



wwPDB X-ray Structure Validation Summary Report

Dec 3, 2023 – 06:53 pm GMT

PDB ID : 2CDP
Title : Structure of a CBM6 in complex with neoagarohexaose
Authors : Henshaw, J.; Horne, A.; Van Bueren, A.L.; Money, V.A.; Bolam, D.N.; Czjzek, M.; Weiner, R.M.; Hutcheson, S.W.; Davies, G.J.; Boraston, A.B.; Gilbert, H.J.
Deposited on : 2006-01-26
Resolution : 1.59 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  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](#) ) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : **FAILED**
Xtrriage (Phenix) : 1.13
EDS : 2.36
buster-report : **FAILED**
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) : 2.36

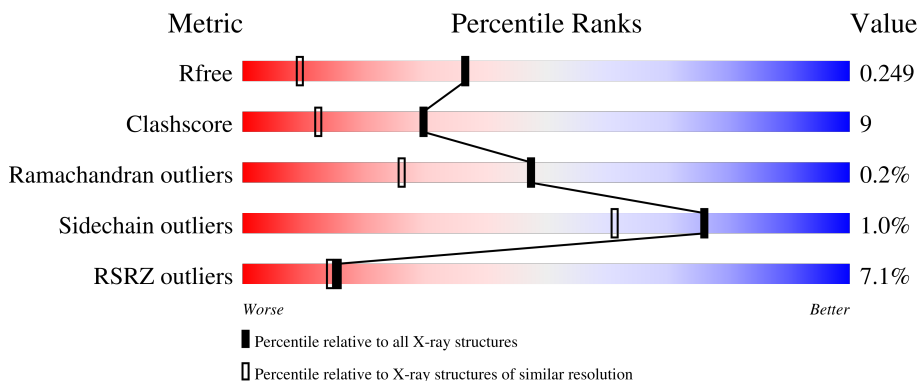
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.59 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	3398 (1.60-1.60)
Clashscore	141614	3665 (1.60-1.60)
Ramachandran outliers	138981	3564 (1.60-1.60)
Sidechain outliers	138945	3563 (1.60-1.60)
RSRZ outliers	127900	3321 (1.60-1.60)

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 $\leq 5\%$. 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	160	 6% 74% 12% 14%
1	B	160	 2% 69% 11% 20%
1	C	160	 4% 71% 9% 21%
1	D	160	 10% 62% 18% 19%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	E	6	 100%
2	F	6	 100%
2	G	6	 100%
2	H	6	 83% 17%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	GAL	E	1	X	-	-	-
2	GAL	F	1	X	-	-	-
2	GAL	G	1	X	-	-	-
2	GAL	H	1	X	-	-	-
4	EDO	A	1141	-	-	X	-

2 Entry composition [i](#)

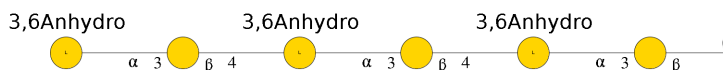
There are 6 unique types of molecules in this entry. The entry contains 4877 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 BETA-AGARASE 1.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
1	A	138	1041	648	177	216	0	4	0
1	B	128	977	613	166	198	0	5	0
1	C	127	978	613	165	200	0	7	0
1	D	130	1012	638	170	204	0	9	0

- Molecule 2 is an oligosaccharide called 3,6-anhydro-alpha-L-galactopyranose-(1-3)-beta-D-galactopyranose-(1-4)-3,6-anhydro-alpha-L-galactopyranose-(1-3)-beta-D-galactopyranose-(1-4)-3,6-anhydro-alpha-L-galactopyranose-(1-3)-beta-D-galactopyranose.



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace
			Total	C	O			
2	E	6	64	36	28	0	0	0
2	F	6	64	36	28	0	0	0
2	G	6	64	36	28	0	0	0
2	H	6	64	36	28	0	0	0

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

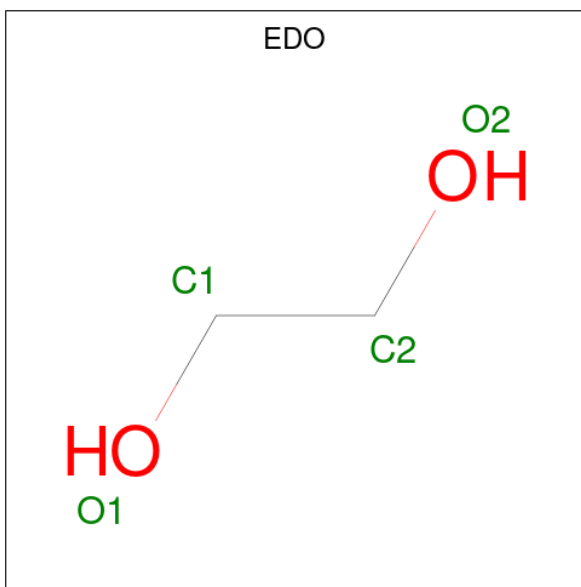
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	2	Total	Ca	0	0
			2	2		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	2	Total Ca 2 2	0	0
3	C	2	Total Ca 2 2	0	0
3	D	2	Total Ca 2 2	0	0

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C₂H₆O₂).



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

- Molecule 5 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	3	Total Cl 3 3	0	0
5	B	3	Total Cl 3 3	0	0
5	C	2	Total Cl 2 2	0	0
5	D	1	Total Cl 1 1	0	0

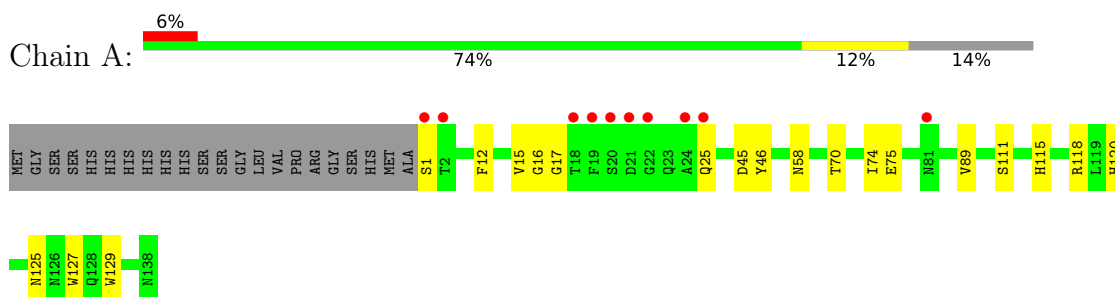
- Molecule 6 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	156	Total O 156 156	0	0
6	B	140	Total O 140 140	0	0
6	C	136	Total O 136 136	0	0
6	D	140	Total O 140 140	0	0

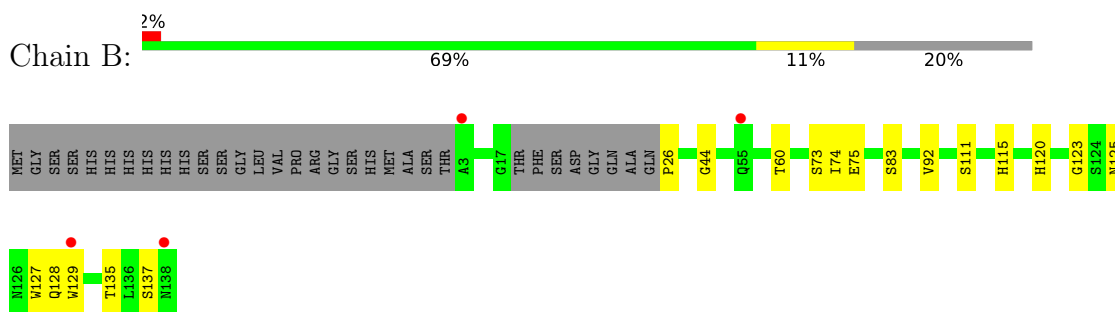
3 Residue-property plots [i](#)

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 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.

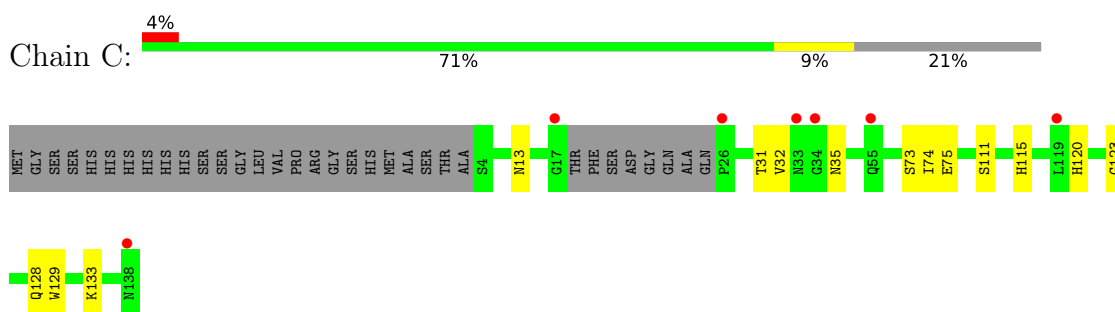
- Molecule 1: BETA-AGARASE 1



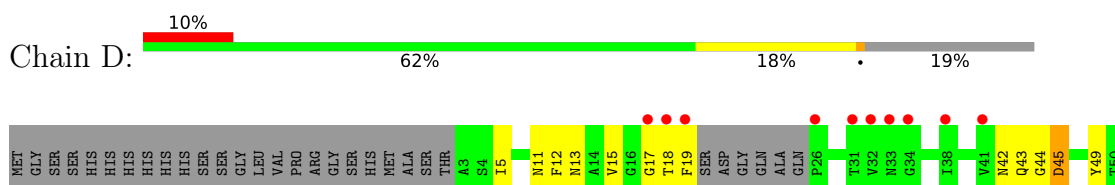
- Molecule 1: BETA-AGARASE 1



- Molecule 1: BETA-AGARASE 1



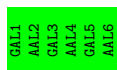
- Molecule 1: BETA-AGARASE 1





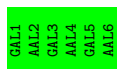
- Molecule 2: 3,6-anhydro-alpha-L-galactopyranose-(1-3)-beta-D-galactopyranose-(1-4)-3,6-anhydro-alpha-L-galactopyranose-(1-3)-beta-D-galactopyranose

Chain E:  100%



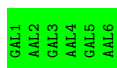
- Molecule 2: 3,6-anhydro-alpha-L-galactopyranose-(1-3)-beta-D-galactopyranose-(1-4)-3,6-anhydro-alpha-L-galactopyranose-(1-3)-beta-D-galactopyranose

Chain F:  100%





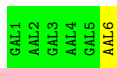
- Molecule 2: 3,6-anhydro-alpha-L-galactopyranose-(1-3)-beta-D-galactopyranose-(1-4)-3,6-anhydro-alpha-L-galactopyranose-(1-3)-beta-D-galactopyranose

Chain G:  100%



- Molecule 2: 3,6-anhydro-alpha-L-galactopyranose-(1-3)-beta-D-galactopyranose-(1-4)-3,6-anhydro-alpha-L-galactopyranose-(1-3)-beta-D-galactopyranose

Chain H:  83%  17%



4 Data and refinement statistics i

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	53.72Å 55.00Å 195.91Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.62 – 1.59 19.61 – 1.59	Depositor EDS
% Data completeness (in resolution range)	98.3 (19.62-1.59) 98.3 (19.61-1.59)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.13 (at 1.59Å)	Xtrriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.210 , 0.253 0.206 , 0.249	Depositor DCC
R_{free} test set	3905 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å ²)	22.0	Xtrriage
Anisotropy	0.019	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.37 , 46.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.027 for k,h,-l	Xtrriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	4877	wwPDB-VP
Average B, all atoms (Å ²)	21.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.80% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

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: EDO, CL, AAL, CA, GAL

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	0.92	0/1077	0.81	1/1474 (0.1%)
1	B	0.95	0/1014	0.76	0/1386
1	C	0.96	0/1021	0.79	0/1396
1	D	0.90	0/1062	0.82	1/1453 (0.1%)
All	All	0.93	0/4174	0.80	2/5709 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	D	0	1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	D	45	ASP	CB-CG-OD1	6.09	123.78	118.30
1	A	118	ARG	NE-CZ-NH2	-5.65	117.48	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	42	ASN	Peptide

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	1041	0	963	16	0
1	B	977	0	916	15	0
1	C	978	0	916	11	0
1	D	1012	0	955	36	0
2	E	64	0	51	0	0
2	F	64	0	51	0	0
2	G	64	0	51	0	0
2	H	64	0	51	1	0
3	A	2	0	0	0	0
3	B	2	0	0	0	0
3	C	2	0	0	0	0
3	D	2	0	0	0	0
4	A	4	0	6	4	0
4	B	4	0	6	0	0
4	C	4	0	6	0	0
4	D	12	0	18	3	0
5	A	3	0	0	1	0
5	B	3	0	0	1	0
5	C	2	0	0	0	0
5	D	1	0	0	0	0
6	A	156	0	0	5	0
6	B	140	0	0	5	0
6	C	136	0	0	1	2
6	D	140	0	0	8	2
All	All	4877	0	3990	78	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 9.

The worst 5 of 78 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:D:19:PHE:CE1	1:D:126[B]:ASN:OD1	1.82	1.30
1:D:51[A]:ILE:HD12	1:D:117[A]:VAL:HG21	1.24	1.09
1:D:13[A]:ASN:ND2	6:D:2014:HOH:O	1.75	0.98

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:51[A]:ILE:HD12	1:D:117[A]:VAL:CG2	1.95	0.94
1:D:13[B]:ASN:ND2	6:D:2013:HOH:O	1.99	0.88

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)
6:C:2053:HOH:O	6:D:2111:HOH:O[1_455]	2.12	0.08
6:C:2054:HOH:O	6:D:2111:HOH:O[1_455]	2.19	0.01

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	140/160 (88%)	137 (98%)	3 (2%)	0	100	100
1	B	129/160 (81%)	126 (98%)	3 (2%)	0	100	100
1	C	130/160 (81%)	126 (97%)	4 (3%)	0	100	100
1	D	135/160 (84%)	130 (96%)	4 (3%)	1 (1%)	22	7
All	All	534/640 (83%)	519 (97%)	14 (3%)	1 (0%)	47	26

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	43	GLN

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	110/124 (89%)	110 (100%)	0	100	100
1	B	103/124 (83%)	102 (99%)	1 (1%)	76	61
1	C	105/124 (85%)	104 (99%)	1 (1%)	76	61
1	D	109/124 (88%)	107 (98%)	2 (2%)	59	36
All	All	427/496 (86%)	423 (99%)	4 (1%)	76	65

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	83	SER
1	C	32	VAL
1	D	18	THR
1	D	133	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 20 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	33	ASN
1	D	120	HIS
1	D	138	ASN
1	D	128	GLN
1	B	120	HIS

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

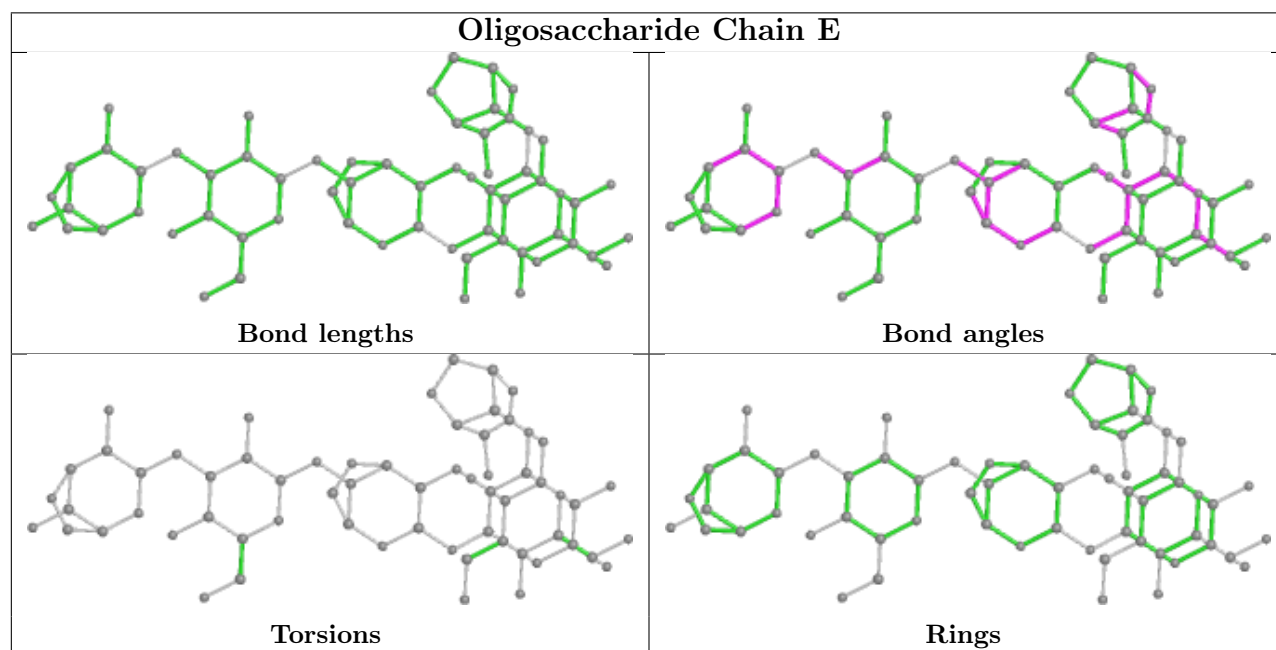
5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

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

5.5 Carbohydrates [i](#)

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

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 [i](#)

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

5.7 Other polymers [i](#)

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

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](#)

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(Å ²)	Q<0.9
1	A	138/160 (86%)	0.47	10 (7%) 15 14	14, 19, 32, 36	0
1	B	128/160 (80%)	0.25	4 (3%) 49 46	14, 18, 27, 36	0
1	C	127/160 (79%)	0.47	7 (5%) 25 22	14, 19, 29, 36	0
1	D	130/160 (81%)	0.66	16 (12%) 4 3	13, 21, 32, 38	2 (1%)
All	All	523/640 (81%)	0.47	37 (7%) 16 14	13, 19, 31, 38	2 (0%)

The worst 5 of 37 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	18	THR	8.0
1	A	24	ALA	8.0
1	C	34	GLY	6.4
1	B	3	ALA	5.9
1	D	34	GLY	4.4

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

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

6.3 Carbohydrates [i](#)

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
2	GAL	H	1	12/12	0.57	0.24	40,45,48,50	0
2	GAL	F	1	12/12	0.62	0.26	45,54,56,56	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
2	GAL	E	1	12/12	0.65	0.33	48,57,60,61	0
2	GAL	G	1	12/12	0.86	0.17	29,31,34,36	0
2	AAL	H	2	10/11	0.88	0.16	24,30,34,35	0
2	AAL	F	2	10/11	0.89	0.12	24,30,35,38	0
2	AAL	E	2	10/11	0.90	0.20	25,33,37,41	0
2	GAL	G	5	11/12	0.91	0.09	16,18,19,20	0
2	GAL	F	3	11/12	0.91	0.12	17,20,24,28	0
2	GAL	G	3	11/12	0.91	0.12	19,21,28,29	0
2	AAL	G	6	10/11	0.92	0.09	19,21,23,23	0
2	AAL	G	2	10/11	0.93	0.11	22,25,26,27	0
2	AAL	F	4	10/11	0.94	0.09	15,17,18,19	0
2	AAL	H	6	10/11	0.94	0.09	19,22,24,25	0
2	AAL	G	4	10/11	0.95	0.07	17,19,19,20	0
2	AAL	E	4	10/11	0.95	0.07	16,17,19,20	0
2	GAL	H	3	11/12	0.95	0.08	18,23,28,29	0
2	GAL	E	3	11/12	0.95	0.07	17,22,27,31	0
2	AAL	H	4	10/11	0.96	0.06	15,18,18,19	0
2	GAL	H	5	11/12	0.96	0.07	16,18,21,22	0
2	AAL	F	6	10/11	0.96	0.08	14,16,17,17	0
2	AAL	E	6	10/11	0.97	0.06	15,16,19,19	0
2	GAL	E	5	11/12	0.97	0.07	15,16,17,18	0
2	GAL	F	5	11/12	0.98	0.08	11,16,18,20	0

6.4 Ligands [i](#)

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(\AA^2)	Q<0.9
4	EDO	D	1144	4/4	0.80	0.14	34,36,36,36	0
4	EDO	D	1143	4/4	0.85	0.12	28,28,29,31	0
4	EDO	A	1141	4/4	0.87	0.15	21,23,25,28	0
4	EDO	D	1142	4/4	0.87	0.11	33,35,35,35	0
4	EDO	B	1141	4/4	0.97	0.09	19,22,23,23	0
3	CA	C	1140	1/1	0.97	0.07	24,24,24,24	0
5	CL	C	1143	1/1	0.97	0.05	30,30,30,30	0
3	CA	D	1139	1/1	0.98	0.05	30,30,30,30	0
3	CA	C	1139	1/1	0.98	0.04	27,27,27,27	0
3	CA	A	1139	1/1	0.98	0.05	24,24,24,24	0
5	CL	A	1143	1/1	0.98	0.08	28,28,28,28	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
5	CL	B	1142	1/1	0.98	0.08	27,27,27,27	0
5	CL	B	1143	1/1	0.98	0.03	23,23,23,23	0
5	CL	C	1142	1/1	0.98	0.08	26,26,26,26	1
4	EDO	C	1141	4/4	0.98	0.11	17,17,19,19	0
5	CL	D	1141	1/1	0.98	0.07	32,32,32,32	0
3	CA	D	1140	1/1	0.99	0.21	17,17,17,17	0
3	CA	A	1140	1/1	0.99	0.04	21,21,21,21	0
5	CL	B	1144	1/1	0.99	0.07	26,26,26,26	1
3	CA	B	1139	1/1	0.99	0.05	23,23,23,23	0
3	CA	B	1140	1/1	0.99	0.20	17,17,17,17	0
5	CL	A	1144	1/1	0.99	0.04	27,27,27,27	0
5	CL	A	1142	1/1	1.00	0.04	24,24,24,24	0

6.5 Other polymers [i](#)

There are no such residues in this entry.