



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 2, 2016 – 11:04 AM EDT

PDB ID : 5CB7  
Title : Crystallographic structure of human rotavirus K8 VP8\* in complex with A-type HBGA  
Authors : Yu, X.; Blanchard, H.  
Deposited on : 2015-06-30  
Resolution : 1.35 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  
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity : 4.02b-467  
Mogul : 1.7.1 (RC1), CSD as537be (2016)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20027674  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20027674

**i**

## X-RAY DIFFRACTION

A.

 $R_{free}$ 

		c

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
6	GOL	B	307[A]	-	-	-	X
6	GOL	B	308[B]	-	-	-	X

## 2 Entry composition [i](#)

There are 8 unique types of molecules in this entry. The entry contains 3399 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.

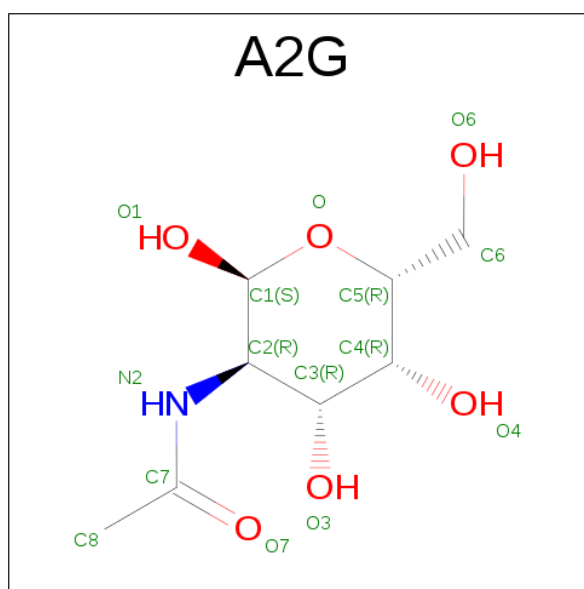
- Molecule 1 is a protein called Outer capsid protein VP4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	160	Total	C	N	O	S	0	14	0
			1346	856	219	263	8			
1	B	160	Total	C	N	O	S	0	11	0
			1333	847	217	262	7			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	149	ASP	TYR	engineered mutation	UNP Q01641
B	149	ASP	TYR	engineered mutation	UNP Q01641

- Molecule 2 is N-ACETYL-2-DEOXY-2-AMINO-GALACTOSE (three-letter code: A2G) (formula: C<sub>8</sub>H<sub>15</sub>NO<sub>6</sub>).



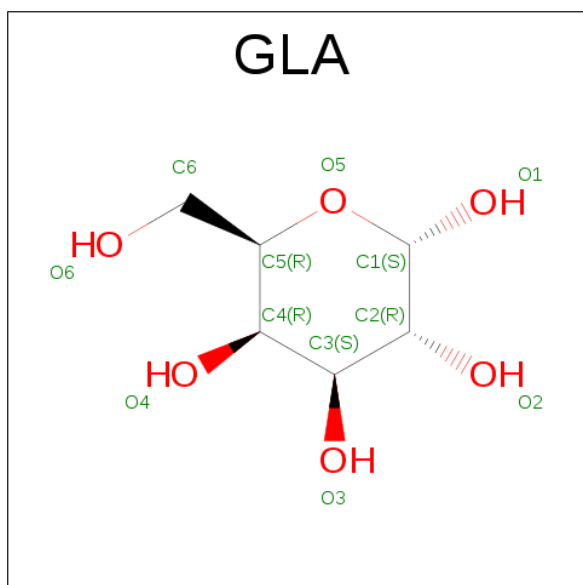
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			14	8	1	5		

*Continued on next page...*

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	B	1	Total	C	N	O	0	1
			14	8	1	5		
2	B	1	Total	C	N	O	0	1
			14	8	1	5		

- Molecule 3 is ALPHA D-GALACTOSE (three-letter code: GLA) (formula: C<sub>6</sub>H<sub>12</sub>O<sub>6</sub>).



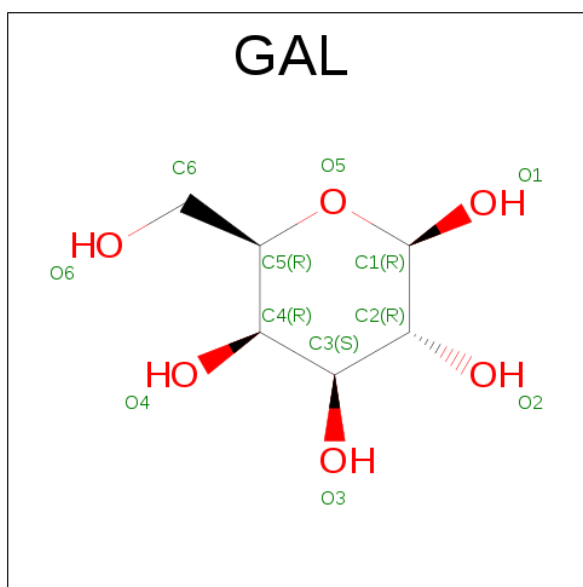
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			12	6	6		
3	B	1	Total	C	O	0	1
			12	6	6		

- Molecule 4 is ALPHA-L-FUCOSE (three-letter code: FUC) (formula: C<sub>6</sub>H<sub>12</sub>O<sub>5</sub>).



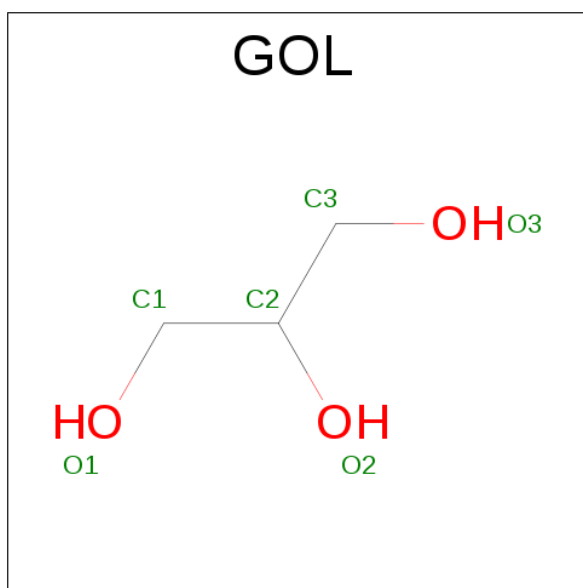
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			10	6	4		
4	B	1	Total	C	O	0	1
			10	6	4		
4	B	1	Total	C	O	0	1
			10	6	4		

- Molecule 5 is BETA-D-GALACTOSE (three-letter code: GAL) (formula: C<sub>6</sub>H<sub>12</sub>O<sub>6</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	B	1	Total	C	O	0	1
			12	6	6		

- Molecule 6 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	B	1	Total	C	O	0	1
			6	3	3		
6	B	1	Total	C	O	0	1
			6	3	3		

- Molecule 7 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	B	2	Total	Cl	0	0
			2	2		

- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	294	Total	O	0	0
			294	294		
8	B	304	Total	O	0	0
			304	304		

### 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 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: Outer capsid protein VP4



- Molecule 1: Outer capsid protein VP4





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	42.27Å 83.19Å 47.75Å 90.00° 94.03° 90.00°	Depositor
Resolution (Å)	47.63 – 1.35 21.39 – 1.35	Depositor EDS
% Data completeness (in resolution range)	95.8 (47.63-1.35) 95.9 (21.39-1.35)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	9.08 (at 1.35Å)	Xtriage
Refinement program	REFMAC 5.8.0073	Depositor
R, $R_{free}$	0.170 , 0.192 0.178 , 0.194	Depositor DCC
$R_{free}$ test set	3479 reflections (5.29%)	DCC
Wilson B-factor (Å <sup>2</sup> )	7.7	Xtriage
Anisotropy	0.043	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.40 , 38.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	3399	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	10.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.54% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: GOL, CL, GLA, GAL, FUC, A2G

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	0.43	0/1417	0.68	0/1936
1	B	0.47	0/1398	0.69	1/1911 (0.1%)
All	All	0.45	0/2815	0.68	1/3847 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	149	ASP	CB-CG-OD1	5.03	122.83	118.30

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-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 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	1346	0	1311	5	0
1	B	1333	0	1291	1	0
2	A	14	0	13	0	0
2	B	28	0	26	0	0
3	A	12	0	10	0	0
3	B	12	0	10	0	0
4	A	10	0	10	0	0
4	B	20	0	20	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	B	12	0	10	0	0
6	B	12	0	16	1	0
7	B	2	0	0	0	0
8	A	294	0	0	1	0
8	B	304	0	0	1	0
All	All	3399	0	2717	7	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 1.

The worst 5 of 7 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:141:ILE:HD13	8:A:410:HOH:O	2.06	0.56
1:B:116:GLN:HG3	1:B:127:GLN:HB3	1.91	0.52
6:B:308[B]:GOL:H11	8:B:610:HOH:O	2.11	0.50
1:A:116:GLN:HG2	1:A:129:HIS:CD2	2.49	0.47
1:A:105[B]:CYS:SG	1:A:140:PHE:HD1	2.38	0.47

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	172/161 (107%)	166 (96%)	6 (4%)	0	100	100
1	B	169/161 (105%)	164 (97%)	5 (3%)	0	100	100
All	All	341/322 (106%)	330 (97%)	11 (3%)	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	157/146 (108%)	153 (98%)	4 (2%)	55	17
1	B	155/146 (106%)	150 (97%)	5 (3%)	46	11
All	All	312/292 (107%)	303 (97%)	9 (3%)	57	13

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	73[A]	SER
1	B	190	LEU
1	B	116	GLN
1	A	167	TRP
1	B	73[B]	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

4 non-standard protein/DNA/RNA residues 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
4	FUC	A	303	3	10,10,11	0.73	0	13,14,16	0.77	0
4	FUC	B	303[A]	3	10,10,11	0.87	1 (10%)	13,14,16	1.09	1 (7%)
5	GAL	B	305[B]	2,4	12,12,12	0.48	0	17,17,17	1.56	3 (17%)
4	FUC	B	306[B]	5	10,10,11	0.73	0	13,14,16	1.33	2 (15%)

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
4	FUC	A	303	3	-	0/0/17/20	0/1/1/1
4	FUC	B	303[A]	3	-	0/0/17/20	0/1/1/1
5	GAL	B	305[B]	2,4	-	0/2/22/22	0/1/1/1
4	FUC	B	306[B]	5	-	0/0/17/20	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	303[A]	FUC	O5-C1	-2.35	1.39	1.43

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	B	305[B]	GAL	C6-C5-C4	-4.12	102.66	112.99
4	B	306[B]	FUC	C3-C4-C5	-2.33	106.19	109.66
5	B	305[B]	GAL	O5-C1-C2	2.09	113.66	110.00
4	B	303[A]	FUC	C3-C4-C5	2.10	112.80	109.66
5	B	305[B]	GAL	O5-C5-C6	2.72	113.42	106.38

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry

Of 13 ligands modelled in this entry, 2 are monoatomic - leaving 11 for Mogul analysis.

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	A2G	A	301	3	14,14,15	0.48	0	15,19,21	0.80	0
3	GLA	A	302	2,4	12,12,12	0.50	0	17,17,17	0.60	0
4	FUC	A	303	3	10,10,11	0.73	0	13,14,16	0.77	0
2	A2G	B	301[A]	3	14,14,15	0.29	0	15,19,21	0.73	0
3	GLA	B	302[A]	2,4	12,12,12	0.62	0	17,17,17	0.55	0
4	FUC	B	303[A]	3	10,10,11	0.87	1 (10%)	13,14,16	1.09	1 (7%)
2	A2G	B	304[B]	5	14,14,15	0.27	0	15,19,21	0.53	0
5	GAL	B	305[B]	2,4	12,12,12	0.48	0	17,17,17	1.56	3 (17%)
4	FUC	B	306[B]	5	10,10,11	0.73	0	13,14,16	1.33	2 (15%)
6	GOL	B	307[A]	-	5,5,5	0.28	0	5,5,5	0.56	0
6	GOL	B	308[B]	-	5,5,5	0.33	0	5,5,5	0.41	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
2	A2G	A	301	3	-	0/6/23/26	0/1/1/1
3	GLA	A	302	2,4	-	0/2/22/22	0/1/1/1
4	FUC	A	303	3	-	0/0/17/20	0/1/1/1
2	A2G	B	301[A]	3	-	0/6/23/26	0/1/1/1
3	GLA	B	302[A]	2,4	-	0/2/22/22	0/1/1/1
4	FUC	B	303[A]	3	-	0/0/17/20	0/1/1/1
2	A2G	B	304[B]	5	-	0/6/23/26	0/1/1/1
5	GAL	B	305[B]	2,4	-	0/2/22/22	0/1/1/1
4	FUC	B	306[B]	5	-	0/0/17/20	0/1/1/1
6	GOL	B	307[A]	-	-	0/4/4/4	0/0/0/0
6	GOL	B	308[B]	-	-	0/4/4/4	0/0/0/0

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	303[A]	FUC	O5-C1	-2.35	1.39	1.43

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	B	305[B]	GAL	C6-C5-C4	-4.12	102.66	112.99
4	B	306[B]	FUC	C3-C4-C5	-2.33	106.19	109.66
5	B	305[B]	GAL	O5-C1-C2	2.09	113.66	110.00
4	B	303[A]	FUC	C3-C4-C5	2.10	112.80	109.66
5	B	305[B]	GAL	O5-C5-C6	2.72	113.42	106.38

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	B	308[B]	GOL	1	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

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	160/161 (99%)	0.27	2 (1%) 79 80	4, 7, 13, 21	0
1	B	160/161 (99%)	0.34	5 (3%) 52 54	4, 7, 15, 24	0
All	All	320/322 (99%)	0.31	7 (2%) 65 67	4, 7, 14, 24	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	65	LEU	4.6
1	A	65	LEU	4.1
1	B	183	ASN	3.5
1	B	88	ARG	3.1
1	B	133	ASP	2.5

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

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	RSCC	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
4	FUC	B	303[A]	10/11	0.97	0.08	-1.18	6,6,6,7	10
4	FUC	B	306[B]	10/11	0.96	0.07	-1.59	6,6,7,7	10
4	FUC	A	303	10/11	0.87	0.13	-	15,16,16,18	0
5	GAL	B	305[B]	12/12	0.97	0.09	-	5,6,6,6	12



### 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

### 6.4 Ligands [i](#)

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	RSCC	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
6	GOL	B	307[A]	6/6	0.90	0.18	7.85	17,17,17,18	6
6	GOL	B	308[B]	6/6	0.91	0.15	4.99	10,10,10,10	6
2	A2G	A	301	14/15	0.91	0.11	1.84	8,11,13,14	0
4	FUC	B	303[A]	10/11	0.97	0.08	-1.18	6,6,6,7	10
2	A2G	B	304[B]	14/15	0.97	0.07	-1.46	5,6,6,6	14
4	FUC	B	306[B]	10/11	0.96	0.07	-1.59	6,6,7,7	10
2	A2G	B	301[A]	14/15	0.97	0.07	-1.70	5,5,5,5	14
7	CL	B	310	1/1	0.99	0.03	-1.84	13,13,13,13	0
7	CL	B	309	1/1	0.98	0.07	-	11,11,11,11	0
5	GAL	B	305[B]	12/12	0.97	0.09	-	5,6,6,6	12
3	GLA	B	302[A]	12/12	0.94	0.10	-	5,6,6,7	12
3	GLA	A	302	12/12	0.88	0.12	-	13,14,16,16	0
4	FUC	A	303	10/11	0.87	0.13	-	15,16,16,18	0

### 6.5 Other polymers [i](#)

There are no such residues in this entry.