



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 8, 2020 – 01:40 AM BST

PDB ID : 4UIP
Title : The complex structure of extracellular domain of EGFR with Repebody (rAC1).
Authors : Kang, Y.J.; Cha, Y.J.; Cho, H.S.; Lee, J.J.; Kim, H.S.
Deposited on : 2015-03-31
Resolution : 2.95 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 following versions of software and data (see [references ⓘ](#)) 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

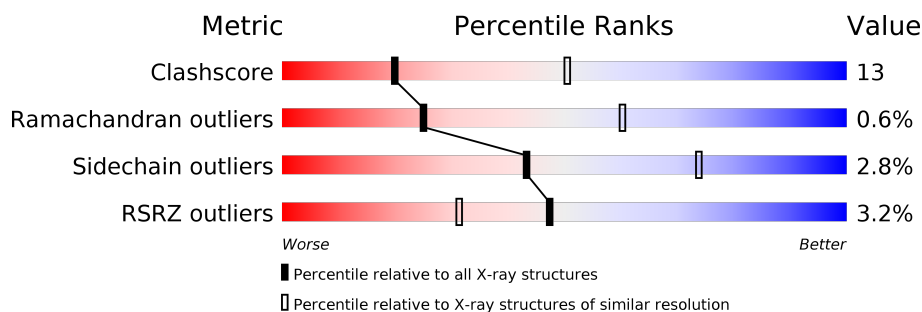
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.95 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	3462 (3.00-2.92)
Ramachandran outliers	138981	3340 (3.00-2.92)
Sidechain outliers	138945	3343 (3.00-2.92)
RSRZ outliers	127900	2986 (3.00-2.92)

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 $\leq 5\%$. 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	618	
2	B	251	
3	C	2	

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
4	NAG	A	1614	X	-	-	-

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 6680 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 EPIDERMAL GROWTH FACTOR RECEPTOR.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	612	Total	C	N	O	S	0	0	0
			4708	2904	839	905	60			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	614	HIS	-	expression tag	UNP P00533
A	615	HIS	-	expression tag	UNP P00533
A	616	HIS	-	expression tag	UNP P00533
A	617	HIS	-	expression tag	UNP P00533
A	618	HIS	-	expression tag	UNP P00533
A	619	HIS	-	expression tag	UNP P00533

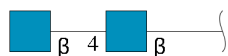
- Molecule 2 is a protein called REPEBODY (RAC1).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	238	Total	C	N	O	S	0	0	0
			1875	1190	321	359	5			

There are 10 discrepancies between the modelled and reference sequences:

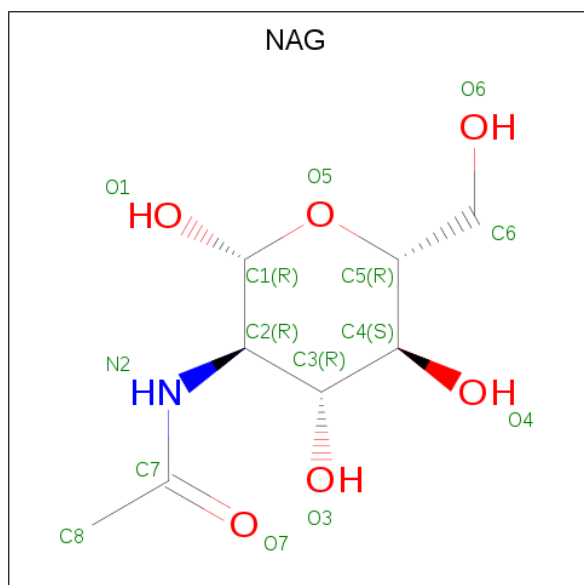
Chain	Residue	Modelled	Actual	Comment	Reference
B	1	MET	-	expression tag	UNP E0ACT6
B	25	ALA	ASP	conflict	UNP E0ACT6
B	244	LEU	-	expression tag	UNP Q4G1L3
B	245	GLU	-	expression tag	UNP Q4G1L3
B	246	HIS	-	expression tag	UNP Q4G1L3
B	247	HIS	-	expression tag	UNP Q4G1L3
B	248	HIS	-	expression tag	UNP Q4G1L3
B	249	HIS	-	expression tag	UNP Q4G1L3
B	250	HIS	-	expression tag	UNP Q4G1L3
B	251	HIS	-	expression tag	UNP Q4G1L3

- 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	C	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: $C_8H_{15}NO_6$).



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		

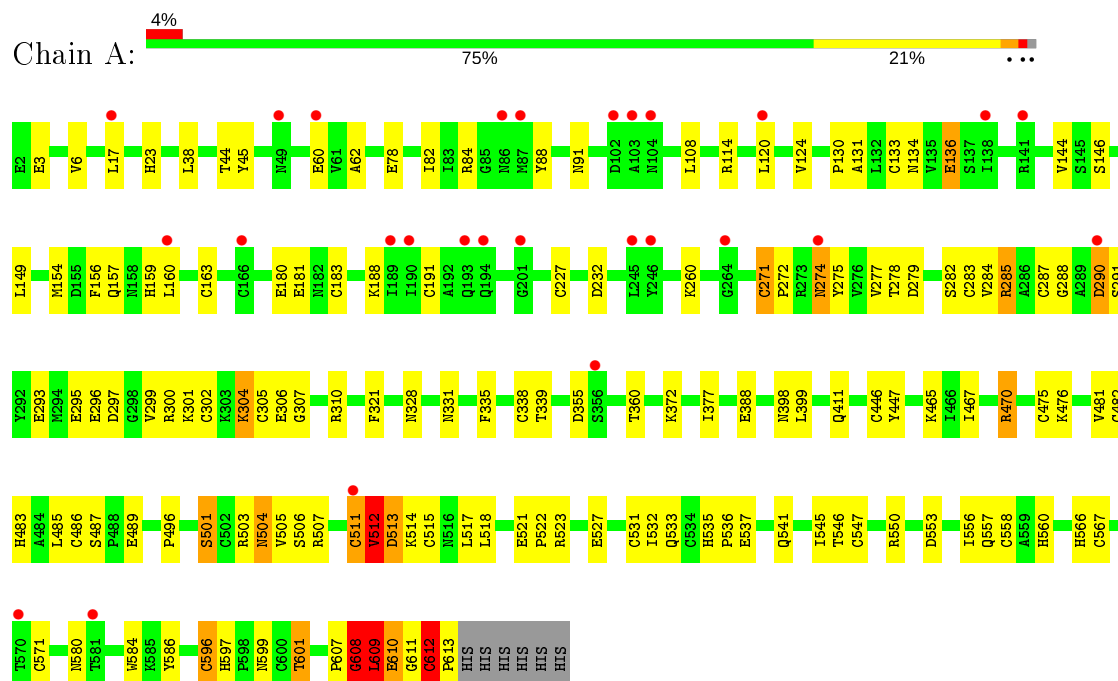
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	33	Total	O	0	0
			33	33		
5	B	8	Total	O	0	0
			8	8		

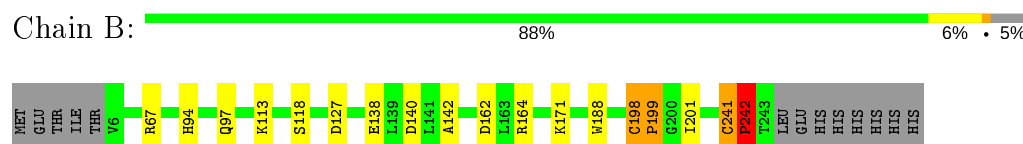
3 Residue-property plots

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 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: EPIDERMAL GROWTH FACTOR RECEPTOR



• Molecule 2: REPEBODY (RAC1)



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



4 Data and refinement statistics

Property	Value	Source
Space group	P 2 21 21	Depositor
Cell constants a, b, c, α , β , γ	66.79 Å 88.34 Å 189.21 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	53.28 – 2.95 53.28 – 2.95	Depositor EDS
% Data completeness (in resolution range)	99.8 (53.28-2.95) 99.9 (53.28-2.95)	Depositor EDS
R_{merge}	0.17	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.04 (at 2.96 Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, R_{free}	0.237 , 0.297 0.246 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å ²)	45.5	Xtriage
Anisotropy	0.053	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 38.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.88	EDS
Total number of atoms	6680	wwPDB-VP
Average B, all atoms (Å ²)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.89% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: NAG

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	0.75	1/4800 (0.0%)	0.83	14/6493 (0.2%)
2	B	0.67	1/1910 (0.1%)	0.74	4/2595 (0.2%)
All	All	0.73	2/6710 (0.0%)	0.80	18/9088 (0.2%)

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	A	0	5

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	388	GLU	CD-OE1	5.33	1.31	1.25
2	B	199	PRO	N-CD	5.00	1.54	1.47

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	164	ARG	NE-CZ-NH2	6.92	123.76	120.30
2	B	241	CYS	C-N-CD	6.69	142.45	128.40
1	A	596	CYS	CA-CB-SG	-6.62	102.08	114.00
1	A	571	CYS	CA-CB-SG	-6.45	102.39	114.00
1	A	547	CYS	CA-CB-SG	6.42	125.55	114.00

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	274	ASN	Peptide
1	A	290	ASP	Peptide
1	A	512	VAL	Peptide
1	A	521	GLU	Peptide
1	A	609	LEU	Peptide

5.2 Too-close contacts

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	4708	0	4536	146	0
2	B	1875	0	1899	21	0
3	C	28	0	25	1	0
4	A	28	0	26	0	0
5	A	33	0	0	4	0
5	B	8	0	0	3	0
All	All	6680	0	6486	164	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 13.

The worst 5 of 164 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:470:ARG:HG2	1:A:475:CYS:SG	1.44	1.53
1:A:133:CYS:HB2	1:A:163:CYS:SG	1.57	1.42
2:B:198:CYS:SG	2:B:241:CYS:CB	2.30	1.17
2:B:198:CYS:CB	2:B:241:CYS:SG	2.34	1.16
1:A:287:CYS:SG	1:A:302:CYS:SG	1.26	1.16

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	610/618 (99%)	568 (93%)	38 (6%)	4 (1%)	22	56
2	B	236/251 (94%)	223 (94%)	12 (5%)	1 (0%)	34	69
All	All	846/869 (97%)	791 (94%)	50 (6%)	5 (1%)	25	60

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	608	GLY
1	A	522	PRO
2	B	242	PRO
1	A	512	VAL
1	A	612	CYS

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	534/540 (99%)	517 (97%)	17 (3%)	39	71
2	B	215/228 (94%)	211 (98%)	4 (2%)	57	81
All	All	749/768 (98%)	728 (97%)	21 (3%)	43	74

5 of 21 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	501	SER

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	513	ASP
2	B	127	ASP
1	A	476	LYS
2	B	171	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	346	HIS
1	A	480	GLN
1	A	535	HIS
1	A	334	HIS
1	A	533	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 ⓘ

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	C	1	1,3	14,14,15	1.14	2 (14%)	17,19,21	1.10	0
3	NAG	C	2	3	14,14,15	0.55	0	17,19,21	1.38	4 (23%)

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	C	1	1,3	-	0/6/23/26	0/1/1/1
3	NAG	C	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	C	1	NAG	C1-C2	-2.93	1.48	1.52
3	C	1	NAG	O5-C1	-2.34	1.40	1.43

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	2	NAG	O4-C4-C5	2.58	115.70	109.30
3	C	2	NAG	C1-C2-N2	-2.35	106.47	110.49
3	C	2	NAG	C1-O5-C5	2.19	115.17	112.19
3	C	2	NAG	O5-C1-C2	-2.12	107.94	111.29

There are no chirality outliers.

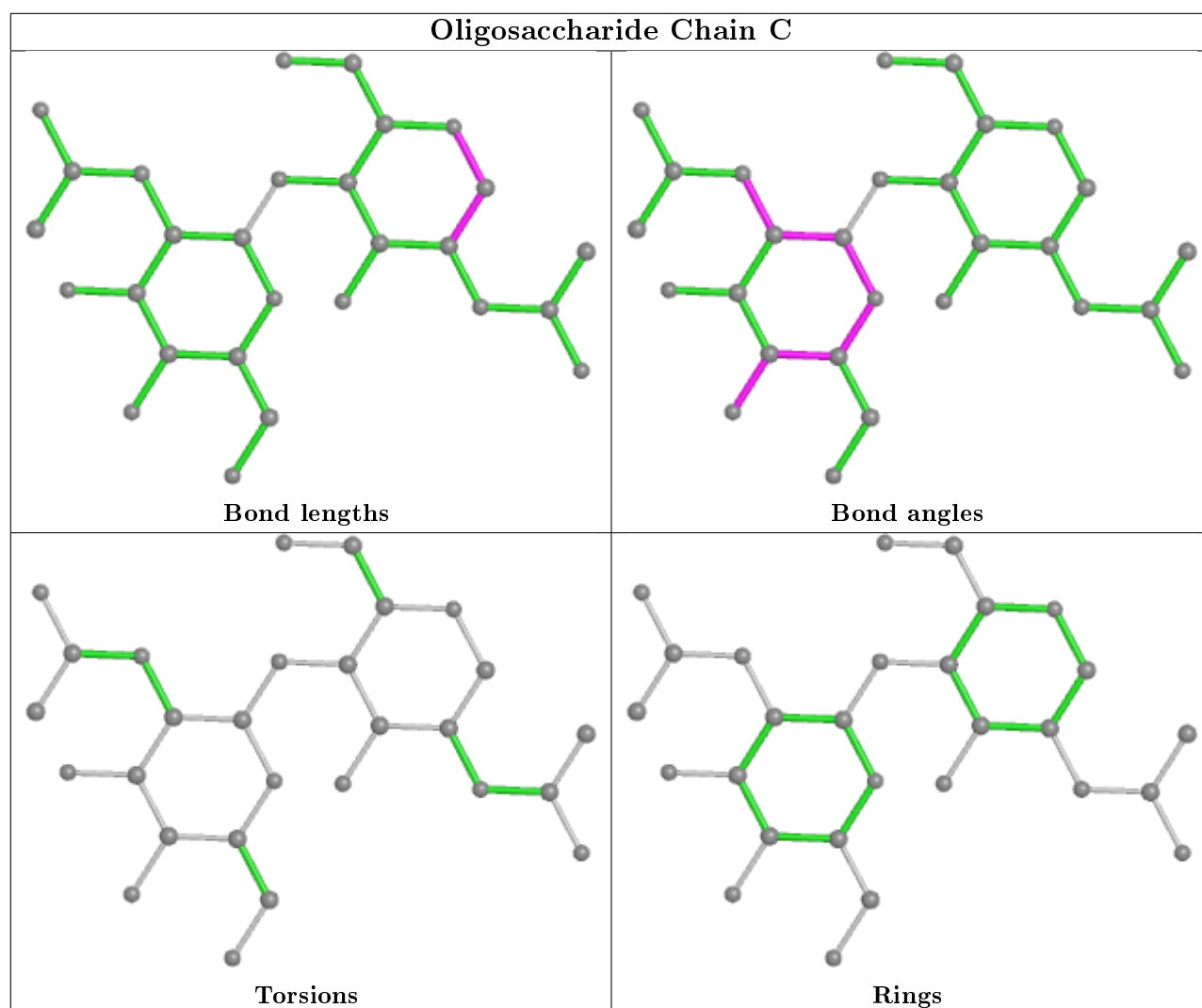
There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	C	1	NAG	1	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.



5.6 Ligand geometry [i](#)

2 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	$\# Z > 2$
4	NAG	A	1615	1	14,14,15	0.50	0	17,19,21	2.92	7 (41%)
4	NAG	A	1614	1	14,14,15	0.56	0	17,19,21	2.31	5 (29%)

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	1615	1	-	2/6/23/26	0/1/1/1
4	NAG	A	1614	1	1/1/5/7	0/6/23/26	0/1/1/1

There are no bond length outliers.

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	1615	NAG	C1-O5-C5	7.36	122.17	112.19
4	A	1615	NAG	C4-C3-C2	6.06	119.90	111.02
4	A	1614	NAG	C4-C3-C2	5.93	119.72	111.02
4	A	1615	NAG	O3-C3-C4	-3.84	101.46	110.35
4	A	1615	NAG	O5-C1-C2	-3.62	105.58	111.29

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
4	A	1614	NAG	C1

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	1615	NAG	C8-C7-N2-C2
4	A	1615	NAG	O7-C7-N2-C2

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](#)

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	612/618 (99%)	0.41	27 (4%) 34 21	16, 48, 74, 95	0
2	B	238/251 (94%)	-0.19	0 100 100	17, 29, 51, 76	0
All	All	850/869 (97%)	0.24	27 (3%) 47 31	16, 42, 72, 95	0

The worst 5 of 27 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	193	GLN	4.4
1	A	201	GLY	3.8
1	A	17	LEU	3.6
1	A	104	ASN	3.4
1	A	103	ALA	3.4

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

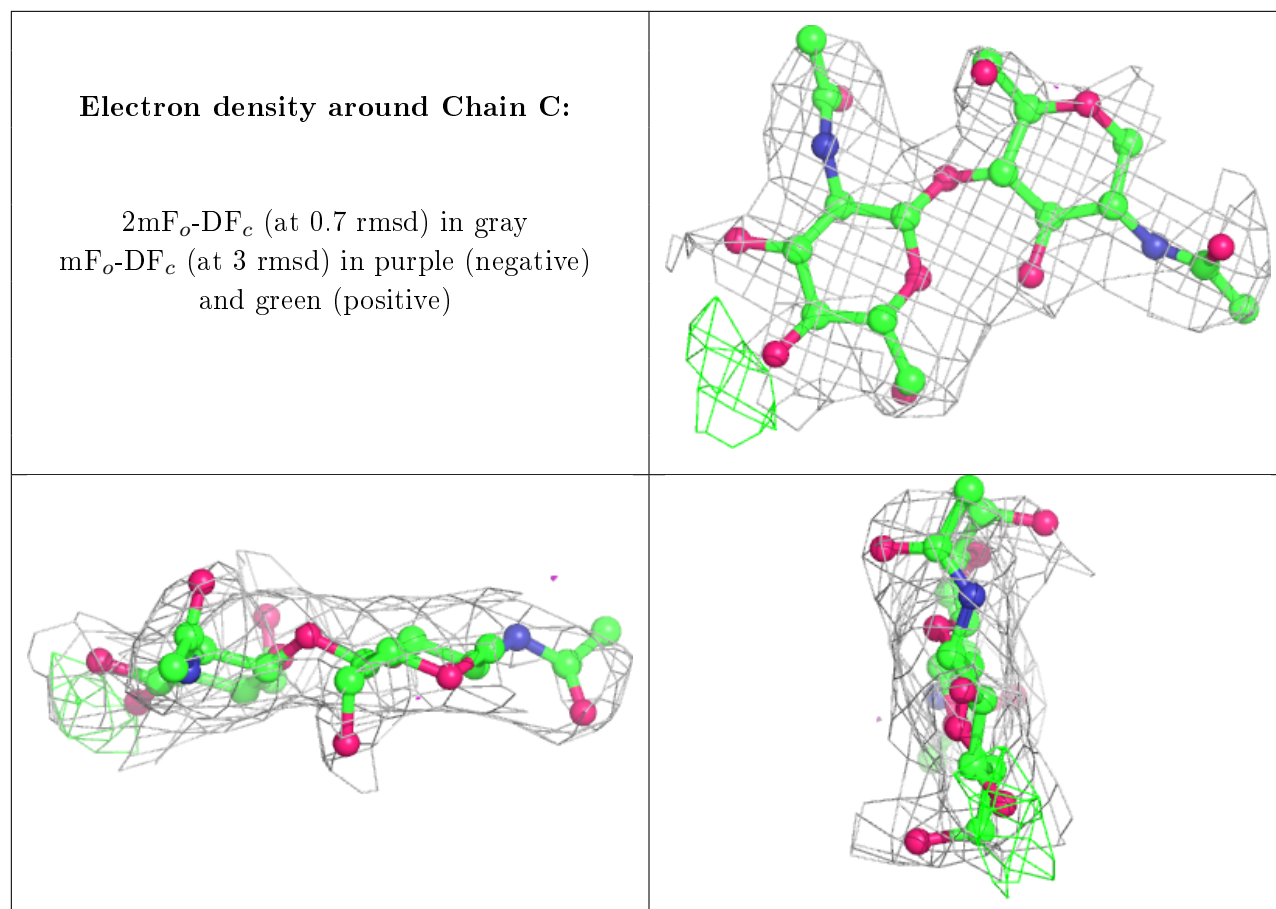
There are no non-standard protein/DNA/RNA residues in this entry.

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
3	NAG	C	2	14/15	0.80	0.26	49,56,63,64	0
3	NAG	C	1	14/15	0.93	0.31	49,55,61,64	0

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 [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	NAG	A	1615	14/15	0.70	0.23	57,64,66,67	0
4	NAG	A	1614	14/15	0.83	0.27	46,56,74,86	0

6.5 Other polymers [i](#)

There are no such residues in this entry.