



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 7, 2020 – 04:32 PM BST

PDB ID : 6RJM
Title : Complex structure of virulence factor SghA and its hydrolysis product glucose
Authors : Ye, F.Z.; Wang, C.; Chang, C.Q.; Zhang, L.H.; Gao, Y.G.
Deposited on : 2019-04-27
Resolution : 2.11 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 : 2.13.1
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) : 2.13.1

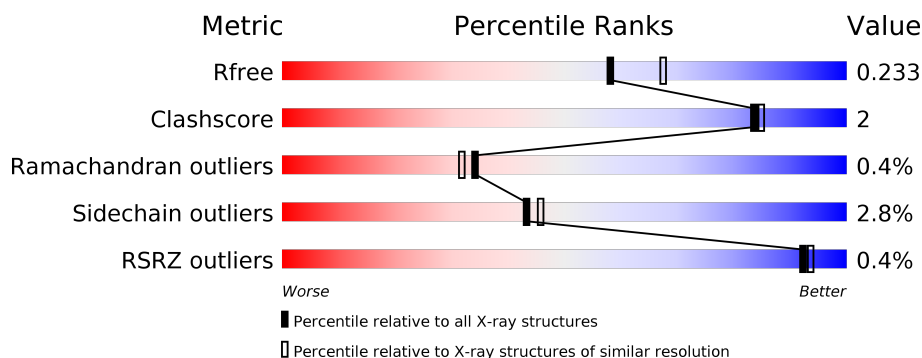
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.11 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	6241 (2.14-2.10)
Clashscore	141614	6778 (2.14-2.10)
Ramachandran outliers	138981	6705 (2.14-2.10)
Sidechain outliers	138945	6706 (2.14-2.10)
RSRZ outliers	127900	6112 (2.14-2.10)

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 ≥ 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	490	
1	B	490	

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 7647 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 Beta-glucosidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	453	Total	C	N	O	S	0	0	0
			3584	2294	610	659	21			
1	B	456	Total	C	N	O	S	0	0	0
			3598	2299	616	662	21			

There are 46 discrepancies between the modelled and reference sequences:

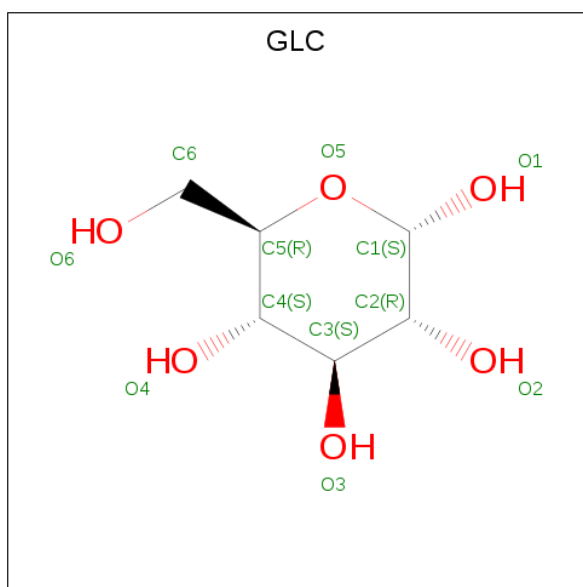
Chain	Residue	Modelled	Actual	Comment	Reference
A	-22	MET	-	initiating methionine	UNP A0A2I4PGZ0
A	-21	GLY	-	expression tag	UNP A0A2I4PGZ0
A	-20	SER	-	expression tag	UNP A0A2I4PGZ0
A	-19	SER	-	expression tag	UNP A0A2I4PGZ0
A	-18	HIS	-	expression tag	UNP A0A2I4PGZ0
A	-17	HIS	-	expression tag	UNP A0A2I4PGZ0
A	-16	HIS	-	expression tag	UNP A0A2I4PGZ0
A	-15	HIS	-	expression tag	UNP A0A2I4PGZ0
A	-14	HIS	-	expression tag	UNP A0A2I4PGZ0
A	-13	HIS	-	expression tag	UNP A0A2I4PGZ0
A	-12	SER	-	expression tag	UNP A0A2I4PGZ0
A	-11	SER	-	expression tag	UNP A0A2I4PGZ0
A	-10	GLY	-	expression tag	UNP A0A2I4PGZ0
A	-9	LEU	-	expression tag	UNP A0A2I4PGZ0
A	-8	VAL	-	expression tag	UNP A0A2I4PGZ0
A	-7	PRO	-	expression tag	UNP A0A2I4PGZ0
A	-6	ARG	-	expression tag	UNP A0A2I4PGZ0
A	-5	GLY	-	expression tag	UNP A0A2I4PGZ0
A	-4	SER	-	expression tag	UNP A0A2I4PGZ0
A	-3	HIS	-	expression tag	UNP A0A2I4PGZ0
A	-2	MET	-	expression tag	UNP A0A2I4PGZ0
A	-1	LEU	-	expression tag	UNP A0A2I4PGZ0
A	0	GLU	-	expression tag	UNP A0A2I4PGZ0
B	-22	MET	-	initiating methionine	UNP A0A2I4PGZ0
B	-21	GLY	-	expression tag	UNP A0A2I4PGZ0

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-20	SER	-	expression tag	UNP A0A2I4PGZ0
B	-19	SER	-	expression tag	UNP A0A2I4PGZ0
B	-18	HIS	-	expression tag	UNP A0A2I4PGZ0
B	-17	HIS	-	expression tag	UNP A0A2I4PGZ0
B	-16	HIS	-	expression tag	UNP A0A2I4PGZ0
B	-15	HIS	-	expression tag	UNP A0A2I4PGZ0
B	-14	HIS	-	expression tag	UNP A0A2I4PGZ0
B	-13	HIS	-	expression tag	UNP A0A2I4PGZ0
B	-12	SER	-	expression tag	UNP A0A2I4PGZ0
B	-11	SER	-	expression tag	UNP A0A2I4PGZ0
B	-10	GLY	-	expression tag	UNP A0A2I4PGZ0
B	-9	LEU	-	expression tag	UNP A0A2I4PGZ0
B	-8	VAL	-	expression tag	UNP A0A2I4PGZ0
B	-7	PRO	-	expression tag	UNP A0A2I4PGZ0
B	-6	ARG	-	expression tag	UNP A0A2I4PGZ0
B	-5	GLY	-	expression tag	UNP A0A2I4PGZ0
B	-4	SER	-	expression tag	UNP A0A2I4PGZ0
B	-3	HIS	-	expression tag	UNP A0A2I4PGZ0
B	-2	MET	-	expression tag	UNP A0A2I4PGZ0
B	-1	LEU	-	expression tag	UNP A0A2I4PGZ0
B	0	GLU	-	expression tag	UNP A0A2I4PGZ0

- Molecule 2 is alpha-D-glucopyranose (three-letter code: GLC) (formula: C₆H₁₂O₆).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			12	6	6		

Continued on next page...

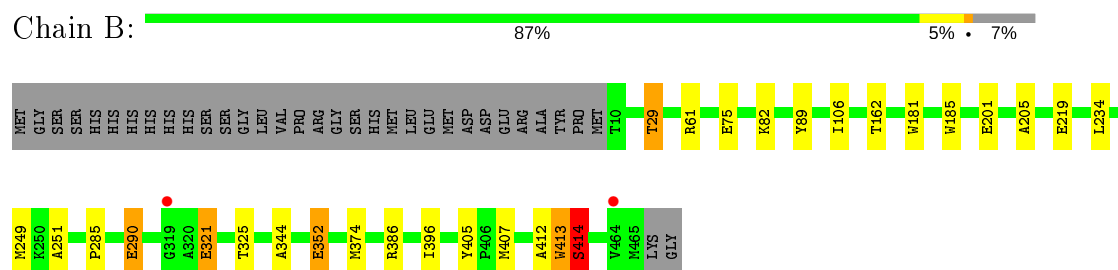
Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	B	1	Total	C	O	0	0
			12	6	6		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	195	Total	O	0	0
			195	195		
3	B	246	Total	O	0	0
			246	246		

- Molecule 1: Beta-glucosidase



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	59.03Å 84.59Å 184.73Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.78 – 2.11 49.78 – 2.11	Depositor EDS
% Data completeness (in resolution range)	96.0 (49.78-2.11) 94.5 (49.78-2.11)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.07	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.32 (at 2.12Å)	Xtriage
Refinement program	PHENIX 1.8_1069	Depositor
R, R_{free}	0.181 , 0.230 0.184 , 0.233	Depositor DCC
R_{free} test set	2560 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å ²)	34.0	Xtriage
Anisotropy	0.303	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 38.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	7647	wwPDB-VP
Average B, all atoms (Å ²)	19.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.36% 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: GLC

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	0.45	1/3695 (0.0%)	0.57	5/5026 (0.1%)
1	B	0.57	5/3708 (0.1%)	0.65	4/5042 (0.1%)
All	All	0.52	6/7403 (0.1%)	0.61	9/10068 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	B	2	0

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	414	SER	CA-CB	-14.97	1.30	1.52
1	B	414	SER	CA-C	-8.31	1.31	1.52
1	B	413	TRP	CA-CB	-8.13	1.36	1.53
1	B	413	TRP	N-CA	-7.84	1.30	1.46
1	A	413	TRP	CB-CG	7.03	1.62	1.50

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	414	SER	CB-CA-C	18.86	145.94	110.10
1	B	413	TRP	N-CA-CB	15.53	138.56	110.60
1	A	413	TRP	N-CA-CB	7.34	123.81	110.60
1	A	413	TRP	O-C-N	6.59	133.24	122.70
1	A	413	TRP	CA-C-N	-6.22	103.52	117.20

All (2) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	B	413	TRP	CA
1	B	414	SER	CA

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	3584	0	3388	22	0
1	B	3598	0	3400	14	0
2	A	12	0	11	0	0
2	B	12	0	11	0	0
3	A	195	0	0	1	0
3	B	246	0	0	2	0
All	All	7647	0	6810	34	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.

The worst 5 of 34 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:13:LYS:HE2	1:A:13:LYS:HA	1.30	1.08
1:A:13:LYS:CE	1:A:13:LYS:HA	2.10	0.82
1:A:61:ARG:NH2	1:B:201:GLU:OE1	2.20	0.75
1:A:18:ARG:NH2	1:A:399:ASP:OD1	2.20	0.71
1:B:386:ARG:NH2	3:B:610:HOH:O	2.28	0.61

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	451/490 (92%)	433 (96%)	16 (4%)	2 (0%)	34	32
1	B	454/490 (93%)	439 (97%)	13 (3%)	2 (0%)	34	32
All	All	905/980 (92%)	872 (96%)	29 (3%)	4 (0%)	34	32

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	8	PRO
1	A	9	MET
1	B	414	SER
1	B	413	TRP

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	362/392 (92%)	352 (97%)	10 (3%)	43	46
1	B	363/392 (93%)	353 (97%)	10 (3%)	43	46
All	All	725/784 (92%)	705 (97%)	20 (3%)	43	46

5 of 20 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	437	GLU
1	B	29	THR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	321	GLU
1	A	374	MET
1	A	426	ARG

Some 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 ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

2 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	GLC	B	501	-	12,12,12	1.23	1 (8%)	17,17,17	2.94	6 (35%)
2	GLC	A	501	-	12,12,12	1.03	0	17,17,17	2.12	7 (41%)

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	GLC	B	501	-	-	2/2/22/22	0/1/1/1
2	GLC	A	501	-	-	2/2/22/22	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	501	GLC	O5-C5	-3.05	1.37	1.44

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	501	GLC	O1-C1-O5	-7.90	86.67	110.38
2	B	501	GLC	O5-C5-C6	-5.97	91.58	106.44
2	A	501	GLC	C6-C5-C4	-4.18	103.20	113.00
2	B	501	GLC	O2-C2-C1	4.14	118.77	109.16
2	A	501	GLC	O5-C5-C6	-3.59	97.50	106.44

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	501	GLC	O5-C5-C6-O6
2	A	501	GLC	O5-C5-C6-O6
2	A	501	GLC	C4-C5-C6-O6
2	B	501	GLC	C4-C5-C6-O6

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, 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	453/490 (92%)	-0.35	2 (0%) 92 93	7, 20, 36, 52	0
1	B	456/490 (93%)	-0.47	2 (0%) 92 93	5, 16, 30, 50	0
All	All	909/980 (92%)	-0.41	4 (0%) 92 93	5, 18, 34, 52	0

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	414	SER	4.3
1	A	333	ASP	3.6
1	B	319	GLY	3.2
1	B	464	VAL	2.9

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
2	GLC	B	501	12/12	0.87	0.14	12,16,22,27	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
2	GLC	A	501	12/12	0.92	0.12	15,17,24,26	0

6.5 Other polymers [i](#)

There are no such residues in this entry.