Full wwPDB X-ray Structure Validation Report ⓘ

May 23, 2020 – 04:12 pm BST

PDB ID	:	5HU7
Title	:	Crystal structure of the trans-AT PKS dehydratase domain of C0ZGQ4 from Brevibacillus brevis
Authors	:	Jakob, R.P.; Hauswirth, P.; Dilmi, J.; Herbst, D.A.; Maier, T.
Deposited on	:	2016-01-27
Resolution	:	2.40 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
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:

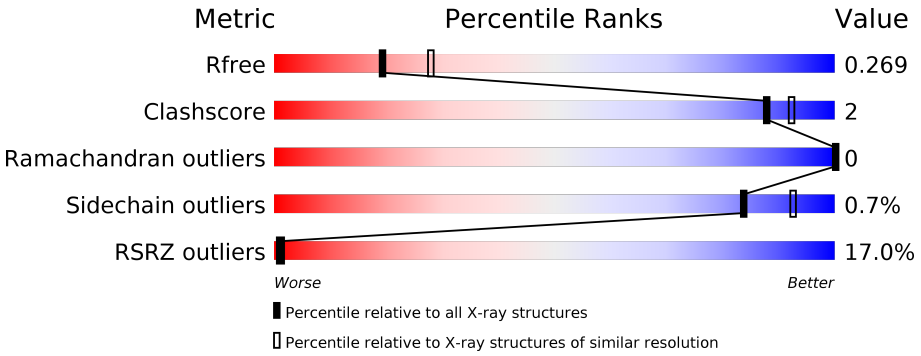
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 2.40 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	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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	322	9% 75% 6% 19%
1	B	322	18% 76% . 20%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 4124 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 Putative polyketide synthase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	261	Total	C	N	O	S	0	0	0
			2050	1302	338	400	10			
1	B	256	Total	C	N	O	S	0	0	0
			2009	1282	331	387	9			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	2839	SER	-	expression tag	UNP C0ZGQ4
A	2840	MET	-	expression tag	UNP C0ZGQ4
B	2839	SER	-	expression tag	UNP C0ZGQ4
B	2840	MET	-	expression tag	UNP C0ZGQ4

- Molecule 2 is water.

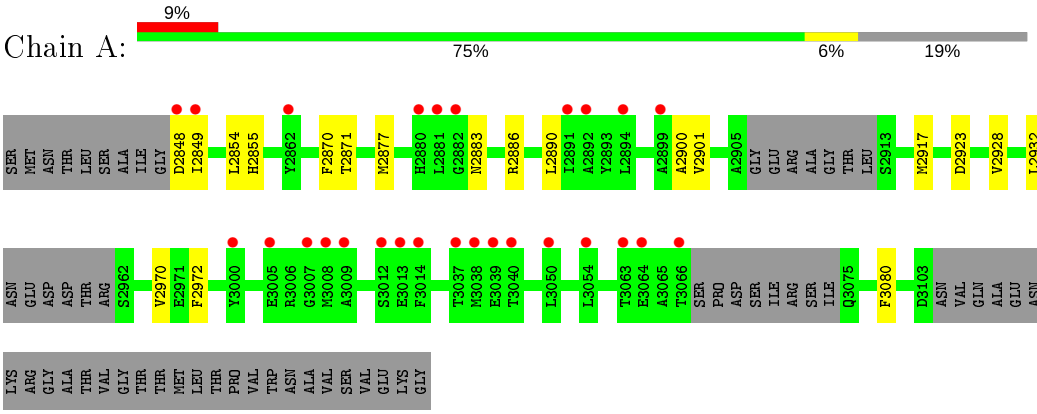
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	42	Total	O	0	0
			42	42		
2	B	23	Total	O	0	0
			23	23		

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.

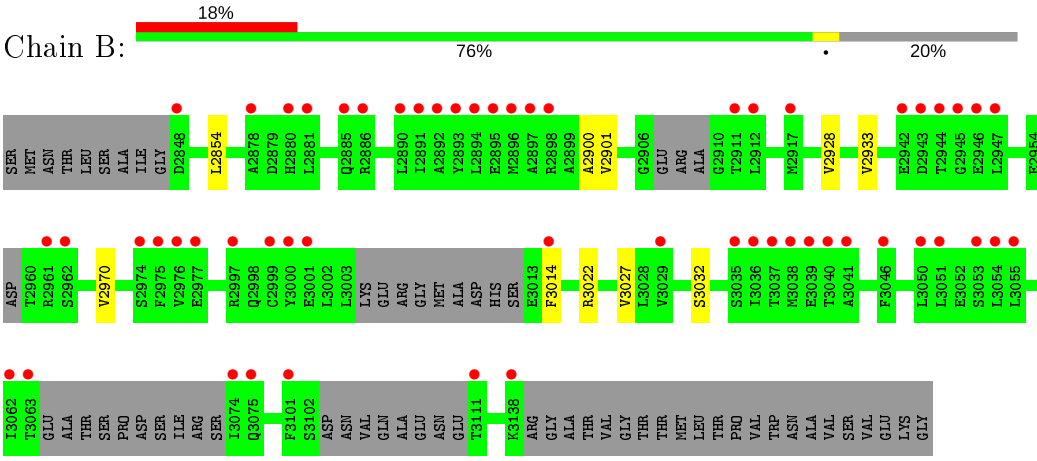
- Molecule 1: Putative polyketide synthase

Chain A:



- Molecule 1: Putative polyketide synthase

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	160.61Å 83.01Å 48.81Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	28.44 – 2.40 27.27 – 2.40	Depositor EDS
% Data completeness (in resolution range)	99.6 (28.44-2.40) 99.6 (27.27-2.40)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.05 (at 2.39Å)	Xtriage
Refinement program	BUSTER 2.10.2	Depositor
R, R_{free}	0.224 , 0.264 0.233 , 0.269	Depositor DCC
R_{free} test set	1311 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	66.4	Xtriage
Anisotropy	0.697	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 73.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	4124	wwPDB-VP
Average B, all atoms (Å^2)	98.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.81% 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	0.52	0/2089	0.69	0/2828
1	B	0.47	0/2046	0.68	0/2769
All	All	0.49	0/4135	0.68	0/5597

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

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	2050	0	2002	12	0
1	B	2009	0	1982	4	0
2	A	42	0	0	0	0
2	B	23	0	0	0	0
All	All	4124	0	3984	16	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 2.

All (16) 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:2901:VAL:HG11	1:A:2970:VAL:HG11	1.71	0.72

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:2890:LEU:HD22	1:A:3080:PHE:HD2	1.63	0.63
1:A:2890:LEU:HD22	1:A:3080:PHE:CD2	2.33	0.63
1:A:2848:ASP:O	1:A:2849:ILE:HD13	2.09	0.51
1:A:2917:MET:HG2	1:A:2970:VAL:HG22	1.95	0.49
1:A:2883:ASN:HB2	1:A:2886:ARG:HH12	1.80	0.47
1:A:2871:THR:HG22	1:A:2932:LEU:CD1	2.46	0.46
1:B:3022:ARG:HG3	1:B:3027:VAL:HG12	1.98	0.46
1:A:2901:VAL:HG13	1:A:2972:PHE:CZ	2.52	0.45
1:B:2854:LEU:HD11	1:B:2900:ALA:HA	2.00	0.44
1:A:2854:LEU:HD11	1:A:2900:ALA:HA	2.01	0.43
1:A:2928:VAL:HG12	1:A:2933:VAL:HG21	2.02	0.41
1:A:2849:ILE:HG23	1:A:2855:HIS:HA	2.02	0.41
1:B:2901:VAL:HG11	1:B:2970:VAL:HG11	2.03	0.41
1:A:2870:PHE:HB3	1:A:2877:MET:HE2	2.03	0.41
1:B:2928:VAL:HG12	1:B:2933:VAL:HG21	2.02	0.40

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	251/322 (78%)	244 (97%)	7 (3%)	0	100	100
1	B	244/322 (76%)	237 (97%)	7 (3%)	0	100	100
All	All	495/644 (77%)	481 (97%)	14 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	225/275 (82%)	224 (100%)	1 (0%)	91 96
1	B	221/275 (80%)	219 (99%)	2 (1%)	78 90
All	All	446/550 (81%)	443 (99%)	3 (1%)	84 92

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	2923	ASP
1	B	3014	PHE
1	B	3032	SER

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

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	261/322 (81%)	0.52	29 (11%)	5	4	51, 80, 140, 172	0
1	B	256/322 (79%)	1.08	59 (23%)	0	0	72, 103, 161, 197	0
All	All	517/644 (80%)	0.79	88 (17%)	1	1	51, 91, 155, 197	0

All (88) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	3007	GLY	14.6
1	B	2911	THR	9.1
1	A	3008	MET	7.9
1	A	2881	LEU	7.8
1	B	2943	ASP	7.5
1	B	2976	VAL	7.3
1	B	3039	GLU	7.1
1	A	2848	ASP	6.6
1	A	3038	MET	5.7
1	B	3037	THR	5.7
1	B	3074	ILE	5.4
1	B	2975	PHE	5.3
1	B	3111	THR	5.2
1	B	3035	SER	5.2
1	A	3012	SER	5.1
1	B	3051	LEU	5.0
1	B	2881	LEU	4.8
1	B	3058	THR	4.8
1	B	2891	ILE	4.7
1	B	3038	MET	4.7
1	B	3060	TYR	4.7
1	B	3036	ILE	4.6
1	B	3040	THR	4.6
1	A	3014	PHE	4.5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	3039	GLU	4.4
1	B	2944	THR	4.4
1	B	3001	GLU	4.3
1	B	3138	LYS	4.2
1	B	2942	GLU	4.2
1	B	2977	GLU	4.1
1	B	2894	LEU	4.0
1	B	2912	LEU	3.9
1	A	2891	ILE	3.8
1	B	2945	GLY	3.7
1	B	3101	PHE	3.6
1	B	3055	LEU	3.5
1	A	2849	ILE	3.4
1	B	2880	HIS	3.4
1	B	2896	MET	3.4
1	B	2974	SER	3.3
1	A	2882	GLY	3.3
1	B	2848	ASP	3.2
1	B	3053	SER	3.1
1	A	2899	ALA	3.1
1	B	2897	ALA	3.0
1	A	3063	THR	3.0
1	A	3000	TYR	3.0
1	B	2898	ARG	3.0
1	B	2997	ARG	3.0
1	B	2878	ALA	3.0
1	B	3000	TYR	2.9
1	B	3063	THR	2.9
1	A	3066	THR	2.8
1	B	2946	GLU	2.7
1	B	3041	ALA	2.7
1	A	2894	LEU	2.7
1	B	2961	ARG	2.6
1	A	2954	GLU	2.6
1	B	2890	LEU	2.6
1	A	2862	TYR	2.5
1	A	3037	THR	2.5
1	B	2885	GLN	2.5
1	A	3137	GLU	2.5
1	B	2893	TYR	2.5
1	A	3050	LEU	2.5
1	B	2917	MET	2.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	3054	LEU	2.4
1	A	2880	HIS	2.4
1	A	3013	GLU	2.4
1	B	2999	CYS	2.4
1	B	3061	LEU	2.3
1	B	2962	SER	2.2
1	B	2892	ALA	2.2
1	A	3054	LEU	2.2
1	B	3050	LEU	2.2
1	B	3075	GLN	2.2
1	B	2886	ARG	2.2
1	B	3029	VAL	2.2
1	A	2892	ALA	2.2
1	B	2895	GLU	2.2
1	B	3062	ILE	2.1
1	A	3064	GLU	2.1
1	B	3046	PHE	2.1
1	A	3040	THR	2.1
1	B	3014	PHE	2.1
1	A	3009	ALA	2.0
1	A	3005	GLU	2.0
1	B	2947	LEU	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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.