wwPDB X-ray Structure Validation Summary Report (i)

Aug 12, 2024 – 09:47 am BST

PDB ID	:	8QQH
Title	:	Structure of beta-galactosidase from Desulfurococcus amyloliticus
Authors	:	Samygina, V.R.; Kil, Y.; Sergeev, R.S.; Rychkov, G.N.
Deposited on	:	2023-10-04
Resolution	:	2.15 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 (i) 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 (i)) were used in the production of this report:

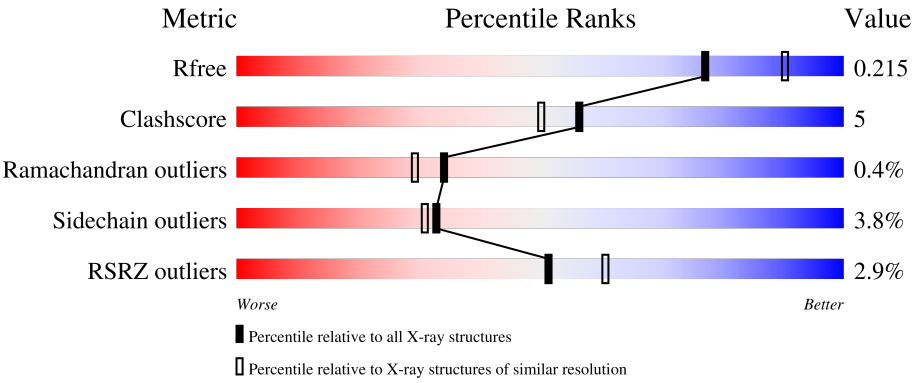
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.15 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	1479 (2.16-2.16)
Clashscore	141614	1585 (2.16-2.16)
Ramachandran outliers	138981	1560 (2.16-2.16)
Sidechain outliers	138945	1559 (2.16-2.16)
RSRZ outliers	127900	1456 (2.16-2.16)

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	AAA	988	2% 83% 15%
1	BBB	988	3% 86% 11%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 16160 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
1	AAA	971	Total	C	N	O	S	0	0	0
			7863	5071	1324	1446	22			
1	BBB	969	Total	C	N	O	S	0	0	0
			7846	5065	1319	1440	22			

There are 38 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	43	ASP	GLY	conflict	UNP B8D3P7
AAA	50	ASN	ASP	conflict	UNP B8D3P7
AAA	513	GLU	GLY	conflict	UNP B8D3P7
AAA	629	LYS	ARG	conflict	UNP B8D3P7
AAA	716	PRO	SER	conflict	UNP B8D3P7
AAA	751	ASP	GLY	conflict	UNP B8D3P7
AAA	976	LYS	-	expression tag	UNP B8D3P7
AAA	977	LEU	-	expression tag	UNP B8D3P7
AAA	978	ALA	-	expression tag	UNP B8D3P7
AAA	979	ALA	-	expression tag	UNP B8D3P7
AAA	980	ALA	-	expression tag	UNP B8D3P7
AAA	981	LEU	-	expression tag	UNP B8D3P7
AAA	982	GLU	-	expression tag	UNP B8D3P7
AAA	983	HIS	-	expression tag	UNP B8D3P7
AAA	984	HIS	-	expression tag	UNP B8D3P7
AAA	985	HIS	-	expression tag	UNP B8D3P7
AAA	986	HIS	-	expression tag	UNP B8D3P7
AAA	987	HIS	-	expression tag	UNP B8D3P7
AAA	988	HIS	-	expression tag	UNP B8D3P7
BBB	43	ASP	GLY	conflict	UNP B8D3P7
BBB	50	ASN	ASP	conflict	UNP B8D3P7
BBB	513	GLU	GLY	conflict	UNP B8D3P7
BBB	629	LYS	ARG	conflict	UNP B8D3P7
BBB	716	PRO	SER	conflict	UNP B8D3P7
BBB	751	ASP	GLY	conflict	UNP B8D3P7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
BBB	976	LYS	-	expression tag	UNP B8D3P7
BBB	977	LEU	-	expression tag	UNP B8D3P7
BBB	978	ALA	-	expression tag	UNP B8D3P7
BBB	979	ALA	-	expression tag	UNP B8D3P7
BBB	980	ALA	-	expression tag	UNP B8D3P7
BBB	981	LEU	-	expression tag	UNP B8D3P7
BBB	982	GLU	-	expression tag	UNP B8D3P7
BBB	983	HIS	-	expression tag	UNP B8D3P7
BBB	984	HIS	-	expression tag	UNP B8D3P7
BBB	985	HIS	-	expression tag	UNP B8D3P7
BBB	986	HIS	-	expression tag	UNP B8D3P7
BBB	987	HIS	-	expression tag	UNP B8D3P7
BBB	988	HIS	-	expression tag	UNP B8D3P7

- Molecule 2 is water.

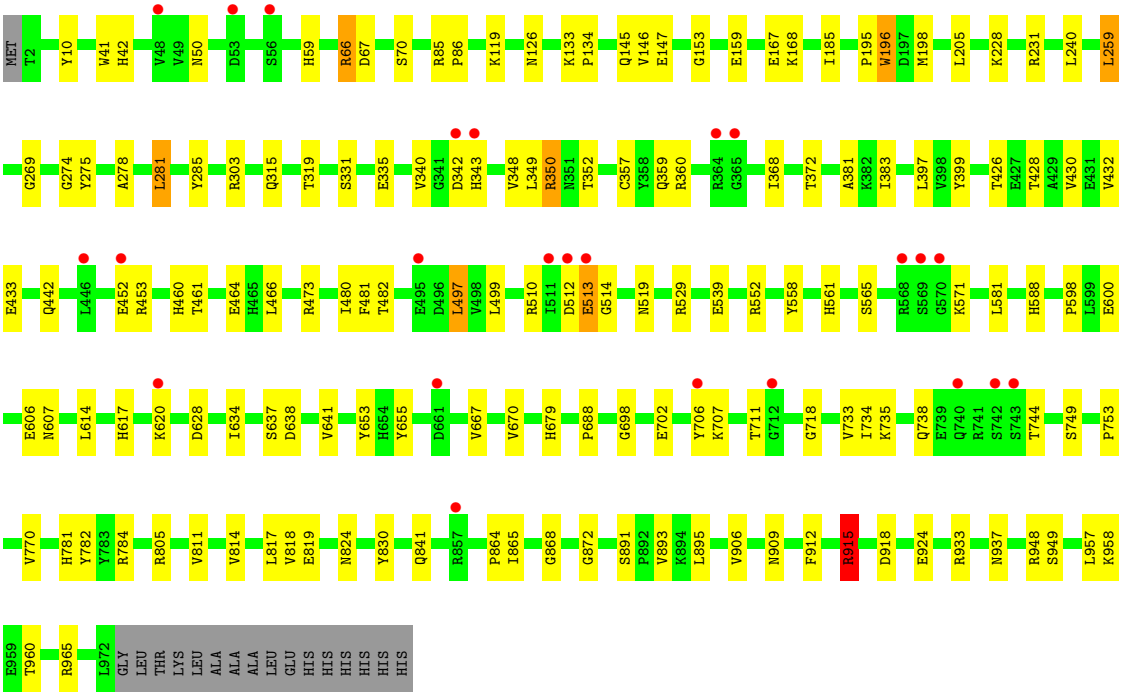
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	AAA	257	Total O 257 257	0	0
2	BBB	194	Total O 194 194	0	0

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.

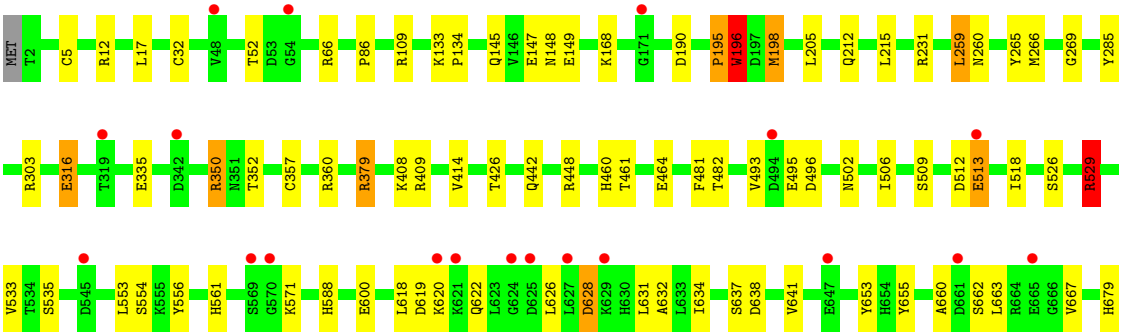
• Molecule 1: Beta-galactosidase

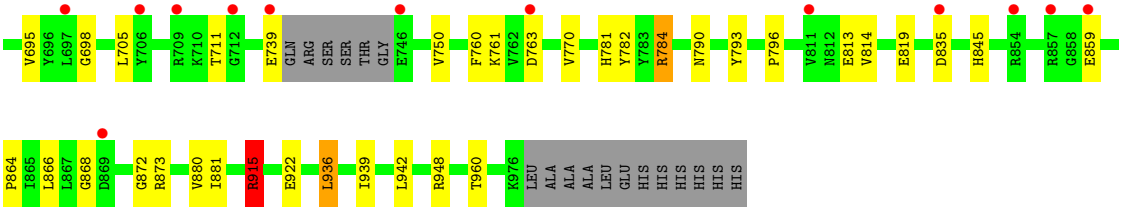
Chain AAA:



• Molecule 1: Beta-galactosidase

Chain BBB:




V695
Y696
L697
G698
L705
Y706
R709
K710
T711
G712
E739
GLN
ARG
SER
SER
THR
GLY
E746
V750
F760
K761
V762
D763
V770
H781
Y782
Y783
R784
N790
Y793
P796
V811
N812
E813
V814
E819
D835
H845
R854
R857
G858
E859
P864
I865
L866
L867
G868
D869
G872
R873
V880
I881
R915
E922
L936
I939
L942
R948
T960
K976
LEU
ALA
ALA
ALA
LEU
GLU
HIS
HIS
HIS
HIS
HIS
HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 2 21 21	Depositor
Cell constants a, b, c, α, β, γ	81.23Å 140.76Å 207.77Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	15.00 – 2.15 14.99 – 2.15	Depositor EDS
% Data completeness (in resolution range)	98.4 (15.00-2.15) 98.7 (14.99-2.15)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.15 (at 2.16Å)	Xtriage
Refinement program	REFMAC 5.8.0258	Depositor
R, R_{free}	0.165 , 0.211 0.175 , 0.215	Depositor DCC
R_{free} test set	6351 reflections (4.96%)	wwPDB-VP
Wilson B-factor (Å^2)	39.0	Xtriage
Anisotropy	0.023	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 41.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	16160	wwPDB-VP
Average B, all atoms (Å^2)	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.54% 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	AAA	0.72	0/8067	0.92	11/10968 (0.1%)
1	BBB	0.73	0/8049	0.92	11/10942 (0.1%)
All	All	0.72	0/16116	0.92	22/21910 (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	AAA	0	1
1	BBB	0	1
All	All	0	2

There are no bond length outliers.

The worst 5 of 22 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	BBB	12	ARG	NE-CZ-NH2	-12.40	114.10	120.30
1	BBB	12	ARG	NE-CZ-NH1	11.11	125.86	120.30
1	AAA	915	ARG	NE-CZ-NH2	-10.99	114.81	120.30
1	BBB	350	ARG	NE-CZ-NH2	-9.32	115.64	120.30
1	AAA	350	ARG	NE-CZ-NH2	-9.11	115.75	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	AAA	195	PRO	Peptide
1	BBB	195	PRO	Peptide

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	AAA	7863	0	7788	89	0
1	BBB	7846	0	7775	74	0
2	AAA	257	0	0	10	0
2	BBB	194	0	0	5	0
All	All	16160	0	15563	157	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 5.

The worst 5 of 157 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:BBB:619:ASP:H	1:BBB:622:GLN:HE21	1.25	0.84
1:AAA:634:ILE:HG21	1:AAA:770:VAL:HG11	1.59	0.82
1:AAA:933:ARG:H	1:AAA:937:ASN:HD21	1.29	0.81
1:BBB:641:VAL:HG11	1:BBB:872:GLY:HA2	1.67	0.76
1:AAA:278:ALA:HB3	1:AAA:281:LEU:HD13	1.66	0.76

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	AAA	969/988 (98%)	927 (96%)	38 (4%)	4 (0%)	34	29
1	BBB	965/988 (98%)	921 (95%)	40 (4%)	4 (0%)	34	29

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	1934/1976 (98%)	1848 (96%)	78 (4%)	8 (0%)	34 29

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	AAA	512	ASP
1	BBB	513	GLU
1	AAA	196	TRP
1	BBB	196	TRP
1	AAA	571	LYS

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	AAA	856/869 (98%)	822 (96%)	34 (4%)	31 29
1	BBB	853/869 (98%)	822 (96%)	31 (4%)	35 33
All	All	1709/1738 (98%)	1644 (96%)	65 (4%)	33 31

5 of 65 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	BBB	784	ARG
1	BBB	859	GLU
1	AAA	707	LYS
1	AAA	702	GLU
1	BBB	866	LEU

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)

There are no ligands in this entry.

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^{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	AAA	971/988 (98%)	-0.21	24 (2%)	57	65	25, 39, 70, 111	0
1	BBB	969/988 (98%)	-0.14	32 (3%)	46	55	26, 42, 79, 110	0
All	All	1940/1976 (98%)	-0.17	56 (2%)	51	61	25, 41, 74, 111	0

The worst 5 of 56 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AAA	513	GLU	5.2
1	AAA	742	SER	4.9
1	AAA	712	GLY	4.7
1	BBB	545	ASP	4.6
1	BBB	712	GLY	4.4

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

There are no such residues in this entry.