



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 01:54 PM GMT

PDB ID : 3VGB  
Title : Crystal structure of glycosyltrehalose trehalohydrolase (GTHase) from *Sulfolobus solfataricus* KM1  
Authors : Okazaki, N.; Tamada, T.; Feese, M.D.; Kato, M.; Miura, Y.; Komeda, T.; Kobayashi, K.; Kondo, K.; Kuroki, R.  
Deposited on : 2011-08-09  
Resolution : 2.65 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](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/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 : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865

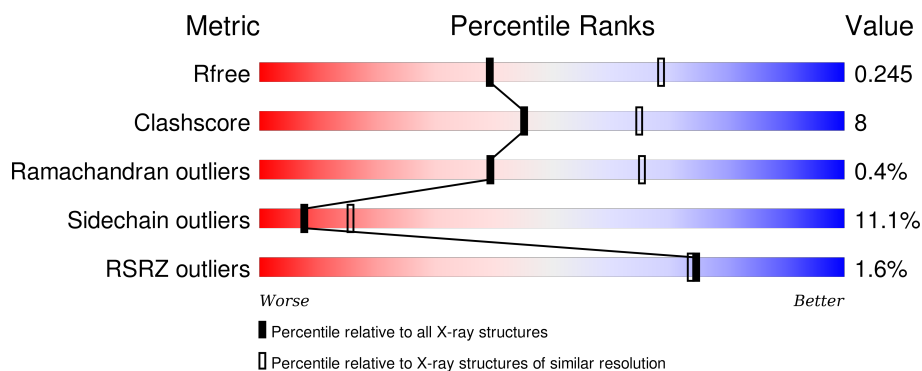
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.65 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))
$R_{free}$	91344	3152 (2.70-2.62)
Clashscore	102246	3524 (2.70-2.62)
Ramachandran outliers	100387	3469 (2.70-2.62)
Sidechain outliers	100360	3469 (2.70-2.62)
RSRZ outliers	91569	3161 (2.70-2.62)

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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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	558	

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	GOL	A	1101	-	-	-	X

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	GOL	A	1102	-	-	-	X
3	GOL	A	1103	-	-	-	X

## 2 Entry composition [i](#)

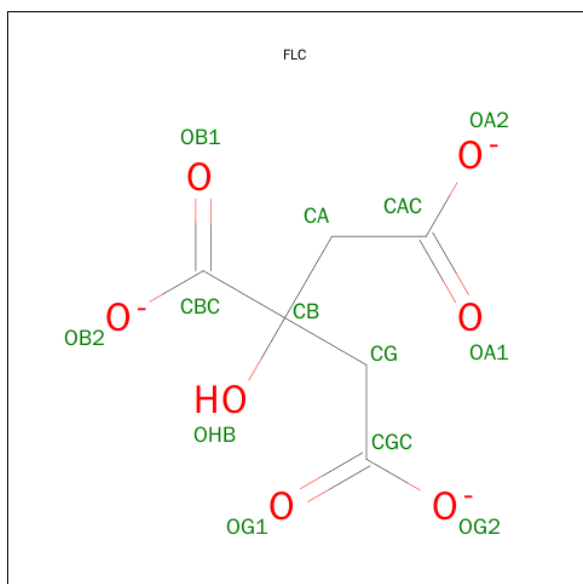
There are 4 unique types of molecules in this entry. The entry contains 4720 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Malto-oligosyltrehalose trehalohydrolase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	550	Total	C	N	O	S	0	0	0
			4512	2907	741	855	9			

- Molecule 2 is CITRATE ANION (three-letter code: FLC) (formula:  $C_6H_5O_7$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			13	6	7		

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



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

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	177	Total	O	0	0
			177	177		

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

- Chain A:
- 
- 2% 75% 10%
- THR PHE ALA TYR LYS ILE ASP G8 I19 E10 W16 Q20 K24 L25 Y33 I45 T46 L47 M48 V52 R53 V59 L60 D61 D62 E65 P69 A70 S71 R72 Y73 Q74 I84 E87 S88 K89 N92 I93 E94 K98 K99 E100 I104 Y105 Y161
- I122 I123 K124 L125 L128 I133 I138 I141 A142 Q143 V155 Y156 L157 V160 Q161 M162 V174 K179 V184 I185 L186 V193 G194 P195 N198 L203 K210 Y211 K212 W215 G216 L217 E225 S226 V229 V237 K242 E243 D252 E264
- L265 L266 E267 E268 V272 V273 R278 L279 V280 R290 V291 V292 K295 E296 K297 C298 D308 R311 R312 E321 R322 Q323 L332 D333 V336 Y339 K340 R354 R357 V361 V370 V371 V372 R387 V392 E395 S396 A402 Y409 W412
- I413 F414 Y415 Y416 Y423 F427 R435 L436 L437 Q438 R444 K445 K446 E447 Q450 D453 P454 Q455 D456 E457 S458 L480 Y483 R484 L487 R489 R494 V497 E501 L504 R509 E510 Y511 S518 K525 L531 L532 L535 Y536 S537 E544
- F553 A554 I555 Y556 K557 LEU

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	78.27Å 78.27Å 282.10Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	67.78 – 2.65 67.78 – 2.65	Depositor EDS
% Data completeness (in resolution range)	97.4 (67.78-2.65) 97.4 (67.78-2.65)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.64 (at 2.65Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, $R_{free}$	0.202 , 0.243 0.206 , 0.245	Depositor DCC
$R_{free}$ test set	1486 reflections (5.33%)	DCC
Wilson B-factor (Å <sup>2</sup> )	33.5	Xtriage
Anisotropy	0.013	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 45.0	EDS
Estimated twinning fraction	0.057 for -h,-k,l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.28$	Xtriage
Outliers	0 of 29341 reflections	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	4720	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.06% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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, FLC

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	0.39	0/4628	0.55	0/6253

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	4512	0	4369	70	0
2	A	13	0	5	0	0
3	A	18	0	24	3	0
4	A	177	0	0	7	0
All	All	4720	0	4398	70	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 8.

The worst 5 of 70 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:A:133:ILE:HD11	1:A:413:ILE:HD12	1.48	0.95
1:A:354:ARG:HG3	1:A:354:ARG:HH11	1.49	0.76
1:A:264:HIS:HD2	1:A:266:LEU:H	1.36	0.73
1:A:105:TYR:HB2	1:A:133:ILE:HG12	1.71	0.71
1:A:321:GLU:HB3	1:A:323:GLN:NE2	2.06	0.71

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	A	548/558 (98%)	524 (96%)	22 (4%)	2 (0%)	39 65

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	210	LYS
1	A	62	ASP

### 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	A	487/494 (99%)	433 (89%)	54 (11%)	8 15

5 of 54 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	237	VAL
1	A	336	VAL
1	A	525	LYS
1	A	280	VAL
1	A	292	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	264	HIS
1	A	301	ASN
1	A	357	HIS
1	A	245	ASN
1	A	323	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 carbohydrates 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
2	FLC	A	1001	-	3,12,12	0.57	0	3,17,17	0.59	0
3	GOL	A	1101	-	5,5,5	0.34	0	5,5,5	0.25	0
3	GOL	A	1102	-	5,5,5	0.33	0	5,5,5	0.28	0
3	GOL	A	1103	-	5,5,5	0.32	0	5,5,5	0.36	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	FLC	A	1001	-	-	0/6/16/16	0/0/0/0
3	GOL	A	1101	-	-	0/4/4/4	0/0/0/0
3	GOL	A	1102	-	-	0/4/4/4	0/0/0/0
3	GOL	A	1103	-	-	0/4/4/4	0/0/0/0

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.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1101	GOL	2	0
3	A	1103	GOL	1	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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	550/558 (98%)	0.24	9 (1%) 74 74	24, 40, 78, 101	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	8	GLY	4.7
1	A	211	TYR	3.9
1	A	212	LYS	2.8
1	A	46	THR	2.7
1	A	217	LEU	2.7

### 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 carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
3	GOL	A	1102	6/6	0.94	0.38	6.45	65,66,66,67	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	GOL	A	1101	6/6	0.82	0.33	4.28	79,80,81,81	0
3	GOL	A	1103	6/6	0.90	0.29	3.47	58,58,59,59	0
2	FLC	A	1001	13/13	0.96	0.14	-1.10	49,50,50,50	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.