



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 9, 2026 – 09:00 AM UTC

PDB ID : 4R1O / pdb_00004r1o
Title : Crystal Structure of Thermophilic Geobacillus kaustophilus L-Arabinose isomerase
Authors : Choi, J.M.; Lee, Y.J.; Lee, D.W.; Lee, S.H.
Deposited on : 2014-08-07
Resolution : 2.40 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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

MolProbity	:	4-5-2 with Phenix2.0
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

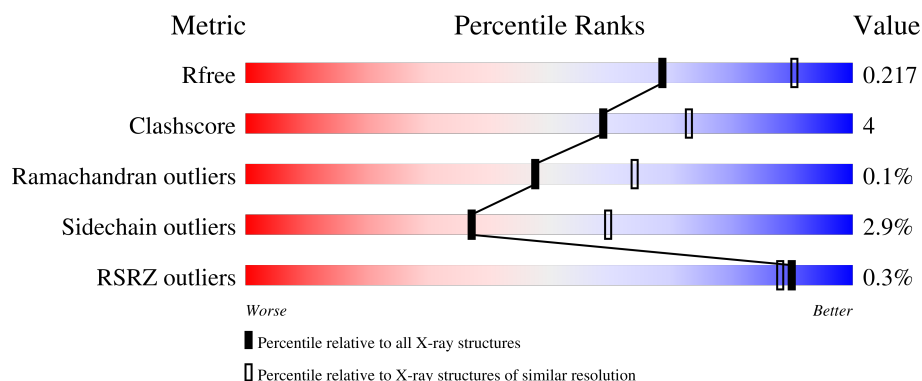
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.40 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}	180053	4912 (2.40-2.40)
Clashscore	190562	5391 (2.40-2.40)
Ramachandran outliers	187476	5320 (2.40-2.40)
Sidechain outliers	187428	5321 (2.40-2.40)
RSRZ outliers	180081	4916 (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 of 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	497	89% 10% .
1	B	497	88% 12%
1	C	497	% 89% 10% .
1	D	497	87% 12%
1	E	497	84% 15% .

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	497	% 84% 14% ..

2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 24843 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 L-arabinose isomerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	496	Total	C	N	O	S	0	0	0
			3953	2514	693	724	22			
1	B	495	Total	C	N	O	S	0	0	0
			3942	2508	689	723	22			
1	C	496	Total	C	N	O	S	0	0	0
			3953	2514	693	724	22			
1	D	496	Total	C	N	O	S	0	0	0
			3953	2514	693	724	22			
1	E	494	Total	C	N	O	S	0	0	0
			3934	2503	688	722	21			
1	F	493	Total	C	N	O	S	0	0	0
			3926	2497	687	721	21			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	SEE REMARK 999	UNP Q5KYP7
B	1	MET	-	SEE REMARK 999	UNP Q5KYP7
C	1	MET	-	SEE REMARK 999	UNP Q5KYP7
D	1	MET	-	SEE REMARK 999	UNP Q5KYP7
E	1	MET	-	SEE REMARK 999	UNP Q5KYP7
F	1	MET	-	SEE REMARK 999	UNP Q5KYP7

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	230	Total	O	0	0
			230	230		
2	B	213	Total	O	0	0
			213	213		
2	C	220	Total	O	0	0
			220	220		

Continued on next page...

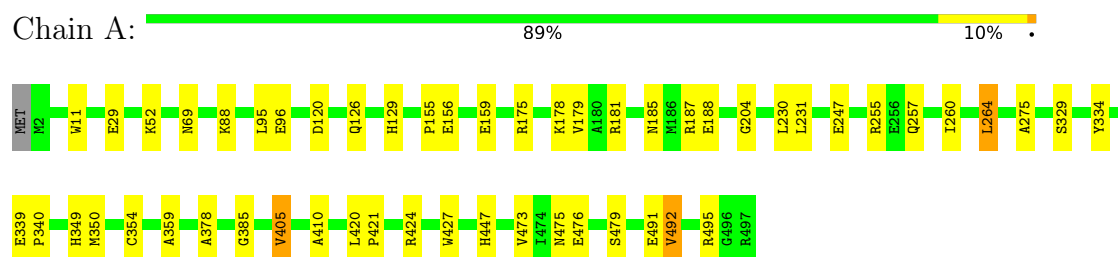
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	D	191	Total 191	O 191	0	0
2	E	160	Total 160	O 160	0	0
2	F	168	Total 168	O 168	0	0

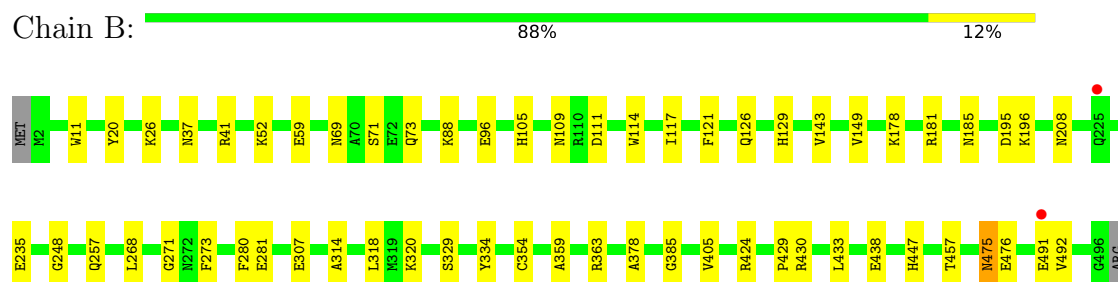
3 Residue-property plots

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

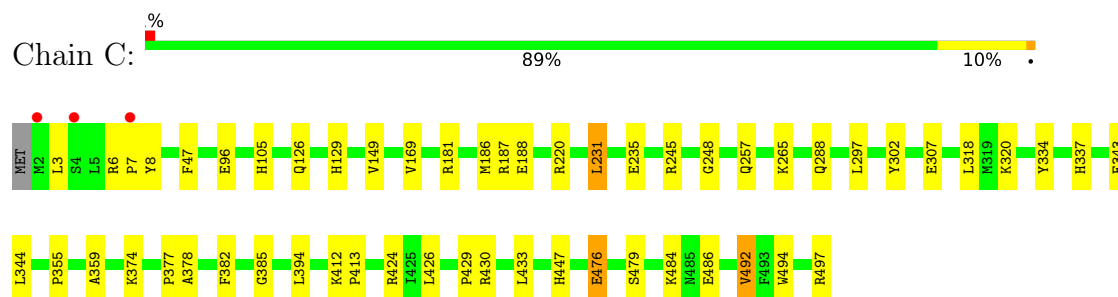
- Molecule 1: L-arabinose isomerase



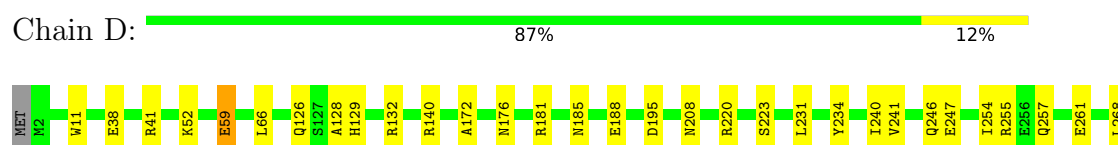
- Molecule 1: L-arabinose isomerase

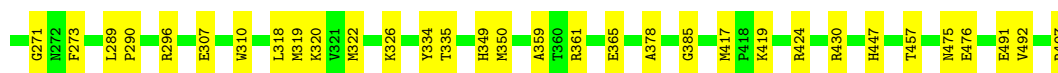


- Molecule 1: L-arabinose isomerase



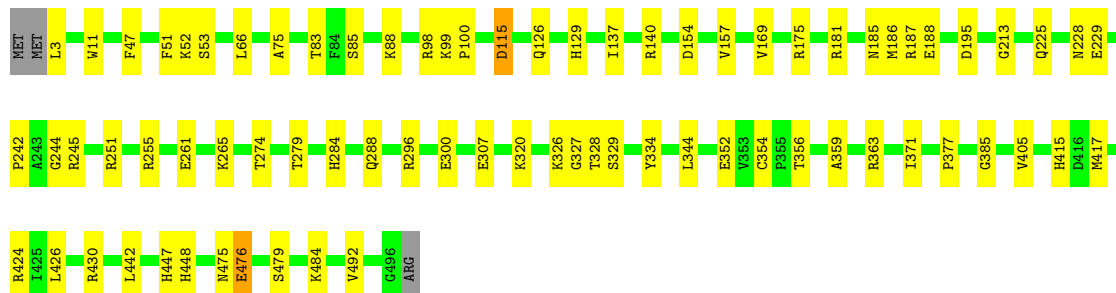
- Molecule 1: L-arabinose isomerase





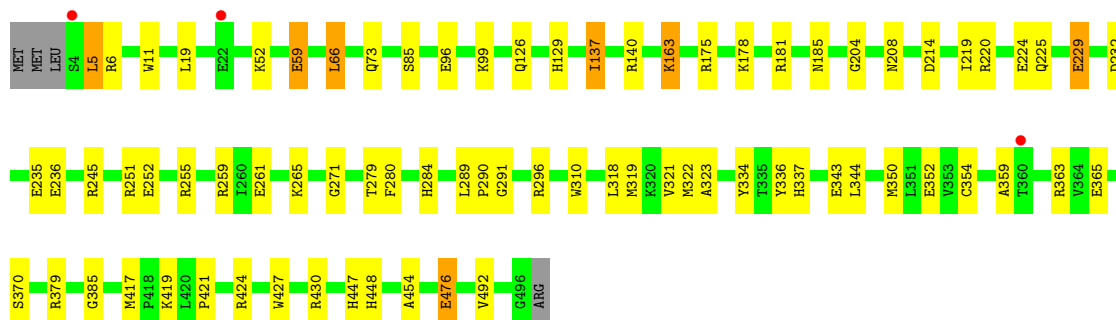
- Molecule 1: L-arabinose isomerase

Chain E: 84% 15%



- Molecule 1: L-arabinose isomerase

Chain F: 84% 14%



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	116.72Å 140.75Å 215.82Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.32 – 2.40 45.32 – 2.40	Depositor EDS
% Data completeness (in resolution range)	94.8 (45.32-2.40) 94.9 (45.32-2.40)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	0.12	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.66 (at 2.39Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, R_{free}	0.164 , 0.216 0.167 , 0.217	Depositor DCC
R_{free} test set	6608 reflections (4.75%)	wwPDB-VP
Wilson B-factor (Å ²)	25.6	Xtriage
Anisotropy	0.185	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.41 , 40.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	24843	wwPDB-VP
Average B, all atoms (Å ²)	17.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.79% 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](#)

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	0.50	0/4052	0.81	0/5486
1	B	0.50	0/4041	0.82	3/5472 (0.1%)
1	C	0.51	0/4052	0.82	2/5486 (0.0%)
1	D	0.49	0/4052	0.81	2/5486 (0.0%)
1	E	0.46	0/4033	0.81	4/5462 (0.1%)
1	F	0.47	0/4025	0.85	9/5451 (0.2%)
All	All	0.49	0/24255	0.82	20/32843 (0.1%)

There are no bond length outliers.

The worst 5 of 20 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	248	GLY	CA-C-N	6.82	126.06	118.97
1	C	248	GLY	C-N-CA	6.82	126.06	118.97
1	B	248	GLY	CA-C-N	6.33	126.53	119.32
1	B	248	GLY	C-N-CA	6.33	126.53	119.32
1	F	85	SER	CA-C-N	6.03	127.37	119.84

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	A	3953	0	3846	33	0
1	B	3942	0	3833	34	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	3953	0	3846	32	0
1	D	3953	0	3846	43	0
1	E	3934	0	3824	49	0
1	F	3926	0	3813	43	0
2	A	230	0	0	3	0
2	B	213	0	0	5	0
2	C	220	0	0	4	0
2	D	191	0	0	5	0
2	E	160	0	0	2	0
2	F	168	0	0	3	0
All	All	24843	0	23008	206	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 4.

The worst 5 of 206 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:F:6:ARG:HH21	1:F:73:GLN:HG3	1.39	0.85
1:F:430:ARG:HD3	1:F:476:GLU:HG3	1.62	0.80
1:C:447:HIS:HB2	1:E:129:HIS:HB2	1.63	0.80
1:B:129:HIS:HB2	1:E:447:HIS:HB2	1.65	0.79
1:D:129:HIS:HB2	1:F:447:HIS:HB2	1.66	0.78

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	494/497 (99%)	480 (97%)	14 (3%)	0	100	100
1	B	493/497 (99%)	475 (96%)	18 (4%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	494/497 (99%)	473 (96%)	20 (4%)	1 (0%)	43	58
1	D	494/497 (99%)	479 (97%)	15 (3%)	0	100	100
1	E	492/497 (99%)	473 (96%)	19 (4%)	0	100	100
1	F	491/497 (99%)	474 (96%)	16 (3%)	1 (0%)	43	58
All	All	2958/2982 (99%)	2854 (96%)	102 (3%)	2 (0%)	48	64

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	7	PRO
1	F	291	GLY

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	414/415 (100%)	403 (97%)	11 (3%)	39	62
1	B	413/415 (100%)	401 (97%)	12 (3%)	37	60
1	C	414/415 (100%)	403 (97%)	11 (3%)	39	62
1	D	414/415 (100%)	401 (97%)	13 (3%)	35	57
1	E	412/415 (99%)	404 (98%)	8 (2%)	50	71
1	F	411/415 (99%)	394 (96%)	17 (4%)	27	46
All	All	2478/2490 (100%)	2406 (97%)	72 (3%)	37	60

5 of 72 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	66	LEU
1	F	492	VAL
1	F	163	LYS
1	F	318	LEU
1	C	231	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
1	F	18	HIS
1	F	202	GLN
1	C	477	HIS
1	D	27	GLN
1	D	125	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 oligosaccharides 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	496/497 (99%)	-0.66	0 100 100	5, 12, 28, 43	0
1	B	495/497 (99%)	-0.60	2 (0%) 88 86	6, 14, 30, 44	0
1	C	496/497 (99%)	-0.66	3 (0%) 85 83	5, 12, 27, 63	0
1	D	496/497 (99%)	-0.58	0 100 100	6, 15, 32, 48	0
1	E	494/497 (99%)	-0.45	0 100 100	8, 18, 40, 63	0
1	F	493/497 (99%)	-0.50	3 (0%) 85 83	7, 17, 38, 57	0
All	All	2970/2982 (99%)	-0.58	8 (0%) 90 88	5, 14, 33, 63	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	4	SER	3.1
1	C	7	PRO	2.8
1	C	2	MET	2.5
1	B	491	GLU	2.3
1	F	4	SER	2.2

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 oligosaccharides 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.