Full wwPDB X-ray Structure Validation Report (i)

May 21, 2020 – 06:05 am BST

PDB ID	:	5M4X
Title	:	Mutant glyceraldehyde dehydrogenase (F34M+Y399C+S405N) from Thermoplasma acidophilum
Authors	:	Iermak, I.; Mesters, J.R.; Kuta Smatanova, I.
Deposited on	:	2016-10-19
Resolution	:	3.56 Å(reported)

The following versions of software and data (see references (i)) 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 (i)

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

The reported resolution of this entry is 3.56 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.
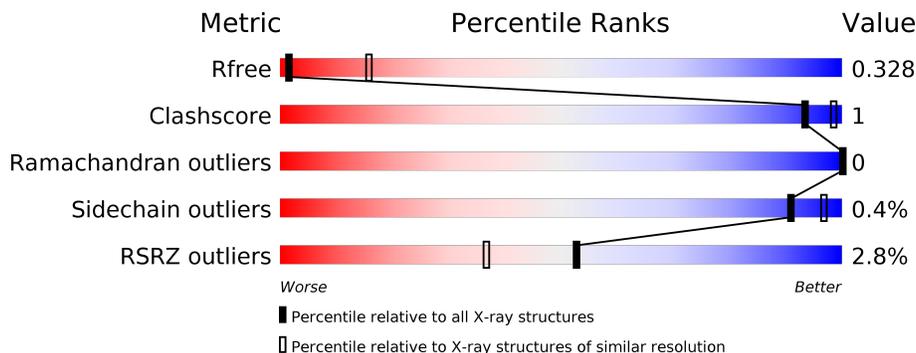
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1020 (3.62-3.50)
Clashscore	141614	1100 (3.62-3.50)
Ramachandran outliers	138981	1065 (3.62-3.50)
Sidechain outliers	138945	1066 (3.62-3.50)
RSRZ outliers	127900	1009 (3.64-3.48)

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	508	% 86% • 9%
1	B	508	89% • 8%
1	C	508	5% 87% • 9%
1	D	508	4% 88% 5% 7%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 13897 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 D-glyceraldehyde dehydrogenase (NADP(+)).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	461	Total	C	N	O	S	0	0	0
			3447	2199	570	665	13			
1	B	468	Total	C	N	O	S	0	0	0
			3501	2231	579	679	12			
1	C	461	Total	C	N	O	S	0	0	0
			3444	2195	572	664	13			
1	D	472	Total	C	N	O	S	0	0	0
			3505	2234	578	680	13			

There are 72 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	34	MET	PHE	engineered mutation	UNP Q9HK01
A	399	CYS	TYR	engineered mutation	UNP Q9HK01
A	405	ASN	SER	engineered mutation	UNP Q9HK01
A	494	SER	-	expression tag	UNP Q9HK01
A	495	GLY	-	expression tag	UNP Q9HK01
A	496	ARG	-	expression tag	UNP Q9HK01
A	497	PRO	-	expression tag	UNP Q9HK01
A	498	VAL	-	expression tag	UNP Q9HK01
A	499	LEU	-	expression tag	UNP Q9HK01
A	500	GLY	-	expression tag	UNP Q9HK01
A	501	SER	-	expression tag	UNP Q9HK01
A	502	SER	-	expression tag	UNP Q9HK01
A	503	HIS	-	expression tag	UNP Q9HK01
A	504	HIS	-	expression tag	UNP Q9HK01
A	505	HIS	-	expression tag	UNP Q9HK01
A	506	HIS	-	expression tag	UNP Q9HK01
A	507	HIS	-	expression tag	UNP Q9HK01
A	508	HIS	-	expression tag	UNP Q9HK01
B	34	MET	PHE	engineered mutation	UNP Q9HK01
B	399	CYS	TYR	engineered mutation	UNP Q9HK01
B	405	ASN	SER	engineered mutation	UNP Q9HK01

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	494	SER	-	expression tag	UNP Q9HK01
B	495	GLY	-	expression tag	UNP Q9HK01
B	496	ARG	-	expression tag	UNP Q9HK01
B	497	PRO	-	expression tag	UNP Q9HK01
B	498	VAL	-	expression tag	UNP Q9HK01
B	499	LEU	-	expression tag	UNP Q9HK01
B	500	GLY	-	expression tag	UNP Q9HK01
B	501	SER	-	expression tag	UNP Q9HK01
B	502	SER	-	expression tag	UNP Q9HK01
B	503	HIS	-	expression tag	UNP Q9HK01
B	504	HIS	-	expression tag	UNP Q9HK01
B	505	HIS	-	expression tag	UNP Q9HK01
B	506	HIS	-	expression tag	UNP Q9HK01
B	507	HIS	-	expression tag	UNP Q9HK01
B	508	HIS	-	expression tag	UNP Q9HK01
C	34	MET	PHE	engineered mutation	UNP Q9HK01
C	399	CYS	TYR	engineered mutation	UNP Q9HK01
C	405	ASN	SER	engineered mutation	UNP Q9HK01
C	494	SER	-	expression tag	UNP Q9HK01
C	495	GLY	-	expression tag	UNP Q9HK01
C	496	ARG	-	expression tag	UNP Q9HK01
C	497	PRO	-	expression tag	UNP Q9HK01
C	498	VAL	-	expression tag	UNP Q9HK01
C	499	LEU	-	expression tag	UNP Q9HK01
C	500	GLY	-	expression tag	UNP Q9HK01
C	501	SER	-	expression tag	UNP Q9HK01
C	502	SER	-	expression tag	UNP Q9HK01
C	503	HIS	-	expression tag	UNP Q9HK01
C	504	HIS	-	expression tag	UNP Q9HK01
C	505	HIS	-	expression tag	UNP Q9HK01
C	506	HIS	-	expression tag	UNP Q9HK01
C	507	HIS	-	expression tag	UNP Q9HK01
C	508	HIS	-	expression tag	UNP Q9HK01
D	34	MET	PHE	engineered mutation	UNP Q9HK01
D	399	CYS	TYR	engineered mutation	UNP Q9HK01
D	405	ASN	SER	engineered mutation	UNP Q9HK01
D	494	SER	-	expression tag	UNP Q9HK01
D	495	GLY	-	expression tag	UNP Q9HK01
D	496	ARG	-	expression tag	UNP Q9HK01
D	497	PRO	-	expression tag	UNP Q9HK01
D	498	VAL	-	expression tag	UNP Q9HK01
D	499	LEU	-	expression tag	UNP Q9HK01

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	500	GLY	-	expression tag	UNP Q9HK01
D	501	SER	-	expression tag	UNP Q9HK01
D	502	SER	-	expression tag	UNP Q9HK01
D	503	HIS	-	expression tag	UNP Q9HK01
D	504	HIS	-	expression tag	UNP Q9HK01
D	505	HIS	-	expression tag	UNP Q9HK01
D	506	HIS	-	expression tag	UNP Q9HK01
D	507	HIS	-	expression tag	UNP Q9HK01
D	508	HIS	-	expression tag	UNP Q9HK01

3 Residue-property plots (i)

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: D-glyceraldehyde dehydrogenase (NADP(+))

- Molecule 1: D-glyceraldehyde dehydrogenase (NADP(+))

- Molecule 1: D-glyceraldehyde dehydrogenase (NADP(+))

- Molecule 1: D-glyceraldehyde dehydrogenase (NADP(+))

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α, β, γ	120.29Å 120.29Å 344.88Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.73 – 3.56 48.73 – 3.56	Depositor EDS
% Data completeness (in resolution range)	99.5 (48.73-3.56) 99.6 (48.73-3.56)	Depositor EDS
R_{merge}	0.20	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	1.84 (at 3.57Å)	Xtriage
Refinement program	REFMAC 5.8.0155	Depositor
R, R_{free}	0.312 , 0.319 0.312 , 0.328	Depositor DCC
R_{free} test set	1093 reflections (3.50%)	wwPDB-VP
Wilson B-factor (Å^2)	117.6	Xtriage
Anisotropy	0.230	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 2.3	EDS
L-test for twinning[2]	$<|L|> = 0.47$, $<L^2> = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.88	EDS
Total number of atoms	13897	wwPDB-VP
Average B, all atoms (Å^2)	55.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.48% 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	0.55	0/3516	0.81	4/4778 (0.1%)
1	B	0.54	0/3571	0.79	2/4856 (0.0%)
1	C	0.51	0/3514	0.78	2/4776 (0.0%)
1	D	0.54	0/3576	0.79	2/4866 (0.0%)
All	All	0.53	0/14177	0.79	10/19276 (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	A	0	2
1	B	0	1
1	C	0	1
All	All	0	4

There are no bond length outliers.

All (10) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	241	MET	CG-SD-CE	6.94	111.30	100.20
1	D	241	MET	CG-SD-CE	5.69	109.30	100.20
1	C	212	ILE	N-CA-CB	5.61	123.69	110.80
1	A	212	ILE	N-CA-CB	5.59	123.67	110.80
1	D	212	ILE	N-CA-CB	5.59	123.67	110.80
1	B	212	ILE	N-CA-CB	5.52	123.50	110.80
1	C	241	MET	CG-SD-CE	5.51	109.02	100.20
1	B	358	ARG	NE-CZ-NH1	-5.44	117.58	120.30
1	A	135	TYR	CB-CG-CD2	-5.22	117.87	121.00
1	A	135	TYR	CB-CG-CD1	5.10	124.06	121.00

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	492	PHE	Peptide
1	A	493	GLN	Peptide
1	B	492	PHE	Peptide
1	C	492	PHE	Peptide

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	3447	0	3276	10	1
1	B	3501	0	3329	7	1
1	C	3444	0	3259	10	0
1	D	3505	0	3305	14	0
All	All	13897	0	13169	37	1

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 1.

All (37) 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:434:TYR:CE1	1:C:439:GLY:HA2	2.11	0.85
1:B:409:LEU:HD12	1:B:409:LEU:O	1.81	0.79
1:D:434:TYR:CE1	1:D:439:GLY:HA2	2.21	0.76
1:D:249:GLY:HA2	1:D:281:CYS:HB3	1.79	0.64
1:C:411:SER:N	1:C:432:GLU:O	2.32	0.62
1:D:282:ILE:HG22	1:D:410:ALA:HB2	1.83	0.61
1:A:409:LEU:C	1:A:409:LEU:HD12	2.20	0.60
1:B:409:LEU:HD12	1:B:409:LEU:C	2.23	0.57
1:D:175:SER:O	1:D:178:PRO:HD3	2.06	0.55
1:A:280:SER:OG	1:A:283:ALA:HB2	2.07	0.54
1:D:6:TYR:HB2	1:D:11:TRP:CZ3	2.44	0.52
1:B:247:GLU:O	1:B:248:LEU:HD23	2.10	0.51
1:D:299:SER:O	1:D:302:VAL:N	2.45	0.50
1:D:299:SER:O	1:D:300:ARG:C	2.51	0.49

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:131:LYS:HE3	1:C:131:LYS:CE	2.43	0.48
1:A:321:LEU:HD12	1:A:327:LEU:HA	1.94	0.48
1:D:321:LEU:HD12	1:D:327:LEU:HA	1.97	0.47
1:D:6:TYR:CZ	1:D:9:GLY:HA2	2.50	0.47
1:B:321:LEU:HD12	1:B:327:LEU:HA	1.97	0.46
1:A:338:LYS:CG	1:A:344:ILE:HD11	2.46	0.45
1:B:338:LYS:CG	1:B:344:ILE:HD11	2.47	0.45
1:C:321:LEU:HD12	1:C:327:LEU:HA	1.97	0.45
1:D:338:LYS:CG	1:D:344:ILE:HD11	2.47	0.44
1:A:65:SER:HB3	1:A:115:ALA:HB2	2.00	0.44
1:C:235:GLN:HG3	1:D:238:SER:CB	2.47	0.44
1:D:338:LYS:C	1:D:340:SER:H	2.22	0.43
1:D:65:SER:HB3	1:D:115:ALA:HB2	1.99	0.43
1:A:338:LYS:C	1:A:340:SER:H	2.22	0.43
1:B:65:SER:HB3	1:B:115:ALA:HB2	2.00	0.42
1:C:249:GLY:HA2	1:C:281:CYS:HB3	2.02	0.42
1:A:145:LEU:HD23	1:A:145:LEU:HA	1.78	0.42
1:B:338:LYS:C	1:B:340:SER:H	2.22	0.41
1:C:65:SER:HB3	1:C:115:ALA:HB2	2.00	0.41
1:A:148:TRP:CZ3	1:A:322:ILE:CB	3.03	0.41
1:C:163:LEU:HA	1:C:163:LEU:HD23	1.98	0.41
1:A:418:PRO:HB3	1:C:418:PRO:HB3	2.03	0.40
1:C:271:TRP:HB2	1:D:484:ILE:HG21	2.03	0.40

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:17:GLY:O	1:B:358:ARG:NH2[6_454]	1.74	0.46

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	455/508 (90%)	444 (98%)	11 (2%)	0	100	100
1	B	464/508 (91%)	453 (98%)	11 (2%)	0	100	100
1	C	455/508 (90%)	446 (98%)	9 (2%)	0	100	100
1	D	468/508 (92%)	458 (98%)	10 (2%)	0	100	100
All	All	1842/2032 (91%)	1801 (98%)	41 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	339/419 (81%)	338 (100%)	1 (0%)	92	97
1	B	346/419 (83%)	345 (100%)	1 (0%)	92	97
1	C	337/419 (80%)	335 (99%)	2 (1%)	86	94
1	D	342/419 (82%)	340 (99%)	2 (1%)	86	94
All	All	1364/1676 (81%)	1358 (100%)	6 (0%)	91	97

All (6) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	383	PHE
1	B	383	PHE
1	C	281	CYS
1	C	383	PHE
1	D	175	SER
1	D	383	PHE

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 (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	461/508 (90%)	-0.06	4 (0%)	84	72	32, 39, 49, 66	0
1	B	468/508 (92%)	-0.05	2 (0%)	92	86	36, 43, 53, 70	0
1	C	461/508 (90%)	0.38	27 (5%)	22	13	66, 85, 96, 113	0
1	D	472/508 (92%)	0.12	19 (4%)	38	25	42, 50, 60, 77	0
All	All	1862/2032 (91%)	0.10	52 (2%)	53	37	32, 48, 89, 113	0

All (52) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	439	GLY	4.8
1	C	37	ALA	4.7
1	C	354	SER	4.0
1	A	494	SER	4.0
1	D	16	SER	4.0
1	D	354	SER	3.9
1	D	494	SER	3.8
1	C	494	SER	3.6
1	C	4	LYS	3.6
1	C	2	ASP	3.4
1	C	268	THR	3.3
1	D	425	SER	3.2
1	D	486	THR	3.1
1	B	116	ARG	3.0
1	C	256	VAL	2.8
1	A	129	HIS	2.7
1	C	357	TYR	2.7
1	D	358	ARG	2.6
1	D	26	VAL	2.6
1	C	202	THR	2.6
1	D	357	TYR	2.6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	D	116	ARG	2.5
1	C	266	LEU	2.5
1	D	356	PRO	2.5
1	C	200	PHE	2.5
1	D	353	LEU	2.4
1	C	205	GLY	2.4
1	C	342	ALA	2.4
1	A	362	PHE	2.4
1	D	362	PHE	2.4
1	C	403	ASN	2.4
1	C	355	GLY	2.3
1	C	44	ARG	2.3
1	D	361	TYR	2.3
1	D	218	VAL	2.3
1	D	118	LEU	2.3
1	D	439	GLY	2.3
1	C	156	ALA	2.3
1	B	425	SER	2.2
1	C	188	PHE	2.2
1	A	59	LEU	2.2
1	C	310	LEU	2.2
1	D	112	ALA	2.2
1	C	255	MET	2.1
1	D	19	THR	2.1
1	C	414	PHE	2.1
1	C	274	TYR	2.1
1	C	239	ALA	2.1
1	C	140	GLY	2.1
1	D	18	LYS	2.1
1	C	385	PRO	2.0
1	C	387	ILE	2.0

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.