wwPDB X-ray Structure Validation Summary Report (i)

May 28, 2020 – 07:51 pm BST

PDB ID	:	1GYH
Title	:	Structure of D158A Cellvibrio cellulosa alpha-L-arabinanase mutant
Authors	:	Nurizzo, D.; Turkenburg, J.P.; Charnock, S.J.; Roberts, S.M.; Dodson, E.J.; McKie, V.A.; Taylor, E.J.; Gilbert, H.J.; Davies, G.J.
Deposited on	:	2002-04-23
Resolution	:	1.89 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

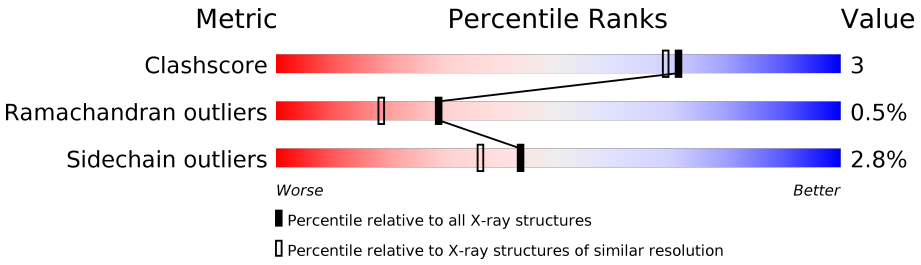
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 1.89 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))
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)

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 on 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%

Mol	Chain	Length	Quality of chain
1	A	318	92% 7% •
1	B	318	92% 6% ••
1	C	318	94% • •
1	D	318	91% 8% ••
1	E	318	88% 10% ••
1	F	318	86% 12% ••

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	CL	A	1348	-	-	X	-
2	CL	B	1348	-	-	X	-
2	CL	C	1348	-	-	X	-
2	CL	D	1347	-	-	X	-
2	CL	E	1348	-	-	X	-
2	CL	F	1347	-	-	X	-

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 17338 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 ARABINAN ENDO-1,5-ALPHA-L-ARABINOSIDASE A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	318	Total	C	N	O	S	0	8	0
			2589	1654	452	477	6			
1	B	317	Total	C	N	O	S	0	11	0
			2592	1658	450	478	6			
1	C	318	Total	C	N	O	S	0	7	0
			2576	1646	444	480	6			
1	D	316	Total	C	N	O	S	0	2	0
			2548	1634	436	471	7			
1	E	318	Total	C	N	O	S	0	3	0
			2570	1647	441	476	6			
1	F	316	Total	C	N	O	S	0	2	0
			2547	1632	436	473	6			

- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	D	1	Total	Cl	0	0
			1	1		
2	E	1	Total	Cl	0	0
			1	1		
2	B	1	Total	Cl	0	0
			1	1		
2	C	1	Total	Cl	0	0
			1	1		
2	A	1	Total	Cl	0	0
			1	1		
2	F	1	Total	Cl	0	0
			1	1		

- Molecule 3 is water.

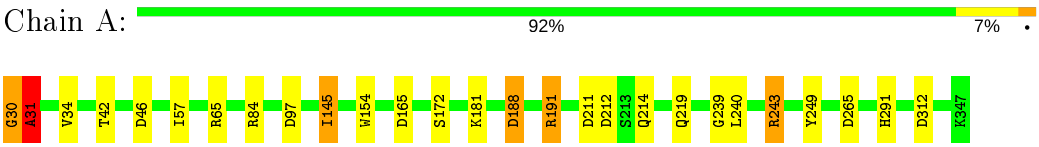
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	447	Total O 447 447	0	0
3	B	356	Total O 356 356	0	0
3	C	371	Total O 371 371	0	0
3	D	277	Total O 277 277	0	0
3	E	291	Total O 291 291	0	0
3	F	168	Total O 168 168	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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. 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. 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.

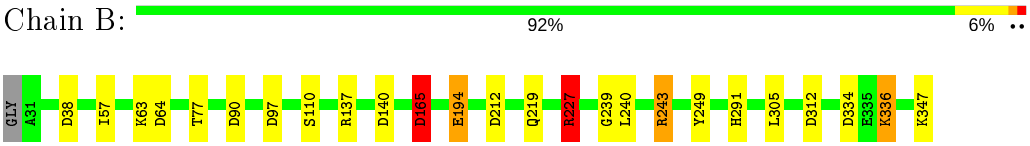
- Molecule 1: ARABINAN ENDO-1,5-ALPHA-L-ARABINOSIDASE A

Chain A:



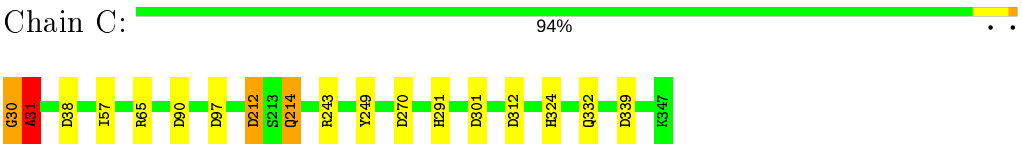
- Molecule 1: ARABINAN ENDO-1,5-ALPHA-L-ARABINOSIDASE A

Chain B:



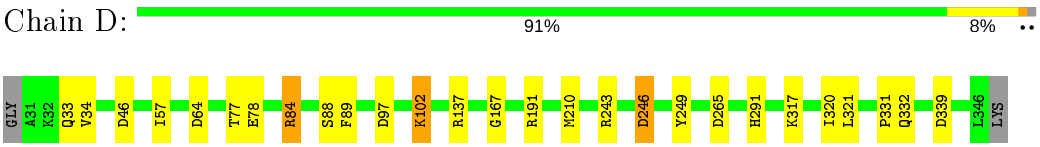
- Molecule 1: ARABINAN ENDO-1,5-ALPHA-L-ARABINOSIDASE A

Chain C:



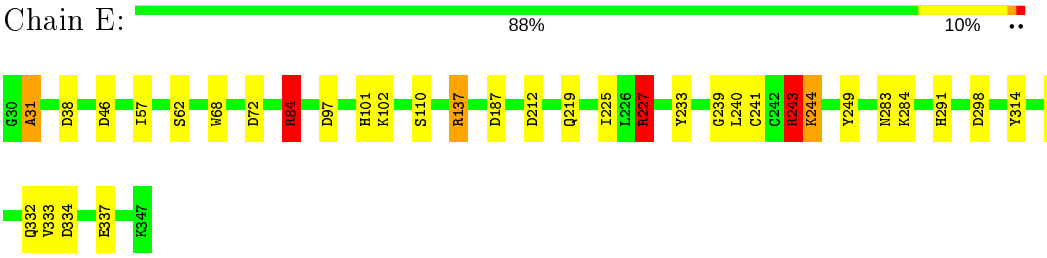
- Molecule 1: ARABINAN ENDO-1,5-ALPHA-L-ARABINOSIDASE A

Chain D:



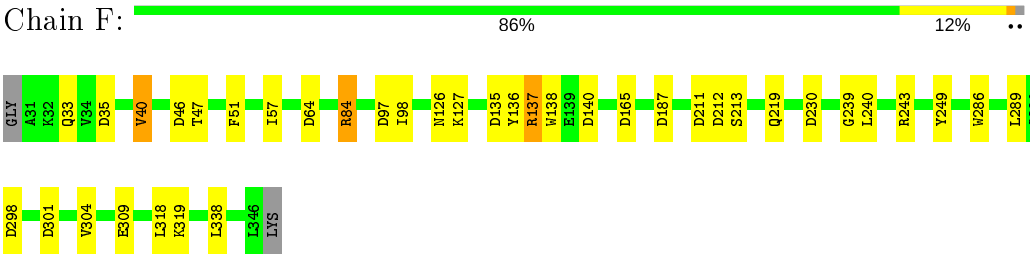
- Molecule 1: ARABINAN ENDO-1,5-ALPHA-L-ARABINOSIDASE A

Chain E:



- Molecule 1: ARABINAN ENDO-1,5-ALPHA-L-ARABINOSIDASE A

Chain F:


86%
12%
GLY
A31
K32
Q33
V34
D35
V40
D46
T47
F51
I57
D64
R84
D97
I98
N126
K127
D135
Y136
R137
W138
E139
D140
D165
D187
D211
D212
S213
Q219
D230
G239
L240
R243
Y249
W286
L289
G290
H291
N292
S293
D298
D301
V304
E309
I318
K319
I338
L346
LYS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	49.56Å 79.56Å 132.40Å 88.83° 89.58° 83.58°	Depositor
Resolution (Å)	20.00 – 1.89 19.85 – 1.89	Depositor EDS
% Data completeness (in resolution range)	96.6 (20.00-1.89) 96.6 (19.85-1.89)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.65 (at 1.89Å)	Xtriage
Refinement program	REFMAC 5.1.13	Depositor
R, R_{free}	0.147 , 0.203 0.226 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	17.8	Xtriage
Anisotropy	0.349	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 70.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.007 for -h,-k,l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	17338	wwPDB-VP
Average B, all atoms (Å^2)	18.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.32% 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

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

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	0.86	2/2712 (0.1%)	0.98	14/3679 (0.4%)
1	B	0.77	0/2729	0.94	15/3702 (0.4%)
1	C	0.79	2/2692 (0.1%)	0.91	12/3656 (0.3%)
1	D	0.73	1/2636 (0.0%)	0.88	9/3584 (0.3%)
1	E	0.74	1/2664 (0.0%)	0.92	14/3619 (0.4%)
1	F	0.63	0/2635	0.86	12/3582 (0.3%)
All	All	0.76	6/16068 (0.0%)	0.92	76/21822 (0.3%)

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	30	GLY	N-CA	-8.36	1.33	1.46
1	A	30	GLY	N-CA	-7.63	1.34	1.46
1	A	31	ALA	CA-CB	-7.44	1.36	1.52
1	D	191	ARG	CG-CD	-6.38	1.36	1.51
1	C	31	ALA	CA-CB	-5.67	1.40	1.52

The worst 5 of 76 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	188[A]	ASP	CB-CG-OD2	12.28	129.35	118.30
1	A	188[B]	ASP	CB-CG-OD2	12.28	129.35	118.30
1	B	165[A]	ASP	CB-CG-OD2	9.41	126.77	118.30
1	B	165[B]	ASP	CB-CG-OD2	9.41	126.77	118.30
1	D	84	ARG	NE-CZ-NH1	8.30	124.45	120.30

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	2589	0	2485	12	0
1	B	2592	0	2495	16	0
1	C	2576	0	2462	8	0
1	D	2548	0	2441	11	0
1	E	2570	0	2462	17	0
1	F	2547	0	2438	16	0
2	A	1	0	0	6	0
2	B	1	0	0	4	0
2	C	1	0	0	5	0
2	D	1	0	0	6	0
2	E	1	0	0	5	0
2	F	1	0	0	5	0
3	A	447	0	0	8	0
3	B	356	0	0	7	0
3	C	371	0	0	7	0
3	D	277	0	0	7	0
3	E	291	0	0	8	0
3	F	168	0	0	4	0
All	All	17338	0	14783	105	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 105 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:165[A]:ASP:OD1	1:B:227[A]:ARG:CZ	1.69	1.39
1:B:165[A]:ASP:OD1	1:B:227[A]:ARG:NH2	1.58	1.34
1:B:336[A]:LYS:HE2	3:B:2344:HOH:O	1.41	1.17
1:B:194[A]:GLU:OE1	3:B:2200:HOH:O	1.72	1.06
2:C:1348:CL:CL	3:C:2179:HOH:O	2.23	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	324/318 (102%)	305 (94%)	17 (5%)	2 (1%)	25	15
1	B	326/318 (102%)	311 (95%)	14 (4%)	1 (0%)	41	31
1	C	323/318 (102%)	305 (94%)	16 (5%)	2 (1%)	25	15
1	D	316/318 (99%)	300 (95%)	15 (5%)	1 (0%)	41	31
1	E	319/318 (100%)	304 (95%)	13 (4%)	2 (1%)	25	15
1	F	316/318 (99%)	300 (95%)	15 (5%)	1 (0%)	41	31
All	All	1924/1908 (101%)	1825 (95%)	90 (5%)	9 (0%)	29	18

5 of 9 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	31	ALA
1	A	31	ALA
1	E	31	ALA
1	A	57	ILE
1	B	57	ILE

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	275/267 (103%)	264 (96%)	11 (4%)	31	22
1	B	277/267 (104%)	264 (95%)	13 (5%)	26	16
1	C	274/267 (103%)	271 (99%)	3 (1%)	73	73

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	D	268/267 (100%)	260 (97%)	8 (3%)	41 33
1	E	270/267 (101%)	261 (97%)	9 (3%)	38 29
1	F	268/267 (100%)	260 (97%)	8 (3%)	41 33
All	All	1632/1602 (102%)	1580 (97%)	52 (3%)	43 30

5 of 52 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	336[B]	LYS
1	D	102	LYS
1	F	240	LEU
1	C	214[A]	GLN
1	C	249	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	214	GLN
1	D	219	GLN
1	E	219	GLN
1	C	332	GLN
1	E	214	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 carbohydrates in this entry.

5.6 Ligand geometry

Of 6 ligands modelled in this entry, 6 are monoatomic - leaving 0 for Mogul analysis.

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

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.