



# wwPDB X-ray Structure Validation Summary Report

Feb 27, 2014 – 12:46 AM GMT

PDB ID : 2O1G  
Title : Natural occurring mutant of Human ABO(H) Galactosyltransferase :GTB/M214T  
Authors : Letts, J.A.; Borisova, S.N.; Evans, S.V.  
Deposited on : 2006-11-28  
Resolution : 1.71 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>

---

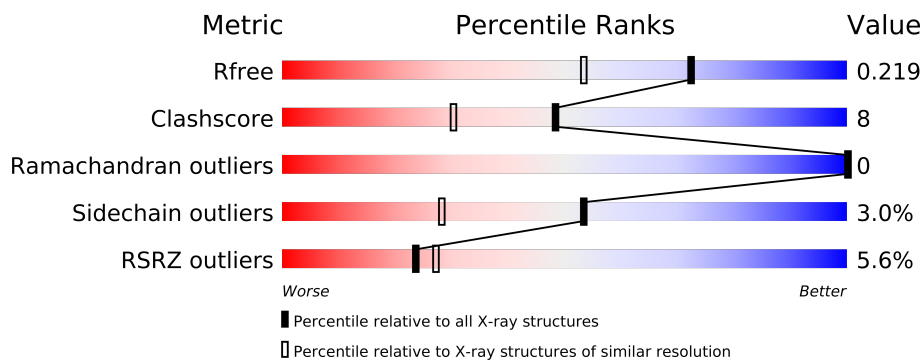
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22683

# 1 Overall quality at a glance

The reported resolution of this entry is 1.71 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	2979 (1.74-1.70)
Clashscore	79885	3506 (1.74-1.70)
Ramachandran outliers	78287	3449 (1.74-1.70)
Sidechain outliers	78261	3449 (1.74-1.70)
RSRZ outliers	66119	2979 (1.74-1.70)

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	297	

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 2388 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 ABO glycosyltransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	264	Total	C	N	O	S	0	3	0
			2185	1420	371	381	13			

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	58	ALA	-	CLONING ARTIFACT	UNP Q70V27
A	59	ILE	-	CLONING ARTIFACT	UNP Q70V27
A	60	GLY	-	CLONING ARTIFACT	UNP Q70V27
A	61	GLU	-	CLONING ARTIFACT	UNP Q70V27
A	62	PHE	-	CLONING ARTIFACT	UNP Q70V27
A	63	MET	-	CLONING ARTIFACT	UNP Q70V27
A	214	THR	MET	ENGINEERED	UNP Q70V27

- Molecule 2 is MERCURY (II) ION (three-letter code: HG) (formula: Hg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	5	Total	Hg	0	0
			5	5		

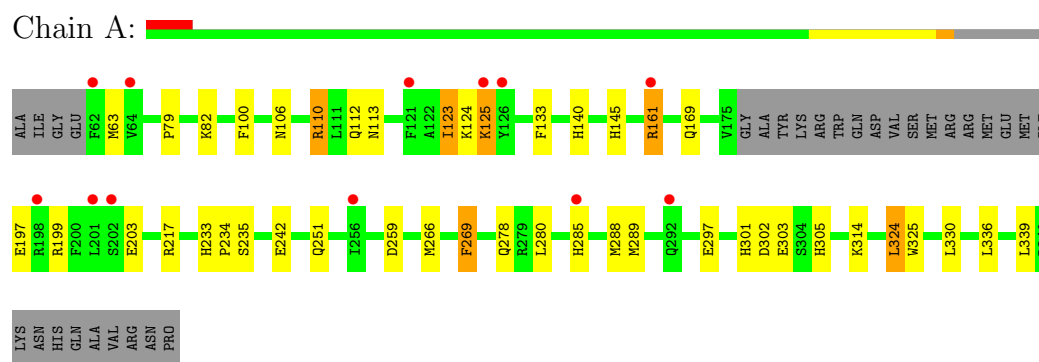
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	198	Total	O	0	0
			198	198		

### 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: ABO glycosyltransferase



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	52.72Å 150.16Å 79.55Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 1.71 19.22 – 1.71	Depositor EDS
% Data completeness (in resolution range)	96.7 (20.00-1.71) 96.7 (19.22-1.71)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.53 (at 1.71Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.183 , 0.218 0.182 , 0.219	Depositor DCC
$R_{free}$ test set	1688 reflections (5.32%)	DCC
Wilson B-factor (Å <sup>2</sup> )	24.9	Xtriage
Anisotropy	0.384	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.42 , 46.0	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtriage
Outliers	1 of 33405 reflections (0.003%)	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	2388	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	25.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.79% 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: HG

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	1.26	5/2251 (0.2%)	1.02	6/3060 (0.2%)

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	269	PHE	CE1-CZ	6.52	1.49	1.37
1	A	169	GLN	CB-CG	-6.47	1.35	1.52
1	A	242	GLU	CG-CD	5.49	1.60	1.51
1	A	314	LYS	CD-CE	5.29	1.64	1.51
1	A	341	PHE	CD1-CE1	5.12	1.49	1.39

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	259	ASP	CB-CG-OD1	6.32	123.99	118.30
1	A	110	ARG	NE-CZ-NH2	6.28	123.44	120.30
1	A	100	PHE	CB-CG-CD2	-5.73	116.79	120.80
1	A	133	PHE	CB-CG-CD1	-5.63	116.86	120.80
1	A	110	ARG	NE-CZ-NH1	-5.50	117.55	120.30

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	2185	0	2163	33	0
2	A	5	0	0	1	0
3	A	198	0	0	8	1
All	All	2388	0	2163	34	1

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 8.

The worst 5 of 34 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:288:MET:HE2	3:A:544:HOH:O	1.32	1.24
2:A:404:HG:HG	3:A:517:HOH:O	1.38	1.01
1:A:125:LYS:H	1:A:125:LYS:HD3	1.29	0.97
1:A:297:GLU:OE2	1:A:301:HIS:HD2	1.62	0.82
1:A:112:GLN:HG3	3:A:529:HOH:O	1.83	0.78

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
3:A:545:HOH:O	3:A:602:HOH:O[3_654]	2.16	0.04

## 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	264/297 (89%)	259 (98%)	5 (2%)	0	100 100

There are no Ramachandran outliers to report.

### 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	239/263 (91%)	232 (97%)	7 (3%)	55 29

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	161	ARG
1	A	324	LEU
1	A	251	GLN
1	A	123	ILE
1	A	269	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	251	GLN
1	A	305	HIS
1	A	286	GLN
1	A	155	GLN
1	A	278	GLN

### 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 ⓘ

There are no carbohydrates in this entry.



## 5.6 Ligand geometry ⓘ

Of 5 ligands modelled in this entry, 5 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.

## 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	264/297 (88%)	0.11	15 (5%) 23 26	15, 23, 38, 69	0

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	196	CYS	9.0
1	A	345	PRO	4.9
1	A	62	PHE	4.2
1	A	64	VAL	4.0
1	A	161	ARG	3.6

### 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 ⓘ

There are no carbohydrates in this entry.

### 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
2	HG	A	401	1/1	0.07	-1.04	32,32,32,32	1

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	HG	A	402	1/1	0.04	-1.53	39,39,39,39	1
2	HG	A	403	1/1	0.03	-1.92	33,33,33,33	1
2	HG	A	405	1/1	0.03	-1.99	30,30,30,30	1
2	HG	A	404	1/1	0.02	-6.15	30,30,30,30	1

## 6.5 Other polymers ⓘ

There are no such residues in this entry.