



wwPDB X-ray Structure Validation Summary Report

Jun 17, 2024 – 07:21 pm BST

PDB ID : 8PA1
Title : Crystal structure of alcohol dehydrogenase/ketoreductase variant from
Thermus thermophilus apo form
Authors : Rozeboom, H.J.; Janssen, D.B.
Deposited on : 2023-06-07
Resolution : 2.60 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 : 4.02b-467
Mogul : 1.8.4, CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.37.1
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) : 2.37.1

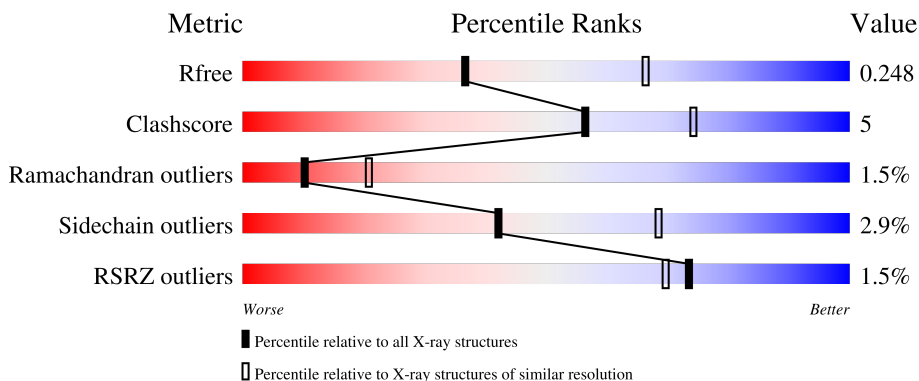
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.60 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	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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	264	 84% 9% • 6%
1	B	264	 2% 81% 12% • 6%
1	C	264	 2% 82% 11% 6%
1	D	264	 2% 81% 12% 6%

2 Entry composition i

There are 4 unique types of molecules in this entry. The entry contains 7660 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 Oxidoreductase, short-chain dehydrogenase/reductase family.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	249	1856	1175	334	342	5	0	0	0
1	B	249	1856	1175	334	342	5	0	0	0
1	C	249	1856	1175	334	342	5	0	0	0
1	D	249	1856	1175	334	342	5	0	0	0

There are 68 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-7	MET	-	initiating methionine	UNP Q5SLC4
A	-6	GLY	-	expression tag	UNP Q5SLC4
A	-5	HIS	-	expression tag	UNP Q5SLC4
A	-4	HIS	-	expression tag	UNP Q5SLC4
A	-3	HIS	-	expression tag	UNP Q5SLC4
A	-2	HIS	-	expression tag	UNP Q5SLC4
A	-1	HIS	-	expression tag	UNP Q5SLC4
A	0	HIS	-	expression tag	UNP Q5SLC4
A	126	PRO	VAL	engineered mutation	UNP Q5SLC4
A	128	PRO	GLY	engineered mutation	UNP Q5SLC4
A	137	TYR	VAL	engineered mutation	UNP Q5SLC4
A	138	LEU	GLN	engineered mutation	UNP Q5SLC4
A	181	PHE	ALA	engineered mutation	UNP Q5SLC4
A	187	LEU	VAL	engineered mutation	UNP Q5SLC4
A	190	LEU	ALA	engineered mutation	UNP Q5SLC4
A	204	MET	TRP	conflict	UNP Q5SLC4
A	245	LEU	MET	engineered mutation	UNP Q5SLC4
B	-7	MET	-	initiating methionine	UNP Q5SLC4
B	-6	GLY	-	expression tag	UNP Q5SLC4
B	-5	HIS	-	expression tag	UNP Q5SLC4
B	-4	HIS	-	expression tag	UNP Q5SLC4

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-3	HIS	-	expression tag	UNP Q5SLC4
B	-2	HIS	-	expression tag	UNP Q5SLC4
B	-1	HIS	-	expression tag	UNP Q5SLC4
B	0	HIS	-	expression tag	UNP Q5SLC4
B	126	PRO	VAL	engineered mutation	UNP Q5SLC4
B	128	PRO	GLY	engineered mutation	UNP Q5SLC4
B	137	TYR	VAL	engineered mutation	UNP Q5SLC4
B	138	LEU	GLN	engineered mutation	UNP Q5SLC4
B	181	PHE	ALA	engineered mutation	UNP Q5SLC4
B	187	LEU	VAL	engineered mutation	UNP Q5SLC4
B	190	LEU	ALA	engineered mutation	UNP Q5SLC4
B	204	MET	TRP	conflict	UNP Q5SLC4
B	245	LEU	MET	engineered mutation	UNP Q5SLC4
C	-7	MET	-	initiating methionine	UNP Q5SLC4
C	-6	GLY	-	expression tag	UNP Q5SLC4
C	-5	HIS	-	expression tag	UNP Q5SLC4
C	-4	HIS	-	expression tag	UNP Q5SLC4
C	-3	HIS	-	expression tag	UNP Q5SLC4
C	-2	HIS	-	expression tag	UNP Q5SLC4
C	-1	HIS	-	expression tag	UNP Q5SLC4
C	0	HIS	-	expression tag	UNP Q5SLC4
C	126	PRO	VAL	engineered mutation	UNP Q5SLC4
C	128	PRO	GLY	engineered mutation	UNP Q5SLC4
C	137	TYR	VAL	engineered mutation	UNP Q5SLC4
C	138	LEU	GLN	engineered mutation	UNP Q5SLC4
C	181	PHE	ALA	engineered mutation	UNP Q5SLC4
C	187	LEU	VAL	engineered mutation	UNP Q5SLC4
C	190	LEU	ALA	engineered mutation	UNP Q5SLC4
C	204	MET	TRP	conflict	UNP Q5SLC4
C	245	LEU	MET	engineered mutation	UNP Q5SLC4
D	-7	MET	-	initiating methionine	UNP Q5SLC4
D	-6	GLY	-	expression tag	UNP Q5SLC4
D	-5	HIS	-	expression tag	UNP Q5SLC4
D	-4	HIS	-	expression tag	UNP Q5SLC4
D	-3	HIS	-	expression tag	UNP Q5SLC4
D	-2	HIS	-	expression tag	UNP Q5SLC4
D	-1	HIS	-	expression tag	UNP Q5SLC4
D	0	HIS	-	expression tag	UNP Q5SLC4
D	126	PRO	VAL	engineered mutation	UNP Q5SLC4
D	128	PRO	GLY	engineered mutation	UNP Q5SLC4
D	137	TYR	VAL	engineered mutation	UNP Q5SLC4
D	138	LEU	GLN	engineered mutation	UNP Q5SLC4

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	181	PHE	ALA	engineered mutation	UNP Q5SLC4
D	187	LEU	VAL	engineered mutation	UNP Q5SLC4
D	190	LEU	ALA	engineered mutation	UNP Q5SLC4
D	204	MET	TRP	conflict	UNP Q5SLC4
D	245	LEU	MET	engineered mutation	UNP Q5SLC4

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	S		
2	A	1	5	4	1	0	0
2	B	1	5	4	1	0	0
2	C	1	5	4	1	0	0
2	D	1	5	4	1	0	0

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 6 3 3	0	0
3	B	1	Total C O 6 3 3	0	0
3	C	1	Total C O 6 3 3	0	0
3	D	1	Total C O 6 3 3	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	45	Total O 45 45	0	0
4	B	54	Total O 54 54	0	0
4	C	46	Total O 46 46	0	0
4	D	47	Total O 47 47	0	0

A168	P169	E189	L190	E221	K230	I238	S248	F249	MET	MET	ALA	GLY	ARG	PRO	VAL
------	------	------	------	------	------	------	------	------	-----	-----	-----	-----	-----	-----	-----

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	71.54Å 95.45Å 79.02Å 90.00° 108.06° 90.00°	Depositor
Resolution (Å)	47.80 – 2.60 47.73 – 2.60	Depositor EDS
% Data completeness (in resolution range)	99.7 (47.80-2.60) 99.7 (47.73-2.60)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.60 (at 2.61Å)	Xtrriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.177 , 0.241 0.185 , 0.248	Depositor DCC
R_{free} test set	1540 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å ²)	41.6	Xtrriage
Anisotropy	0.681	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 46.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	7660	wwPDB-VP
Average B, all atoms (Å ²)	50.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.08% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: SO4, GOL

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	0.71	0/1884	0.84	0/2554
1	B	0.74	0/1884	0.84	0/2554
1	C	0.72	0/1884	0.84	0/2554
1	D	0.72	0/1884	0.86	0/2554
All	All	0.72	0/7536	0.85	0/10216

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 [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	1856	0	1906	16	0
1	B	1856	0	1906	26	0
1	C	1856	0	1906	22	0
1	D	1856	0	1906	20	0
2	A	5	0	0	0	0
2	B	5	0	0	0	0
2	C	5	0	0	1	0
2	D	5	0	0	0	0
3	A	6	0	8	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	6	0	8	0	0
3	C	6	0	8	0	0
3	D	6	0	8	1	0
4	A	45	0	0	0	0
4	B	54	0	0	3	0
4	C	46	0	0	0	0
4	D	47	0	0	0	0
All	All	7660	0	7656	76	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 76 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:125:LYS:HB2	1:B:126:PRO:HD3	1.49	0.92
1:A:125:LYS:HB2	1:A:126:PRO:HD3	1.55	0.88
1:B:248:SER:HB2	1:D:165:LEU:HD11	1.69	0.74
1:A:248:SER:HB2	1:C:165:LEU:HD11	1.70	0.73
1:C:16:ARG:HD2	1:C:17:GLY:N	2.11	0.66

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	247/264 (94%)	236 (96%)	7 (3%)	4 (2%)	9 19
1	B	247/264 (94%)	236 (96%)	7 (3%)	4 (2%)	9 19
1	C	247/264 (94%)	236 (96%)	8 (3%)	3 (1%)	13 27
1	D	247/264 (94%)	235 (95%)	8 (3%)	4 (2%)	9 19

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	988/1056 (94%)	943 (95%)	30 (3%)	15 (2%)	10 21

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	125	LYS
1	B	125	LYS
1	C	125	LYS
1	D	125	LYS
1	A	127	GLY

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	183/195 (94%)	178 (97%)	5 (3%)	44 71
1	B	183/195 (94%)	178 (97%)	5 (3%)	44 71
1	C	183/195 (94%)	179 (98%)	4 (2%)	52 76
1	D	183/195 (94%)	176 (96%)	7 (4%)	33 59
All	All	732/780 (94%)	711 (97%)	21 (3%)	42 68

5 of 21 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	16	ARG
1	D	43	LYS
1	D	190	LEU
1	D	125	LYS
1	D	41	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	150	ASN
1	B	24	GLN
1	C	150	ASN
1	D	24	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 monosaccharides in this entry.

5.6 Ligand geometry [i](#)

8 ligands are modelled in this entry.

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	$\# Z > 2$
3	GOL	B	302	-	5,5,5	0.25	0	5,5,5	0.60	0
2	SO4	B	301	-	4,4,4	0.31	0	6,6,6	0.18	0
3	GOL	C	301	-	5,5,5	0.18	0	5,5,5	0.34	0
2	SO4	C	302	-	4,4,4	0.27	0	6,6,6	0.13	0
3	GOL	A	302	-	5,5,5	0.10	0	5,5,5	0.31	0
3	GOL	D	302	-	5,5,5	0.27	0	5,5,5	0.74	0
2	SO4	A	301	-	4,4,4	0.26	0	6,6,6	0.18	0
2	SO4	D	301	-	4,4,4	0.21	0	6,6,6	0.20	0

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
3	GOL	A	302	-	-	4/4/4/4	-
3	GOL	D	302	-	-	3/4/4/4	-
3	GOL	B	302	-	-	2/4/4/4	-
3	GOL	C	301	-	-	3/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 12 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	302	GOL	C1-C2-C3-O3
3	B	302	GOL	C1-C2-C3-O3
3	A	302	GOL	O1-C1-C2-C3
3	D	302	GOL	O1-C1-C2-C3
3	A	302	GOL	O2-C2-C3-O3

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	C	302	SO4	1	0
3	D	302	GOL	1	0

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	249/264 (94%)	-0.12	0 100 100	29, 47, 74, 125	0
1	B	249/264 (94%)	-0.14	4 (1%) 72 68	27, 45, 76, 102	0
1	C	249/264 (94%)	0.04	5 (2%) 65 60	29, 48, 80, 140	0
1	D	249/264 (94%)	-0.07	6 (2%) 59 53	29, 46, 74, 118	0
All	All	996/1056 (94%)	-0.07	15 (1%) 73 70	27, 47, 77, 140	0

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	125	LYS	3.6
1	D	73	TYR	3.2
1	D	2	GLY	3.2
1	B	1	MET	3.2
1	B	140	LEU	3.1

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 monosaccharides 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	SO4	A	301	5/5	0.67	0.21	88,92,116,129	0
2	SO4	C	302	5/5	0.73	0.26	83,108,128,133	0
2	SO4	B	301	5/5	0.82	0.21	75,85,118,130	0
2	SO4	D	301	5/5	0.83	0.20	68,79,104,108	0
3	GOL	C	301	6/6	0.86	0.43	56,65,73,73	0
3	GOL	A	302	6/6	0.91	0.40	57,62,64,71	0
3	GOL	D	302	6/6	0.94	0.21	46,47,51,55	0
3	GOL	B	302	6/6	0.95	0.15	48,55,56,56	0

6.5 Other polymers [\(i\)](#)

There are no such residues in this entry.