wwPDB EM Validation Summary Report

Feb 25, 2024 – 11:20 AM EST

PDB ID	:	6X1Q
EMDB ID	:	EMD-21995
Title	:	1.8 Angstrom resolution structure of b-galactosidase with a 200 kV cryoARM electron microscope
Authors	:	Merk, A.; Fukumura, T.; Zhu, X.; Darling, J.; Grisshammer, R.; Ognjenovic, J.; Subramaniam, S.
Deposited on	:	2020-05-19
Resolution	:	1.80 Å(reported)
Based on initial model	:	5A1A

This is a wwPDB EM 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/EMValidationReportHelp
with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

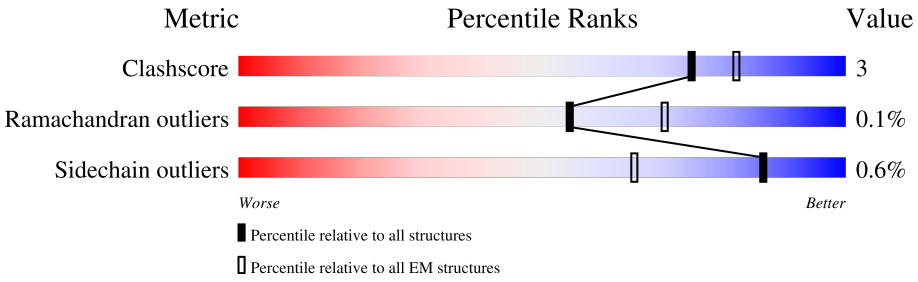
EMDB validation analysis	:	FAILED
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
MapQ	:	FAILED
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
ELECTRON MICROSCOPY

The reported resolution of this entry is 1.80 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)	EM structures (#Entries)
Clashscore	158937	4297
Ramachandran outliers	154571	4023
Sidechain outliers	154315	3826

The table below summarises the geometric issues observed across the polymeric chains and their fit to the map. The red, orange, yellow and green segments of the 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%

Mol	Chain	Length	Quality of chain
1	A	1021	91% 8% .
1	B	1021	92% 7% .
1	C	1021	91% 8% .
1	D	1021	92% 7% .

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 66138 atoms, of which 30520 are hydrogens and 0 are deuteriums.

In the tables below, 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						AltConf	Trace
1	A	1014	Total	C	H	N	O	S	0	0
			15626	5078	7630	1412	1469	37		
1	B	1014	Total	C	H	N	O	S	0	0
			15626	5078	7630	1412	1469	37		
1	C	1014	Total	C	H	N	O	S	0	0
			15626	5078	7630	1412	1469	37		
1	D	1014	Total	C	H	N	O	S	0	0
			15626	5078	7630	1412	1469	37		

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	334	VAL	GLU	conflict	UNP P00722
A	871	VAL	GLU	conflict	UNP P00722
B	334	VAL	GLU	conflict	UNP P00722
B	871	VAL	GLU	conflict	UNP P00722
C	334	VAL	GLU	conflict	UNP P00722
C	871	VAL	GLU	conflict	UNP P00722
D	334	VAL	GLU	conflict	UNP P00722
D	871	VAL	GLU	conflict	UNP P00722

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		AltConf
2	A	2	Total	Mg	0
			2	2	
2	B	2	Total	Mg	0
			2	2	
2	C	2	Total	Mg	0
			2	2	
2	D	2	Total	Mg	0
			2	2	

- Molecule 3 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms	AltConf
3	A	2	Total Na 2 2	0
3	B	2	Total Na 2 2	0
3	C	2	Total Na 2 2	0
3	D	2	Total Na 2 2	0

- Molecule 4 is water.

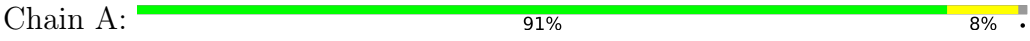
Mol	Chain	Residues	Atoms	AltConf
4	A	906	Total O 906 906	0
4	B	901	Total O 901 901	0
4	C	909	Total O 909 909	0
4	D	902	Total O 902 902	0

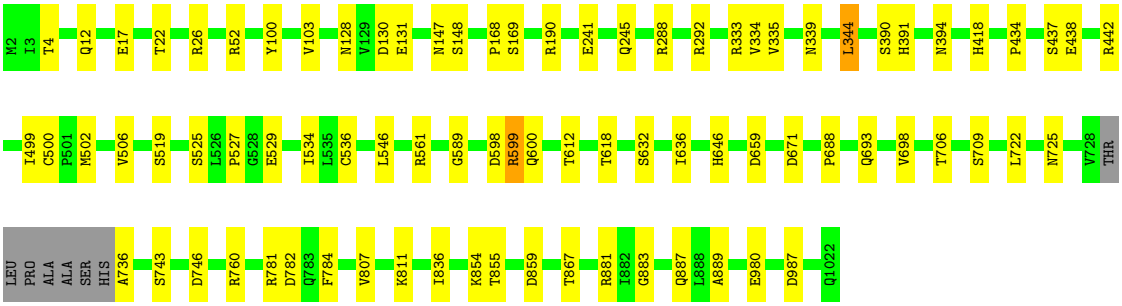
3 Residue-property plots (i)

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. 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. 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

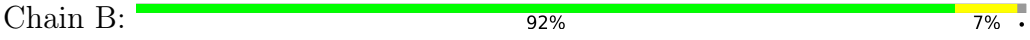
Chain A:





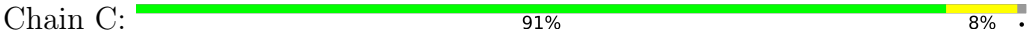
• Molecule 1: Beta-galactosidase

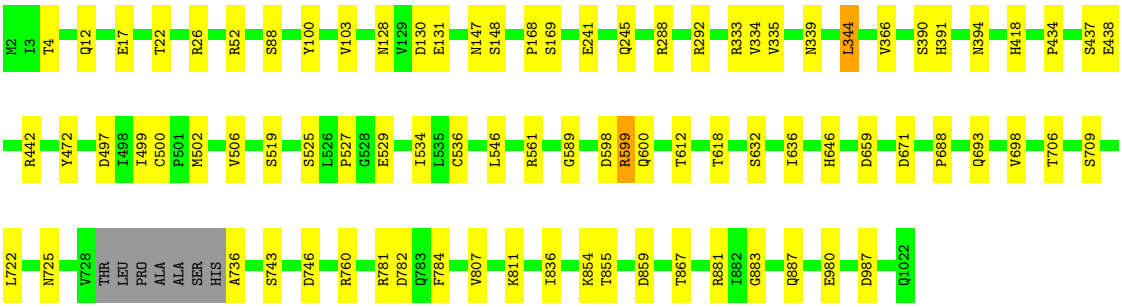
Chain B:



• Molecule 1: Beta-galactosidase

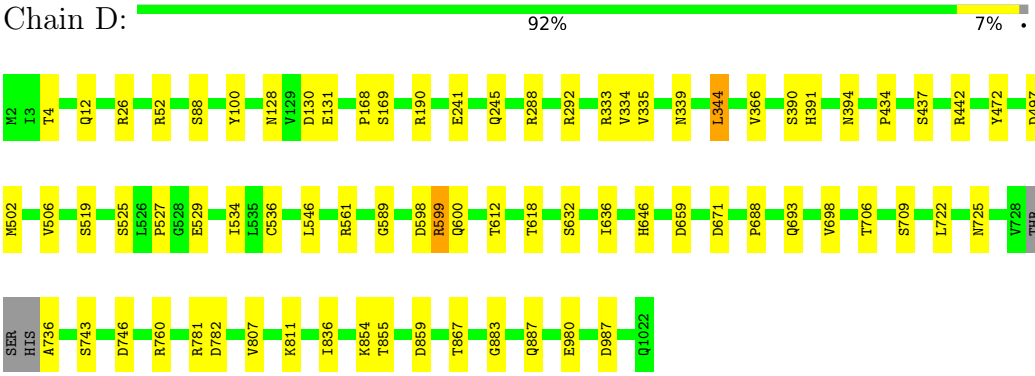
Chain C:





• Molecule 1: Beta-galactosidase

Chain D:



4 Experimental information (i)

Property	Value	Source
EM reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, D2	Depositor
Number of particles used	257202	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	JEOL CRYO ARM 200	Depositor
Voltage (kV)	200	Depositor
Electron dose ($e^-/Å^2$)	40	Depositor
Minimum defocus (nm)	-800	Depositor
Maximum defocus (nm)	-950	Depositor
Magnification	Not provided	
Image detector	GATAN K3 (6k x 4k)	Depositor

5 Model quality

5.1 Standard geometry

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

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	0.41	0/8234	0.52	0/11253
1	B	0.41	0/8234	0.52	0/11253
1	C	0.41	0/8234	0.52	0/11253
1	D	0.41	0/8234	0.52	0/11253
All	All	0.41	0/32936	0.52	0/45012

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	7996	7630	7518	54	0
1	B	7996	7630	7518	51	0
1	C	7996	7630	7518	54	0
1	D	7996	7630	7518	51	0
2	A	2	0	0	0	0
2	B	2	0	0	0	0
2	C	2	0	0	0	0
2	D	2	0	0	0	0
3	A	2	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	2	0	0	0	0
3	C	2	0	0	0	0
3	D	2	0	0	0	0
4	A	906	0	0	13	0
4	B	901	0	0	10	0
4	C	909	0	0	11	0
4	D	902	0	0	11	0
All	All	35618	30520	30072	198	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 3.

The worst 5 of 198 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:855:THR:HG1	1:B:867:THR:HG1	1.13	0.73
1:A:859:ASP:OD1	4:A:1202:HOH:O	2.08	0.71
1:C:859:ASP:OD1	4:C:1202:HOH:O	2.08	0.71
1:D:859:ASP:OD1	4:D:1202:HOH:O	2.08	0.70
1:B:859:ASP:OD1	4:B:1202:HOH:O	2.08	0.70

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 PDB entries followed by that with respect to all EM entries.

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	1010/1021 (99%)	991 (98%)	18 (2%)	1 (0%)	51 36
1	B	1010/1021 (99%)	991 (98%)	18 (2%)	1 (0%)	51 36
1	C	1010/1021 (99%)	991 (98%)	18 (2%)	1 (0%)	51 36
1	D	1010/1021 (99%)	991 (98%)	18 (2%)	1 (0%)	51 36

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	4040/4084 (99%)	3964 (98%)	72 (2%)	4 (0%)	54 36

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	599	ARG
1	B	599	ARG
1	C	599	ARG
1	D	599	ARG

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 PDB entries followed by that with respect to all EM entries.

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	827/873 (95%)	822 (99%)	5 (1%)	86 84
1	B	827/873 (95%)	822 (99%)	5 (1%)	86 84
1	C	827/873 (95%)	822 (99%)	5 (1%)	86 84
1	D	827/873 (95%)	822 (99%)	5 (1%)	86 84
All	All	3308/3492 (95%)	3288 (99%)	20 (1%)	86 84

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

Mol	Chain	Res	Type
1	C	632	SER
1	D	394	ASN
1	D	632	SER
1	D	546	LEU
1	B	344	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 28 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	262	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	D	945	ASN
1	C	817	GLN
1	D	817	GLN
1	C	761	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)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 16 ligands modelled in this entry, 16 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.