wwPDB X-ray Structure Validation Summary Report (i)

Dec 19, 2023 – 03:45 PM EST

PDB ID	:	1UC2
Title	:	Hypothetical Extein Protein of PH1602 from Pyrococcus horikoshii
Authors	:	Okada, C.; Maegawa, Y.; Yao, M.; Tanaka, I.
Deposited on	:	2003-04-08
Resolution	:	2.15 Å(reported)

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

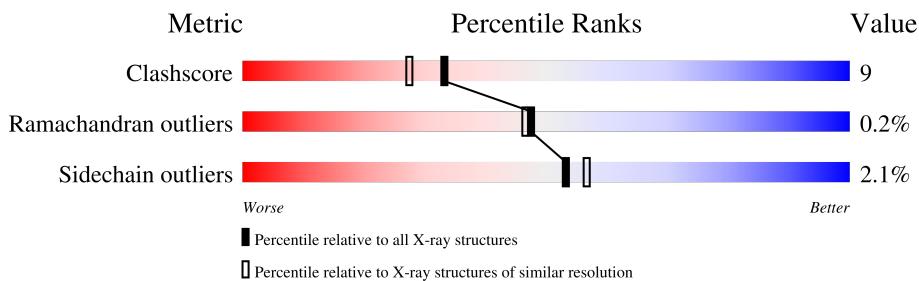|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.15 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
Clashscore		9
Ramachandran outliers		0.2%
Sidechain outliers		2.1%

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))
Clashscore	141614	1585 (2.16-2.16)
Ramachandran outliers	138981	1560 (2.16-2.16)
Sidechain outliers	138945	1559 (2.16-2.16)

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 was not executed.

| Mol	Chain	Length	Quality of chain
1	A	481	84% 15% •
1	B	481	82% 16% •
2	C	2	50% 50%
2	D	2	100%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 8480 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 hypothetical protein PH1602.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	480	Total 3759	C 2370	N 693	O 680	S 16	0	0	0
1	B	480	Total 3759	C 2370	N 693	O 680	S 16	0	0	0

- Molecule 2 is an oligosaccharide called beta-D-fructofuranose-(2-1)-alpha-D-glucopyranose.
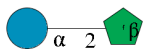
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace
2	C	2	Total 23	C 12	O 11	0	0	0
2	D	2	Total 23	C 12	O 11	0	0	0

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O_4S).

SO4

O1

O

O3 O⁻ — S — O⁻ O4

S

O

O2

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total 5	O 4	S 1	0	0
3	A	1	Total 5	O 4	S 1	0	0
3	A	1	Total 5	O 4	S 1	0	0
3	A	1	Total 5	O 4	S 1	0	0
3	A	1	Total 5	O 4	S 1	0	0
3	B	1	Total 5	O 4	S 1	0	0
3	B	1	Total 5	O 4	S 1	0	0
3	B	1	Total 5	O 4	S 1	0	0
3	B	1	Total 5	O 4	S 1	0	0

- Molecule 4 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	447	Total 447	O 447	0	0
4	B	424	Total 424	O 424	0	0

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. 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 was not executed.

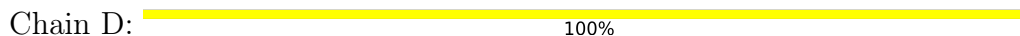• Molecule 1: hypothetical protein PH1602
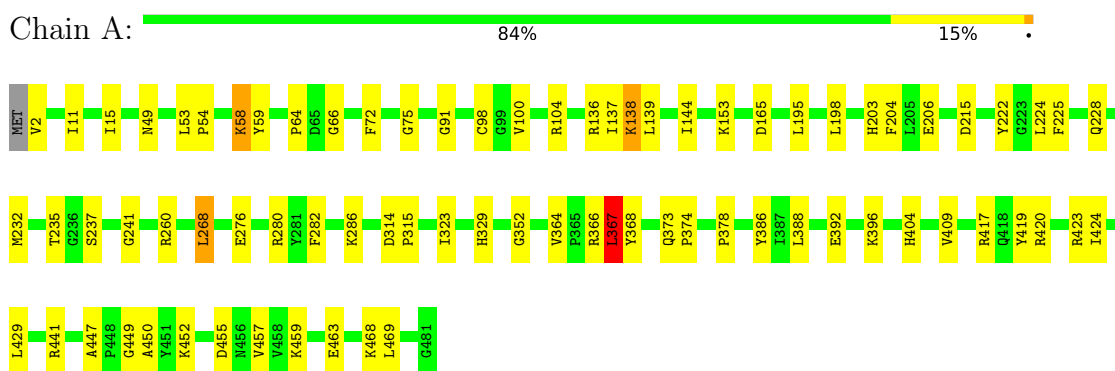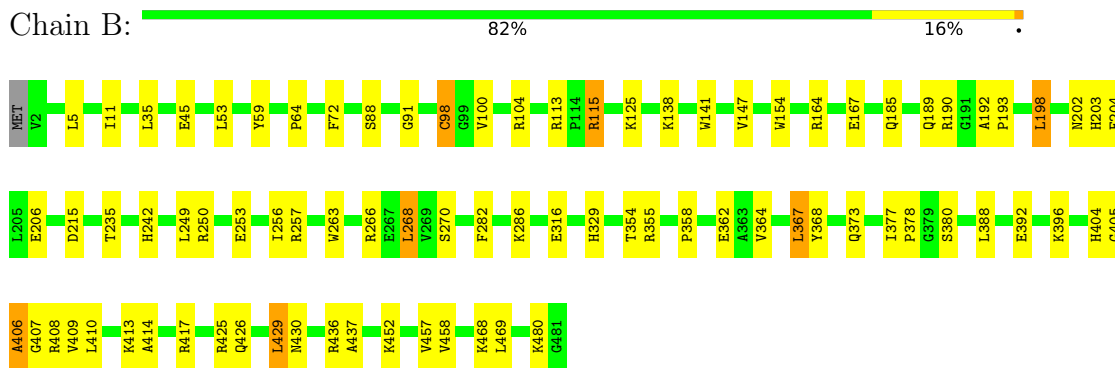
Chain A:

• Molecule 1: hypothetical protein PH1602

Chain B:

• Molecule 2: beta-D-fructofuranose-(2-1)-alpha-D-glucopyranose

Chain C:

• Molecule 2: beta-D-fructofuranose-(2-1)-alpha-D-glucopyranose

Chain D:

GLC1
FRU2

4 Data and refinement statistics ⓘ

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	81.25Å 137.59Å 148.50Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	10.00 – 2.15	Depositor
% Data completeness (in resolution range)	99.7 (10.00-2.15)	Depositor
R_{merge}	0.09	Depositor
R_{sym}	0.08	Depositor
Refinement program	CNS	Depositor
R, R_{free}	0.168 , 0.204	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	8480	wwPDB-VP
Average B, all atoms (Å^2)	28.0	wwPDB-VP

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: GLC, FRU, SO4

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	0.56	0/3836	0.75	3/5175 (0.1%)
1	B	0.59	0/3836	0.74	1/5175 (0.0%)
All	All	0.57	0/7672	0.74	4/10350 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	98	CYS	N-CA-C	-7.20	91.56	111.00
1	A	98	CYS	N-CA-C	-6.56	93.30	111.00
1	A	367	LEU	CA-CB-CG	5.13	127.09	115.30
1	A	15	ILE	N-CA-C	-5.02	97.45	111.00

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	3759	0	3783	62	0
1	B	3759	0	3783	74	0
2	C	23	0	21	0	0
2	D	23	0	21	0	0
3	A	25	0	0	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	20	0	0	0	0
4	A	447	0	0	6	0
4	B	424	0	0	9	0
All	All	8480	0	7608	131	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 9.

The worst 5 of 131 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:53:LEU:HD11	1:B:268:LEU:HD11	1.34	1.05
1:B:329:HIS:HD2	1:B:404:HIS:CD2	1.77	1.01
1:B:329:HIS:CD2	1:B:404:HIS:CD2	2.60	0.90
1:B:100:VAL:HG21	1:B:378:PRO:HG3	1.56	0.87
1:B:185:GLN:HG2	1:B:189:GLN:HE21	1.39	0.85

There are no symmetry-related clashes.

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	478/481 (99%)	464 (97%)	13 (3%)	1 (0%)	47	46
1	B	478/481 (99%)	463 (97%)	14 (3%)	1 (0%)	47	46
All	All	956/962 (99%)	927 (97%)	27 (3%)	2 (0%)	47	46

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	406	ALA

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	58	LYS

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	389/390 (100%)	383 (98%)	6 (2%)	65	69
1	B	389/390 (100%)	379 (97%)	10 (3%)	46	47
All	All	778/780 (100%)	762 (98%)	16 (2%)	53	57

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	367	LEU
1	B	362	GLU
1	B	104	ARG
1	B	358	PRO
1	B	59	TYR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	202	ASN
1	B	203	HIS
1	B	426	GLN
1	B	404	HIS
1	B	418	GLN

5.3.3   RNA ⓘ

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates ⓘ

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	GLC	C	1	2	11,11,12	2.08	4 (36%)	15,15,17	0.63	0
2	FRU	C	2	2	11,12,12	0.95	0	10,18,18	0.61	0
2	GLC	D	1	2	11,11,12	2.01	4 (36%)	15,15,17	0.69	0
2	FRU	D	2	2	11,12,12	1.15	2 (18%)	10,18,18	0.73	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
2	GLC	C	1	2	-	0/2/19/22	0/1/1/1
2	FRU	C	2	2	-	0/5/24/24	0/1/1/1
2	GLC	D	1	2	-	0/2/19/22	0/1/1/1
2	FRU	D	2	2	-	1/5/24/24	0/1/1/1

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	1	GLC	O5-C5	4.52	1.52	1.43
2	D	1	GLC	O5-C5	4.31	1.52	1.43
2	C	1	GLC	O5-C1	3.00	1.48	1.43
2	D	1	GLC	C2-C3	2.91	1.56	1.52
2	C	1	GLC	C2-C3	2.72	1.56	1.52

There are no bond angle outliers.

WORLDWIDE wwPDB PROTEIN DATA BANK

There are no chirality outliers.

All (1) torsion outliers are listed below:

| Mol	Chain	Res	Type	Atoms
2	D	2	FRU	O5-C5-C6-O6

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 C

Bond lengths

Bond angles

Torsions

Rings

Oligosaccharide Chain D

Bond lengths

Bond angles

Torsions

Rings
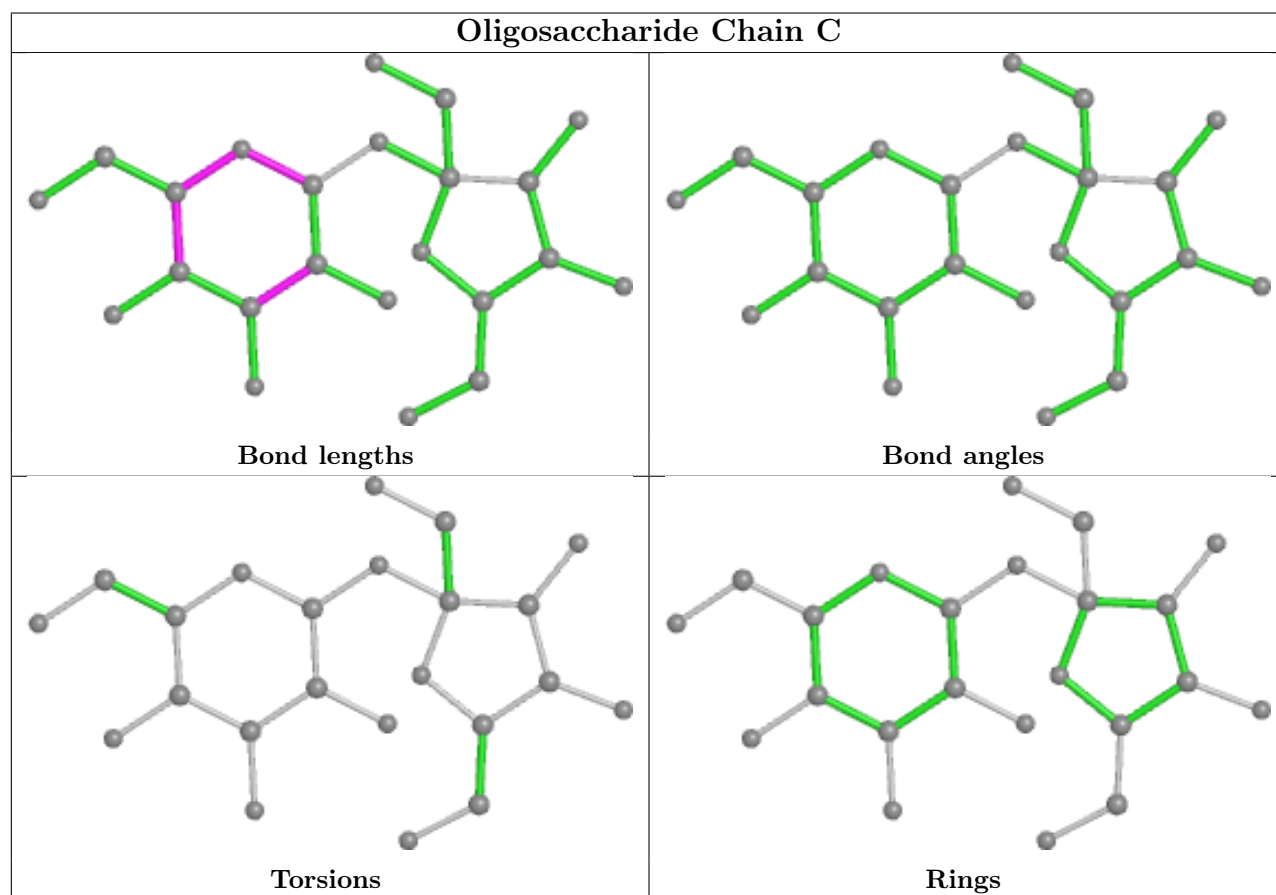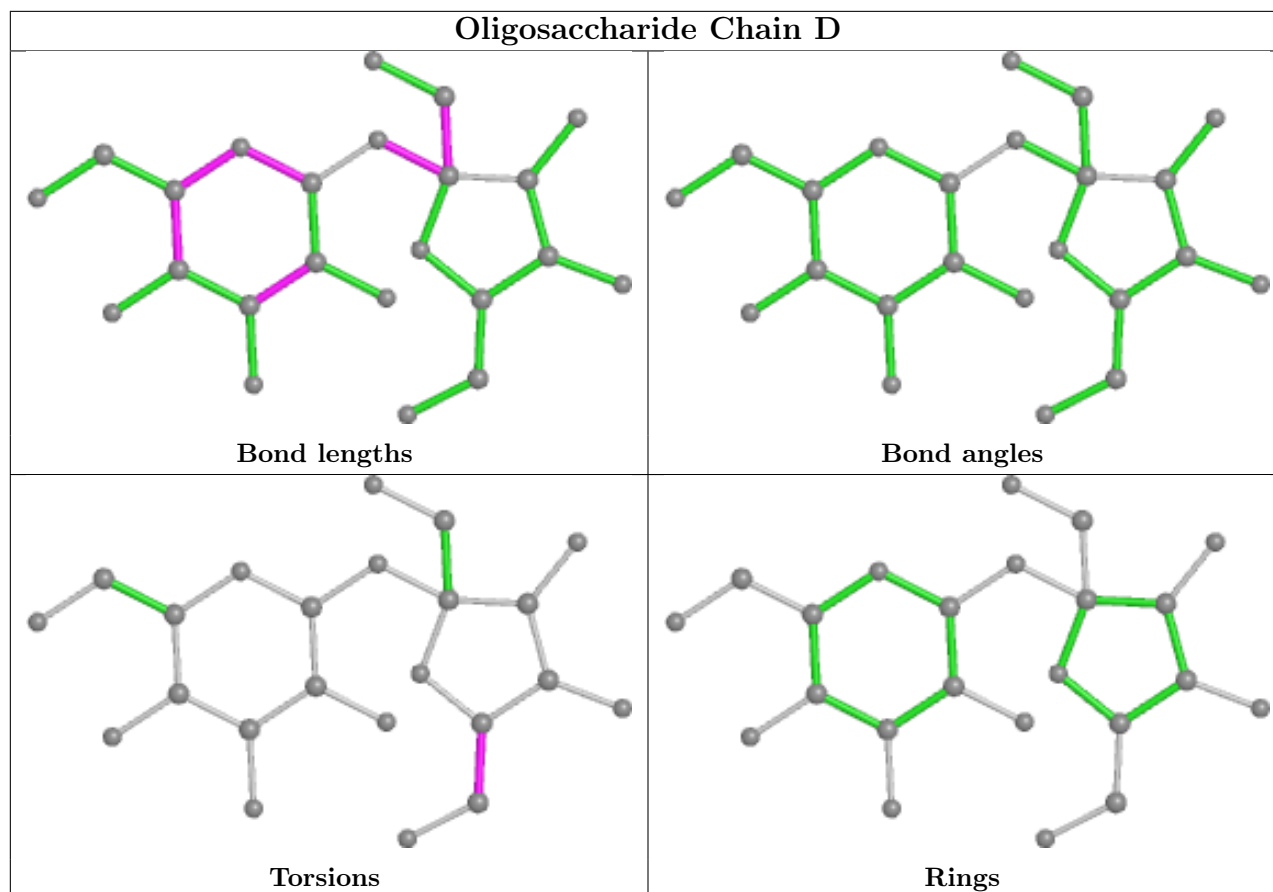
5.6   Ligand geometry ⓘ

9 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	$\#|Z| > 2$
3	SO4	B	1104	-	4,4,4	0.22	0	6,6,6	0.23	0
3	SO4	A	1105	-	4,4,4	0.28	0	6,6,6	0.15	0
3	SO4	B	1106	-	4,4,4	0.26	0	6,6,6	0.15	0
3	SO4	A	1101	-	4,4,4	0.36	0	6,6,6	0.73	0
3	SO4	A	1103	-	4,4,4	0.25	0	6,6,6	0.15	0
3	SO4	B	1108	-	4,4,4	0.30	0	6,6,6	0.08	0
3	SO4	A	1107	-	4,4,4	0.24	0	6,6,6	0.10	0
3	SO4	B	1109	-	4,4,4	0.29	0	6,6,6	0.13	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#|Z| > 2$	Counts	RMSZ	$\#|Z| > 2$
3	SO4	A	1102	-	4,4,4	0.32	0	6,6,6	0.21	0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1105	SO4	1	0
3	A	1103	SO4	1	0

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 ⓘ

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates ⓘ

EDS was not executed - this section is therefore empty.

6.4 Ligands ⓘ

EDS was not executed - this section is therefore empty.

6.5 Other polymers ⓘ

EDS was not executed - this section is therefore empty.