wwPDB X-ray Structure Validation Summary Report (i)

Jan 30, 2021 – 02:29 PM EST

PDB ID	:	3DHU
Title	:	Crystal structure of an alpha-amylase from Lactobacillus plantarum
Authors	:	Bonanno, J.B.; Dickey, M.; Bain, K.T.; Iizuka, M.; Ozyurt, S.; Smith, D.; Wasserman, S.; Sauder, J.M.; Burley, S.K.; Almo, S.C.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on	:	2008-06-18
Resolution	:	2.00 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 following versions of software and data (see references (1)) were used in the production of this report:

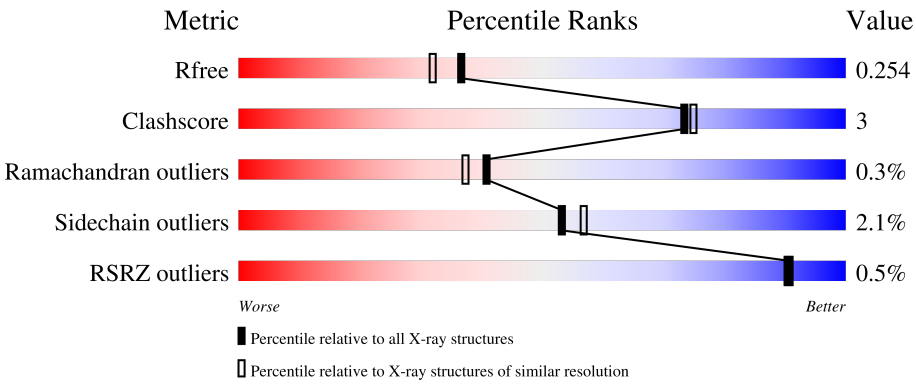
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.16
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.16

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 2.00 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	449	83% 10% 6%
1	B	449	86% 7% 7%
1	C	449	% 86% 7% • 6%
1	D	449	% 85% 7% • 7%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 14732 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 Alpha-amylase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	422	3361	2155	563	632	11	0	0	0
1	B	418	3338	2142	558	627	11	0	0	0
1	C	422	3364	2156	563	634	11	0	0	0
1	D	418	3338	2142	558	627	11	0	0	0

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	MET	-	insertion	UNP Q88ZW5
A	1	SER	-	insertion	UNP Q88ZW5
A	2	LEU	-	insertion	UNP Q88ZW5
A	441	GLU	-	insertion	UNP Q88ZW5
A	442	GLY	-	insertion	UNP Q88ZW5
A	443	HIS	-	insertion	UNP Q88ZW5
A	444	HIS	-	insertion	UNP Q88ZW5
A	445	HIS	-	insertion	UNP Q88ZW5
A	446	HIS	-	insertion	UNP Q88ZW5
A	447	HIS	-	insertion	UNP Q88ZW5
A	448	HIS	-	insertion	UNP Q88ZW5
B	0	MET	-	insertion	UNP Q88ZW5
B	1	SER	-	insertion	UNP Q88ZW5
B	2	LEU	-	insertion	UNP Q88ZW5
B	441	GLU	-	insertion	UNP Q88ZW5
B	442	GLY	-	insertion	UNP Q88ZW5
B	443	HIS	-	insertion	UNP Q88ZW5
B	444	HIS	-	insertion	UNP Q88ZW5
B	445	HIS	-	insertion	UNP Q88ZW5
B	446	HIS	-	insertion	UNP Q88ZW5
B	447	HIS	-	insertion	UNP Q88ZW5

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	448	HIS	-	insertion	UNP Q88ZW5
C	0	MET	-	insertion	UNP Q88ZW5
C	1	SER	-	insertion	UNP Q88ZW5
C	2	LEU	-	insertion	UNP Q88ZW5
C	441	GLU	-	insertion	UNP Q88ZW5
C	442	GLY	-	insertion	UNP Q88ZW5
C	443	HIS	-	insertion	UNP Q88ZW5
C	444	HIS	-	insertion	UNP Q88ZW5
C	445	HIS	-	insertion	UNP Q88ZW5
C	446	HIS	-	insertion	UNP Q88ZW5
C	447	HIS	-	insertion	UNP Q88ZW5
C	448	HIS	-	insertion	UNP Q88ZW5
D	0	MET	-	insertion	UNP Q88ZW5
D	1	SER	-	insertion	UNP Q88ZW5
D	2	LEU	-	insertion	UNP Q88ZW5
D	441	GLU	-	insertion	UNP Q88ZW5
D	442	GLY	-	insertion	UNP Q88ZW5
D	443	HIS	-	insertion	UNP Q88ZW5
D	444	HIS	-	insertion	UNP Q88ZW5
D	445	HIS	-	insertion	UNP Q88ZW5
D	446	HIS	-	insertion	UNP Q88ZW5
D	447	HIS	-	insertion	UNP Q88ZW5
D	448	HIS	-	insertion	UNP Q88ZW5

- Molecule 2 is water.

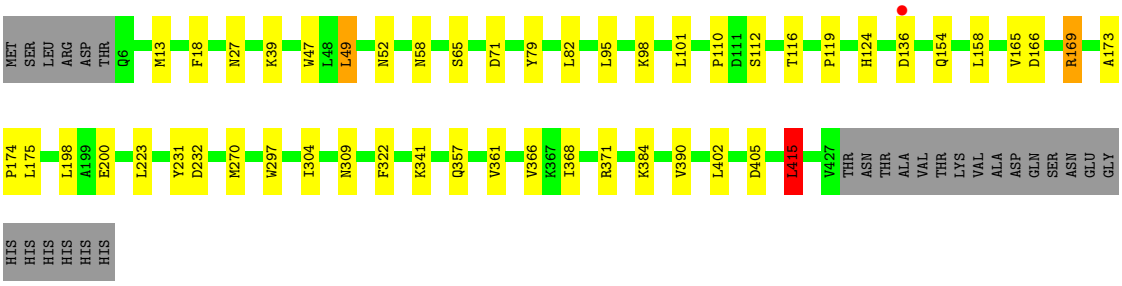
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	340	340	340	0	0
			Total	O		
2	B	314	314	314	0	0
			Total	O		
2	C	357	357	357	0	0
			Total	O		
2	D	320	320	320	0	0

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.

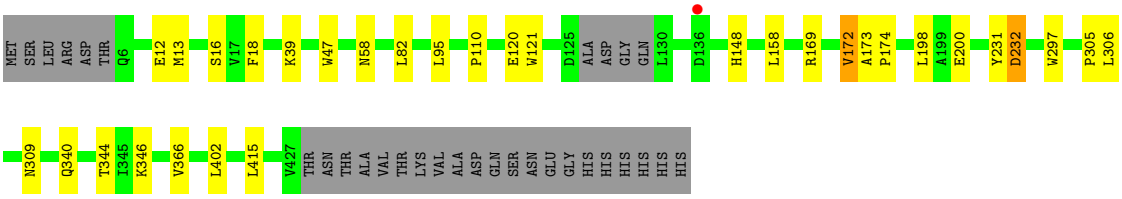
• Molecule 1: Alpha-amylase

Chain A:



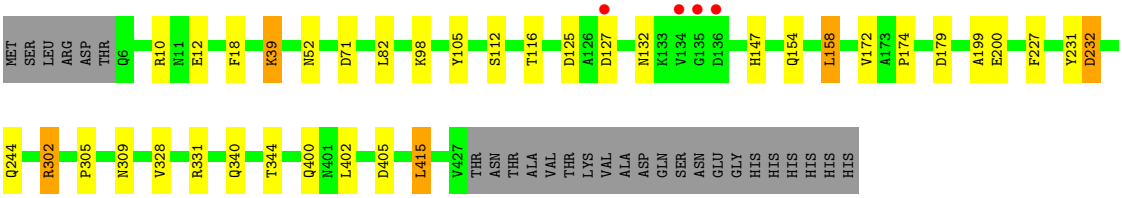
• Molecule 1: Alpha-amylase

Chain B:



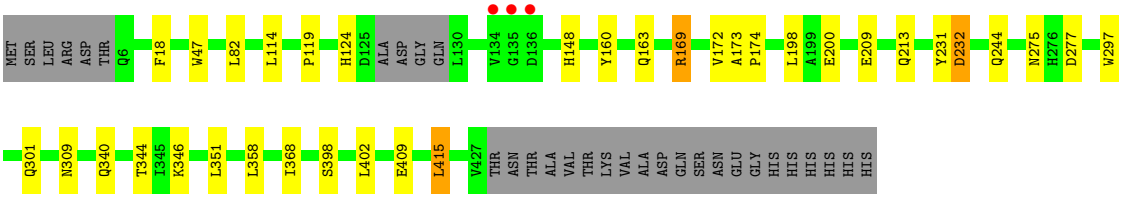
• Molecule 1: Alpha-amylase

Chain C:



• Molecule 1: Alpha-amylase

Chain D:


MET
SER
LEU
ARG
ASP
THR
Q6
F18
W47
L82
L114
P119
H124
D125
ALA
ASP
GLY
GLN
L130
V134
G135
D136
H148
Y160
Q163
R169
V172
A173
P174
L198
A199
E200
E209
Q213
Y231
D232
Q244
N275
H276
D277
W297
Q301
N309
Q340
T344
I345
K346
L351
L358
I368
S398
L402
E409
L415
V427
THR
ASN
THR
ALA
VAL
THR
LYS
VAL
ALA
ASP
GLN
SER
ASN
GLU
GLY
HIS
HIS
HIS
HIS
HIS
HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	202.28Å 63.77Å 155.68Å 90.00° 102.93° 90.00°	Depositor
Resolution (Å)	20.00 – 2.00 24.41 – 2.00	Depositor EDS
% Data completeness (in resolution range)	98.6 (20.00-2.00) 98.6 (24.41-2.00)	Depositor EDS
R_{merge}	0.18	Depositor
R_{sym}	0.18	Depositor
$< I/\sigma(I) >$ [1]	3.10 (at 1.99Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.200 , 0.251 0.205 , 0.254	Depositor DCC
R_{free} test set	6500 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	12.4	Xtriage
Anisotropy	0.489	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 44.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	14732	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 3.71% 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 (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	0.65	0/3446	0.73	4/4695 (0.1%)
1	B	0.66	0/3422	0.71	2/4661 (0.0%)
1	C	0.68	0/3449	0.72	2/4699 (0.0%)
1	D	0.64	0/3422	0.71	1/4661 (0.0%)
All	All	0.66	0/13739	0.72	9/18716 (0.0%)

There are no bond length outliers.

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	415	LEU	CA-CB-CG	8.58	135.03	115.30
1	A	169	ARG	NE-CZ-NH1	8.33	124.47	120.30
1	D	415	LEU	CA-CB-CG	7.83	133.31	115.30
1	C	415	LEU	CA-CB-CG	7.09	131.60	115.30
1	B	415	LEU	CA-CB-CG	6.65	130.59	115.30

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	3361	0	3243	31	0
1	B	3338	0	3224	17	0
1	C	3364	0	3245	21	0
1	D	3338	0	3224	19	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	340	0	0	4	0
2	B	314	0	0	1	0
2	C	357	0	0	6	0
2	D	320	0	0	2	0
All	All	14732	0	12936	88	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 3.

The worst 5 of 88 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:244:GLN:NE2	2:D:552:HOH:O	2.04	0.91
1:D:47:TRP:CZ2	1:D:169:ARG:HD3	2.07	0.90
1:B:169:ARG:HG3	1:B:198:LEU:HD23	1.57	0.86
1:A:47:TRP:CZ2	1:A:169:ARG:HD3	2.13	0.83
1:C:112:SER:O	1:C:116:THR:HG23	1.78	0.81

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	420/449 (94%)	407 (97%)	12 (3%)	1 (0%)	47	44
1	B	414/449 (92%)	402 (97%)	10 (2%)	2 (0%)	29	23
1	C	420/449 (94%)	406 (97%)	13 (3%)	1 (0%)	47	44
1	D	414/449 (92%)	401 (97%)	12 (3%)	1 (0%)	47	44
All	All	1668/1796 (93%)	1616 (97%)	47 (3%)	5 (0%)	41	37

All (5) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	232	ASP
1	D	232	ASP
1	C	232	ASP
1	A	361	VAL
1	B	172	VAL

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	352/384 (92%)	343 (97%)	9 (3%)	46	48
1	B	351/384 (91%)	345 (98%)	6 (2%)	60	65
1	C	353/384 (92%)	345 (98%)	8 (2%)	50	53
1	D	351/384 (91%)	344 (98%)	7 (2%)	55	58
All	All	1407/1536 (92%)	1377 (98%)	30 (2%)	53	57

5 of 30 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	309	ASN
1	C	39	LYS
1	D	309	ASN
1	C	18	PHE
1	C	82	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	118	HIS
1	C	124	HIS
1	D	124	HIS
1	B	244	GLN
1	D	132	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 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	422/449 (93%)	-0.38	1 (0%)	95	94	7, 15, 27, 36	0
1	B	418/449 (93%)	-0.37	1 (0%)	95	94	8, 15, 27, 33	0
1	C	422/449 (93%)	-0.36	4 (0%)	84	83	7, 14, 26, 41	0
1	D	418/449 (93%)	-0.30	3 (0%)	87	87	8, 16, 27, 48	0
All	All	1680/1796 (93%)	-0.35	9 (0%)	91	90	7, 15, 27, 48	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	134	VAL	5.1
1	C	134	VAL	4.6
1	D	136	ASP	4.5
1	D	135	GLY	4.3
1	C	136	ASP	3.7

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.