



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 24, 2021 – 01:08 pm GMT

PDB ID : 7B0M
Title : Sugar transaminase from a metagenome collected from troll oil field production water
Authors : Littlechild, J.A.; De Rose, S.A.; Isupov, M.N.; Sayer, C.; Karki, S.
Deposited on : 2020-11-20
Resolution : 1.43 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 : 1.8.4 (270009), 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 : 5.8.0267
CCP4 : 7.1.010 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.23.2

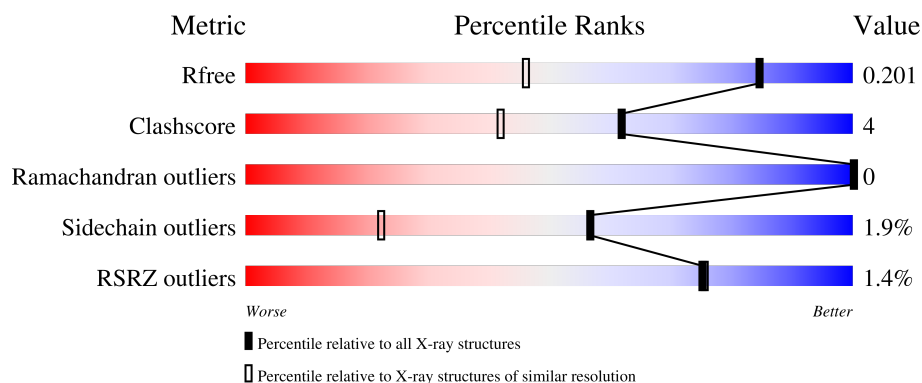
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.43 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	2021 (1.46-1.42)
Clashscore	141614	2086 (1.46-1.42)
Ramachandran outliers	138981	2047 (1.46-1.42)
Sidechain outliers	138945	2047 (1.46-1.42)
RSRZ outliers	127900	1993 (1.46-1.42)

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 $\leq 5\%$. 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	375	<div> <div></div> <div>88%</div> <div>9%</div> <div>..</div> </div>

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 6691 atoms, of which 3222 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 Sugar aminotransferase.

Mol	Chain	Residues	Atoms								ZeroOcc	AltConf	Trace
1	AAA	369	Total	C	H	N	O	P	S		81	24	0
			6203	1992	3138	490	570	1	12				

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



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	AAA	1	Total	C	H	O	1	0
			10	2	6	2		
2	AAA	1	Total	C	H	O	1	0
			10	2	6	2		
2	AAA	1	Total	C	H	O	1	0
			10	2	6	2		
2	AAA	1	Total	C	H	O	1	0
			10	2	6	2		
2	AAA	1	Total	C	H	O	1	0
			10	2	6	2		

Continued on next page...

Continued from previous page...

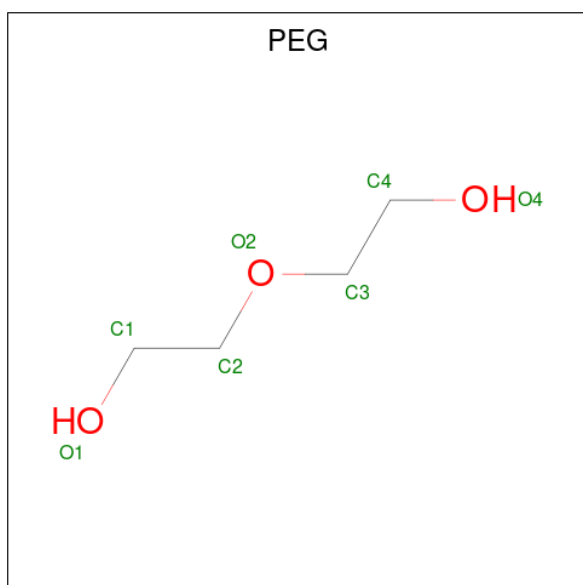
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	AAA	1	Total	C	H	O	1	0
			10	2	6	2		

- Molecule 3 is TETRAETHYLENE GLYCOL (three-letter code: PG4) (formula: $C_8H_{18}O_5$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	AAA	1	Total	C	H	O	1	0
			31	8	18	5		

- Molecule 4 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	AAA	1	Total	C	H	O	1	0
			17	4	10	3		
4	AAA	1	Total	C	H	O	1	0
			17	4	10	3		
4	AAA	1	Total	C	H	O	1	0
			17	4	10	3		

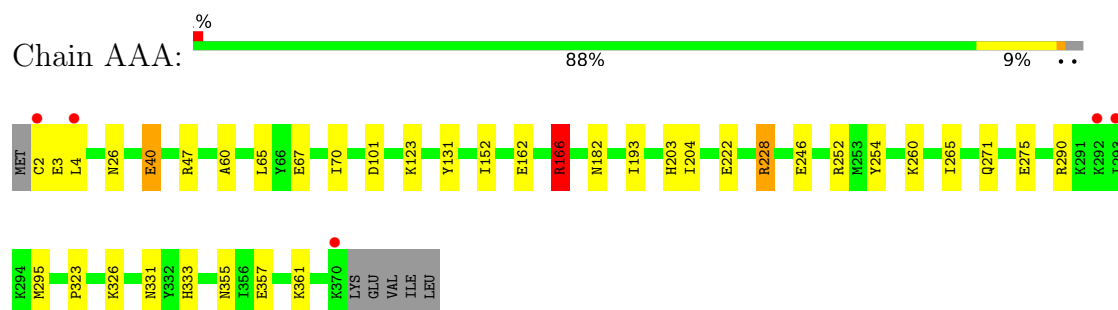
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	AAA	346	Total	O	0	0
			346	346		

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: Sugar aminotransferase



4 Data and refinement statistics

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, α , β , γ	76.41Å 76.41Å 134.44Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	59.37 – 1.43 59.37 – 1.43	Depositor EDS
% Data completeness (in resolution range)	100.0 (59.37-1.43) 100.0 (59.37-1.43)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.19 (at 1.43Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.165 , 0.197 0.173 , 0.201	Depositor DCC
R_{free} test set	4195 reflections (4.96%)	wwPDB-VP
Wilson B-factor (Å ²)	25.7	Xtriage
Anisotropy	0.279	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	(Not available) , (Not available)	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.023 for -h,-k,l	Xtriage
F_o, F_c correlation	0.98	EDS
Total number of atoms	6691	wwPDB-VP
Average B, all atoms (Å ²)	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.51% 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: PEG, LLP, EDO, PG4

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	0.90	2/3175 (0.1%)	1.07	9/4287 (0.2%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	AAA	246	GLU	CD-OE1	6.77	1.33	1.25
1	AAA	162	GLU	CD-OE2	-5.65	1.19	1.25

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	AAA	166[A]	ARG	NE-CZ-NH2	-10.09	115.26	120.30
1	AAA	166[B]	ARG	NE-CZ-NH2	-10.09	115.26	120.30
1	AAA	166[A]	ARG	NE-CZ-NH1	6.82	123.71	120.30
1	AAA	166[B]	ARG	NE-CZ-NH1	6.82	123.71	120.30
1	AAA	47[A]	ARG	NE-CZ-NH1	5.67	123.14	120.30

There are no chirality outliers.

There are no planarity outliers.

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	3065	3138	3151	25	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	AAA	24	36	36	6	0
3	AAA	13	18	18	0	0
4	AAA	21	30	30	1	0
5	AAA	346	0	0	10	0
All	All	3469	3222	3235	28	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 4.

The worst 5 of 28 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:166[B]:ARG:NH2	5:AAA:503:HOH:O	1.79	1.13
1:AAA:275[A]:GLU:OE1	5:AAA:502:HOH:O	1.74	1.05
1:AAA:166[A]:ARG:HG2	1:AAA:166[A]:ARG:HH11	1.44	0.79
1:AAA:67:GLU:OE1	1:AAA:203:HIS:HD2	1.72	0.73
2:AAA:407:EDO:H12	5:AAA:583:HOH:O	1.93	0.69

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	390/375 (104%)	382 (98%)	8 (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	337/319 (106%)	329 (98%)	8 (2%)	49 15

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	AAA	265	ILE
1	AAA	260	LYS
1	AAA	166[B]	ARG
1	AAA	166[A]	ARG
1	AAA	228	ARG

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 ⓘ

1 non-standard protein/DNA/RNA residue is modelled in this entry.

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.

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry

10 ligands are modelled in this entry.

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 [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	368/375 (98%)	0.04	5 (1%) 75 75	20, 30, 61, 121	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AAA	2	CYS	7.7
1	AAA	293	ILE	4.1
1	AAA	4	LEU	3.5
1	AAA	370	LYS	2.6
1	AAA	292	LYS	2.2

6.2 Non-standard residues in protein, DNA, RNA chains [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
1	LLP	AAA	183	24/25	0.99	0.08	20,25,33,36	1

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	AAA	402	4/4	0.68	0.14	62,66,70,72	1
4	PEG	AAA	406	7/7	0.77	0.16	68,82,92,93	1
4	PEG	AAA	410	7/7	0.86	0.10	56,64,66,70	1
2	EDO	AAA	409	4/4	0.90	0.15	51,55,57,61	1
2	EDO	AAA	401	4/4	0.90	0.11	42,54,59,68	1
2	EDO	AAA	404	4/4	0.90	0.14	50,58,73,87	1
4	PEG	AAA	408	7/7	0.91	0.20	59,64,75,82	1
3	PG4	AAA	405	13/13	0.93	0.08	38,45,54,65	1
2	EDO	AAA	403	4/4	0.93	0.23	39,47,51,54	1
2	EDO	AAA	407	4/4	0.95	0.15	36,45,87,87	1

6.5 Other polymers ⓘ

There are no such residues in this entry.