



wwPDB X-ray Structure Validation Summary Report ⓘ

Apr 27, 2022 – 03:35 PM EDT

PDB ID : 7MKR
Title : Crystal structure of the GH12 domain from *Acidothermus cellulolyticus* GuxA
Authors : Lunin, V.V.
Deposited on : 2021-04-26
Resolution : 1.50 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	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.28.1
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)	:	2.28.1

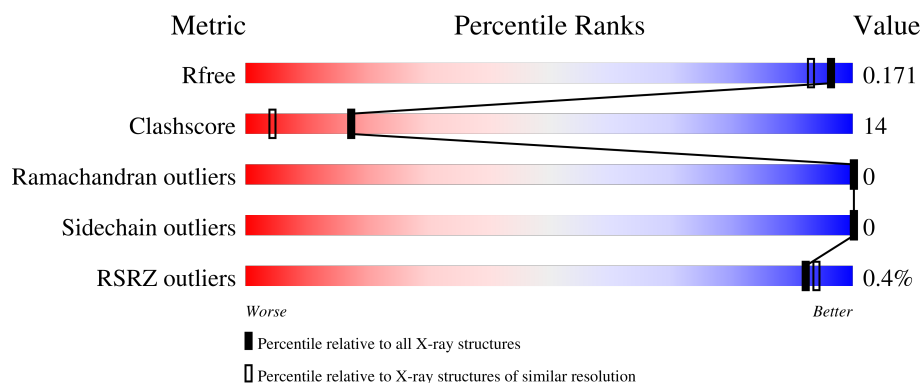
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.50 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	2936 (1.50-1.50)
Clashscore	141614	3144 (1.50-1.50)
Ramachandran outliers	138981	3066 (1.50-1.50)
Sidechain outliers	138945	3064 (1.50-1.50)
RSRZ outliers	127900	2884 (1.50-1.50)

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	244	<div> <div></div> <div>86%</div> <div>11%</div> <div>..</div> </div>

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
4	ACT	AAA	306	-	-	X	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
6	CL	AAA	319	-	-	X	-

2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 2438 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 Glycoside hydrolase, family 6.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	AAA	240	Total	C	N	O	S	4	26	0
			1908	1192	318	388	10			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	237	SER	-	expression tag	UNP A0LSH8
AAA	238	ALA	-	expression tag	UNP A0LSH8
AAA	239	HIS	-	expression tag	UNP A0LSH8
AAA	240	HIS	-	expression tag	UNP A0LSH8
AAA	241	HIS	-	expression tag	UNP A0LSH8
AAA	242	HIS	-	expression tag	UNP A0LSH8
AAA	243	HIS	-	expression tag	UNP A0LSH8
AAA	244	HIS	-	expression tag	UNP A0LSH8

- Molecule 2 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	AAA	5	Total	Na	0	0
			5	5		

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	AAA	4	Total	Zn	0	0
			4	4		

- Molecule 4 is ACETATE ION (three-letter code: ACT) (formula: C₂H₃O₂).



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

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).

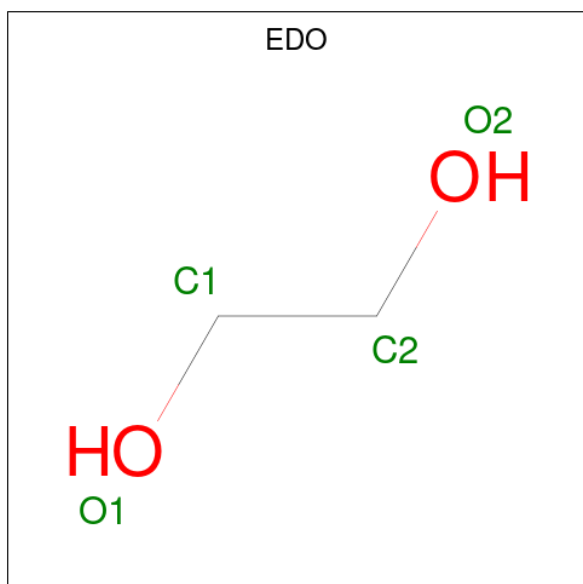


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	AAA	1	Total	C	O	0	0
			6	3	3		

- Molecule 6 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	AAA	2	Total	Cl	0	0
			2	2		

- Molecule 7 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
7	AAA	1	Total	C	O	0	0
			4	2	2		
7	AAA	1	Total	C	O	0	0
			4	2	2		
7	AAA	1	Total	C	O	0	0
			4	2	2		
7	AAA	1	Total	C	O	0	1
			8	4	4		
7	AAA	1	Total	C	O	0	0
			4	2	2		
7	AAA	1	Total	C	O	0	0
			4	2	2		

- Molecule 8 is water.

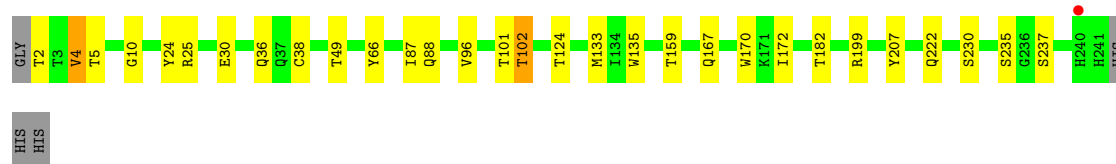
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	AAA	473	Total	O	0	8
			481	481		

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: Glycoside hydrolase, family 6

Chain AAA:  86% 11% ..



4 Data and refinement statistics

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α , β , γ	66.18Å 66.18Å 127.89Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	27.98 – 1.50 27.96 – 1.50	Depositor EDS
% Data completeness (in resolution range)	99.7 (27.98-1.50) 99.7 (27.96-1.50)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.97 (at 1.50Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.129 , 0.171 0.130 , 0.171	Depositor DCC
R_{free} test set	2663 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å ²)	6.4	Xtriage
Anisotropy	0.728	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 65.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.036 for -h,-k,l	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	2438	wwPDB-VP
Average B, all atoms (Å ²)	14.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.78% 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: NA, ACT, ZN, GOL, CL, 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	0.85	1/2011 (0.0%)	1.01	4/2759 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	AAA	4	VAL	CB-CG2	-5.02	1.42	1.52

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	AAA	25	ARG	NE-CZ-NH1	-8.76	115.92	120.30
1	AAA	199	ARG	NE-CZ-NH1	-5.38	117.61	120.30
1	AAA	102	THR	CA-CB-CG2	-5.37	104.88	112.40
1	AAA	24	TYR	CB-CG-CD1	-5.30	117.82	121.00

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	1908	0	1834	42	0
2	AAA	5	0	0	1	0
3	AAA	4	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	AAA	4	0	3	2	0
5	AAA	6	0	8	2	0
6	AAA	2	0	0	2	0
7	AAA	28	0	40	11	0
8	AAA	481	0	0	28	0
All	All	2438	0	1885	53	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 14.

The worst 5 of 53 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:182[A]:THR:CG2	8:AAA:424:HOH:O	1.80	1.25
1:AAA:88[B]:GLN:OE1	8:AAA:404:HOH:O	1.63	1.12
1:AAA:167[B]:GLN:HG3	1:AAA:172[B]:ILE:HD11	1.30	1.07
1:AAA:222[A]:GLN:OE1	8:AAA:405:HOH:O	1.76	1.03
1:AAA:167[B]:GLN:CG	1:AAA:172[B]:ILE:HD11	1.91	1.00

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	265/244 (109%)	261 (98%)	4 (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	218/195 (112%)	218 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

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 20 ligands modelled in this entry, 11 are monoatomic - leaving 9 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	$\# Z > 2$
7	EDO	AAA	313[A]	-	3,3,3	0.50	0	2,2,2	0.83	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
5	GOL	AAA	308	2	5,5,5	0.27	0	5,5,5	1.47	1 (20%)
7	EDO	AAA	313[B]	-	3,3,3	0.20	0	2,2,2	0.16	0
7	EDO	AAA	315	3	3,3,3	0.48	0	2,2,2	0.69	0
7	EDO	AAA	312	-	3,3,3	0.75	0	2,2,2	0.70	0
7	EDO	AAA	310	-	3,3,3	0.15	0	2,2,2	0.50	0
7	EDO	AAA	311	-	3,3,3	0.79	0	2,2,2	0.70	0
4	ACT	AAA	306	3	1,3,3	1.64	0	0,3,3	-	-
7	EDO	AAA	314	-	3,3,3	0.98	0	2,2,2	1.02	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
7	EDO	AAA	313[A]	-	-	1/1/1/1	-
5	GOL	AAA	308	2	-	2/4/4/4	-
7	EDO	AAA	313[B]	-	-	1/1/1/1	-
7	EDO	AAA	315	3	-	1/1/1/1	-
7	EDO	AAA	312	-	-	1/1/1/1	-
7	EDO	AAA	310	-	-	0/1/1/1	-
7	EDO	AAA	311	-	-	1/1/1/1	-
7	EDO	AAA	314	-	-	1/1/1/1	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	AAA	308	GOL	O1-C1-C2	2.58	122.56	110.20

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	AAA	308	GOL	C1-C2-C3-O3
7	AAA	311	EDO	O1-C1-C2-O2
7	AAA	312	EDO	O1-C1-C2-O2
7	AAA	313[A]	EDO	O1-C1-C2-O2
7	AAA	313[B]	EDO	O1-C1-C2-O2

There are no ring outliers.

7 monomers are involved in 15 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
7	AAA	313[A]	EDO	2	0
5	AAA	308	GOL	2	0
7	AAA	313[B]	EDO	3	0
7	AAA	315	EDO	2	0
7	AAA	312	EDO	2	0
4	AAA	306	ACT	2	0
7	AAA	314	EDO	2	0

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	240/244 (98%)	-0.85	1 (0%) 92 94	4, 8, 19, 57	6 (2%)

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AAA	240[A]	HIS	4.7

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
6	CL	AAA	319	1/1	0.87	0.09	43,43,43,43	0
7	EDO	AAA	314	4/4	0.88	0.13	17,18,24,29	4
7	EDO	AAA	315	4/4	0.88	0.10	30,33,38,38	4
3	ZN	AAA	305	1/1	0.92	0.12	31,31,31,31	1
7	EDO	AAA	312	4/4	0.92	0.11	18,20,24,25	4
2	NA	AAA	307	1/1	0.95	0.07	27,27,27,27	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
2	NA	AAA	301	1/1	0.95	0.08	30,30,30,30	1
7	EDO	AAA	313[B]	4/4	0.96	0.14	30,31,31,37	4
5	GOL	AAA	308	6/6	0.96	0.12	7,19,19,23	6
7	EDO	AAA	313[A]	4/4	0.96	0.14	14,15,17,23	4
7	EDO	AAA	311	4/4	0.97	0.10	19,20,27,31	4
7	EDO	AAA	310	4/4	0.98	0.12	18,23,25,26	4
2	NA	AAA	318	1/1	0.99	0.03	28,28,28,28	0
2	NA	AAA	316	1/1	0.99	0.04	31,31,31,31	0
4	ACT	AAA	306	4/4	0.99	0.04	11,15,18,21	4
2	NA	AAA	317	1/1	0.99	0.03	27,27,27,27	0
3	ZN	AAA	302	1/1	1.00	0.03	9,9,9,9	1
3	ZN	AAA	303	1/1	1.00	0.04	13,13,13,13	1
3	ZN	AAA	304	1/1	1.00	0.02	12,12,12,12	1
6	CL	AAA	309	1/1	1.00	0.03	9,9,9,9	1

6.5 Other polymers [i](#)

There are no such residues in this entry.