wwPDB X-ray Structure Validation Summary Report (i)

Aug 10, 2020 – 02:44 AM BST

PDB ID	:	4G68
Title	:	Biochemical and structural insights into xylan utilization by the thermophilic bacteriumcaldanaerobius polysaccharolyticus
Authors	:	Agarwal, V.; Nair, S.K.
Deposited on	:	2012-07-18
Resolution	:	1.80 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 (1)) were used in the production of this report:

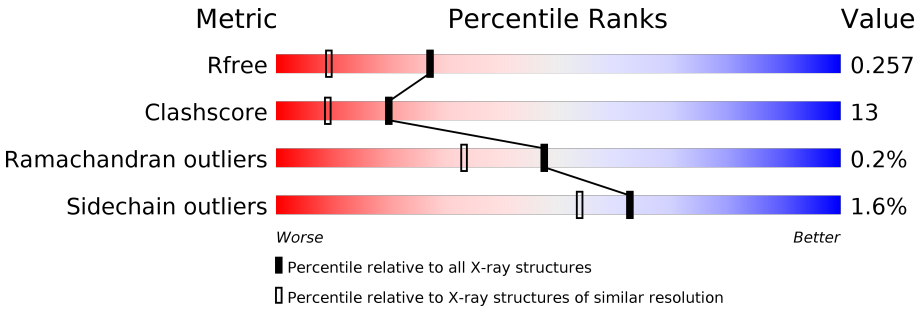
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.13.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.13.1

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 1.80 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	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)

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%

Mol	Chain	Length	Quality of chain
1	A	456	76% 9% • 14%
2	B	432	81% 9% • 9%
3	C	432	63% 26% • 9%
4	D	3	100%
4	E	3	100%

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 10210 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 ABC transporter.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	392	3058	1957	494	597	4	6	0	2	0

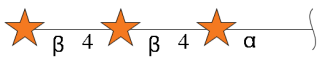
- Molecule 2 is a protein called ABC transporter.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
2	B	392	3050	1951	494	596	4	5	0	0	0

- Molecule 3 is a protein called ABC transporter.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
3	C	392	3050	1951	494	596	8	1	0	0	0

- Molecule 4 is an oligosaccharide called beta-D-xylopyranose-(1-4)-beta-D-xylopyranose-(1-4)-alpha-D-xylopyranose.



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace
			Total	C	O			
4	D	3	28	15	13	0	0	0
4	E	3	28	15	13	0	0	0

- Molecule 5 is water.

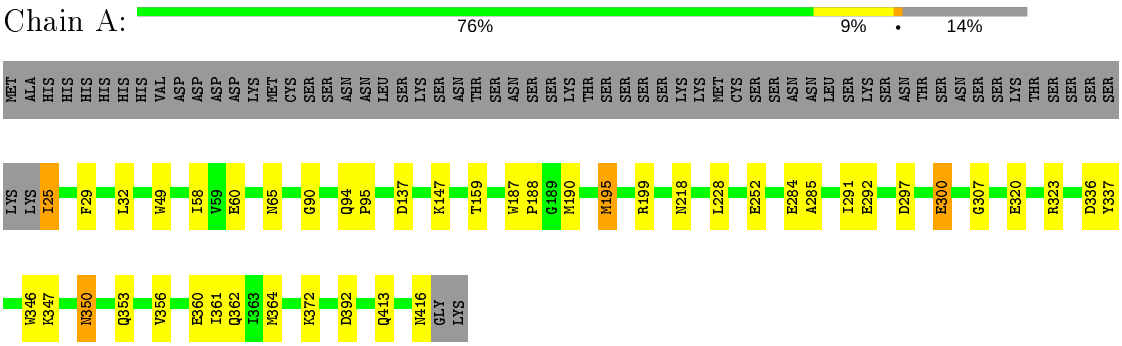
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	364	Total O 364 364	0	0
5	B	363	Total O 363 363	0	0
5	C	269	Total O 269 269	0	0

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.

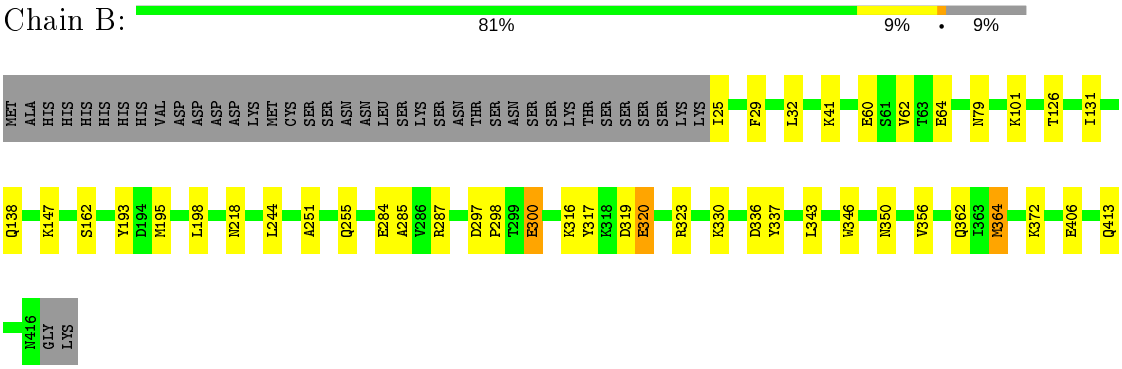
- Molecule 1: ABC transporter

Chain A:



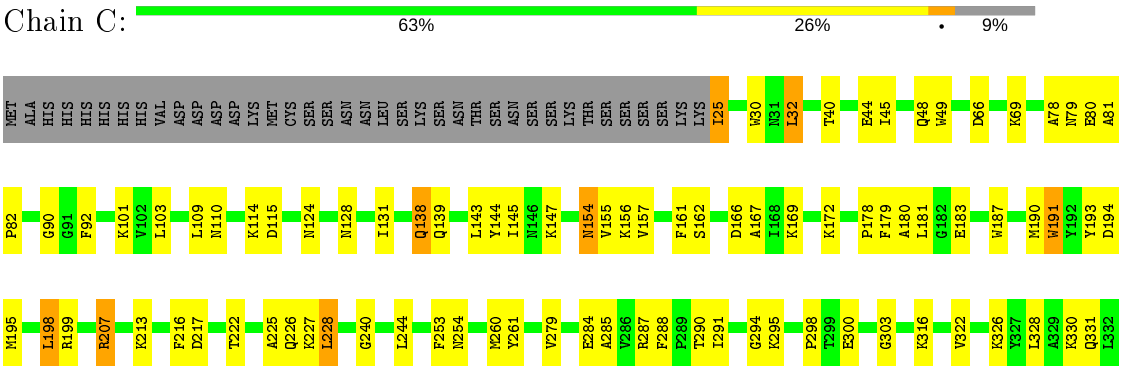
- Molecule 2: ABC transporter

Chain B:



- Molecule 3: ABC transporter

Chain C:

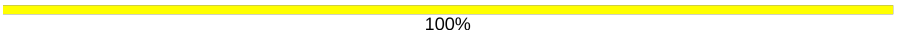


- Molecule 4: beta-D-xylopyranose-(1-4)-beta-D-xylopyranose-(1-4)-alpha-D-xylopyranose

Chain D:

- Molecule 4: beta-D-xylopyranose-(1-4)-beta-D-xylopyranose-(1-4)-alpha-D-xylopyranose

Chain E:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	59.33Å 150.86Å 150.88Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	25.00 – 1.80 38.36 – 1.80	Depositor EDS
% Data completeness (in resolution range)	100.0 (25.00-1.80) 99.7 (38.36-1.80)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	6.42 (at 1.81Å)	Xtriage
Refinement program	REFMAC 5.5.0056	Depositor
R, R_{free}	0.230 , 0.261 0.215 , 0.257	Depositor DCC
R_{free} test set	6314 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	13.8	Xtriage
Anisotropy	0.257	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 19.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.43$, $< L^2 > = 0.25$	Xtriage
Estimated twinning fraction	0.478 for -h,l,k	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	10210	wwPDB-VP
Average B, all atoms (Å^2)	17.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.92% 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: XYP, XYS

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	1.11	4/3125 (0.1%)	0.99	6/4224 (0.1%)
2	B	1.08	5/3112 (0.2%)	0.97	9/4209 (0.2%)
3	C	0.91	5/3116 (0.2%)	0.87	3/4221 (0.1%)
All	All	1.04	14/9353 (0.1%)	0.94	18/12654 (0.1%)

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	320	GLU	CD-OE1	6.94	1.33	1.25
2	B	406	GLU	CD-OE2	-6.76	1.18	1.25
2	B	346	TRP	CD2-CE2	6.33	1.49	1.41
2	B	320	GLU	CD-OE1	6.33	1.32	1.25
2	B	64	GLU	CD-OE1	6.33	1.32	1.25

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	195	MET	CG-SD-CE	-11.49	81.82	100.20
1	A	323	ARG	NE-CZ-NH1	8.62	124.61	120.30
1	A	323	ARG	NE-CZ-NH2	-7.25	116.67	120.30
3	C	207	ARG	NE-CZ-NH2	-6.72	116.94	120.30
1	A	195	MSE	CG-SE-CE	-6.64	84.30	98.90

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	3058	0	2995	32	1
2	B	3050	0	2981	34	1
3	C	3050	0	2981	165	2
4	D	28	0	9	0	0
4	E	28	0	9	0	0
5	A	364	0	0	14	0
5	B	363	0	0	14	0
5	C	269	0	0	126	0
All	All	10210	0	8975	229	2

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 13.

The worst 5 of 229 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)
2:B:218:ASN:HB3	5:B:1245:HOH:O	1.21	1.29
3:C:363:ILE:HG22	5:C:764:HOH:O	1.13	1.27
3:C:69:LYS:HD2	5:C:701:HOH:O	1.25	1.27
3:C:198:LEU:HG	5:C:765:HOH:O	1.28	1.27
3:C:279:VAL:HG21	5:C:584:HOH:O	1.26	1.26

All (2) 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)
2:B:162:SER:O	3:C:154:ASN:ND2[1_455]	2.02	0.18
1:A:347:LYS:NZ	3:C:295:LYS:NZ[1_455]	2.13	0.07

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	392/456 (86%)	386 (98%)	5 (1%)	1 (0%)	41	27
2	B	390/432 (90%)	382 (98%)	8 (2%)	0	100	100
3	C	390/432 (90%)	379 (97%)	10 (3%)	1 (0%)	41	27
All	All	1172/1320 (89%)	1147 (98%)	23 (2%)	2 (0%)	47	33

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	154	ASN
1	A	65	ASN

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	321/375 (86%)	317 (99%)	4 (1%)	71	65
2	B	319/352 (91%)	315 (99%)	4 (1%)	69	62
3	C	319/356 (90%)	312 (98%)	7 (2%)	52	39
All	All	959/1083 (89%)	944 (98%)	15 (2%)	62	54

5 of 15 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	356	VAL

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	B	372	LYS
3	C	228	LEU
2	B	244	LEU
3	C	198	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 23 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	65	ASN
2	B	110	ASN
3	C	350	ASN
2	B	79	ASN
2	B	219	GLN

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	$\#\|Z\|>2$
4	XYS	D	1	4	10,10,10	2.13	3 (30%)	14,14,14	3.30	8 (57%)
4	XYP	D	2	4	9,9,10	0.98	0	10,12,14	2.20	4 (40%)
4	XYP	D	3	4	9,9,10	1.03	0	10,12,14	1.43	1 (10%)
4	XYS	E	1	4	10,10,10	2.07	3 (30%)	14,14,14	3.21	7 (50%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\|>2$	Counts	RMSZ	$\#\|Z\|>2$
4	XYP	E	2	4	9,9,10	1.23	1 (11%)	10,12,14	2.12	3 (30%)
4	XYP	E	3	4	9,9,10	0.96	0	10,12,14	1.11	1 (10%)

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	XYS	D	1	4	-	-	0/1/1/1
4	XYP	D	2	4	-	-	0/1/1/1
4	XYP	D	3	4	-	-	0/1/1/1
4	XYS	E	1	4	-	-	0/1/1/1
4	XYP	E	2	4	-	-	0/1/1/1
4	XYP	E	3	4	-	-	0/1/1/1

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	D	1	XYS	O5-C5	-5.14	1.35	1.43
4	E	1	XYS	O5-C5	-4.53	1.36	1.43
4	E	1	XYS	O5-C1	-3.12	1.38	1.43
4	D	1	XYS	O5-C1	-2.55	1.39	1.43
4	E	2	XYP	C2-C3	2.54	1.56	1.52

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	D	1	XYS	C5-O5-C1	7.63	125.54	112.71
4	E	1	XYS	C5-O5-C1	6.81	124.17	112.71
4	E	1	XYS	O5-C1-C2	6.77	119.49	109.43
4	D	1	XYS	O5-C5-C4	5.33	118.99	110.77
4	E	2	XYP	C4-C3-C2	-4.77	105.26	110.92

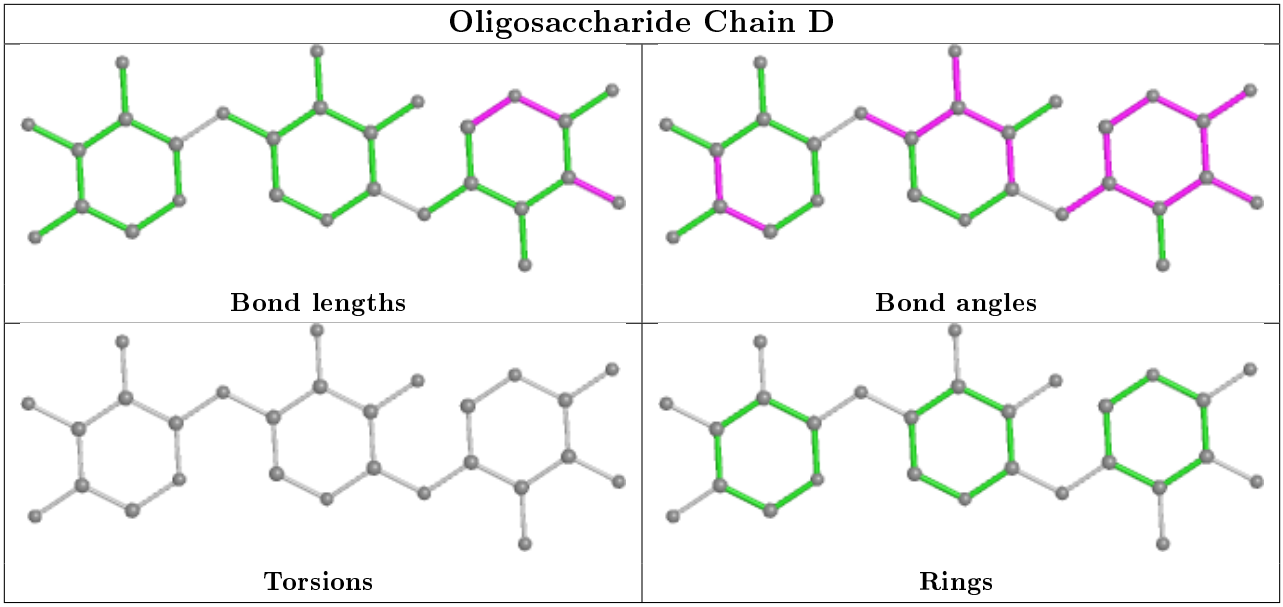
There are no chirality outliers.

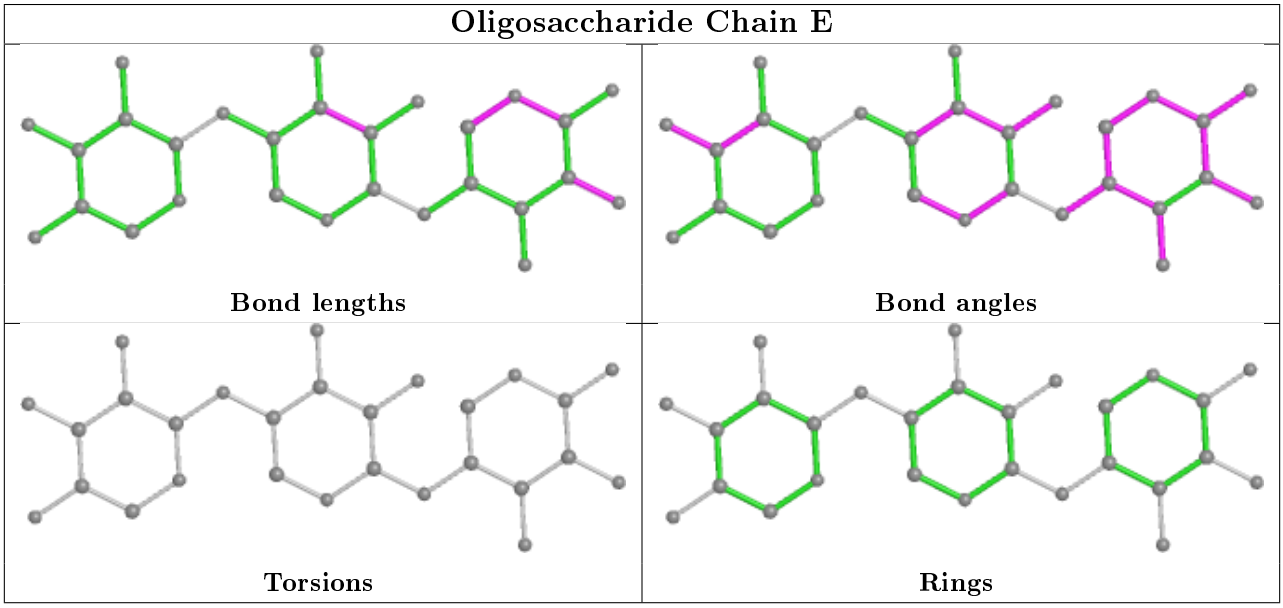
There are no torsion outliers.

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.





5.6 Ligand geometry

There are no ligands in this entry.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

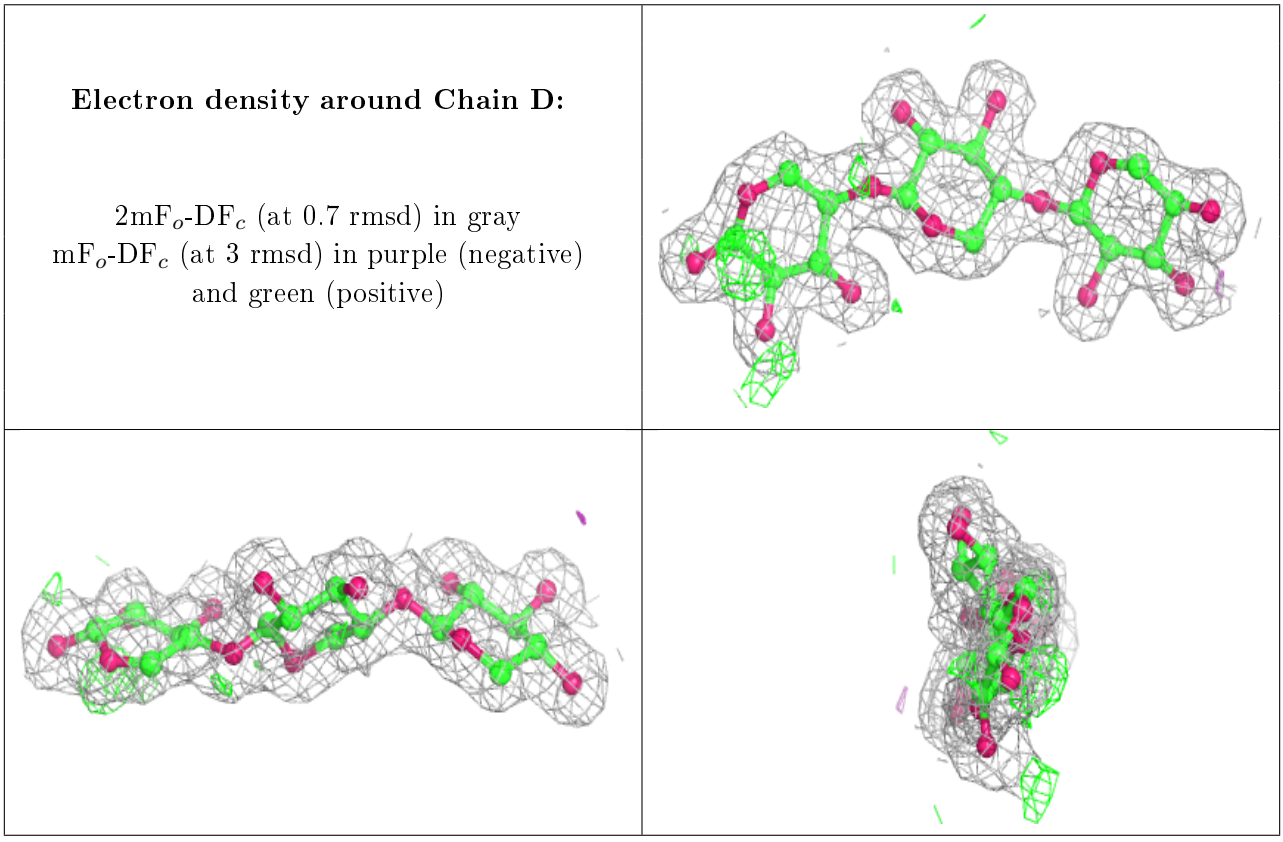
6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

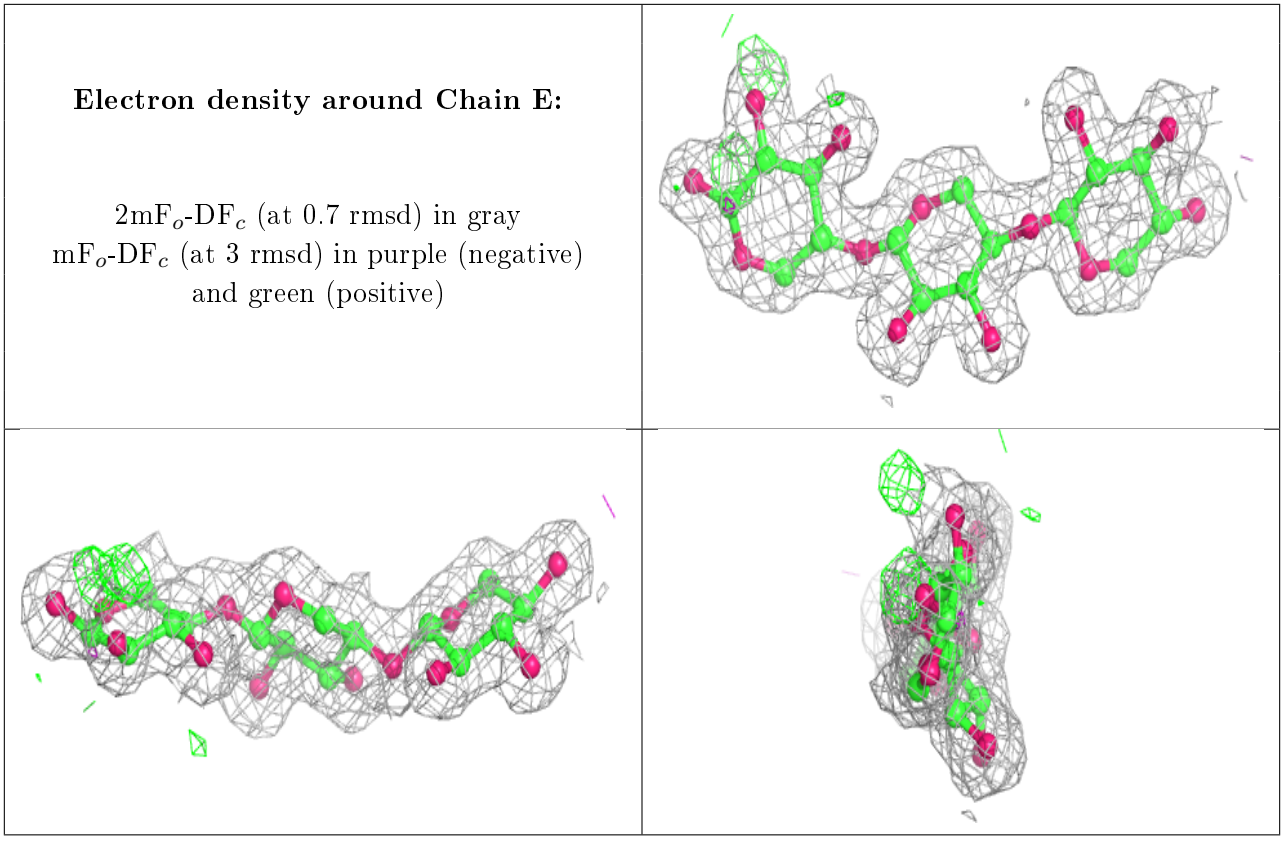
Unable to reproduce the depositors R factor - this section is therefore empty.

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)



6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.