



# Full wwPDB X-ray Structure Validation Report ⓘ

Jun 27, 2023 – 07:17 pm BST

PDB ID : 8A8Y  
Title : Structure of truncated 1-deoxy-D-xylulose 5-phosphate synthase (DXS) from *Klebsiella pneumoniae*  
Authors : Adam, S.; Hirsch, A.  
Deposited on : 2022-06-24  
Resolution : 2.10 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](mailto: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 : 2.33  
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) : 2.33

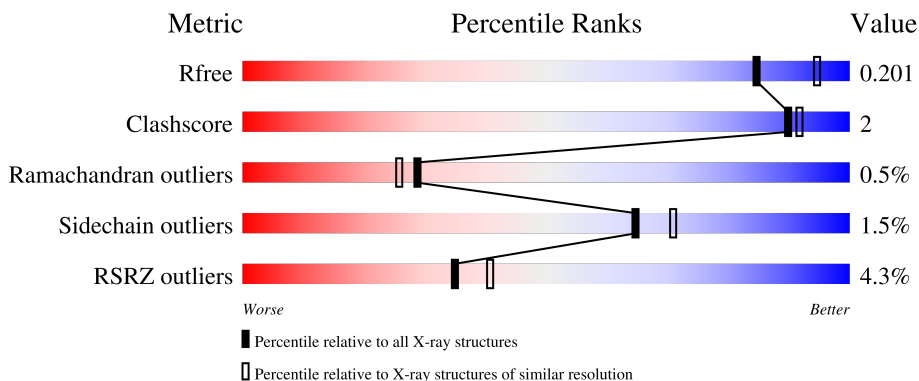
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

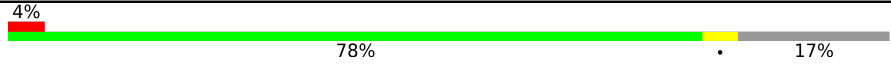

The reported resolution of this entry is 2.10 Å.

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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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  $\geq 3$ , 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	586	 4% 78% 17%
1	B	586	 3% 78% 17%

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 14982 atoms, of which 7206 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 1-deoxy-D-xylulose-5-phosphate synthase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	485	7311	2352	3622	628	690	19	0	0	0
1	B	485	7279	2358	3584	626	692	19	0	0	0

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	198	GLY	-	linker	UNP A0A0S3F934
A	199	GLY	-	linker	UNP A0A0S3F934
A	200	GLY	-	linker	UNP A0A0S3F934
A	201	GLY	-	linker	UNP A0A0S3F934
A	202	GLY	-	linker	UNP A0A0S3F934
A	203	GLY	-	linker	UNP A0A0S3F934
A	204	GLY	-	linker	UNP A0A0S3F934
B	198	GLY	-	linker	UNP A0A0S3F934
B	199	GLY	-	linker	UNP A0A0S3F934
B	200	GLY	-	linker	UNP A0A0S3F934
B	201	GLY	-	linker	UNP A0A0S3F934
B	202	GLY	-	linker	UNP A0A0S3F934
B	203	GLY	-	linker	UNP A0A0S3F934
B	204	GLY	-	linker	UNP A0A0S3F934

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	203	Total	O	0	0
			203	203		
2	B	189	Total	O	0	0
			189	189		



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	119.16Å 144.25Å 142.91Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.93 – 2.10 91.87 – 2.10	Depositor EDS
% Data completeness (in resolution range)	99.5 (45.93-2.10) 99.5 (91.87-2.10)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.08 (at 2.10Å)	Xtrriage
Refinement program	PHENIX 1.11.1_2575	Depositor
R, $R_{free}$	0.179 , 0.201 0.179 , 0.201	Depositor DCC
$R_{free}$ test set	3564 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	35.4	Xtrriage
Anisotropy	0.439	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.40 , 50.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	14982	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	52.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.58% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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	0.25	0/3760	0.43	0/5097
1	B	0.25	0/3766	0.43	0/5107
All	All	0.25	0/7526	0.43	0/10204

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	3689	3622	3700	14	0
1	B	3695	3584	3712	15	0
2	A	203	0	0	2	0
2	B	189	0	0	2	0
All	All	7776	7206	7412	27	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 2.

All (27) 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:78:VAL:HG13	1:B:80:HIS:H	1.25	0.97

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:284:LEU:H	1:A:285:PRO:HD2	1.57	0.68
1:B:78:VAL:CG1	1:B:80:HIS:H	2.09	0.61
1:A:524:GLU:OE1	1:A:567:ARG:NH1	2.35	0.59
1:B:78:VAL:HG11	1:B:80:HIS:CE1	2.38	0.58
1:A:284:LEU:H	1:A:285:PRO:CD	2.17	0.57
1:A:348:ILE:CD1	1:B:124:SER:HB3	2.34	0.57
1:B:299:ALA:O	1:B:328:ARG:NH2	2.36	0.56
1:B:17:THR:OG1	1:B:19:GLU:OE2	2.22	0.54
1:A:313:ARG:O	1:A:314:GLU:HB2	2.08	0.53
1:B:316:SER:HB2	2:B:769:HOH:O	2.11	0.51
1:B:524:GLU:OE1	1:B:567:ARG:NH1	2.45	0.50
1:A:81:GLN:HA	1:A:81:GLN:OE1	2.14	0.46
1:A:83:TYR:N	1:A:84:PRO:CD	2.79	0.45
1:A:159:MET:SD	1:A:364:ARG:NH2	2.89	0.45
1:A:402:ASP:HA	1:A:405:PHE:CE2	2.53	0.44
1:B:350:ASP:OD1	2:B:601:HOH:O	2.21	0.43
1:B:310:PRO:O	1:B:311:ALA:HB3	2.19	0.43
1:A:41:ASP:HB3	2:A:741:HOH:O	2.18	0.42
1:A:310:PRO:O	1:A:311:ALA:HB3	2.20	0.42
1:A:28:LEU:N	1:A:29:PRO:CD	2.84	0.41
1:B:286:SER:HA	1:B:446:SER:HA	2.02	0.41
1:B:152:ASP:OD2	1:B:182:GLU:HG3	2.20	0.41
1:A:466:LYS:NZ	2:A:613:HOH:O	2.51	0.41
1:A:122:HIS:ND1	1:B:367:ASP:OD2	2.45	0.40
1:B:18:GLN:HG3	1:B:18:GLN:O	2.21	0.40
1:B:14:VAL:O	1:B:18:GLN:N	2.55	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	A	473/586 (81%)	453 (96%)	16 (3%)	4 (1%)	19	15
1	B	475/586 (81%)	458 (96%)	16 (3%)	1 (0%)	47	49
All	All	948/1172 (81%)	911 (96%)	32 (3%)	5 (0%)	29	26

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	81	GLN
1	A	284	LEU
1	A	82	ALA
1	A	311	ALA
1	B	311	ALA

### 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	387/466 (83%)	381 (98%)	6 (2%)	62	69
1	B	388/466 (83%)	382 (98%)	6 (2%)	65	71
All	All	775/932 (83%)	763 (98%)	12 (2%)	65	71

All (12) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	15	ASP
1	A	73	ARG
1	A	143	GLN
1	A	284	LEU
1	A	375	ILE
1	A	441	ARG
1	B	13	LEU
1	B	25	LYS
1	B	78	VAL
1	B	233	LYS
1	B	375	ILE

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
1	B	441	ARG

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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	485/586 (82%)	0.16	25 (5%) 27 32	28, 41, 95, 141	0
1	B	485/586 (82%)	0.09	17 (3%) 44 50	28, 40, 86, 135	0
All	All	970/1172 (82%)	0.13	42 (4%) 35 41	28, 41, 93, 141	0

All (42) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	13	LEU	8.3
1	B	19	GLU	6.2
1	A	315	GLY	5.6
1	A	83	TYR	5.3
1	B	395	GLN	5.1
1	A	39	LEU	4.8
1	A	395	GLN	4.4
1	A	11	LEU	4.3
1	B	124	SER	4.2
1	A	80	HIS	4.0
1	B	393	ASP	3.9
1	A	54	LEU	3.9
1	B	17	THR	3.9
1	B	560	GLN	3.8
1	A	35	LEU	3.7
1	B	16	SER	3.5
1	A	15	ASP	3.5
1	A	121	GLY	3.2
1	A	394	GLY	3.2
1	A	38	TYR	3.1
1	A	283	GLY	2.9
1	A	314	GLU	2.9
1	A	12	ALA	2.8
1	B	314	GLU	2.8

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	B	394	GLY	2.7
1	B	83	TYR	2.7
1	A	37	ARG	2.7
1	B	5	ILE	2.7
1	A	8	TYR	2.6
1	A	88	LEU	2.6
1	B	396	THR	2.5
1	A	31	LEU	2.4
1	A	316	SER	2.4
1	A	41	ASP	2.3
1	B	316	SER	2.3
1	A	6	ALA	2.3
1	B	9	PRO	2.3
1	A	87	ILE	2.2
1	A	82	ALA	2.2
1	B	79	GLY	2.1
1	B	182	GLU	2.1
1	B	53	GLY	2.0

## 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.