WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Apr 21, 2024 – 08:18 am BST

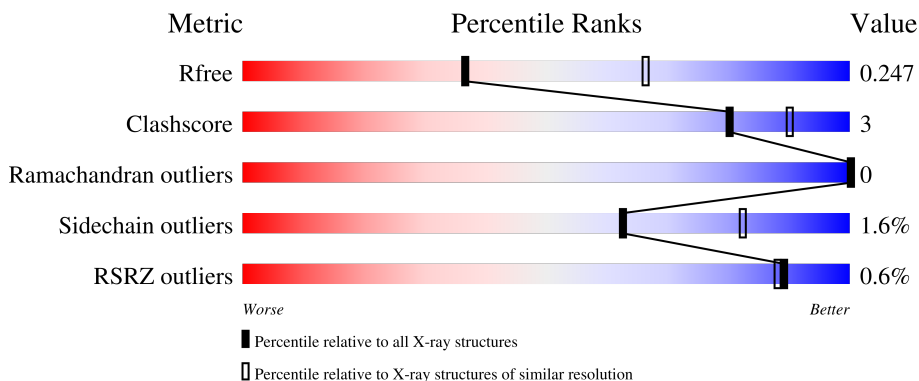|		
PDB ID	:	7ANB
Title	:	A single sulfatase is required for metabolism of colonic mucin O-glycans and intestinal colonization by a symbiotic human gut bacterium (BT1622-S1_20)
Authors	:	Sofia de Jesus Vaz Luis, A.; Basle, A.; Martens, E.C.; Cartmell, A.
Deposited on	:	2020-10-11
Resolution	:	2.60 Å(reported)

---

1 Overall quality at a glance ⓘ

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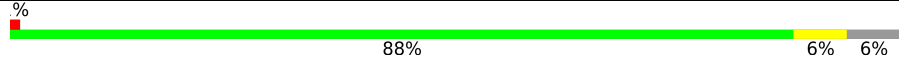*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.60 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		0.247
Clashscore		3
Ramachandran outliers		0
Sidechain outliers		1.6%
RSRZ outliers		0.6%

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	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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	503	88% 6% 6%
1	BBB	503	87% 7% 6%

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
3	EDO	BBB	602	-	-	X	-

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 7623 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 N-acetylgalactosamine-6-sulfatase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAA	475	Total	C	N	O	S	0	0	0
			3785	2412	646	705	22			
1	BBB	475	Total	C	N	O	S	0	0	0
			3785	2412	646	705	22			

There are 46 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	-2	MET	-	initiating methionine	UNP Q8A7A3
AAA	-1	GLY	-	expression tag	UNP Q8A7A3
AAA	0	SER	-	expression tag	UNP Q8A7A3
AAA	1	SER	-	expression tag	UNP Q8A7A3
AAA	2	HIS	-	expression tag	UNP Q8A7A3
AAA	3	HIS	-	expression tag	UNP Q8A7A3
AAA	4	HIS	-	expression tag	UNP Q8A7A3
AAA	5	HIS	-	expression tag	UNP Q8A7A3
AAA	6	HIS	-	expression tag	UNP Q8A7A3
AAA	7	HIS	-	expression tag	UNP Q8A7A3
AAA	8	SER	-	expression tag	UNP Q8A7A3
AAA	9	SER	-	expression tag	UNP Q8A7A3
AAA	10	GLY	-	expression tag	UNP Q8A7A3
AAA	11	LEU	-	expression tag	UNP Q8A7A3
AAA	12	VAL	-	expression tag	UNP Q8A7A3
AAA	13	PRO	-	expression tag	UNP Q8A7A3
AAA	14	ARG	-	expression tag	UNP Q8A7A3
AAA	15	GLY	-	expression tag	UNP Q8A7A3
AAA	16	SER	-	expression tag	UNP Q8A7A3
AAA	17	HIS	-	expression tag	UNP Q8A7A3
AAA	18	MET	-	expression tag	UNP Q8A7A3
AAA	19	ALA	-	expression tag	UNP Q8A7A3
AAA	20	SER	-	expression tag	UNP Q8A7A3
BBB	-2	MET	-	initiating methionine	UNP Q8A7A3
BBB	-1	GLY	-	expression tag	UNP Q8A7A3

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
BBB	0	SER	-	expression tag	UNP Q8A7A3
BBB	1	SER	-	expression tag	UNP Q8A7A3
BBB	2	HIS	-	expression tag	UNP Q8A7A3
BBB	3	HIS	-	expression tag	UNP Q8A7A3
BBB	4	HIS	-	expression tag	UNP Q8A7A3
BBB	5	HIS	-	expression tag	UNP Q8A7A3
BBB	6	HIS	-	expression tag	UNP Q8A7A3
BBB	7	HIS	-	expression tag	UNP Q8A7A3
BBB	8	SER	-	expression tag	UNP Q8A7A3
BBB	9	SER	-	expression tag	UNP Q8A7A3
BBB	10	GLY	-	expression tag	UNP Q8A7A3
BBB	11	LEU	-	expression tag	UNP Q8A7A3
BBB	12	VAL	-	expression tag	UNP Q8A7A3
BBB	13	PRO	-	expression tag	UNP Q8A7A3
BBB	14	ARG	-	expression tag	UNP Q8A7A3
BBB	15	GLY	-	expression tag	UNP Q8A7A3
BBB	16	SER	-	expression tag	UNP Q8A7A3
BBB	17	HIS	-	expression tag	UNP Q8A7A3
BBB	18	MET	-	expression tag	UNP Q8A7A3
BBB	19	ALA	-	expression tag	UNP Q8A7A3
BBB	20	SER	-	expression tag	UNP Q8A7A3

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).
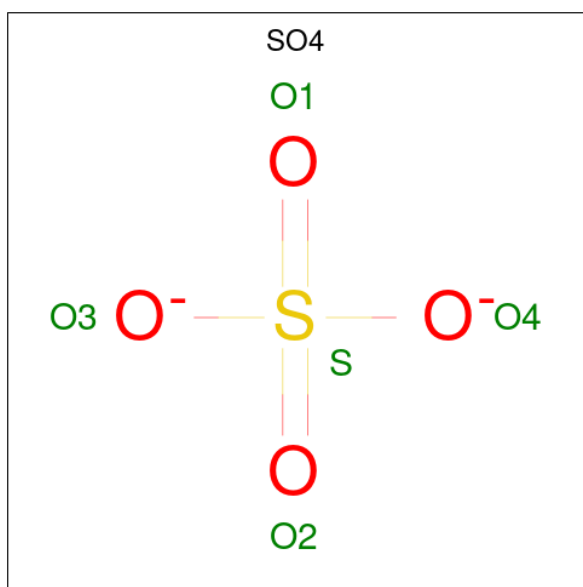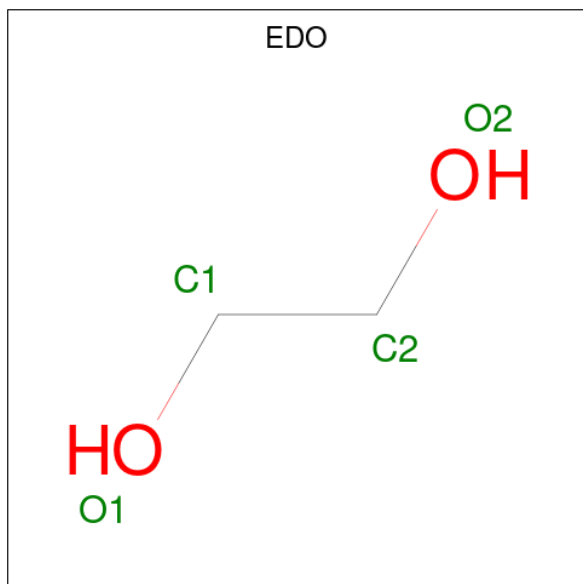

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	AAA	1	Total	O	S	0	0
			5	4	1		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	BBB	1	Total 5	O 4	S 1	0	0

- Molecule 3 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).

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

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

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

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	AAA	9	Total 9	O 9	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	BBB	20	Total 20	O 20	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.
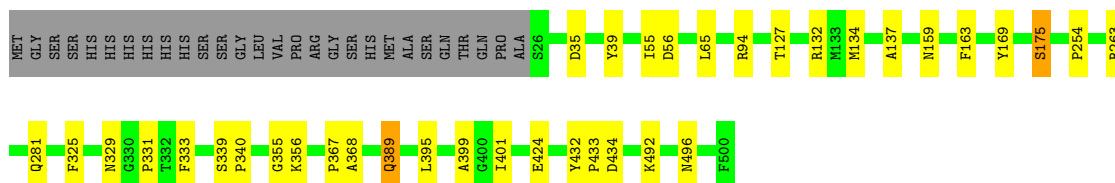
• Molecule 1: N-acetylgalactosamine-6-sulfatase

Chain AAA:
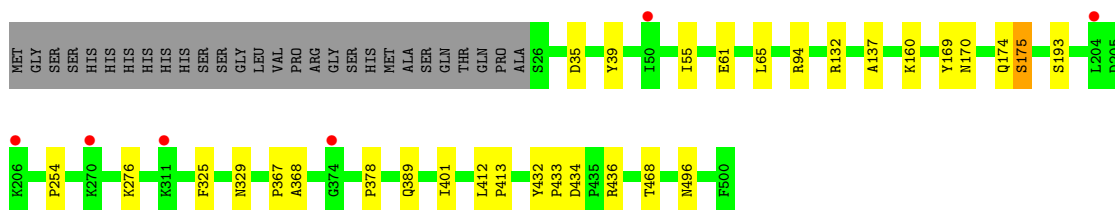
• Molecule 1: N-acetylgalactosamine-6-sulfatase

Chain BBB:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	81.27Å 81.27Å 383.74Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	95.93 – 2.60 95.93 – 2.60	Depositor EDS
% Data completeness (in resolution range)	97.6 (95.93-2.60) 97.7 (95.93-2.60)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.60 (at 2.62Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.219 , 0.245 0.222 , 0.247	Depositor DCC
R_{free} test set	1975 reflections (4.94%)	wwPDB-VP
Wilson B-factor (Å^2)	47.4	Xtriage
Anisotropy	0.490	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 37.1	EDS
L-test for twinning[2]	$< |L| > = 0.48, < L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	7623	wwPDB-VP
Average B, all atoms (Å^2)	61.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.51% 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: SO4, 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	0.64	0/3899	0.74	0/5289
1	BBB	0.63	0/3899	0.74	0/5289
All	All	0.63	0/7798	0.74	0/10578

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 ⓘ

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	3785	0	3612	17	0
1	BBB	3785	0	3612	24	0
2	AAA	5	0	0	0	0
2	BBB	5	0	0	0	0
3	AAA	8	0	12	3	0
3	BBB	4	0	6	5	0
4	AAA	1	0	0	0	0
4	BBB	1	0	0	0	0
5	AAA	9	0	0	0	0
5	BBB	20	0	0	0	0
All	All	7623	0	7242	40	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 3.

The worst 5 of 40 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:174:GLN:HG3	3:AAA:603:EDO:H22	1.47	0.96
1:BBB:333:PHE:CD1	3:BBB:602:EDO:H21	2.28	0.68
1:BBB:399:ALA:HB3	1:BBB:401:ILE:HD12	1.83	0.60
1:BBB:333:PHE:HD1	3:BBB:602:EDO:H21	1.66	0.56
1:AAA:137:ALA:CB	1:AAA:401:ILE:CD1	2.85	0.54

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	473/503 (94%)	454 (96%)	19 (4%)	0	100	100
1	BBB	473/503 (94%)	454 (96%)	19 (4%)	0	100	100
All	All	946/1006 (94%)	908 (96%)	38 (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	AAA	400/425 (94%)	394 (98%)	6 (2%)	65	83
1	BBB	400/425 (94%)	393 (98%)	7 (2%)	59	80
All	All	800/850 (94%)	787 (98%)	13 (2%)	62	82

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	BBB	263	ARG
1	BBB	281	GLN
1	BBB	492	LYS
1	BBB	389	GLN
1	BBB	424	GLU

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

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 ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

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	$\#\|Z\| > 2$
3	EDO	AAA	603	-	3,3,3	0.12	0	2,2,2	0.42	0
3	EDO	BBB	602	-	3,3,3	0.15	0	2,2,2	0.27	0
2	SO4	AAA	601	4	4,4,4	0.42	0	6,6,6	0.16	0
2	SO4	BBB	601	4	4,4,4	0.38	0	6,6,6	0.18	0
3	EDO	AAA	602	-	3,3,3	0.07	0	2,2,2	0.27	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
3	EDO	AAA	603	-	-	0/1/1/1	-
3	EDO	BBB	602	-	-	1/1/1/1	-
3	EDO	AAA	602	-	-	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
3	BBB	602	EDO	O1-C1-C2-O2

There are no ring outliers.

2 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	AAA	603	EDO	3	0
3	BBB	602	EDO	5	0

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(Å^2)	Q<0.9
1	AAA	475/503 (94%)	0.25	6 (1%)	77	73	44, 65, 98, 113	0
1	BBB	475/503 (94%)	0.13	0	100	100	39, 53, 71, 81	0
All	All	950/1006 (94%)	0.19	6 (0%)	89	88	39, 58, 90, 113	0

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AAA	270	LYS	2.7
1	AAA	204	LEU	2.5
1	AAA	206	LYS	2.5
1	AAA	50	ILE	2.2
1	AAA	374	GLY	2.1

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 ⓘ

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, 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($\mathring{A}^2$)	Q<0.9
3	EDO	BBB	602	4/4	0.92	0.34	54,54,55,55	0
3	EDO	AAA	603	4/4	0.93	0.59	63,64,64,64	0
3	EDO	AAA	602	4/4	0.93	0.20	53,53,53,53	0
2	SO4	BBB	601	5/5	0.98	0.19	49,49,50,51	0
2	SO4	AAA	601	5/5	0.98	0.16	52,52,52,54	0
4	CA	AAA	604	1/1	0.98	0.10	49,49,49,49	0
4	CA	BBB	603	1/1	0.98	0.13	39,39,39,39	0

6.5 Other polymers ⓘ

There are no such residues in this entry.