



# wwPDB X-ray Structure Validation Summary Report

Feb 28, 2014 – 04:12 PM GMT

PDB ID : 3BCV  
Title : Crystal structure of a putative glycosyltransferase from Bacteroides fragilis  
Authors : Palani, K.; Kumaran, D.; Burley, S.K.; Swaminathan, S.; New York SGX  
Research Center for Structural Genomics (NYSGXRC)  
Deposited on : 2007-11-13  
Resolution : 2.35 Å(reported)

This is a wwPDB validation summary 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 <http://wwpdb.org/ValidationPDFNotes.html>

---

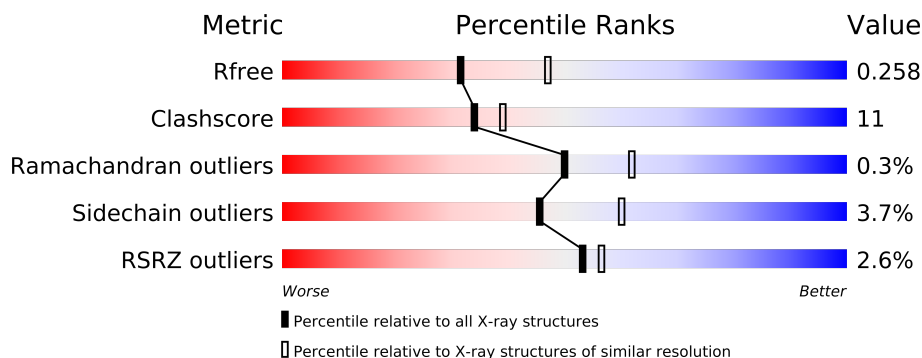
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22683

# 1 Overall quality at a glance

The reported resolution of this entry is 2.35 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))
$R_{free}$	66092	3327 (2.40-2.32)
Clashscore	79885	1064 (2.38-2.34)
Ramachandran outliers	78287	1048 (2.38-2.34)
Sidechain outliers	78261	1049 (2.38-2.34)
RSRZ outliers	66119	3330 (2.40-2.32)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	240	
1	B	240	

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3214 atoms, of which 0 are hydrogen and 0 are deuterium.

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 glycosyltransferase protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	196	Total	C	N	O	S	Se	0	0	0
			1555	993	252	299	7	4			
1	B	196	Total	C	N	O	S	Se	0	0	0
			1555	993	252	299	7	4			

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	-	EXPRESSION TAG	UNP Q5LBM4
A	2	SER	-	EXPRESSION TAG	UNP Q5LBM4
A	3	LEU	-	EXPRESSION TAG	UNP Q5LBM4
A	233	GLU	-	EXPRESSION TAG	UNP Q5LBM4
A	234	GLY	-	EXPRESSION TAG	UNP Q5LBM4
A	235	HIS	-	EXPRESSION TAG	UNP Q5LBM4
A	236	HIS	-	EXPRESSION TAG	UNP Q5LBM4
A	237	HIS	-	EXPRESSION TAG	UNP Q5LBM4
A	238	HIS	-	EXPRESSION TAG	UNP Q5LBM4
A	239	HIS	-	EXPRESSION TAG	UNP Q5LBM4
A	240	HIS	-	EXPRESSION TAG	UNP Q5LBM4
B	1	MSE	-	EXPRESSION TAG	UNP Q5LBM4
B	2	SER	-	EXPRESSION TAG	UNP Q5LBM4
B	3	LEU	-	EXPRESSION TAG	UNP Q5LBM4
B	233	GLU	-	EXPRESSION TAG	UNP Q5LBM4
B	234	GLY	-	EXPRESSION TAG	UNP Q5LBM4
B	235	HIS	-	EXPRESSION TAG	UNP Q5LBM4
B	236	HIS	-	EXPRESSION TAG	UNP Q5LBM4
B	237	HIS	-	EXPRESSION TAG	UNP Q5LBM4
B	238	HIS	-	EXPRESSION TAG	UNP Q5LBM4
B	239	HIS	-	EXPRESSION TAG	UNP Q5LBM4
B	240	HIS	-	EXPRESSION TAG	UNP Q5LBM4

- Molecule 2 is water.

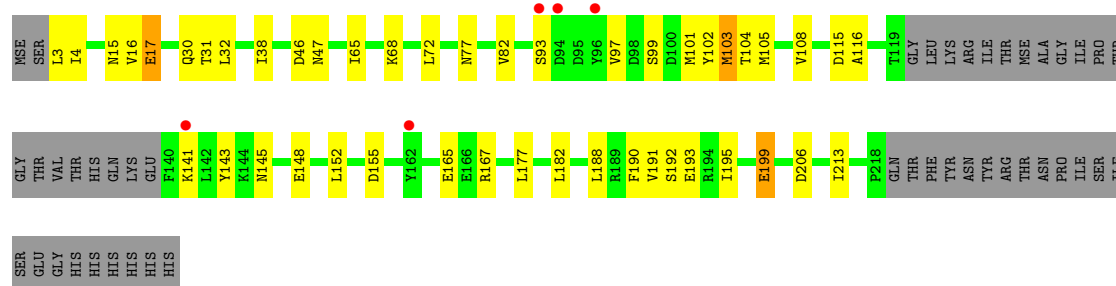
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	42	Total 42	O 42	0	0
2	B	62	Total 62	O 62	0	0

### 3 Residue-property plots

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.

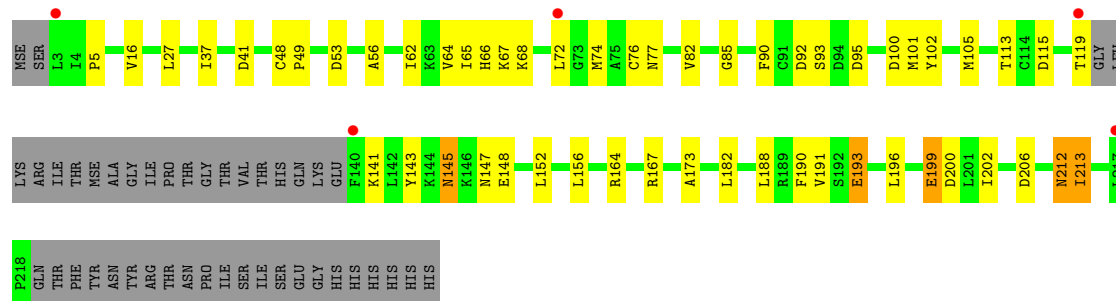
- Molecule 1: Putative glycosyltransferase protein

Chain A: 



- Molecule 1: Putative glycosyltransferase protein

Chain B: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	I 2 3	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	152.43Å 152.43Å 152.43Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	27.83 – 2.35 48.20 – 2.35	Depositor EDS
% Data completeness (in resolution range)	94.4 (27.83-2.35) 94.4 (48.20-2.35)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.33 (at 2.34Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.230 , 0.257 0.230 , 0.258	Depositor DCC
$R_{free}$ test set	1119 reflections (4.82%)	DCC
Wilson B-factor (Å <sup>2</sup> )	38.9	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 28.6	EDS
Estimated twinning fraction	0.031 for -l,-k,-h	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 24161 reflections	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	3214	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	39.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.26% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

## 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	A	0.40	0/1577	0.63	1/2134 (0.0%)
1	B	0.42	0/1577	0.67	1/2134 (0.0%)
All	All	0.41	0/3154	0.65	2/4268 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	212	ASN	N-CA-C	-6.16	94.38	111.00
1	A	68	LYS	N-CA-C	-5.09	97.26	111.00

There are no chirality outliers.

There are no planarity outliers.

### 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1555	0	1562	32	0
1	B	1555	0	1562	37	0
2	A	42	0	0	0	0
2	B	62	0	0	2	0
All	All	3214	0	3124	68	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 11.

The worst 5 of 68 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:101:MSE:HG2	1:A:105:MSE:HE2	1.27	1.12
1:B:101:MSE:HG2	1:B:105:MSE:HE2	1.25	1.09
1:B:77:ASN:HD21	1:B:190:PHE:H	1.19	0.89
1:A:77:ASN:HD21	1:A:190:PHE:H	1.23	0.87
1:B:101:MSE:CG	1:B:105:MSE:HE2	2.12	0.79

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	A	192/240 (80%)	185 (96%)	7 (4%)	0	100	100
1	B	192/240 (80%)	186 (97%)	5 (3%)	1 (0%)	38	45
All	All	384/480 (80%)	371 (97%)	12 (3%)	1 (0%)	50	62

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	213	ILE

### 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	177/210 (84%)	170 (96%)	7 (4%)	42	56
1	B	177/210 (84%)	171 (97%)	6 (3%)	49	64
All	All	354/420 (84%)	341 (96%)	13 (4%)	45	60



5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	193	GLU
1	A	199	GLU
1	B	145	ASN
1	A	141	LYS
1	B	113	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	58	GLN
1	B	181	ASN
1	B	77	ASN
1	A	170	GLN
1	B	66	HIS

### 5.3.3 RNA ⓘ

There are no RNA chains in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

There are no ligands in this entry.

## 5.7 Other polymers ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues

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	196/240 (81%)	0.21	5 (2%) 53 56	26, 40, 54, 64	0
1	B	196/240 (81%)	0.06	5 (2%) 53 56	25, 37, 50, 59	0
All	All	392/480 (81%)	0.13	10 (2%) 53 56	25, 38, 53, 64	0

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	96	TYR	4.2
1	B	72	LEU	3.5
1	B	3	LEU	3.4
1	A	162	TYR	3.2
1	B	119	THR	3.2

### 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 carbohydrates 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.