



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 22, 2023 – 07:59 PM JST

PDB ID : 7YOB
Title : Crystal structure of Aldehyde dehydrogenase 1A1 from mouse
Authors : Zhang, X.Y.; Ouyang, Z.Q.
Deposited on : 2022-08-01
Resolution : 2.89 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 : 4.02b-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

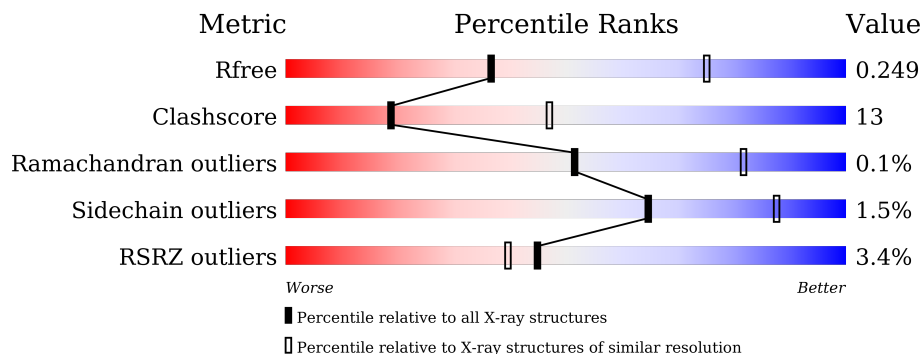
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.89 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.



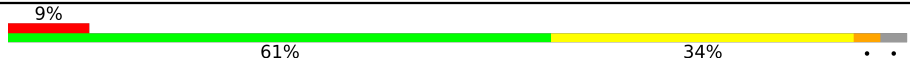

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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	507	 % 75% 22% ..
1	B	507	 % 74% 22% ..
1	C	507	 2% 73% 23% ..
1	D	507	 5% 75% 22% ..
1	E	507	 3% 70% 26% ..
1	F	507	 3% 75% 21% ..

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	507	
1	H	507	

2 Entry composition [i](#)

There is only 1 type of molecule in this entry. The entry contains 29957 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 1A1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	494	3760	2395	646	696	23	0	0	0
1	B	494	3745	2385	641	696	23	0	0	0
1	C	494	3751	2387	642	699	23	0	0	0
1	D	494	3744	2381	642	698	23	0	0	0
1	E	494	3747	2387	639	698	23	0	0	0
1	G	492	3698	2350	630	695	23	0	0	0
1	H	494	3765	2398	644	700	23	0	0	0
1	F	494	3747	2384	642	698	23	0	0	0

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-5	HIS	-	expression tag	UNP P24549
A	-4	HIS	-	expression tag	UNP P24549
A	-3	HIS	-	expression tag	UNP P24549
A	-2	HIS	-	expression tag	UNP P24549
A	-1	HIS	-	expression tag	UNP P24549
A	0	HIS	-	expression tag	UNP P24549
B	-5	HIS	-	expression tag	UNP P24549
B	-4	HIS	-	expression tag	UNP P24549
B	-3	HIS	-	expression tag	UNP P24549
B	-2	HIS	-	expression tag	UNP P24549
B	-1	HIS	-	expression tag	UNP P24549
B	0	HIS	-	expression tag	UNP P24549
C	-5	HIS	-	expression tag	UNP P24549

Continued on next page...

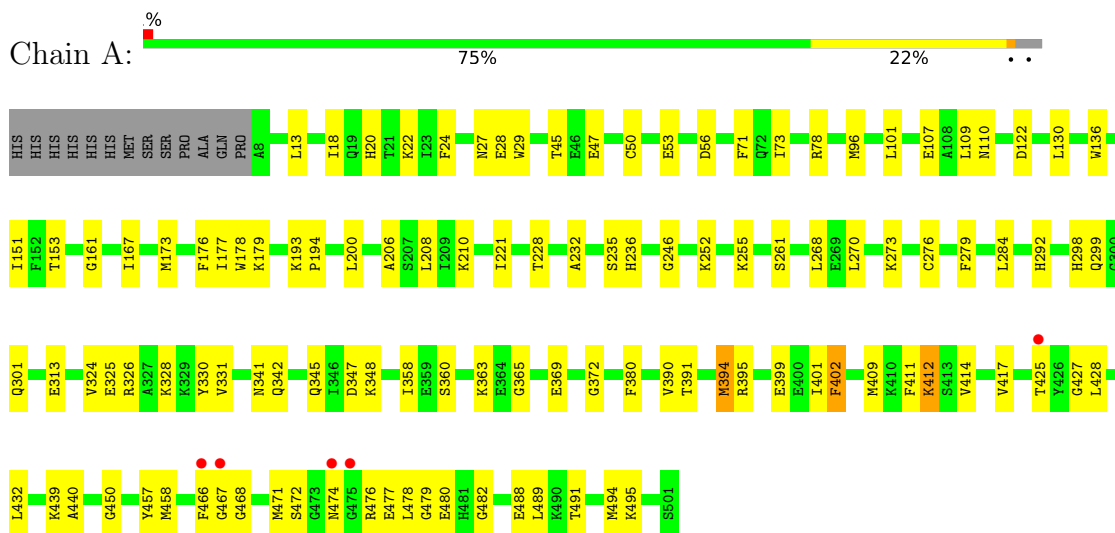
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	-4	HIS	-	expression tag	UNP P24549
C	-3	HIS	-	expression tag	UNP P24549
C	-2	HIS	-	expression tag	UNP P24549
C	-1	HIS	-	expression tag	UNP P24549
C	0	HIS	-	expression tag	UNP P24549
D	-5	HIS	-	expression tag	UNP P24549
D	-4	HIS	-	expression tag	UNP P24549
D	-3	HIS	-	expression tag	UNP P24549
D	-2	HIS	-	expression tag	UNP P24549
D	-1	HIS	-	expression tag	UNP P24549
D	0	HIS	-	expression tag	UNP P24549
E	-5	HIS	-	expression tag	UNP P24549
E	-4	HIS	-	expression tag	UNP P24549
E	-3	HIS	-	expression tag	UNP P24549
E	-2	HIS	-	expression tag	UNP P24549
E	-1	HIS	-	expression tag	UNP P24549
E	0	HIS	-	expression tag	UNP P24549
G	-5	HIS	-	expression tag	UNP P24549
G	-4	HIS	-	expression tag	UNP P24549
G	-3	HIS	-	expression tag	UNP P24549
G	-2	HIS	-	expression tag	UNP P24549
G	-1	HIS	-	expression tag	UNP P24549
G	0	HIS	-	expression tag	UNP P24549
H	-5	HIS	-	expression tag	UNP P24549
H	-4	HIS	-	expression tag	UNP P24549
H	-3	HIS	-	expression tag	UNP P24549
H	-2	HIS	-	expression tag	UNP P24549
H	-1	HIS	-	expression tag	UNP P24549
H	0	HIS	-	expression tag	UNP P24549
F	-5	HIS	-	expression tag	UNP P24549
F	-4	HIS	-	expression tag	UNP P24549
F	-3	HIS	-	expression tag	UNP P24549
F	-2	HIS	-	expression tag	UNP P24549
F	-1	HIS	-	expression tag	UNP P24549
F	0	HIS	-	expression tag	UNP P24549

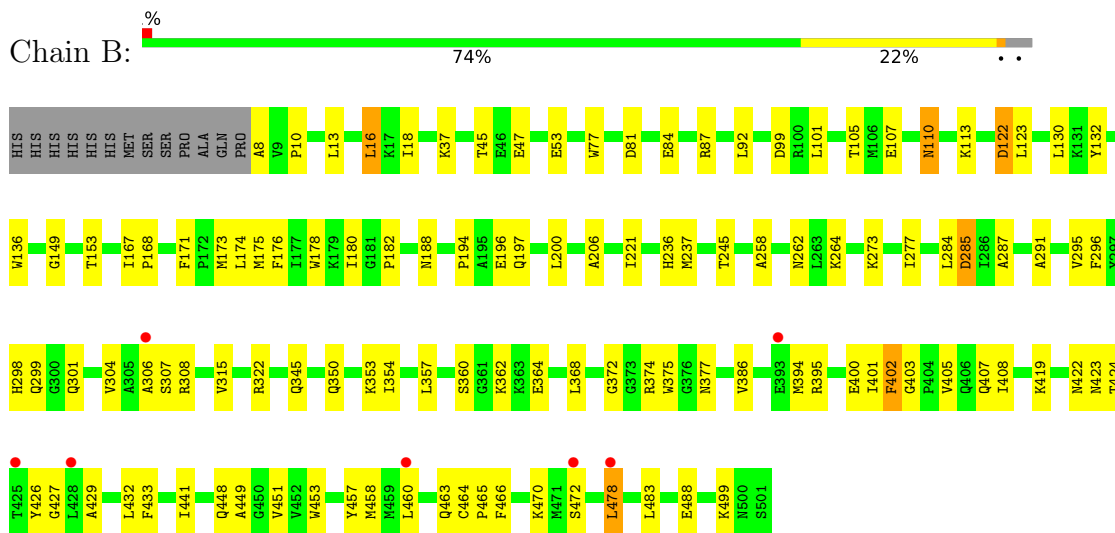
3 Residue-property plots i

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.

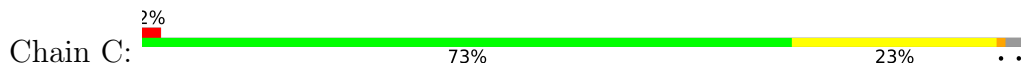
• Molecule 1: Aldehyde dehydrogenase 1A1

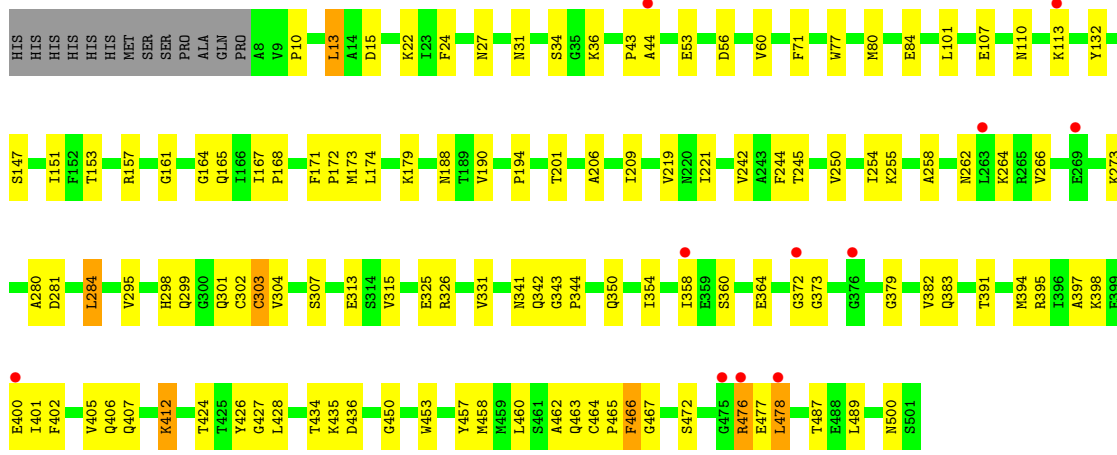


• Molecule 1: Aldehyde dehydrogenase 1A1

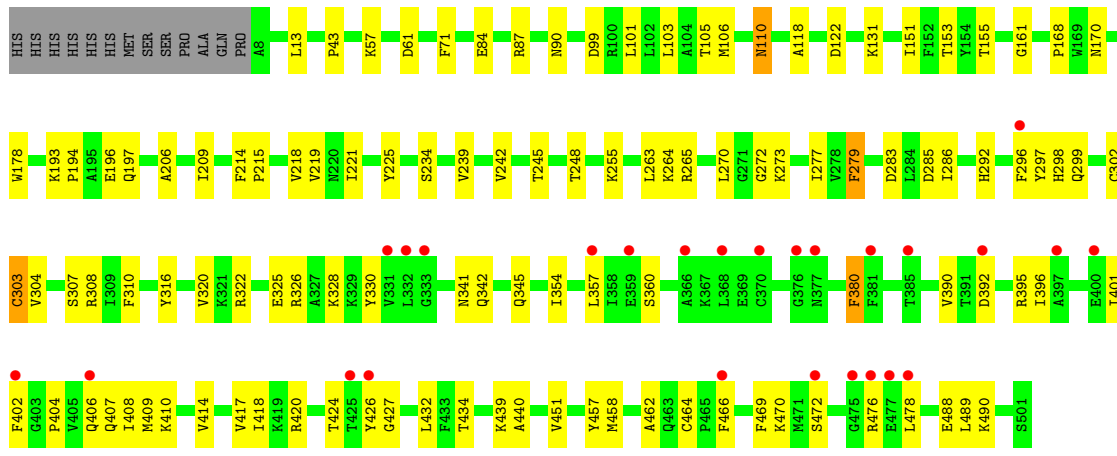
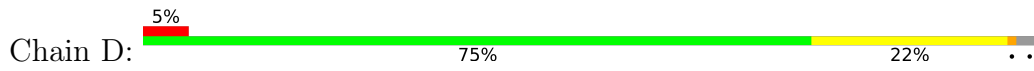


• Molecule 1: Aldehyde dehydrogenase 1A1

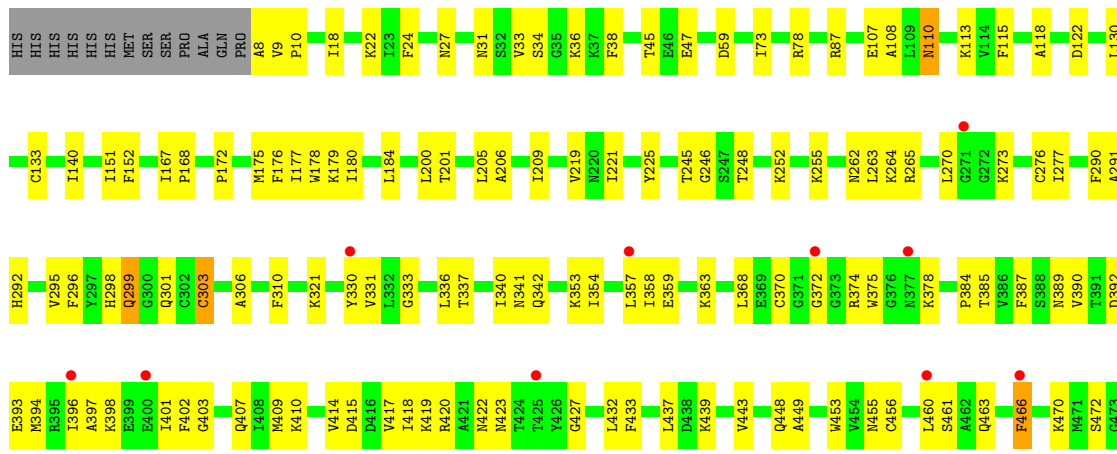




• Molecule 1: Aldehyde dehydrogenase 1A1

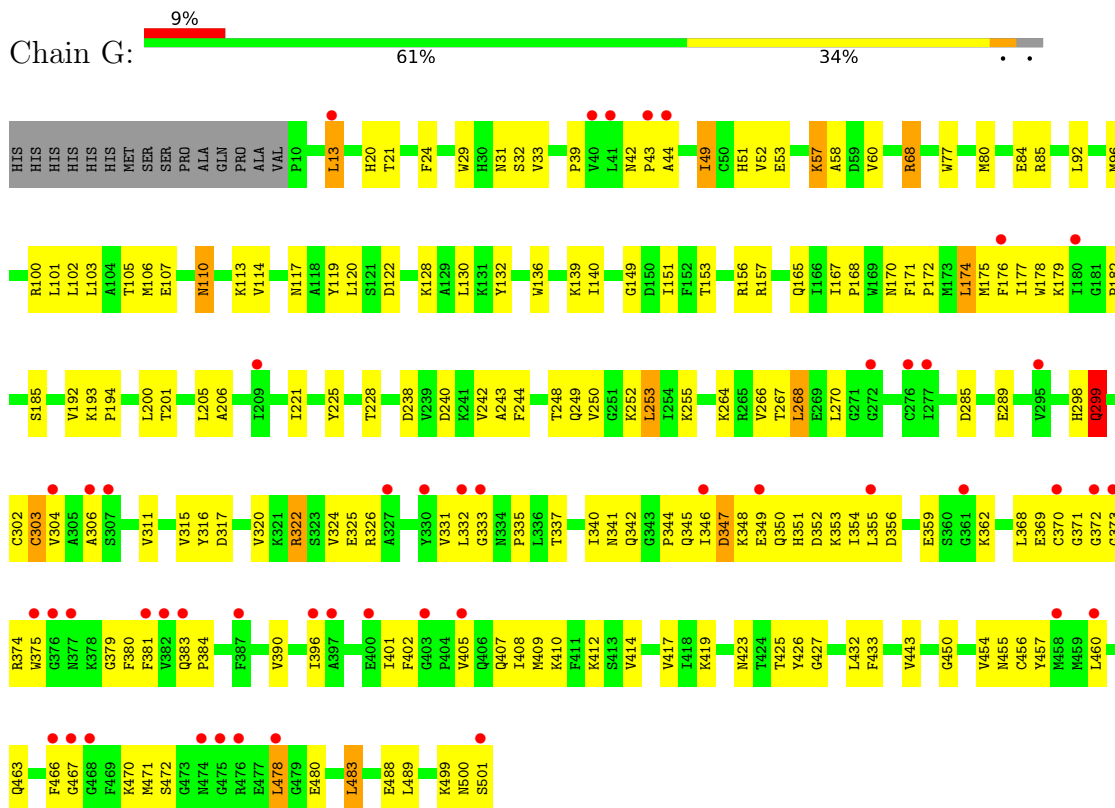


• Molecule 1: Aldehyde dehydrogenase 1A1

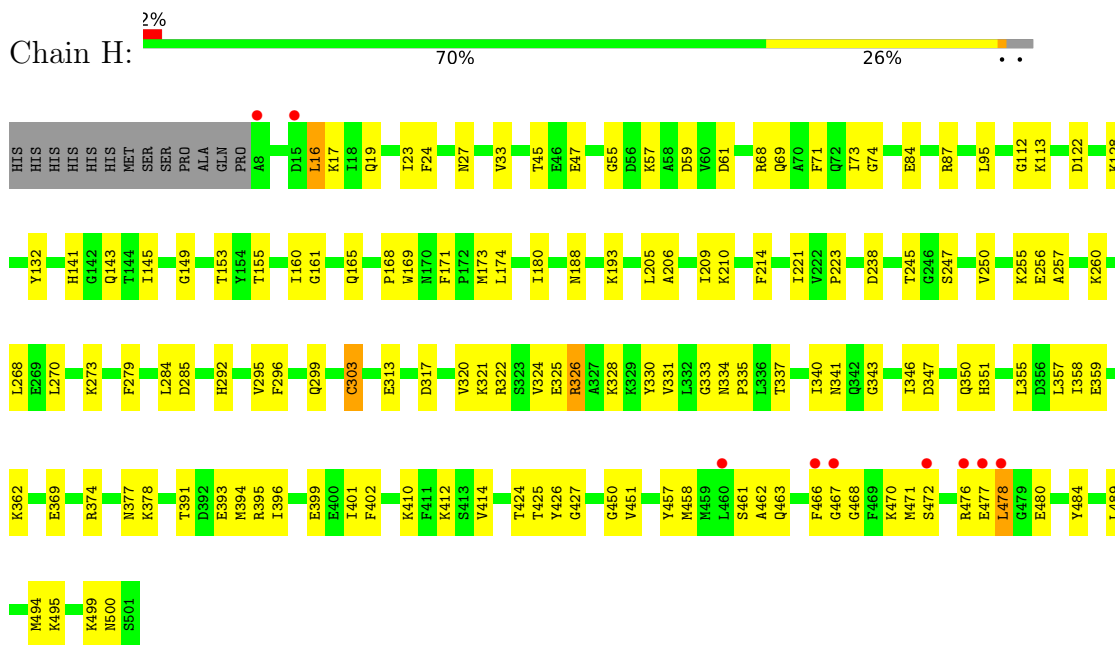




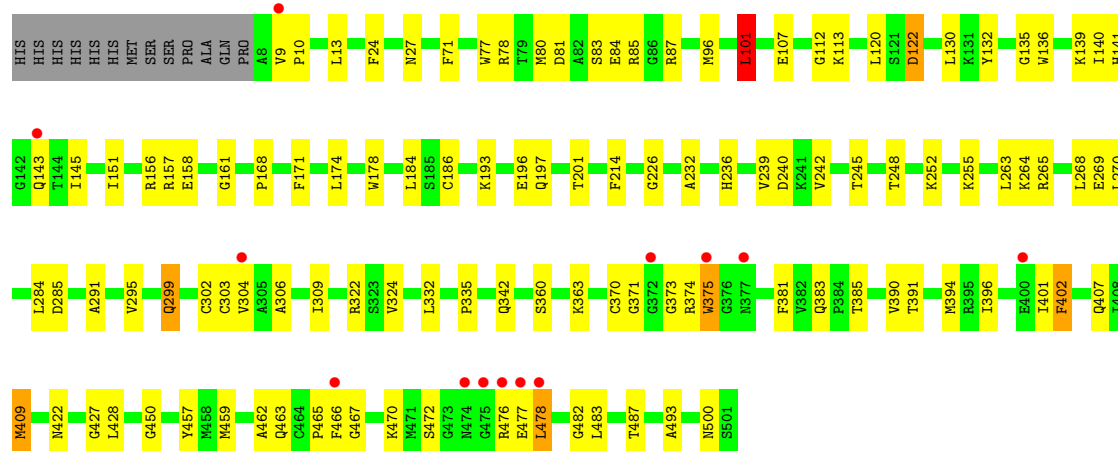
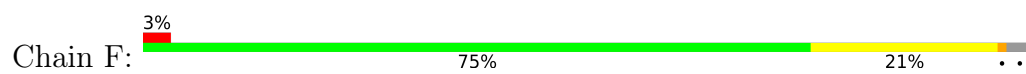
- Molecule 1: Aldehyde dehydrogenase 1A1



- Molecule 1: Aldehyde dehydrogenase 1A1



- Molecule 1: Aldehyde dehydrogenase 1A1



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	104.06Å 156.48Å 129.21Å 90.00° 97.97° 90.00°	Depositor
Resolution (Å)	127.96 – 2.89 127.96 – 2.89	Depositor EDS
% Data completeness (in resolution range)	93.2 (127.96-2.89) 93.3 (127.96-2.89)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.43 (at 2.91Å)	Xtrriage
Refinement program	PHENIX 1.19_4092	Depositor
R, R_{free}	0.195 , 0.249 0.196 , 0.249	Depositor DCC
R_{free} test set	2000 reflections (2.34%)	wwPDB-VP
Wilson B-factor (Å ²)	69.2	Xtrriage
Anisotropy	0.209	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 50.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.62$, $\langle L^2 \rangle = 0.48$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	29957	wwPDB-VP
Average B, all atoms (Å ²)	76.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.67% 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](#)

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	0.59	3/3840 (0.1%)	0.80	4/5195 (0.1%)
1	B	0.58	0/3825	0.82	2/5179 (0.0%)
1	C	0.58	2/3831 (0.1%)	0.81	5/5187 (0.1%)
1	D	0.52	1/3824 (0.0%)	0.77	3/5179 (0.1%)
1	E	0.54	1/3827 (0.0%)	0.76	6/5181 (0.1%)
1	F	0.57	1/3827 (0.0%)	0.80	5/5183 (0.1%)
1	G	0.51	1/3778 (0.0%)	0.84	10/5125 (0.2%)
1	H	0.59	2/3845 (0.1%)	0.81	6/5201 (0.1%)
All	All	0.56	11/30597 (0.0%)	0.80	41/41430 (0.1%)

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	G	0	1

The worst 5 of 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	303	CYS	CB-SG	11.05	2.01	1.82
1	H	303	CYS	CB-SG	8.86	1.97	1.82
1	H	326	ARG	CG-CD	-7.74	1.32	1.51
1	C	303	CYS	CB-SG	7.59	1.95	1.82
1	D	303	CYS	CB-SG	7.58	1.95	1.82

The worst 5 of 41 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	G	478	LEU	CB-CG-CD1	-13.03	88.84	111.00
1	B	478	LEU	CB-CG-CD1	-12.54	89.68	111.00
1	G	483	LEU	CB-CG-CD1	-10.83	92.59	111.00

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	439	LYS	CD-CE-NZ	-10.75	86.97	111.70
1	G	13	LEU	CB-CG-CD2	-9.70	94.51	111.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	G	57	LYS	Mainchain

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	3760	0	3761	82	0
1	B	3745	0	3726	96	0
1	C	3751	0	3728	101	0
1	D	3744	0	3710	91	0
1	E	3747	0	3730	99	0
1	F	3747	0	3719	100	0
1	G	3698	0	3614	150	0
1	H	3765	0	3767	106	0
All	All	29957	0	29755	781	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 13.

The worst 5 of 781 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:424:THR:O	1:B:470:LYS:NZ	1.82	1.11
1:C:434:THR:HG22	1:C:436:ASP:H	1.22	1.00
1:C:424:THR:HG22	1:C:426:TYR:H	1.26	0.98
1:H:467:GLY:HA3	1:H:476:ARG:HD3	1.51	0.92
1:B:285:ASP:OD1	1:B:322:ARG:NH1	2.03	0.91

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	492/507 (97%)	484 (98%)	8 (2%)	0	100	100
1	B	492/507 (97%)	480 (98%)	11 (2%)	1 (0%)	47	78
1	C	492/507 (97%)	480 (98%)	12 (2%)	0	100	100
1	D	492/507 (97%)	486 (99%)	6 (1%)	0	100	100
1	E	492/507 (97%)	486 (99%)	6 (1%)	0	100	100
1	F	492/507 (97%)	484 (98%)	7 (1%)	1 (0%)	47	78
1	G	490/507 (97%)	479 (98%)	11 (2%)	0	100	100
1	H	492/507 (97%)	481 (98%)	11 (2%)	0	100	100
All	All	3934/4056 (97%)	3860 (98%)	72 (2%)	2 (0%)	51	82

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	16	LEU
1	F	10	PRO

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	396/413 (96%)	393 (99%)	3 (1%)	81	94
1	B	393/413 (95%)	389 (99%)	4 (1%)	76	92
1	C	394/413 (95%)	386 (98%)	8 (2%)	55	82

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	D	392/413 (95%)	386 (98%)	6 (2%)	65	87
1	E	394/413 (95%)	388 (98%)	6 (2%)	65	87
1	F	393/413 (95%)	386 (98%)	7 (2%)	59	85
1	G	383/413 (93%)	374 (98%)	9 (2%)	50	80
1	H	398/413 (96%)	394 (99%)	4 (1%)	76	92
All	All	3143/3304 (95%)	3096 (98%)	47 (2%)	65	87

5 of 47 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	110	ASN
1	G	402	PHE
1	G	268	LEU
1	G	322	ARG
1	H	303	CYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:

Mol	Chain	Res	Type
1	G	383	GLN
1	H	292	HIS
1	F	500	ASN
1	F	141	HIS
1	D	407	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 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 [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	494/507 (97%)	0.08	5 (1%) 82 82	37, 59, 84, 100	0
1	B	494/507 (97%)	0.14	7 (1%) 75 75	34, 61, 98, 116	0
1	C	494/507 (97%)	0.15	11 (2%) 62 59	46, 71, 96, 114	0
1	D	494/507 (97%)	0.31	26 (5%) 26 22	35, 69, 127, 146	0
1	E	494/507 (97%)	0.28	15 (3%) 50 45	49, 78, 105, 127	0
1	F	494/507 (97%)	0.23	13 (2%) 56 52	54, 77, 99, 113	0
1	G	492/507 (97%)	0.61	48 (9%) 7 5	45, 100, 157, 174	0
1	H	494/507 (97%)	0.20	9 (1%) 68 67	42, 73, 104, 119	0
All	All	3950/4056 (97%)	0.25	134 (3%) 45 40	34, 73, 120, 174	0

The worst 5 of 134 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	466	PHE	6.9
1	E	476	ARG	6.4
1	D	333	GLY	5.5
1	G	377	ASN	5.2
1	G	476	ARG	5.2

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

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.