wwPDB X-ray Structure Validation Summary Report (i)

Aug 7, 2020 – 02:50 AM BST

PDB ID	:	1CGY
Title	:	SITE DIRECTED MUTATIONS OF THE ACTIVE SITE RESIDUE TYROSINE 195 OF CYCLODEXTRIN GLYXOSYLTRANSFERASE FROM BACILLUS CIRCULANS STRAIN 251 AFFECTING ACTIVITY AND PRODUCT SPECIFICITY
Authors	:	Strokopytov, B.V.; Dijkstra, B.W.
Deposited on	:	1994-08-05
Resolution	:	2.50 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 following versions of software and data (see references (i)) were used in the production of this report:

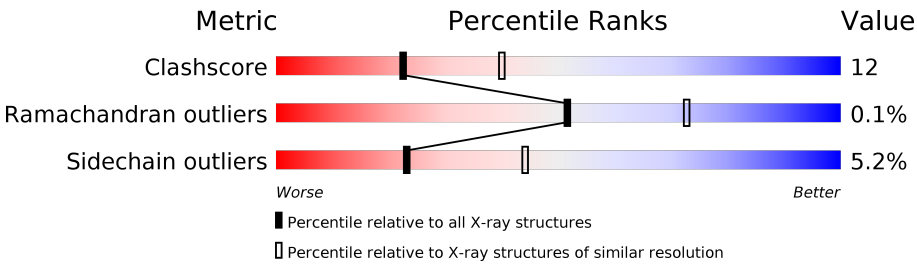
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.13.1

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 2.50 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	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)

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 on 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	686	71% 26% •
2	B	2	100%
2	C	2	100%
2	D	2	100%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 5510 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 CYCLOMALTODEXTRIN GLUCANOTRANSFERASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	686	5266	3323	901	1026	16	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	195	TRP	TYR	conflict	UNP P43379

- Molecule 2 is an oligosaccharide called alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace
			Total	C	O			
2	B	2	23	12	11	0	0	0
2	C	2	23	12	11	0	0	0
2	D	2	23	12	11	0	0	0

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Ca		
3	A	2	2	2	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	173	173	173	0	0

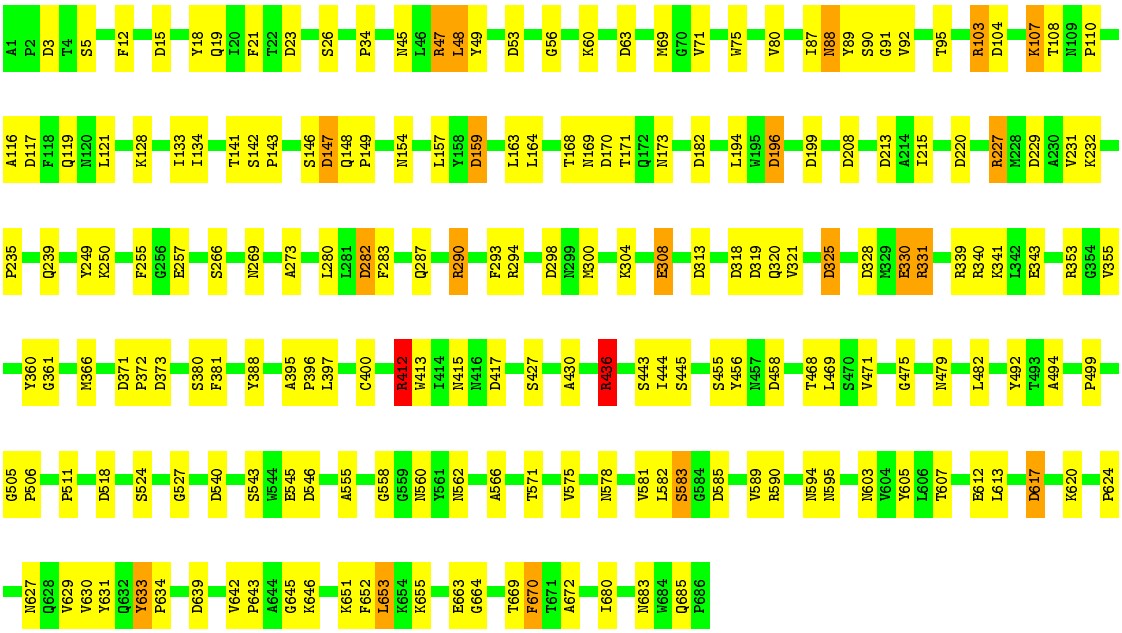
3 Residue-property plots (i)

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 was not executed.

- Molecule 1: CYCLOMALTODEXTRIN GLUCANOTRANSFERASE

Chain A:

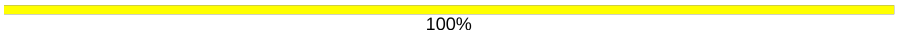


- Molecule 2: alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose

Chain B:

- Molecule 2: alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose

Chain C:



- Molecule 2: alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose

Chain D: 100%

GLC1
GLC2

4 Data and refinement statistics (i)

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	120.18Å 110.70Å 66.21Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.50	Depositor
% Data completeness (in resolution range)	(Not available) (20.00-2.50)	Depositor
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	PROLSQ, TNT	Depositor
R, R_{free}	0.162 , (Not available)	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	5510	wwPDB-VP
Average B, all atoms (Å^2)	22.0	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: CA, GLC

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	0.95	0/5397	1.50	68/7357 (0.9%)

There are no bond length outliers.

The worst 5 of 68 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	331	ARG	NE-CZ-NH2	-18.50	111.05	120.30
1	A	331	ARG	NE-CZ-NH1	15.67	128.13	120.30
1	A	412	ARG	NE-CZ-NH1	13.77	127.18	120.30
1	A	196	ASP	CB-CG-OD1	12.62	129.66	118.30
1	A	63	ASP	CB-CG-OD1	11.57	128.72	118.30

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	5266	0	5026	118	0
2	B	23	0	21	3	0
2	C	23	0	21	0	0
2	D	23	0	21	5	0
3	A	2	0	0	0	0
4	A	173	0	0	4	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	5510	0	5089	121	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 12.

The worst 5 of 121 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:560:ASN:HD21	1:A:578:ASN:HA	1.24	1.01
1:A:88:ASN:HD21	1:A:91:GLY:H	1.04	0.97
1:A:142:SER:HB2	1:A:143:PRO:HD2	1.61	0.82
1:A:88:ASN:HD21	1:A:91:GLY:N	1.81	0.76
1:A:159:ASP:HB2	1:A:164:LEU:HD11	1.73	0.70

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	684/686 (100%)	647 (95%)	36 (5%)	1 (0%)	51 73

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	629	VAL

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	556/556 (100%)	527 (95%)	29 (5%)	23 44

5 of 29 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	227	ARG
1	A	330	GLU
1	A	613	LEU
1	A	308	GLU
1	A	353	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	169	ASN
1	A	172	GLN
1	A	479	ASN
1	A	120	ASN
1	A	416	ASN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates (i)

6 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	B	1	2	12,12,12	0.45	0	17,17,17	1.07	1 (5%)
2	GLC	B	2	2	11,11,12	0.45	0	15,15,17	1.23	3 (20%)
2	GLC	C	1	2	12,12,12	0.31	0	17,17,17	1.17	1 (5%)
2	GLC	C	2	2	11,11,12	0.53	0	15,15,17	1.12	2 (13%)
2	GLC	D	1	2	12,12,12	0.47	0	17,17,17	1.65	4 (23%)
2	GLC	D	2	2	11,11,12	0.25	0	15,15,17	1.92	3 (20%)

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	B	1	2	-	0/2/22/22	0/1/1/1
2	GLC	B	2	2	-	2/2/19/22	0/1/1/1
2	GLC	C	1	2	-	2/2/22/22	0/1/1/1
2	GLC	C	2	2	-	0/2/19/22	0/1/1/1
2	GLC	D	1	2	-	0/2/22/22	0/1/1/1
2	GLC	D	2	2	-	2/2/19/22	0/1/1/1

There are no bond length outliers.

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	2	GLC	C1-O5-C5	4.87	118.79	112.19
2	D	2	GLC	O5-C1-C2	4.00	116.95	110.77
2	D	1	GLC	O4-C4-C3	-3.51	102.24	110.35
2	D	1	GLC	O5-C1-C2	3.38	116.31	110.28
2	C	1	GLC	O5-C5-C4	-3.02	104.21	109.69

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	C	1	GLC	O5-C5-C6-O6
2	C	1	GLC	C4-C5-C6-O6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
2	D	2	GLC	O5-C5-C6-O6
2	D	2	GLC	C4-C5-C6-O6
2	B	2	GLC	C4-C5-C6-O6

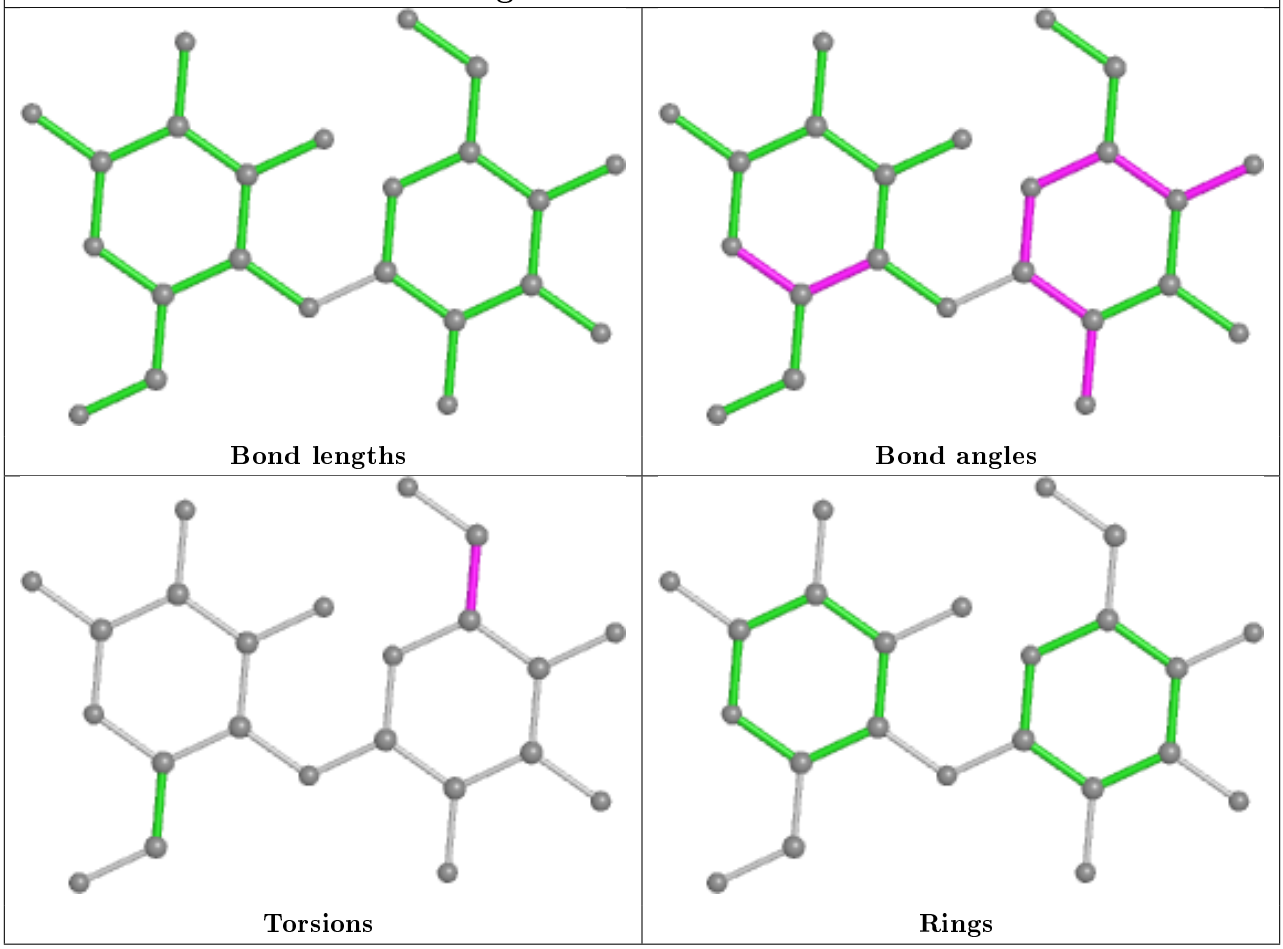
There are no ring outliers.

4 monomers are involved in 8 short contacts:

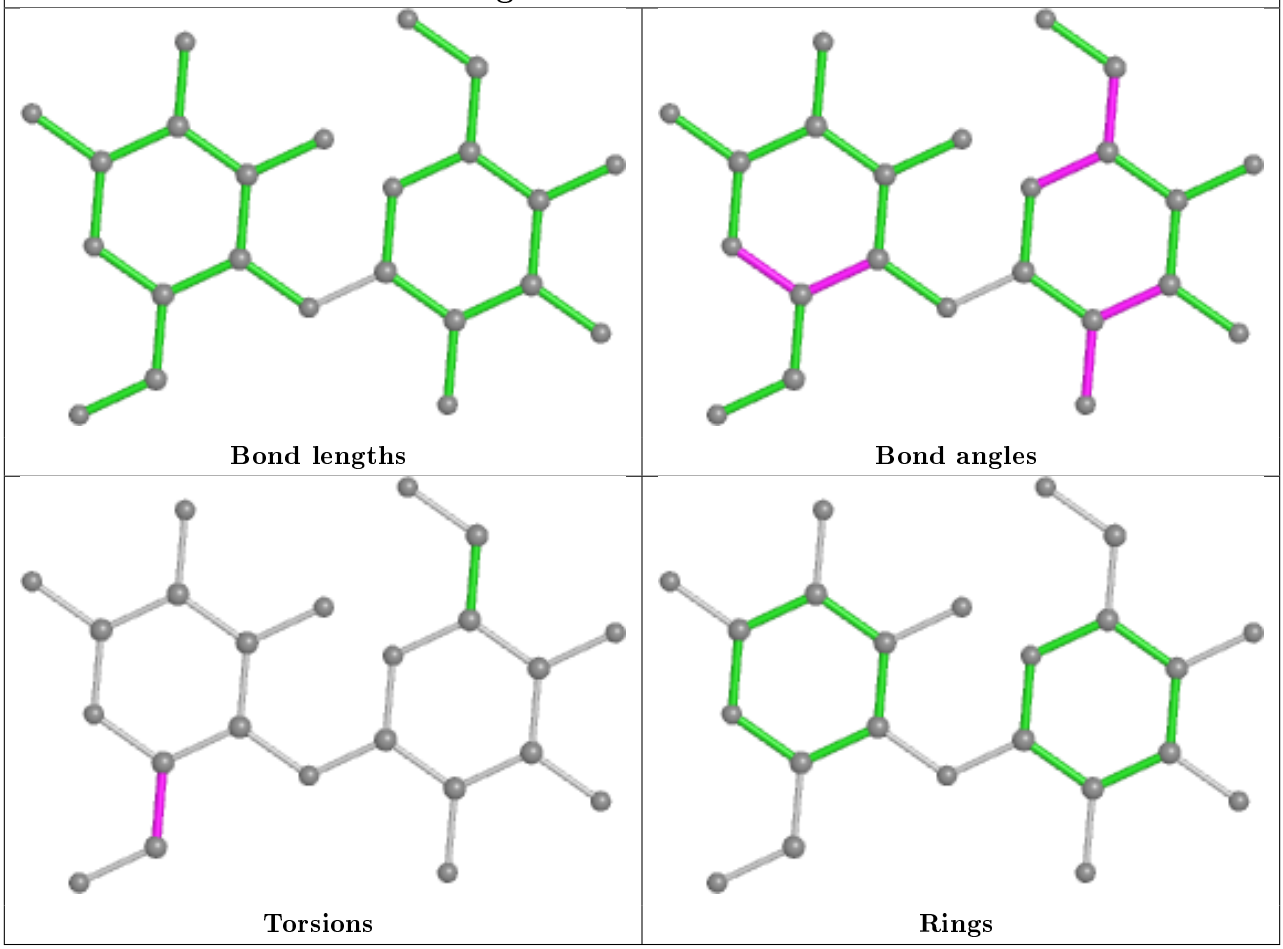
Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	D	1	GLC	3	0
2	D	2	GLC	5	0
2	B	2	GLC	1	0
2	B	1	GLC	2	0

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

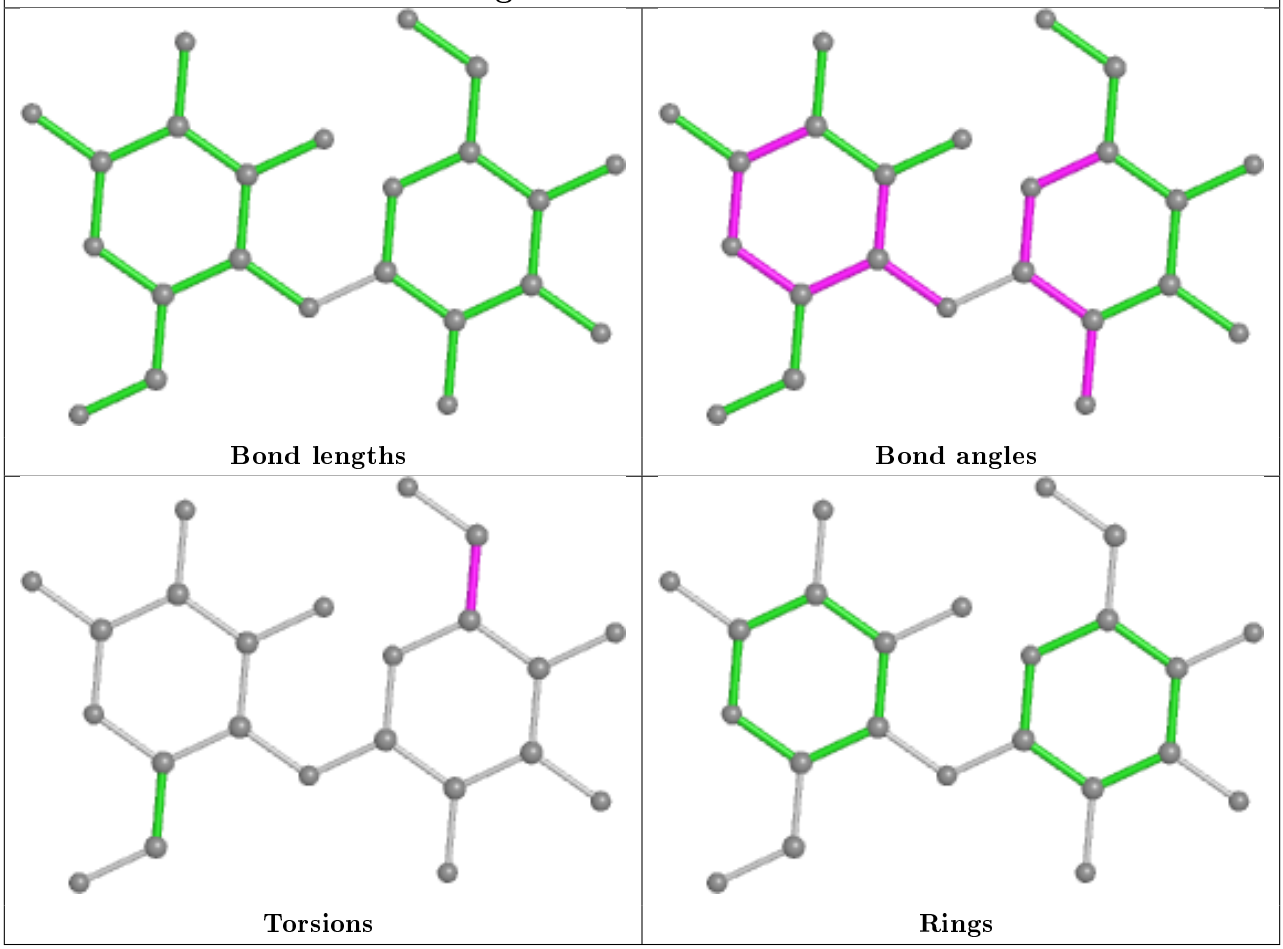
Oligosaccharide Chain B



Oligosaccharide Chain C



Oligosaccharide Chain D



5.6 Ligand geometry (i)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

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.

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 (i)

6.1 Protein, DNA and RNA chains (i)

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates (i)

EDS was not executed - this section is therefore empty.

6.4 Ligands (i)

EDS was not executed - this section is therefore empty.

6.5 Other polymers (i)

EDS was not executed - this section is therefore empty.