wwPDB X-ray Structure Validation Summary Report (i)

Aug 27, 2023 – 05:10 PM EDT

PDB ID	:	3IMF
Title	:	1.99 Angstrom resolution crystal structure of a short chain dehydrogenase from Bacillus anthracis str. 'Ames Ancestor'
Authors	:	Halavaty, A.S.; Minasov, G.; Skarina, T.; Onopriyenko, O.; Gordon, E.; Peterson, S.; Savchenko, A.; Anderson, W.F.; Center for Structural Genomics of Infectious Diseases (CSGID)
Deposited on	:	2009-08-10
Resolution	:	1.99 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 (i) 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 (i)) were used in the production of this report:

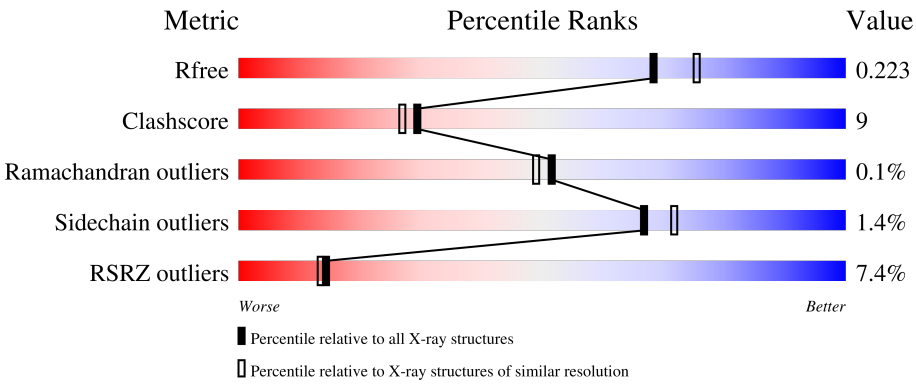
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.99 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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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	257	7%, 80%, 16%, • •
1	B	257	7%, 83%, 14%, •
1	C	257	7%, 82%, 11%, 6%
1	D	257	6%, 81%, 14%, • •

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	ACT	D	255	-	-	-	X

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 8338 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.

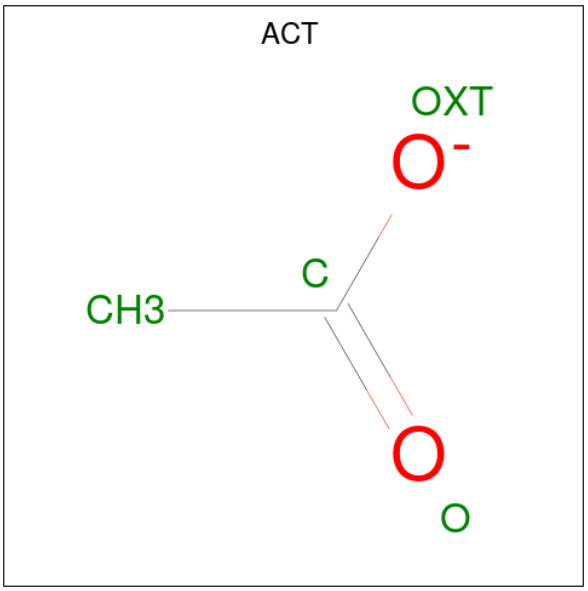
- Molecule 1 is a protein called Short chain dehydrogenase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	248	1957	1229	342	370	4	12	0	9	0
1	B	249	1950	1225	339	370	4	12	0	8	0
1	C	241	1898	1198	331	354	4	11	0	8	0
1	D	248	1927	1212	337	362	4	12	0	6	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	SER	-	expression tag	UNP Q81MP2
A	-1	ASN	-	expression tag	UNP Q81MP2
A	0	ALA	-	expression tag	UNP Q81MP2
B	-2	SER	-	expression tag	UNP Q81MP2
B	-1	ASN	-	expression tag	UNP Q81MP2
B	0	ALA	-	expression tag	UNP Q81MP2
C	-2	SER	-	expression tag	UNP Q81MP2
C	-1	ASN	-	expression tag	UNP Q81MP2
C	0	ALA	-	expression tag	UNP Q81MP2
D	-2	SER	-	expression tag	UNP Q81MP2
D	-1	ASN	-	expression tag	UNP Q81MP2
D	0	ALA	-	expression tag	UNP Q81MP2

- Molecule 2 is ACETATE ION (three-letter code: ACT) (formula: $C_2H_3O_2$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
2	B	1	4	2	2	0	0
			Total	C	O		
2	D	1	4	2	2	0	0

- Molecule 3 is water.

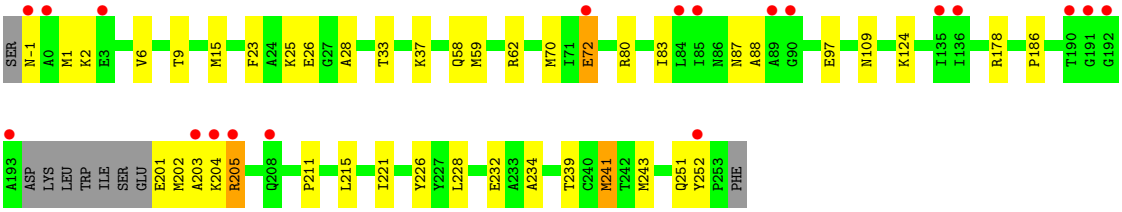
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	146	148	148	0	5
			Total	O		
3	B	152	153	153	0	6
			Total	O		
3	C	155	157	157	0	8
			Total	O		
3	D	139	140	140	0	4

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.

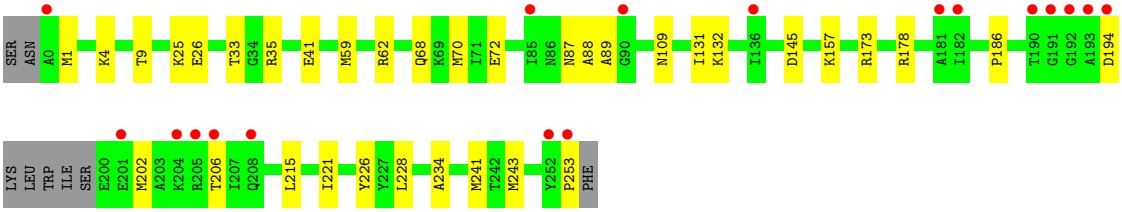
• Molecule 1: Short chain dehydrogenase

Chain A:



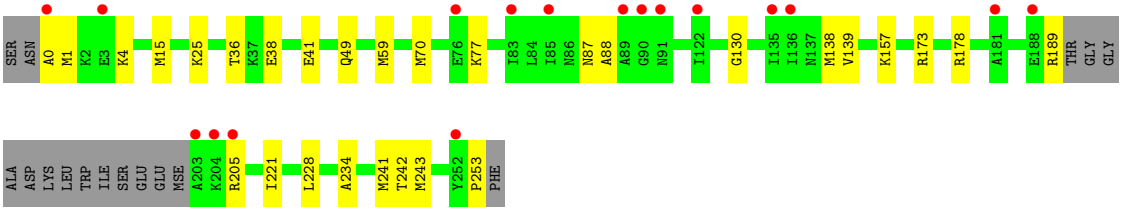
• Molecule 1: Short chain dehydrogenase

Chain B:



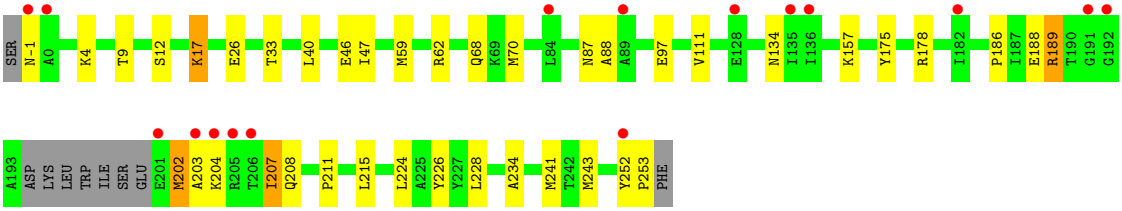
• Molecule 1: Short chain dehydrogenase

Chain C:



• Molecule 1: Short chain dehydrogenase

Chain D:


SER
N-1
A0
K4
T9
S12
K17
E26
T33
L40
E46
I47
M59
R62
Q68
K69
M70
L84
N87
A88
A89
E97
V111
E128
N134
I135
I136
K157
Y175
R178
I182
P186
I187
E188
R189
T190
G191
G192
A193
ASP
LYS
LEU
TRP
ILE
SER
GLU
E201
M202
A203
K204
R205
T206
I207
Q208
P211
L215
L224
A225
Y226
Y227
L228
A234
M241
T242
M243
Y252
P253
PHE

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	174.24Å 89.56Å 71.78Å 90.00° 111.60° 90.00°	Depositor
Resolution (Å)	30.00 – 1.99 29.36 – 1.99	Depositor EDS
% Data completeness (in resolution range)	98.8 (30.00-1.99) 98.9 (29.36-1.99)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.21 (at 2.00Å)	Xtriage
Refinement program	REFMAC 5.5.0102	Depositor
R, R_{free}	0.166 , 0.206 0.189 , 0.223	Depositor DCC
R_{free} test set	3494 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	26.4	Xtriage
Anisotropy	0.313	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 53.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.088 for -h-2*l,-k,l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	8338	wwPDB-VP
Average B, all atoms (Å^2)	22.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 67.44 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 5.1468e-06. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: ACT

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	0.52	0/1977	0.66	0/2647
1	B	0.58	0/1970	0.70	1/2639 (0.0%)
1	C	0.55	0/1919	0.70	0/2572
1	D	0.53	0/1947	0.68	0/2608
All	All	0.55	0/7813	0.68	1/10466 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	145	ASP	CB-CG-OD2	-5.10	113.71	118.30

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	1957	0	1963	51	0
1	B	1950	0	1955	48	0
1	C	1898	0	1921	35	0
1	D	1927	0	1941	48	0
2	B	4	0	3	0	0
2	D	4	0	3	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	148	0	0	4	0
3	B	153	0	0	13	0
3	C	157	0	0	9	0
3	D	140	0	0	4	0
All	All	8338	0	7786	148	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 9.

The worst 5 of 148 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:241[B]:MSE:HE3	1:B:241[B]:MSE:CE	1.76	1.15
1:A:241[B]:MSE:CE	1:B:241[B]:MSE:HE2	1.82	1.08
1:A:241[A]:MSE:CE	1:B:241[A]:MSE:HE3	1.91	1.00
1:C:241[A]:MSE:HE1	1:D:241[A]:MSE:HB3	1.44	0.96
1:A:241[A]:MSE:SE	1:B:241[A]:MSE:HE1	2.20	0.92

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	253/257 (98%)	246 (97%)	6 (2%)	1 (0%)	34 30
1	B	253/257 (98%)	245 (97%)	8 (3%)	0	100 100
1	C	245/257 (95%)	238 (97%)	7 (3%)	0	100 100
1	D	250/257 (97%)	244 (98%)	6 (2%)	0	100 100
All	All	1001/1028 (97%)	973 (97%)	27 (3%)	1 (0%)	51 49

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	204	LYS

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	202/192 (105%)	198 (98%)	4 (2%)	55	58
1	B	202/192 (105%)	201 (100%)	1 (0%)	88	92
1	C	197/192 (103%)	194 (98%)	3 (2%)	65	69
1	D	199/192 (104%)	195 (98%)	4 (2%)	55	58
All	All	800/768 (104%)	788 (98%)	12 (2%)	67	69

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	242	THR
1	D	17	LYS
1	D	207	ILE
1	D	189	ARG
1	A	241[B]	MSE

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)

2 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	ACT	B	255	-	3,3,3	0.79	0	3,3,3	1.38	0
2	ACT	D	255	-	3,3,3	0.77	0	3,3,3	1.48	0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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	238/257 (92%)	0.41	19 (7%)	12	11	11, 21, 32, 41	0
1	B	239/257 (92%)	0.37	18 (7%)	14	13	2, 20, 30, 45	0
1	C	232/257 (90%)	0.26	17 (7%)	15	14	5, 20, 32, 51	0
1	D	238/257 (92%)	0.37	16 (6%)	17	17	6, 20, 32, 45	0
All	All	947/1028 (92%)	0.35	70 (7%)	14	13	2, 20, 32, 51	0

The worst 5 of 70 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	203	ALA	8.8
1	D	252	TYR	6.7
1	A	191	GLY	6.1
1	D	191	GLY	5.7
1	A	252	TYR	5.4

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)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	ACT	D	255	4/4	0.48	0.42	76,77,77,77	0
2	ACT	B	255	4/4	0.87	0.51	53,53,53,53	4

6.5 Other polymers (i)

There are no such residues in this entry.