WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Nov 21, 2023 – 05:24 PM JST

PDB ID	:	7W8A
Title	:	Babesia orientalis lactate dehydrogenase, BoLDH apo
Authors	:	Yu, L.
Deposited on	:	2021-12-07
Resolution	:	2.67 Å(reported)

The following versions of software and data (see references (i)) were used in the production of this report:

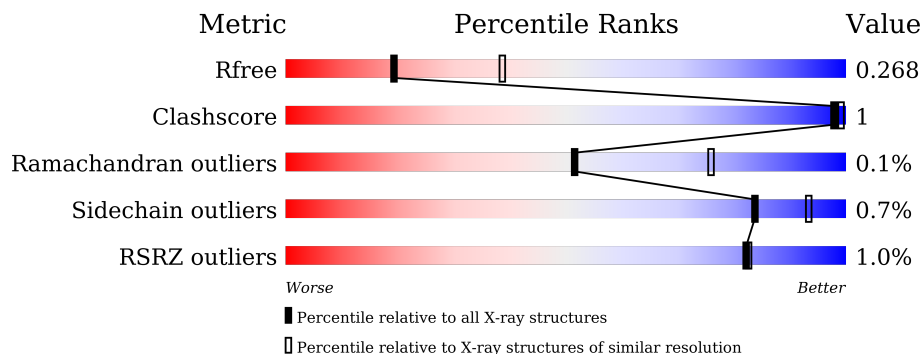|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 2.67 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.
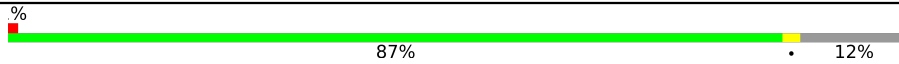
Metric	Percentile Ranks	Value
Rfree		0.268
Clashscore		1
Ramachandran outliers		0.1%
Sidechain outliers		0.7%
RSRZ outliers		1.0%

Worse　　　　　　　　　　　　　　　　Better

▌Percentile relative to all X-ray structures

▯Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	130704	3863 (2.70-2.66)
Clashscore	141614	4210 (2.70-2.66)
Ramachandran outliers	138981	4141 (2.70-2.66)
Sidechain outliers	138945	4141 (2.70-2.66)
RSRZ outliers	127900	3780 (2.70-2.66)

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	A	330	.% 91% · · 6%
1	B	330	.% 87% · 11%
1	C	330	89% · 9%
1	D	330	2% 86% · 12%
1	E	330	.% 84% 5% 11%
1	F	330	87% · 12%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	330	% 87%　·　12%
1	H	330	89%　·　9%

2 Entry composition ⓘ

There is only 1 type of molecule in this entry. The entry contains 35845 atoms, of which 18104 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 lactate dehydrogenase.

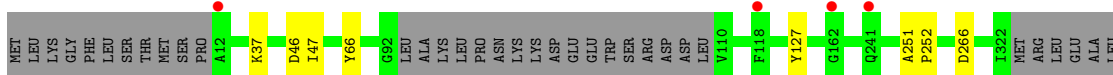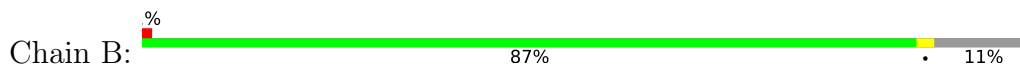| Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	311	Total	C	H	N	O	S	3	0	0
			4719	1483	2382	393	445	16			
1	B	294	Total	C	H	N	O	S	0	0	0
			4433	1393	2240	368	416	16			
1	C	301	Total	C	H	N	O	S	3	0	0
			4531	1425	2288	375	426	17			
1	D	292	Total	C	H	N	O	S	0	0	0
			4410	1386	2228	366	414	16			
1	E	293	Total	C	H	N	O	S	4	0	0
			4416	1388	2230	367	415	16			
1	F	292	Total	C	H	N	O	S	0	0	0
			4402	1383	2223	366	414	16			
1	G	292	Total	C	H	N	O	S	0	1	0
			4411	1386	2229	366	414	16			
1	H	300	Total	C	H	N	O	S	0	0	0
			4523	1423	2284	374	425	17			

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.

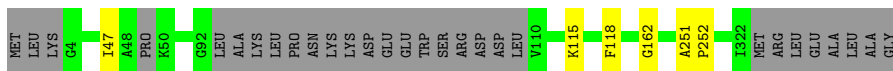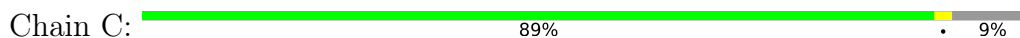- Molecule 1: lactate dehydrogenase

Chain A:

- Molecule 1: lactate dehydrogenase

Chain B:

- Molecule 1: lactate dehydrogenase

Chain C:

- Molecule 1: lactate dehydrogenase

Chain D:

- Molecule 1: lactate dehydrogenase

Chain E:

MET LEU LYS GLY PHE LEU SER THR MET SER PRO A12 K17 D46 I47 A48 P49 Y66 A82 D83 G92 LEU ALA LYS LEU PRO ASN LYS LYS ASP GLU GLU TRP SER ARG ASP ASP LEU VAL A111 P112 F118 V159 N179 N245 A251 P252 Y263

I299 I302 K306 D316 I322 MET ARG LEU GLU ALA LEU ALA GLY

• Molecule 1: lactate dehydrogenase

Chain F:
87%   •   12%

MET LEU LYS GLY PHE LEU SER THR MET SER PRO A12 K17 A48 PRO K50 Y66 G92 LEU ALA LYS LEU PRO ASN LYS LYS ASP GLU GLU TRP SER ARG ASP ASP LEU VAL A111 E216 R217 I322 MET ARG LEU GLU ALA LEU ALA GLY
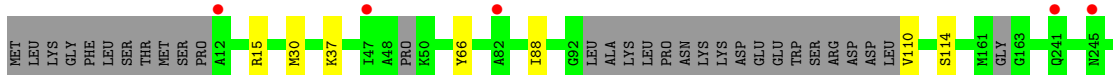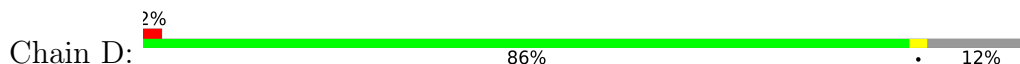
• Molecule 1: lactate dehydrogenase

Chain G:
%
87%   •   12%

MET LEU LYS GLY PHE LEU SER THR MET SER PRO A12 V13 Y66 E79 G92 LEU ALA LYS LEU PRO ASN LYS LYS ASP GLU GLU TRP SER ARG ASP ASP LEU VAL A111 F118 L140 V144 M161 GLY G163 Q241 T290 K306 I322 MET ARG LEU GLU
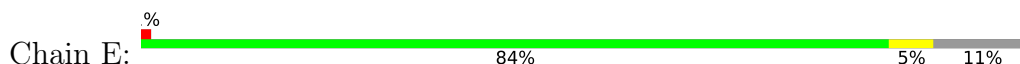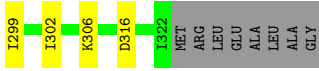
ALA LEU ALA GLY

• Molecule 1: lactate dehydrogenase

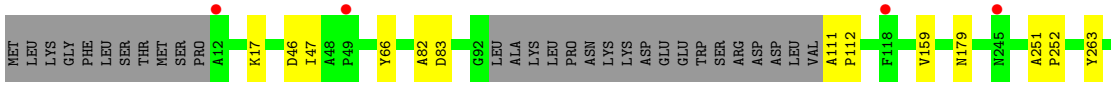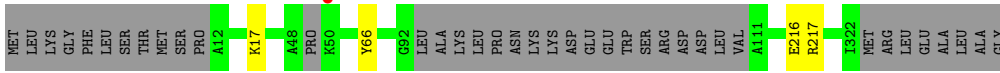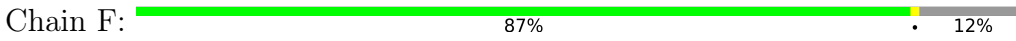Chain H:
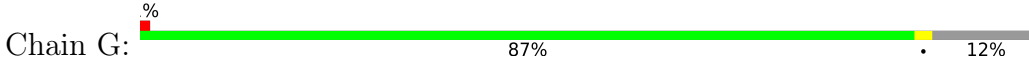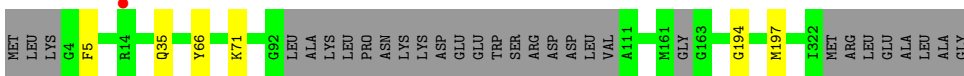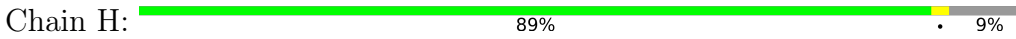89%   •   9%

MET LEU LYS G4 F5 R14 Q35 Y66 K71 G92 LEU ALA LYS LEU PRO ASN LYS LYS ASP GLU GLU TRP SER ARG ASP ASP LEU VAL A111 M161 GLY G163 G194 M197 I322 MET ARG LEU GLU ALA LEU ALA GLY

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	81.85Å 92.30Å 109.89Å 77.47° 70.69° 64.36°	Depositor
Resolution (Å)	45.00 – 2.67 45.00 – 2.67	Depositor EDS
% Data completeness (in resolution range)	93.5 (45.00-2.67) 93.5 (45.00-2.67)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.81 (at 2.69Å)	Xtriage
Refinement program	PHENIX 1.12_2829	Depositor
R, R_{free}	0.236 , 0.268 0.236 , 0.268	Depositor DCC
R_{free} test set	3682 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å^2)	27.3	Xtriage
Anisotropy	0.394	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.42 , 19.9	EDS
L-test for twinning[2]	$< |L| > = 0.48, < L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.011 for h,h-k,h-l	Xtriage
F_o,F_c correlation	0.87	EDS
Total number of atoms	35845	wwPDB-VP
Average B, all atoms (Å^2)	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.08% 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	A	0.25	0/2374	0.42	0/3212
1	B	0.24	0/2226	0.41	0/3011
1	C	0.24	0/2276	0.40	0/3076
1	D	0.25	0/2212	0.41	0/2988
1	E	0.25	0/2219	0.40	0/3001
1	F	0.24	0/2210	0.41	0/2986
1	G	0.24	0/2223	0.40	0/3004
1	H	0.24	0/2273	0.41	0/3073
All	All	0.24	0/18013	0.41	0/24351

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	A	2337	2382	2381	6	0
1	B	2193	2240	2238	3	1
1	C	2243	2288	2286	3	0
1	D	2182	2228	2226	4	0
1	E	2186	2230	2229	8	1
1	F	2179	2223	2221	2	0
1	G	2182	2229	2212	2	0
1	H	2239	2284	2281	3	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	17741	18104	18074	25	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 1.

The worst 5 of 25 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:E:46:ASP:OD1	1:E:47:ILE:N	2.39	0.56
1:E:306:LYS:NZ	1:F:216:GLU:OE1	2.41	0.54
1:D:30:MET:HE2	1:D:88:ILE:HG21	1.92	0.51
1:A:135:CYS:HB3	1:A:144:VAL:HG22	1.97	0.46
1:A:101:ASP:O	1:A:102:GLU:HB2	2.16	0.45

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:B:127:TYR:OH	1:E:83:ASP:OD2[1_456]	2.07	0.13

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	309/330 (94%)	297 (96%)	11 (4%)	1 (0%)	41	64
1	B	290/330 (88%)	280 (97%)	10 (3%)	0	100	100
1	C	295/330 (89%)	285 (97%)	9 (3%)	1 (0%)	41	64
1	D	284/330 (86%)	269 (95%)	15 (5%)	0	100	100
1	E	289/330 (88%)	277 (96%)	12 (4%)	0	100	100

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	F	286/330 (87%)	273 (96%)	13 (4%)	0	100	100
1	G	287/330 (87%)	271 (94%)	16 (6%)	0	100	100
1	H	294/330 (89%)	284 (97%)	10 (3%)	0	100	100
All	All	2334/2640 (88%)	2236 (96%)	96 (4%)	2 (0%)	51	76

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	102	GLU
1	C	162	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	250/265 (94%)	248 (99%)	2 (1%)	81	92
1	B	234/265 (88%)	232 (99%)	2 (1%)	78	91
1	C	240/265 (91%)	240 (100%)	0	100	100
1	D	233/265 (88%)	232 (100%)	1 (0%)	91	96
1	E	233/265 (88%)	230 (99%)	3 (1%)	69	86
1	F	232/265 (88%)	230 (99%)	2 (1%)	78	91
1	G	234/265 (88%)	231 (99%)	3 (1%)	69	86
1	H	240/265 (91%)	239 (100%)	1 (0%)	91	96
All	All	1896/2120 (89%)	1882 (99%)	14 (1%)	84	93

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	316	ASP
1	F	17	LYS
1	H	66	TYR
1	G	118	PHE

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	G	290	THR

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

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

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, 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	311/330 (94%)	-0.15	4 (1%) 77	78	21, 25, 32, 46	0
1	B	294/330 (89%)	-0.16	4 (1%) 75	76	21, 25, 29, 38	0
1	C	301/330 (91%)	-0.16	0 100	100	20, 25, 30, 37	0
1	D	292/330 (88%)	-0.10	5 (1%) 70	71	21, 26, 32, 38	0
1	E	293/330 (88%)	-0.06	4 (1%) 75	76	22, 27, 31, 34	0
1	F	292/330 (88%)	-0.07	1 (0%) 94	95	22, 27, 32, 41	0
1	G	292/330 (88%)	-0.10	4 (1%) 75	76	22, 26, 31, 41	0
1	H	300/330 (90%)	-0.10	1 (0%) 94	95	22, 25, 29, 38	0
All	All	2375/2640 (89%)	-0.11	23 (0%) 82	82	20, 26, 31, 46	0

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	12	ALA	4.6
1	E	118	PHE	3.7
1	A	101	ASP	3.7
1	D	241	GLN	3.2
1	B	12	ALA	3.2

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 ⓘ

There are no monosaccharides in this entry.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.