



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jul 19, 2022 – 07:26 AM JST

PDB ID : 7FF4  
Title : The crystal structure of Clostridium cellulolyticum LacI family transcriptional regulator Ccel\_1438  
Authors : Zhang, N.; Ge, H.  
Deposited on : 2021-07-22  
Resolution : 2.19 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 types of validation reports are described at <http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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	:	2.29
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)	:	2.29

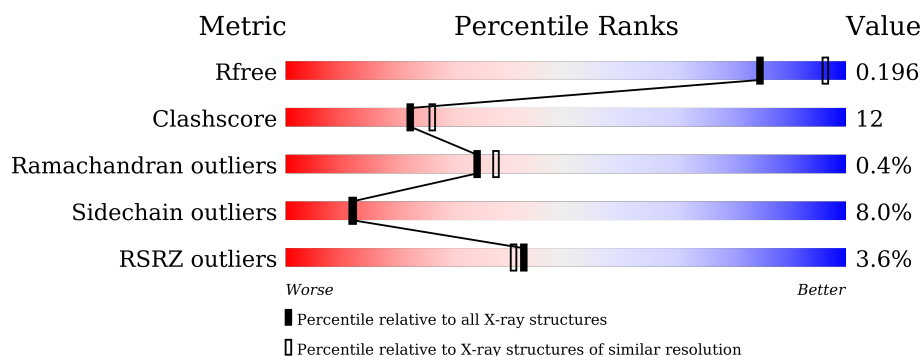
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.19 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	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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  $\geq 3$ , 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	343	<div> <div>3%</div> <div> <div></div> <div>63%</div> <div>14%</div> <div>•</div> <div>18%</div> </div> </div>
1	B	343	<div> <div>2%</div> <div> <div></div> <div>63%</div> <div>15%</div> <div>•</div> <div>19%</div> </div> </div>
1	C	343	<div> <div>3%</div> <div> <div></div> <div>63%</div> <div>12%</div> <div>6%</div> <div>19%</div> </div> </div>
1	D	343	<div> <div>3%</div> <div> <div></div> <div>63%</div> <div>15%</div> <div>•</div> <div>18%</div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 9043 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 Transcriptional regulator, LacI family.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	280	Total	C	N	O	S	0	0	0
			2202	1398	374	422	8			
1	B	278	Total	C	N	O	S	0	0	0
			2191	1392	372	419	8			
1	C	279	Total	C	N	O	S	0	1	0
			2200	1397	373	421	9			
1	D	280	Total	C	N	O	S	0	1	0
			2205	1400	374	422	9			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	338	HIS	-	expression tag	UNP B8I1W6
A	339	HIS	-	expression tag	UNP B8I1W6
A	340	HIS	-	expression tag	UNP B8I1W6
A	341	HIS	-	expression tag	UNP B8I1W6
A	342	HIS	-	expression tag	UNP B8I1W6
A	343	HIS	-	expression tag	UNP B8I1W6
B	338	HIS	-	expression tag	UNP B8I1W6
B	339	HIS	-	expression tag	UNP B8I1W6
B	340	HIS	-	expression tag	UNP B8I1W6
B	341	HIS	-	expression tag	UNP B8I1W6
B	342	HIS	-	expression tag	UNP B8I1W6
B	343	HIS	-	expression tag	UNP B8I1W6
C	338	HIS	-	expression tag	UNP B8I1W6
C	339	HIS	-	expression tag	UNP B8I1W6
C	340	HIS	-	expression tag	UNP B8I1W6
C	341	HIS	-	expression tag	UNP B8I1W6
C	342	HIS	-	expression tag	UNP B8I1W6
C	343	HIS	-	expression tag	UNP B8I1W6
D	338	HIS	-	expression tag	UNP B8I1W6
D	339	HIS	-	expression tag	UNP B8I1W6
D	340	HIS	-	expression tag	UNP B8I1W6

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
D	341	HIS	-	expression tag	UNP B8I1W6
D	342	HIS	-	expression tag	UNP B8I1W6
D	343	HIS	-	expression tag	UNP B8I1W6

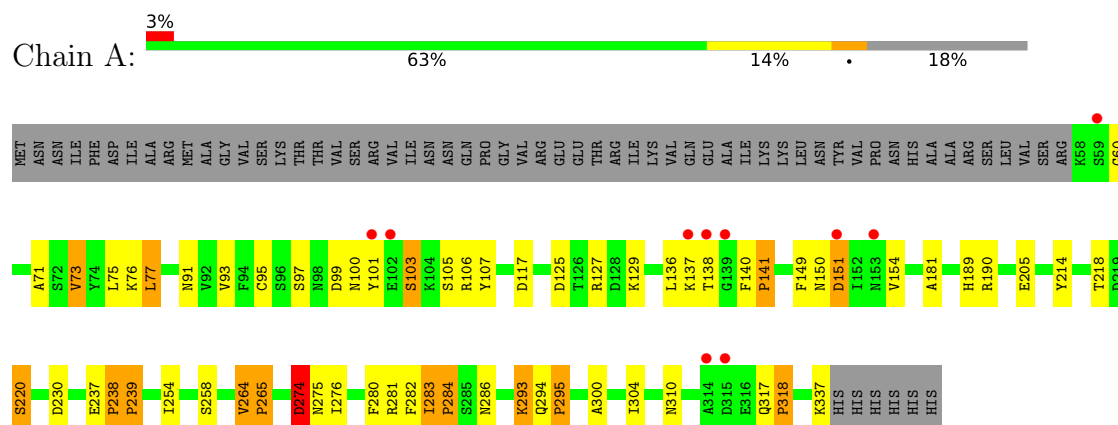
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	89	Total	O	0	0
			89	89		
2	B	53	Total	O	0	0
			53	53		
2	C	62	Total	O	0	0
			62	62		
2	D	41	Total	O	0	0
			41	41		

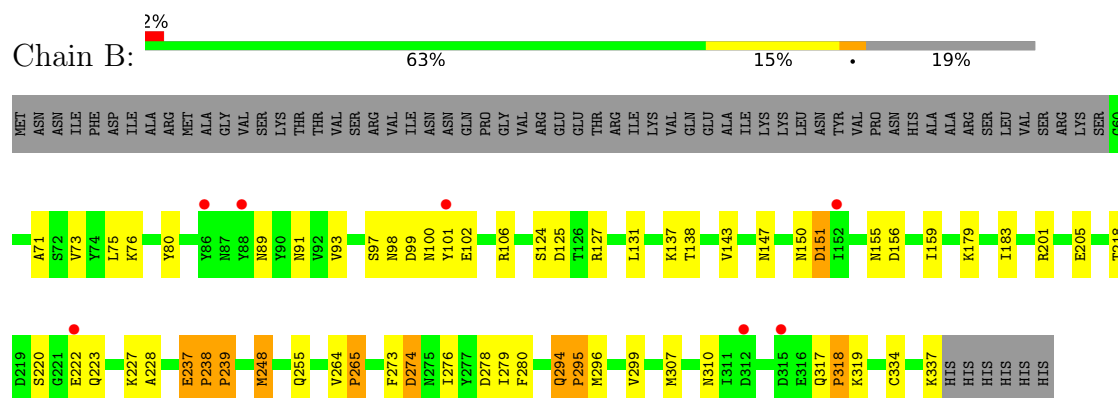
### 3 Residue-property plots

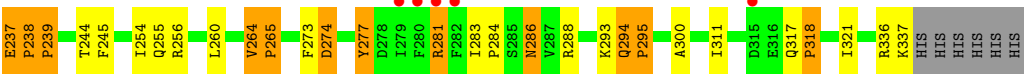
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: Transcriptional regulator, LacI family

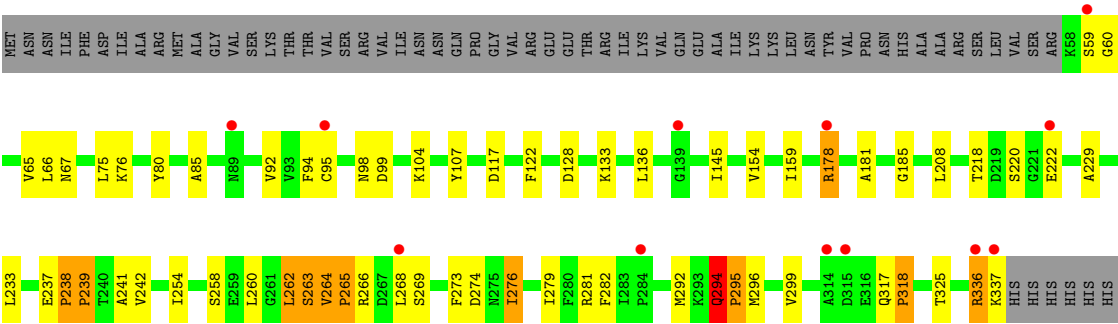


- Molecule 1: Transcriptional regulator, LacI family





● Molecule 1: Transcriptional regulator, LacI family



## 4 Data and refinement statistics

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	194.52Å 194.52Å 118.64Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	48.63 – 2.19 48.63 – 2.19	Depositor EDS
% Data completeness (in resolution range)	99.5 (48.63-2.19) 99.5 (48.63-2.19)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.08 (at 2.18Å)	Xtriage
Refinement program	PHENIX 1.14 _3260	Depositor
R, $R_{free}$	0.178 , 0.208 0.181 , 0.196	Depositor DCC
$R_{free}$ test set	4310 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	48.9	Xtriage
Anisotropy	0.294	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 46.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.006 for -2/3*h-1/3*k-4/3*l,-1/3*h-2/3*k+4/3*l,-1/3*h+1/3*k+1/3*l 0.004 for -h,1/3*h-1/3*k-4/3*l,-1/3*h-2/3*k+1/3*l 0.002 for -1/3*h+1/3*k+4/3*l,-k,2/3*h+1/3*k+1/3*l 0.004 for -h,2/3*h+1/3*k+4/3*l,1/3*h+2/3*k-1/3*l 0.009 for -1/3*h-2/3*k+4/3*l,-2/3*h-1/3*k-4/3*l,1/3*h-1/3*k-1/3*l 0.003 for 1/3*h+2/3*k-4/3*l,-k,-2/3*h-1/3*k-1/3*l 0.014 for h,-h-k,-l	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	9043	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	60.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.02% 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 [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	1.10	14/2242 (0.6%)	0.88	9/3035 (0.3%)
1	B	0.99	10/2231 (0.4%)	0.86	7/3020 (0.2%)
1	C	1.04	12/2243 (0.5%)	0.85	8/3036 (0.3%)
1	D	0.98	10/2248 (0.4%)	0.85	5/3043 (0.2%)
All	All	1.03	46/8964 (0.5%)	0.86	29/12134 (0.2%)

The worst 5 of 46 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	265	PRO	N-CA	13.23	1.69	1.47
1	A	141	PRO	N-CA	13.13	1.69	1.47
1	B	265	PRO	N-CA	13.13	1.69	1.47
1	D	238	PRO	N-CA	12.92	1.69	1.47
1	D	295	PRO	N-CA	12.89	1.69	1.47

The worst 5 of 29 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	318	PRO	CA-N-CD	-8.28	99.91	111.50
1	B	318	PRO	CA-N-CD	-8.12	100.14	111.50
1	D	318	PRO	CA-N-CD	-7.67	100.76	111.50
1	B	150	ASN	CB-CA-C	-7.38	95.63	110.40
1	D	265	PRO	CA-N-CD	-7.18	101.44	111.50

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	2202	0	2165	63	0
1	B	2191	0	2158	53	0
1	C	2200	0	2168	50	0
1	D	2205	0	2170	58	0
2	A	89	0	0	4	0
2	B	53	0	0	0	0
2	C	62	0	0	0	0
2	D	41	0	0	0	0
All	All	9043	0	8661	205	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 12.

The worst 5 of 205 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:C:265:PRO:CA	1:C:265:PRO:N	1.69	1.48
1:B:238:PRO:CA	1:B:238:PRO:N	1.68	1.47
1:A:284:PRO:N	1:A:284:PRO:CA	1.69	1.46
1:C:238:PRO:N	1:C:238:PRO:CA	1.68	1.46
1:A:318:PRO:N	1:A:318:PRO:CA	1.68	1.45

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	278/343 (81%)	273 (98%)	4 (1%)	1 (0%)	34	37
1	B	276/343 (80%)	271 (98%)	4 (1%)	1 (0%)	34	37

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	278/343 (81%)	273 (98%)	4 (1%)	1 (0%)	34	37
1	D	279/343 (81%)	273 (98%)	5 (2%)	1 (0%)	34	37
All	All	1111/1372 (81%)	1090 (98%)	17 (2%)	4 (0%)	34	37

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	274	ASP
1	B	274	ASP
1	C	274	ASP
1	D	274	ASP

### 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	239/296 (81%)	220 (92%)	19 (8%)	12	12
1	B	238/296 (80%)	221 (93%)	17 (7%)	14	16
1	C	240/296 (81%)	217 (90%)	23 (10%)	8	8
1	D	240/296 (81%)	223 (93%)	17 (7%)	14	16
All	All	957/1184 (81%)	881 (92%)	76 (8%)	12	12

5 of 76 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	336	ARG
1	D	276	ILE
1	D	59	SER
1	D	222	GLU
1	D	337	LYS

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

Mol	Chain	Res	Type
1	A	189	HIS
1	B	67	ASN
1	B	155	ASN
1	B	310	ASN
1	D	294	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 [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	280/343 (81%)	0.20	10 (3%) 42 41	35, 53, 85, 113	0
1	B	278/343 (81%)	0.12	7 (2%) 57 55	37, 57, 94, 115	0
1	C	279/343 (81%)	0.29	11 (3%) 39 37	36, 56, 87, 109	0
1	D	280/343 (81%)	0.37	12 (4%) 35 33	42, 61, 96, 121	0
All	All	1117/1372 (81%)	0.25	40 (3%) 42 41	35, 56, 92, 121	0

The worst 5 of 40 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	59	SER	6.1
1	D	337	LYS	4.9
1	B	315	ASP	4.5
1	D	139	GLY	3.6
1	A	151	ASP	3.5

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.