



Full wwPDB X-ray Structure Validation Report ⓘ

Feb 13, 2017 – 08:24 PM EST

PDB ID : 5I0G
Title : Cycloalternan-degrading enzyme from *Trueperella pyogenes* in complex with cycloalternan
Authors : Light, S.H.; Minasov, G.; Anderson, W.F.; Center for Structural Genomics of Infectious Diseases (CSGID)
Deposited on : 2016-02-03
Resolution : 2.15 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20028442
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
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-20028442

i

X-RAY DIFFRACTION

A.

Metric

(#Entries)

 $(\# \text{Entries, resolution range}(\text{\AA}))$ R_{free}

9134

Clashscore

10224

Ramachandran outliers

10038

Sidechain outliers

10036

RSRZ outliers

91569

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

0.0000

1

B

733

96%

●

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 6224 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.

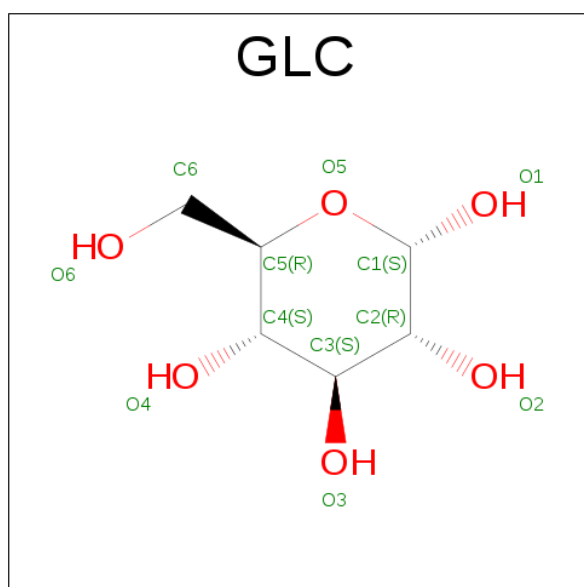
- Molecule 1 is a protein called Glycoside hydrolase family 31.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	B	730	5789	3667	1017	1086	19	0	0	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	532	ALA	ASP	engineered mutation	UNP X4QP62
B	730	GLY	-	expression tag	UNP X4QP62
B	731	ALA	-	expression tag	UNP X4QP62
B	732	ALA	-	expression tag	UNP X4QP62
B	733	LEU	-	expression tag	UNP X4QP62

- Molecule 2 is ALPHA-D-GLUCOSE (three-letter code: GLC) (formula: C₆H₁₂O₆).



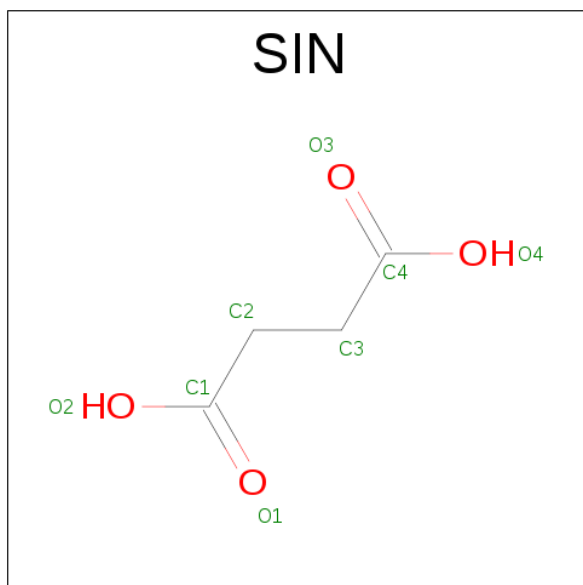
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
2	B	1	11	6	5	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	B	1	Total	C	O	0	0
			11	6	5		
2	B	1	Total	C	O	0	0
			11	6	5		
2	B	1	Total	C	O	0	0
			11	6	5		

- Molecule 3 is SUCCINIC ACID (three-letter code: SIN) (formula: C₄H₆O₄).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	B	1	Total	C	O	0	0
			8	4	4		

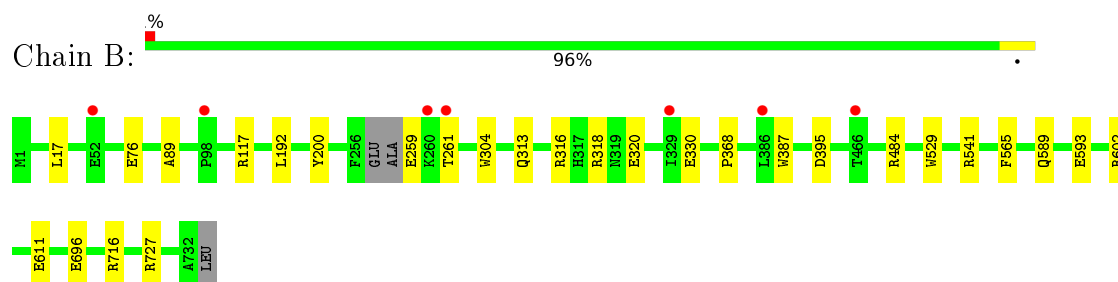
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	383	Total	O	0	0
			383	383		

3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: Glycoside hydrolase family 31



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	196.37Å 103.90Å 44.80Å 90.00° 90.32° 90.00°	Depositor
Resolution (Å)	28.48 – 2.15 28.48 – 2.15	Depositor EDS
% Data completeness (in resolution range)	99.4 (28.48-2.15) 99.4 (28.48-2.15)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.06 (at 2.16Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
R, R_{free}	0.172 , 0.210 0.178 , 0.218	Depositor DCC
R_{free} test set	2526 reflections (5.48%)	DCC
Wilson B-factor (Å ²)	28.2	Xtriage
Anisotropy	0.215	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 33.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.033 for -h,-k,l	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	6224	wwPDB-VP
Average B, all atoms (Å ²)	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.46% 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](#)

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

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	B	0.50	0/5970	0.73	7/8151 (0.1%)

There are no bond length outliers.

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	B	484	ARG	NE-CZ-NH1	6.39	123.50	120.30
1	B	484	ARG	NE-CZ-NH2	-5.91	117.34	120.30
1	B	727	ARG	NE-CZ-NH2	-5.63	117.48	120.30
1	B	318	ARG	NE-CZ-NH2	5.46	123.03	120.30
1	B	541	ARG	NE-CZ-NH2	5.36	122.98	120.30
1	B	117	ARG	NE-CZ-NH1	5.35	122.98	120.30
1	B	727	ARG	NE-CZ-NH1	5.27	122.94	120.30

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	B	5789	0	5442	5	0
2	B	44	0	36	0	0
3	B	8	0	4	0	0
4	B	383	0	0	0	1

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	6224	0	5482	5	1

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 0.

All (5) 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:316:ARG:NH1	1:B:320:GLU:OE2	2.29	0.65
1:B:304:TRP:CE2	1:B:313:GLN:HG3	2.51	0.45
1:B:200:TYR:OH	1:B:593:GLU:OE2	2.31	0.45
1:B:17:LEU:HD11	1:B:89:ALA:HB1	2.00	0.43
1:B:330:GLU:HA	1:B:387:TRP:HB2	2.03	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:B:1219:HOH:O	4:B:1219:HOH:O[2_557]	1.79	0.41

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	B	726/733 (99%)	708 (98%)	18 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	B	593/596 (100%)	580 (98%)	13 (2%)	60	63

All (13) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	76	GLU
1	B	192	LEU
1	B	259	GLU
1	B	261	THR
1	B	368	PRO
1	B	395	ASP
1	B	529	TRP
1	B	565	PHE
1	B	589	GLN
1	B	602	ARG
1	B	611	GLU
1	B	696	GLU
1	B	716	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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

5 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
2	GLC	B	801	2	11,11,12	0.47	0	15,15,17	1.15	0
2	GLC	B	802	2	11,11,12	0.31	0	15,15,17	0.92	1 (6%)
2	GLC	B	803	2	11,11,12	0.25	0	15,15,17	0.78	0
2	GLC	B	804	2	11,11,12	0.32	0	15,15,17	0.86	0
3	SIN	B	805	-	1,7,7	0.10	0	2,8,8	1.09	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	GLC	B	801	2	-	0/2/19/22	0/1/1/1
2	GLC	B	802	2	-	0/2/19/22	0/1/1/1
2	GLC	B	803	2	-	0/2/19/22	0/1/1/1
2	GLC	B	804	2	-	0/2/19/22	0/1/1/1
3	SIN	B	805	-	-	0/1/5/5	0/0/0/0

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	802	GLC	C1-O5-C5	2.81	116.28	112.14

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, 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	B	730/733 (99%)	-0.29	7 (0%) 84 88	20, 29, 45, 83	0

All (7) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	52	GLU	4.1
1	B	260	LYS	3.6
1	B	386	LEU	2.3
1	B	98	PRO	2.2
1	B	261	THR	2.1
1	B	329	ILE	2.0
1	B	466	THR	2.0

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](#)

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th 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	LLDF	B-factors(\AA^2)	Q<0.9
2	GLC	B	804	11/12	0.94	0.11	-0.54	27,27,29,30	0
2	GLC	B	801	11/12	0.96	0.08	-0.87	25,27,28,30	0
2	GLC	B	803	11/12	0.95	0.08	-1.38	29,30,32,33	0
2	GLC	B	802	11/12	0.94	0.12	-	30,33,35,36	0
3	SIN	B	805	8/8	0.73	0.17	-	62,65,67,68	0

6.5 Other polymers [i](#)

There are no such residues in this entry.