



Full wwPDB X-ray Structure Validation Report ⓘ

May 27, 2020 – 10:40 pm BST

PDB ID : 2ZYF
Title : Crystal structure of homocitrate synthase from *Thermus thermophilus* complexed with magnesium ion and alpha-ketoglutarate
Authors : Okada, T.; Tomita, T.; Kuzuyama, T.; Nishiyama, M.
Deposited on : 2009-01-20
Resolution : 2.15 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
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

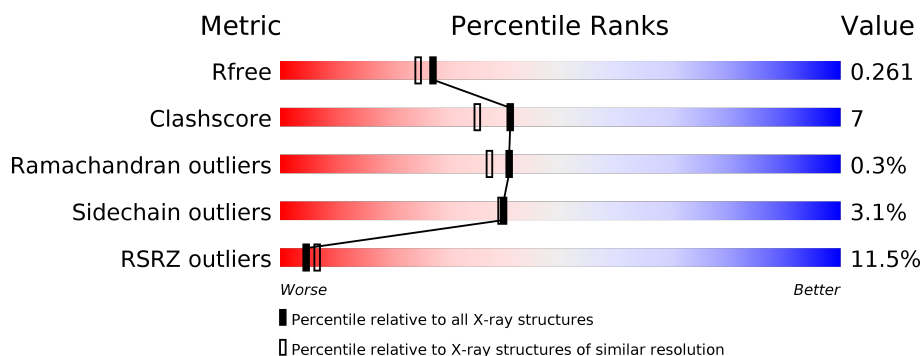
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.15 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	1479 (2.16-2.16)
Clashscore	141614	1585 (2.16-2.16)
Ramachandran outliers	138981	1560 (2.16-2.16)
Sidechain outliers	138945	1559 (2.16-2.16)
RSRZ outliers	127900	1456 (2.16-2.16)

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	382	<div> <div>9%</div> <div>71%</div> <div>10%</div> <div>18%</div> </div>

2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 2655 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.

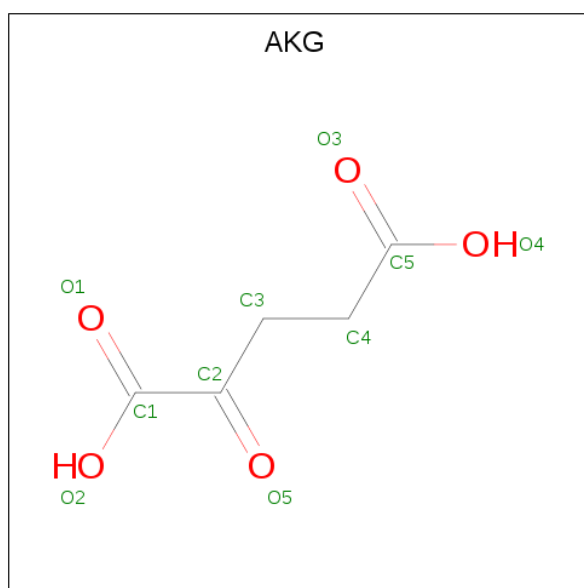
- Molecule 1 is a protein called Homocitrate synthase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	314	2502	1591	435	468	8	0	6	0

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	104	PRO	ALA	SEE REMARK 999	UNP O87198
A	377	HIS	-	EXPRESSION TAG	UNP O87198
A	378	HIS	-	EXPRESSION TAG	UNP O87198
A	379	HIS	-	EXPRESSION TAG	UNP O87198
A	380	HIS	-	EXPRESSION TAG	UNP O87198
A	381	HIS	-	EXPRESSION TAG	UNP O87198
A	382	HIS	-	EXPRESSION TAG	UNP O87198

- Molecule 2 is 2-OXOGLUTARIC ACID (three-letter code: AKG) (formula: C₅H₆O₅).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			10	5	5		

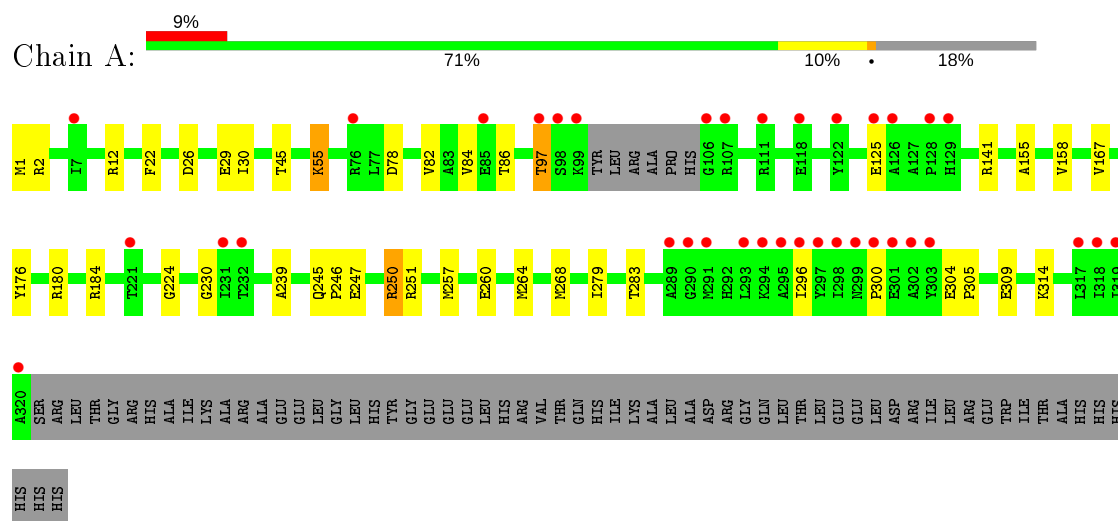
- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Mg	0	0
			1	1		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	142	Total	O	0	0
			142	142		

- Molecule 1: Homocitrate synthase



4 Data and refinement statistics

Property	Value	Source
Space group	P 63 2 2	Depositor
Cell constants a, b, c, α , β , γ	135.47Å 135.47Å 126.61Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	31.51 – 2.15 30.57 – 2.15	Depositor EDS
% Data completeness (in resolution range)	99.9 (31.51-2.15) 99.9 (30.57-2.15)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	0.11	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.51 (at 2.16Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.203 , 0.232 0.233 , 0.261	Depositor DCC
R_{free} test set	1882 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	41.1	Xtriage
Anisotropy	0.100	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 51.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	2655	wwPDB-VP
Average B, all atoms (Å ²)	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.90% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: MG, AKG

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	0.69	0/2566	0.69	0/3473

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	2502	0	2534	36	0
2	A	10	0	4	0	0
3	A	1	0	0	0	0
4	A	142	0	0	3	0
All	All	2655	0	2538	36	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 7.

All (36) 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:26:ASP:HB3	1:A:268:MET:HE1	1.43	0.99
1:A:26:ASP:HB3	1:A:268:MET:CE	2.00	0.90
1:A:30:ILE:HA	1:A:264:MET:HE1	1.54	0.88
1:A:176:TYR:CE2	1:A:180:ARG:HD2	2.11	0.84
1:A:30:ILE:HA	1:A:264:MET:CE	2.18	0.73
1:A:246:PRO:O	1:A:250:ARG:HG2	1.88	0.73
1:A:251[A]:ARG:NH2	4:A:470:HOH:O	2.10	0.70
1:A:22:PHE:CD2	1:A:268:MET:HE2	2.31	0.66
1:A:55:LYS:HE3	4:A:481:HOH:O	1.97	0.64
1:A:247:GLU:OE1	1:A:250:ARG:NH1	2.33	0.62
1:A:30:ILE:CG1	1:A:264:MET:HE1	2.29	0.62
1:A:30:ILE:CA	1:A:264:MET:HE1	2.32	0.59
1:A:30:ILE:CG1	1:A:264:MET:CE	2.82	0.57
1:A:22:PHE:CE2	1:A:268:MET:HE2	2.39	0.56
1:A:30:ILE:HG12	1:A:264:MET:HE1	1.85	0.56
1:A:30:ILE:HG13	1:A:264:MET:CE	2.38	0.53
1:A:1:MET:HE2	4:A:515:HOH:O	2.08	0.53
1:A:239:ALA:HB2	1:A:279:ILE:HD11	1.90	0.52
1:A:30:ILE:CA	1:A:264:MET:CE	2.90	0.49
1:A:26:ASP:CB	1:A:268:MET:CE	2.85	0.48
1:A:82:VAL:O	1:A:86:THR:HG23	2.13	0.48
1:A:224:GLY:O	1:A:230:GLY:HA3	2.13	0.48
1:A:26:ASP:HB3	1:A:268:MET:HE3	1.91	0.47
1:A:12:ARG:HG2	1:A:45:THR:HA	1.97	0.45
1:A:247:GLU:OE2	1:A:251[B]:ARG:HD2	2.17	0.45
1:A:257:MET:O	1:A:260:GLU:HG2	2.16	0.45
1:A:141:ARG:NH2	1:A:167:VAL:HG21	2.33	0.43
1:A:29:GLU:HG2	1:A:264:MET:HE3	2.01	0.43
1:A:155:ALA:HA	1:A:158:VAL:HG22	2.02	0.42
1:A:167:VAL:HG12	1:A:167:VAL:O	2.21	0.41
1:A:2:ARG:HB2	1:A:245:GLN:NE2	2.36	0.41
1:A:30:ILE:HG13	1:A:264:MET:HE2	2.03	0.41
1:A:296:ILE:HG23	1:A:300:PRO:HA	2.03	0.41
1:A:304:GLU:HA	1:A:305:PRO:HD2	1.91	0.41
1:A:309:GLU:H	1:A:309:GLU:CD	2.24	0.40
1:A:180:ARG:O	1:A:184[B]:ARG:HG3	2.20	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	316/382 (83%)	301 (95%)	14 (4%)	1 (0%)	41	37

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	97	THR

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	264/315 (84%)	256 (97%)	8 (3%)	41	40

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	55	LYS
1	A	78	ASP
1	A	84	VAL
1	A	97	THR
1	A	125	GLU
1	A	250	ARG
1	A	283	THR
1	A	314	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	145	GLN

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](#)

Of 2 ligands modelled in this entry, 1 is monoatomic - leaving 1 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# $ Z > 2$	Counts	RMSZ	# $ Z > 2$
2	AKG	A	383	3	3,9,9	0.42	0	4,11,11	2.38	1 (25%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	AKG	A	383	3	-	0/3/9/9	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
2	A	383	AKG	C4-C3-C2	-4.32	103.82	113.14

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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	314/382 (82%)	0.59	36 (11%) 4 6	14, 24, 49, 61	0

All (36) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	297	TYR	13.4
1	A	298	ILE	8.8
1	A	319	ILE	8.7
1	A	300	PRO	8.2
1	A	296	ILE	8.0
1	A	291	MET	7.5
1	A	293	LEU	6.9
1	A	98	SER	6.9
1	A	320	ALA	6.8
1	A	299	ASN	6.6
1	A	295	ALA	6.5
1	A	294	LYS	5.2
1	A	301	GLU	5.2
1	A	289	ALA	4.9
1	A	97	THR	4.9
1	A	106	GLY	4.5
1	A	318	ILE	4.4
1	A	290	GLY	4.0
1	A	317	LEU	3.8
1	A	111	ARG	3.6
1	A	122	TYR	3.5
1	A	129	HIS	3.4
1	A	125	GLU	3.3
1	A	99	LYS	3.3
1	A	107	ARG	3.2
1	A	231	ILE	3.2
1	A	302	ALA	3.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	128	PRO	3.1
1	A	7	ILE	3.0
1	A	85	GLU	2.7
1	A	232	THR	2.6
1	A	303	TYR	2.5
1	A	76	ARG	2.3
1	A	126	ALA	2.2
1	A	221	THR	2.1
1	A	118	GLU	2.1

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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
2	AKG	A	383	10/10	0.88	0.30	53,54,58,58	0
3	MG	A	384	1/1	0.95	0.31	37,37,37,37	0

6.5 Other polymers [i](#)

There are no such residues in this entry.