W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

May 21, 2020 – 04:25 pm BST

PDB ID	:	4QK0
Title	:	Crystal structure of Ara127N-Se, a GH127 beta-L-arabinofuranosidase from Geobacillus Stearothermophilus T6
Authors	:	Lansky, S.; Salama, R.; Dann, R.; Shner, I.; Manjasetty, B.; Belrhali, H.; Shoham, Y.; Shoham, G.
Deposited on	:	2014-06-05
Resolution	:	2.26 Å(reported)

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

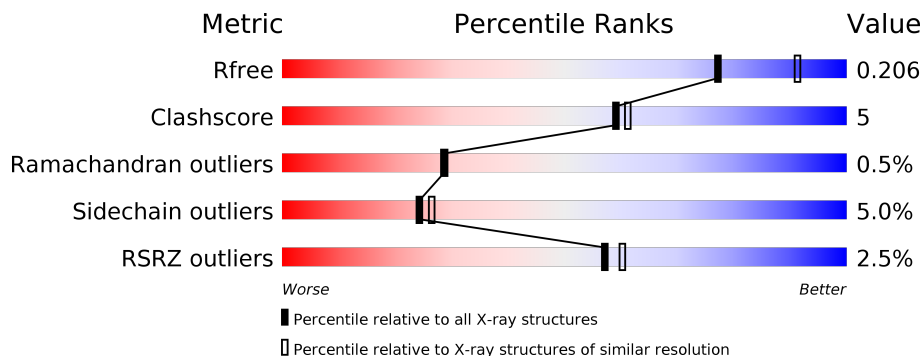|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry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

1 Overall quality at a glance ⓘ

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

The reported resolution of this entry is 2.26 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	Percentile Ranks	Value
Rfree		0.206
Clashscore		5
Ramachandran outliers		0.5%
Sidechain outliers		5.0%
RSRZ outliers		2.5%

Worse   *Better*

■ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	1377 (2.26-2.26)
Clashscore	141614	1487 (2.26-2.26)
Ramachandran outliers	138981	1449 (2.26-2.26)
Sidechain outliers	138945	1450 (2.26-2.26)
RSRZ outliers	127900	1356 (2.26-2.26)

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	648	2% 86% 9% ··
1	B	648	4% 85% 11% ··
1	C	648	2% 85% 11% ··
1	D	648	2% 87% 10% ··

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 22177 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 GH127 beta-L-arabinofuranoside.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	635	Total 5152	C 3297	N 880	O 949	S 11	Se 15	0	1	0
1	B	635	Total 5179	C 3316	N 883	O 952	S 11	Se 17	0	6	0
1	C	635	Total 5165	C 3306	N 883	O 949	S 11	Se 16	0	3	0
1	D	635	Total 5163	C 3304	N 881	O 952	S 11	Se 15	0	3	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	HIS	-	EXPRESSION TAG	UNP B3EYN9
A	-2	HIS	-	EXPRESSION TAG	UNP B3EYN9
A	-1	HIS	-	EXPRESSION TAG	UNP B3EYN9
A	0	HIS	-	EXPRESSION TAG	UNP B3EYN9
A	1	HIS	-	EXPRESSION TAG	UNP B3EYN9
A	2	HIS	-	EXPRESSION TAG	UNP B3EYN9
B	-3	HIS	-	EXPRESSION TAG	UNP B3EYN9
B	-2	HIS	-	EXPRESSION TAG	UNP B3EYN9
B	-1	HIS	-	EXPRESSION TAG	UNP B3EYN9
B	0	HIS	-	EXPRESSION TAG	UNP B3EYN9
B	1	HIS	-	EXPRESSION TAG	UNP B3EYN9
B	2	HIS	-	EXPRESSION TAG	UNP B3EYN9
C	-3	HIS	-	EXPRESSION TAG	UNP B3EYN9
C	-2	HIS	-	EXPRESSION TAG	UNP B3EYN9
C	-1	HIS	-	EXPRESSION TAG	UNP B3EYN9
C	0	HIS	-	EXPRESSION TAG	UNP B3EYN9
C	1	HIS	-	EXPRESSION TAG	UNP B3EYN9
C	2	HIS	-	EXPRESSION TAG	UNP B3EYN9
D	-3	HIS	-	EXPRESSION TAG	UNP B3EYN9
D	-2	HIS	-	EXPRESSION TAG	UNP B3EYN9
D	-1	HIS	-	EXPRESSION TAG	UNP B3EYN9

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	0	HIS	-	EXPRESSION TAG	UNP B3EYN9
D	1	HIS	-	EXPRESSION TAG	UNP B3EYN9
D	2	HIS	-	EXPRESSION TAG	UNP B3EYN9

- Molecule 2 is water.

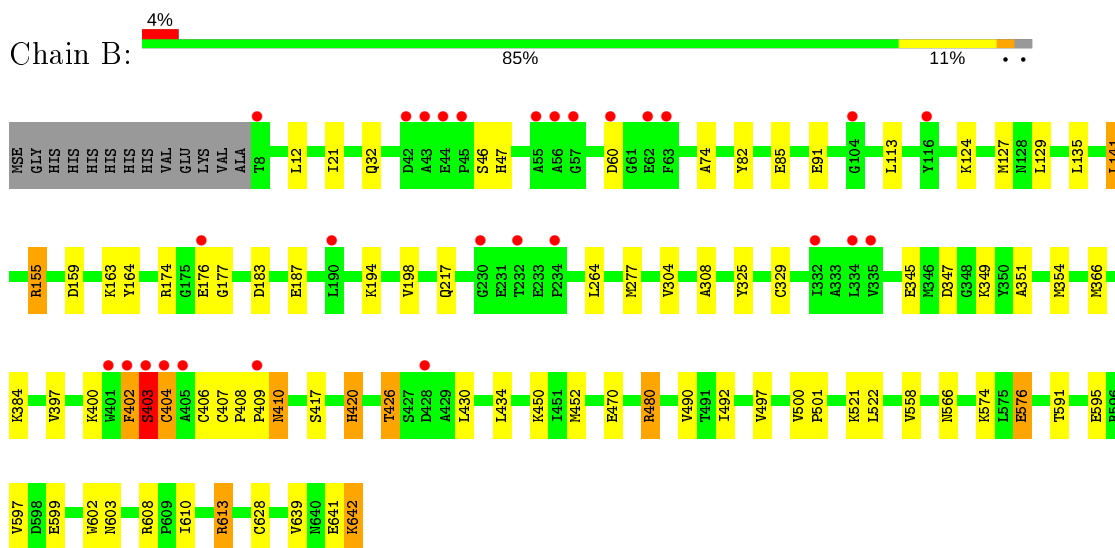| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	343	Total 343	O 343	0	0
2	B	311	Total 311	O 311	0	0
2	C	401	Total 401	O 401	0	0
2	D	463	Total 463	O 463	0	0

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.
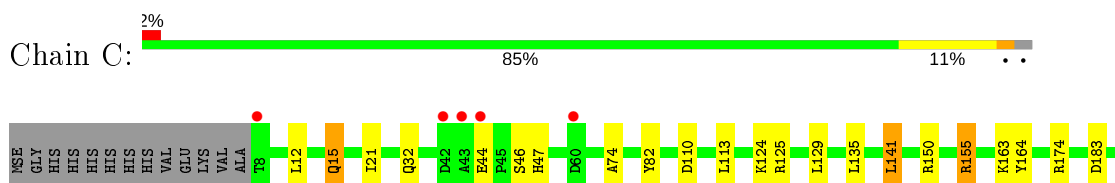
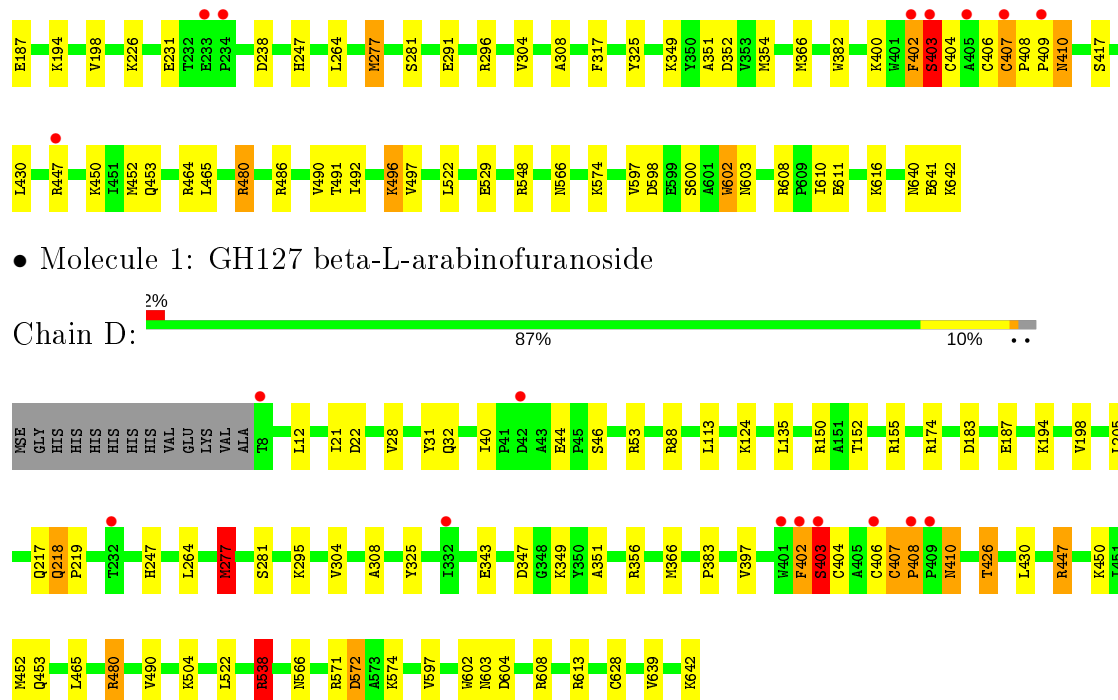- Molecule 1: GH127 beta-L-arabinofuranoside

- Molecule 1: GH127 beta-L-arabinofuranoside

- Molecule 1: GH127 beta-L-arabinofuranoside

E187 K194 V198 K226 E231 T232 E233 P234 D238 H247 L264 M277 S281 E291 R296 V304 A308 F317 Y325 K349 Y350 A351 D352 V353 M354 M366 W382 K400 W401 F402 S403 C404 A405 C406 C407 P408 P409 N410 S417

L430 R447 K450 I451 M452 Q453 R464 L465 R480 R486 V490 T491 I492 K496 V497 L522 E529 R548 N566 K574 V597 D598 E599 S600 A601 W602 N603 R608 P609 I610 E611 K616 N640 E641 K642

● Molecule 1: GH127 beta-L-arabinofuranoside

Chain D:

2%
87%   10%

MSE GLY HIS HIS HIS HIS HIS HIS VAL GLU LYS VAL ALA T8 L12 I21 D22 V28 Y31 Q32 I40 P41 D42 A43 E44 P45 S46 R53 R88 L113 K124 L135 R150 A151 T152 R155 R174 D183 E187 K194 V198 L205

Q217 Q218 P219 T232 H247 L264 M277 S281 K295 V304 A308 Y325 I332 E343 D347 G348 K349 Y350 A351 R356 M366 P383 V397 W401 F402 S403 C404 A405 C406 C407 P408 P409 M410 T426 L430 R447 K450 I451

M452 Q453 L465 R480 V490 K504 L522 R538 N566 R571 D572 A573 K574 V597 W602 N603 D604 R608 R613 C628 V639 K642

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	105.38Å 132.27Å 109.05Å 90.00° 113.06° 90.00°	Depositor
Resolution (Å)	40.00 – 2.26 39.97 – 2.26	Depositor EDS
% Data completeness (in resolution range)	99.5 (40.00-2.26) 99.6 (39.97-2.26)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	3.65 (at 2.27Å)	Xtriage
Refinement program	REFMAC 5.7.0032	Depositor
R, R_{free}	0.157 , 0.200 0.170 , 0.206	Depositor DCC
R_{free} test set	6452 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	29.0	Xtriage
Anisotropy	0.326	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 36.2	EDS
L-test for twinning[2]	$<|L|> = 0.49, <L^2> = 0.32$	Xtriage
Estimated twinning fraction	0.009 for l,-k,h	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	22177	wwPDB-VP
Average B, all atoms (Å^2)	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 17.18% 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 (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	0.80	0/5278	0.84	5/7137 (0.1%)
1	B	0.78	0/5320	0.82	3/7192 (0.0%)
1	C	0.85	0/5297	0.85	8/7161 (0.1%)
1	D	0.85	0/5295	0.87	9/7160 (0.1%)
All	All	0.82	0/21190	0.85	25/28650 (0.1%)

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	D	0	2

There are no bond length outliers.

The worst 5 of 25 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	538	ARG	NE-CZ-NH1	9.43	125.01	120.30
1	D	538	ARG	NE-CZ-NH2	-8.82	115.89	120.30
1	C	480	ARG	NE-CZ-NH2	-8.01	116.29	120.30
1	D	150	ARG	NE-CZ-NH2	-6.90	116.85	120.30
1	D	480	ARG	NE-CZ-NH2	-6.45	117.08	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	572[A]	ASP	Mainchain
1	D	572[B]	ASP	Mainchain

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	5152	0	4996	40	0
1	B	5179	0	5036	46	0
1	C	5165	0	5018	59	0
1	D	5163	0	5008	54	0
2	A	343	0	0	14	0
2	B	311	0	0	17	0
2	C	401	0	0	18	0
2	D	463	0	0	11	0
All	All	22177	0	20058	190	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 5.

The worst 5 of 190 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:642:LYS:HE2	1:D:642:LYS:HB3	1.44	0.98
1:A:329:CYS:SG	2:A:1019:HOH:O	2.22	0.97
1:C:642:LYS:HE2	1:D:642:LYS:CB	1.97	0.95
1:B:426:THR:HG21	2:B:895:HOH:O	1.65	0.94
1:B:595:GLU:OE1	1:B:642:LYS:HE3	1.75	0.87

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	634/648 (98%)	612 (96%)	19 (3%)	3 (0%)	29	29
1	B	639/648 (99%)	618 (97%)	18 (3%)	3 (0%)	29	29
1	C	636/648 (98%)	614 (96%)	19 (3%)	3 (0%)	29	29
1	D	636/648 (98%)	614 (96%)	19 (3%)	3 (0%)	29	29
All	All	2545/2592 (98%)	2458 (97%)	75 (3%)	12 (0%)	29	29

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	187	GLU
1	B	187	GLU
1	C	187	GLU
1	D	187	GLU
1	A	403	SER

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	535/530 (101%)	507 (95%)	28 (5%)	23	24
1	B	540/530 (102%)	510 (94%)	30 (6%)	21	21
1	C	537/530 (101%)	513 (96%)	24 (4%)	27	31
1	D	537/530 (101%)	512 (95%)	25 (5%)	26	29
All	All	2149/2120 (101%)	2042 (95%)	107 (5%)	24	26

5 of 107 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	430	LEU
1	C	46	SER
1	D	426	THR
1	B	450	LYS
1	B	610	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 25 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	15	GLN
1	C	247	HIS
1	D	515	GLN
1	C	217	GLN
1	C	377	ASN

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 carbohydrates 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, 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(Å²)	Q<0.9
1	A	621/648 (95%)	-0.33	11 (1%)	68	71	18, 32, 63, 108	0
1	B	621/648 (95%)	-0.03	28 (4%)	33	36	19, 37, 84, 118	0
1	C	621/648 (95%)	-0.37	13 (2%)	63	66	14, 27, 64, 101	0
1	D	621/648 (95%)	-0.47	10 (1%)	72	74	15, 25, 56, 99	0
All	All	2484/2592 (95%)	-0.30	62 (2%)	57	60	14, 30, 70, 118	0

The worst 5 of 62 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	409	PRO	7.0
1	B	404	CYS	5.8
1	A	409	PRO	5.7
1	B	403	SER	5.7
1	B	62	GLU	5.2

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.