wwPDB X-ray Structure Validation Summary Report (i)

May 26, 2020 – 08:30 pm BST

PDB ID	:	5WQE
Title	:	Crystal structure of Alicyclobacillus acidoterrestris C2c1 in complex with single-guide RNA at 3.1 Angstrom resolution
Authors	:	Liu, L.; Wang, Y.L.
Deposited on	:	2016-11-26
Resolution	:	3.13 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

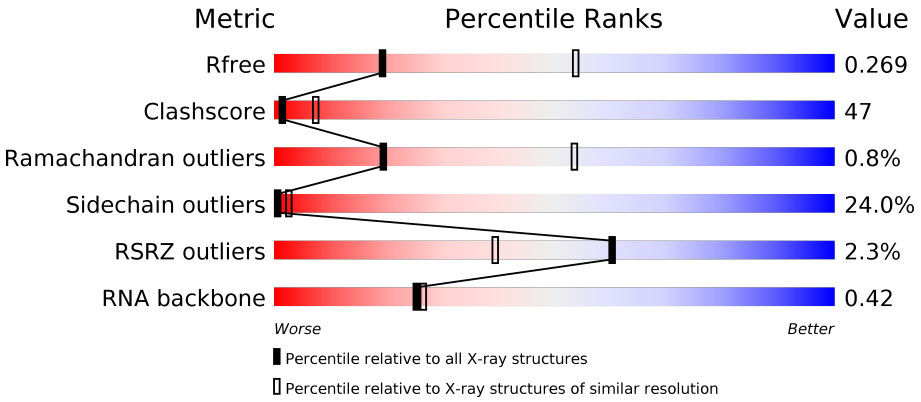
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.13 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	1292 (3.14-3.10)
Clashscore	141614	1389 (3.14-3.10)
Ramachandran outliers	138981	1337 (3.14-3.10)
Sidechain outliers	138945	1337 (3.14-3.10)
RSRZ outliers	127900	1260 (3.14-3.10)
RNA backbone	3102	1134 (3.44-2.80)

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 on 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 <=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	1137	2% 38% 39% 11% • 11%
2	B	112	12% 19% 21% • 46%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9308 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 CRISPR-associated endonuclease C2c1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	1011	Total	C	N	O	S	Se	0	0	0
			7988	5033	1476	1452	8	19			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1130	LEU	-	expression tag	UNP T0D7A2
A	1131	GLU	-	expression tag	UNP T0D7A2
A	1132	HIS	-	expression tag	UNP T0D7A2
A	1133	HIS	-	expression tag	UNP T0D7A2
A	1134	HIS	-	expression tag	UNP T0D7A2
A	1135	HIS	-	expression tag	UNP T0D7A2
A	1136	HIS	-	expression tag	UNP T0D7A2
A	1137	HIS	-	expression tag	UNP T0D7A2

- Molecule 2 is a RNA chain called RNA (60-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	60	Total	C	N	O	P	0	0	0
			1278	567	241	410	60			

- Molecule 3 is water.

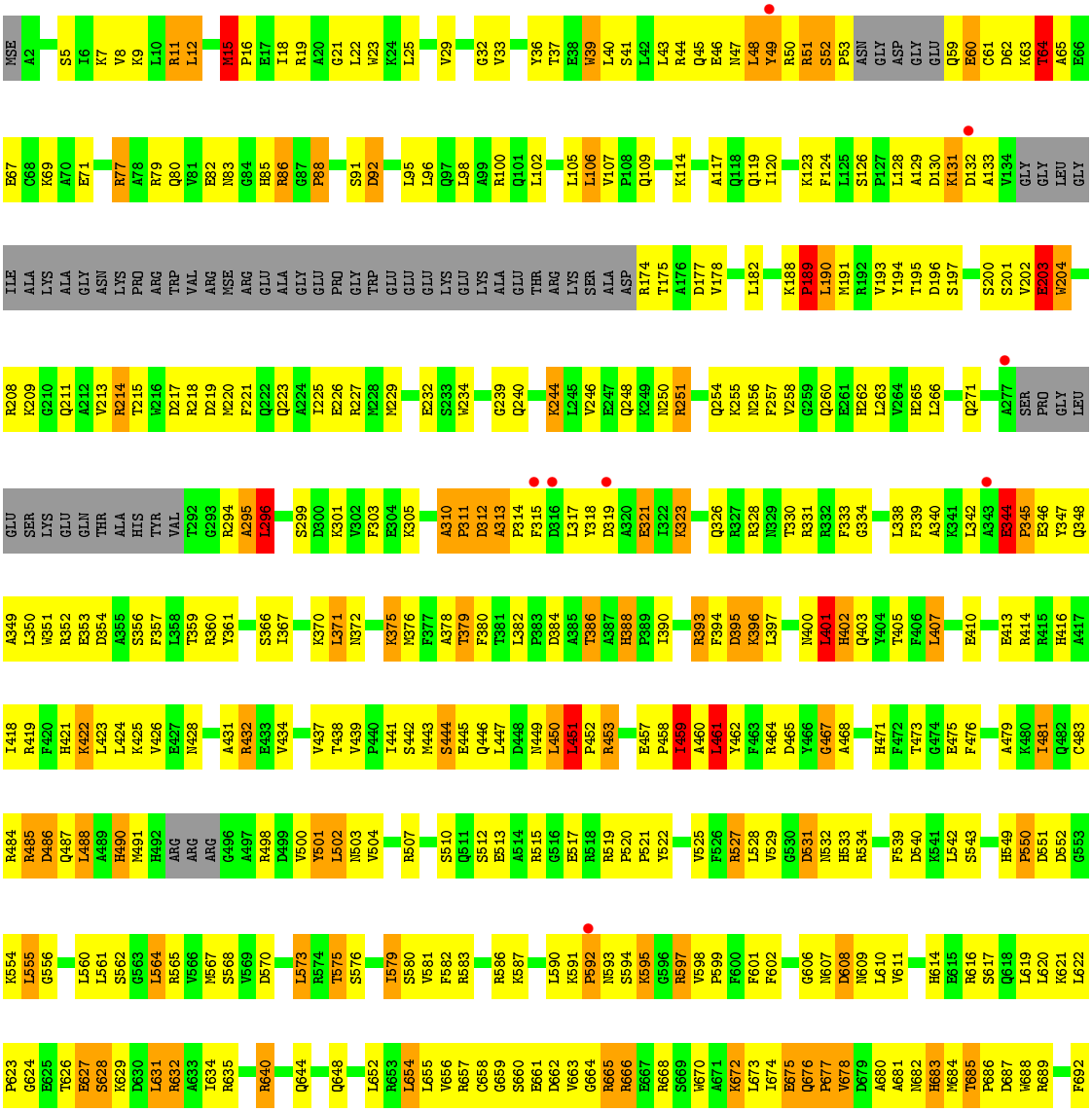
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	39	Total	O	0	0
			39	39		
3	B	3	Total	O	0	0
			3	3		

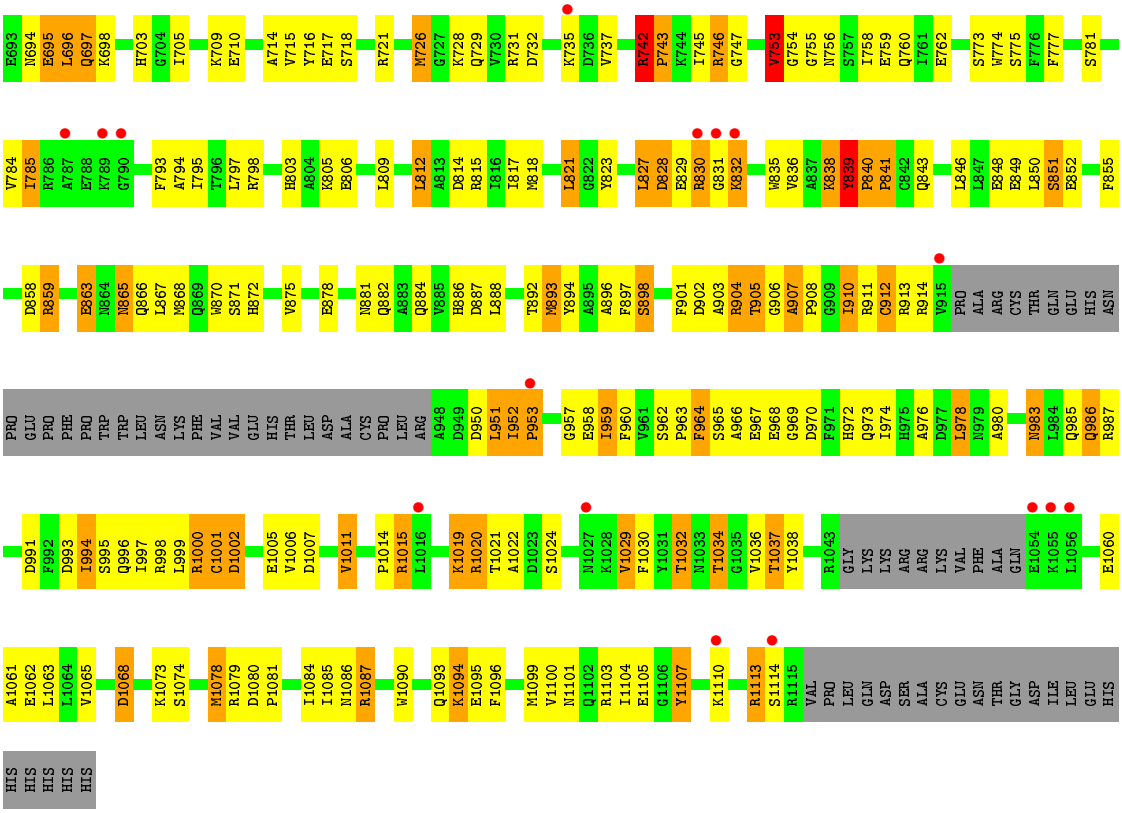
3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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: CRISPR-associated endonuclease C2c1

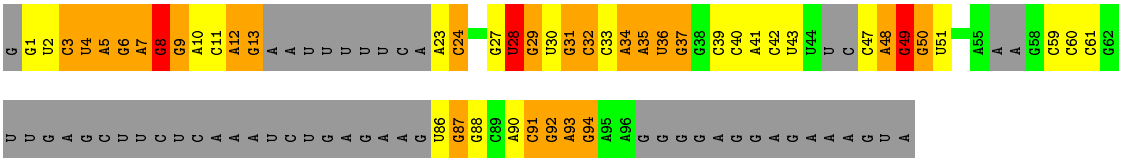
Chain A:





- Molecule 2: RNA (60-MER)

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	194.55Å 129.84Å 84.00Å 90.00° 110.00° 90.00°	Depositor
Resolution (Å)	48.64 – 3.13 48.64 – 3.13	Depositor EDS
% Data completeness (in resolution range)	79.3 (48.64-3.13) 79.3 (48.64-3.13)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.87 (at 3.12Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.238 , 0.268 0.239 , 0.269	Depositor DCC
R_{free} test set	1340 reflections (4.86%)	wwPDB-VP
Wilson B-factor (Å^2)	58.1	Xtriage
Anisotropy	0.358	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.23 , 21.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.41$, $< L^2 > = 0.24$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.86	EDS
Total number of atoms	9308	wwPDB-VP
Average B, all atoms (Å^2)	48.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.54% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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

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	0.70	12/8134 (0.1%)	0.88	34/10953 (0.3%)
2	B	0.52	1/1427 (0.1%)	0.85	5/2217 (0.2%)
All	All	0.67	13/9561 (0.1%)	0.87	39/13170 (0.3%)

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	A	0	5

The worst 5 of 13 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	841	PRO	N-CD	-6.47	1.38	1.47
1	A	677	PRO	N-CD	5.42	1.55	1.47
1	A	345	PRO	N-CD	5.40	1.55	1.47
1	A	88	PRO	N-CD	5.39	1.55	1.47
1	A	520	PRO	N-CD	5.34	1.55	1.47

The worst 5 of 39 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	203	GLU	CB-CA-C	-13.91	82.58	110.40
1	A	296	LEU	N-CA-CB	13.47	137.34	110.40
1	A	753	VAL	N-CA-C	12.27	144.12	111.00
1	A	295	ALA	N-CA-C	-12.09	78.36	111.00
1	A	421	HIS	N-CA-C	-10.52	82.60	111.00

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	467	GLY	Peptide
1	A	531	ASP	Peptide
1	A	64	THR	Peptide
1	A	784	VAL	Peptide
1	A	912	CYS	Peptide

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	7988	0	7713	724	0
2	B	1278	0	645	129	0
3	A	39	0	0	3	0
3	B	3	0	0	0	0
All	All	9308	0	8358	825	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 47.

The worst 5 of 825 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)
2:B:8:G:H3'	2:B:9:G:C5'	1.71	1.19
1:A:575:THR:CG2	1:A:621:LYS:HE2	1.73	1.18
1:A:758:ILE:HG23	1:A:867:LEU:HD23	1.27	1.14
1:A:1073:LYS:HE2	1:A:1073:LYS:HA	1.15	1.13
2:B:11:C:H2'	2:B:12:A:H5'	1.31	1.11

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	997/1137 (88%)	965 (97%)	24 (2%)	8 (1%)	19 53

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	296	LEU
1	A	401	LEU
1	A	459	ILE
1	A	952	ILE
1	A	344	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	804/955 (84%)	611 (76%)	193 (24%)	0 2

5 of 193 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	540	ASP
1	A	654	LEU
1	A	1024	SER
1	A	555	LEU
1	A	597	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 23 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	533	HIS
1	A	648	GLN
1	A	881	ASN
1	A	618	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	697	GLN

5.3.3 RNA

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	B	55/112 (49%)	25 (45%)	9 (16%)

5 of 25 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	B	4	U
2	B	5	A
2	B	6	G
2	B	7	A
2	B	8	G

5 of 9 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	B	36	U
2	B	92	G
2	B	50	G
2	B	28	U
2	B	49	G

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^{th} 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	992/1137 (87%)	-0.21	24 (2%) 59 37	11, 41, 87, 134	6 (0%)
2	B	60/112 (53%)	-0.35	0 100 100	22, 61, 124, 146	0
All	All	1052/1249 (84%)	-0.22	24 (2%) 60 39	11, 43, 90, 146	6 (0%)

The worst 5 of 24 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	831	GLY	7.4
1	A	790	GLY	6.6
1	A	1055	LYS	5.3
1	A	830	ARG	4.2
1	A	316	ASP	3.7

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.