W O R L D W I D E
wwPDB
P R O T E I N D A T A B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Oct 31, 2021 – 03:02 PM EDT

PDB ID	:	1VFO
Title	:	Crystal structure of Thermoactinomyces vulgaris R-47 alpha-amylase 2/beta-cyclodextrin complex
Authors	:	Ohtaki, A.; Mizuno, M.; Tonozuka, T.; Sakano, Y.; Kamitori, S.
Deposited on	:	2004-04-16
Resolution	:	2.81 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

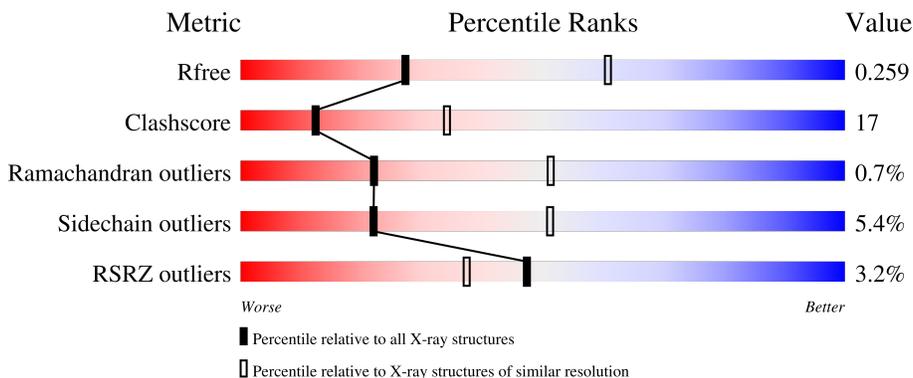|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 2.81 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	Percentile Ranks	Value
Rfree		0.259
Clashscore		17
Ramachandran outliers		0.7%
Sidechain outliers		5.4%
RSRZ outliers		3.2%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3617 (2.84-2.80)
Clashscore	141614	4060 (2.84-2.80)
Ramachandran outliers	138981	3978 (2.84-2.80)
Sidechain outliers	138945	3980 (2.84-2.80)
RSRZ outliers	127900	3552 (2.84-2.80)

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 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 <=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	585	2% — 69% / 28% / •
1	B	585	5% — 65% / 31% / •
2	C	7	57% / 29% / 14%
3	D	7	43% / 57%

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 9936 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 Neopullulanase 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	585	Total	C	N	O	S	0	0	0
			4776	3056	833	872	15			
1	B	585	Total	C	N	O	S	0	0	0
			4776	3056	833	872	15			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	325	ASN	ASP	engineered mutation	UNP Q08751
A	421	ASN	ASP	engineered mutation	UNP Q08751
B	325	ASN	ASP	engineered mutation	UNP Q08751
B	421	ASN	ASP	engineered mutation	UNP Q08751

- Molecule 2 is an oligosaccharide called Cycloheptakis-(1-4)-(alpha-D-glucopyranose).
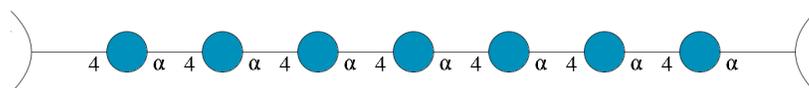
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace
2	C	7	Total	C	O	0	0	0
			77	42	35			

- Molecule 3 is an oligosaccharide called Cyclic alpha-D-glucopyranose-(1-4)-beta-D-glucopyranose-(1-4)-alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose.
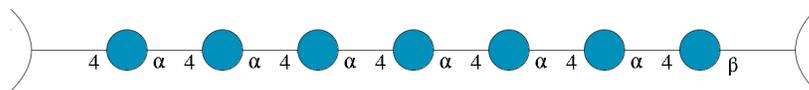
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace
3	D	7	Total 77	C 42	O 35	0	0	0

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total 1	Ca 1	0	0
4	B	1	Total 1	Ca 1	0	0

- Molecule 5 is water.

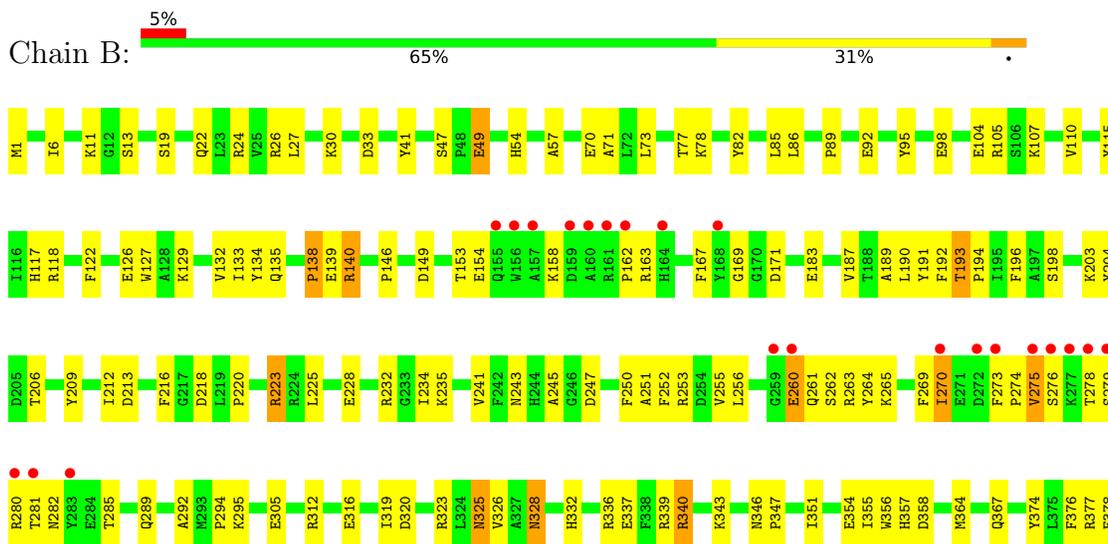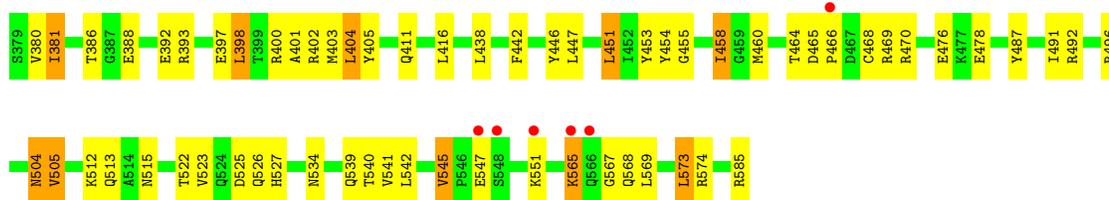| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	125	Total 125	O 125	0	0
5	B	103	Total 103	O 103	0	0

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: Neopullulanase 2

Chain A:

• Molecule 1: Neopullulanase 2

Chain B:

S379 V380 I381 T386 G387 E388 E392 R393 E397 L398 T399 R400 A401 R402 M403 L404 Y405 Q411 L416 L438 F442 Y446 L447 L451 I452 Y453 Y454 G455 I458 G459 M460 T464 D465 P466 D467 C468 R469 R470 E476 K477 E478 Y487 I491 R492 R496

N504 V505 K512 Q513 A514 N515 T522 V523 Q524 D525 Q526 H527 N534 Q539 T540 V541 L542 V545 P546 E547 S548 K551 K565 Q566 Q567 Q568 L569 L573 R574 R585
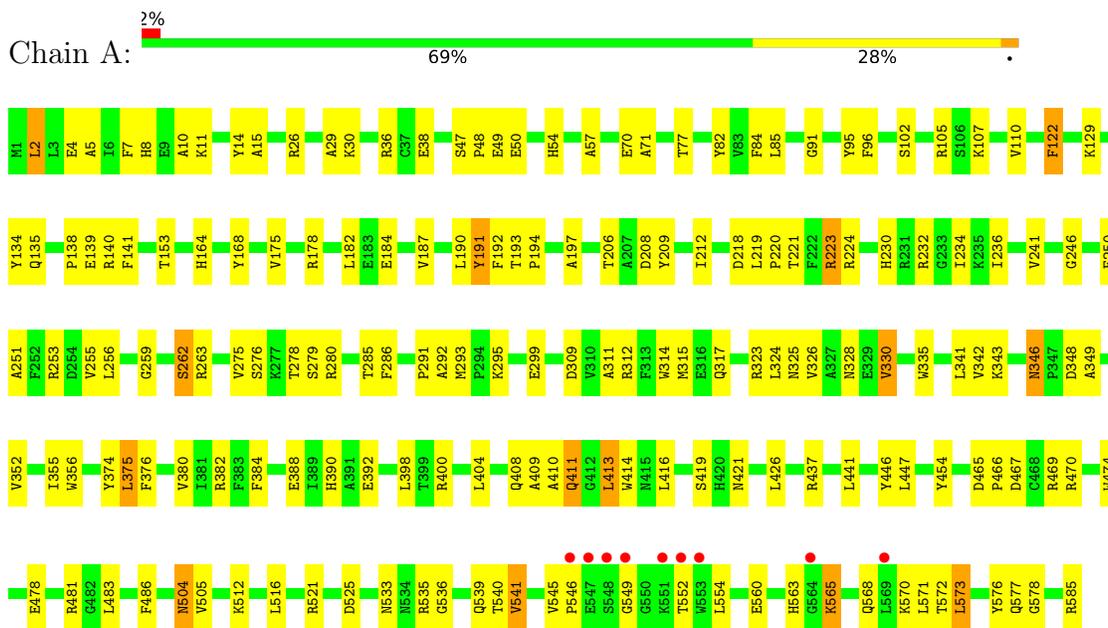
• Molecule 2: Cycloheptakis-(1-4)-(alpha-D-glucopyranose)

Chain C: 57%   29%   14%

GLC1 GLC2 GLC3 GLC4 GLC5 GLC6 GLC7

• Molecule 3: Cyclic alpha-D-glucopyranose-(1-4)-beta-D-glucopyranose-(1-4)-alpha-D-glucopyra nose-(1-4)-alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose-(1-4)-alpha-D-glucopyranose-(1-4) -alpha-D-glucopyranose

Chain D: 43%   57%

GLC1 GLC2 GLC3 GLC4 GLC5 BGC6 GLC7

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	115.08Å 118.81Å 114.28Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	38.87 – 2.81 38.87 – 2.81	Depositor EDS
% Data completeness (in resolution range)	99.3 (38.87-2.81) 99.4 (38.87-2.81)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.45 (at 2.81Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.200 , 0.256 0.208 , 0.259	Depositor DCC
R_{free} test set	3864 reflections (10.02%)	wwPDB-VP
Wilson B-factor (Å^2)	45.7	Xtriage
Anisotropy	0.564	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 64.4	EDS
L-test for twinning[2]	$< |L| > = 0.48, < L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.023 for k,h,-l 0.023 for -h,-l,-k 0.023 for l,-k,h 0.009 for k,l,h 0.009 for l,h,k	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	9936	wwPDB-VP
Average B, all atoms (Å^2)	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.22% 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: CA, BGC, 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	0.39	0/4906	0.64	1/6641 (0.0%)
1	B	0.38	0/4906	0.61	0/6641
All	All	0.39	0/9812	0.63	1/13282 (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	A	0	1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	323	ARG	N-CA-C	-5.06	97.34	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	14	TYR	Sidechain

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	4776	0	4611	129	0
1	B	4776	0	4611	195	0
2	C	77	0	63	3	0
3	D	77	0	63	2	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
5	A	125	0	0	9	0
5	B	103	0	0	10	0
All	All	9936	0	9348	322	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 17.

The worst 5 of 322 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:328:ASN:HB3	1:A:355:ILE:HD12	1.37	1.02
1:B:138:PRO:HD2	1:B:193:THR:HG22	1.50	0.91
1:B:512:LYS:HE3	1:B:513:GLN:HE22	1.37	0.90
1:A:278:THR:HG23	1:A:280:ARG:H	1.36	0.89
1:B:476:GLU:HB3	5:B:1052:HOH:O	1.73	0.87

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	583/585 (100%)	534 (92%)	47 (8%)	2 (0%)	41	70
1	B	583/585 (100%)	522 (90%)	55 (9%)	6 (1%)	15	42
All	All	1166/1170 (100%)	1056 (91%)	102 (9%)	8 (1%)	22	51

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	2	LEU
1	B	11	LYS
1	B	162	PRO
1	B	193	THR
1	B	464	THR

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	493/493 (100%)	469 (95%)	24 (5%)	25	55
1	B	493/493 (100%)	464 (94%)	29 (6%)	19	47
All	All	986/986 (100%)	933 (95%)	53 (5%)	22	51

5 of 53 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	122	PHE
1	B	325	ASN
1	B	539	GLN
1	B	140	ARG
1	B	260	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 31 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	509	HIS
1	B	513	GLN
1	B	135	GLN
1	B	534	ASN
1	B	357	HIS

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 ⓘ

14 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	C	1	2	11,11,12	0.45	0	15,15,17	0.74	1 (6%)
2	GLC	C	2	2	11,11,12	0.59	0	15,15,17	0.74	1 (6%)
2	GLC	C	3	2	11,11,12	0.77	0	15,15,17	0.72	0
2	GLC	C	4	2	11,11,12	0.60	0	15,15,17	0.64	0
2	GLC	C	5	2	11,11,12	0.57	0	15,15,17	0.65	0
2	GLC	C	6	2	11,11,12	0.44	0	15,15,17	0.60	0
2	GLC	C	7	2	11,11,12	0.56	0	15,15,17	0.63	0
3	GLC	D	1	3	11,11,12	0.63	0	15,15,17	0.61	0
3	GLC	D	2	3	11,11,12	0.59	0	15,15,17	0.70	0
3	GLC	D	3	3	11,11,12	0.55	0	15,15,17	0.57	0
3	GLC	D	4	3	11,11,12	0.58	0	15,15,17	0.68	1 (6%)
3	GLC	D	5	3	11,11,12	0.69	0	15,15,17	1.02	1 (6%)
3	BGC	D	6	3	11,11,12	0.66	0	15,15,17	0.78	0
3	GLC	D	7	3	11,11,12	0.59	0	15,15,17	0.67	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
2	GLC	C	1	2	-	0/2/19/22	0/1/1/1
2	GLC	C	2	2	-	0/2/19/22	0/1/1/1
2	GLC	C	3	2	-	0/2/19/22	0/1/1/1
2	GLC	C	4	2	-	0/2/19/22	0/1/1/1
2	GLC	C	5	2	-	2/2/19/22	0/1/1/1
2	GLC	C	6	2	-	2/2/19/22	0/1/1/1
2	GLC	C	7	2	-	2/2/19/22	0/1/1/1
3	GLC	D	1	3	-	0/2/19/22	0/1/1/1
3	GLC	D	2	3	-	2/2/19/22	0/1/1/1
3	GLC	D	3	3	-	0/2/19/22	0/1/1/1
3	GLC	D	4	3	-	2/2/19/22	0/1/1/1
3	GLC	D	5	3	-	2/2/19/22	0/1/1/1
3	BGC	D	6	3	-	2/2/19/22	0/1/1/1
3	GLC	D	7	3	-	2/2/19/22	0/1/1/1

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	D	5	GLC	C1-O5-C5	2.64	115.77	112.19
2	C	2	GLC	C1-O5-C5	2.25	115.24	112.19
3	D	4	GLC	C1-O5-C5	2.03	114.95	112.19
2	C	1	GLC	C1-O5-C5	2.00	114.91	112.19

There are no chirality outliers.

5 of 16 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	C	6	GLC	O5-C5-C6-O6
2	C	7	GLC	C4-C5-C6-O6
3	D	4	GLC	O5-C5-C6-O6
2	C	6	GLC	C4-C5-C6-O6
3	D	4	GLC	C4-C5-C6-O6

There are no ring outliers.

4 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	D	3	GLC	2	0
2	C	3	GLC	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	D	2	GLC	2	0
2	C	1	GLC	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.

Oligosaccharide Chain C

Bond lengths

Bond angles

Torsions

Rings

Oligosaccharide Chain D

Bond lengths

Bond angles

Torsions

Rings
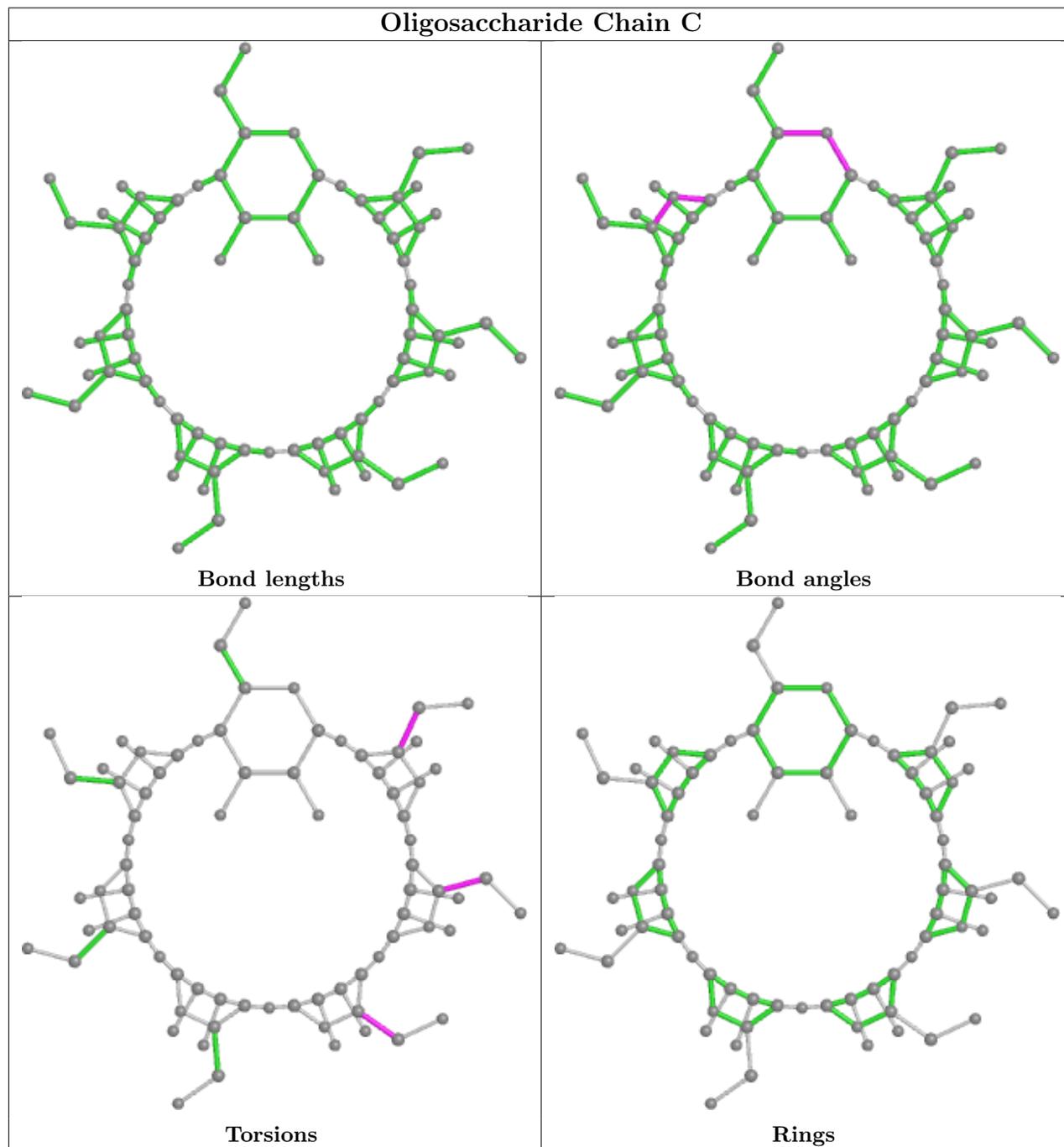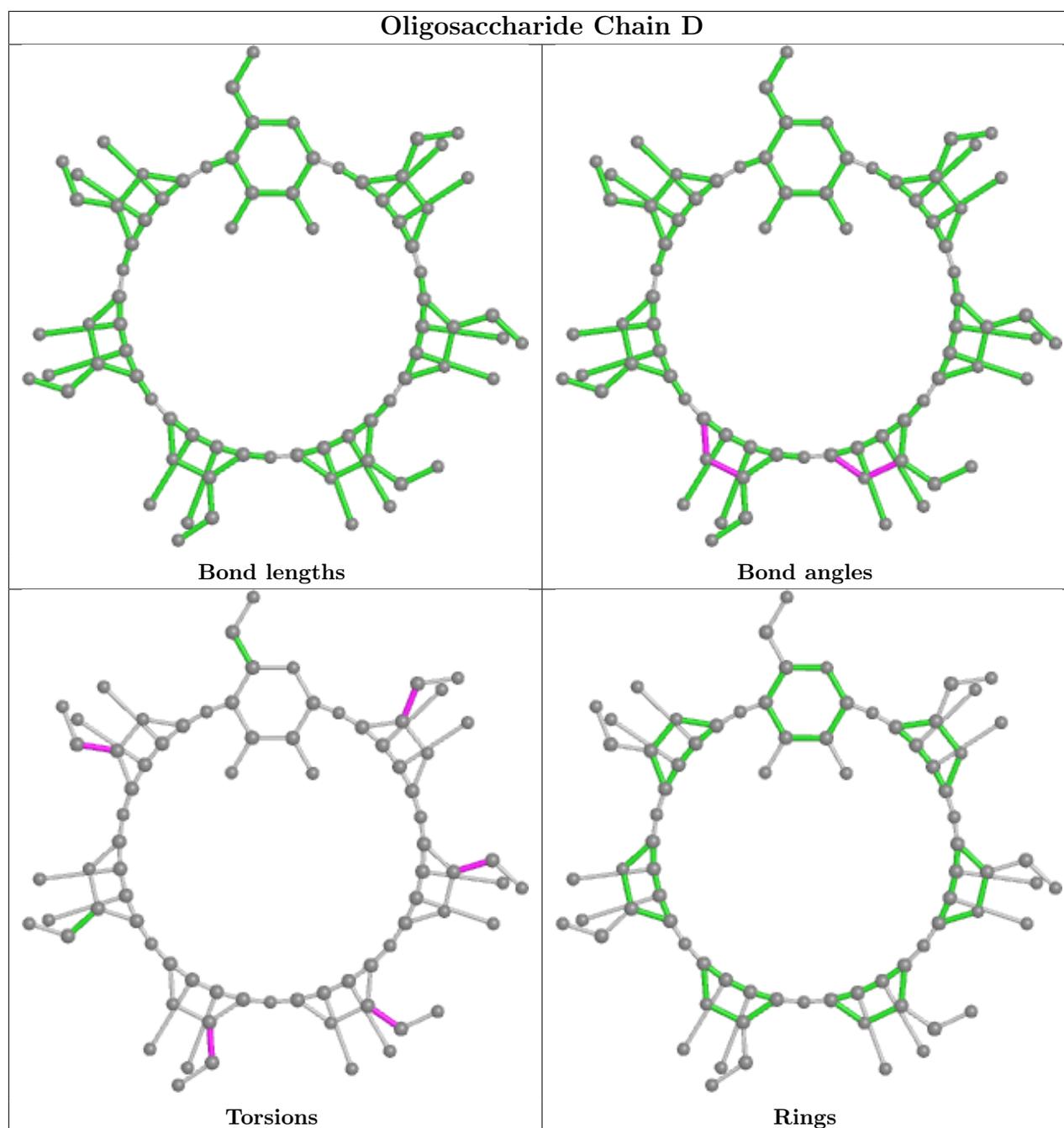
5.6 Ligand geometry ⓘ

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 ⓘ

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 ⓘ

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(Å²)	Q<0.9
1	A	585/585 (100%)	-0.34	9 (1%)	73	67	14, 38, 69, 85	0
1	B	585/585 (100%)	-0.05	28 (4%)	30	21	18, 45, 83, 100	0
All	All	1170/1170 (100%)	-0.20	37 (3%)	47	37	14, 41, 78, 100	0

The worst 5 of 37 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	273	PHE	6.2
1	B	275	VAL	5.0
1	B	279	SER	4.3
1	B	272	ASP	4.2
1	B	155	GLN	4.1

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

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

6.3 Carbohydrates ⓘ

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(Å²)	Q<0.9
3	GLC	D	5	11/12	0.76	0.31	96,97,98,98	0
3	GLC	D	3	11/12	0.82	0.23	91,93,94,97	0
3	GLC	D	4	11/12	0.84	0.31	97,98,99,99	0
3	BGC	D	6	11/12	0.84	0.29	95,97,98,98	0
3	GLC	D	7	11/12	0.87	0.29	92,95,96,96	0

*Continued on next page...*

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	GLC	D	2	11/12	0.88	0.21	90,92,93,93	0
3	GLC	D	1	11/12	0.88	0.20	93,94,94,95	0
2	GLC	C	7	11/12	0.93	0.17	44,46,48,48	0
2	GLC	C	2	11/12	0.94	0.17	46,47,49,50	0
2	GLC	C	3	11/12	0.95	0.15	40,45,47,49	0
2	GLC	C	4	11/12	0.95	0.15	37,38,39,40	0
2	GLC	C	1	11/12	0.96	0.14	45,46,48,50	0
2	GLC	C	5	11/12	0.96	0.13	44,47,49,51	0
2	GLC	C	6	11/12	0.97	0.13	49,49,51,51	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.

Electron density around Chain C:

$2mF_o\text{-}DF_c$ (at 0.7 rmsd) in gray
$mF_o\text{-}DF_c$ (at 3 rmsd) in purple (negative)
and green (positive)

Electron density around Chain D:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)
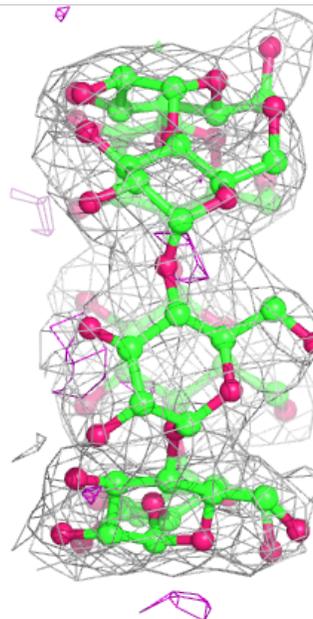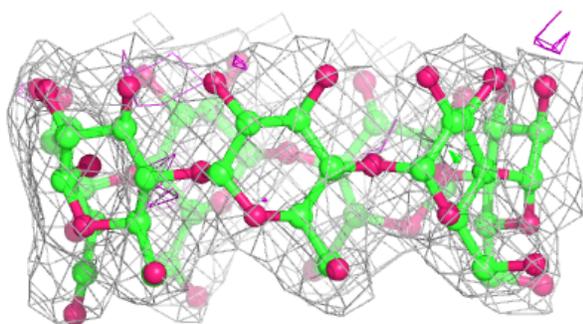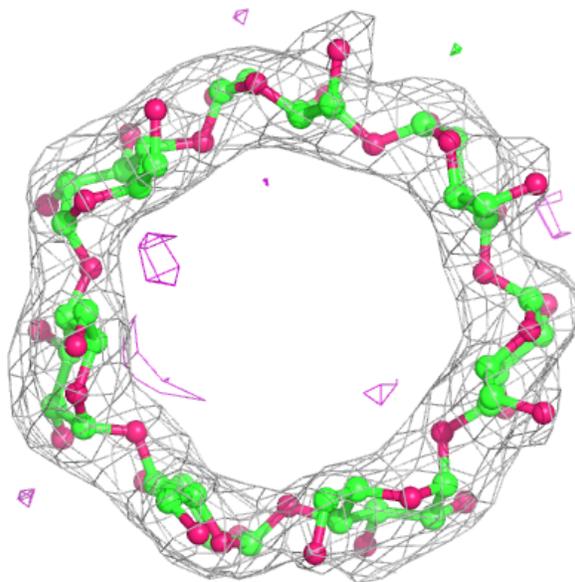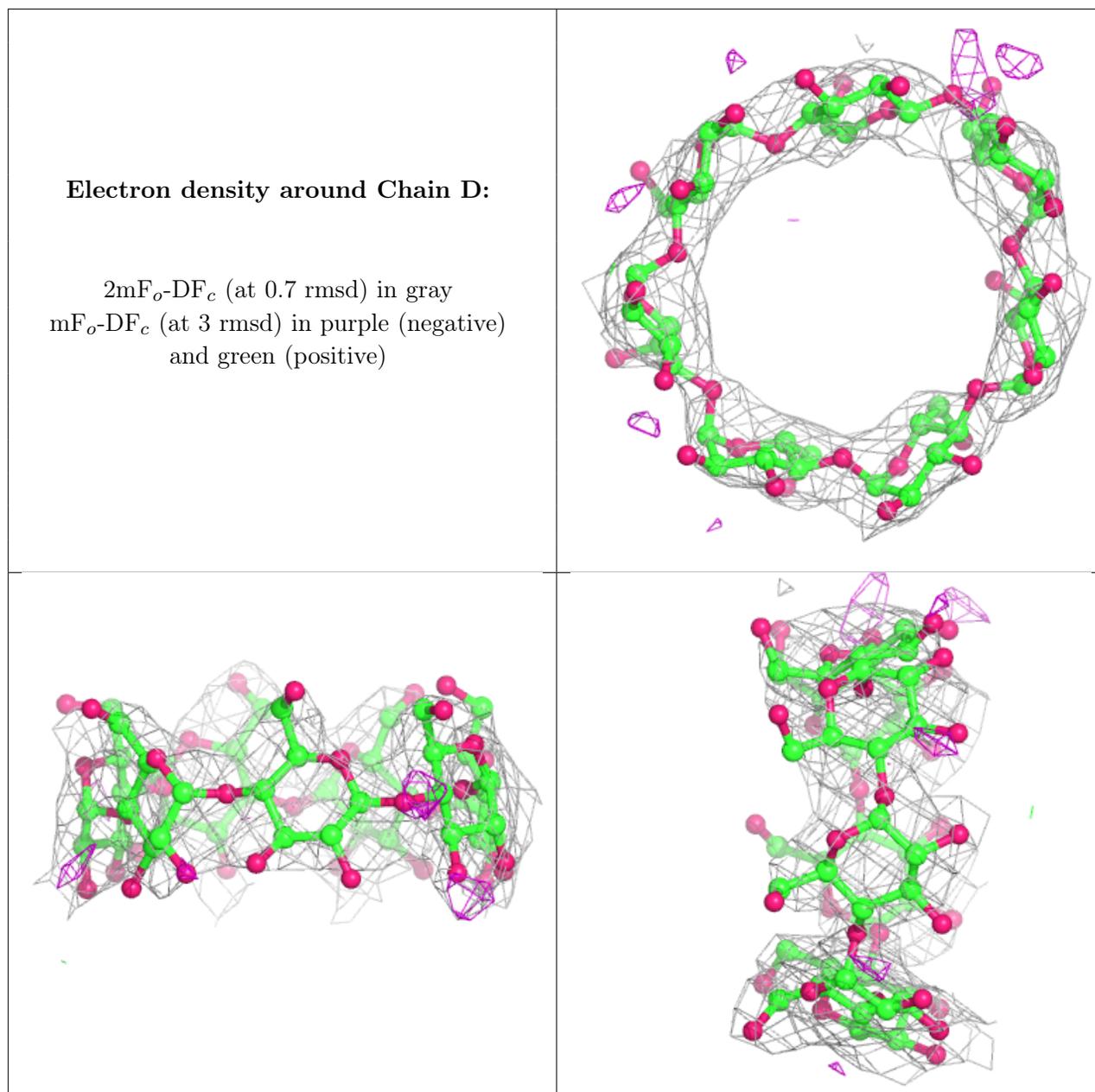
6.4 Ligands (i)

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
4	CA	A	1001	1/1	0.97	0.08	35,35,35,35	0
4	CA	B	1002	1/1	0.98	0.07	66,66,66,66	0

6.5 Other polymers (i)

There are no such residues in this entry.