



wwPDB X-ray Structure Validation Summary Report

Oct 6, 2021 – 05:10 am BST

PDB ID : 7B5T
Title : *S. agalactiae* BusR transcription factor
Authors : Bandera, A.M.; Witte, G.
Deposited on : 2020-12-07
Resolution : 2.80 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  symbol.

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

MolProbity : 4.02b-467
Mogul : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.23.2
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac : 5.8.0267
CCP4 : 7.1.010 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.23.2

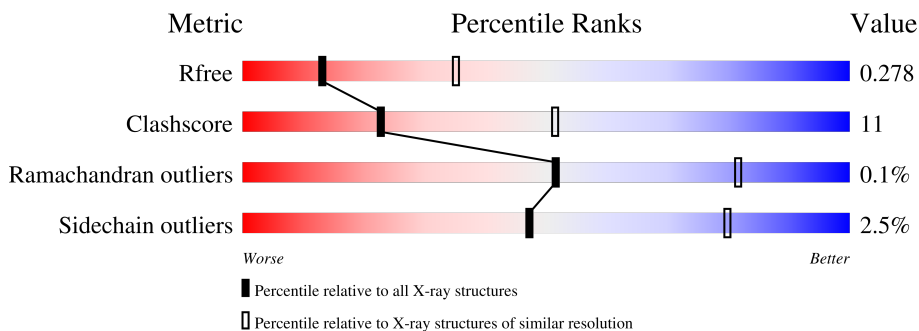
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.80 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)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	215	
1	B	215	
1	C	215	
1	D	215	

2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 6528 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 GntR family transcriptional regulator.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	Se			
1	A	211	Total 1656	C 1053	N 279	O 321	Se 3	0	0	0
1	B	207	Total 1628	C 1035	N 275	O 316	Se 2	0	0	0
1	D	200	Total 1573	C 1002	N 267	O 302	Se 2	0	0	0
1	C	210	Total 1649	C 1048	N 278	O 320	Se 3	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	expression tag	UNP K0JNC6
A	0	PRO	-	expression tag	UNP K0JNC6
B	-1	GLY	-	expression tag	UNP K0JNC6
B	0	PRO	-	expression tag	UNP K0JNC6
D	-1	GLY	-	expression tag	UNP K0JNC6
D	0	PRO	-	expression tag	UNP K0JNC6
C	-1	GLY	-	expression tag	UNP K0JNC6
C	0	PRO	-	expression tag	UNP K0JNC6

- Molecule 2 is water.

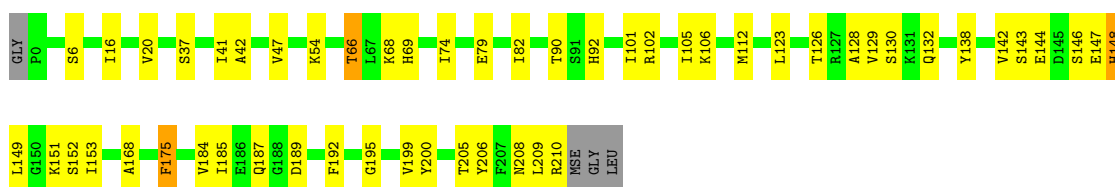
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	6	Total 6	O 6	0	0
2	B	2	Total 2	O 2	0	0
2	D	5	Total 5	O 5	0	0
2	C	9	Total 9	O 9	0	0

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. 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.

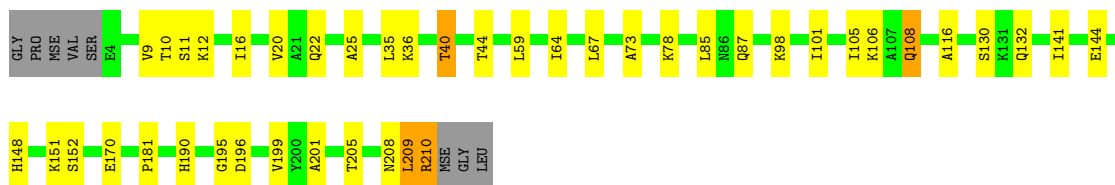
- Molecule 1: GntR family transcriptional regulator

Chain A: 



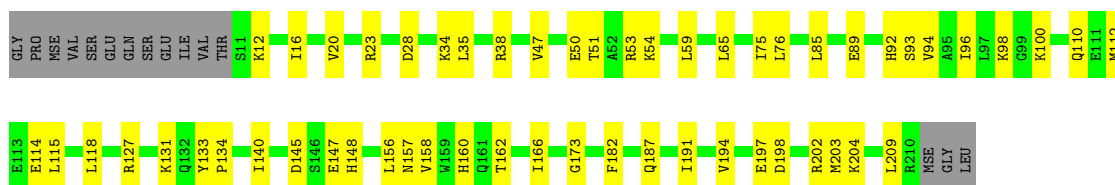
- Molecule 1: GntR family transcriptional regulator

Chain B: 



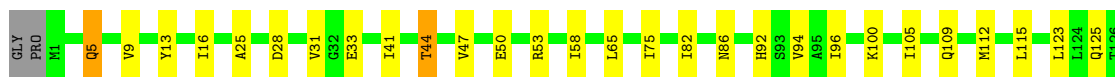
- Molecule 1: GntR family transcriptional regulator

Chain D: 



- Molecule 1: GntR family transcriptional regulator

Chain C: 



R127	A128	V129	S130	Y133	P134	L135	A136	P137	I140	E147	S152	V158	T162	G163	A164	T165	I166	E170	H171	K174	F175	I176	P179	G180	P181	F182	S183	E186	H190	I191	F192	F193	D196	E197	D198	R202	M203	K204	L209	R210	MSF	GLY	LEU
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-----	-----	-----

4 Data and refinement statistics i

Property	Value	Source
Space group	I 4	Depositor
Cell constants a, b, c, α , β , γ	114.18Å 114.18Å 240.48Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.22 – 2.80 44.32 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.1 (48.22-2.80) 99.1 (44.32-2.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.22 (at 2.81Å)	Xtrriage
Refinement program	PHENIX 1.18.2_3874	Depositor
R, R_{free}	0.257 , 0.269 0.257 , 0.278	Depositor DCC
R_{free} test set	1996 reflections (5.35%)	wwPDB-VP
Wilson B-factor (Å ²)	104.6	Xtrriage
Anisotropy	0.429	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	(Not available) , (Not available)	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.030 for h,-k,-l	Xtrriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	6528	wwPDB-VP
Average B, all atoms (Å ²)	127.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.42% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

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	0.31	0/1681	0.54	0/2266
1	B	0.30	0/1653	0.52	0/2230
1	C	0.37	0/1674	0.60	1/2258 (0.0%)
1	D	0.33	0/1598	0.57	0/2155
All	All	0.33	0/6606	0.56	1/8909 (0.0%)

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	B	0	1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	C	5	GLN	N-CA-CB	-5.22	101.21	110.60

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	208	ASN	Mainchain

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	1656	0	1671	49	0
1	B	1628	0	1638	32	0
1	C	1649	0	1664	54	0
1	D	1573	0	1586	40	0
2	A	6	0	0	0	0
2	B	2	0	0	0	0
2	C	9	0	0	0	0
2	D	5	0	0	1	0
All	All	6528	0	6559	150	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 11.

The worst 5 of 150 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:A:90:THR:HG23	1:A:92:HIS:H	1.37	0.88
1:A:147:GLU:HG2	1:A:148:HIS:H	1.42	0.85
1:C:204:LYS:HE2	1:C:209:LEU:HD23	1.65	0.77
1:C:13:TYR:HD1	1:C:58:ILE:HD11	1.49	0.77
1:C:198:ASP:O	1:C:202:ARG:HG3	1.87	0.73

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	209/215 (97%)	196 (94%)	13 (6%)	0	100 100
1	B	205/215 (95%)	195 (95%)	9 (4%)	1 (0%)	29 61
1	C	208/215 (97%)	200 (96%)	8 (4%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	D	198/215 (92%)	189 (96%)	9 (4%)	0	100	100
All	All	820/860 (95%)	780 (95%)	39 (5%)	1 (0%)	51	81

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	209	LEU

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	181/179 (101%)	175 (97%)	6 (3%)	38	72
1	B	177/179 (99%)	172 (97%)	5 (3%)	43	77
1	C	180/179 (101%)	176 (98%)	4 (2%)	52	83
1	D	170/179 (95%)	167 (98%)	3 (2%)	59	86
All	All	708/716 (99%)	690 (98%)	18 (2%)	47	80

5 of 18 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	44	THR
1	C	196	ASP
1	C	175	PHE
1	B	108	GLN
1	D	194	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	171	HIS
1	D	14	GLN
1	C	187	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	C	109	GLN
1	C	125	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](#)

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.