



# Full wwPDB X-ray Structure Validation Report i

Nov 21, 2014 – 02:20 AM EST

PDB ID : 4D6F  
Title : Crystal structure of a family 98 glycoside hydrolase catalytic module (Sp3GH98) in complex with the blood group A-tetrasaccharide (E558A, X01 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 : 2.03 Å(reported)

This is a full wwPDB validation 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>

---

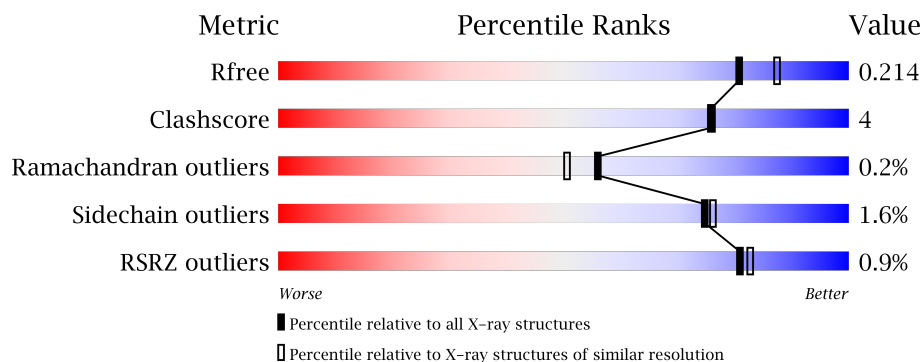
The following versions of software and data (see [references](#)) were used in the production of this report:

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 2.03 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	6003 (2.04-2.00)
Clashscore	79885	7467 (2.04-2.00)
Ramachandran outliers	78287	7370 (2.04-2.00)
Sidechain outliers	78261	7368 (2.04-2.00)
RSRZ outliers	66119	6006 (2.04-2.00)

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	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
3	A2G	A	2010	-	X
4	EDO	A	2012	-	X

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 5478 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	586	Total	C	N	O	S	0	2	0
			4784	3071	803	897	13			

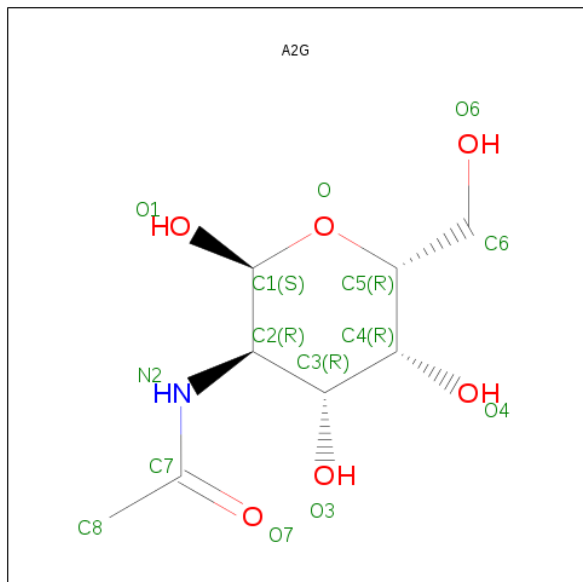
There are 19 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	558	ALA	GLU	ENGINEERED MUTATION	UNP A5LBQ0
A	559	SER	ASN	ENGINEERED MUTATION	UNP A5LBQ0
A	592	VAL	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).

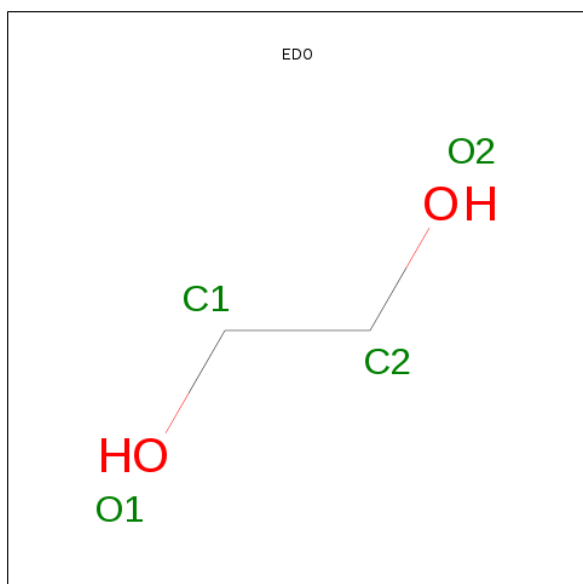
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	3	Total	C	N	O	0	0
			36	20	1	15		

- 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.

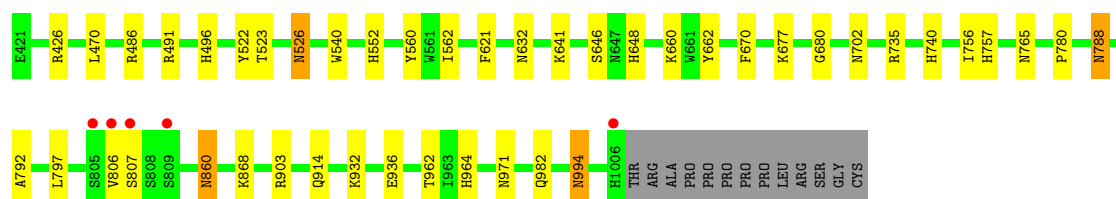
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	636	Total	O	0	0
			636	636		

### 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, $\alpha$ , $\beta$ , $\gamma$	92.16Å 154.78Å 97.26Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.28 – 2.03 30.28 – 2.03	Depositor EDS
% Data completeness (in resolution range)	97.3 (30.28-2.03) 97.3 (30.28-2.03)	Depositor EDS
$R_{merge}$	0.14	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.76 (at 2.03Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, $R_{free}$	0.167 , 0.215 0.166 , 0.214	Depositor DCC
$R_{free}$ test set	2224 reflections (5.32%)	DCC
Wilson B-factor (Å <sup>2</sup> )	15.9	Xtriage
Anisotropy	0.048	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 32.6	EDS
Estimated twinning fraction	0.020 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.028 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 44001 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	5478	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	14.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.44% 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	0.51	0/4918	0.57	0/6666

There are no bond length outliers.

There are no bond angle outliers.

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	4784	0	4610	33	1
2	A	36	0	33	4	0
3	A	14	0	13	3	0
4	A	8	0	12	1	0
5	A	636	0	0	7	2
All	All	5478	0	4668	36	2

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 4.

All (36) close contacts within the same asymmetric unit are listed below.



Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:426:ARG:HH11	1:A:552:HIS:HD2	1.27	0.81
1:A:903:ARG:H	1:A:994:ASN:HD21	1.28	0.79
1:A:670:PHE:H	1:A:860:ASN:ND2	1.86	0.73
1:A:560:TYR:OH	2:A:2009:NAG:H81	1.88	0.72
1:A:670:PHE:H	1:A:860:ASN:HD21	1.36	0.71
1:A:646:SER:OG	1:A:648:HIS:HD2	1.84	0.61
1:A:962:THR:OG1	1:A:964:HIS:HE1	1.87	0.57
1:A:903:ARG:H	1:A:994:ASN:ND2	2.01	0.57
1:A:740:HIS:HE1	5:A:3366:HOH:O	1.90	0.53
3:A:2010:A2G:H8B	5:A:3317:HOH:O	2.08	0.53
1:A:932:LYS:HE2	1:A:936:GLU:OE2	2.10	0.52
1:A:641:LYS:NZ	1:A:702:ASN:HD22	2.07	0.52
1:A:522:TYR:HA	1:A:526:ASN:ND2	2.26	0.51
1:A:522:TYR:HA	1:A:526:ASN:HD21	1.77	0.50
1:A:702:ASN:HD21	1:A:788:ASN:HB2	1.77	0.49
1:A:662:TYR:CD1	1:A:680:GLY:HA2	2.48	0.48
1:A:560:TYR:OH	2:A:2009:NAG:C8	2.61	0.48
1:A:662:TYR:CZ	1:A:677:LYS:HB3	2.49	0.48
1:A:426:ARG:HH11	1:A:552:HIS:CD2	2.17	0.48
1:A:491:ARG:HD2	1:A:540:TRP:CD1	2.50	0.46
1:A:632:ASN:ND2	1:A:765:ASN:HB3	2.31	0.46
1:A:486:ARG:CZ	4:A:2012:EDO:H11	2.46	0.46
1:A:971:ASN:ND2	5:A:3598:HOH:O	2.45	0.45
1:A:757:HIS:CE1	1:A:797:LEU:O	2.71	0.44
2:A:2008:GAL:C3	3:A:2010:A2G:C1	2.87	0.44
1:A:868:LYS:NZ	5:A:3483:HOH:O	2.37	0.43
2:A:2008:GAL:O3	3:A:2010:A2G:C2	2.62	0.43
1:A:660:LYS:HE2	1:A:780:PRO:HG3	2.02	0.42
1:A:648:HIS:HE1	5:A:3271:HOH:O	2.02	0.42
1:A:994:ASN:HD22	1:A:994:ASN:C	2.23	0.41
1:A:641:LYS:HZ1	1:A:702:ASN:HD22	1.69	0.41
1:A:470:LEU:HD21	1:A:735:ARG:HG3	2.03	0.41
1:A:523:THR:H	1:A:526:ASN:ND2	2.19	0.40
1:A:552:HIS:HE1	5:A:3159:HOH:O	2.04	0.40
1:A:756:ILE:O	1:A:792:ALA:HA	2.21	0.40
1:A:496:HIS:HE1	5:A:3116:HOH:O	2.05	0.40

All (2) 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)
5:A:3146:HOH:O	5:A:3534:HOH:O[8_445]	1.86	0.34
1:A:914:GLN:OE1	5:A:3232:HOH:O[6_445]	2.17	0.03

## 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	586/599 (98%)	562 (96%)	23 (4%)	1 (0%)	56	50

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	512/526 (97%)	504 (98%)	8 (2%)	75	76

All (8) 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	806	VAL
1	A	807	SER
1	A	860	ASN
1	A	982	GLN
1	A	994	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (21) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	496	HIS
1	A	512	ASN
1	A	526	ASN
1	A	552	HIS
1	A	581	ASN
1	A	609	GLN
1	A	618	ASN
1	A	632	ASN
1	A	648	HIS
1	A	702	ASN
1	A	740	HIS
1	A	757	HIS
1	A	788	ASN
1	A	860	ASN
1	A	898	ASN
1	A	917	ASN
1	A	924	GLN
1	A	964	HIS
1	A	971	ASN
1	A	992	ASN
1	A	994	ASN

### 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	2007	2	9,10,11	0.87	0	10,14,16	0.82	0
2	GAL	A	2008	3,2	10,11,12	0.97	1 (10%)	11,15,17	1.85	3 (27%)
2	NAG	A	2009	2	15,15,15	0.44	0	21,21,21	1.39	3 (14%)

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	2007	2	-	0/0/17/20	0/1/1/1
2	GAL	A	2008	3,2	-	0/2/19/22	0/1/1/1
2	NAG	A	2009	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	2008	GAL	O5-C5	-2.63	1.41	1.45

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	2008	GAL	C4-C3-C2	4.09	115.17	110.61
2	A	2009	NAG	O5-C1-C2	3.66	113.05	109.50
2	A	2009	NAG	C4-C3-C2	-2.29	107.24	110.51
2	A	2008	GAL	O3-C3-C2	-2.20	105.74	109.74
2	A	2009	NAG	C1-C2-C3	2.15	113.44	110.42
2	A	2008	GAL	O4-C4-C5	2.06	114.71	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	2010	2	12,14,15	0.64	0	15,19,21	0.86	1 (6%)
4	EDO	A	2011	-	3,3,3	0.54	0	2,2,2	0.30	0
4	EDO	A	2012	-	3,3,3	0.55	0	2,2,2	0.18	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	2010	2	-	0/6/23/26	0/1/1/1
4	EDO	A	2011	-	-	0/1/1/1	0/0/0/0
4	EDO	A	2012	-	-	0/1/1/1	0/0/0/0

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	2010	A2G	O-C5-C6	2.71	109.82	106.98

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	586/599 (97%)	-0.40	5 (0%) 81 83	6, 12, 24, 45	8 (1%)

All (5) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	806	VAL	6.2
1	A	807	SER	3.9
1	A	805	SER	2.7
1	A	809	SER	2.6
1	A	1006	HIS	2.5

### 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	2009	15/15	0.11	1.29	15,18,21,23	0
2	GAL	A	2008	11/12	0.09	-0.10	10,11,12,13	0
2	FUC	A	2007	10/11	0.09	-0.60	10,11,12,12	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	0.20	3.67	26,27,28,28	0
3	A2G	A	2010	14/15	0.12	2.37	9,11,23,27	0
4	EDO	A	2011	4/4	0.08	-0.46	10,11,11,13	0

## 6.5 Other polymers

There are no such residues in this entry.