

# wwPDB X-ray Structure Validation Summary Report (i)

#### Jan 4, 2021 – 11:11 am GMT

PDB ID	:	6TRT
Title	:	Chaetomium thermophilum UDP-Glucose Glucosyl Transferase (UGGT) dou-
		ble cysteine mutant $S180C/T742C$ .
Authors	:	Roversi, P.; Zitzmann, N.; Ibba, R.; Hensen, M.
Deposited on	:	2019-12-19
Resolution	:	4.58  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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
$\mathrm{EDS}$	:	2.16
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
$\operatorname{Refmac}$	:	5.8.0158
$\operatorname{CCP4}$	:	7.0.044  (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.16

# 1 Overall quality at a glance (i)

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

The reported resolution of this entry is 4.58 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 <sub>free</sub>	130704	1060 (5.30-3.80)
Clashscore	141614	1128 (5.30-3.80)
Ramachandran outliers	138981	1072(5.30-3.80)
Sidechain outliers	138945	1053 (5.30-3.80)
RSRZ outliers	127900	1102(5.46-3.70)

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	А	1494	9%	20%	• 8%				
2	В	3	100%						
2	D	3	100%						
3	С	8	75%	2	25%				

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard



Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	NAG	D	1	-	-	-	Х
2	NAG	D	2	-	-	-	Х
3	MAN	С	4	-	-	-	Х
3	MAN	С	5	-	-	-	Х
5	NAG	А	1607	_	-	_	Х

residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:



# 2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 11210 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 UDP-glucose-glycoprotein glucosyltransferase-like protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	А	1369	Total 11007	C 7045	N 1875	O 2053	${ m S} { m 34}$	0	0	0

Chain	Residue	Modelled	Actual	Comment	Reference
А	21	GLU	-	expression tag	UNP G0SB58
А	22	THR	-	expression tag	UNP G0SB58
А	23	GLY	-	expression tag	UNP G0SB58
А	180	CYS	SER	engineered mutation	UNP G0SB58
А	742	CYS	THR	engineered mutation	UNP G0SB58
A	1506	GLY	-	expression tag	UNP G0SB58
A	1507	THR	-	expression tag	UNP G0SB58
А	1508	LYS	-	expression tag	UNP G0SB58
А	1509	HIS	-	expression tag	UNP G0SB58
A	1510	HIS	-	expression tag	UNP G0SB58
A	1511	HIS	-	expression tag	UNP G0SB58
А	1512	HIS	-	expression tag	UNP G0SB58
A	1513	HIS	-	expression tag	UNP G0SB58
A	1514	HIS	-	expression tag	UNP G0SB58

There are 14 discrepancies between the modelled and reference sequences:

• Molecule 2 is an oligosaccharide called beta-D-mannopyranose-(1-4)-2-acetamido-2-deoxy-b eta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose.



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	В	3	Total 39	С 22	N 2	0 15	0	0	0

Continued on next page...



Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf	Trace		
2	D	3	Total 39	С 22	N 2	O 15	0	0	0

• Molecule 3 is an oligosaccharide called alpha-D-mannopyranose-(1-2)-alpha-D-mannopyran ose-(1-2)-alpha-D-mannopyranose-(1-6)-[alpha-D-mannopyranose-(1-2)-alpha-D-mannopyranose-(1-3)]beta-D-mannopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose.



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	С	8	Total 94	C 52	N 2	O 40	0	0	0

• Molecule 4 is TERBIUM(III) ION (three-letter code: TB) (formula: Tb).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	А	3	Total Tb 3 3	0	0

• Molecule 5 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula:  $C_8H_{15}NO_6$ ).





Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	А	1	Total         C         N         O           14         8         1         5	0	0
5	А	1	Total         C         N         O           14         8         1         5	0	0



# 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: UDP-glucose-glycoprotein glucosyltransferase-like protein



• Molecule 2: beta-D-mannopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose

Chain B:

100%

#### NAG 1 NAG 2 BMA 3

• Molecule 2: beta-D-mannopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose

Chain D:

100%

#### NAG 1 NAG 2 BMA 3

• Molecule 3: alpha-D-mannopyranose-(1-2)-alpha-D-mannopyranose-(1-2)-alpha-D-mannopyranose-(1-6)-[alpha-D-mannopyranose-(1-2)-alpha-D-mannopyranose-(1-3)]beta-D-mannopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose-(1-4)-2-acetamido-2-deoxy-beta-D-glucopyranose e

Chain C:

75%

25%





# 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32 1 2	Depositor
Cell constants	$148.80 \text{\AA}$ 148.80 Å $235.55 \text{\AA}$	Deperitor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	$90.00^{\circ}$ $90.00^{\circ}$ $120.00^{\circ}$	Depositor
$\mathbf{P}_{\text{assolution}}(\hat{\mathbf{A}})$	128.86 - 4.58	Depositor
Resolution (A)	128.86 - 4.56	EDS
% Data completeness	54.9(128.86-4.58)	Depositor
(in resolution range)	54.9(128.86-4.56)	EDS
R <sub>merge</sub>	0.12	Depositor
R <sub>sym</sub>	(Not available)	Depositor
$< I/\sigma(I) > 1$	$1.46 (at 4.47 \text{\AA})$	Xtriage
Refinement program	BUSTER 2.10.3	Depositor
D D .	0.291 , $0.297$	Depositor
II, II, <i>free</i>	0.281 , $0.285$	DCC
$R_{free}$ test set	418 reflections $(4.39\%)$	wwPDB-VP
Wilson B-factor $(Å^2)$	263.8	Xtriage
Anisotropy	0.046	Xtriage
Bulk solvent $k_{sol}(e/A^3), B_{sol}(A^2)$	0.31 , 177.0	EDS
L-test for $twinning^2$	$< L >=0.41, < L^2>=0.24$	Xtriage
Estimated twinning fraction	0.083 for -h,-k,l	Xtriage
$F_o, F_c$ correlation	0.82	EDS
Total number of atoms	11210	wwPDB-VP
Average B, all atoms $(Å^2)$	142.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 2.91% of the height of the origin peak. No significant pseudotranslation is detected.

<sup>&</sup>lt;sup>2</sup>Theoretical 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.



<sup>&</sup>lt;sup>1</sup>Intensities estimated from amplitudes.

# 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: BMA, TB, NAG, MAN

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).

Mal	Chain	Bond	lengths	Bo	ond angles
	RMSZ	# Z  > 5	RMSZ	# Z  > 5	
1	А	0.29	0/11272	0.61	3/15289~(0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	$Observed(^{o})$	$Ideal(^{o})$
1	А	305	GLU	C-N-CA	6.93	139.03	121.70
1	А	243	TYR	C-N-CA	5.82	136.25	121.70
1	А	568	PHE	C-N-CA	5.43	135.28	121.70

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	А	11007	0	10875	225	6
2	В	39	0	34	0	0
2	D	39	0	34	0	0
3	С	94	0	79	3	0
4	А	3	0	0	0	0
5	А	28	0	26	0	0
All	All	11210	0	11048	228	6



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 10.

The worst 5 of 228 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:1225:HIS:ND1	1:A:1308:ARG:HA	1.64	1.11
1:A:338:HIS:HD2	1:A:898:THR:HG23	1.24	1.02
3:C:3:BMA:H4	3:C:6:MAN:H61	1.44	0.97
1:A:1295:LEU:HD21	1:A:1298:VAL:CG2	1.95	0.94
1:A:1225:HIS:CE1	1:A:1308:ARG:HA	2.05	0.91

The worst 5 of 6 symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:1442:THR:OG1	1:A:1442:THR:CG2[5_554]	1.39	0.81
1:A:1442:THR:CB	1:A:1442:THR:CB[5_554]	1.39	0.81
1:A:1442:THR:CB	1:A:1442:THR:CG2[5_554]	1.49	0.71
1:A:1442:THR:CG2	1:A:1442:THR:CG2[5_554]	1.51	0.69
1:A:1442:THR:CB	1:A:1442:THR:OG1[5_554]	1.55	0.65

### 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	А	1363/1494~(91%)	1284 (94%)	65~(5%)	14 (1%)	15 54

5 of 14 Ramachandran outliers are listed below:

Mol	Chain	$\mathbf{Res}$	Type
1	А	288	SER
1	А	569	LYS
1	А	656	ASN

Continued on next page...



Continued from previous page...

Mol	Chain	$\mathbf{Res}$	Type
1	А	657	THR
1	А	659	ILE

#### 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	Percentil	$\mathbf{es}$
1	А	1196/1297~(92%)	1163~(97%)	33~(3%)	43 65	

5 of 33 residues with a non-rotameric sidechain are listed below:

Mol	Chain	$\mathbf{Res}$	Type
1	А	758	ASP
1	А	873	ASP
1	А	1437	CYS
1	А	759	LYS
1	А	770	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	А	513	GLN
1	А	656	ASN
1	А	1305	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)

14 monosaccharides 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.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for oligosaccharide.











# 5.6 Ligand geometry (i)

Of 5 ligands modelled in this entry, 3 are monoatomic - leaving 2 for Mogul analysis.

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 (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,  $95^{th}$  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	< <b>RSRZ</b> >	#RSRZ>2	$\mathbf{OWAB}(\mathbf{A}^2)$	Q<0.9
1	А	1369/1494~(91%)	0.54	134 (9%) 7 7	54, 128, 265, 300	1 (0%)

The worst 5 of 134 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	А	132	SER	14.0
1	А	133	GLY	11.2
1	А	1084	LYS	6.0
1	А	279	ILE	5.8
1	А	1205	VAL	5.7

# 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)

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,  $95^{th}$  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 <sup>2</sup> )	Q<0.9
3	MAN	C	4	11/12	0.13	0.49	300,300,300,300	0
3	BMA	С	3	11/12	0.32	0.36	291,296,298,298	0
3	MAN	C	5	11/12	0.48	0.55	300,300,300,300	0
2	BMA	D	3	11/12	0.57	0.36	291,292,292,293	0
2	NAG	D	2	14/15	0.58	0.81	290,292,295,296	0
3	MAN	С	6	11/12	0.60	0.30	300,300,300,300	0
2	BMA	В	3	11/12	0.60	0.39	167,261,296,297	0

Continued on next page...



Mol	Type	Chain	Res	Atoms	RSCC	RSR	$\mathbf{B} ext{-factors}(\mathbf{A}^2)$	Q<0.9
3	MAN	С	8	11/12	0.71	0.27	298,298,299,299	0
2	NAG	D	1	14/15	0.72	0.74	286,292,293,294	0
3	MAN	С	7	11/12	0.83	0.23	297,298,299,299	0
3	NAG	С	2	14/15	0.89	0.33	191,214,266,286	0
2	NAG	В	1	14/15	0.91	0.27	$95,\!109,\!135,\!138$	0
2	NAG	В	2	14/15	0.92	0.22	113,153,272,293	0
3	NAG	С	1	14/15	0.93	0.41	132,156,192,273	0

Continued from previous page...

The following is a graphical depiction of the model fit to experimental electron density for oligosaccharide. Each fit is shown from different orientation to approximate a three-dimensional view.









# 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,  $95^{th}$  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	${f B}$ -factors(Å <sup>2</sup> )	Q<0.9
5	NAG	А	1607	14/15	<mark>0.39</mark>	0.52	114,191,280,293	0
4	TB	А	1601	1/1	0.73	0.36	213,213,213,213	0
5	NAG	А	1619	14/15	0.87	0.56	289,290,291,291	0
4	TB	А	1602	1/1	0.99	0.28	166, 166, 166, 166	0
4	TB	А	1603	1/1	0.99	0.31	83,83,83,83	0

### 6.5 Other polymers (i)

There are no such residues in this entry.

