



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 5, 2026 – 05:16 PM UTC

PDB ID : 7KQV / pdb_00007kqv
Title : Crystal Structure of aldehyde dehydrogenase (ChALDH) from Cladosporium herbarum
Authors : Lee, S.G.; Jez, J.M.
Deposited on : 2020-11-17
Resolution : 3.18 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

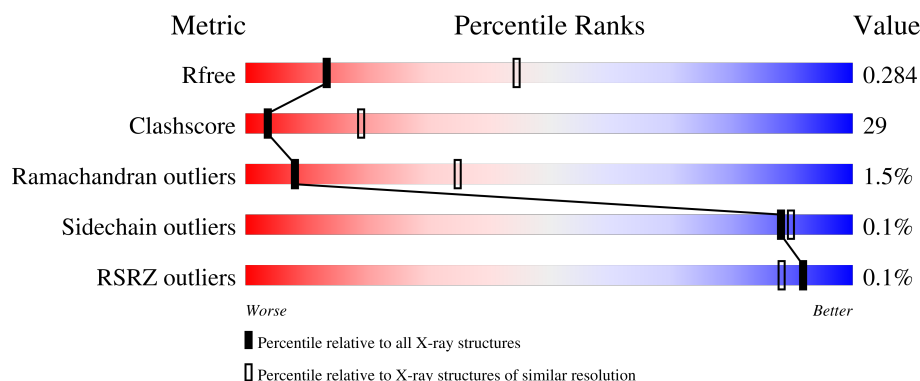
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.18 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	2001 (3.20-3.16)
Clashscore	190562	2119 (3.20-3.16)
Ramachandran outliers	187476	2070 (3.20-3.16)
Sidechain outliers	187428	2069 (3.20-3.16)
RSRZ outliers	180081	2001 (3.20-3.16)

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	496	
1	B	496	
1	C	496	
1	D	496	

2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 13484 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.

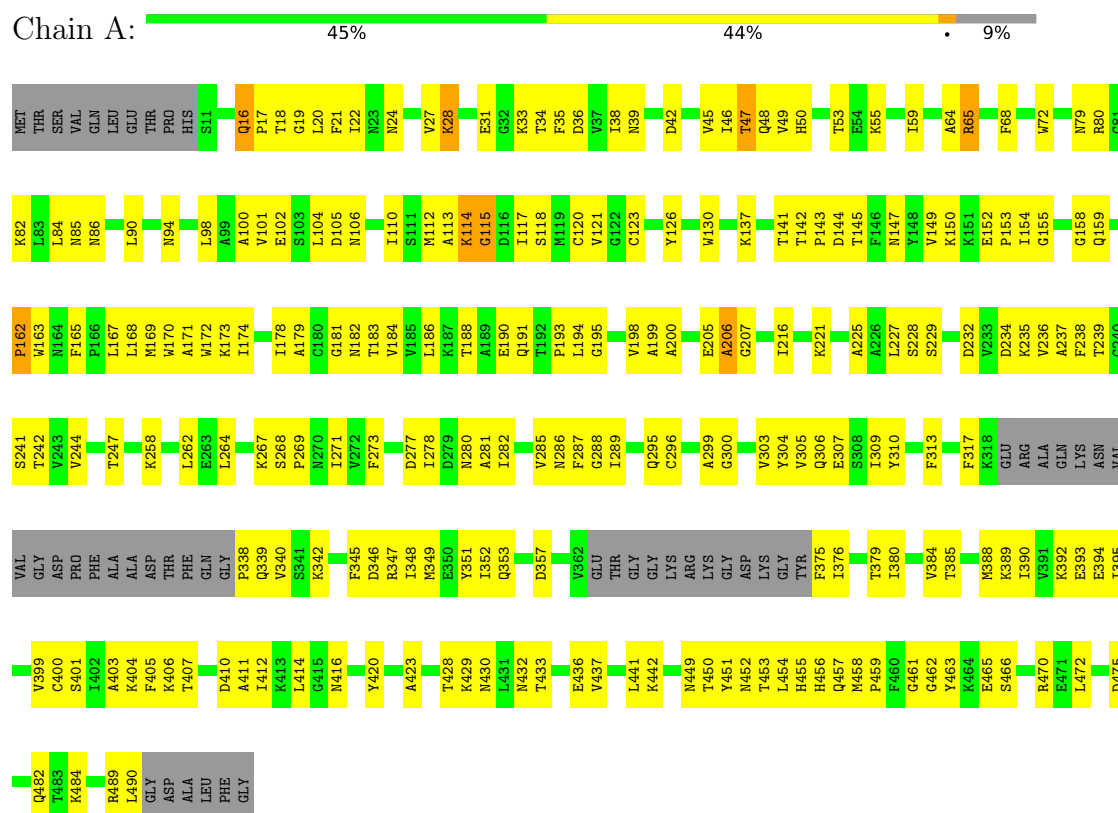
- Molecule 1 is a protein called Aldehyde dehydrogenase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	449	Total	C	N	O	S	0	0	0
			3418	2181	572	650	15			
1	B	460	Total	C	N	O	S	0	0	0
			3485	2219	582	669	15			
1	C	399	Total	C	N	O	S	0	0	0
			3016	1923	507	573	13			
1	D	469	Total	C	N	O	S	0	0	0
			3565	2269	601	680	15			

3 Residue-property plots

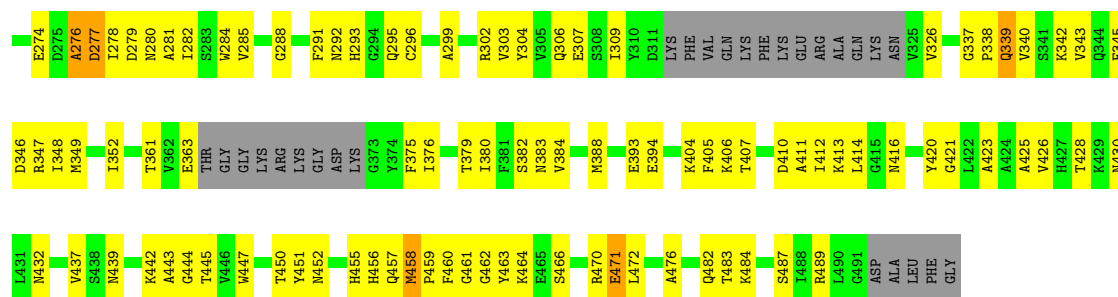
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: Aldehyde dehydrogenase

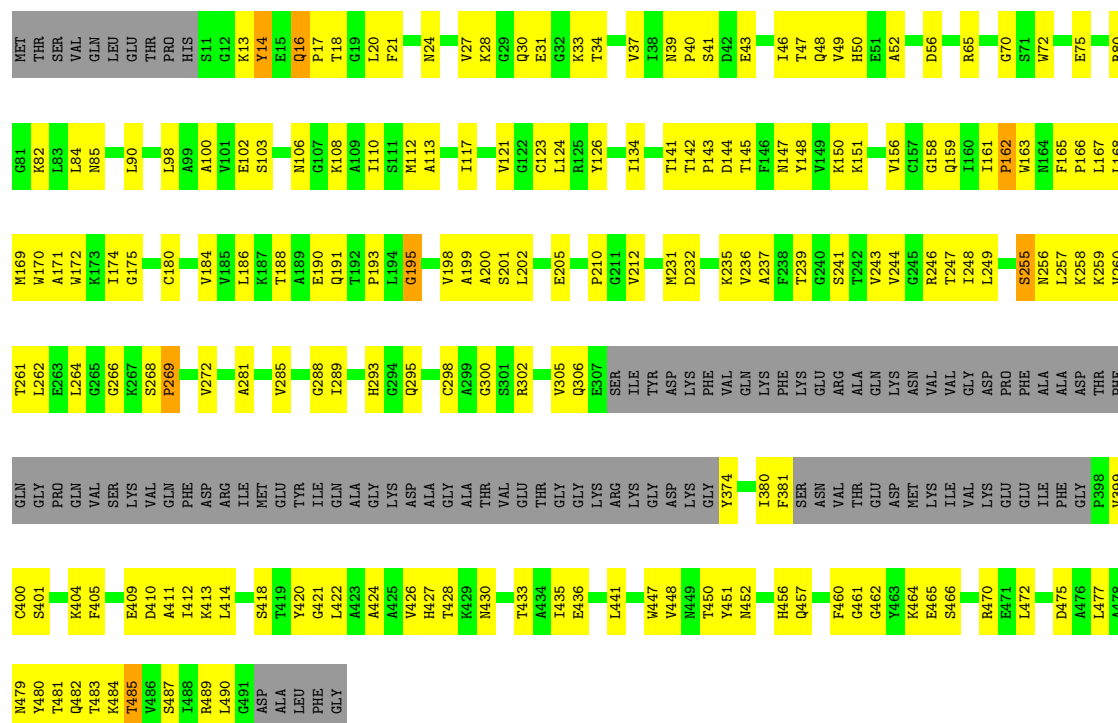


• Molecule 1: Aldehyde dehydrogenase

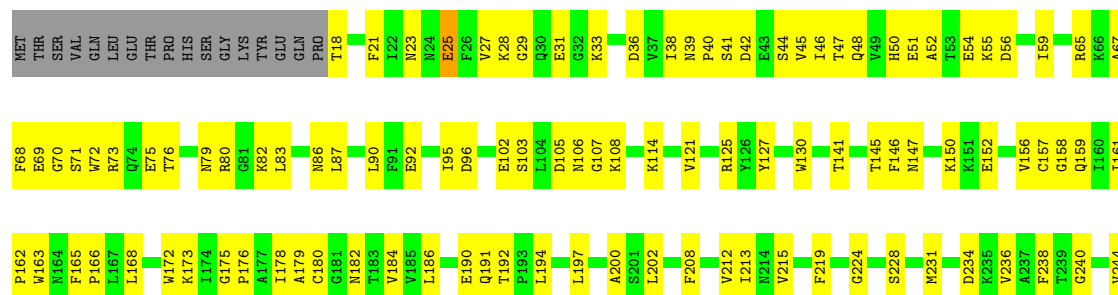




• Molecule 1: Aldehyde dehydrogenase



• Molecule 1: Aldehyde dehydrogenase



A417	S418	T419	Y420	G421	A425	T428	N432	T433	E436	V437	S438	N439	A440	L441	G444	T445	V446	T450	Y451	H455	H456	Q457	M458	G462	Y463	K464	E465	R470	E471	G472	G473	E474	D475	A476	L477	T481	Q482	T483	K484	T488	R489	L490	G491	ASP	ALA	LEU	PHE	GLY		
G337	P338	Q339	V340	S341	K342	Y343	Q344	F345	I348	I352	Q353	K356	D357	G365	G366	K367	R368	D371	K372	G373	Y374	F375	I376	E377	P378	T379	I380	T385	M388	K389	I390	V391	K392	E393	E394	I395	F396	I402	F405	K406	T407	K408	E409	D410	A411	I412	K413	L414	G415	N416
T247	I248	L249	K258	K259	V260	T261	L262	E263	L264	K267	S268	P269	V272	F273	E274	D275	A276	A281	V284	V285	N286	F287	G288	I289	F290	F291	N292	H293	G294	Q295	S301	R302	V303	Y304	V305	Q306	E307	S308	I309	K312	F313	Q322	LYS	ASN	VAL	VAL	GLY	D328	P329	

4 Data and refinement statistics

Property	Value	Source
Space group	I 4	Depositor
Cell constants a, b, c, α , β , γ	157.18Å 157.18Å 164.62Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.70 – 3.18 49.70 – 3.18	Depositor EDS
% Data completeness (in resolution range)	99.5 (49.70-3.18) 99.5 (49.70-3.18)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.27 (at 3.19Å)	Xtriage
Refinement program	PHENIX 1.14 _3260	Depositor
R, R_{free}	0.233 , 0.281 0.234 , 0.284	Depositor DCC
R_{free} test set	1693 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å ²)	79.3	Xtriage
Anisotropy	0.517	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 76.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.009 for l,-k,h 0.014 for -l,-k,-h 0.008 for -h,-l,-k 0.005 for -h,l,k 0.458 for -h,k,-l	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	13484	wwPDB-VP
Average B, all atoms (Å ²)	83.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.63% 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	0.48	1/3484 (0.0%)	0.80	1/4716 (0.0%)
1	B	0.49	0/3554	0.80	4/4817 (0.1%)
1	C	0.50	1/3076 (0.0%)	0.81	0/4170
1	D	0.49	0/3634	0.78	2/4917 (0.0%)
All	All	0.49	2/13748 (0.0%)	0.80	7/18620 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	B	0	2

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	65	ARG	CZ-NH1	5.94	1.41	1.32
1	C	485	THR	CB-CG2	5.17	1.69	1.52

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	28	LYS	CD-CE-NZ	6.79	133.64	111.90
1	D	471	GLU	CA-C-N	-6.77	110.69	121.86
1	D	471	GLU	C-N-CA	-6.77	110.69	121.86
1	B	458	MET	CB-CA-C	5.70	116.19	109.47
1	B	471	GLU	CA-C-N	-5.33	113.21	122.64

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	276	ALA	Peptide
1	B	277	ASP	Peptide

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	3418	0	3422	229	1
1	B	3485	0	3457	204	1
1	C	3016	0	3011	184	0
1	D	3565	0	3557	203	0
All	All	13484	0	13447	787	1

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 29.

The worst 5 of 787 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:356:LYS:NZ	1:D:357:ASP:OD1	1.74	1.20
1:D:353:GLN:HG3	1:D:356:LYS:CE	1.77	1.15
1:B:343:VAL:O	1:B:347:ARG:HD3	1.50	1.09
1:D:353:GLN:HG3	1:D:356:LYS:HE3	1.28	1.08
1:D:353:GLN:CG	1:D:356:LYS:HE3	1.84	1.08

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:429:LYS:NZ	1:B:432:ASN:OD1[6_455]	2.05	0.15

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	443/496 (89%)	401 (90%)	36 (8%)	6 (1%)	9	36
1	B	454/496 (92%)	411 (90%)	38 (8%)	5 (1%)	11	41
1	C	393/496 (79%)	351 (89%)	34 (9%)	8 (2%)	6	28
1	D	465/496 (94%)	428 (92%)	30 (6%)	7 (2%)	8	35
All	All	1755/1984 (88%)	1591 (91%)	138 (8%)	26 (2%)	8	35

5 of 26 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	16	GLN
1	B	276	ALA
1	B	326	VAL
1	D	70	GLY
1	A	115	GLY

5.3.2 Protein sidechains ⓘ

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	365/400 (91%)	365 (100%)	0	100	100
1	B	370/400 (92%)	369 (100%)	1 (0%)	86	86
1	C	320/400 (80%)	320 (100%)	0	100	100
1	D	377/400 (94%)	376 (100%)	1 (0%)	86	86
All	All	1432/1600 (90%)	1430 (100%)	2 (0%)	88	90

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	43	GLU
1	D	25	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 24 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	280	ASN
1	D	50	HIS
1	C	430	ASN
1	D	74	GLN
1	A	452	ASN

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(Å²)	Q<0.9	
1	A	449/496 (90%)	-1.07	0	100	100	54, 85, 137, 251	0
1	B	460/496 (92%)	-1.08	0	100	100	48, 74, 118, 222	0
1	C	399/496 (80%)	-1.10	0	100	100	54, 79, 130, 177	0
1	D	469/496 (94%)	-1.12	1 (0%)	91	85	47, 74, 119, 195	0
All	All	1777/1984 (89%)	-1.09	1 (0%)	92	88	47, 78, 129, 251	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	308	SER	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 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.