Full wwPDB X-ray Structure Validation Report ⓘ

May 26, 2020 – 11:19 pm BST

PDB ID	:	5HST
Title	:	Crystal structure of the dehydratase domain of MlnB from Bacillus amyloliquefaciens
Authors	:	Jakob, R.P.; Dilmi, J.; Hauswirth, P.; Herbst, D.A.; Maier, T.
Deposited on	:	2016-01-26
Resolution	:	2.10 Å(reported)

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

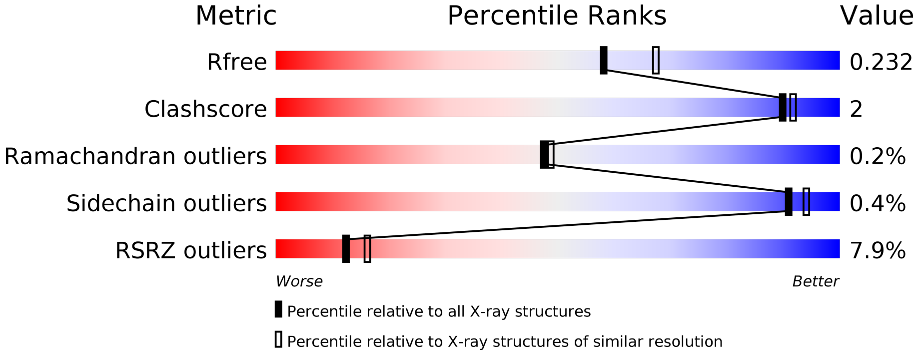
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

1 Overall quality at a glance

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

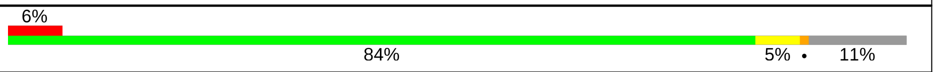
The reported resolution of this entry is 2.10 Å.

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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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	303	
1	B	303	

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 8744 atoms, of which 4159 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 Polyketide synthase type I.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	270	4217	1348	2096	358	406	9	0	1	0
1	B	265	4149	1328	2063	350	399	9	0	1	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	SER	-	expression tag	UNP Q1RS66
A	2	MET	-	expression tag	UNP Q1RS66
B	1	SER	-	expression tag	UNP Q1RS66
B	2	MET	-	expression tag	UNP Q1RS66

- Molecule 2 is water.

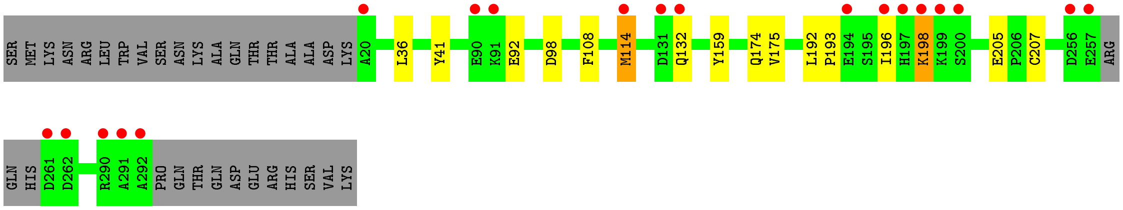
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	206	206	206	0	0
2	B	172	172	172	0	0

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 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.

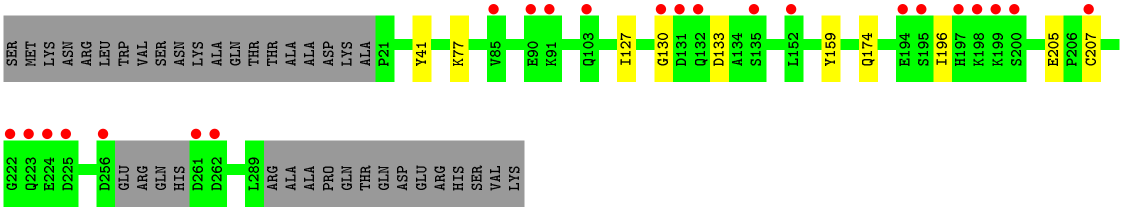
- Molecule 1: Polyketide synthase type I

Chain A:



- Molecule 1: Polyketide synthase type I

Chain B:



4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 2 21 21	Depositor
Cell constants a, b, c, α, β, γ	57.66Å 77.44Å 138.90Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	36.10 – 2.10 36.10 – 2.10	Depositor EDS
% Data completeness (in resolution range)	99.7 (36.10-2.10) 99.7 (36.10-2.10)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.62 (at 2.10Å)	Xtriage
Refinement program	PHENIX (1.10pre_2131: ???)	Depositor
R, R_{free}	0.186 , 0.228 0.188 , 0.232	Depositor DCC
R_{free} test set	2000 reflections (5.41%)	wwPDB-VP
Wilson B-factor (Å^2)	27.9	Xtriage
Anisotropy	0.504	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 45.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	8744	wwPDB-VP
Average B, all atoms (Å^2)	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.94% 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	0.28	0/2166	0.51	0/2928
1	B	0.29	0/2131	0.49	0/2880
All	All	0.28	0/4297	0.50	0/5808

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	2121	2096	2097	10	0
1	B	2086	2063	2064	5	0
2	A	206	0	0	1	0
2	B	172	0	0	0	0
All	All	4585	4159	4161	15	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 (15) 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:41:TYR:OH	1:A:205:GLU:OE1	2.02	0.78

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:41:TYR:OH	1:B:205:GLU:OE1	2.11	0.67
1:A:114:MET:SD	1:A:114:MET:N	2.81	0.53
1:B:77:LYS:HB3	1:B:127:ILE:HB	1.95	0.49
1:A:132:GLN:NE2	2:A:409:HOH:O	2.46	0.48
1:A:198:LYS:HB3	1:A:205:GLU:HG2	1.96	0.47
1:A:192:LEU:HD22	1:A:207[B]:CYS:SG	2.56	0.45
1:A:132:GLN:HG2	1:A:132:GLN:O	2.16	0.44
1:A:159:TYR:CZ	1:A:174:GLN:HA	2.53	0.44
1:A:36:LEU:O	1:A:92:GLU:HA	2.17	0.44
1:B:130:GLY:H	1:B:133:ASP:CG	2.21	0.43
1:A:175:VAL:HG12	1:A:193:PRO:HD3	2.00	0.42
1:B:196:ILE:HD13	1:B:207[A]:CYS:SG	2.59	0.42
1:B:159:TYR:CZ	1:B:174:GLN:HA	2.54	0.41
1:A:98:ASP:O	1:A:108:PHE:HA	2.20	0.41

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	267/303 (88%)	260 (97%)	6 (2%)	1 (0%)	34	32
1	B	262/303 (86%)	257 (98%)	5 (2%)	0	100	100
All	All	529/606 (87%)	517 (98%)	11 (2%)	1 (0%)	47	49

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	196	ILE

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	229/258 (89%)	227 (99%)	2 (1%)	78	84
1	B	227/258 (88%)	227 (100%)	0	100	100
All	All	456/516 (88%)	454 (100%)	2 (0%)	91	94

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	114	MET
1	A	198	LYS

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, 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	270/303 (89%)	0.32	19 (7%)	16	20	16, 33, 79, 125	0
1	B	265/303 (87%)	0.43	23 (8%)	10	13	17, 33, 82, 120	0
All	All	535/606 (88%)	0.38	42 (7%)	12	16	16, 33, 81, 125	0

All (42) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	131	ASP	8.5
1	A	197	HIS	7.6
1	B	132	GLN	6.4
1	A	131	ASP	6.1
1	A	196	ILE	6.0
1	A	194	GLU	6.0
1	A	292	ALA	5.9
1	B	198	LYS	5.7
1	B	130	GLY	5.3
1	B	225	ASP	4.9
1	B	223	GLN	4.9
1	A	200	SER	4.8
1	A	261	ASP	4.7
1	B	200	SER	4.5
1	A	257	GLU	4.1
1	B	224	GLU	4.0
1	A	132	GLN	4.0
1	B	256	ASP	3.7
1	A	91	LYS	3.6
1	B	197	HIS	3.5
1	B	262	ASP	3.5
1	B	135	SER	3.3
1	B	91	LYS	3.1
1	A	291	ALA	3.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	198	LYS	3.0
1	A	199	LYS	3.0
1	A	90	GLU	2.8
1	B	222	GLY	2.8
1	A	256	ASP	2.7
1	A	20	ALA	2.7
1	B	199	LYS	2.6
1	B	90	GLU	2.6
1	B	194	GLU	2.4
1	B	152	LEU	2.3
1	B	103	GLN	2.3
1	A	262	ASP	2.2
1	A	114	MET	2.2
1	B	85	VAL	2.1
1	B	207[A]	CYS	2.1
1	B	195	SER	2.1
1	A	290	ARG	2.0
1	B	261	ASP	2.0

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.