



# wwPDB X-ray Structure Validation Summary Report

Nov 20, 2014 – 05:13 PM EST

PDB ID : 4D6I  
Title : Crystal structure of a family 98 glycoside hydrolase catalytic module (Sp3GH98) in complex with the type 1 blood group A-tetrasaccharide (E558A L19 mutant)  
Authors : Kwan, D.H.; Constantinescu, I.; Chapanian, R.; Higgins, M.A.; Samain, E.; Boraston, A.B.; Kizhakkedathu, J.N.; Withers, S.G.  
Deposited on : 2014-11-11  
Resolution : 1.99 Å(reported)

This is a wwPDB 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 <http://wwpdb.org/ValidationPDFNotes.html>

---

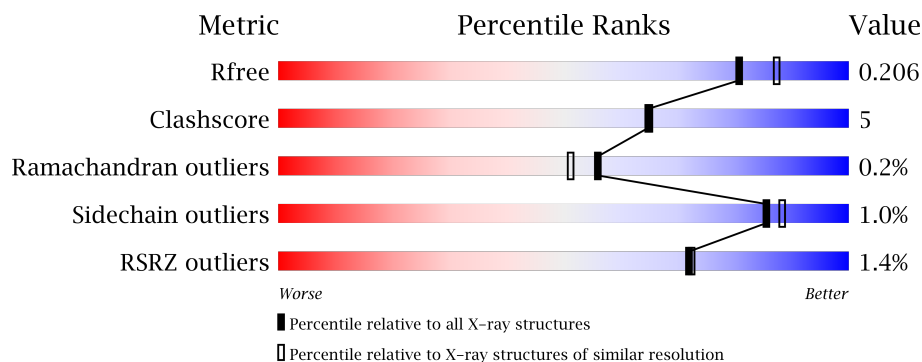
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.16 November 2013  
Xtriage (Phenix) : dev-1439  
EDS : stable24195  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.1.3  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable24195

# 1 Overall quality at a glance


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))
$R_{free}$	66092	4888 (2.00-2.00)
Clashscore	79885	6188 (2.00-2.00)
Ramachandran outliers	78287	6102 (2.00-2.00)
Sidechain outliers	78261	6100 (2.00-2.00)
RSRZ outliers	66119	4890 (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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	599	

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 5548 atoms, of which 0 are hydrogen and 0 are deuterium.

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 GLYCOSIDE HYDROLASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	587	Total	C	N	O	S	0	5	0
			4810	3083	811	903	13			

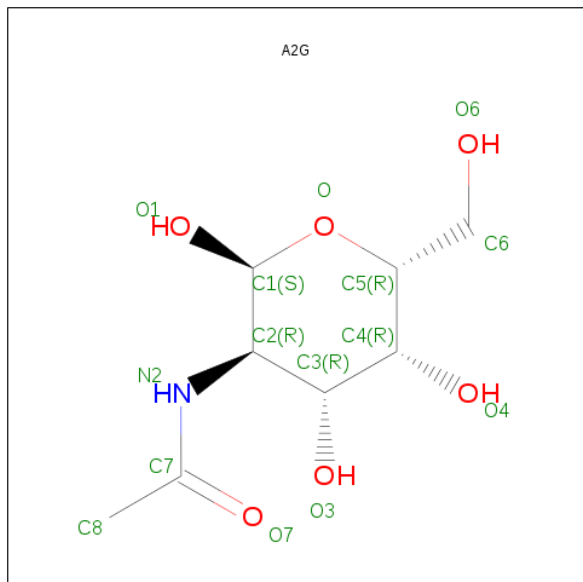
There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	530	HIS	TYR	ENGINEERED MUTATION	UNP A5LBQ0
A	558	ALA	GLU	ENGINEERED MUTATION	UNP A5LBQ0
A	559	SER	ASN	ENGINEERED MUTATION	UNP A5LBQ0
A	592	SER	ASN	ENGINEERED MUTATION	UNP A5LBQ0
A	1004	LEU	-	EXPRESSION TAG	UNP A5LBQ0
A	1005	GLU	-	EXPRESSION TAG	UNP A5LBQ0
A	1006	HIS	-	EXPRESSION TAG	UNP A5LBQ0
A	1007	THR	-	EXPRESSION TAG	UNP A5LBQ0
A	1008	ARG	-	EXPRESSION TAG	UNP A5LBQ0
A	1009	ALA	-	EXPRESSION TAG	UNP A5LBQ0
A	1010	PRO	-	EXPRESSION TAG	UNP A5LBQ0
A	1011	PRO	-	EXPRESSION TAG	UNP A5LBQ0
A	1012	PRO	-	EXPRESSION TAG	UNP A5LBQ0
A	1013	PRO	-	EXPRESSION TAG	UNP A5LBQ0
A	1014	PRO	-	EXPRESSION TAG	UNP A5LBQ0
A	1015	LEU	-	EXPRESSION TAG	UNP A5LBQ0
A	1016	ARG	-	EXPRESSION TAG	UNP A5LBQ0
A	1017	SER	-	EXPRESSION TAG	UNP A5LBQ0
A	1018	GLY	-	EXPRESSION TAG	UNP A5LBQ0
A	1019	CYS	-	EXPRESSION TAG	UNP A5LBQ0

- Molecule 2 is a polymer of unknown type called SUGAR (3-MER).

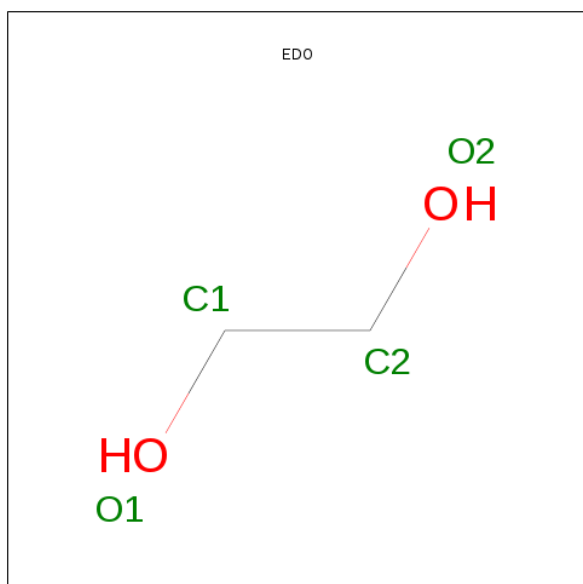
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	3	Total	C	N	O	0	0
			36	20	1	15		

- Molecule 3 is SUGAR (N-ACETYL-2-DEOXY-2-AMINO-GALACTOSE) (three-letter code: A2G) (formula:  $C_8H_{15}NO_6$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			14	8	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		

*Continued on next page...*

*Continued from previous page...*

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

- Molecule 5 is water.

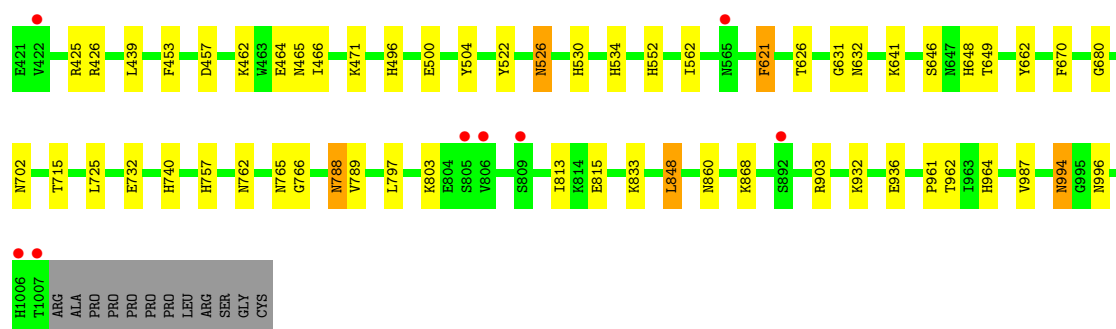
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	680	Total	O	0	0
			680	680		

### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: GLYCOSIDE HYDROLASE

Chain A: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	92.30Å 154.48Å 97.06Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.95 – 1.99 29.95 – 1.99	Depositor EDS
% Data completeness (in resolution range)	99.2 (29.95-1.99) 99.2 (29.95-1.99)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.69 (at 2.00Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, $R_{free}$	0.160 , 0.206 0.160 , 0.206	Depositor DCC
$R_{free}$ test set	2403 reflections (5.33%)	DCC
Wilson B-factor (Å <sup>2</sup> )	16.5	Xtriage
Anisotropy	0.046	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 31.7	EDS
Estimated twinning fraction	0.015 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.025 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 47497 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	5548	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	15.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.53% of the height of the origin peak. No significant pseudotranslation is detected.*

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

## 5 Model quality ⓘ

### 5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: FUC, GAL, NAG, A2G, EDO

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	0.56	0/4944	0.60	2/6699 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	425	ARG	NE-CZ-NH2	-5.88	117.36	120.30
1	A	457	ASP	CB-CG-OD1	5.22	123.00	118.30

There are no chirality outliers.

There are no planarity outliers.

### 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	4810	0	4631	48	0
2	A	36	0	33	3	0
3	A	14	0	13	2	0
4	A	8	0	12	2	0
5	A	680	0	0	12	0
All	All	5548	0	4689	50	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 5.

The worst 5 of 50 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:462:LYS:HE3	5:A:3069:HOH:O	1.54	1.06
1:A:670:PHE:H	1:A:860:ASN:HD21	1.28	0.81
1:A:868:LYS:HD2	5:A:3543:HOH:O	1.81	0.80
1:A:903:ARG:H	1:A:994:ASN:HD21	1.29	0.79
1:A:426:ARG:HH11	1:A:552:HIS:HD2	1.29	0.78

There are no symmetry-related clashes.

## 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	590/599 (98%)	569 (96%)	20 (3%)	1 (0%)	56	51

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	562	ILE

### 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	515/526 (98%)	510 (99%)	5 (1%)	85	88

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	526	ASN
1	A	621	PHE
1	A	788	ASN
1	A	848	LEU
1	A	994	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	648	HIS
1	A	757	HIS
1	A	992	ASN
1	A	702	ASN
1	A	740	HIS

### 5.3.3 RNA ⓘ

There are no RNA chains 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 ⓘ

3 carbohydrates 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	FUC	A	2008	2	9,10,11	0.82	0	10,14,16	0.66	0
2	GAL	A	2009	3,2	10,11,12	1.13	1 (10%)	11,15,17	1.74	3 (27%)
2	NAG	A	2010	2	15,15,15	0.44	0	21,21,21	1.22	1 (4%)

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. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	FUC	A	2008	2	-	0/0/17/20	0/1/1/1
2	GAL	A	2009	3,2	-	0/2/19/22	0/1/1/1
2	NAG	A	2010	2	-	0/6/26/26	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	2009	GAL	O5-C5	-2.89	1.40	1.45

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	2009	GAL	C4-C3-C2	3.37	114.37	110.61
2	A	2010	NAG	O5-C5-C6	2.86	113.47	106.34
2	A	2009	GAL	O2-C2-C3	-2.70	104.39	110.10
2	A	2009	GAL	O4-C4-C5	2.37	115.53	109.25

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 5.6 Ligand geometry

3 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	$\# Z  > 2$
3	A2G	A	2011	2	12,14,15	1.05	1 (8%)	15,19,21	1.35	1 (6%)
4	EDO	A	2012	-	3,3,3	0.43	0	2,2,2	0.49	0
4	EDO	A	2013	-	3,3,3	0.61	0	2,2,2	0.26	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. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	A2G	A	2011	2	-	0/6/23/26	0/1/1/1
4	EDO	A	2012	-	-	0/1/1/1	0/0/0/0
4	EDO	A	2013	-	-	0/1/1/1	0/0/0/0

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	2011	A2G	C8-C7	2.61	1.56	1.50

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	2011	A2G	C2-N2-C7	-4.47	117.99	123.39

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	587/599 (97%)	-0.27	8 (1%) 72 72	7, 13, 26, 38	7 (1%)

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	806	VAL	4.3
1	A	892	SER	3.1
1	A	1007	THR	3.1
1	A	1006	HIS	2.6
1	A	422	VAL	2.4

### 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 ⓘ

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
2	NAG	A	2010	15/15	0.14	4.01	22,29,32,32	0
2	GAL	A	2009	11/12	0.09	-0.12	15,16,18,19	0
2	FUC	A	2008	10/11	0.08	-0.67	12,13,15,15	0

## 6.4 Ligands

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
4	EDO	A	2012	4/4	0.12	1.73	25,25,26,27	0
4	EDO	A	2013	4/4	0.08	0.25	10,13,13,15	0
3	A2G	A	2011	14/15	0.09	0.06	10,11,18,18	0

## 6.5 Other polymers

There are no such residues in this entry.