



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 08:41 AM GMT

PDB ID : 3FN9
Title : Crystal structure of putative beta-galactosidase from bacteroides fragilis
Authors : Ramagopal, U.A.; Toro, R.; Burley, S.K.; Almo, S.C.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on : 2008-12-23
Resolution : 2.70 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
<http://wwpdb.org/validation/2016/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 : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk26865

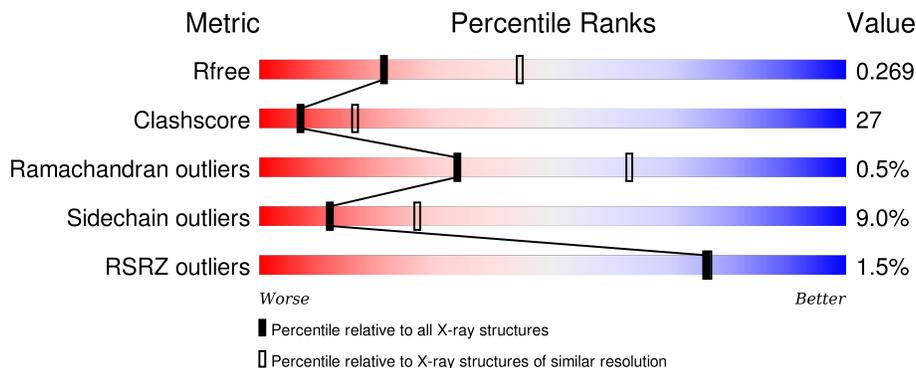
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.70 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}	91344	2103 (2.70-2.70)
Clashscore	102246	2422 (2.70-2.70)
Ramachandran outliers	100387	2382 (2.70-2.70)
Sidechain outliers	100360	2382 (2.70-2.70)
RSRZ outliers	91569	2107 (2.70-2.70)

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 ≥ 3 , 2, 1 and 0 types of geometric quality criteria. 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	692	 60% 34% . .
1	B	692	 50% 40% 7% .
1	C	692	 64% 30% . .
1	D	692	 58% 34% . .

2 Entry composition i

There are 3 unique types of molecules in this entry. The entry contains 21740 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 Putative beta-galactosidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	672	5426	3469	926	1013	18	0	0	0
1	B	672	5410	3463	923	1006	18	0	0	0
1	C	673	5431	3471	927	1015	18	0	0	0
1	D	666	5374	3439	916	1001	18	0	0	0

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	20	MET	-	EXPRESSION TAG	UNP Q5LJ68
A	21	SER	-	EXPRESSION TAG	UNP Q5LJ68
A	22	LEU	-	EXPRESSION TAG	UNP Q5LJ68
A	704	GLU	-	EXPRESSION TAG	UNP Q5LJ68
A	705	GLY	-	EXPRESSION TAG	UNP Q5LJ68
A	706	HIS	-	EXPRESSION TAG	UNP Q5LJ68
A	707	HIS	-	EXPRESSION TAG	UNP Q5LJ68
A	708	HIS	-	EXPRESSION TAG	UNP Q5LJ68
A	709	HIS	-	EXPRESSION TAG	UNP Q5LJ68
A	710	HIS	-	EXPRESSION TAG	UNP Q5LJ68
A	711	HIS	-	EXPRESSION TAG	UNP Q5LJ68
B	20	MET	-	EXPRESSION TAG	UNP Q5LJ68
B	21	SER	-	EXPRESSION TAG	UNP Q5LJ68
B	22	LEU	-	EXPRESSION TAG	UNP Q5LJ68
B	704	GLU	-	EXPRESSION TAG	UNP Q5LJ68
B	705	GLY	-	EXPRESSION TAG	UNP Q5LJ68
B	706	HIS	-	EXPRESSION TAG	UNP Q5LJ68
B	707	HIS	-	EXPRESSION TAG	UNP Q5LJ68
B	708	HIS	-	EXPRESSION TAG	UNP Q5LJ68
B	709	HIS	-	EXPRESSION TAG	UNP Q5LJ68
B	710	HIS	-	EXPRESSION TAG	UNP Q5LJ68

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	711	HIS	-	EXPRESSION TAG	UNP Q5LJ68
C	20	MET	-	EXPRESSION TAG	UNP Q5LJ68
C	21	SER	-	EXPRESSION TAG	UNP Q5LJ68
C	22	LEU	-	EXPRESSION TAG	UNP Q5LJ68
C	704	GLU	-	EXPRESSION TAG	UNP Q5LJ68
C	705	GLY	-	EXPRESSION TAG	UNP Q5LJ68
C	706	HIS	-	EXPRESSION TAG	UNP Q5LJ68
C	707	HIS	-	EXPRESSION TAG	UNP Q5LJ68
C	708	HIS	-	EXPRESSION TAG	UNP Q5LJ68
C	709	HIS	-	EXPRESSION TAG	UNP Q5LJ68
C	710	HIS	-	EXPRESSION TAG	UNP Q5LJ68
C	711	HIS	-	EXPRESSION TAG	UNP Q5LJ68
D	20	MET	-	EXPRESSION TAG	UNP Q5LJ68
D	21	SER	-	EXPRESSION TAG	UNP Q5LJ68
D	22	LEU	-	EXPRESSION TAG	UNP Q5LJ68
D	704	GLU	-	EXPRESSION TAG	UNP Q5LJ68
D	705	GLY	-	EXPRESSION TAG	UNP Q5LJ68
D	706	HIS	-	EXPRESSION TAG	UNP Q5LJ68
D	707	HIS	-	EXPRESSION TAG	UNP Q5LJ68
D	708	HIS	-	EXPRESSION TAG	UNP Q5LJ68
D	709	HIS	-	EXPRESSION TAG	UNP Q5LJ68
D	710	HIS	-	EXPRESSION TAG	UNP Q5LJ68
D	711	HIS	-	EXPRESSION TAG	UNP Q5LJ68

- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	3	Total Cl 3 3	0	0
2	D	3	Total Cl 3 3	0	0
2	C	1	Total Cl 1 1	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	31	Total O 31 31	0	0
3	B	18	Total O 18 18	0	0

Continued on next page...

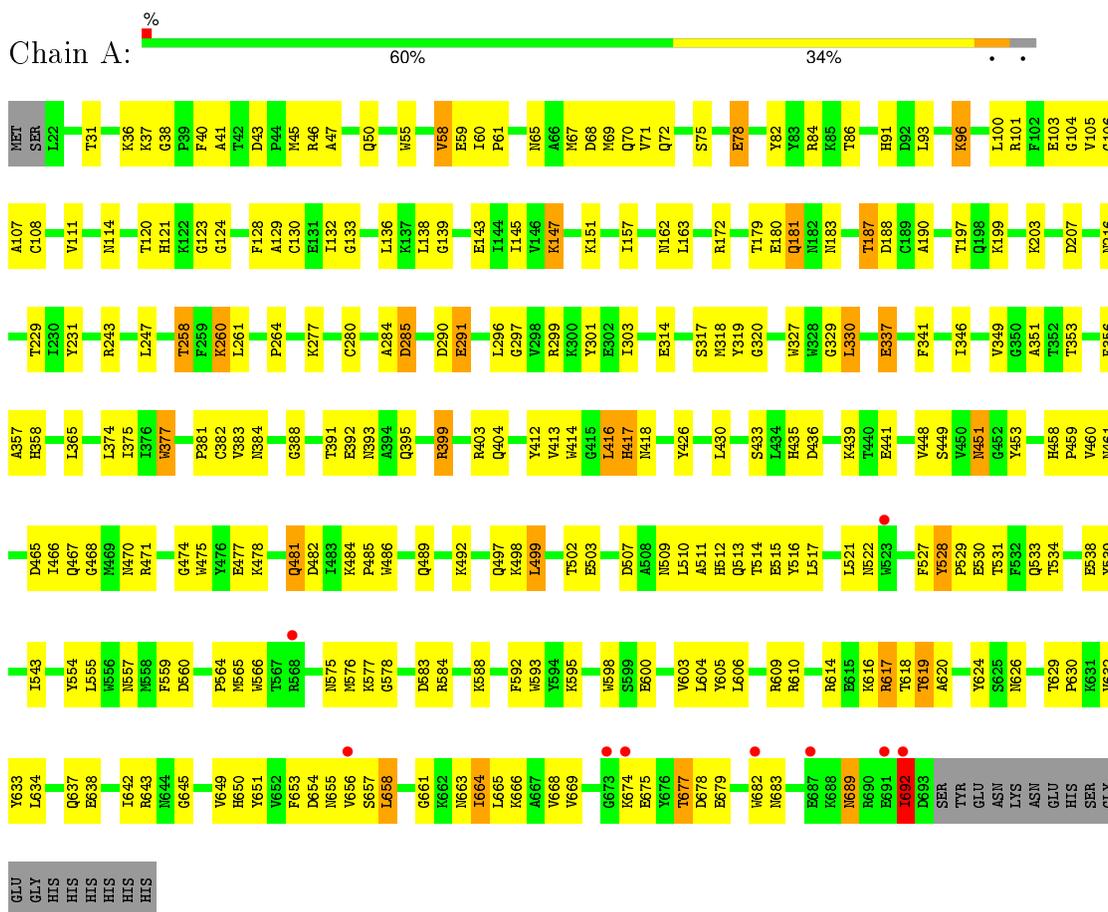
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	C	23	Total	O	0	0
			23	23		
3	D	20	Total	O	0	0
			20	20		

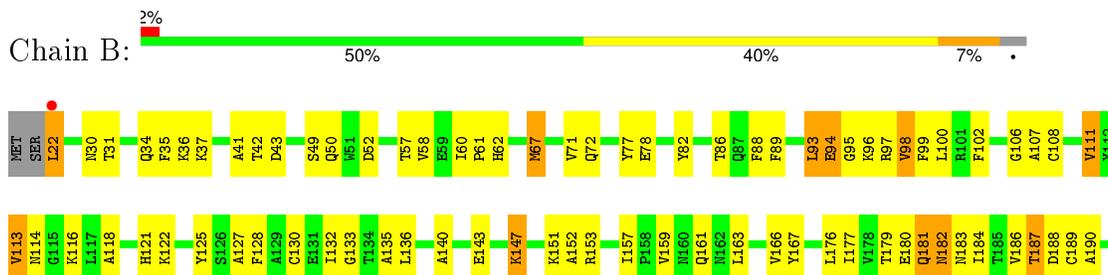
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: Putative beta-galactosidase



- Molecule 1: Putative beta-galactosidase



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	109.45Å 133.07Å 221.19Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.39 – 2.70 46.28 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.9 (49.39-2.70) 99.9 (46.28-2.70)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	0.10	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.12 (at 2.69Å)	Xtrriage
Refinement program	REFMAC 5.5.0066	Depositor
R, R_{free}	0.201 , 0.270 0.205 , 0.269	Depositor DCC
R_{free} test set	4414 reflections (5.21%)	DCC
Wilson B-factor (Å ²)	42.2	Xtrriage
Anisotropy	0.063	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 33.2	EDS
Estimated twinning fraction	No twinning to report.	Xtrriage
L-test for twinning ²	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.34$	Xtrriage
Outliers	0 of 89139 reflections	Xtrriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	21740	wwPDB-VP
Average B, all atoms (Å ²)	39.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.14% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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:
CL

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	0.55	1/5577 (0.0%)	0.67	0/7579
1	B	0.54	0/5561	0.67	0/7558
1	C	0.55	0/5582	0.66	0/7586
1	D	0.54	0/5525	0.67	0/7509
All	All	0.55	1/22245 (0.0%)	0.67	0/30232

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	108	CYS	CB-SG	-5.66	1.72	1.81

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	5426	0	5216	274	0
1	B	5410	0	5189	372	0
1	C	5431	0	5219	205	0
1	D	5374	0	5167	287	0
2	A	3	0	0	0	0
2	C	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	D	3	0	0	1	0
3	A	31	0	0	4	0
3	B	18	0	0	1	0
3	C	23	0	0	2	0
3	D	20	0	0	5	0
All	All	21740	0	20791	1125	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 27.

The worst 5 of 1125 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:630:PRO:HA	1:B:669:VAL:CG2	1.65	1.26
1:D:347:MET:CE	1:D:374:LEU:HD13	1.69	1.20
1:D:347:MET:HE1	1:D:374:LEU:CD1	1.72	1.19
1:D:347:MET:CE	1:D:374:LEU:CD1	2.24	1.15
1:B:208:ILE:CD1	1:B:259:PHE:HB2	1.76	1.15

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	670/692 (97%)	617 (92%)	50 (8%)	3 (0%)	39 69
1	B	670/692 (97%)	613 (92%)	51 (8%)	6 (1%)	21 49
1	C	671/692 (97%)	626 (93%)	44 (7%)	1 (0%)	56 83
1	D	664/692 (96%)	620 (93%)	40 (6%)	4 (1%)	30 59
All	All	2675/2768 (97%)	2476 (93%)	185 (7%)	14 (0%)	34 63

5 of 14 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	689	ASN
1	C	238	VAL
1	A	692	ILE
1	B	627	ILE
1	B	638	GLU

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	570/591 (96%)	524 (92%)	46 (8%)	15	33
1	B	564/591 (95%)	493 (87%)	71 (13%)	5	13
1	C	571/591 (97%)	530 (93%)	41 (7%)	18	41
1	D	564/591 (95%)	518 (92%)	46 (8%)	14	32
All	All	2269/2364 (96%)	2065 (91%)	204 (9%)	12	27

5 of 204 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	417	HIS
1	B	677	THR
1	D	557	ASN
1	B	436	ASP
1	B	601	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 58 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	336	ASN
1	B	637	GLN
1	D	557	ASN
1	B	393	ASN
1	B	455	HIS

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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 7 ligands modelled in this entry, 7 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, 95th 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	672/692 (97%)	-0.14	9 (1%) 79 79	23, 36, 61, 81	0
1	B	672/692 (97%)	-0.00	14 (2%) 67 68	22, 40, 62, 87	0
1	C	673/692 (97%)	-0.25	3 (0%) 93 94	22, 36, 53, 72	0
1	D	666/692 (96%)	-0.08	13 (1%) 68 69	21, 36, 61, 77	0
All	All	2683/2768 (96%)	-0.12	39 (1%) 76 76	21, 37, 60, 87	0

The worst 5 of 39 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	612	ALA	4.2
1	D	661	GLY	3.7
1	B	692	ILE	3.6
1	B	516	TYR	3.3
1	D	686	GLY	3.2

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 carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th 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	LLDF	B-factors(Å ²)	Q<0.9
2	CL	A	802	1/1	0.96	0.29	-	50,50,50,50	0
2	CL	A	801	1/1	0.97	0.24	-	39,39,39,39	0
2	CL	A	807	1/1	0.85	0.29	-	78,78,78,78	0
2	CL	C	805	1/1	0.96	0.34	-	59,59,59,59	0
2	CL	D	803	1/1	0.89	0.25	-	56,56,56,56	0
2	CL	D	804	1/1	0.84	0.15	-	74,74,74,74	0
2	CL	D	806	1/1	0.98	0.21	-	48,48,48,48	0

6.5 Other polymers [\(i\)](#)

There are no such residues in this entry.