



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 5, 2026 – 10:41 AM UTC

PDB ID : 4HBK / pdb_00004hbk
Title : Structure of the Aldose Reductase from Schistosoma japonicum
Authors : Liu, J.; Cheng, J.; Zhang, X.; Yang, Z.; Hu, W.; Xu, Y.
Deposited on : 2012-09-28
Resolution : 2.20 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

<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	:	4-5-2 with Phenix2.0
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

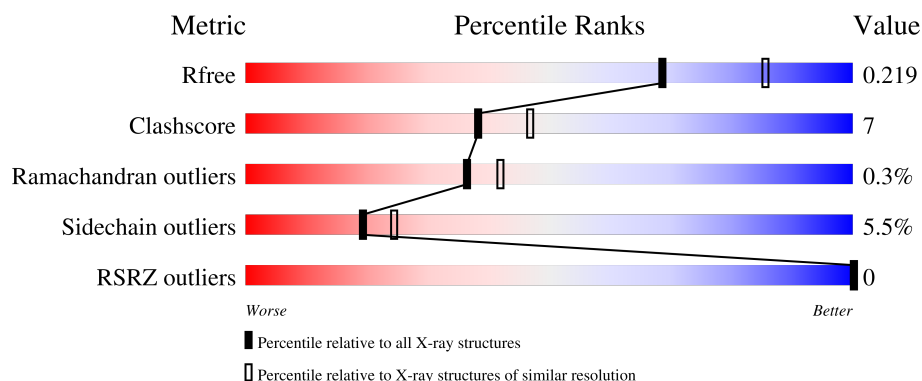
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.20 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}	180053	6164 (2.20-2.20)
Clashscore	190562	6851 (2.20-2.20)
Ramachandran outliers	187476	6768 (2.20-2.20)
Sidechain outliers	187428	6769 (2.20-2.20)
RSRZ outliers	180081	6166 (2.20-2.20)

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 $\leq 5\%$. 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	344	 72% 13% • 12%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 2677 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 Aldo-keto reductase family 1, member B4 (Aldose reductase).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	301	Total	C	N	O	S	0	0	0
			2452	1568	427	447	10			

There are 34 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-33	MET	-	expression tag	UNP Q5DD64
A	-32	GLY	-	expression tag	UNP Q5DD64
A	-31	SER	-	expression tag	UNP Q5DD64
A	-30	SER	-	expression tag	UNP Q5DD64
A	-29	HIS	-	expression tag	UNP Q5DD64
A	-28	HIS	-	expression tag	UNP Q5DD64
A	-27	HIS	-	expression tag	UNP Q5DD64
A	-26	HIS	-	expression tag	UNP Q5DD64
A	-25	HIS	-	expression tag	UNP Q5DD64
A	-24	HIS	-	expression tag	UNP Q5DD64
A	-23	SER	-	expression tag	UNP Q5DD64
A	-22	SER	-	expression tag	UNP Q5DD64
A	-21	GLY	-	expression tag	UNP Q5DD64
A	-20	LEU	-	expression tag	UNP Q5DD64
A	-19	VAL	-	expression tag	UNP Q5DD64
A	-18	PRO	-	expression tag	UNP Q5DD64
A	-17	ARG	-	expression tag	UNP Q5DD64
A	-16	GLY	-	expression tag	UNP Q5DD64
A	-15	SER	-	expression tag	UNP Q5DD64
A	-14	HIS	-	expression tag	UNP Q5DD64
A	-13	MET	-	expression tag	UNP Q5DD64
A	-12	ALA	-	expression tag	UNP Q5DD64
A	-11	SER	-	expression tag	UNP Q5DD64
A	-10	MET	-	expression tag	UNP Q5DD64
A	-9	THR	-	expression tag	UNP Q5DD64
A	-8	GLY	-	expression tag	UNP Q5DD64
A	-7	GLY	-	expression tag	UNP Q5DD64

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	-6	GLN	-	expression tag	UNP Q5DD64
A	-5	GLN	-	expression tag	UNP Q5DD64
A	-4	MET	-	expression tag	UNP Q5DD64
A	-3	GLY	-	expression tag	UNP Q5DD64
A	-2	ARG	-	expression tag	UNP Q5DD64
A	-1	GLY	-	expression tag	UNP Q5DD64
A	0	SER	-	expression tag	UNP Q5DD64

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	225	Total 225	O 225	0	0

i

- Molecule 1: Aldo-keto reductase family 1, member B4 (Aldose reductase)

R291	Q292	F293	T294	L295	M298	S299	Y304	K307	E308	E309	Y310	F83	T94	L96	L104	P114	L115	L121	F122	P123	T124	D125	S126	N127	G128	Q129	L130	C131	L132	D133	R164	E171	P189	N190	I191	G211	SER	PRO	ALA	HIS	PRO	PRO	GLY	LYS	VAL	N221	T224	K257	R263	I264	E265	Q274	S287	L288	N289	F290
Met	GLY	SER	SER	HIS	HIS	HIS	HIS	SER	SER	GLY	VAL	PRO	ARG	GLY	SER	HIS	Met	ALA	SER	Met	THR	GLY	GLY	GLN	Met	GLY	ARG	GLY	SER	M1	E2	P3	L4	S11	N21	S22	P23	K32	I37	H41	A45	Y46	V47	Y48	R49	L58	N50									

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α , β , γ	67.49Å 90.99Å 54.67Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	28.71 – 2.20 28.71 – 2.20	Depositor EDS
% Data completeness (in resolution range)	95.3 (28.71-2.20) 99.4 (28.71-2.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.96 (at 2.20Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.6.1_353)	Depositor
R, R_{free}	0.177 , 0.217 0.176 , 0.219	Depositor DCC
R_{free} test set	896 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å ²)	16.0	Xtriage
Anisotropy	0.770	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 32.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	2677	wwPDB-VP
Average B, all atoms (Å ²)	17.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.44% 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

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	A	0.46	0/2507	0.77	1/3392 (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	A	288	LEU	N-CA-C	6.66	118.20	111.07

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	A	2452	0	2464	34	0
2	A	225	0	0	3	0
All	All	2677	0	2464	34	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 34 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:A:125:ASP:HB3	1:A:127:ASN:H	1.51	0.75

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:123:PRO:HB2	1:A:131:CYS:HB2	1.78	0.66
1:A:295:LEU:HG	1:A:298:MET:HE2	1.78	0.66
1:A:289:ASN:O	1:A:289:ASN:OD1	2.22	0.58
1:A:287:SER:OG	1:A:289:ASN:HB3	2.08	0.53

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	297/344 (86%)	293 (99%)	3 (1%)	1 (0%)	36	42

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	125	ASP

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	271/304 (89%)	256 (94%)	15 (6%)	19	24

5 of 15 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	130	LEU
1	A	290	GLU
1	A	132	LEU
1	A	295	LEU
1	A	263	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	21	ASN
1	A	179	ASN
1	A	281	HIS
1	A	289	ASN
1	A	302	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 oligosaccharides 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	301/344 (87%)	-0.50	0 100 100	8, 16, 31, 38	0

There are no RSRZ outliers to report.

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 oligosaccharides 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.