



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 3, 2022 – 12:23 PM EST

PDB ID : 7LHA  
Title : Structure of the Exo-L-galactose-6-sulfatase BuS1\_11 from *Bacteroides uniformis*  
Authors : Robb, C.S.; Boraston, A.B.  
Deposited on : 2021-01-21  
Resolution : 1.95 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](mailto: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
Xtriage (Phenix)	:	1.13
EDS	:	2.26
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)	:	2.26

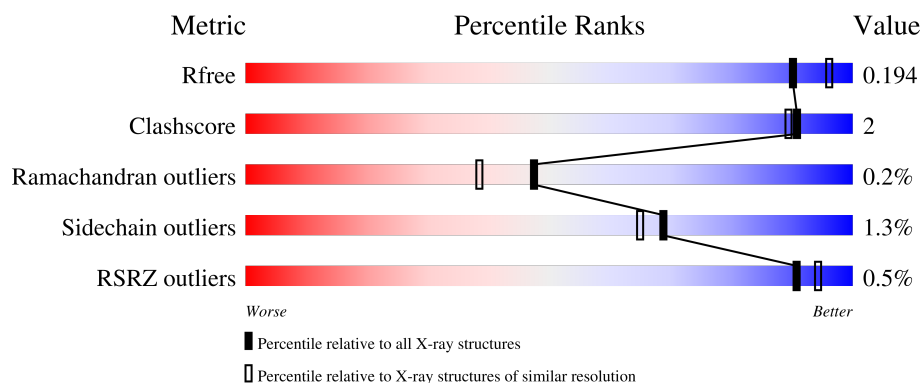
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*



The reported resolution of this entry is 1.95 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	2580 (1.96-1.96)
Clashscore	141614	2705 (1.96-1.96)
Ramachandran outliers	138981	2678 (1.96-1.96)
Sidechain outliers	138945	2678 (1.96-1.96)
RSRZ outliers	127900	2539 (1.96-1.96)

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  $\geq 3$ , 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	533	
1	B	533	

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 9126 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 Exo-L-galactose-6-sulfatase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	501	Total	C	N	O	S	0	1	0
			4104	2617	699	762	26			
1	B	499	Total	C	N	O	S	0	1	0
			4074	2597	690	761	26			

There are 66 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	3	MET	-	initiating methionine	UNP A0A4Y1VMZ7
A	4	GLY	-	expression tag	UNP A0A4Y1VMZ7
A	5	SER	-	expression tag	UNP A0A4Y1VMZ7
A	6	SER	-	expression tag	UNP A0A4Y1VMZ7
A	7	HIS	-	expression tag	UNP A0A4Y1VMZ7
A	8	HIS	-	expression tag	UNP A0A4Y1VMZ7
A	9	HIS	-	expression tag	UNP A0A4Y1VMZ7
A	10	HIS	-	expression tag	UNP A0A4Y1VMZ7
A	11	HIS	-	expression tag	UNP A0A4Y1VMZ7
A	12	HIS	-	expression tag	UNP A0A4Y1VMZ7
A	13	SER	-	expression tag	UNP A0A4Y1VMZ7
A	14	SER	-	expression tag	UNP A0A4Y1VMZ7
A	15	GLY	-	expression tag	UNP A0A4Y1VMZ7
A	16	LEU	-	expression tag	UNP A0A4Y1VMZ7
A	17	VAL	-	expression tag	UNP A0A4Y1VMZ7
A	18	PRO	-	expression tag	UNP A0A4Y1VMZ7
A	19	ARG	-	expression tag	UNP A0A4Y1VMZ7
A	20	GLY	-	expression tag	UNP A0A4Y1VMZ7
A	21	SER	-	expression tag	UNP A0A4Y1VMZ7
A	22	HIS	-	expression tag	UNP A0A4Y1VMZ7
A	23	MET	-	expression tag	UNP A0A4Y1VMZ7
A	24	ALA	-	expression tag	UNP A0A4Y1VMZ7
A	25	GLN	-	expression tag	UNP A0A4Y1VMZ7
A	26	GLU	-	expression tag	UNP A0A4Y1VMZ7
A	27	LYS	-	expression tag	UNP A0A4Y1VMZ7

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	28	PRO	-	expression tag	UNP A0A4Y1VMZ7
A	29	MET	-	expression tag	UNP A0A4Y1VMZ7
A	30	ASN	-	expression tag	UNP A0A4Y1VMZ7
A	31	VAL	-	expression tag	UNP A0A4Y1VMZ7
A	32	ILE	-	expression tag	UNP A0A4Y1VMZ7
A	33	PHE	-	expression tag	UNP A0A4Y1VMZ7
A	34	ILE	-	expression tag	UNP A0A4Y1VMZ7
A	392	LYS	GLU	conflict	UNP A0A4Y1VMZ7
B	3	MET	-	initiating methionine	UNP A0A4Y1VMZ7
B	4	GLY	-	expression tag	UNP A0A4Y1VMZ7
B	5	SER	-	expression tag	UNP A0A4Y1VMZ7
B	6	SER	-	expression tag	UNP A0A4Y1VMZ7
B	7	HIS	-	expression tag	UNP A0A4Y1VMZ7
B	8	HIS	-	expression tag	UNP A0A4Y1VMZ7
B	9	HIS	-	expression tag	UNP A0A4Y1VMZ7
B	10	HIS	-	expression tag	UNP A0A4Y1VMZ7
B	11	HIS	-	expression tag	UNP A0A4Y1VMZ7
B	12	HIS	-	expression tag	UNP A0A4Y1VMZ7
B	13	SER	-	expression tag	UNP A0A4Y1VMZ7
B	14	SER	-	expression tag	UNP A0A4Y1VMZ7
B	15	GLY	-	expression tag	UNP A0A4Y1VMZ7
B	16	LEU	-	expression tag	UNP A0A4Y1VMZ7
B	17	VAL	-	expression tag	UNP A0A4Y1VMZ7
B	18	PRO	-	expression tag	UNP A0A4Y1VMZ7
B	19	ARG	-	expression tag	UNP A0A4Y1VMZ7
B	20	GLY	-	expression tag	UNP A0A4Y1VMZ7
B	21	SER	-	expression tag	UNP A0A4Y1VMZ7
B	22	HIS	-	expression tag	UNP A0A4Y1VMZ7
B	23	MET	-	expression tag	UNP A0A4Y1VMZ7
B	24	ALA	-	expression tag	UNP A0A4Y1VMZ7
B	25	GLN	-	expression tag	UNP A0A4Y1VMZ7
B	26	GLU	-	expression tag	UNP A0A4Y1VMZ7
B	27	LYS	-	expression tag	UNP A0A4Y1VMZ7
B	28	PRO	-	expression tag	UNP A0A4Y1VMZ7
B	29	MET	-	expression tag	UNP A0A4Y1VMZ7
B	30	ASN	-	expression tag	UNP A0A4Y1VMZ7
B	31	VAL	-	expression tag	UNP A0A4Y1VMZ7
B	32	ILE	-	expression tag	UNP A0A4Y1VMZ7
B	33	PHE	-	expression tag	UNP A0A4Y1VMZ7
B	34	ILE	-	expression tag	UNP A0A4Y1VMZ7
B	392	LYS	GLU	conflict	UNP A0A4Y1VMZ7

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total 1	Ca 1	0	0
2	B	1	Total 1	Ca 1	0	0

- Molecule 3 is NICKEL (II) ION (three-letter code: NI) (formula: Ni).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	2	Total 2	Ni 2	0	0
3	B	2	Total 2	Ni 2	0	0

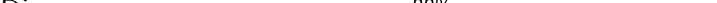
- Molecule 4 is water.

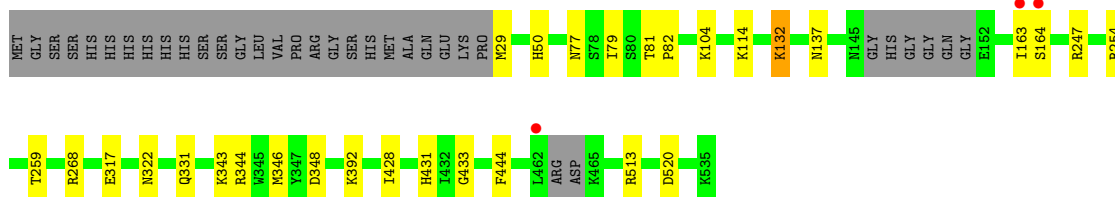
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	493	Total 493	O 493	0	0
4	B	449	Total 449	O 449	0	0



- Molecule 1: Exo-L-galactose-6-sulfatase

S243	MET
F256	GLY
Y274	SER
K283	HIS
R291	HIS
G329	HIS
D330	SER
Q331	GLY
K343	LEU
D348	PHE
K361	ARG
E451	GLY
K465	SER
H468	HIS
E478	MET
K535	ALA
	GLN
	GUJ
	LYS
	M29
	D37
	I79
	S90
	T81
	P82
	A124
	N145
	GLY
	HIS
	GLY
	GLY
	GLN
	GLY
	E152
	Y153
	K170
	L203
	P210
	W213
	F214
	E215
	E225
	D226

Chain B:  88% 5% 6%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	89.41Å 89.41Å 264.79Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.60 – 1.95 40.60 – 1.95	Depositor EDS
% Data completeness (in resolution range)	100.0 (40.60-1.95) 100.0 (40.60-1.95)	Depositor EDS
$R_{merge}$	0.17	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.44 (at 1.95Å)	Xtriage
Refinement program	PHENIX 1.13_2998	Depositor
R, $R_{free}$	0.148 , 0.194 0.148 , 0.194	Depositor DCC
$R_{free}$ test set	3978 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	16.0	Xtriage
Anisotropy	0.031	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 51.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	9126	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	18.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.59% 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, 0.333 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: CA, NI

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	0.47	0/4214	0.59	0/5700
1	B	0.44	0/4183	0.59	0/5663
All	All	0.46	0/8397	0.59	0/11363

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	4104	0	3935	15	0
1	B	4074	0	3874	12	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	2	0	0	0	0
3	B	2	0	0	0	0
4	A	493	0	0	4	0
4	B	449	0	0	0	0
All	All	9126	0	7809	27	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 (27) 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:29:MET:HE2	1:B:322:ASN:HA	1.54	0.89
1:A:170:LYS:NZ	4:A:708:HOH:O	2.36	0.58
1:A:283:LYS:NZ	4:A:714:HOH:O	2.39	0.54
1:A:225:GLU:HG3	1:A:226:ASP:N	2.24	0.51
1:A:451:GLU:OE2	4:A:701:HOH:O	2.20	0.49
1:A:37:ASP:O	1:A:210:PRO:HD2	2.13	0.48
1:A:124:ALA:HB3	1:A:203:LEU:HD12	1.96	0.47
1:B:431:HIS:HB3	1:B:444:PHE:CD2	2.50	0.46
1:B:513:ARG:NH2	1:B:520:ASP:OD2	2.35	0.46
1:A:274:TYR:CZ	1:A:291:ARG:HG2	2.52	0.45
1:A:243:SER:HB3	1:A:478:GLU:HG2	1.99	0.45
1:B:104:LYS:HB3	1:B:132:LYS:HB2	1.99	0.45
1:B:247:ARG:HB3	1:B:254:ARG:HD3	1.99	0.44
1:A:465:LYS:HB3	1:A:465:LYS:HE2	1.84	0.44
1:A:81:THR:HB	1:A:82:PRO:HD3	1.99	0.43
1:A:274:TYR:CE2	1:A:291:ARG:HG2	2.53	0.43
1:A:153:TYR:CD2	1:A:215:GLU:HB2	2.54	0.43
1:A:361:LYS:HA	1:A:361:LYS:HD3	1.71	0.42
1:B:317:GLU:O	1:B:317:GLU:HG3	2.19	0.42
1:B:163:ILE:O	1:B:164:SER:HB2	2.19	0.42
1:B:346:MET:SD	1:B:433:GLY:HA3	2.60	0.42
1:B:259:THR:HG22	1:B:268:ARG:HB3	2.02	0.41
1:A:215:GLU:HA	4:A:1099:HOH:O	2.19	0.41
1:A:256:PHE:HA	1:A:468:HIS:CE1	2.55	0.41
1:B:81:THR:HB	1:B:82:PRO:HD3	2.02	0.41
1:B:344:ARG:HD2	1:B:428:ILE:HD12	2.03	0.40
1:B:114:LYS:HE2	1:B:137:ASN:ND2	2.36	0.40

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	498/533 (93%)	483 (97%)	15 (3%)	0	100	100
1	B	494/533 (93%)	480 (97%)	12 (2%)	2 (0%)	34	22
All	All	992/1066 (93%)	963 (97%)	27 (3%)	2 (0%)	47	38

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	132	LYS
1	B	77	ASN

### 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	439/475 (92%)	434 (99%)	5 (1%)	73	71
1	B	434/475 (91%)	428 (99%)	6 (1%)	67	62
All	All	873/950 (92%)	862 (99%)	11 (1%)	69	65

All (11) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	79	ILE
1	A	291	ARG
1	A	331	GLN
1	A	343	LYS
1	A	348	ASP
1	B	50	HIS
1	B	79	ILE
1	B	331	GLN
1	B	343	LYS
1	B	348	ASP
1	B	392	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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 monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 6 ligands modelled in this entry, 6 are monoatomic - leaving 0 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<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	501/533 (93%)	-0.37	2 (0%) 92 95	6, 13, 31, 49	0
1	B	499/533 (93%)	-0.27	3 (0%) 89 93	6, 15, 33, 60	0
All	All	1000/1066 (93%)	-0.32	5 (0%) 91 94	6, 14, 32, 60	0

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	163	ILE	3.2
1	A	329	GLY	2.1
1	A	213	MET	2.1
1	B	164	SER	2.1
1	B	462	LEU	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 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, 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	B-factors( $\text{\AA}^2$ )	Q<0.9
2	CA	A	601	1/1	0.99	0.05	25,25,25,25	0
2	CA	B	601	1/1	0.99	0.12	24,24,24,24	0
3	NI	A	602	1/1	1.00	0.04	22,22,22,22	0
3	NI	A	603	1/1	1.00	0.05	27,27,27,27	0
3	NI	B	602	1/1	1.00	0.06	24,24,24,24	0
3	NI	B	603	1/1	1.00	0.05	31,31,31,31	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.