Full wwPDB X-ray Structure Validation Report ⓘ

Jan 20, 2024 – 12:10 pm GMT

PDB ID	:	7P75
Title	:	Re-engineered 2-deoxy-D-ribose-5-phosphate aldolase catalysing asymmetric Michael addition reactions in substrate-free state
Authors	:	Thunnissen, A.M.W.H.; Rozeboom, H.J.; Kunzendorf, A.; Poelarends, G.J.
Deposited on	:	2021-07-19
Resolution	:	1.23 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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:

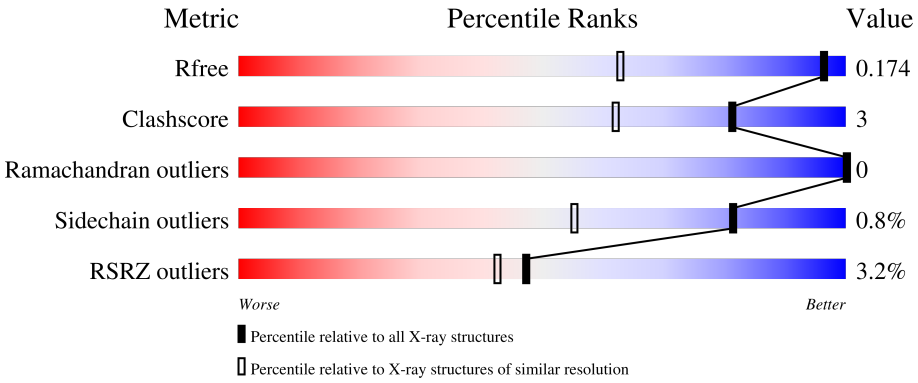
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 1.23 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	2024 (1.28-1.20)
Clashscore	141614	1007 (1.26-1.22)
Ramachandran outliers	138981	2053 (1.28-1.20)
Sidechain outliers	138945	2051 (1.28-1.20)
RSRZ outliers	127900	1987 (1.28-1.20)

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 <=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	AAA	267	3% 88% 6% 6%
1	BBB	267	3% 90% · 5%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 8217 atoms, of which 3906 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 Deoxyribose-phosphate aldolase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	AAA	250	Total	C	H	N	O	S	46	5	0
			3853	1198	1957	327	362	9			
1	BBB	253	Total	C	H	N	O	S	47	3	0
			3851	1198	1949	327	368	9			

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
AAA	18	SER	THR	engineered mutation	UNP B1IS38
AAA	22	GLY	ASP	engineered mutation	UNP B1IS38
AAA	24	TYR	ASP	engineered mutation	UNP B1IS38
AAA	47	SER	CYS	engineered mutation	UNP B1IS38
AAA	52	SER	PHE	engineered mutation	UNP B1IS38
AAA	142	SER	THR	engineered mutation	UNP B1IS38
AAA	172	LEU	LYS	engineered mutation	UNP B1IS38
AAA	197	SER	THR	engineered mutation	UNP B1IS38
AAA	202	VAL	PRO	engineered mutation	UNP B1IS38
AAA	203	THR	ALA	engineered mutation	UNP B1IS38
AAA	206	ALA	VAL	engineered mutation	UNP B1IS38
AAA	239	GLY	SER	engineered mutation	UNP B1IS38
AAA	260	LEU	-	expression tag	UNP B1IS38
AAA	261	GLU	-	expression tag	UNP B1IS38
AAA	262	HIS	-	expression tag	UNP B1IS38
AAA	263	HIS	-	expression tag	UNP B1IS38
AAA	264	HIS	-	expression tag	UNP B1IS38
AAA	265	HIS	-	expression tag	UNP B1IS38
AAA	266	HIS	-	expression tag	UNP B1IS38
AAA	267	HIS	-	expression tag	UNP B1IS38
BBB	18	SER	THR	engineered mutation	UNP B1IS38
BBB	22	GLY	ASP	engineered mutation	UNP B1IS38
BBB	24	TYR	ASP	engineered mutation	UNP B1IS38
BBB	47	SER	CYS	engineered mutation	UNP B1IS38
BBB	52	SER	PHE	engineered mutation	UNP B1IS38

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
BBB	142	SER	THR	engineered mutation	UNP B1IS38
BBB	172	LEU	LYS	engineered mutation	UNP B1IS38
BBB	197	SER	THR	engineered mutation	UNP B1IS38
BBB	202	VAL	PRO	engineered mutation	UNP B1IS38
BBB	203	THR	ALA	engineered mutation	UNP B1IS38
BBB	206	ALA	VAL	engineered mutation	UNP B1IS38
BBB	239	GLY	SER	engineered mutation	UNP B1IS38
BBB	260	LEU	-	expression tag	UNP B1IS38
BBB	261	GLU	-	expression tag	UNP B1IS38
BBB	262	HIS	-	expression tag	UNP B1IS38
BBB	263	HIS	-	expression tag	UNP B1IS38
BBB	264	HIS	-	expression tag	UNP B1IS38
BBB	265	HIS	-	expression tag	UNP B1IS38
BBB	266	HIS	-	expression tag	UNP B1IS38
BBB	267	HIS	-	expression tag	UNP B1IS38

- Molecule 2 is water.

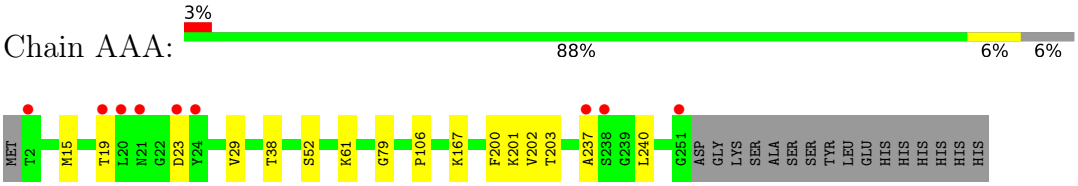
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	AAA	277	Total O 277 277	0	0
2	BBB	236	Total O 236 236	0	0

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.

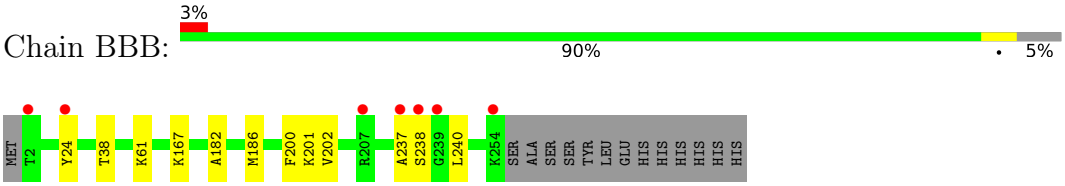
• Molecule 1: Deoxyribose-phosphate aldolase

Chain AAA:



• Molecule 1: Deoxyribose-phosphate aldolase

Chain BBB:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	47.63Å 67.50Å 73.66Å 90.00° 101.05° 90.00°	Depositor
Resolution (Å)	47.00 – 1.23 46.75 – 1.23	Depositor EDS
% Data completeness (in resolution range)	96.8 (47.00-1.23) 96.8 (46.75-1.23)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.39 (at 1.23Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.148 , 0.171 0.153 , 0.174	Depositor DCC
R_{free} test set	6430 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	12.0	Xtriage
Anisotropy	0.159	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.46 , 47.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	8217	wwPDB-VP
Average B, all atoms (Å^2)	18.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.71% 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	AAA	0.61	0/1936	0.70	0/2615
1	BBB	0.60	0/1936	0.69	0/2615
All	All	0.60	0/3872	0.70	0/5230

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

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	AAA	1896	1957	1957	13	0
1	BBB	1902	1949	1947	7	0
2	AAA	277	0	0	1	0
2	BBB	236	0	0	2	0
All	All	4311	3906	3904	20	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 3.

All (20) 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:AAA:15[A]:MET:HE3	1:AAA:237:ALA:HB3	1.85	0.58

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:AAA:61:LYS:HG3	2:AAA:474:HOH:O	2.04	0.56
1:BBB:182:ALA:O	1:BBB:186[B]:MET:HG3	2.08	0.54
1:AAA:200:PHE:CE2	1:AAA:202[A]:VAL:HG22	2.45	0.52
1:AAA:19:THR:OG1	1:AAA:23:ASP:OD1	2.21	0.50
1:AAA:15[A]:MET:CE	1:AAA:237:ALA:HB3	2.43	0.49
1:AAA:200:PHE:HE2	1:AAA:202[A]:VAL:HG22	1.78	0.47
1:AAA:23:ASP:O	1:AAA:29:VAL:HG22	2.16	0.45
1:BBB:200:PHE:CE2	1:BBB:202:VAL:HG22	2.52	0.45
1:BBB:61:LYS:HG3	2:BBB:466:HOH:O	2.15	0.45
1:BBB:167:LYS:HD2	1:BBB:201:LYS:HD3	1.99	0.45
1:AAA:237:ALA:HB1	1:AAA:240:LEU:HB3	1.99	0.45
1:AAA:19:THR:CB	1:AAA:23:ASP:OD1	2.65	0.44
1:AAA:167:LYS:HD2	1:AAA:201:LYS:HD3	1.98	0.44
1:BBB:237:ALA:HB1	1:BBB:240:LEU:HB3	1.99	0.44
1:AAA:29:VAL:HG21	1:AAA:52:SER:HB3	2.01	0.43
1:AAA:167:LYS:HE3	1:AAA:203:THR:OG1	2.19	0.42
1:BBB:238:SER:N	2:BBB:302:HOH:O	2.48	0.41
1:BBB:200:PHE:HE2	1:BBB:202:VAL:HG22	1.86	0.41
1:AAA:79:GLY:O	1:AAA:106:PRO:HA	2.22	0.40

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	AAA	253/267 (95%)	247 (98%)	6 (2%)	0	100	100
1	BBB	254/267 (95%)	249 (98%)	5 (2%)	0	100	100
All	All	507/534 (95%)	496 (98%)	11 (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	AAA	197/207 (95%)	196 (100%)	1 (0%)	88 68
1	BBB	197/207 (95%)	195 (99%)	2 (1%)	76 45
All	All	394/414 (95%)	391 (99%)	3 (1%)	81 55

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

Mol	Chain	Res	Type
1	AAA	38	THR
1	BBB	24	TYR
1	BBB	38	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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)

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

6.1 Protein, DNA and RNA chains

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, 95^{th} 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	AAA	250/267 (93%)	0.07	9 (3%) 42 38	11, 15, 28, 55	0
1	BBB	253/267 (94%)	0.05	7 (2%) 53 48	12, 17, 28, 47	0
All	All	503/534 (94%)	0.06	16 (3%) 47 42	11, 16, 28, 55	0

All (16) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	AAA	20	LEU	9.6
1	AAA	21	ASN	8.7
1	AAA	19	THR	5.6
1	BBB	237	ALA	5.2
1	AAA	2	THR	5.1
1	BBB	238	SER	4.6
1	AAA	24	TYR	4.5
1	BBB	207	ARG	3.6
1	AAA	237	ALA	3.4
1	BBB	254	LYS	3.1
1	BBB	239	GLY	3.0
1	BBB	24	TYR	2.9
1	BBB	2	THR	2.4
1	AAA	238	SER	2.3
1	AAA	251	GLY	2.2
1	AAA	23	ASP	2.1

6.2 Non-standard residues in protein, DNA, RNA chains

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.