



wwPDB X-ray Structure Validation Summary Report ⓘ

May 21, 2020 – 08:35 am BST

PDB ID : 4O9P
Title : Crystal structure of Thermus thermophilis transhydrogenase domain II dimer
SeMet derivative
Authors : Leung, J.H.; Yamaguchi, M.; Moeller, A.; Schurig-Briccio, L.A.; Gennis, R.B.;
Potter, C.S.; Carragher, B.; Stout, C.D.
Deposited on : 2014-01-02
Resolution : 2.89 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
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

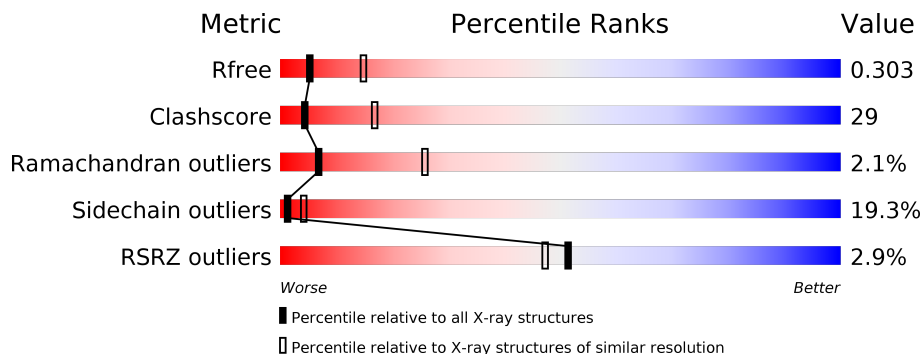
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.89 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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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	100	 45% 36% 12% • 6%
1	C	100	 2% 42% 36% 11% 11%
2	B	283	 2% 51% 34% 8% • 7%
2	D	283	 4% 51% 32% 8% • 8%

2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 5191 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 NAD(P) transhydrogenase subunit alpha 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	Se			
1	C	89	666	444	105	112	5	0	0	0
1	A	94	713	475	114	119	5	0	0	0

- Molecule 2 is a protein called NAD(P) transhydrogenase subunit beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	Se			
2	D	260	1893	1262	302	313	16	0	0	0
2	B	264	1919	1280	306	317	16	0	0	0

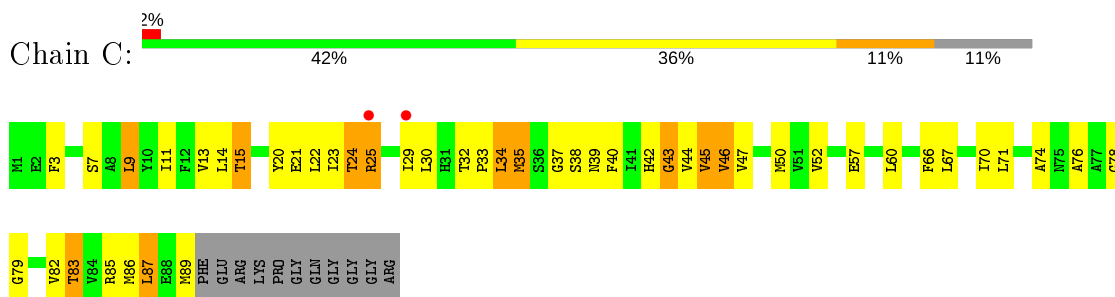
There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	276	HIS	-	EXPRESSION TAG	UNP Q72GS0
D	277	HIS	-	EXPRESSION TAG	UNP Q72GS0
D	278	HIS	-	EXPRESSION TAG	UNP Q72GS0
D	279	HIS	-	EXPRESSION TAG	UNP Q72GS0
D	280	HIS	-	EXPRESSION TAG	UNP Q72GS0
D	281	HIS	-	EXPRESSION TAG	UNP Q72GS0
D	282	HIS	-	EXPRESSION TAG	UNP Q72GS0
D	283	HIS	-	EXPRESSION TAG	UNP Q72GS0
B	276	HIS	-	EXPRESSION TAG	UNP Q72GS0
B	277	HIS	-	EXPRESSION TAG	UNP Q72GS0
B	278	HIS	-	EXPRESSION TAG	UNP Q72GS0
B	279	HIS	-	EXPRESSION TAG	UNP Q72GS0
B	280	HIS	-	EXPRESSION TAG	UNP Q72GS0
B	281	HIS	-	EXPRESSION TAG	UNP Q72GS0
B	282	HIS	-	EXPRESSION TAG	UNP Q72GS0
B	283	HIS	-	EXPRESSION TAG	UNP Q72GS0

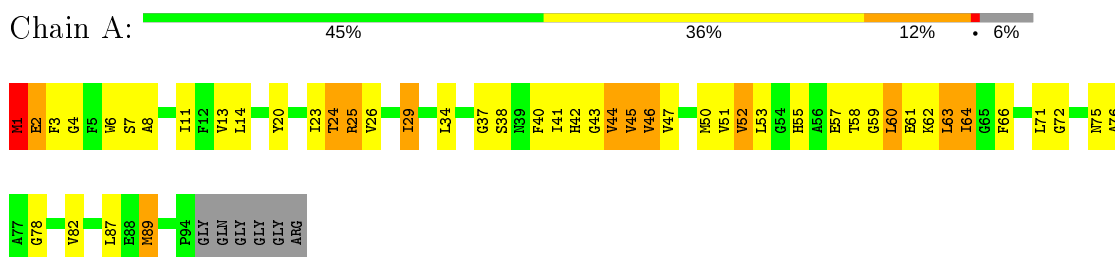
3 Residue-property plots [i](#)

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.

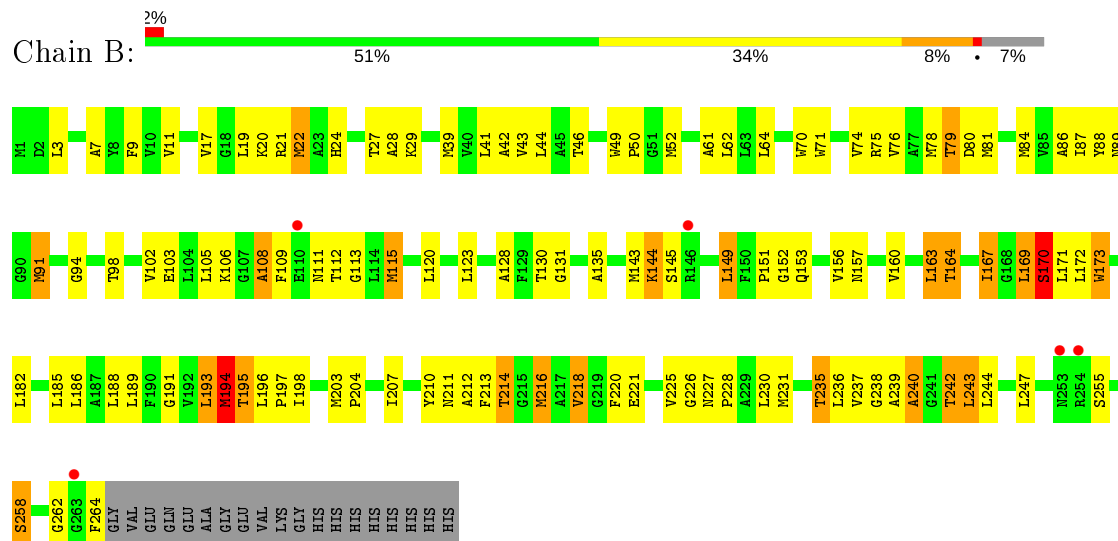
- Molecule 1: NAD(P) transhydrogenase subunit alpha 2



- Molecule 1: NAD(P) transhydrogenase subunit alpha 2



- Molecule 2: NAD(P) transhydrogenase subunit beta



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	126.57Å 86.87Å 99.16Å 90.00° 94.91° 90.00°	Depositor
Resolution (Å)	97.93 – 2.89 37.84 – 2.89	Depositor EDS
% Data completeness (in resolution range)	93.6 (97.93-2.89) 93.8 (37.84-2.89)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.53 (at 2.90Å)	Xtrriage
Refinement program	REFMAC 5.8.0049	Depositor
R, R_{free}	0.222 , 0.288 0.235 , 0.303	Depositor DCC
R_{free} test set	1159 reflections (5.13%)	wwPDB-VP
Wilson B-factor (Å ²)	49.0	Xtrriage
Anisotropy	0.288	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 45.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.87	EDS
Total number of atoms	5191	wwPDB-VP
Average B, all atoms (Å ²)	37.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.48% 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	0.67	0/726	0.96	3/978 (0.3%)
1	C	0.54	0/677	0.85	1/913 (0.1%)
2	B	0.68	0/1946	0.92	3/2626 (0.1%)
2	D	0.62	0/1919	0.86	5/2590 (0.2%)
All	All	0.64	0/5268	0.90	12/7107 (0.2%)

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
2	B	0	1

There are no bond length outliers.

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
2	D	75	ARG	NE-CZ-NH1	6.83	123.71	120.30
1	A	89	MSE	CG-SE-CE	-6.69	84.18	98.90
1	A	1	MSE	CG-SE-CE	6.53	113.27	98.90
2	B	216	MSE	CG-SE-CE	6.39	112.96	98.90
2	D	149	LEU	CA-CB-CG	6.01	129.12	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	226	GLY	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	713	0	739	56	0
1	C	666	0	691	76	0
2	B	1919	0	2030	111	0
2	D	1893	0	2006	148	0
All	All	5191	0	5466	308	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 29.

The worst 5 of 308 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)
2:B:135:ALA:CB	2:B:203:MSE:HE2	1.53	1.37
1:C:87:LEU:HD23	2:D:32:ILE:CD1	1.83	1.09
2:D:71:TRP:HE1	2:D:75:ARG:HD2	0.95	1.08
1:C:82:VAL:HB	2:D:81:MSE:HE1	1.32	1.08
2:B:135:ALA:CB	2:B:203:MSE:CE	2.30	1.08

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	92/100 (92%)	82 (89%)	7 (8%)	3 (3%)	4 15
1	C	87/100 (87%)	82 (94%)	3 (3%)	2 (2%)	6 23

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	262/283 (93%)	232 (88%)	24 (9%)	6 (2%)	6	23
2	D	258/283 (91%)	236 (92%)	18 (7%)	4 (2%)	9	32
All	All	699/766 (91%)	632 (90%)	52 (7%)	15 (2%)	7	26

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	D	144	LYS
2	D	174	ASN
1	A	3	PHE
1	A	58	THR
2	B	170	SER

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	73/70 (104%)	61 (84%)	12 (16%)	2	7
1	C	68/70 (97%)	52 (76%)	16 (24%)	1	2
2	B	189/188 (100%)	154 (82%)	35 (18%)	1	5
2	D	187/188 (100%)	150 (80%)	37 (20%)	1	4
All	All	517/516 (100%)	417 (81%)	100 (19%)	1	4

5 of 100 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	189	LEU
1	A	25	ARG
2	B	214	THR
2	D	193	LEU
2	D	243	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
2	D	227	ASN
2	B	153	GLN
2	B	89	ASN
2	D	5	GLN
2	B	24	HIS

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](#)

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	89/100 (89%)	-0.27	0 100 100	17, 27, 62, 87	0
1	C	84/100 (84%)	0.05	2 (2%) 59 56	21, 35, 60, 78	0
2	B	248/283 (87%)	-0.15	5 (2%) 65 63	14, 29, 54, 71	0
2	D	244/283 (86%)	0.02	12 (4%) 29 26	21, 38, 72, 100	0
All	All	665/766 (86%)	-0.08	19 (2%) 51 47	14, 33, 64, 100	0

The worst 5 of 19 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	151	PRO	5.5
1	C	29	ILE	4.7
2	D	253	ASN	4.7
2	B	253	ASN	3.9
2	D	150	PHE	3.8

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

There are no such residues in this entry.