



wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 22, 2022 – 04:26 am BST

PDB ID : 8AE5
Title : Crystal structure of human legumain in complex with macrocypin 1a
Authors : Elamin, T.; Brandstetter, H.; Dall, E.
Deposited on : 2022-07-12
Resolution : 2.29 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 : 1.8.4, CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.30
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac : 5.8.0267
CCP4 : 7.1.010 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.30

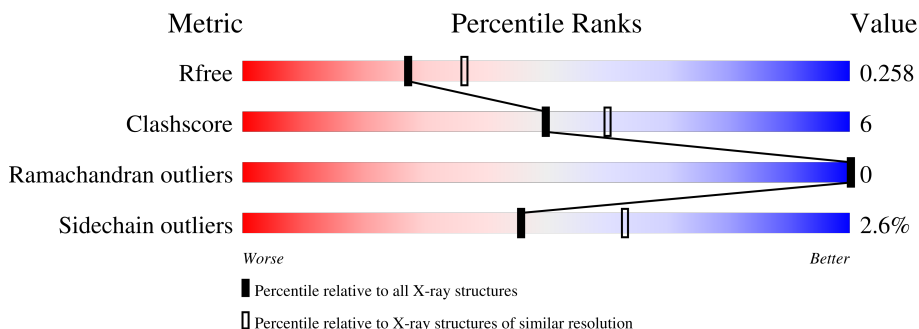
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.29 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	5042 (2.30-2.30)
Clashscore	141614	5643 (2.30-2.30)
Ramachandran outliers	138981	5575 (2.30-2.30)
Sidechain outliers	138945	5575 (2.30-2.30)

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 $\leq 5\%$.

Mol	Chain	Length	Quality of chain
1	A	263	
1	B	263	
2	C	177	
2	D	177	
3	E	2	

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 7053 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 Legumain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	263	Total	C	N	O	S	7	0	0
			2115	1335	361	403	16			
1	B	263	Total	C	N	O	S	41	0	0
			2115	1335	361	403	16			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	147	SNN	ASP	modified residue	UNP Q99538
B	147	SNN	ASP	modified residue	UNP Q99538

- Molecule 2 is a protein called Macrocypin-1a.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	168	Total	C	N	O	S	7	0	0
			1349	865	225	255	4			
2	D	160	Total	C	N	O	S	3	0	0
			1285	825	212	244	4			

There are 16 discrepancies between the modelled and reference sequences:

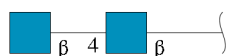
Chain	Residue	Modelled	Actual	Comment	Reference
C	170	LEU	-	expression tag	UNP B9V973
C	171	GLU	-	expression tag	UNP B9V973
C	172	HIS	-	expression tag	UNP B9V973
C	173	HIS	-	expression tag	UNP B9V973
C	174	HIS	-	expression tag	UNP B9V973
C	175	HIS	-	expression tag	UNP B9V973
C	176	HIS	-	expression tag	UNP B9V973
C	177	HIS	-	expression tag	UNP B9V973
D	170	LEU	-	expression tag	UNP B9V973
D	171	GLU	-	expression tag	UNP B9V973

Continued on next page...

Continued from previous page...

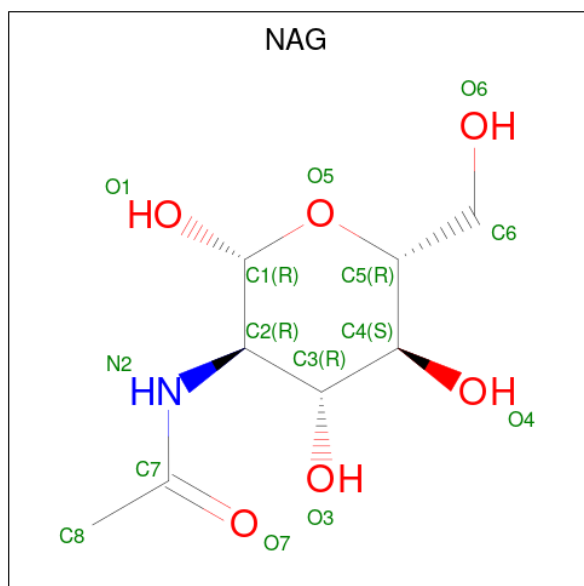
Chain	Residue	Modelled	Actual	Comment	Reference
D	172	HIS	-	expression tag	UNP B9V973
D	173	HIS	-	expression tag	UNP B9V973
D	174	HIS	-	expression tag	UNP B9V973
D	175	HIS	-	expression tag	UNP B9V973
D	176	HIS	-	expression tag	UNP B9V973
D	177	HIS	-	expression tag	UNP B9V973

- Molecule 3 is an oligosaccharide called 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose.



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	E	2	Total	C	N	O	0	0	0
			28	16	2	10			

- Molecule 4 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: C₈H₁₅NO₆).



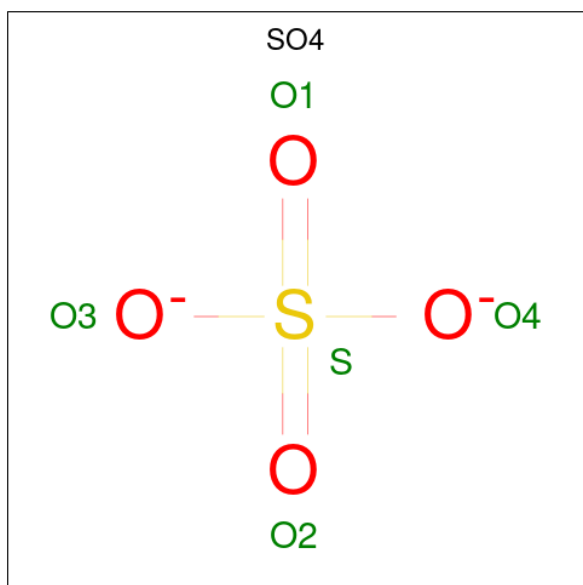
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total	C	N	O	0	0
			14	8	1	5		
4	A	1	Total	C	N	O	0	0
			14	8	1	5		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	B	1	Total	C	N	O	0	0
			14	8	1	5		
4	B	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 5 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



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

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	17	Total	O	0	0
			17	17		
6	C	32	Total	O	0	0
			32	32		

Continued on next page...

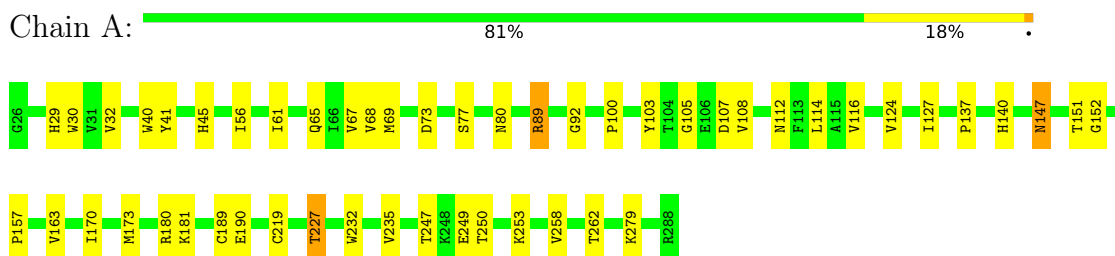
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	B	11	Total	O	0	0
			11	11		
6	D	20	Total	O	0	0
			20	20		

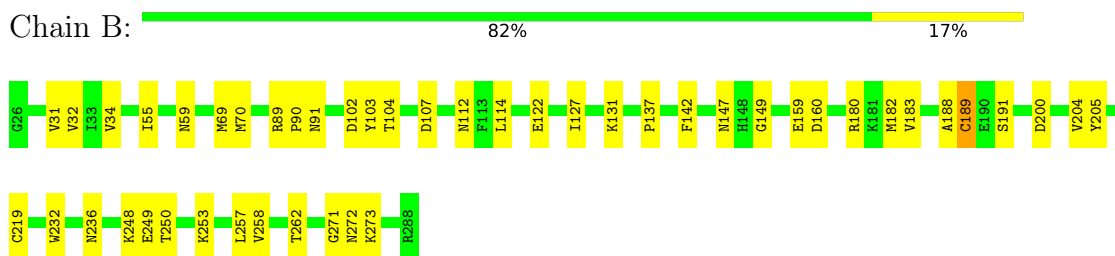
3 Residue-property plots [i](#)

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.

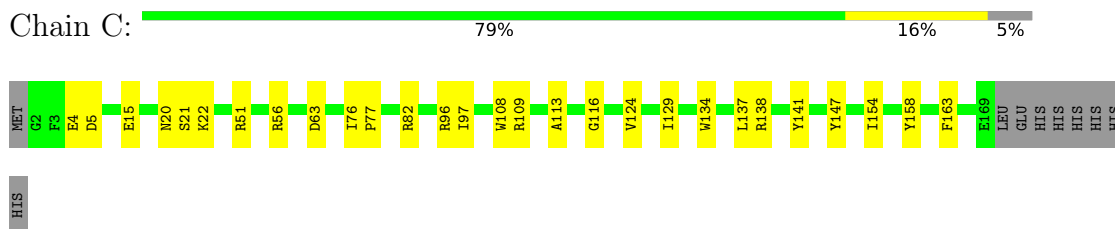
- Molecule 1: Legumain



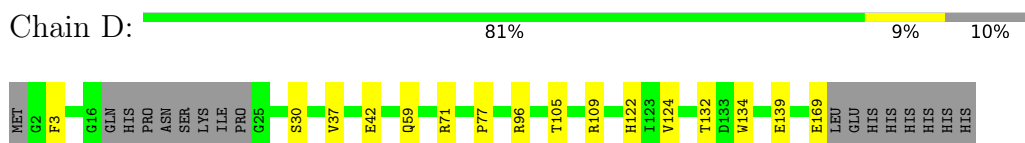
- Molecule 1: Legumain



- Molecule 2: Macrocypin-1a



- Molecule 2: Macrocypin-1a



- Molecule 3: 2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose

Chain E:

100%

MAG1
MAG2

4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α , β , γ	159.18Å 174.12Å 113.00Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.39 – 2.29 49.11 – 2.29	Depositor EDS
% Data completeness (in resolution range)	99.2 (47.39-2.29) 99.4 (49.11-2.29)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	0.95 (at 2.29Å)	Xtriage
Refinement program	PHENIX 1.18.2_3874, PHENIX 1.18.2_3874	Depositor
R, R_{free}	0.221 , 0.258 0.221 , 0.258	Depositor DCC
R_{free} test set	3674 reflections (5.23%)	wwPDB-VP
Wilson B-factor (Å ²)	73.8	Xtriage
Anisotropy	0.381	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	(Not available) , (Not available)	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	7053	wwPDB-VP
Average B, all atoms (Å ²)	93.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.55% 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: NAG, SO4, SNN, SCH

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	0.42	0/2154	0.58	0/2918
1	B	0.40	0/2154	0.61	0/2918
2	C	0.52	0/1394	0.69	0/1905
2	D	0.47	0/1326	0.63	0/1810
All	All	0.44	0/7028	0.62	0/9551

There are no bond length outliers.

There are no bond angle outliers.

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	2115	0	2000	31	0
1	B	2115	0	1999	24	0
2	C	1349	0	1274	18	0
2	D	1285	0	1209	7	0
3	E	28	0	25	0	0
4	A	28	0	26	1	0
4	B	28	0	26	0	0
5	A	5	0	0	0	0
5	B	5	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	C	5	0	0	0	0
5	D	10	0	0	0	0
6	A	17	0	0	0	0
6	B	11	0	0	0	0
6	C	32	0	0	2	0
6	D	20	0	0	0	0
All	All	7053	0	6559	78	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 6.

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:B:250:THR:HG23	1:B:253:LYS:H	1.48	0.77
2:C:56:ARG:NH2	2:C:63:ASP:H	1.83	0.76
1:B:253:LYS:HE2	1:B:257:LEU:HD11	1.71	0.72
1:A:250:THR:HG23	1:A:253:LYS:H	1.58	0.69
1:A:170:ILE:HA	1:A:173:MET:HE2	1.78	0.65

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	257/263 (98%)	247 (96%)	10 (4%)	0	100	100
1	B	257/263 (98%)	246 (96%)	11 (4%)	0	100	100
2	C	166/177 (94%)	159 (96%)	7 (4%)	0	100	100
2	D	156/177 (88%)	148 (95%)	8 (5%)	0	100	100
All	All	836/880 (95%)	800 (96%)	36 (4%)	0	100	100

There are no Ramachandran outliers to report.

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	227/228 (100%)	223 (98%)	4 (2%)	59	75
1	B	227/228 (100%)	222 (98%)	5 (2%)	52	69
2	C	142/151 (94%)	138 (97%)	4 (3%)	43	60
2	D	134/151 (89%)	128 (96%)	6 (4%)	27	39
All	All	730/758 (96%)	711 (97%)	19 (3%)	46	63

5 of 19 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	37	VAL
2	D	132	THR
2	D	169	GLU
2	D	105	THR
1	B	32	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	140	HIS

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

4 non-standard protein/DNA/RNA residues 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$
1	SCH	A	189	1	6,7,8	1.24	0	3,7,9	2.44	1 (33%)
1	SNN	B	147	1	7,8,8	1.34	1 (14%)	7,11,11	2.65	4 (57%)
1	SNN	A	147	1	7,8,8	1.23	2 (28%)	7,11,11	3.24	3 (42%)
1	SCH	B	189	1	6,7,8	1.19	0	3,7,9	2.19	1 (33%)

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
1	SCH	A	189	1	-	1/2/6/8	-
1	SNN	B	147	1	-	-	0/1/1/1
1	SNN	A	147	1	-	-	0/1/1/1
1	SCH	B	189	1	-	0/2/6/8	-

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	147	SNN	C2-N1	-2.69	1.34	1.37
1	A	147	SNN	C2-N1	-2.05	1.34	1.37
1	A	147	SNN	C5-N1	-2.05	1.34	1.37

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	147	SNN	C3-C2-N1	5.40	111.47	107.30
1	A	147	SNN	O2-C2-C3	-5.13	122.47	126.18
1	B	147	SNN	C3-C2-N1	4.80	111.01	107.30
1	A	189	SCH	CB-SG-SD	3.97	114.10	103.82
1	A	147	SNN	O5-C5-C4	-3.59	121.66	126.39

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	189	SCH	CA-CB-SG-SD

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	A	147	SNN	1	0
1	B	189	SCH	1	0

5.5 Carbohydrates [i](#)

2 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$
3	NAG	E	1	3,1	14,14,15	1.02	2 (14%)	17,19,21	1.27	2 (11%)
3	NAG	E	2	3	14,14,15	0.55	0	17,19,21	1.04	1 (5%)

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	NAG	E	1	3,1	-	3/6/23/26	0/1/1/1
3	NAG	E	2	3	-	0/6/23/26	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	E	1	NAG	C1-C2	2.87	1.56	1.52
3	E	1	NAG	O5-C1	2.22	1.47	1.43

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
3	E	1	NAG	O4-C4-C5	3.82	118.79	109.30
3	E	2	NAG	C1-O5-C5	3.81	117.36	112.19
3	E	1	NAG	C1-O5-C5	2.47	115.54	112.19

There are no chirality outliers.

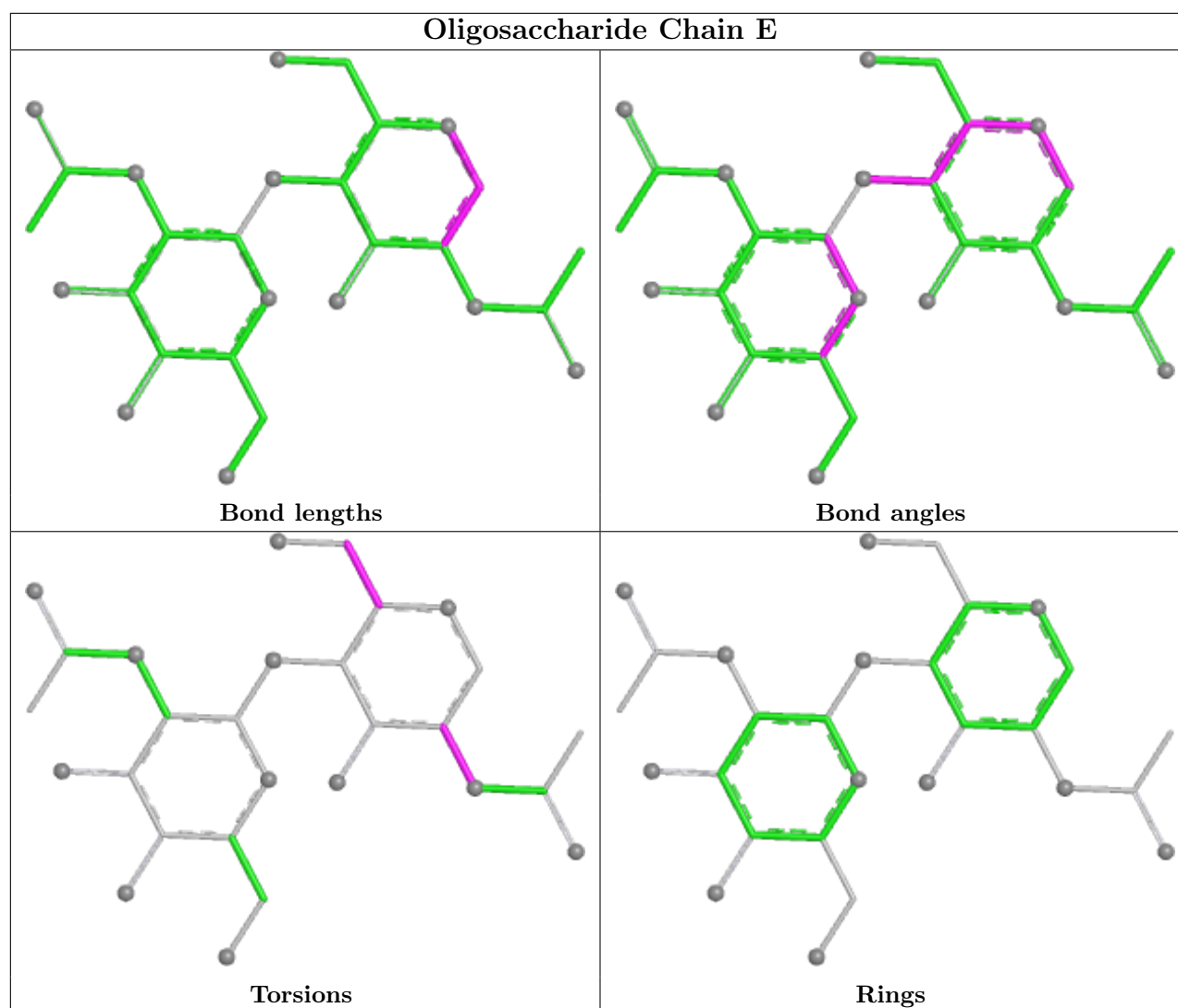
All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	E	1	NAG	C4-C5-C6-O6
3	E	1	NAG	C3-C2-N2-C7
3	E	1	NAG	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.



5.6 Ligand geometry [i](#)

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	$\# Z > 2$
5	SO4	A	303	-	4,4,4	0.23	0	6,6,6	0.31	0
4	NAG	A	302	1	14,14,15	0.68	0	17,19,21	0.61	0
4	NAG	B	301	1	14,14,15	0.36	0	17,19,21	0.77	1 (5%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
4	NAG	B	302	1	14,14,15	0.59	0	17,19,21	0.44	0
5	SO4	D	202	-	4,4,4	0.19	0	6,6,6	0.53	0
5	SO4	B	303	-	4,4,4	0.11	0	6,6,6	0.37	0
4	NAG	A	301	1	14,14,15	1.78	1 (7%)	17,19,21	0.83	1 (5%)
5	SO4	C	201	-	4,4,4	0.15	0	6,6,6	0.35	0
5	SO4	D	201	-	4,4,4	0.20	0	6,6,6	0.28	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	NAG	A	301	1	-	1/6/23/26	0/1/1/1
4	NAG	B	302	1	-	0/6/23/26	0/1/1/1
4	NAG	A	302	1	-	0/6/23/26	0/1/1/1
4	NAG	B	301	1	-	2/6/23/26	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	301	NAG	O5-C1	6.34	1.53	1.43

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	B	301	NAG	C1-O5-C5	2.58	115.69	112.19
4	A	301	NAG	C1-O5-C5	2.32	115.33	112.19

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	B	301	NAG	O5-C5-C6-O6
4	A	301	NAG	O5-C5-C6-O6
4	B	301	NAG	C4-C5-C6-O6

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	301	NAG	1	0

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](#)

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

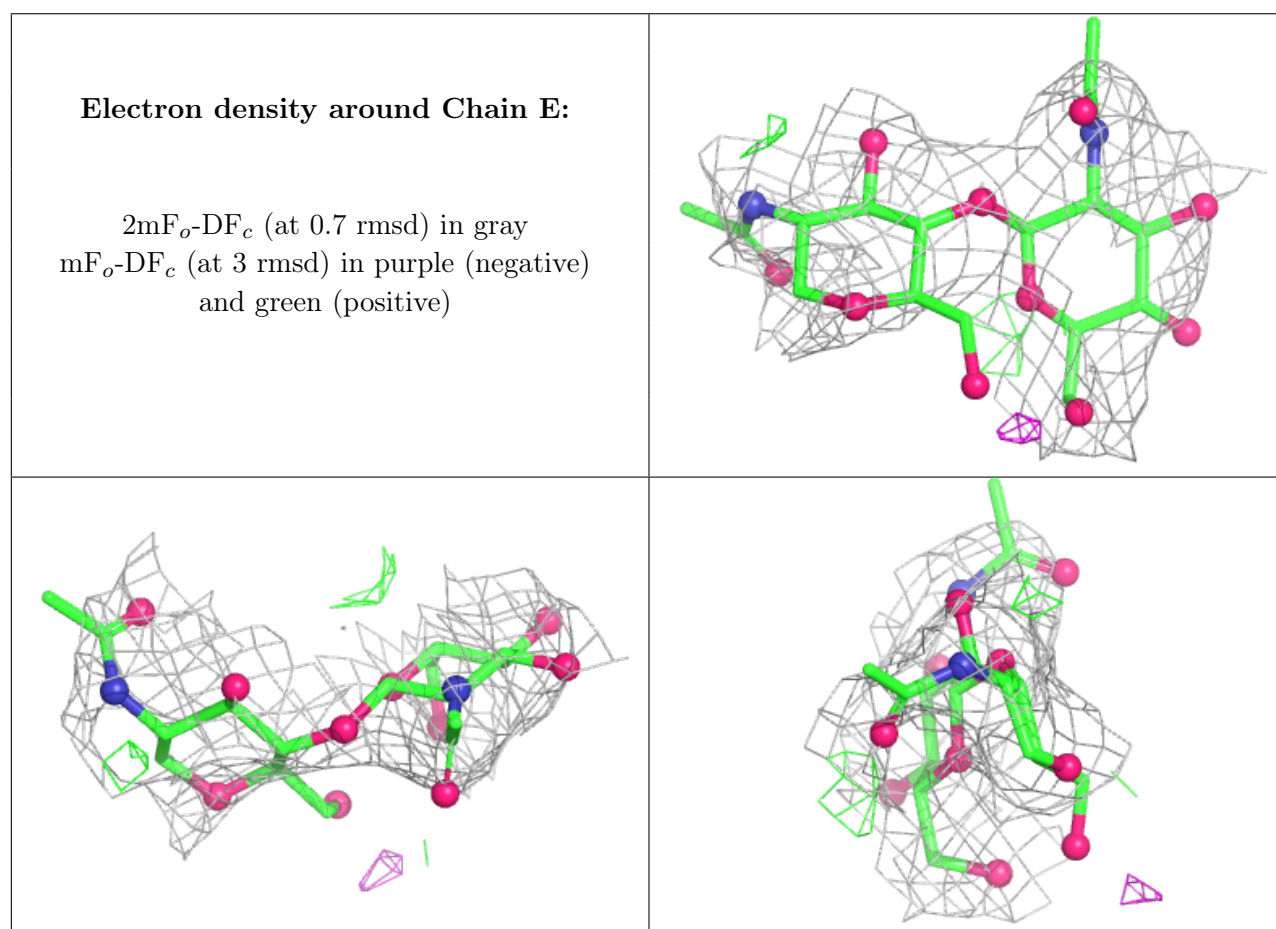
6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

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

6.3 Carbohydrates [i](#)

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.



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.