wwPDB X-ray Structure Validation Summary Report (i)

Dec 13, 2022 – 12:09 pm GMT

PDB ID	:	7ZIN
Title	:	JC Polyomavirus VP1 in complex with 6'-Sialyllactose glycomacromolecules (aliphatic linker)
Authors	:	Freytag, J.; Mueller, J.C.; Stehle, T.
Deposited on	:	2022-04-08
Resolution	:	1.65 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:

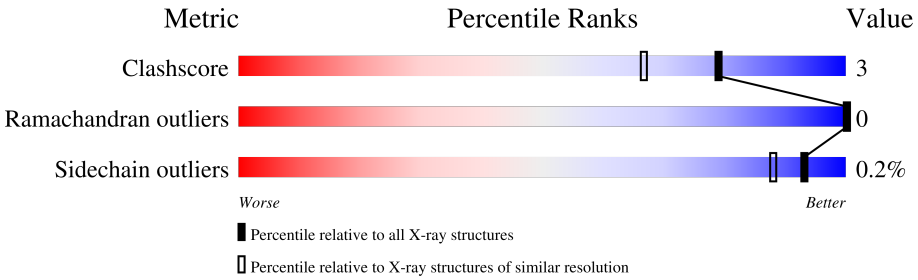
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	FAILED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.31.3

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 1.65 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	141614	3268 (1.66-1.62)
Ramachandran outliers	138981	3215 (1.66-1.62)
Sidechain outliers	138945	3215 (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%

Note EDS failed to run properly.

Mol	Chain	Length	Quality of chain
1	AAA	272	93%
1	BBB	272	90% 5%
1	CCC	272	91% 6%
1	DDD	272	90% 8%
1	EEE	272	91%
2	BaB	3	33% 67%
2	CaC	3	33% 67%
3	A	2	50% 50%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 12341 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 Major capsid protein VP1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAA	263	Total	C	N	O	S	0	10	0
			2097	1317	358	411	11			
1	BBB	260	Total	C	N	O	S	0	8	0
			2058	1301	352	394	11			
1	CCC	265	Total	C	N	O	S	0	14	0
			2136	1346	367	411	12			
1	DDD	265	Total	C	N	O	S	0	13	0
			2128	1336	363	416	13			
1	EEE	260	Total	C	N	O	S	0	9	0
			2081	1308	358	404	11			

There are 20 discrepancies between the modelled and reference sequences:

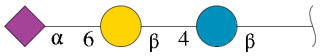
Chain	Residue	Modelled	Actual	Comment	Reference
AAA	18	GLY	-	expression tag	UNP P03089
AAA	19	SER	-	expression tag	UNP P03089
AAA	20	HIS	-	expression tag	UNP P03089
AAA	21	MET	-	expression tag	UNP P03089
BBB	18	GLY	-	expression tag	UNP P03089
BBB	19	SER	-	expression tag	UNP P03089
BBB	20	HIS	-	expression tag	UNP P03089
BBB	21	MET	-	expression tag	UNP P03089
CCC	18	GLY	-	expression tag	UNP P03089
CCC	19	SER	-	expression tag	UNP P03089
CCC	20	HIS	-	expression tag	UNP P03089
CCC	21	MET	-	expression tag	UNP P03089
DDD	18	GLY	-	expression tag	UNP P03089
DDD	19	SER	-	expression tag	UNP P03089
DDD	20	HIS	-	expression tag	UNP P03089
DDD	21	MET	-	expression tag	UNP P03089
EEE	18	GLY	-	expression tag	UNP P03089
EEE	19	SER	-	expression tag	UNP P03089
EEE	20	HIS	-	expression tag	UNP P03089

Continued on next page...

Continued from previous page...

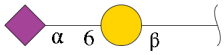
Chain	Residue	Modelled	Actual	Comment	Reference
EEE	21	MET	-	expression tag	UNP P03089

- Molecule 2 is an oligosaccharide called N-acetyl-alpha-neuraminic acid-(2-6)-beta-D-galacto pyranose-(1-4)-beta-D-glucopyranose.



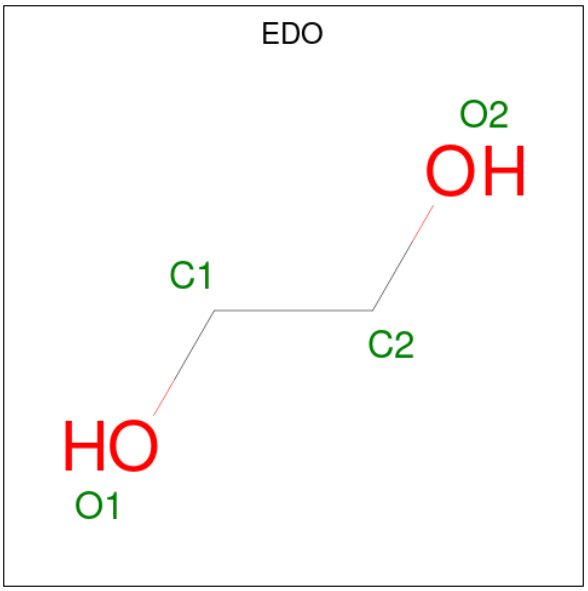
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	BaB	3	Total	C	N	O	0	0	0
			43	23	1	19			
2	CaC	3	Total	C	N	O	0	0	1
			32	17	1	14			

- Molecule 3 is an oligosaccharide called N-acetyl-alpha-neuraminic acid-(2-6)-beta-D-galacto pyranose.



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	A	2	Total	C	N	O	0	0	1
			21	11	1	9			

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



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	AAA	1	Total C O 4 2 2	0	0
4	BBB	1	Total C O 4 2 2	0	0
4	BBB	1	Total C O 4 2 2	0	0
4	CCC	1	Total C O 4 2 2	0	0
4	DDD	1	Total C O 4 2 2	0	0
4	EEE	1	Total C O 4 2 2	0	0
4	EEE	1	Total C O 4 2 2	0	0
4	EEE	1	Total C O 4 2 2	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	AAA	336	Total O 336 336	0	0
5	BBB	316	Total O 316 316	0	0
5	CCC	375	Total O 375 375	0	0
5	DDD	354	Total O 354 354	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	EEE	332	Total 332	O 332	0	0

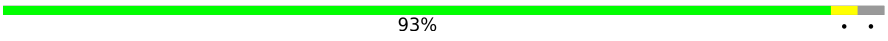
3 Residue-property plots (i)

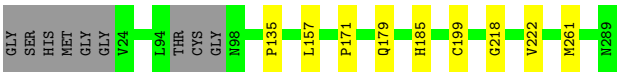
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

Note EDS failed to run properly.

• Molecule 1: Major capsid protein VP1

Chain AAA:





• Molecule 1: Major capsid protein VP1

Chain BBB:

• Molecule 1: Major capsid protein VP1

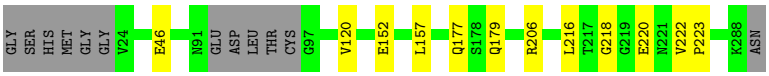
Chain CCC:

• Molecule 1: Major capsid protein VP1

Chain DDD:

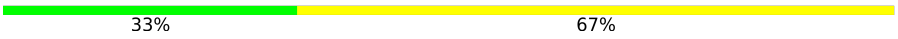
• Molecule 1: Major capsid protein VP1

Chain EEE:



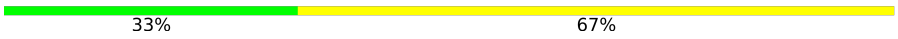
• Molecule 2: N-acetyl-alpha-neuraminic acid-(2-6)-beta-D-galactopyranose-(1-4)-beta-D-glucopyranose

Chain BaB:



• Molecule 2: N-acetyl-alpha-neuraminic acid-(2-6)-beta-D-galactopyranose-(1-4)-beta-D-glucopyranose

Chain CaC:



• Molecule 3: N-acetyl-alpha-neuraminic acid-(2-6)-beta-D-galactopyranose

Chain A:

4 Data and refinement statistics (i)

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

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	150.13Å 96.69Å 128.45Å 90.00° 110.62° 90.00°	Depositor
Resolution (Å)	44.85 – 1.65	Depositor
% Data completeness (in resolution range)	98.8 (44.85-1.65)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.13 (at 1.65Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.166 , 0.191	Depositor
Wilson B-factor (Å^2)	17.2	Xtriage
Anisotropy	0.010	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	12341	wwPDB-VP
Average B, all atoms (Å^2)	20.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.66% 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: EDO, SIA, GAL, BGC

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	AAA	0.66	0/2145	0.82	1/2920 (0.0%)
1	BBB	0.66	0/2112	0.82	0/2871
1	CCC	0.66	0/2199	0.83	0/2988
1	DDD	0.67	0/2185	0.82	0/2974
1	EEE	0.67	0/2129	0.81	0/2898
All	All	0.66	0/10770	0.82	1/14651 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	AAA	199	CYS	CB-CA-C	-5.06	100.29	110.40

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	AAA	2097	0	1991	7	0
1	BBB	2058	0	1984	12	0
1	CCC	2136	0	2050	19	0
1	DDD	2128	0	2041	20	0
1	EEE	2081	0	1995	8	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	BaB	43	0	37	0	0
2	CaC	32	0	26	0	0
3	A	21	0	17	4	0
4	AAA	4	0	6	0	0
4	BBB	8	0	12	0	0
4	CCC	4	0	6	0	0
4	DDD	4	0	6	0	0
4	EEE	12	0	18	0	0
5	AAA	336	0	0	0	0
5	BBB	316	0	0	5	0
5	CCC	375	0	0	6	0
5	DDD	354	0	0	3	0
5	EEE	332	0	0	2	0
All	All	12341	0	10189	58	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 3.

The worst 5 of 58 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:CCC:112:ILE:HG21	1:DDD:181[B]:MET:HE1	1.51	0.92
1:CCC:46:GLU:HG3	5:CCC:602:HOH:O	1.86	0.74
1:DDD:60[B]:SER:HG	3:A:2:SIA:HO8	1.35	0.68
1:DDD:264[B]:ASN:HD21	3:A:2:SIA:C1	2.07	0.67
1:CCC:127:THR:HG21	5:DDD:466:HOH:O	2.00	0.61

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	AAA	270/272 (99%)	263 (97%)	7 (3%)	0	100	100
1	BBB	262/272 (96%)	255 (97%)	7 (3%)	0	100	100
1	CCC	275/272 (101%)	266 (97%)	9 (3%)	0	100	100
1	DDD	274/272 (101%)	265 (97%)	9 (3%)	0	100	100
1	EEE	266/272 (98%)	259 (97%)	7 (3%)	0	100	100
All	All	1347/1360 (99%)	1308 (97%)	39 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	AAA	230/237 (97%)	230 (100%)	0	100	100
1	BBB	227/237 (96%)	227 (100%)	0	100	100
1	CCC	235/237 (99%)	235 (100%)	0	100	100
1	DDD	238/237 (100%)	237 (100%)	1 (0%)	91	84
1	EEE	229/237 (97%)	228 (100%)	1 (0%)	91	84
All	All	1159/1185 (98%)	1157 (100%)	2 (0%)	93	88

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	DDD	220	GLU
1	EEE	220	GLU

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

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)

Of 8 monosaccharides modelled in this entry, 6 were used for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
3	SIA	A	2	3	20,20,21	0.95	1 (5%)	24,28,31	1.48	3 (12%)
2	BGC	BaB	1	2	12,12,12	0.73	0	17,17,17	1.16	0
2	GAL	BaB	2	2	11,11,12	0.67	0	15,15,17	1.35	2 (13%)
2	SIA	BaB	3	2	20,20,21	1.03	2 (10%)	24,28,31	1.23	4 (16%)
2	GAL	CaC	2	2	11,11,12	0.62	0	15,15,17	1.38	2 (13%)
2	SIA	CaC	3	2	20,20,21	1.17	1 (5%)	24,28,31	1.08	1 (4%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	SIA	A	2	3	-	0/18/34/38	0/1/1/1
2	BGC	BaB	1	2	-	0/2/22/22	0/1/1/1
2	GAL	BaB	2	2	-	0/2/19/22	0/1/1/1
2	SIA	BaB	3	2	-	0/18/34/38	0/1/1/1
2	GAL	CaC	2	2	-	0/2/19/22	0/1/1/1
2	SIA	CaC	3	2	-	0/18/34/38	0/1/1/1

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	CaC	3	SIA	C2-C1	3.76	1.55	1.52
2	BaB	3	SIA	C3-C4	2.41	1.57	1.52

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	BaB	3	SIA	C3-C2	2.26	1.56	1.52
3	A	2	SIA	C4-C5	2.16	1.55	1.53

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	2	SIA	C6-C5-N5	-4.18	103.97	110.91
2	BaB	3	SIA	O6-C2-C1	3.61	114.79	107.70
2	CaC	2	GAL	C1-C2-C3	3.00	113.35	109.67
3	A	2	SIA	C4-C3-C2	2.71	114.67	109.81
2	CaC	2	GAL	O6-C6-C5	-2.50	102.73	111.29

There are no chirality outliers.

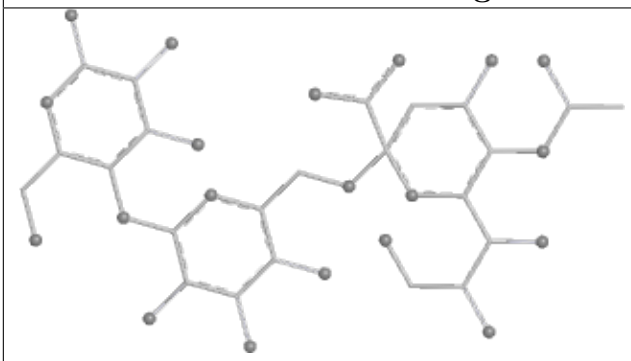
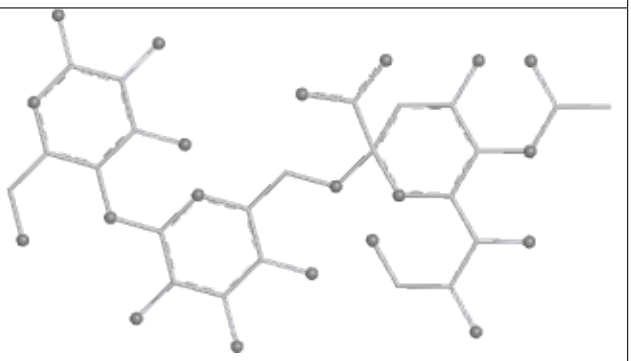
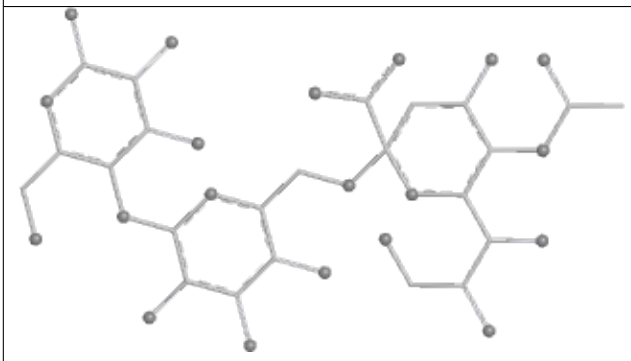
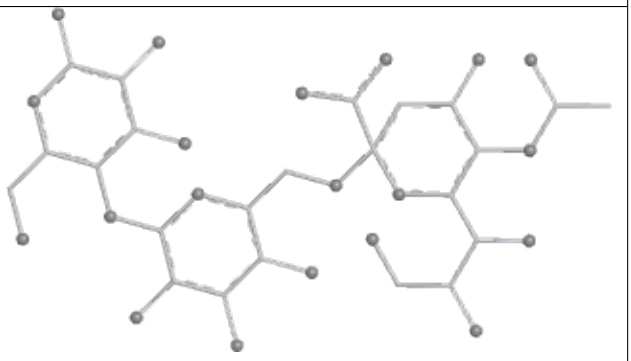
There are no torsion outliers.

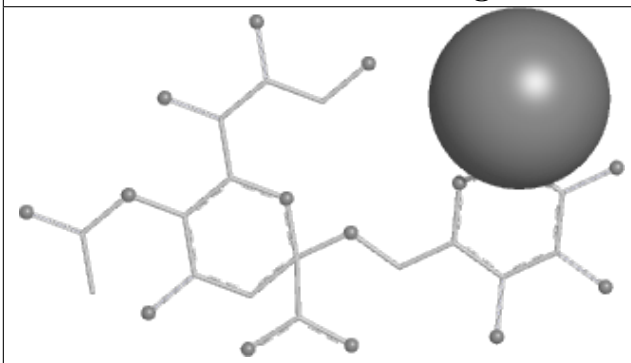
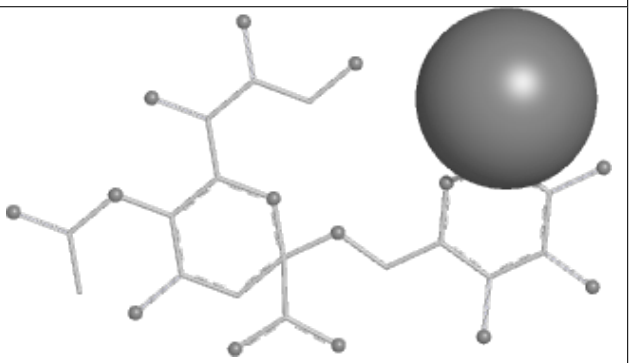
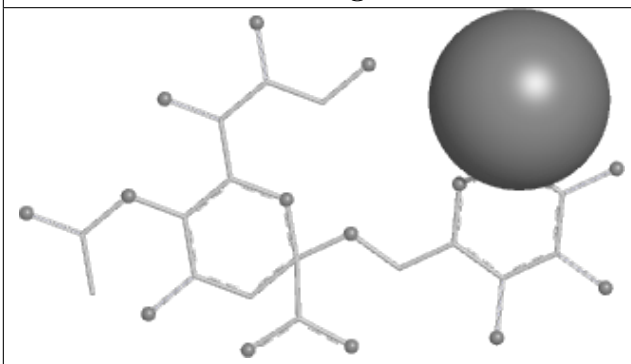
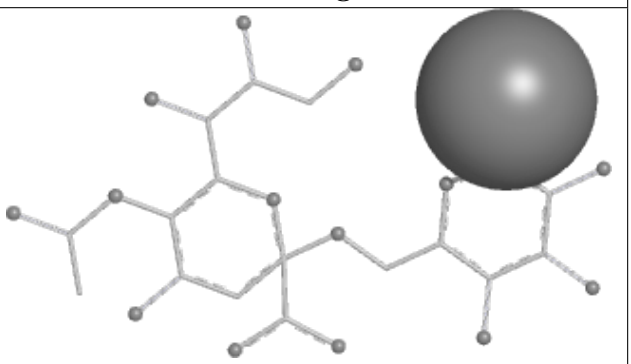
There are no ring outliers.

1 monomer is involved in 4 short contacts:

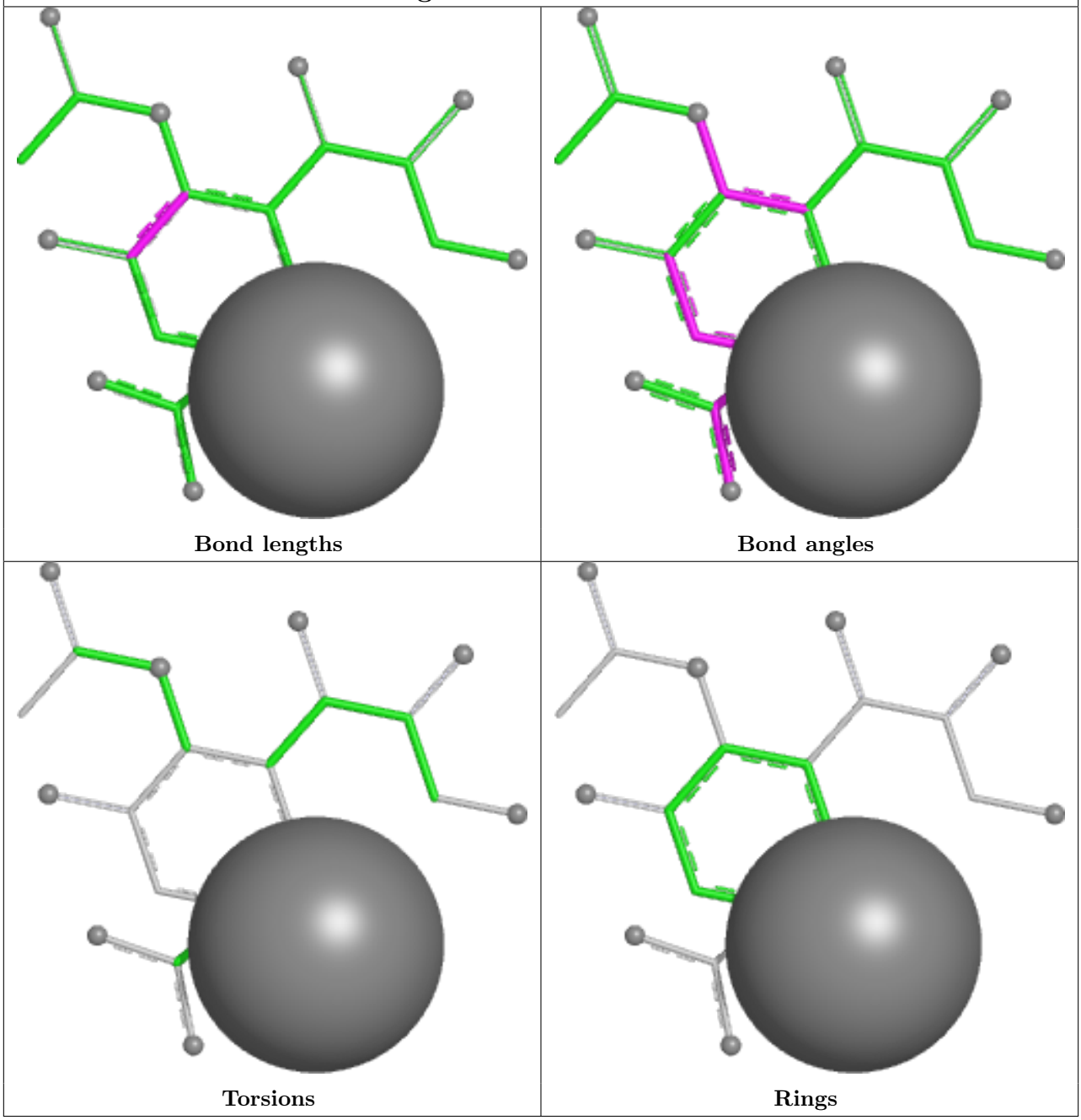
Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	2	SIA	4	0

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 BaB	
	
Bond lengths	**Bond angles**
	
Torsions	**Rings**

Oligosaccharide Chain CaC	
	
Bond lengths	**Bond angles**
	
Torsions	**Rings**

Oligosaccharide Chain A



Bond lengths | **Bond angles**

Torsions | **Rings**

5.6 Ligand geometry (i)

8 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection.

RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\|>2$	Counts	RMSZ	$\#\|Z\|>2$
4	EDO	BBB	501	-	3,3,3	0.12	0	2,2,2	0.28	0
4	EDO	DDD	301	-	3,3,3	0.21	0	2,2,2	0.27	0
4	EDO	EEE	502	-	3,3,3	0.08	0	2,2,2	0.12	0
4	EDO	AAA	501	-	3,3,3	0.16	0	2,2,2	0.11	0
4	EDO	BBB	502	-	3,3,3	0.11	0	2,2,2	0.28	0
4	EDO	EEE	503	-	3,3,3	0.17	0	2,2,2	0.11	0
4	EDO	EEE	501	-	3,3,3	0.09	0	2,2,2	0.15	0
4	EDO	CCC	501	-	3,3,3	0.16	0	2,2,2	0.21	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	BBB	501	-	-	1/1/1/1	-
4	EDO	DDD	301	-	-	1/1/1/1	-
4	EDO	EEE	502	-	-	0/1/1/1	-
4	EDO	AAA	501	-	-	1/1/1/1	-
4	EDO	BBB	502	-	-	1/1/1/1	-
4	EDO	EEE	503	-	-	1/1/1/1	-
4	EDO	EEE	501	-	-	1/1/1/1	-
4	EDO	CCC	501	-	-	1/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 7 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	BBB	502	EDO	O1-C1-C2-O2
4	CCC	501	EDO	O1-C1-C2-O2
4	DDD	301	EDO	O1-C1-C2-O2
4	EEE	503	EDO	O1-C1-C2-O2
4	AAA	501	EDO	O1-C1-C2-O2

There are no ring outliers.

No monomer is involved in short contacts.

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

6.1 Protein, DNA and RNA chains

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

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

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

6.3 Carbohydrates

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

6.4 Ligands

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

6.5 Other polymers

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