A	1413	4/4	0.92	0.13	45,49,55,55	0
8	NA	B	1420	1/1	0.93	0.07	43,43,43,43	0
9	TLA	B	1422	10/10	0.93	0.07	38,43,49,51	0
9	TLA	B	1423	10/10	0.93	0.10	33,47,54,56	0
5	NPO	A	1408	10/10	0.94	0.07	40,42,53,60	0
6	EDO	A	1419	4/4	0.95	0.11	33,36,36,37	0
6	EDO	A	1415	4/4	0.96	0.07	21,33,36,47	0
6	EDO	A	1417	4/4	0.96	0.08	25,26,26,32	0
6	EDO	B	1415	4/4	0.96	0.07	26,26,31,31	0
7	CL	A	1425	1/1	0.97	0.12	45,45,45,45	0
7	CL	B	1418	1/1	0.97	0.06	47,47,47,47	0
4	A3P	B	1401	27/27	0.98	0.05	15,19,22,25	0
7	CL	B	1419	1/1	0.98	0.16	35,35,35,35	0
7	CL	A	1426	1/1	0.98	0.13	52,52,52,52	0
4	A3P	A	1401	27/27	0.99	0.04	16,20,23,25	0
7	CL	A	1424	1/1	1.00	0.05	28,28,28,28	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.
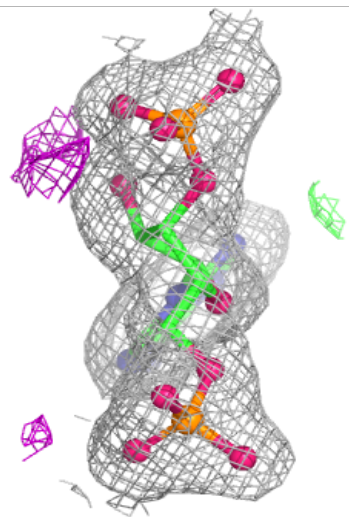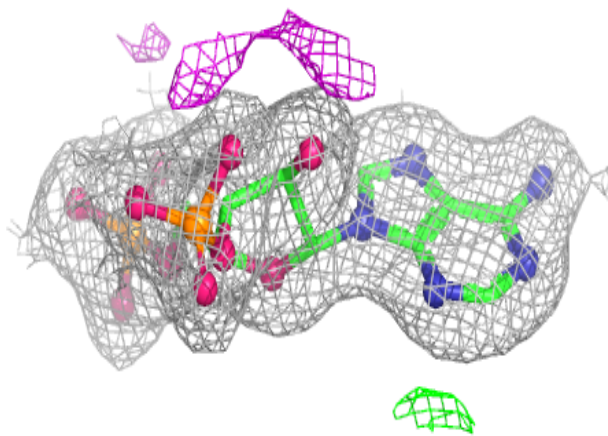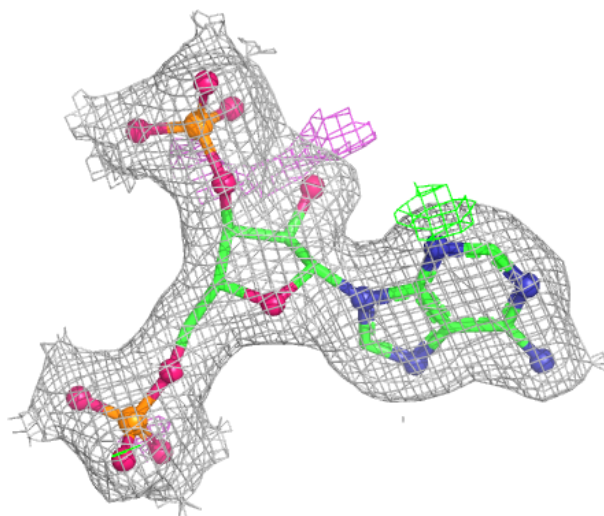
Electron density around A3P B 1401:

$2mF_o\text{-}DF_c$ (at 0.7 rmsd) in gray
$mF_o\text{-}DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)
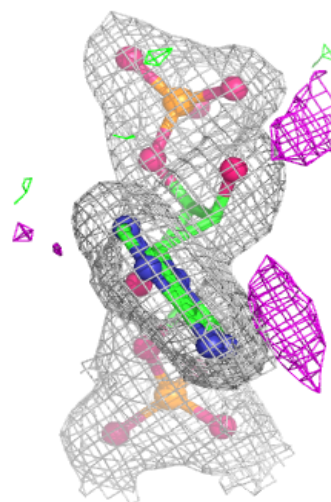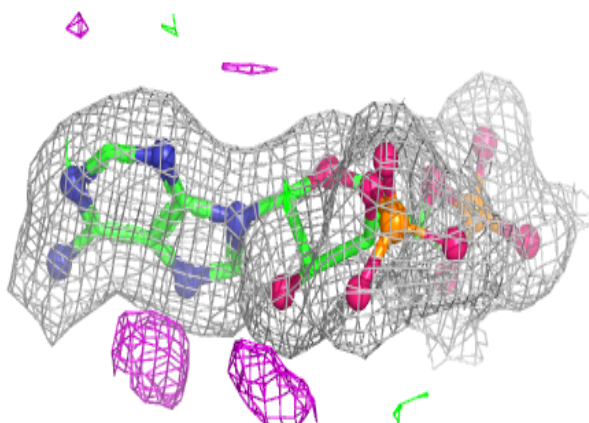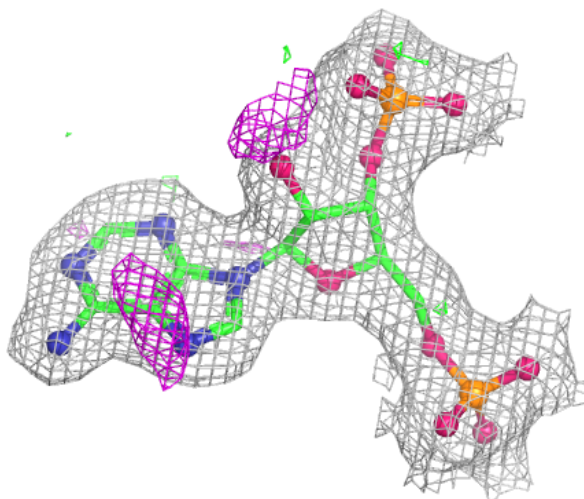
Electron density around A3P A 1401:

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

6.5 Other polymers (i)

There are no such residues in this entry.