



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 15, 2017 – 05:05 AM EDT

PDB ID : 3BIL
Title : Crystal structure of a probable LacI family transcriptional regulator from *Corynebacterium glutamicum*
Authors : Bonanno, J.B.; Freeman, J.; Bain, K.T.; Mendoza, M.; Ozyurt, S.; Smith, D.; Wasserman, S.; Sauder, J.M.; Burley, S.K.; Almo, S.C.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on : unknown
Resolution : 2.50 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
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030345
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	rb-20030345

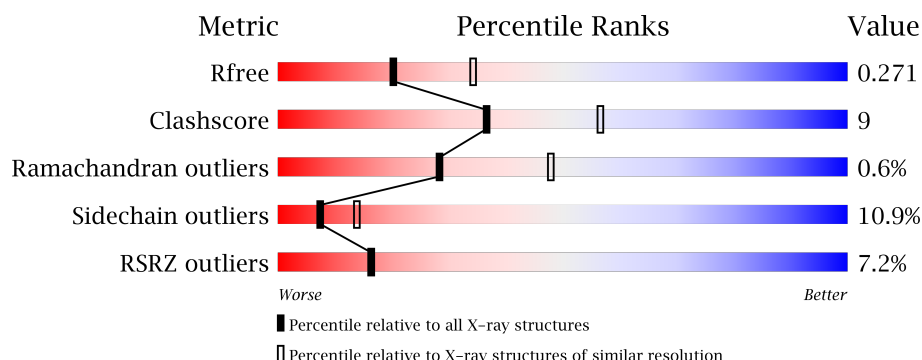
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.50 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}	100719	3846 (2.50-2.50)
Clashscore	112137	4554 (2.50-2.50)
Ramachandran outliers	110173	4463 (2.50-2.50)
Sidechain outliers	110143	4465 (2.50-2.50)
RSRZ outliers	101464	3876 (2.50-2.50)

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	348	<div> <div>4%</div> <div> <div></div> <div>54%</div> <div>17%</div> <div>•</div> <div>26%</div> </div> </div>
1	B	348	<div> <div>7%</div> <div> <div></div> <div>59%</div> <div>13%</div> <div>•</div> <div>26%</div> </div> </div>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3842 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 Probable LacI-family transcriptional regulator.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	257	Total	C	N	O	S	0	1	0
			1911	1202	319	380	10			
1	B	257	Total	C	N	O	S	0	1	0
			1911	1202	319	380	10			

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	MET	-	EXPRESSION TAG	UNP Q8NQQ9
A	1	SER	-	EXPRESSION TAG	UNP Q8NQQ9
A	2	LEU	-	EXPRESSION TAG	UNP Q8NQQ9
A	340	GLU	-	EXPRESSION TAG	UNP Q8NQQ9
A	341	GLY	-	EXPRESSION TAG	UNP Q8NQQ9
A	342	HIS	-	EXPRESSION TAG	UNP Q8NQQ9
A	343	HIS	-	EXPRESSION TAG	UNP Q8NQQ9
A	344	HIS	-	EXPRESSION TAG	UNP Q8NQQ9
A	345	HIS	-	EXPRESSION TAG	UNP Q8NQQ9
A	346	HIS	-	EXPRESSION TAG	UNP Q8NQQ9
A	347	HIS	-	EXPRESSION TAG	UNP Q8NQQ9
B	0	MET	-	EXPRESSION TAG	UNP Q8NQQ9
B	1	SER	-	EXPRESSION TAG	UNP Q8NQQ9
B	2	LEU	-	EXPRESSION TAG	UNP Q8NQQ9
B	340	GLU	-	EXPRESSION TAG	UNP Q8NQQ9
B	341	GLY	-	EXPRESSION TAG	UNP Q8NQQ9
B	342	HIS	-	EXPRESSION TAG	UNP Q8NQQ9
B	343	HIS	-	EXPRESSION TAG	UNP Q8NQQ9
B	344	HIS	-	EXPRESSION TAG	UNP Q8NQQ9
B	345	HIS	-	EXPRESSION TAG	UNP Q8NQQ9
B	346	HIS	-	EXPRESSION TAG	UNP Q8NQQ9
B	347	HIS	-	EXPRESSION TAG	UNP Q8NQQ9

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	11	Total 11	O 11	0	0
2	B	9	Total 9	O 9	0	0

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	75.21Å 54.34Å 78.65Å 90.00° 94.19° 90.00°	Depositor
Resolution (Å)	20.00 – 2.50 29.92 – 2.50	Depositor EDS
% Data completeness (in resolution range)	100.0 (20.00-2.50) 100.0 (29.92-2.50)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	0.13	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.70 (at 2.51Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.218 , 0.278 0.213 , 0.271	Depositor DCC
R_{free} test set	1116 reflections (5.31%)	DCC
Wilson B-factor (Å ²)	45.8	Xtriage
Anisotropy	0.650	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 50.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.026 for l,-k,h	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	3842	wwPDB-VP
Average B, all atoms (Å ²)	47.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.96% 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	0.78	0/1942	0.81	0/2641
1	B	0.77	0/1942	0.79	1/2641 (0.0%)
All	All	0.77	0/3884	0.80	1/5282 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	250	MET	CG-SD-CE	5.05	108.28	100.20

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	1911	0	1914	46	0
1	B	1911	0	1914	29	0
2	A	11	0	0	2	0
2	B	9	0	0	0	0
All	All	3842	0	3828	69	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 9.

The worst 5 of 69 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:149:LEU:HB2	1:B:162:THR:HG22	1.50	0.93
1:A:170:GLY:H	1:A:324:THR:HG22	1.43	0.82
1:B:170:GLY:H	1:B:324:THR:HG22	1.46	0.80
1:A:99:ILE:CG2	1:B:99:ILE:HG23	2.14	0.77
1:A:99:ILE:CG2	1:B:99:ILE:CG2	2.62	0.77

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	250/348 (72%)	236 (94%)	12 (5%)	2 (1%)	22	39
1	B	250/348 (72%)	236 (94%)	13 (5%)	1 (0%)	38	59
All	All	500/696 (72%)	472 (94%)	25 (5%)	3 (1%)	28	48

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	284	GLN
1	A	308	GLU
1	A	132	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	212/282 (75%)	186 (88%)	26 (12%)	5	10
1	B	212/282 (75%)	192 (91%)	20 (9%)	10	19
All	All	424/564 (75%)	378 (89%)	46 (11%)	7	14

5 of 46 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	277	THR
1	A	324	THR
1	B	295	VAL
1	A	295	VAL
1	A	301	ARG

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	120	HIS
1	A	181	ASN
1	A	211	ASN
1	A	334	ASN
1	B	211	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 carbohydrates in this entry.

5.6 Ligand geometry ⓘ

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 [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	257/348 (73%)	0.49	13 (5%) 29 30	16, 45, 78, 88	0
1	B	257/348 (73%)	0.57	24 (9%) 9 9	16, 45, 81, 95	0
All	All	514/696 (73%)	0.53	37 (7%) 16 16	16, 45, 78, 95	0

The worst 5 of 37 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	157	ASP	6.2
1	B	100	ILE	5.5
1	A	100	ILE	5.3
1	B	70	VAL	5.2
1	B	156	GLY	5.0

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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.