



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 10, 2024 – 08:11 AM EST

PDB ID : 2R8P
Title : Transketolase from E. coli in complex with substrate D-fructose-6-phosphate
Authors : Wille, G.; Asztalos, P.; Weiss, M.S.; Tittmann, K.
Deposited on : 2007-09-11
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 ⓘ symbol.

The types of validation reports are described at

<https://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 : **FAILED**
Xtriage (Phenix) : 1.13
EDS : 2.36
buster-report : **FAILED**
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) : 2.36

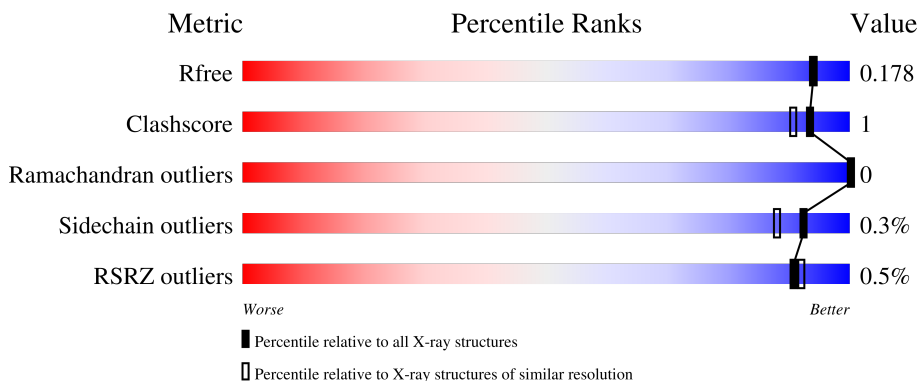
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 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 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	669	98%
1	B	669	96%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	EDO	B	676	-	-	X	-

2 Entry composition i

There are 5 unique types of molecules in this entry. The entry contains 11282 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 Transketolase 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	662	5077	3211	873	967	26	0	1	0
1	B	666	5130	3243	889	971	27	0	2	0

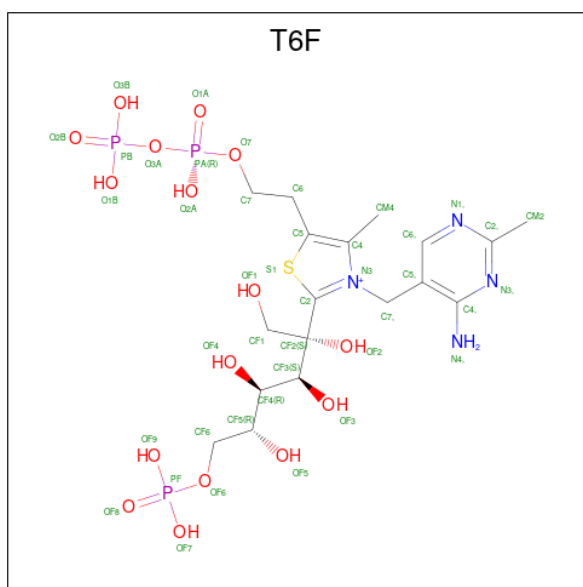
There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	664	HIS	-	expression tag	UNP P27302
A	665	HIS	-	expression tag	UNP P27302
A	666	HIS	-	expression tag	UNP P27302
A	667	HIS	-	expression tag	UNP P27302
A	668	HIS	-	expression tag	UNP P27302
A	669	HIS	-	expression tag	UNP P27302
B	664	HIS	-	expression tag	UNP P27302
B	665	HIS	-	expression tag	UNP P27302
B	666	HIS	-	expression tag	UNP P27302
B	667	HIS	-	expression tag	UNP P27302
B	668	HIS	-	expression tag	UNP P27302
B	669	HIS	-	expression tag	UNP P27302

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

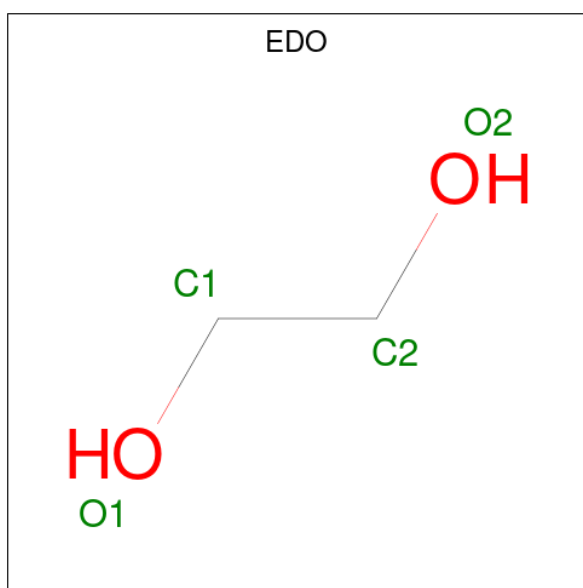
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Ca	0	0
			1	1		
2	B	1	Total	Ca	0	0
			1	1		

- Molecule 3 is 2-C-{3-[(4-amino-2-methylpyrimidin-5-yl)methyl]-5-(2-[(R)-hydroxy(phosphonoxy)phosphoryl]oxy)ethyl)-4-methyl-1,3-thiazol-3-ium-2-yl}-6-O-phosphono-D-glucitol (three-letter code: T6F) (formula: C₁₈H₃₂N₄O₁₆P₃S).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	
			Total	C	N	O	P			S
3	A	1	Total	C	N	O	P	S	0	0
			42	18	4	16	3	1		
3	B	1	Total	C	N	O	P	S	0	0
			42	18	4	16	3	1		

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C₂H₆O₂).



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

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	1	Total C O 4 2 2	0	0
4	B	1	Total C O 4 2 2	0	0
4	B	1	Total C O 4 2 2	0	0
4	B	1	Total C O 4 2 2	0	0
4	B	1	Total C O 4 2 2	0	0
4	B	1	Total C O 4 2 2	0	0
4	B	1	Total C O 4 2 2	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	480	Total O 480 480	0	0
5	B	473	Total O 473 473	0	0

3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

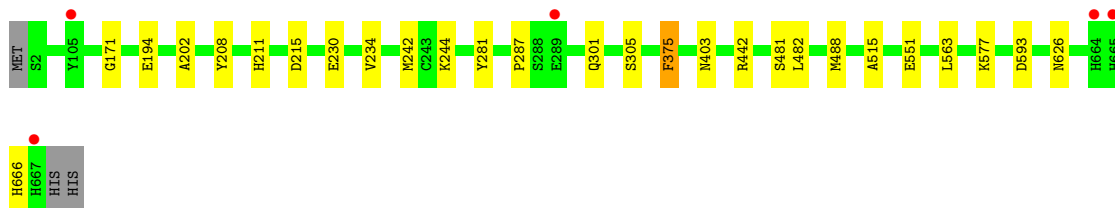
- Molecule 1: Transketolase 1

Chain A:  98%



- Molecule 1: Transketolase 1

Chain B:  96%



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	90.21Å 101.86Å 133.34Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.87 – 1.65 35.08 – 1.65	Depositor EDS
% Data completeness (in resolution range)	99.5 (39.87-1.65) 99.5 (35.08-1.65)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	7.40 (at 1.65Å)	Xtrriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.156 , 0.180 0.155 , 0.178	Depositor DCC
R_{free} test set	2182 reflections (1.48%)	wwPDB-VP
Wilson B-factor (Å ²)	8.4	Xtrriage
Anisotropy	0.031	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 43.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	11282	wwPDB-VP
Average B, all atoms (Å ²)	8.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.67% 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: CA, T6F, EDO

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	0.50	0/5198	0.59	0/7045
1	B	0.51	0/5255	0.61	1/7121 (0.0%)
All	All	0.51	0/10453	0.60	1/14166 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	593	ASP	CB-CG-OD1	5.47	123.22	118.30

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	5077	0	4947	4	0
1	B	5130	0	4991	22	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	42	0	27	1	0
3	B	42	0	27	1	0
4	A	4	0	6	0	0
4	B	32	0	48	4	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	A	480	0	0	1	0
5	B	473	0	0	5	0
All	All	11282	0	10046	28	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 1.

The worst 5 of 28 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:215:ASP:HA	1:B:244:LYS:HD2	1.78	0.64
1:B:403:ASN:O	4:B:676:EDO:H22	2.03	0.58
1:B:194:GLU:HG3	5:B:1054:HOH:O	2.03	0.57
1:B:403:ASN:O	4:B:676:EDO:C2	2.54	0.56
1:B:626:ASN:O	1:B:666:HIS:HE1	1.88	0.56

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	661/669 (99%)	648 (98%)	13 (2%)	0	100 100
1	B	666/669 (100%)	653 (98%)	13 (2%)	0	100 100
All	All	1327/1338 (99%)	1301 (98%)	26 (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	516/522 (99%)	515 (100%)	1 (0%)	93	89
1	B	521/522 (100%)	519 (100%)	2 (0%)	91	85
All	All	1037/1044 (99%)	1034 (100%)	3 (0%)	92	88

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	603	LYS
1	B	375	PHE
1	B	563	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	301	GLN
1	B	453	GLN
1	B	666	HIS
1	B	510	GLN
1	B	258	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](#)

Mogul failed to run properly - this section is therefore empty.

5.5 Carbohydrates [i](#)

Mogul failed to run properly - this section is therefore empty.

5.6 Ligand geometry [i](#)

Mogul failed to run properly - this section is therefore empty.

5.7 Other polymers [i](#)

Mogul failed to run properly - this section is therefore empty.

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	662/669 (98%)	-0.35	2 (0%) 94 94	3, 6, 14, 21	0
1	B	666/669 (99%)	-0.35	5 (0%) 86 88	3, 7, 15, 26	0
All	All	1328/1338 (99%)	-0.35	7 (0%) 91 92	3, 7, 14, 26	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	105	TYR	10.0
1	A	105	TYR	8.2
1	B	665	HIS	3.4
1	B	667	HIS	2.7
1	B	664	HIS	2.2

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
4	EDO	B	674	4/4	0.60	0.25	31,31,32,32	0
4	EDO	B	676	4/4	0.87	0.24	11,15,17,19	0
4	EDO	A	672	4/4	0.88	0.11	20,22,23,23	0
4	EDO	B	678	4/4	0.89	0.17	18,18,18,19	0
3	T6F	A	671	42/42	0.95	0.11	3,12,19,20	16
4	EDO	B	675	4/4	0.96	0.07	11,13,13,14	0
4	EDO	B	672	4/4	0.96	0.07	12,13,13,15	0
4	EDO	B	677	4/4	0.96	0.09	16,17,18,20	0
3	T6F	B	671	42/42	0.96	0.10	4,12,16,17	16
4	EDO	B	673	4/4	0.97	0.11	5,8,12,12	0
4	EDO	B	679	4/4	0.97	0.14	11,13,14,14	0
2	CA	B	670	1/1	1.00	0.04	5,5,5,5	0
2	CA	A	670	1/1	1.00	0.03	5,5,5,5	0

6.5 Other polymers [i](#)

There are no such residues in this entry.