



wwPDB X-ray Structure Validation Summary Report ⓘ

May 22, 2020 – 08:24 pm BST

PDB ID : 3S2C
Title : Structure of the thermostable GH51 alpha-L-arabinofuranosidase from Thermotoga petrophila RKU-1
Authors : Souza, T.A.C.B.; Santos, C.R.; Souza, A.R.; Oldiges, D.P.; Ruller, R.; Prade, R.A.; Squina, F.M.; Murakami, M.T.
Deposited on : 2011-05-16
Resolution : 3.00 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 ⓘ 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

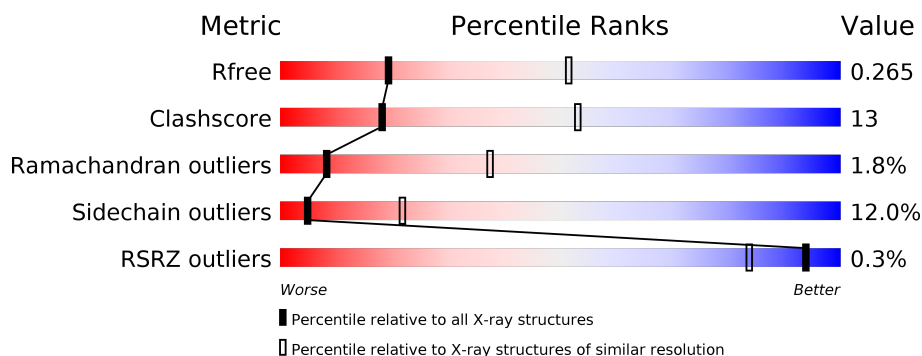
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.00 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	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)

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 $\leq 5\%$. 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
1	A	484	<div> <div>66%</div> <div>27%</div> <div>6%</div> </div>
1	B	484	<div> <div>70%</div> <div>25%</div> <div>5%</div> </div>
1	C	484	<div> <div>69%</div> <div>25%</div> <div>6%</div> </div>
1	D	484	<div> <div>65%</div> <div>29%</div> <div>5%</div> </div>
1	E	484	<div> <div>68%</div> <div>26%</div> <div>5%</div> </div>
1	F	484	<div> <div>66%</div> <div>29%</div> <div>5%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	484	<div><div></div><div>66%26%8%</div></div>
1	H	484	<div><div></div><div>62%30%6%</div></div>
1	I	484	<div><div></div><div>64%30%6%</div></div>
1	J	484	<div><div></div><div>68%26%5%</div></div>
1	K	484	<div><div></div><div>69%25%6%</div></div>
1	L	484	<div><div></div><div>%62%34%5%</div></div>

2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 47152 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 Alpha-N-arabinofuranosidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	483	Total	C	N	O	S	0	0	0
			3893	2499	649	730	15			
1	B	483	Total	C	N	O	S	0	0	0
			3893	2499	649	730	15			
1	C	483	Total	C	N	O	S	0	0	0
			3893	2499	649	730	15			
1	D	483	Total	C	N	O	S	0	0	0
			3893	2499	649	730	15			
1	E	483	Total	C	N	O	S	0	0	0
			3893	2499	649	730	15			
1	F	483	Total	C	N	O	S	0	0	0
			3893	2499	649	730	15			
1	G	483	Total	C	N	O	S	0	1	0
			3904	2505	653	731	15			
1	H	483	Total	C	N	O	S	0	0	0
			3893	2499	649	730	15			
1	I	483	Total	C	N	O	S	0	0	0
			3893	2499	649	730	15			
1	J	481	Total	C	N	O	S	0	0	0
			3876	2488	647	726	15			
1	K	483	Total	C	N	O	S	0	0	0
			3893	2499	649	730	15			
1	L	483	Total	C	N	O	S	0	1	0
			3902	2504	650	733	15			

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	42	Total	O	0	0
			42	42		
2	B	43	Total	O	0	0
			43	43		

Continued on next page...

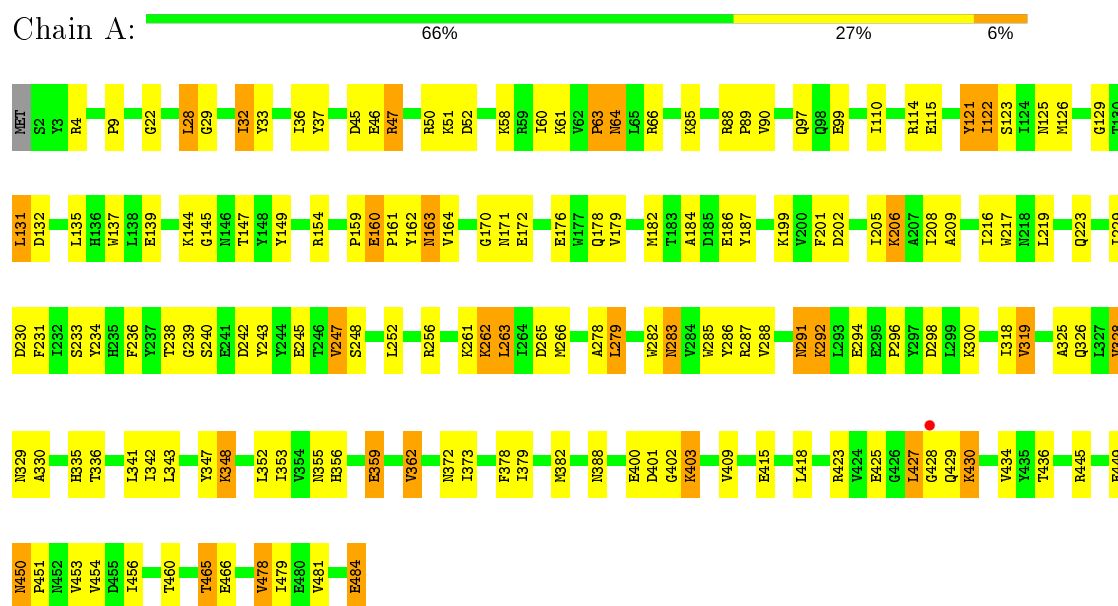
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	C	50	Total	O	0	0
			50	50		
2	D	54	Total	O	0	0
			54	54		
2	E	45	Total	O	0	0
			45	45		
2	F	39	Total	O	0	0
			39	39		
2	G	30	Total	O	0	0
			30	30		
2	H	28	Total	O	0	0
			28	28		
2	I	27	Total	O	0	0
			27	27		
2	J	23	Total	O	0	0
			23	23		
2	K	32	Total	O	0	0
			32	32		
2	L	20	Total	O	0	0
			20	20		

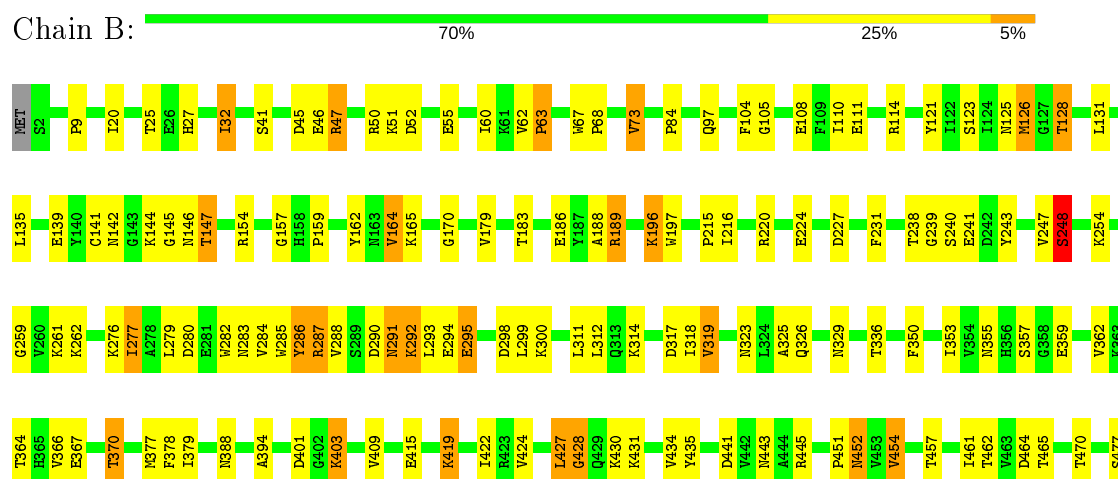
3 Residue-property plots

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: Alpha-N-arabinofuranosidase



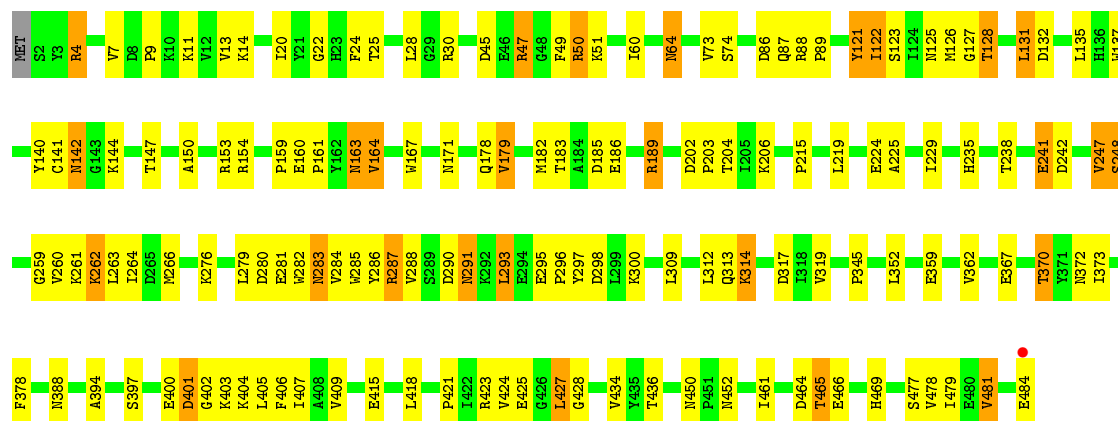
• Molecule 1: Alpha-N-arabinofuranosidase





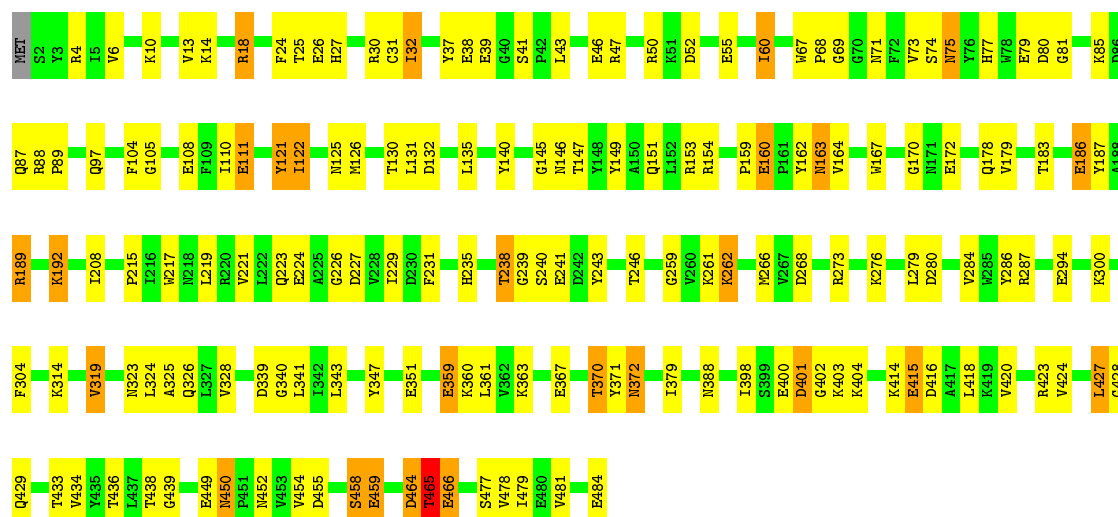
• Molecule 1: Alpha-N-arabinofuranosidase

Chain C: 69% 25% 6%



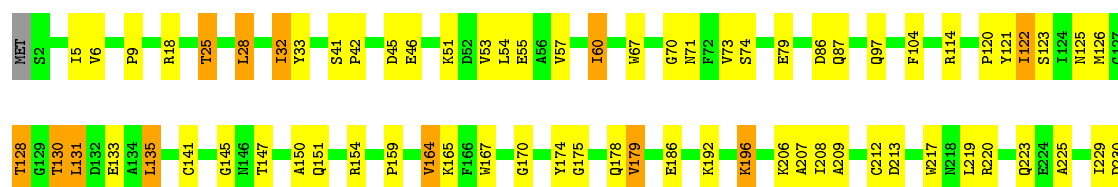
• Molecule 1: Alpha-N-arabinofuranosidase

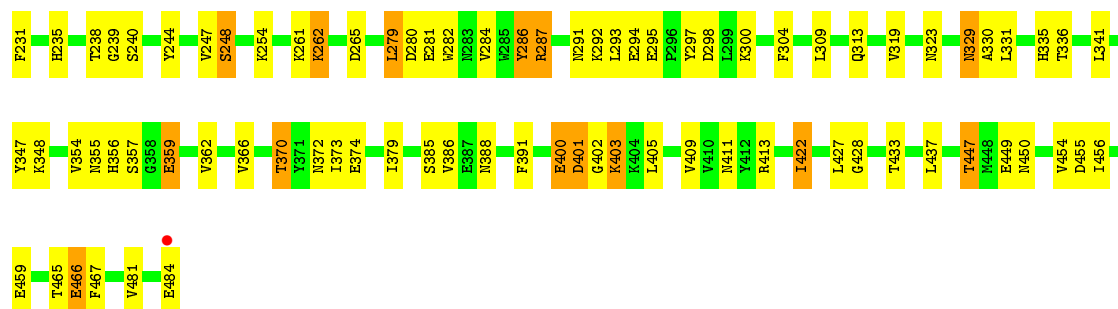
Chain D: 65% 29% 5%



• Molecule 1: Alpha-N-arabinofuranosidase

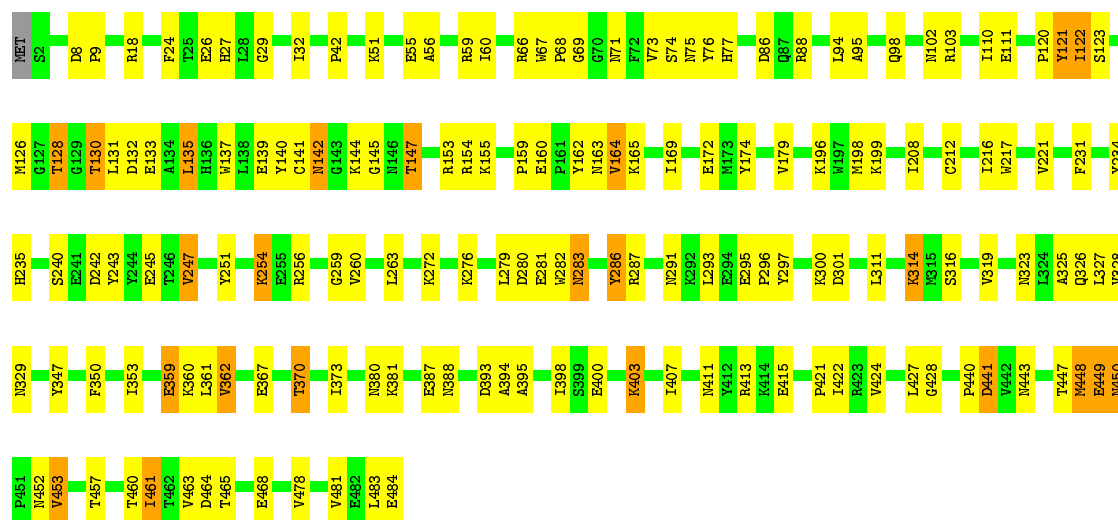
Chain E: 68% 26% 5%





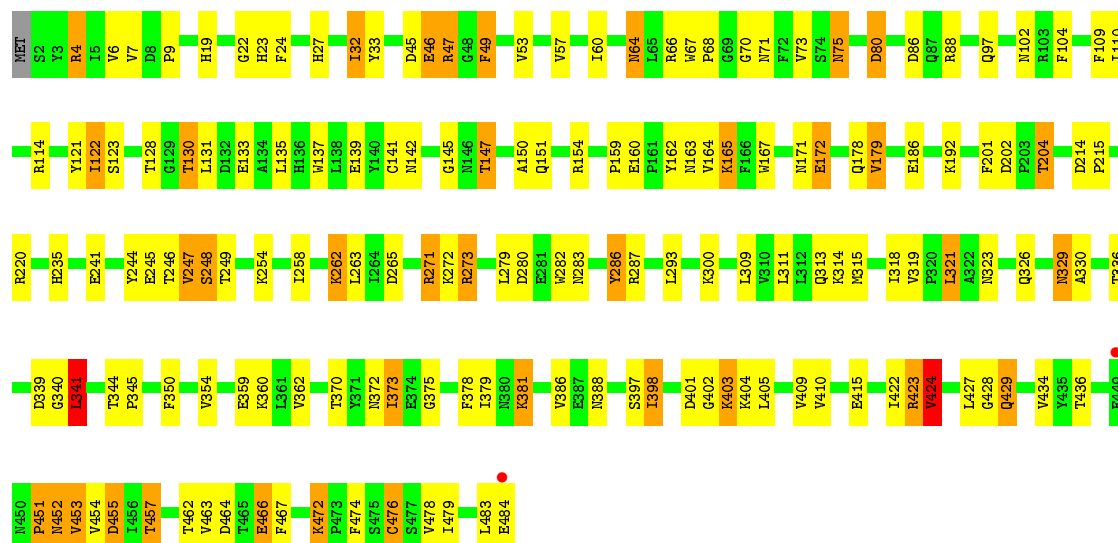
• Molecule 1: Alpha-N-arabinofuranosidase

Chain F: 66% 29% 5%

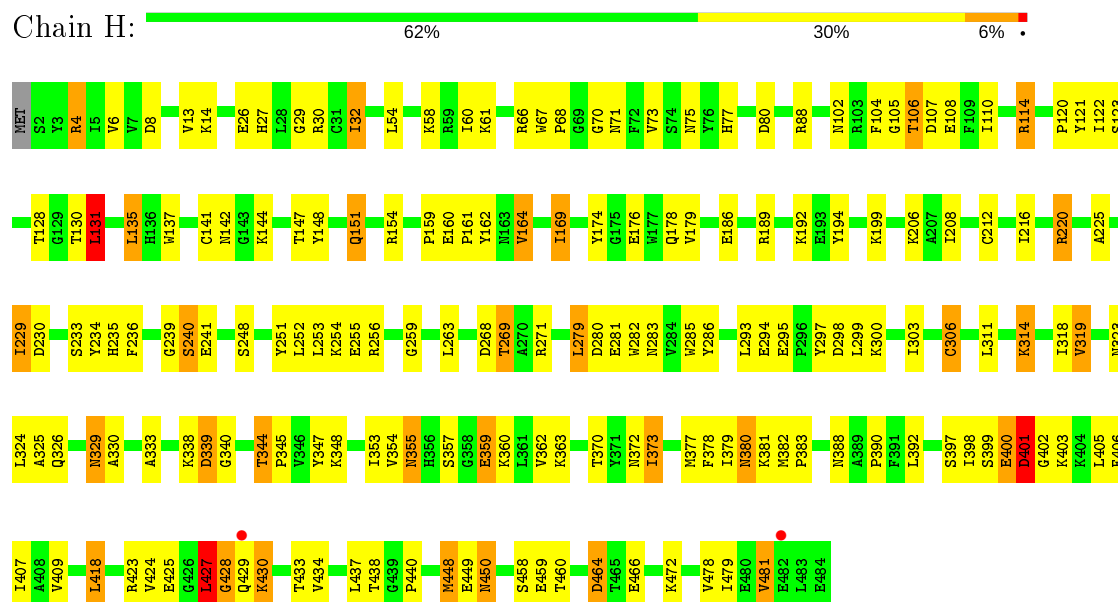


• Molecule 1: Alpha-N-arabinofuranosidase

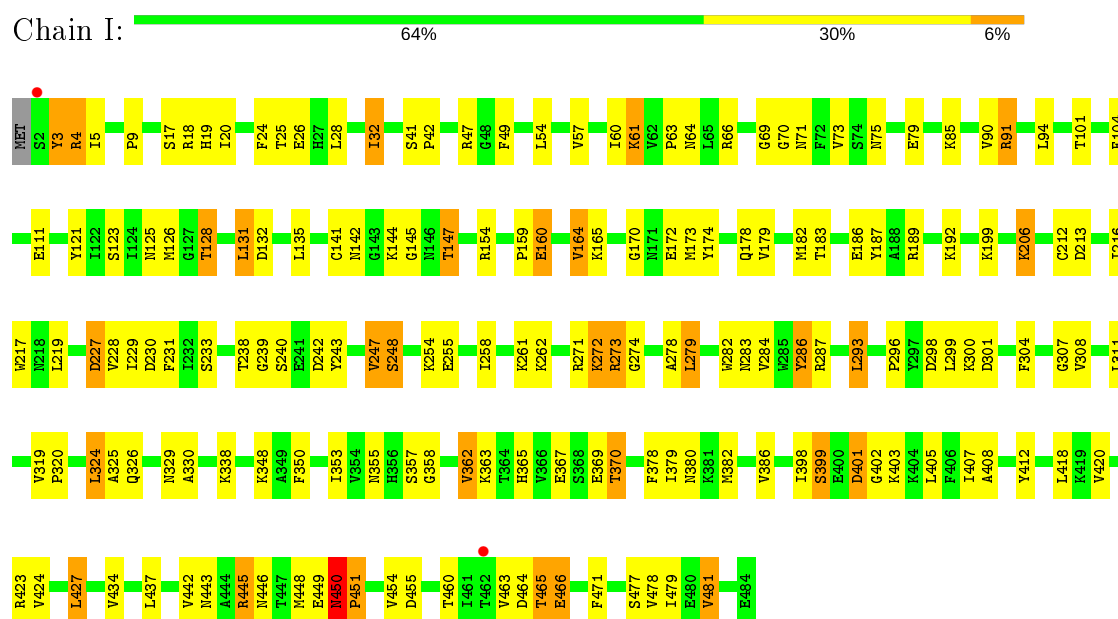
Chain G: 66% 26% 8%



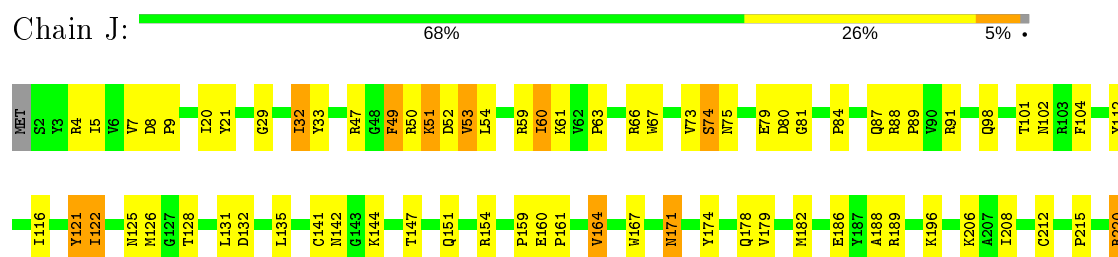
• Molecule 1: Alpha-N-arabinofuranosidase



• Molecule 1: Alpha-N-arabinofuranosidase



• Molecule 1: Alpha-N-arabinofuranosidase



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	105.95Å 187.29Å 180.64Å 90.00° 90.87° 90.00°	Depositor
Resolution (Å)	49.24 – 3.00 48.87 – 2.99	Depositor EDS
% Data completeness (in resolution range)	97.2 (49.24-3.00) 97.2 (48.87-2.99)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.01 (at 3.01Å)	Xtriage
Refinement program	REFMAC 5.5.0102	Depositor
R, R_{free}	0.195 , 0.268 0.195 , 0.265	Depositor DCC
R_{free} test set	6914 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å ²)	54.9	Xtriage
Anisotropy	0.120	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 51.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.000 for -h,-l,-k 0.000 for -h,l,k 0.013 for h,-k,-l	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	47152	wwPDB-VP
Average B, all atoms (Å ²)	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.38% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² Theoretical 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

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	0/3990	0.67	0/5413
1	B	0.55	0/3990	0.68	0/5413
1	C	0.55	0/3990	0.67	0/5413
1	D	0.54	0/3990	0.69	1/5413 (0.0%)
1	E	0.56	0/3990	0.69	0/5413
1	F	0.54	0/3990	0.68	0/5413
1	G	0.52	0/4001	0.69	1/5427 (0.0%)
1	H	0.53	0/3990	0.68	3/5413 (0.1%)
1	I	0.49	0/3990	0.64	1/5413 (0.0%)
1	J	0.48	0/3973	0.63	1/5390 (0.0%)
1	K	0.50	0/3990	0.66	0/5413
1	L	0.49	0/3999	0.66	0/5426
All	All	0.52	0/47883	0.67	7/64960 (0.0%)

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	1
1	B	0	1
1	D	0	2
1	J	0	1
1	K	0	2
All	All	0	7

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	G	341	LEU	CA-CB-CG	6.25	129.68	115.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	239	GLY	N-CA-C	-6.21	97.58	113.10
1	H	131	LEU	CA-CB-CG	6.16	129.46	115.30
1	H	427	LEU	CA-CB-CG	5.86	128.78	115.30
1	J	279	LEU	CA-CB-CG	5.74	128.50	115.30

There are no chirality outliers.

5 of 7 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	63	PRO	Peptide
1	B	63	PRO	Peptide
1	D	238	THR	Peptide
1	D	240	SER	Peptide
1	J	448	MET	Peptide

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	3893	0	3818	116	0
1	B	3893	0	3818	94	0
1	C	3893	0	3818	99	0
1	D	3893	0	3818	120	0
1	E	3893	0	3818	105	0
1	F	3893	0	3818	91	0
1	G	3904	0	3830	103	0
1	H	3893	0	3818	118	0
1	I	3893	0	3818	115	0
1	J	3876	0	3801	84	0
1	K	3893	0	3818	100	0
1	L	3902	0	3824	127	0
2	A	42	0	0	1	0
2	B	43	0	0	3	0
2	C	50	0	0	0	0
2	D	54	0	0	1	0
2	E	45	0	0	3	0
2	F	39	0	0	4	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	G	30	0	0	3	0
2	H	28	0	0	3	0
2	I	27	0	0	2	0
2	J	23	0	0	2	0
2	K	32	0	0	1	0
2	L	20	0	0	1	0
All	All	47152	0	45817	1236	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 13.

The worst 5 of 1236 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:291:ASN:HB3	2:B:495:HOH:O	1.43	1.19
1:E:125:ASN:HD21	1:E:128:THR:HG22	1.05	1.11
1:L:154:ARG:HG2	1:L:159:PRO:HA	1.36	1.05
1:G:4:ARG:HH11	1:G:4:ARG:HG3	1.21	1.02
1:F:154:ARG:HG2	1:F:159:PRO:HA	1.37	1.01

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	481/484 (99%)	434 (90%)	42 (9%)	5 (1%)	15 53
1	B	481/484 (99%)	435 (90%)	39 (8%)	7 (2%)	10 42
1	C	481/484 (99%)	433 (90%)	42 (9%)	6 (1%)	13 48
1	D	481/484 (99%)	434 (90%)	38 (8%)	9 (2%)	8 36
1	E	481/484 (99%)	431 (90%)	41 (8%)	9 (2%)	8 36

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	F	481/484 (99%)	433 (90%)	39 (8%)	9 (2%)	8	36
1	G	482/484 (100%)	424 (88%)	47 (10%)	11 (2%)	6	30
1	H	481/484 (99%)	418 (87%)	53 (11%)	10 (2%)	7	33
1	I	481/484 (99%)	426 (89%)	42 (9%)	13 (3%)	5	26
1	J	479/484 (99%)	416 (87%)	50 (10%)	13 (3%)	5	26
1	K	481/484 (99%)	433 (90%)	42 (9%)	6 (1%)	13	48
1	L	482/484 (100%)	420 (87%)	54 (11%)	8 (2%)	9	39
All	All	5772/5808 (99%)	5137 (89%)	529 (9%)	106 (2%)	8	37

5 of 106 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	450	ASN
1	B	227	ASP
1	B	428	GLY
1	D	450	ASN
1	D	465	THR

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	419/420 (100%)	371 (88%)	48 (12%)	5	24
1	B	419/420 (100%)	366 (87%)	53 (13%)	4	20
1	C	419/420 (100%)	371 (88%)	48 (12%)	5	24
1	D	419/420 (100%)	370 (88%)	49 (12%)	5	22
1	E	419/420 (100%)	374 (89%)	45 (11%)	6	26
1	F	419/420 (100%)	371 (88%)	48 (12%)	5	24
1	G	420/420 (100%)	359 (86%)	61 (14%)	3	15
1	H	419/420 (100%)	362 (86%)	57 (14%)	3	17
1	I	419/420 (100%)	374 (89%)	45 (11%)	6	26

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	J	417/420 (99%)	366 (88%)	51 (12%)	5	21
1	K	419/420 (100%)	367 (88%)	52 (12%)	4	20
1	L	420/420 (100%)	374 (89%)	46 (11%)	6	25
All	All	5028/5040 (100%)	4425 (88%)	603 (12%)	5	22

5 of 603 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	353	ILE
1	G	398	ILE
1	L	11	LYS
1	F	415	GLU
1	G	130	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 126 such sidechains are listed below:

Mol	Chain	Res	Type
1	F	98	GLN
1	G	291	ASN
1	L	97	GLN
1	F	163	ASN
1	G	23	HIS

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 [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(Å ²)	Q<0.9
1	A	483/484 (99%)	-0.57	1 (0%) 95 87	26, 38, 55, 63	6 (1%)
1	B	483/484 (99%)	-0.63	1 (0%) 95 87	23, 36, 51, 68	5 (1%)
1	C	483/484 (99%)	-0.60	1 (0%) 95 87	21, 37, 57, 71	8 (1%)
1	D	483/484 (99%)	-0.58	0 100 100	23, 37, 55, 68	11 (2%)
1	E	483/484 (99%)	-0.54	1 (0%) 95 87	21, 37, 60, 72	8 (1%)
1	F	483/484 (99%)	-0.56	0 100 100	19, 39, 60, 69	14 (2%)
1	G	483/484 (99%)	-0.48	2 (0%) 92 79	30, 46, 60, 72	11 (2%)
1	H	483/484 (99%)	-0.32	2 (0%) 92 79	34, 56, 68, 81	8 (1%)
1	I	483/484 (99%)	-0.38	2 (0%) 92 79	31, 50, 76, 86	22 (4%)
1	J	481/484 (99%)	-0.38	1 (0%) 95 87	30, 55, 78, 87	22 (4%)
1	K	483/484 (99%)	-0.49	1 (0%) 95 87	25, 46, 60, 79	13 (2%)
1	L	483/484 (99%)	-0.30	6 (1%) 79 54	29, 50, 77, 86	12 (2%)
All	All	5794/5808 (99%)	-0.49	18 (0%) 94 84	19, 44, 68, 87	140 (2%)

The worst 5 of 18 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	K	484	GLU	3.4
1	I	462	THR	2.9
1	I	2	SER	2.8
1	H	429	GLN	2.8
1	L	463	VAL	2.8

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.