



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 9, 2020 – 09:05 PM BST

PDB ID : 6PD0  
Title : Crystal structure of the bacterial cellulose synthase subunit G (BcsG) from Escherichia coli, catalytic domain  
Authors : Anderson, A.C.; Brenner, T.; Weadge, J.T.  
Deposited on : 2019-06-18  
Resolution : 1.75 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](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
Xtriage (Phenix)	:	1.13
EDS	:	2.14.3.dev2
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)	:	2.14.3.dev2

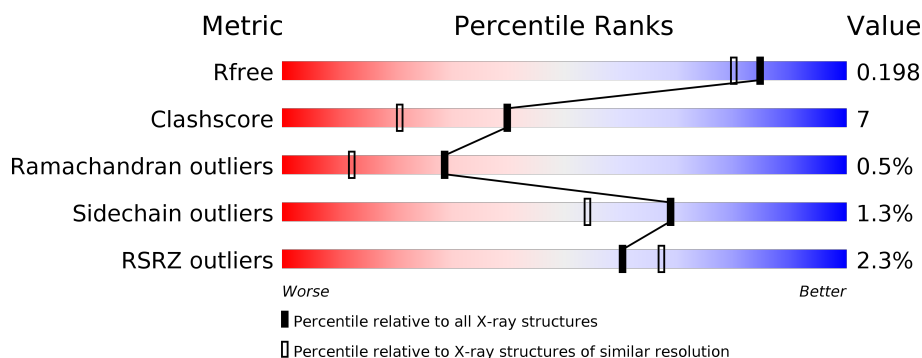
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.75 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	2340 (1.76-1.76)
Clashscore	141614	2466 (1.76-1.76)
Ramachandran outliers	138981	2437 (1.76-1.76)
Sidechain outliers	138945	2437 (1.76-1.76)
RSRZ outliers	127900	2298 (1.76-1.76)

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	417	<div> <div style="width: 100%; height: 10px; background-color: red;"></div> <div style="display: flex; justify-content: space-between; align-items: center;"> <div style="width: 78%; height: 10px; background-color: green;"></div> <div style="width: 10%; height: 10px; background-color: yellow;"></div> <div style="width: 12%; height: 10px; background-color: grey;"></div> </div> <div style="display: flex; justify-content: space-between; align-items: center; margin-top: 5px;"> <span>78%</span> <span>10%</span> <span>12%</span> </div> </div>
1	B	417	<div> <div style="width: 100%; height: 10px; background-color: red;"></div> <div style="display: flex; justify-content: space-between; align-items: center;"> <div style="width: 75%; height: 10px; background-color: green;"></div> <div style="width: 11%; height: 10px; background-color: yellow;"></div> <div style="width: 12%; height: 10px; background-color: grey;"></div> </div> <div style="display: flex; justify-content: space-between; align-items: center; margin-top: 5px;"> <span>75%</span> <span>11%</span> <span>12%</span> </div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 6686 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 Cellulose biosynthesis protein BcsG.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	366	Total	C	N	O	S	0	0	0
			2853	1815	484	542	12			
1	B	366	Total	C	N	O	S	0	0	0
			2853	1815	484	542	12			

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	143	MET	-	initiating methionine	UNP P37659
A	144	GLY	-	expression tag	UNP P37659
A	145	SER	-	expression tag	UNP P37659
A	146	SER	-	expression tag	UNP P37659
A	147	HIS	-	expression tag	UNP P37659
A	148	HIS	-	expression tag	UNP P37659
A	149	HIS	-	expression tag	UNP P37659
A	150	HIS	-	expression tag	UNP P37659
A	151	HIS	-	expression tag	UNP P37659
A	152	HIS	-	expression tag	UNP P37659
A	153	SER	-	expression tag	UNP P37659
A	154	SER	-	expression tag	UNP P37659
A	155	GLY	-	expression tag	UNP P37659
A	156	LEU	-	expression tag	UNP P37659
A	157	VAL	-	expression tag	UNP P37659
A	158	PRO	-	expression tag	UNP P37659
A	159	ARG	-	expression tag	UNP P37659
A	160	GLY	-	expression tag	UNP P37659
A	161	SER	-	expression tag	UNP P37659
A	162	HIS	-	expression tag	UNP P37659
A	163	MET	-	expression tag	UNP P37659
B	143	MET	-	initiating methionine	UNP P37659
B	144	GLY	-	expression tag	UNP P37659
B	145	SER	-	expression tag	UNP P37659
B	146	SER	-	expression tag	UNP P37659

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	147	HIS	-	expression tag	UNP P37659
B	148	HIS	-	expression tag	UNP P37659
B	149	HIS	-	expression tag	UNP P37659
B	150	HIS	-	expression tag	UNP P37659
B	151	HIS	-	expression tag	UNP P37659
B	152	HIS	-	expression tag	UNP P37659
B	153	SER	-	expression tag	UNP P37659
B	154	SER	-	expression tag	UNP P37659
B	155	GLY	-	expression tag	UNP P37659
B	156	LEU	-	expression tag	UNP P37659
B	157	VAL	-	expression tag	UNP P37659
B	158	PRO	-	expression tag	UNP P37659
B	159	ARG	-	expression tag	UNP P37659
B	160	GLY	-	expression tag	UNP P37659
B	161	SER	-	expression tag	UNP P37659
B	162	HIS	-	expression tag	UNP P37659
B	163	MET	-	expression tag	UNP P37659

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

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	1	Total Mg 1 1	0	0
2	A	1	Total Mg 1 1	0	0

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

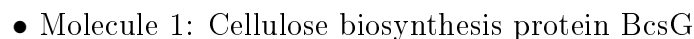
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total Zn 1 1	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	533	Total O 533 533	0	0
4	B	444	Total O 444 444	0	0



- Molecule 1: Cellulose biosynthesis protein BcsG



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	52.51Å 76.92Å 95.92Å 90.00° 90.58° 90.00°	Depositor
Resolution (Å)	47.96 – 1.75 47.96 – 1.47	Depositor EDS
% Data completeness (in resolution range)	96.7 (47.96-1.75) 71.8 (47.96-1.47)	Depositor EDS
$R_{merge}$	0.04	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.09 (at 1.48Å)	Xtriage
Refinement program	PHENIX (1.12_2829: ???)	Depositor
R, $R_{free}$	0.161 , 0.199 0.165 , 0.198	Depositor DCC
$R_{free}$ test set	4610 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	20.4	Xtriage
Anisotropy	0.181	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 39.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.021 for h,-k,-l	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	6686	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	23.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.39% 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

### 5.1 Standard geometry

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

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	0.44	0/2929	0.63	0/3986
1	B	0.45	0/2929	0.64	0/3986
All	All	0.45	0/5858	0.64	0/7972

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	2853	0	2765	31	0
1	B	2853	0	2767	54	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	1	0	0	0	0
4	A	533	0	0	8	0
4	B	444	0	0	13	0
All	All	6686	0	5532	82	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 7.

The worst 5 of 82 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:194:MET:HB2	1:B:195:PRO:CD	1.60	1.29
1:B:194:MET:CB	1:B:195:PRO:HD2	1.68	1.20
1:B:406:LYS:HG3	4:B:979:HOH:O	1.55	1.03
1:B:202:THR:HG23	1:B:205:ASN:H	1.24	0.97
1:A:535:ASN:HB3	1:A:550:ASN:HD22	1.28	0.97

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	364/417 (87%)	358 (98%)	5 (1%)	1 (0%)	41	22
1	B	364/417 (87%)	354 (97%)	7 (2%)	3 (1%)	19	6
All	All	728/834 (87%)	712 (98%)	12 (2%)	4 (0%)	29	12

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	195	PRO
1	B	196	ALA
1	A	196	ALA
1	B	553	ASP

### 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	313/348 (90%)	313 (100%)	0	100	100
1	B	313/348 (90%)	305 (97%)	8 (3%)	46	23
All	All	626/696 (90%)	618 (99%)	8 (1%)	69	54

5 of 8 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	393	LEU
1	B	553	ASP
1	B	540	GLN
1	B	304	ASN
1	B	405	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	550	ASN
1	B	207	ASN
1	B	304	ASN
1	B	540	GLN
1	B	550	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules 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 monosaccharides in this entry.

## 5.6 Ligand geometry

Of 3 ligands modelled in this entry, 3 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

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 [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, 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	366/417 (87%)	-0.29	3 (0%) 86 90	10, 18, 35, 61	0
1	B	366/417 (87%)	-0.07	14 (3%) 40 47	11, 20, 43, 69	0
All	All	732/834 (87%)	-0.18	17 (2%) 60 67	10, 19, 40, 69	0

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	195	PRO	5.9
1	B	196	ALA	4.9
1	B	198	THR	3.0
1	A	195	PRO	3.0
1	B	550	ASN	2.9

### 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.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
3	ZN	A	602	1/1	0.98	0.45	20,20,20,20	0
2	MG	A	601	1/1	0.99	0.04	12,12,12,12	0
2	MG	B	601	1/1	1.00	0.03	14,14,14,14	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.