



Full wwPDB X-ray Structure Validation Report ⓘ

Mar 6, 2026 – 09:59 PM UTC

PDB ID : 5OKH / pdb_00005okh
Title : Conservatively refined structure of Gan1D-WT, a putative 6-phospho-beta-galactosidase from *Geobacillus stearothermophilus*, in the C2 spacegroup
Authors : Lansky, S.; Zehavi, A.; Shoham, Y.; Shoham, G.
Deposited on : 2017-07-25
Resolution : 1.98 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	:	4-5-2 with Phenix2.0
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

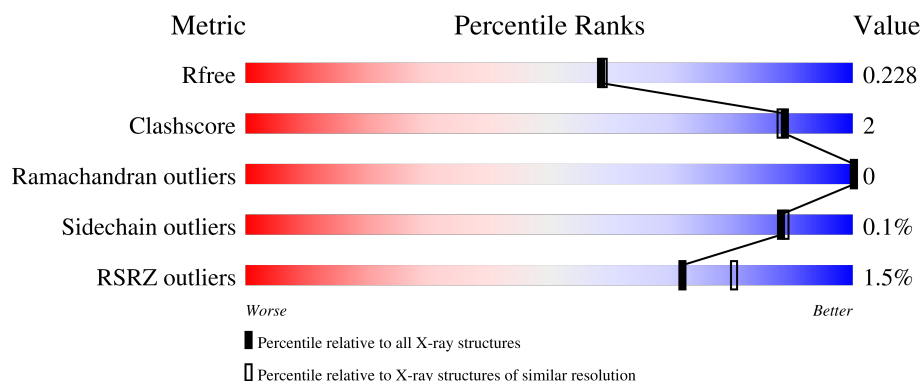
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.98 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}	180053	1506 (1.98-1.98)
Clashscore	190562	1534 (1.98-1.98)
Ramachandran outliers	187476	1518 (1.98-1.98)
Sidechain outliers	187428	1518 (1.98-1.98)
RSRZ outliers	180081	1506 (1.98-1.98)

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	A	485	% 94% • •
1	B	485	% 92% 6% •

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 8355 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 Putative 6-phospho-beta-galactobiosidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	474	Total	C	N	O	S	0	9	0
			3921	2519	674	717	11			
1	B	474	Total	C	N	O	S	0	3	0
			3888	2497	665	715	11			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-6	MET	-	initiating methionine	UNP W8QF82
A	-5	ILE	-	expression tag	UNP W8QF82
A	-4	HIS	-	expression tag	UNP W8QF82
A	-3	HIS	-	expression tag	UNP W8QF82
A	-2	HIS	-	expression tag	UNP W8QF82
A	-1	HIS	-	expression tag	UNP W8QF82
A	0	HIS	-	expression tag	UNP W8QF82
A	1	HIS	-	expression tag	UNP W8QF82
B	-6	MET	-	initiating methionine	UNP W8QF82
B	-5	ILE	-	expression tag	UNP W8QF82
B	-4	HIS	-	expression tag	UNP W8QF82
B	-3	HIS	-	expression tag	UNP W8QF82
B	-2	HIS	-	expression tag	UNP W8QF82
B	-1	HIS	-	expression tag	UNP W8QF82
B	0	HIS	-	expression tag	UNP W8QF82
B	1	HIS	-	expression tag	UNP W8QF82

- Molecule 2 is GLYCEROL (CCD ID: GOL) (formula: C₃H₈O₃).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			6	3	3		

- Molecule 3 is IMIDAZOLE (CCD ID: IMD) (formula: $C_3H_5N_2$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	N	0	0
			5	3	2		
3	A	1	Total	C	N	0	0
			5	3	2		

- Molecule 4 is 2-AMINO-2-HYDROXYMETHYL-PROPANE-1,3-DIOL (CCD ID: TRS)

(formula: $C_4H_{12}NO_3$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total	C	N	O	0	0
			8	4	1	3		

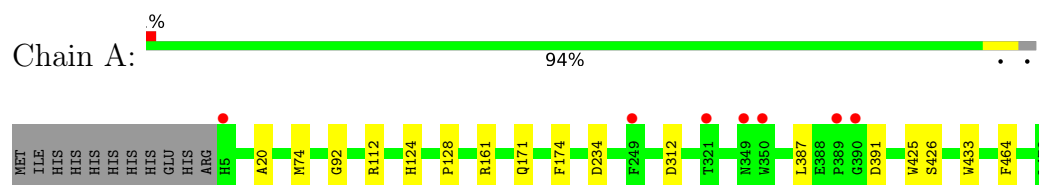
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	355	Total	O	0	0
			355	355		
5	B	167	Total	O	0	0
			167	167		

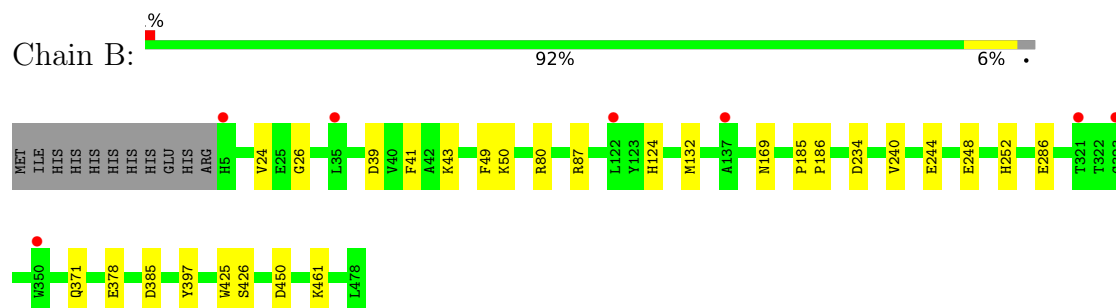
3 Residue-property plots [i](#)

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: Putative 6-phospho-beta-galactobiosidase



- Molecule 1: Putative 6-phospho-beta-galactobiosidase



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	108.02Å 68.81Å 152.77Å 90.00° 100.74° 90.00°	Depositor
Resolution (Å)	25.59 – 1.98 25.59 – 1.98	Depositor EDS
% Data completeness (in resolution range)	99.9 (25.59-1.98) 99.9 (25.59-1.98)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.13 (at 1.98Å)	Xtriage
Refinement program	PHENIX (1.10.1 _2155: ???)	Depositor
R, R_{free}	0.188 , 0.232 0.186 , 0.228	Depositor DCC
R_{free} test set	3856 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å ²)	27.7	Xtriage
Anisotropy	0.930	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 36.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	8355	wwPDB-VP
Average B, all atoms (Å ²)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.46% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, IMD, TRS

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	A	0.38	0/4069	0.52	0/5534
1	B	0.36	0/4018	0.48	0/5467
All	All	0.37	0/8087	0.50	0/11001

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	A	3921	0	3745	11	0
1	B	3888	0	3689	19	0
2	A	6	0	8	0	0
3	A	10	0	10	0	0
4	A	8	0	12	2	0
5	A	355	0	0	0	0
5	B	167	0	0	1	0
All	All	8355	0	7464	29	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 (29) 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:B:132:MET:CE	1:B:186:PRO:HB3	2.27	0.64
1:B:132:MET:HE1	1:B:186:PRO:HB3	1.84	0.59
1:B:80:ARG:HH11	1:B:378:GLU:HG3	1.74	0.53
1:B:50:LYS:HD2	1:B:450:ASP:HB3	1.93	0.50
1:A:433:TRP:HE1	4:A:504:TRS:H12	1.76	0.50
1:B:385:ASP:OD1	1:B:397:TYR:OH	2.29	0.49
1:A:74:MET:HE3	1:A:464:PHE:CD2	2.50	0.47
1:A:433:TRP:HE1	4:A:504:TRS:C1	2.28	0.46
1:B:124:HIS:CE1	1:B:169:ASN:HD22	2.34	0.45
1:B:41:PHE:CG	1:B:185:PRO:HD3	2.52	0.45
1:A:234:ASP:HA	1:B:234:ASP:HA	1.98	0.45
1:A:387:LEU:HD11	1:A:391:ASP:HA	1.99	0.44
1:A:312:ASP:OD1	1:A:312:ASP:N	2.51	0.43
1:B:80:ARG:NH1	1:B:378:GLU:HG3	2.33	0.43
1:A:92:GLY:HA2	1:A:128:PRO:HG2	2.00	0.43
1:A:112[A]:ARG:HD2	1:A:161:ARG:HB3	2.00	0.43
1:A:171:GLN:HA	1:A:174:PHE:CE2	2.53	0.43
1:B:49:PHE:CE2	1:B:50:LYS:HE2	2.54	0.43
1:B:49:PHE:CZ	1:B:50:LYS:HE2	2.54	0.42
1:A:425:TRP:HA	1:A:426:SER:HA	1.82	0.42
1:B:248:GLU:HA	1:B:252:HIS:HB2	2.01	0.42
1:B:425:TRP:HA	1:B:426:SER:HA	1.67	0.42
1:B:240:VAL:O	1:B:244:GLU:HG2	2.20	0.42
1:B:385:ASP:OD2	1:B:461:LYS:NZ	2.47	0.42
1:B:39:ASP:O	1:B:43:LYS:HD2	2.20	0.41
1:B:286:GLU:H	1:B:286:GLU:CD	2.28	0.41
1:B:371:GLN:NE2	5:B:502:HOH:O	2.18	0.41
1:A:20:ALA:HB2	1:A:124:HIS:CE1	2.57	0.40
1:B:26:GLY:O	1:B:87:ARG:NH2	2.55	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	A	481/485 (99%)	463 (96%)	18 (4%)	0	100	100
1	B	475/485 (98%)	456 (96%)	19 (4%)	0	100	100
All	All	956/970 (99%)	919 (96%)	37 (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	A	409/412 (99%)	409 (100%)	0	100	100
1	B	403/412 (98%)	402 (100%)	1 (0%)	87	88
All	All	812/824 (98%)	811 (100%)	1 (0%)	88	89

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	24	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (10) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	214	HIS
1	A	250	GLN
1	A	379	ASN
1	A	435	ASN
1	B	156	GLN
1	B	210	GLN
1	B	250	GLN
1	B	289	GLN
1	B	349	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	438	GLN

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 oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

4 ligands are modelled in this entry.

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	IMD	A	503	-	5,5,5	0.76	0	5,5,5	0.41	0
2	GOL	A	501	-	5,5,5	0.37	0	5,5,5	0.40	0
3	IMD	A	502	-	5,5,5	0.75	0	5,5,5	0.30	0
4	TRS	A	504	-	7,7,7	0.35	0	9,9,9	1.18	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	IMD	A	503	-	-	-	0/1/1/1

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	GOL	A	501	-	-	0/4/4/4	-
3	IMD	A	502	-	-	-	0/1/1/1
4	TRS	A	504	-	-	9/9/9/9	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (9) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	504	TRS	C2-C-C1-O1
4	A	504	TRS	C3-C-C1-O1
4	A	504	TRS	N-C-C1-O1
4	A	504	TRS	C3-C-C2-O2
4	A	504	TRS	N-C-C2-O2
4	A	504	TRS	N-C-C3-O3
4	A	504	TRS	C1-C-C2-O2
4	A	504	TRS	C1-C-C3-O3
4	A	504	TRS	C2-C-C3-O3

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	504	TRS	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	A	474/485 (97%)	0.17	7 (1%) 72 80	12, 25, 48, 71	9 (1%)
1	B	474/485 (97%)	0.51	7 (1%) 72 80	23, 35, 54, 77	3 (0%)
All	All	948/970 (97%)	0.34	14 (1%) 72 80	12, 31, 52, 77	12 (1%)

All (14) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	5	HIS	3.8
1	A	321	THR	3.1
1	B	321	THR	2.7
1	B	5	HIS	2.7
1	A	249	PHE	2.6
1	A	349	ASN	2.5
1	A	389	PRO	2.5
1	B	350	TRP	2.3
1	A	390	GLY	2.3
1	B	35	LEU	2.2
1	B	323	GLY	2.1
1	A	350	TRP	2.1
1	B	122	LEU	2.0
1	B	137	ALA	2.0

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 oligosaccharides 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
4	TRS	A	504	8/8	0.72	0.15	23,34,41,44	0
2	GOL	A	501	6/6	0.74	0.13	39,46,49,50	0
3	IMD	A	502	5/5	0.82	0.18	29,35,40,43	0
3	IMD	A	503	5/5	0.88	0.13	28,30,30,32	0

6.5 Other polymers [i](#)

There are no such residues in this entry.