



Full wwPDB X-ray Structure Validation Report ⓘ

Aug 27, 2020 – 01:51 PM BST

PDB ID : 6T2O
Title : Prominent members of the human gut microbiota express endo-acting O-glycanases to initiate mucin breakdown
Authors : Crouch, L.I.; Liberato, M.V.; Ubranowicz, P.A.; Basle, A.; Lamb, C.A.; Cooke, K.; Doona, M.; Needham, S.; Brady, R.R.; Berrington, J.E.; Madubic, K.; Chater, P.; Zhang, F.; Linhardt, R.J.; Spence, D.I.R.; Bolam, D.N.
Deposited on : 2019-10-09
Resolution : 2.05 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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
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

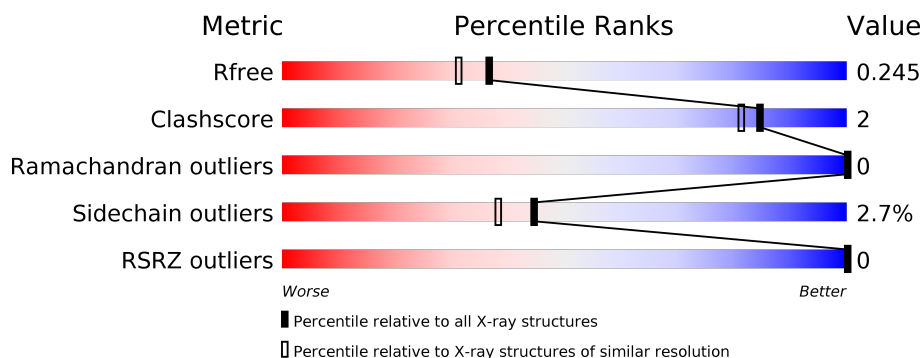
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.05 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	1692 (2.04-2.04)
Clashscore	141614	1773 (2.04-2.04)
Ramachandran outliers	138981	1752 (2.04-2.04)
Sidechain outliers	138945	1752 (2.04-2.04)
RSRZ outliers	127900	1672 (2.04-2.04)

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	AAA	282	 78% 7% 15%
1	BBB	282	 77% 8% 14%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 4071 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 Glycosyl hydrolase family 16.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAA	241	Total	C	N	O	S	0	0	0
			1955	1262	322	361	10			
1	BBB	242	Total	C	N	O	S	0	0	0
			1964	1267	323	364	10			

There are 46 discrepancies between the modelled and reference sequences:

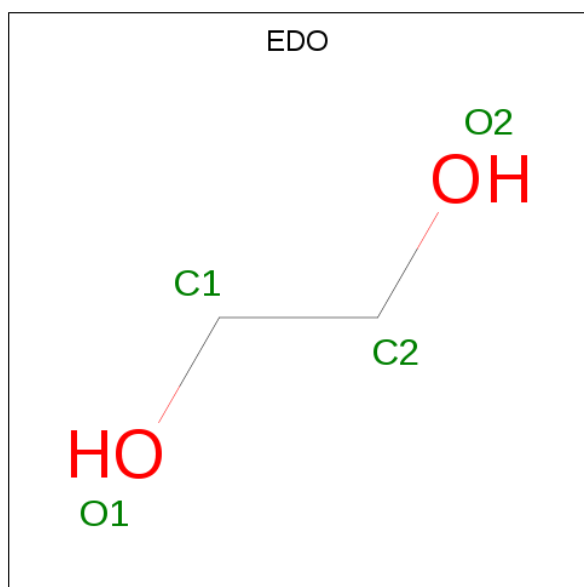
Chain	Residue	Modelled	Actual	Comment	Reference
AAA	-1	MET	-	initiating methionine	UNP A5ZIF0
AAA	0	GLY	-	expression tag	UNP A5ZIF0
AAA	1	SER	-	expression tag	UNP A5ZIF0
AAA	2	SER	-	expression tag	UNP A5ZIF0
AAA	3	HIS	-	expression tag	UNP A5ZIF0
AAA	4	HIS	-	expression tag	UNP A5ZIF0
AAA	5	HIS	-	expression tag	UNP A5ZIF0
AAA	6	HIS	-	expression tag	UNP A5ZIF0
AAA	7	HIS	-	expression tag	UNP A5ZIF0
AAA	8	HIS	-	expression tag	UNP A5ZIF0
AAA	9	SER	-	expression tag	UNP A5ZIF0
AAA	10	SER	-	expression tag	UNP A5ZIF0
AAA	11	GLY	-	expression tag	UNP A5ZIF0
AAA	12	LEU	-	expression tag	UNP A5ZIF0
AAA	13	VAL	-	expression tag	UNP A5ZIF0
AAA	14	PRO	-	expression tag	UNP A5ZIF0
AAA	15	ARG	-	expression tag	UNP A5ZIF0
AAA	16	GLY	-	expression tag	UNP A5ZIF0
AAA	17	SER	-	expression tag	UNP A5ZIF0
AAA	18	HIS	-	expression tag	UNP A5ZIF0
AAA	19	MET	-	expression tag	UNP A5ZIF0
AAA	20	ALA	-	expression tag	UNP A5ZIF0
AAA	21	SER	-	expression tag	UNP A5ZIF0
BBB	-1	MET	-	initiating methionine	UNP A5ZIF0
BBB	0	GLY	-	expression tag	UNP A5ZIF0

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
BBB	1	SER	-	expression tag	UNP A5ZIF0
BBB	2	SER	-	expression tag	UNP A5ZIF0
BBB	3	HIS	-	expression tag	UNP A5ZIF0
BBB	4	HIS	-	expression tag	UNP A5ZIF0
BBB	5	HIS	-	expression tag	UNP A5ZIF0
BBB	6	HIS	-	expression tag	UNP A5ZIF0
BBB	7	HIS	-	expression tag	UNP A5ZIF0
BBB	8	HIS	-	expression tag	UNP A5ZIF0
BBB	9	SER	-	expression tag	UNP A5ZIF0
BBB	10	SER	-	expression tag	UNP A5ZIF0
BBB	11	GLY	-	expression tag	UNP A5ZIF0
BBB	12	LEU	-	expression tag	UNP A5ZIF0
BBB	13	VAL	-	expression tag	UNP A5ZIF0
BBB	14	PRO	-	expression tag	UNP A5ZIF0
BBB	15	ARG	-	expression tag	UNP A5ZIF0
BBB	16	GLY	-	expression tag	UNP A5ZIF0
BBB	17	SER	-	expression tag	UNP A5ZIF0
BBB	18	HIS	-	expression tag	UNP A5ZIF0
BBB	19	MET	-	expression tag	UNP A5ZIF0
BBB	20	ALA	-	expression tag	UNP A5ZIF0
BBB	21	SER	-	expression tag	UNP A5ZIF0

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



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	AAA	1	Total	C	O	0	0
			4	2	2		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	AAA	1	Total	C	O	0	0
			4	2	2		
2	BBB	1	Total	C	O	0	0
			4	2	2		
2	BBB	1	Total	C	O	0	0
			4	2	2		

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	BBB	1	Total	Ca	0	0
			1	1		
3	AAA	1	Total	Ca	0	0
			1	1		


- Molecule 4 is water.

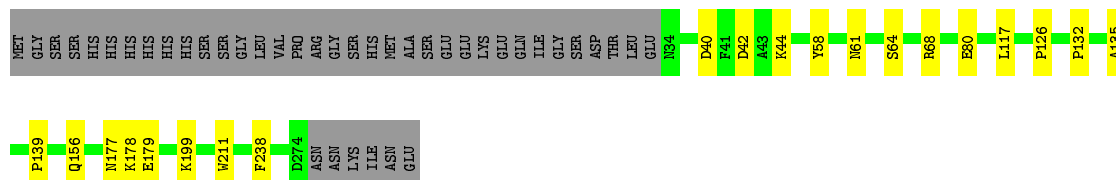
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	AAA	70	Total	O	0	0
			70	70		
4	BBB	64	Total	O	0	0
			64	64		

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.

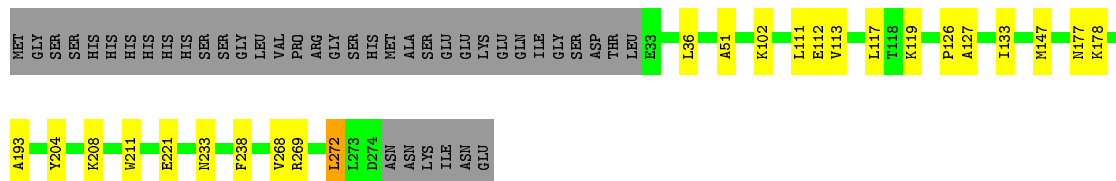
- Molecule 1: Glycosyl hydrolase family 16

Chain AAA: 



- Molecule 1: Glycosyl hydrolase family 16

Chain BBB: 



4 Data and refinement statistics

Property	Value	Source
Space group	P 41	Depositor
Cell constants a, b, c, α , β , γ	82.65Å 82.65Å 121.49Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	82.65 – 2.05 82.65 – 2.05	Depositor EDS
% Data completeness (in resolution range)	99.9 (82.65-2.05) 99.9 (82.65-2.05)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.69 (at 2.05Å)	Xtriage
Refinement program	REFMAC 5.8.0257	Depositor
R, R_{free}	0.199 , 0.240 0.206 , 0.245	Depositor DCC
R_{free} test set	2613 reflections (5.12%)	wwPDB-VP
Wilson B-factor (Å ²)	38.3	Xtriage
Anisotropy	0.112	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 31.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.034 for h,-k,-l	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	4071	wwPDB-VP
Average B, all atoms (Å ²)	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.66% 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: CA, EDO

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	AAA	0.69	0/2011	0.85	0/2728
1	BBB	0.66	0/2020	0.84	0/2740
All	All	0.68	0/4031	0.84	0/5468

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	AAA	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	AAA	156	GLN	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	AAA	1955	0	1876	8	0
1	BBB	1964	0	1882	10	0
2	AAA	8	0	12	0	0
2	BBB	8	0	12	0	0
3	AAA	1	0	0	0	0
3	BBB	1	0	0	0	0
4	AAA	70	0	0	1	0
4	BBB	64	0	0	0	0
All	All	4071	0	3782	18	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 2.

All (18) 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:AAA:40:ASP:HB2	1:AAA:42:ASP:OD1	2.04	0.58
1:BBB:112:GLU:OE1	1:BBB:269:ARG:NH1	2.43	0.52
1:BBB:111:LEU:O	1:BBB:193:ALA:HA	2.12	0.49
1:BBB:177:ASN:HB2	1:BBB:211:TRP:CD1	2.48	0.48
1:BBB:36:LEU:HD22	1:BBB:272:LEU:HD13	1.96	0.48
1:BBB:117:LEU:HD11	1:BBB:126:PRO:HG3	1.94	0.48
1:AAA:117:LEU:HD11	1:AAA:126:PRO:HG3	1.94	0.47
1:BBB:193:ALA:HB3	1:BBB:204:TYR:HB2	1.99	0.45
1:BBB:113:VAL:HG22	1:BBB:268:VAL:HG22	1.99	0.44
1:AAA:61:ASN:OD1	1:AAA:64:SER:HB2	2.17	0.43
1:AAA:132:PRO:HB2	1:AAA:135:ALA:HB2	2.01	0.43
1:AAA:177:ASN:HB2	1:AAA:211:TRP:CD1	2.55	0.42
1:AAA:179:GLU:HA	1:AAA:179:GLU:OE1	2.19	0.42
1:AAA:58:TYR:CD1	1:AAA:68:ARG:HB2	2.55	0.42
1:BBB:133:ILE:HD11	1:BBB:233:ASN:HB3	2.02	0.41
1:AAA:199:LYS:NZ	4:AAA:404:HOH:O	2.48	0.41
1:BBB:51:ALA:O	1:BBB:102:LYS:HE3	2.20	0.40
1:BBB:127:ALA:HA	1:BBB:147:MET:O	2.21	0.40

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	AAA	239/282 (85%)	233 (98%)	6 (2%)	0	100	100
1	BBB	240/282 (85%)	236 (98%)	4 (2%)	0	100	100
All	All	479/564 (85%)	469 (98%)	10 (2%)	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	AAA	207/243 (85%)	202 (98%)	5 (2%)	49	42
1	BBB	208/243 (86%)	202 (97%)	6 (3%)	42	35
All	All	415/486 (85%)	404 (97%)	11 (3%)	44	38

All (11) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	AAA	44	LYS
1	AAA	80	GLU
1	AAA	139	PRO
1	AAA	178	LYS
1	AAA	238	PHE
1	BBB	119	LYS
1	BBB	178	LYS
1	BBB	208	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	BBB	221	GLU
1	BBB	238	PHE
1	BBB	272	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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](#)

There are no monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 6 ligands modelled in this entry, 2 are monoatomic - leaving 4 for Mogul analysis.

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	EDO	AAA	302	-	3,3,3	0.08	0	2,2,2	0.20	0
2	EDO	BBB	301	-	3,3,3	0.21	0	2,2,2	0.14	0
2	EDO	BBB	302	-	3,3,3	0.36	0	2,2,2	0.10	0
2	EDO	AAA	301	-	3,3,3	0.15	0	2,2,2	0.12	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	EDO	AAA	302	-	-	1/1/1/1	-
2	EDO	BBB	301	-	-	1/1/1/1	-
2	EDO	BBB	302	-	-	1/1/1/1	-
2	EDO	AAA	301	-	-	0/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	AAA	302	EDO	O1-C1-C2-O2
2	BBB	301	EDO	O1-C1-C2-O2
2	BBB	302	EDO	O1-C1-C2-O2

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	AAA	241/282 (85%)	-0.10	0 100 100	28, 41, 62, 73	0
1	BBB	242/282 (85%)	-0.18	0 100 100	29, 43, 63, 79	0
All	All	483/564 (85%)	-0.14	0 100 100	28, 42, 63, 79	0

There are no RSRZ outliers to report.

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](#)

There are no monosaccharides in this entry.

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
2	EDO	BBB	301	4/4	0.81	0.14	49,50,51,54	0
2	EDO	AAA	301	4/4	0.93	0.13	42,42,46,48	0
2	EDO	BBB	302	4/4	0.94	0.10	43,46,47,49	0
2	EDO	AAA	302	4/4	0.95	0.14	50,52,52,53	0
3	CA	AAA	303	1/1	0.97	0.06	50,50,50,50	0
3	CA	BBB	303	1/1	0.98	0.06	52,52,52,52	0

6.5 Other polymers [i](#)

There are no such residues in this entry.