



# Full wwPDB X-ray Structure Validation Report ⓘ

Aug 16, 2020 – 07:48 PM BST

PDB ID : 6ZJP  
Title : Cold-adapted beta-D-galactosidase from *Arthrobacter* sp. 32cB mutant E517Q  
Authors : Rutkiewicz, M.; Bujacz, A.; Bujacz, G.  
Deposited on : 2020-06-29  
Resolution : 1.85 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  
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

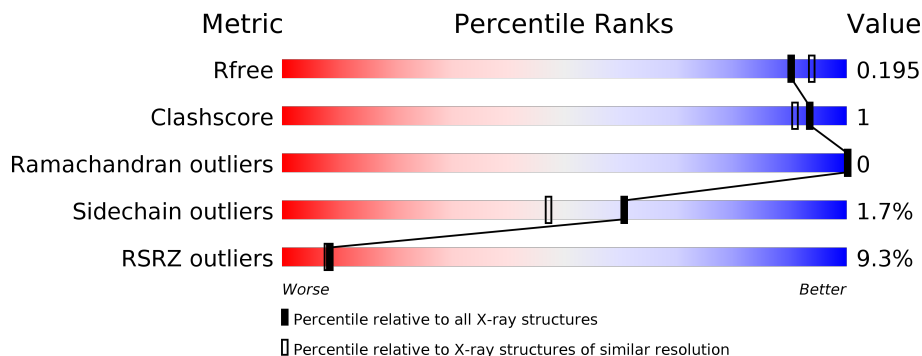
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

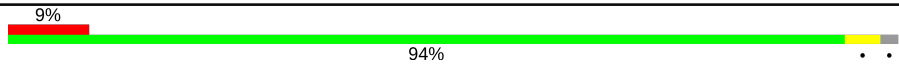
The reported resolution of this entry is 1.85 Å.

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	2469 (1.86-1.86)
Clashscore	141614	2625 (1.86-1.86)
Ramachandran outliers	138981	2592 (1.86-1.86)
Sidechain outliers	138945	2592 (1.86-1.86)
RSRZ outliers	127900	2436 (1.86-1.86)

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 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	1010	

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 8670 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-galactosidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	989	7613	4790	1361	1444	18	0	2	0

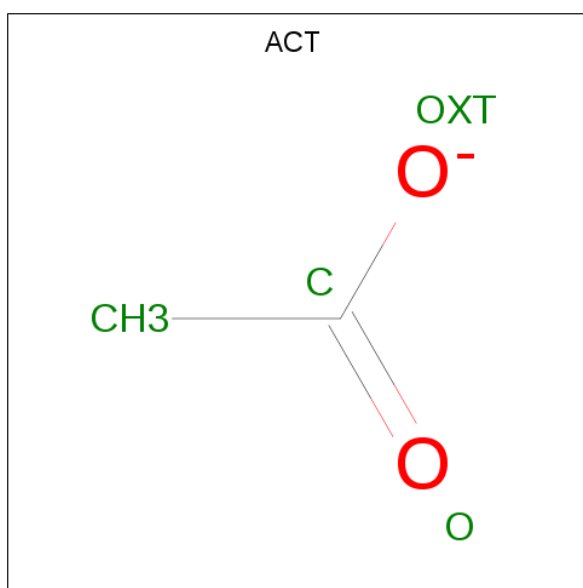
There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	517	GLN	GLU	engineered mutation	UNP A0A023UGN9

- Molecule 2 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	3	Total	Na	0	0
			3	3		

- Molecule 3 is ACETATE ION (three-letter code: ACT) (formula: C<sub>2</sub>H<sub>3</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 4 2 2	0	0

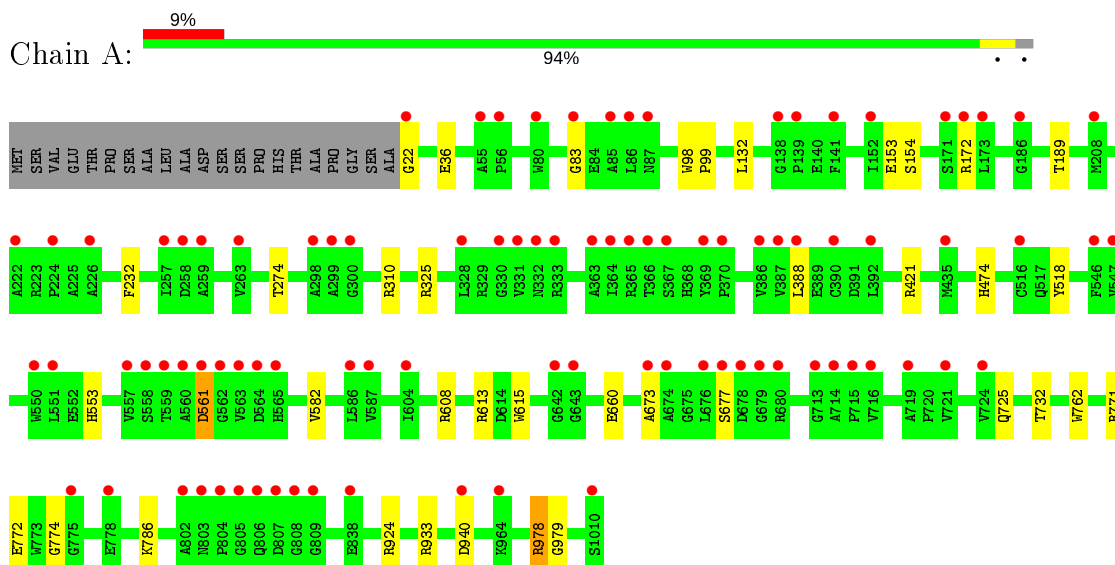
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1050	Total O 1050 1050	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: Beta-galactosidase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	138.00Å 138.00Å 127.22Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	43.55 – 1.85 46.77 – 1.85	Depositor EDS
% Data completeness (in resolution range)	99.0 (43.55-1.85) 99.2 (46.77-1.85)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.12 (at 1.84Å)	Xtrriage
Refinement program	PHENIX 1.12_2829	Depositor
R, $R_{free}$	0.162 , 0.189 0.169 , 0.195	Depositor DCC
$R_{free}$ test set	2100 reflections (1.77%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	30.9	Xtrriage
Anisotropy	0.544	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 55.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.017 for -h,-k,l	Xtrriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	8670	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	44.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.80% 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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: NA, ACT

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	0.31	0/7818	0.53	0/10656

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 [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	7613	0	7306	21	0
2	A	3	0	0	0	0
3	A	4	0	3	1	0
4	A	1050	0	0	7	0
All	All	8670	0	7309	21	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 1.

All (21) 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:553:HIS:HB2	1:A:582:VAL:HG22	1.24	1.08
1:A:553:HIS:HB2	1:A:582:VAL:CG2	2.00	0.91
1:A:561:ASP:N	1:A:561:ASP:OD1	2.22	0.72
1:A:940:ASP:HB2	4:A:1922:HOH:O	1.89	0.71
1:A:22:GLY:N	4:A:1202:HOH:O	2.26	0.68
1:A:325[A]:ARG:NH2	4:A:1203:HOH:O	2.26	0.66
1:A:553:HIS:CB	1:A:582:VAL:HG22	2.16	0.62
1:A:421:ARG:HD2	3:A:1104:ACT:H2	1.84	0.59
1:A:786:LYS:NZ	4:A:1225:HOH:O	2.47	0.48
1:A:725:GLN:HG2	1:A:732:THR:HB	1.97	0.47
1:A:132:LEU:HD11	1:A:189:THR:CG2	2.46	0.45
1:A:153:GLU:HA	1:A:154:SER:HA	1.75	0.45
1:A:978:ARG:HG2	1:A:979:GLY:O	2.17	0.45
1:A:474:HIS:ND1	4:A:1208:HOH:O	2.36	0.44
1:A:615:TRP:CE3	1:A:673:ALA:HB2	2.52	0.44
1:A:772:GLU:C	1:A:774:GLY:H	2.20	0.43
1:A:36:GLU:OE2	4:A:1201:HOH:O	2.21	0.43
1:A:98:TRP:N	1:A:99:PRO:CD	2.82	0.42
1:A:274:THR:HG23	4:A:1665:HOH:O	2.18	0.42
1:A:83:GLY:HA2	1:A:132:LEU:HD21	2.03	0.41
1:A:608:ARG:NH2	1:A:660:GLU:OE2	2.54	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	989/1010 (98%)	960 (97%)	29 (3%)	0	<a href="#">100</a> <a href="#">100</a>

There are no Ramachandran outliers to report.



### 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	777/793 (98%)	764 (98%)	13 (2%)	60 47

All (13) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	172	ARG
1	A	232	PHE
1	A	310	ARG
1	A	388	LEU
1	A	518	TYR
1	A	561	ASP
1	A	613	ARG
1	A	677	SER
1	A	762	TRP
1	A	771	ARG
1	A	924	ARG
1	A	933	ARG
1	A	978	ARG

Some 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 4 ligands modelled in this entry, 3 are monoatomic - leaving 1 for Mogul analysis.

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
3	ACT	A	1104	-	1,3,3	4.79	1 (100%)	0,3,3	0.00	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	1104	ACT	CH3-C	4.79	1.54	1.48

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	1104	ACT	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

### 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<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	989/1010 (97%)	0.24	92 (9%) <b>8</b>   <b>8</b>	23, 40, 72, 131	0

All (92) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	559	THR	10.3
1	A	560	ALA	8.0
1	A	562	GLY	7.9
1	A	561	ASP	7.1
1	A	563	VAL	6.9
1	A	678	ASP	5.3
1	A	85	ALA	4.9
1	A	558	SER	4.9
1	A	677	SER	4.7
1	A	715	PRO	4.5
1	A	676	LEU	4.2
1	A	387	VAL	4.2
1	A	721	VAL	4.2
1	A	809	GLY	4.2
1	A	714	ALA	4.1
1	A	808	GLY	4.0
1	A	1010	SER	3.9
1	A	642	GLY	3.9
1	A	298	ALA	3.9
1	A	226	ALA	3.6
1	A	331	VAL	3.6
1	A	299	ALA	3.6
1	A	366	THR	3.6
1	A	674	ALA	3.5
1	A	328	LEU	3.5
1	A	719	ALA	3.5
1	A	673	ALA	3.5

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Res</b>	<b>Type</b>	<b>RSRZ</b>
1	A	804	PRO	3.5
1	A	547	VAL	3.5
1	A	716	VAL	3.5
1	A	546	PHE	3.4
1	A	87	ASN	3.3
1	A	139	PRO	3.3
1	A	138	GLY	3.2
1	A	388	LEU	3.2
1	A	557	VAL	3.2
1	A	364	ILE	3.2
1	A	369	TYR	3.1
1	A	259	ALA	3.1
1	A	186	GLY	3.1
1	A	806	GLN	3.0
1	A	22	GLY	3.0
1	A	516	CYS	2.9
1	A	173	LEU	2.8
1	A	551	LEU	2.7
1	A	713	GLY	2.7
1	A	258	ASP	2.7
1	A	56	PRO	2.7
1	A	367	SER	2.7
1	A	838	GLU	2.6
1	A	55	ALA	2.6
1	A	724	VAL	2.5
1	A	370	PRO	2.5
1	A	679	GLY	2.5
1	A	86	LEU	2.5
1	A	224	PRO	2.4
1	A	964	LYS	2.4
1	A	564	ASP	2.4
1	A	775	GLY	2.4
1	A	802	ALA	2.4
1	A	586	LEU	2.4
1	A	365	ARG	2.4
1	A	80	TRP	2.3
1	A	171	SER	2.3
1	A	83	GLY	2.3
1	A	778	GLU	2.3
1	A	587	VAL	2.3
1	A	565	HIS	2.3
1	A	392	LEU	2.3

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	805	GLY	2.2
1	A	263	VAL	2.2
1	A	363	ALA	2.2
1	A	330	GLY	2.2
1	A	643	GLY	2.2
1	A	300	GLY	2.2
1	A	386	VAL	2.2
1	A	333	ARG	2.2
1	A	680	ARG	2.2
1	A	940	ASP	2.2
1	A	208	MET	2.1
1	A	152	ILE	2.1
1	A	141	PHE	2.1
1	A	390	CYS	2.1
1	A	435	MET	2.1
1	A	803	ASN	2.1
1	A	550	TRP	2.0
1	A	172	ARG	2.0
1	A	604	ILE	2.0
1	A	222	ALA	2.0
1	A	807	ASP	2.0
1	A	257	ILE	2.0
1	A	332	ASN	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.

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
-----	------	-------	-----	-------	------	-----	-----------------------------	-------

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
3	ACT	A	1104	4/4	0.88	0.14	35,37,40,51	0
2	NA	A	1101	1/1	0.96	0.11	44,44,44,44	0
2	NA	A	1102	1/1	0.97	0.08	42,42,42,42	0
2	NA	A	1103	1/1	0.99	0.16	22,22,22,22	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.