Full wwPDB X-ray Structure Validation Report ⓘ

Jan 16, 2024 – 02:59 am GMT

PDB ID	:	6T2P
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	:	2.10 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 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:

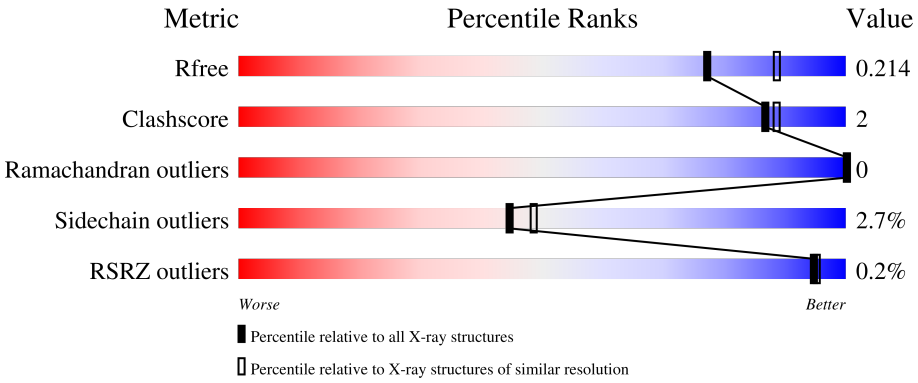
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.10 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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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	AAA	282	80% 5% 15%
1	BBB	282	78% 7% 14%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 4083 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	323	360	10			
1	BBB	242	Total	C	N	O	S	0	0	0
			1964	1267	324	363	10			

There are 48 discrepancies between the modelled and reference sequences:

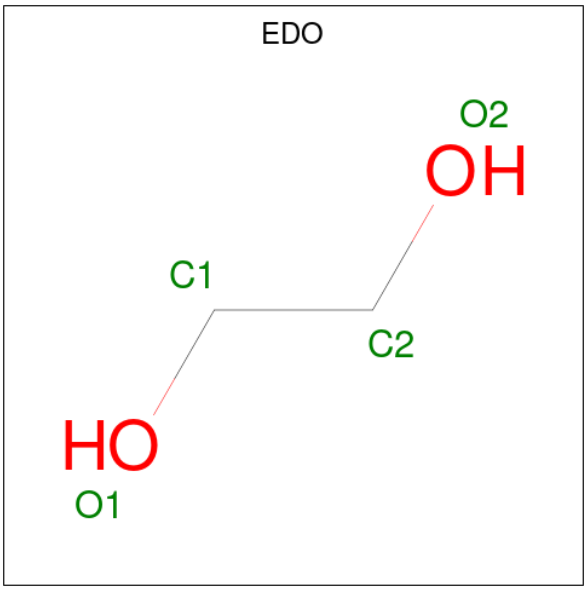
Chain	Residue	Modelled	Actual	Comment	Reference
AAA	-1	MET	-	initiating methionine	UNP A5ZIF0
AAA	0	GLY	-	expression tag	UNP A5ZIF0
AAA	1	SER	-	expression tag	UNP A5ZIF0
AAA	2	SER	-	expression tag	UNP A5ZIF0
AAA	3	HIS	-	expression tag	UNP A5ZIF0
AAA	4	HIS	-	expression tag	UNP A5ZIF0
AAA	5	HIS	-	expression 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	THR	conflict	UNP A5ZIF0
AAA	18	HIS	TYR	conflict	UNP A5ZIF0
AAA	19	MET	VAL	conflict	UNP A5ZIF0
AAA	20	ALA	CYS	conflict	UNP A5ZIF0
AAA	21	SER	ALA	conflict	UNP A5ZIF0
AAA	143	GLN	GLU	conflict	UNP A5ZIF0
BBB	-1	MET	-	initiating methionine	UNP A5ZIF0

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
BBB	0	GLY	-	expression tag	UNP A5ZIF0
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	THR	conflict	UNP A5ZIF0
BBB	18	HIS	TYR	conflict	UNP A5ZIF0
BBB	19	MET	VAL	conflict	UNP A5ZIF0
BBB	20	ALA	CYS	conflict	UNP A5ZIF0
BBB	21	SER	ALA	conflict	UNP A5ZIF0
BBB	143	GLN	GLU	conflict	UNP A5ZIF0

- Molecule 2 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



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

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

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

- Molecule 4 is water.

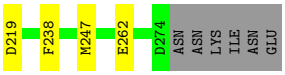
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	AAA	81	Total O 81 81	0	0
4	BBB	61	Total O 61 61	0	0

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: Glycosyl hydrolase family 16

Chain AAA:



- Molecule 1: Glycosyl hydrolase family 16

Chain BBB:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41	Depositor
Cell constants a, b, c, α, β, γ	82.72Å 82.72Å 121.68Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	82.72 – 2.10 82.72 – 2.10	Depositor EDS
% Data completeness (in resolution range)	99.6 (82.72-2.10) 99.6 (82.72-2.10)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.77 (at 2.10Å)	Xtriage
Refinement program	REFMAC 5.8.0257	Depositor
R, R_{free}	0.186 , 0.211 0.193 , 0.214	Depositor DCC
R_{free} test set	2369 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å^2)	34.6	Xtriage
Anisotropy	0.066	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 29.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.035 for h,-k,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	4083	wwPDB-VP
Average B, all atoms (Å^2)	39.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.68% 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: EDO, CA

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	0.68	0/2011	0.89	0/2728
1	BBB	0.70	1/2020 (0.0%)	0.85	0/2740
All	All	0.69	1/4031 (0.0%)	0.87	0/5468

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	BBB	224	GLN	CG-CD	5.01	1.62	1.51

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	AAA	1955	0	1878	6	0
1	BBB	1964	0	1884	10	0
2	AAA	12	0	18	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	81	0	0	1	0
4	BBB	61	0	0	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	4083	0	3792	16	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 (16) 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:BBB:156:GLN:NE2	4:BBB:402:HOH:O	2.30	0.61
1:BBB:112:GLU:OE1	1:BBB:269:ARG:NH1	2.34	0.60
1:BBB:177:ASN:HB2	1:BBB:211:TRP:CD1	2.38	0.59
1:AAA:219:ASP:HB3	4:AAA:475:HOH:O	2.02	0.58
1:BBB:179:GLU:N	1:BBB:179:GLU:OE1	2.43	0.52
1:BBB:111:LEU:O	1:BBB:193:ALA:HA	2.11	0.50
1:AAA:111:LEU:O	1:AAA:193:ALA:HA	2.13	0.48
1:BBB:117:LEU:HD11	1:BBB:126:PRO:HG3	1.98	0.44
1:AAA:85:LEU:O	1:AAA:262:GLU:HA	2.18	0.43
1:BBB:139:PRO:HG2	1:BBB:164:PHE:HA	1.99	0.43
1:BBB:143:GLN:HB3	1:BBB:162:HIS:HB2	2.00	0.43
1:BBB:135:ALA:HB1	1:BBB:136:PRO:HD2	2.01	0.43
1:BBB:85:LEU:O	1:BBB:262:GLU:HA	2.21	0.40
1:AAA:138:TRP:HA	1:AAA:139:PRO:HA	1.90	0.40
1:AAA:157:ILE:HD12	1:AAA:157:ILE:C	2.42	0.40
1:AAA:61:ASN:OD1	1:AAA:64:SER:HB2	2.21	0.40

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	AAA	239/282 (85%)	234 (98%)	5 (2%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	BBB	240/282 (85%)	234 (98%)	6 (2%)	0	100 100
All	All	479/564 (85%)	468 (98%)	11 (2%)	0	100 100

There are no Ramachandran outliers to report.

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	AAA	207/243 (85%)	200 (97%)	7 (3%)	37 39
1	BBB	208/243 (86%)	204 (98%)	4 (2%)	57 63
All	All	415/486 (85%)	404 (97%)	11 (3%)	44 48

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	91	ASN
1	AAA	139	PRO
1	AAA	178	LYS
1	AAA	238	PHE
1	AAA	247	MET
1	BBB	156	GLN
1	BBB	208	LYS
1	BBB	238	PHE
1	BBB	272	LEU

Sometimes 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 7 ligands modelled in this entry, 2 are monoatomic - leaving 5 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	0.18	0
2	EDO	AAA	303	-	3,3,3	0.14	0	2,2,2	0.35	0
2	EDO	AAA	301	-	3,3,3	0.27	0	2,2,2	0.09	0
2	EDO	BBB	301	-	3,3,3	0.20	0	2,2,2	0.23	0
2	EDO	BBB	302	-	3,3,3	0.23	0	2,2,2	0.48	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	AAA	303	-	-	0/1/1/1	-
2	EDO	AAA	301	-	-	0/1/1/1	-
2	EDO	BBB	301	-	-	0/1/1/1	-
2	EDO	BBB	302	-	-	0/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	AAA	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

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, 95^{th} 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%)	0.18	0 100 100	24, 36, 58, 71	0
1	BBB	242/282 (85%)	0.16	1 (0%) 92 93	26, 39, 60, 78	0
All	All	483/564 (85%)	0.17	1 (0%) 95 95	24, 37, 59, 78	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	BBB	179	GLU	2.4

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

There are no monosaccharides in this entry.

6.4 Ligands

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, 95^{th} 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	AAA	302	4/4	0.94	0.12	46,50,52,53	0
2	EDO	BBB	301	4/4	0.95	0.16	33,33,33,33	0
3	CA	AAA	304	1/1	0.96	0.07	48,48,48,48	0
2	EDO	AAA	301	4/4	0.97	0.12	32,33,33,33	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	EDO	BBB	302	4/4	0.97	0.16	32,33,33,33	0
2	EDO	AAA	303	4/4	0.97	0.18	33,33,34,35	0
3	CA	BBB	303	1/1	0.99	0.09	48,48,48,48	0

6.5 Other polymers (i)

There are no such residues in this entry.