wwPDB X-ray Structure Validation Summary Report (i)

May 13, 2020 – 04:59 am BST

PDB ID	:	2ZBT
Title	:	Crystal structure of pyridoxine biosynthesis protein from Thermus thermophilus HB8
Authors	:	Manzoku, M.; Ebihara, A.; Fujimoto, Y.; Yokoyama, S.; Kuramitsu, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2007-10-29
Resolution	:	1.65 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

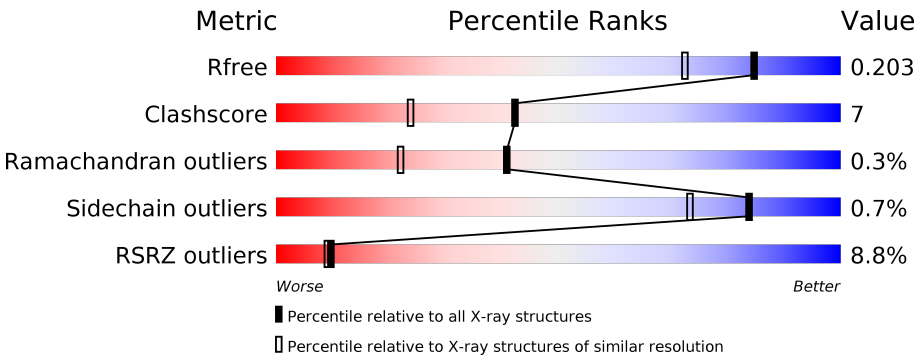
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 1.65 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	1827 (1.66-1.66)
Clashscore	141614	1931 (1.66-1.66)
Ramachandran outliers	138981	1891 (1.66-1.66)
Sidechain outliers	138945	1891 (1.66-1.66)
RSRZ outliers	127900	1791 (1.66-1.66)

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	297	8% 80% 13% 7%
1	B	297	7% 81% 11% 8%
1	C	297	7% 75% 16% 9%
1	D	297	10% 77% 15% 8%

2 Entry composition (i)

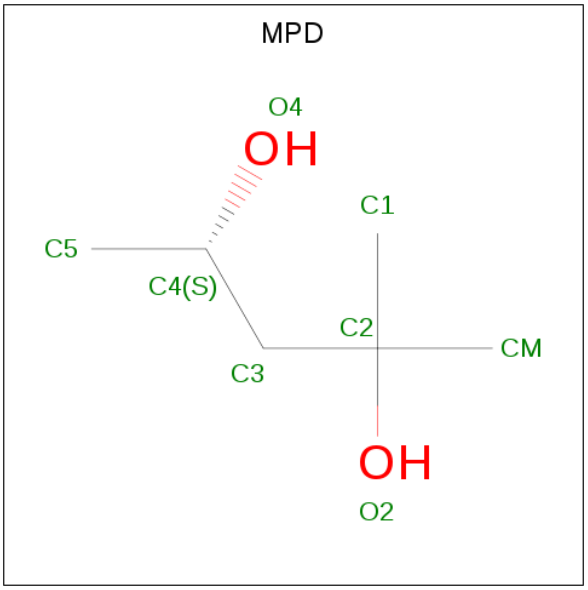
There are 3 unique types of molecules in this entry. The entry contains 8972 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Pyridoxal biosynthesis lyase pdxS.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	276	2095	1327	369	383	16	0	0	0
1	B	274	2083	1320	366	381	16	0	0	0
1	C	270	2057	1305	362	375	15	0	0	0
1	D	274	2083	1320	366	381	16	0	0	0

- Molecule 2 is (4S)-2-METHYL-2,4-PENTANEDIOL (three-letter code: MPD) (formula: $C_6H_{14}O_2$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
2	A	1	8	6	2	0	0
2	B	1	8	6	2	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
2	C	1	8	6	2	0	0
			Total	C	O		
2	D	1	8	6	2	0	0

- Molecule 3 is water.

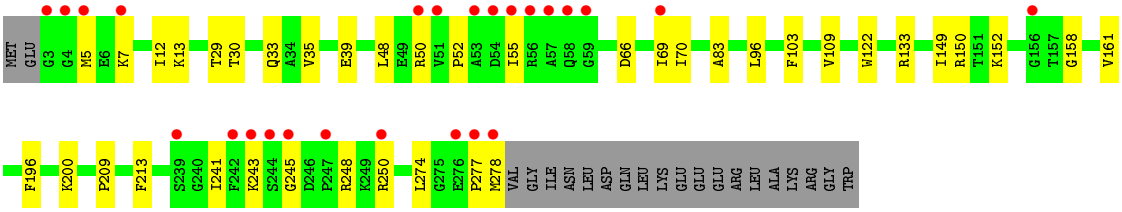
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	173	173	173	0	0
			Total	O		
3	B	171	171	171	0	0
			Total	O		
3	C	143	143	143	0	0
			Total	O		
3	D	135	135	135	0	0

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.

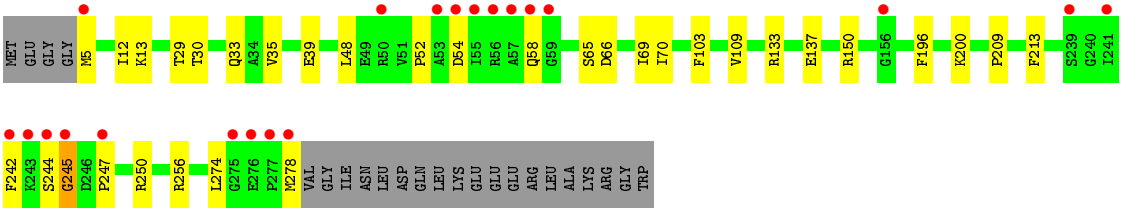
- Molecule 1: Pyridoxal biosynthesis lyase pdxS

Chain A:



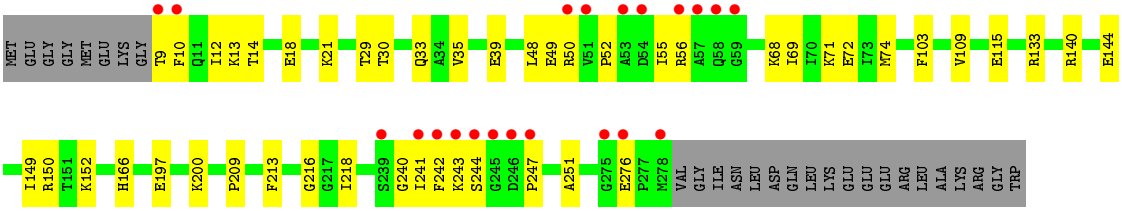
- Molecule 1: Pyridoxal biosynthesis lyase pdxS

Chain B:



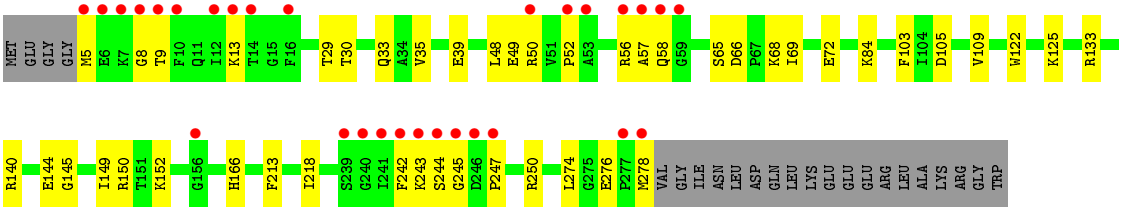
- Molecule 1: Pyridoxal biosynthesis lyase pdxS

Chain C:



- Molecule 1: Pyridoxal biosynthesis lyase pdxS

Chain D:


MET GLU GLY GLY M5 E6 K7 G8 T9 F10 Q11 I12 K13 T14 G15 F16 T29 T30 Q33 A34 V35 E39 L48 E49 R50 V51 P52 A53 R56 A57 Q58 G59 S65 D66 P67 K68 I69 E72 K84 F103 I104 D105 V109 W122 K125 R133
R140 E144 G145 I149 R150 T151 K152 G156 H166 F213 I218 S239 G240 I241 F242 K243 S244 G245 D246 P247 R250 L274 G275 E276 P277 M278 VAL GLY ILE ASN LEU ASP GLN LEU LYS GLU GLU GLU ARG LEU ALA LYS ARG GLY TRP

4 Data and refinement statistics (i)

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, α, β, γ	180.97Å 180.97Å 100.75Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	47.96 – 1.65 47.96 – 1.65	Depositor EDS
% Data completeness (in resolution range)	99.8 (47.96-1.65) 100.0 (47.96-1.65)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	5.34 (at 1.65Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.194 , 0.210 0.188 , 0.203	Depositor DCC
R_{free} test set	14786 reflections (9.99%)	wwPDB-VP
Wilson B-factor (Å^2)	17.1	Xtriage
Anisotropy	0.080	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 43.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	0.010 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	8972	wwPDB-VP
Average B, all atoms (Å^2)	21.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.53% 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 (i)

5.1 Standard geometry (i)

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

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	0.29	0/2134	0.58	0/2879
1	B	0.28	0/2122	0.58	0/2865
1	C	0.28	0/2096	0.58	0/2831
1	D	0.28	0/2122	0.57	0/2865
All	All	0.28	0/8474	0.58	0/11440

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	2095	0	2112	28	0
1	B	2083	0	2095	24	0
1	C	2057	0	2075	38	0
1	D	2083	0	2095	35	0
2	A	8	0	14	0	0
2	B	8	0	13	0	0
2	C	8	0	14	1	0
2	D	8	0	14	0	0
3	A	173	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	171	0	0	0	0
3	C	143	0	0	0	0
3	D	135	0	0	1	0
All	All	8972	0	8432	116	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.

The worst 5 of 116 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:D:242:PHE:HA	1:D:247:PRO:HB3	1.47	0.95
1:C:13:LYS:HA	1:D:5:MET:HG3	1.50	0.91
1:C:133:ARG:HB2	1:C:133:ARG:NH2	2.03	0.73
1:B:242:PHE:HA	1:B:247:PRO:HB3	1.73	0.71
1:D:133:ARG:HH11	1:D:133:ARG:HB2	1.56	0.71

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	274/297 (92%)	270 (98%)	3 (1%)	1 (0%)	34	16
1	B	272/297 (92%)	269 (99%)	2 (1%)	1 (0%)	34	16
1	C	268/297 (90%)	262 (98%)	6 (2%)	0	100	100
1	D	272/297 (92%)	265 (97%)	6 (2%)	1 (0%)	34	16
All	All	1086/1188 (91%)	1066 (98%)	17 (2%)	3 (0%)	41	22

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	245	GLY
1	A	245	GLY
1	D	245	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	211/229 (92%)	210 (100%)	1 (0%)	88	81
1	B	210/229 (92%)	209 (100%)	1 (0%)	88	81
1	C	208/229 (91%)	206 (99%)	2 (1%)	76	62
1	D	210/229 (92%)	208 (99%)	2 (1%)	76	62
All	All	839/916 (92%)	833 (99%)	6 (1%)	84	73

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	56	ARG
1	D	125	LYS
1	C	103	PHE
1	B	103	PHE
1	D	103	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	117	HIS
1	D	58	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)

4 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	$\#\|Z\| > 2$
2	MPD	D	902	-	7,7,7	0.31	0	9,10,10	1.82	3 (33%)
2	MPD	B	904	-	7,7,7	0.30	0	9,10,10	2.63	4 (44%)
2	MPD	C	901	-	7,7,7	0.32	0	9,10,10	2.05	2 (22%)
2	MPD	A	903	-	7,7,7	0.31	0	9,10,10	1.85	3 (33%)

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	MPD	D	902	-	-	2/5/5/5	-
2	MPD	B	904	-	-	1/5/5/5	-
2	MPD	C	901	-	-	3/5/5/5	-
2	MPD	A	903	-	-	3/5/5/5	-

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	904	MPD	O4-C4-C3	5.76	134.64	111.36
2	C	901	MPD	CM-C2-C3	-4.53	88.88	109.96
2	D	902	MPD	CM-C2-C3	-3.77	92.41	109.96
2	B	904	MPD	O2-C2-C3	3.60	123.32	109.80
2	A	903	MPD	CM-C2-C3	-3.30	94.59	109.96

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	D	902	MPD	O2-C2-C3-C4
2	A	903	MPD	O2-C2-C3-C4
2	C	901	MPD	O2-C2-C3-C4
2	D	902	MPD	CM-C2-C3-C4
2	B	904	MPD	CM-C2-C3-C4

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	C	901	MPD	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	276/297 (92%)	0.14	25 (9%)	9	8	11, 17, 42, 59	0
1	B	274/297 (92%)	0.13	21 (7%)	13	13	11, 17, 42, 51	0
1	C	270/297 (90%)	0.19	21 (7%)	13	12	11, 18, 43, 55	0
1	D	274/297 (92%)	0.31	29 (10%)	6	5	12, 18, 48, 60	0
All	All	1094/1188 (92%)	0.19	96 (8%)	10	9	11, 17, 44, 60	0

The worst 5 of 96 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	9	THR	7.9
1	B	56	ARG	7.2
1	D	59	GLY	7.1
1	A	278	MET	6.4
1	D	8	GLY	6.4

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(Å^2)	Q<0.9
2	MPD	A	903	8/8	0.53	0.26	34,37,37,38	0
2	MPD	B	904	8/8	0.71	0.29	34,36,36,37	0
2	MPD	D	902	8/8	0.78	0.18	32,34,34,35	0
2	MPD	C	901	8/8	0.86	0.14	30,33,35,35	0

6.5 Other polymers (i)

There are no such residues in this entry.