



# wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 16, 2023 – 12:32 pm GMT

PDB ID : 4BA0  
Title : Crystal Structure of Agd31B, alpha-transglucosylase, complexed with 5F-alpha-GlcF  
Authors : Larsbrink, J.; Izumi, A.; Hemsworth, G.R.; Davies, G.J.; Brumer, H.  
Deposited on : 2012-09-09  
Resolution : 1.85 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](mailto: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  
Mogul : 1.8.4, CSD as541be (2020)  
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

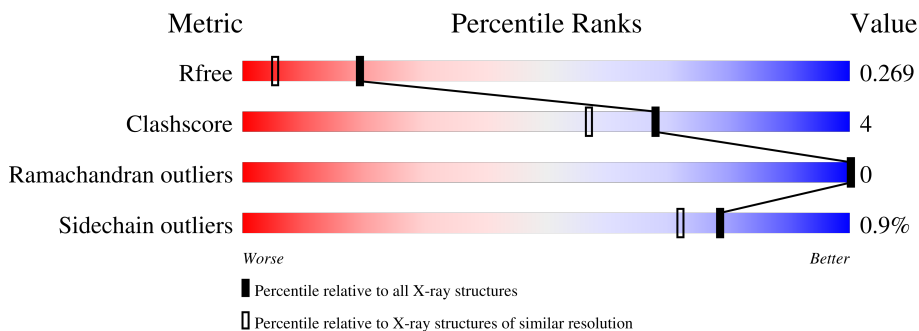
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.85 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	2469 (1.86-1.86)
Clashscore	141614	2625 (1.86-1.86)
Ramachandran outliers	138981	2592 (1.86-1.86)
Sidechain outliers	138945	2592 (1.86-1.86)

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  $\geq 3$ , 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  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	A	817	

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
7	PEG	A	1828	-	-	X	-

## 2 Entry composition [i](#)

There are 8 unique types of molecules in this entry. The entry contains 6830 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.

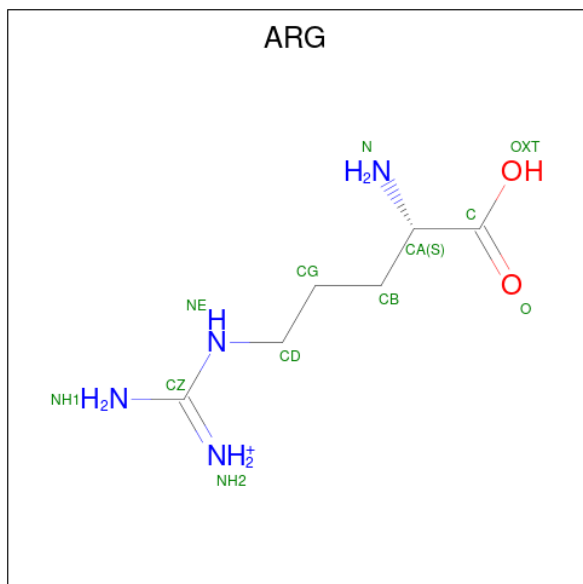
- Molecule 1 is a protein called ALPHA-GLUCOSIDASE, PUTATIVE, ADG31B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	781	6193	3968	1049	1153	23	0	5	0

There is a discrepancy between the modelled and reference sequences:

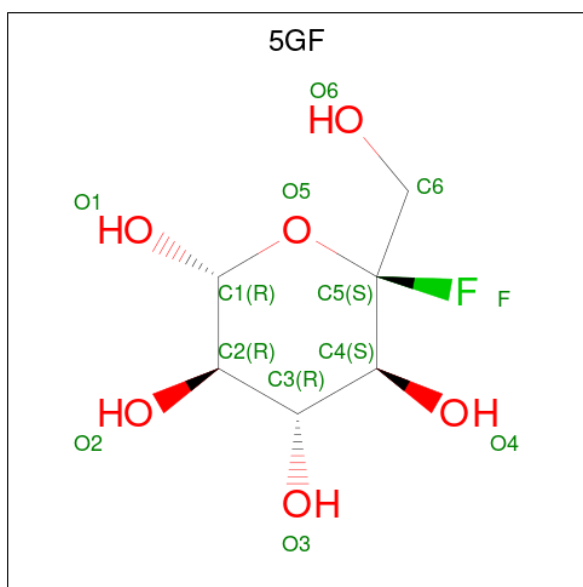
Chain	Residue	Modelled	Actual	Comment	Reference
A	817	LYS	-	expression tag	UNP B3PEE6

- Molecule 2 is ARGinine (three-letter code: ARG) (formula: C<sub>6</sub>H<sub>15</sub>N<sub>4</sub>O<sub>2</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	A	1	12	6	4	2	0	0

- Molecule 3 is 5-fluoro-beta-D-glucopyranose (three-letter code: 5GF) (formula: C<sub>6</sub>H<sub>11</sub>FO<sub>6</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	F	O	0	0
			12	6	1	5		

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula:  $C_2H_6O_2$ ).



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

- Molecule 5 is SULFATE ION (three-letter code: SO4) (formula:  $O_4S$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	O	S	0	0
			5	4	1		
5	A	1	Total	O	S	0	0
			5	4	1		
5	A	1	Total	O	S	0	0
			5	4	1		
5	A	1	Total	O	S	0	0
			5	4	1		
5	A	1	Total	O	S	0	0
			5	4	1		

- Molecule 6 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula: C<sub>6</sub>H<sub>14</sub>O<sub>4</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	A	1	Total	C	O	0	0
			10	6	4		

- Molecule 7 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula:  $C_4H_{10}O_3$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
7	A	1	Total	C	O	0	0
			7	4	3		


- Molecule 8 is water.

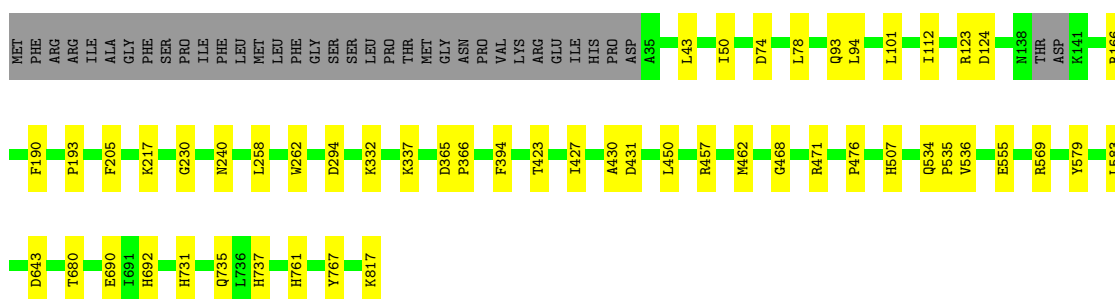
<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
8	A	557	Total 557	O 557	0	0

### 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. 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: ALPHA-GLUCOSIDASE, PUTATIVE, ADG31B

Chain A:  89% 6%





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 6 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	197.23Å 197.23Å 103.06Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	44.48 – 1.85 44.48 – 1.85	Depositor EDS
% Data completeness (in resolution range)	100.0 (44.48-1.85) 100.0 (44.48-1.85)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.97 (at 1.86Å)	Xtrriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, $R_{free}$	0.167 , 0.193 0.262 , 0.269	Depositor DCC
$R_{free}$ test set	5007 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	21.2	Xtrriage
Anisotropy	0.849	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 39.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	6830	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	31.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.92% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: PGE, EDO, PEG, 5GF, SO4

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	0.34	0/6368	0.52	0/8645

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 [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	6193	0	5966	47	0
2	A	12	0	12	0	0
3	A	12	0	9	0	0
4	A	4	0	6	1	0
5	A	35	0	0	0	0
6	A	10	0	14	3	0
7	A	7	0	10	7	0
8	A	557	0	0	5	0
All	All	6830	0	6017	47	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 4.

The worst 5 of 47 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:692:HIS:CD2	1:A:761:HIS:HE1	1.78	1.01
1:A:569:ARG:NH2	6:A:1827:PGE:H12	1.78	0.98
1:A:692:HIS:HD2	1:A:761:HIS:HE1	1.12	0.96
1:A:692:HIS:HD2	1:A:761:HIS:CE1	1.91	0.88
1:A:123[A]:ARG:NH2	1:A:240:ASN:HB3	1.95	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	782/817 (96%)	755 (96%)	27 (4%)	0	<a href="#">100</a> <a href="#">100</a>

There are no Ramachandran outliers to report.

### 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	643/690 (93%)	637 (99%)	6 (1%)	<a href="#">78</a> <a href="#">72</a>

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	294	ASP
1	A	394	PHE
1	A	680	THR
1	A	166	ARG
1	A	74	ASP

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	A	761	HIS
1	A	766	GLN
1	A	769	HIS
1	A	768	GLN
1	A	692	HIS

### 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](#)

12 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
6	PGE	A	1827	-	9,9,9	0.50	0	8,8,8	0.24	0
5	SO4	A	1826	-	4,4,4	0.15	0	6,6,6	0.05	0
2	ARG	A	1829	-	10,11,11	0.75	1 (10%)	11,13,13	1.09	2 (18%)
4	EDO	A	1819	-	3,3,3	0.50	0	2,2,2	0.31	0
7	PEG	A	1828	-	6,6,6	0.68	0	5,5,5	0.49	0
5	SO4	A	1820	-	4,4,4	0.14	0	6,6,6	0.07	0
3	5GF	A	1818	1	9,12,13	1.20	1 (11%)	14,18,20	1.78	4 (28%)
5	SO4	A	1821	-	4,4,4	0.14	0	6,6,6	0.07	0
5	SO4	A	1822	-	4,4,4	0.16	0	6,6,6	0.16	0
5	SO4	A	1823	-	4,4,4	0.16	0	6,6,6	0.15	0
5	SO4	A	1824	-	4,4,4	0.15	0	6,6,6	0.06	0
5	SO4	A	1825	-	4,4,4	0.14	0	6,6,6	0.07	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '2' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
6	PGE	A	1827	-	-	4/7/7/7	-
2	ARG	A	1829	-	-	2/11/11/11	-
7	PEG	A	1828	-	-	2/4/4/4	-
4	EDO	A	1819	-	-	1/1/1/1	-
3	5GF	A	1818	1	-	0/2/23/26	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	1818	5GF	O5-C5	3.23	1.44	1.37
2	A	1829	ARG	OXT-C	-2.08	1.23	1.30

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	1818	5GF	O5-C1-C2	3.56	116.44	111.11
3	A	1818	5GF	C1-O5-C5	3.53	120.41	113.73
3	A	1818	5GF	C1-C2-C3	3.37	113.81	109.67
2	A	1829	ARG	OXT-C-O	-2.74	117.87	124.09
2	A	1829	ARG	OXT-C-CA	2.15	120.70	113.38

There are no chirality outliers.

5 of 9 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	1829	ARG	O-C-CA-N
6	A	1827	PGE	O2-C3-C4-O3
6	A	1827	PGE	O1-C1-C2-O2
7	A	1828	PEG	O1-C1-C2-O2
2	A	1829	ARG	OXT-C-CA-N

There are no ring outliers.

3 monomers are involved in 11 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	A	1827	PGE	3	0
4	A	1819	EDO	1	0
7	A	1828	PEG	7	0

## 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

### 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.