Full wwPDB X-ray Structure Validation Report (i)

Jan 30, 2021 – 04:09 PM EST

PDB ID	:	3G1W
Title	:	Crystal structure of sugar ABC transporter (sugar-binding protein) from Bacillus halodurans
Authors	:	Zhang, Z.; Eswaramoorthy, S.; Swaminathan, S.; Burley, S.K.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on	:	2009-01-30
Resolution	:	2.02 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 (i) symbol.

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

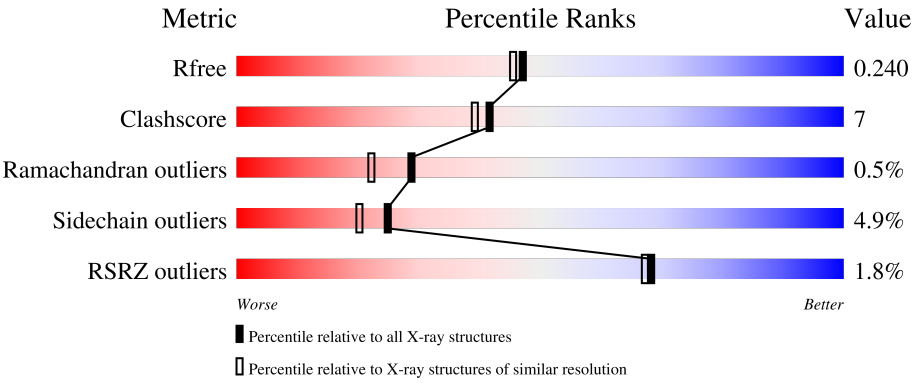
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.16
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA,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

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 2.02 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	10434 (2.04-2.00)
Clashscore	141614	11643 (2.04-2.00)
Ramachandran outliers	138981	11493 (2.04-2.00)
Sidechain outliers	138945	11492 (2.04-2.00)
RSRZ outliers	127900	10220 (2.04-2.00)

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	A	305	79% 15% • 5%
1	B	305	3% 74% 19% • 5%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 4699 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 Sugar ABC transporter.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	291	Total	C	N	O	S	Se	0	0	0
			2245	1417	370	451	1	6			
1	B	290	Total	C	N	O	S	Se	0	0	0
			2225	1403	366	449	1	6			

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	39	MSE	-	expression tag	UNP Q9K7B8
A	40	SER	-	expression tag	UNP Q9K7B8
A	41	LEU	-	expression tag	UNP Q9K7B8
A	336	GLU	-	expression tag	UNP Q9K7B8
A	337	GLY	-	expression tag	UNP Q9K7B8
A	338	HIS	-	expression tag	UNP Q9K7B8
A	339	HIS	-	expression tag	UNP Q9K7B8
A	340	HIS	-	expression tag	UNP Q9K7B8
A	341	HIS	-	expression tag	UNP Q9K7B8
A	342	HIS	-	expression tag	UNP Q9K7B8
A	343	HIS	-	expression tag	UNP Q9K7B8
B	39	MSE	-	expression tag	UNP Q9K7B8
B	40	SER	-	expression tag	UNP Q9K7B8
B	41	LEU	-	expression tag	UNP Q9K7B8
B	336	GLU	-	expression tag	UNP Q9K7B8
B	337	GLY	-	expression tag	UNP Q9K7B8
B	338	HIS	-	expression tag	UNP Q9K7B8
B	339	HIS	-	expression tag	UNP Q9K7B8
B	340	HIS	-	expression tag	UNP Q9K7B8
B	341	HIS	-	expression tag	UNP Q9K7B8
B	342	HIS	-	expression tag	UNP Q9K7B8
B	343	HIS	-	expression tag	UNP Q9K7B8

- Molecule 2 is water.

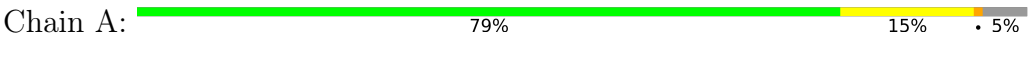
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	130	Total O 130 130	0	0
2	B	99	Total O 99 99	0	0

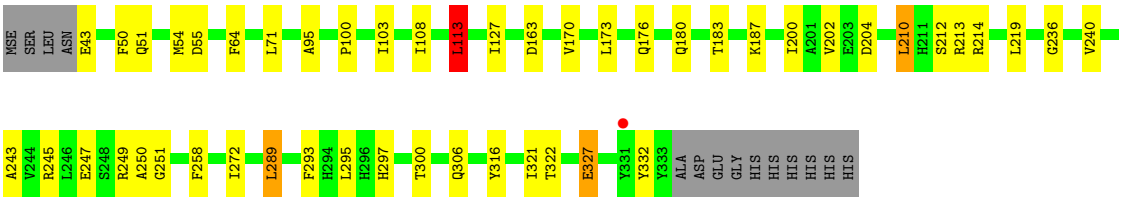
3 Residue-property plots (i)

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: Sugar ABC transporter

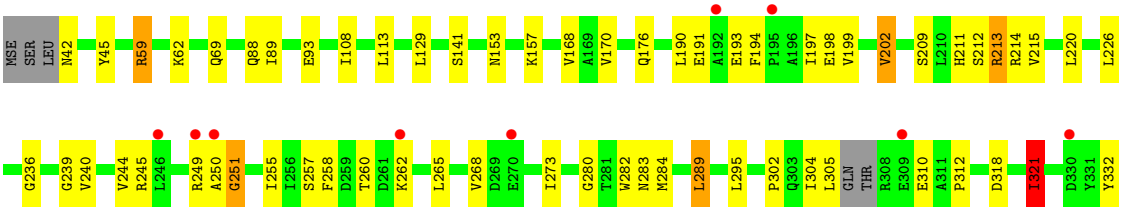
Chain A:





• Molecule 1: Sugar ABC transporter

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	60.46Å 87.86Å 113.58Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	33.92 – 2.02 33.92 – 2.02	Depositor EDS
% Data completeness (in resolution range)	100.0 (33.92-2.02) 87.2 (33.92-2.02)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	0.11	Depositor
$< I/\sigma(I) >$ [1]	1.96 (at 2.03Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.193 , 0.237 0.196 , 0.240	Depositor DCC
R_{free} test set	1760 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	27.4	Xtriage
Anisotropy	0.240	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 49.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	4699	wwPDB-VP
Average B, all atoms (Å^2)	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.94% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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

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	1.01	1/2284 (0.0%)	0.92	3/3102 (0.1%)
1	B	0.93	1/2263 (0.0%)	0.87	5/3073 (0.2%)
All	All	0.97	2/4547 (0.0%)	0.90	8/6175 (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	A	0	1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	45	TYR	CD1-CE1	6.76	1.49	1.39
1	A	64	PHE	CB-CG	5.84	1.61	1.51

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	113	LEU	CB-CG-CD2	-8.76	96.10	111.00
1	A	113	LEU	CA-CB-CG	-8.44	95.89	115.30
1	A	289	LEU	CA-CB-CG	-7.66	97.68	115.30
1	B	289	LEU	CA-CB-CG	-6.07	101.35	115.30
1	B	59	ARG	NE-CZ-NH2	5.98	123.29	120.30
1	B	321	ILE	CG1-CB-CG2	-5.76	98.73	111.40
1	B	59	ARG	NE-CZ-NH1	-5.17	117.72	120.30
1	A	204	ASP	CB-CG-OD2	-5.04	113.77	118.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	250	ALA	Peptide

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	A	2245	0	2155	29	1
1	B	2225	0	2116	35	1
2	A	130	0	0	2	0
2	B	99	0	0	1	0
All	All	4699	0	4271	61	1

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 7.

All (61) 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:304:ILE:HG22	1:B:305:LEU:H	1.47	0.79
1:A:176:GLN:O	1:A:180:GLN:HG3	1.86	0.74
1:B:211:HIS:O	1:B:215:VAL:HG23	1.89	0.72
1:B:302:PRO:HB2	1:B:310:GLU:HB3	1.72	0.71
1:B:153:ASN:HD22	1:B:321:ILE:HG22	1.55	0.70
1:B:260:THR:HG23	1:B:265:LEU:HD11	1.74	0.69
1:B:260:THR:CG2	1:B:265:LEU:HD11	2.24	0.68
1:B:304:ILE:HG22	1:B:305:LEU:N	2.12	0.65
1:B:260:THR:O	1:B:260:THR:HG22	1.98	0.64
1:B:153:ASN:HD22	1:B:321:ILE:CG2	2.11	0.63
1:A:210:LEU:HD13	1:A:214:ARG:HH12	1.63	0.63
1:A:251:GLY:HA2	2:A:368:HOH:O	2.03	0.57
1:B:260:THR:HG23	1:B:265:LEU:HD21	1.85	0.57
1:A:55:ASP:HB3	2:A:425:HOH:O	2.04	0.57
1:B:295:LEU:HD21	1:B:312:PRO:HG2	1.87	0.56
1:B:249:ARG:O	1:B:251:GLY:N	2.31	0.55
1:A:71:LEU:HD12	1:A:289:LEU:HD23	1.87	0.55
1:B:168:VAL:O	1:B:199:VAL:HA	2.08	0.53
1:A:43:GLU:OE2	1:A:297:HIS:HE1	1.90	0.53

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:213:ARG:HG3	1:B:213:ARG:HH11	1.75	0.50
1:A:51:GLN:HB3	1:A:54:MSE:HE3	1.95	0.49
1:B:153:ASN:ND2	1:B:321:ILE:HG22	2.26	0.49
1:B:220:LEU:HD23	1:B:226:LEU:HD23	1.94	0.49
1:B:244:VAL:HG21	1:B:255:ILE:HD11	1.95	0.49
1:A:210:LEU:HD13	1:A:214:ARG:NH1	2.26	0.48
1:A:245:ARG:HD3	1:A:272:ILE:HD13	1.95	0.48
1:A:187:LYS:HB3	1:A:187:LYS:HZ2	1.78	0.48
1:A:316:TYR:HB3	1:B:318:ASP:HB3	1.96	0.48
1:A:108:ILE:HG22	1:A:113:LEU:HD13	1.95	0.48
1:B:213:ARG:HG2	1:B:239:GLY:O	2.14	0.48
1:A:95:ALA:O	1:A:100:PRO:HD3	2.15	0.47
1:A:321:ILE:O	1:A:321:ILE:CG2	2.62	0.47
1:B:302:PRO:CB	1:B:310:GLU:HB3	2.42	0.47
1:A:50:PHE:O	1:A:51:GLN:C	2.53	0.46
1:B:170:VAL:HB	1:B:202:VAL:HB	1.96	0.46
1:A:187:LYS:HB3	1:A:187:LYS:NZ	2.31	0.46
1:A:247:GLU:HG2	1:A:249:ARG:CD	2.46	0.46
1:A:322:THR:HG21	1:A:332:TYR:OH	2.15	0.46
1:B:89:ILE:O	1:B:93:GLU:HG3	2.16	0.46
1:A:289:LEU:O	1:A:289:LEU:HG	2.12	0.45
1:A:213:ARG:HG3	1:A:243:ALA:HB2	1.99	0.44
1:B:211:HIS:HA	1:B:214:ARG:NH1	2.33	0.44
1:A:236:GLY:O	1:A:240:VAL:HG23	2.18	0.44
1:B:212:SER:OG	1:B:236:GLY:HA2	2.17	0.44
1:A:71:LEU:HD13	1:A:293:PHE:CG	2.53	0.43
1:B:280:GLY:O	1:B:283:ASN:HB2	2.19	0.43
1:A:170:VAL:HB	1:A:202:VAL:HG13	2.00	0.43
1:B:42:ASN:HA	2:B:354:HOH:O	2.18	0.43
1:A:183:THR:O	1:A:187:LYS:HG3	2.19	0.42
1:B:268:VAL:HG22	1:B:273:ILE:HG22	2.01	0.42
1:A:200:ILE:HD12	1:A:200:ILE:HG23	1.60	0.42
1:A:300:THR:HA	1:B:282:TRP:CH2	2.54	0.42
1:A:103:ILE:O	1:A:127:ILE:HA	2.20	0.42
1:A:212:SER:OG	1:A:236:GLY:HA2	2.20	0.42
1:B:236:GLY:O	1:B:240:VAL:HG23	2.20	0.41
1:B:260:THR:HG21	1:B:332:TYR:HB3	2.02	0.41
1:B:190:LEU:HD13	1:B:197:ILE:CG2	2.50	0.41
1:A:306:GLN:HE22	1:B:157:LYS:NZ	2.19	0.41
1:B:88:GLN:OE1	1:B:108:ILE:HG22	2.20	0.41
1:B:193:GLU:HB3	1:B:194:PHE:CD2	2.56	0.40

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:280:GLY:O	1:B:284:MSE:HG3	2.21	0.40

All (1) 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:327:GLU:OE2	1:B:262:LYS:NZ[2_665]	1.57	0.63

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	289/305 (95%)	280 (97%)	9 (3%)	0	100 100
1	B	286/305 (94%)	273 (96%)	10 (4%)	3 (1%)	15 9
All	All	575/610 (94%)	553 (96%)	19 (3%)	3 (0%)	29 22

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	250	ALA
1	B	251	GLY
1	B	141	SER

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	235/240 (98%)	227 (97%)	8 (3%)	37	35
1	B	231/240 (96%)	216 (94%)	15 (6%)	17	11
All	All	466/480 (97%)	443 (95%)	23 (5%)	25	20

All (23) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	113	LEU
1	A	163	ASP
1	A	173	LEU
1	A	210	LEU
1	A	219	LEU
1	A	258	PHE
1	A	295	LEU
1	A	327	GLU
1	B	59	ARG
1	B	62	LYS
1	B	69	GLN
1	B	129	LEU
1	B	176	GLN
1	B	191	GLU
1	B	198	GLU
1	B	202	VAL
1	B	209	SER
1	B	213	ARG
1	B	245	ARG
1	B	257	SER
1	B	258	PHE
1	B	289	LEU
1	B	321	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (12) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	225	ASN
1	A	279	GLN
1	A	294	HIS
1	A	297	HIS
1	A	306	GLN
1	B	42	ASN
1	B	153	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	176	GLN
1	B	211	HIS
1	B	218	GLN
1	B	279	GLN
1	B	297	HIS

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

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

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	<RSRZ>	#RSRZ>2			OWAB(Å^2)	Q<0.9
1	A	285/305 (93%)	-0.13	1 (0%)	92	92	18, 28, 39, 50	0
1	B	284/305 (93%)	0.25	9 (3%)	47	47	17, 31, 48, 64	0
All	All	569/610 (93%)	0.06	10 (1%)	68	67	17, 29, 46, 64	0

All (10) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	192	ALA	4.3
1	A	331	TYR	3.7
1	B	262	LYS	3.6
1	B	195	PRO	3.4
1	B	309	GLU	3.2
1	B	246	LEU	3.2
1	B	250	ALA	3.1
1	B	330	ASP	2.5
1	B	270	GLU	2.2
1	B	249	ARG	2.0

6.2 Non-standard residues in protein, DNA, RNA chains

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

6.3 Carbohydrates

There are no monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.