



# wwPDB X-ray Structure Validation Summary Report

Mar 31, 2014 – 03:21 PM BST

PDB ID : 4EEG  
Title : Crystal structure of human M340H-beta-1,4-galactosyltransferase-1(M340H-B4GAL-T1) in complex with GLCNAC-BETA1,6-Gal-Beta  
Authors : Ramakrishnan, B.; Qasba, P.K.  
Deposited on : 2012-03-28  
Resolution : 2.20 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>

---

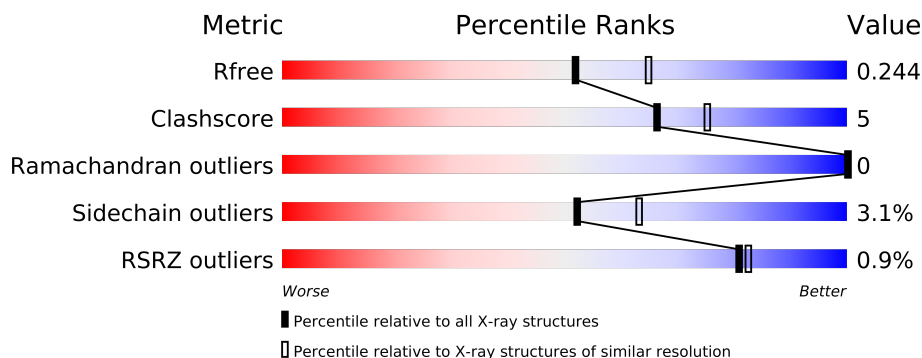
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable23004
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)	:	stable23004

# 1 Overall quality at a glance

The reported resolution of this entry is 2.20 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	2938 (2.20-2.20)
Clashscore	79885	3751 (2.20-2.20)
Ramachandran outliers	78287	3681 (2.20-2.20)
Sidechain outliers	78261	3682 (2.20-2.20)
RSRZ outliers	66119	2939 (2.20-2.20)

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	287	
1	B	287	
1	C	287	

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

Mol	Type	Chain	Res	Geometry	Electron density
5	GOL	A	405	-	X
5	GOL	B	405	-	X
5	GOL	B	406	-	X
5	GOL	C	405	-	X
6	SO4	A	408	-	X
6	SO4	B	411	-	X

## 2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 7417 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 Beta-1,4-galactosyltransferase1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	273	Total	C	N	O	S	0	2	0
			2229	1427	387	404	11			
1	B	273	Total	C	N	O	S	0	1	0
			2223	1423	386	403	11			
1	C	273	Total	C	N	O	S	0	0	0
			2218	1420	386	401	11			

There are 51 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	112	ALA	-	EXPRESSION TAG	UNP P15291
A	113	SER	-	EXPRESSION TAG	UNP P15291
A	114	MET	-	EXPRESSION TAG	UNP P15291
A	115	THR	-	EXPRESSION TAG	UNP P15291
A	116	GLY	-	EXPRESSION TAG	UNP P15291
A	117	GLY	-	EXPRESSION TAG	UNP P15291
A	118	GLN	-	EXPRESSION TAG	UNP P15291
A	119	GLN	-	EXPRESSION TAG	UNP P15291
A	120	MET	-	EXPRESSION TAG	UNP P15291
A	121	GLY	-	EXPRESSION TAG	UNP P15291
A	122	ARG	-	EXPRESSION TAG	UNP P15291
A	123	GLY	-	EXPRESSION TAG	UNP P15291
A	124	SER	-	EXPRESSION TAG	UNP P15291
A	125	ALA	-	EXPRESSION TAG	UNP P15291
A	337	THR	ARG	ENGINEERED MUTATION	UNP P15291
A	338	THR	CYS	ENGINEERED MUTATION	UNP P15291
A	340	HIS	MET	ENGINEERED MUTATION	UNP P15291
B	112	ALA	-	EXPRESSION TAG	UNP P15291
B	113	SER	-	EXPRESSION TAG	UNP P15291
B	114	MET	-	EXPRESSION TAG	UNP P15291
B	115	THR	-	EXPRESSION TAG	UNP P15291
B	116	GLY	-	EXPRESSION TAG	UNP P15291
B	117	GLY	-	EXPRESSION TAG	UNP P15291

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	118	GLN	-	EXPRESSION TAG	UNP P15291
B	119	GLN	-	EXPRESSION TAG	UNP P15291
B	120	MET	-	EXPRESSION TAG	UNP P15291
B	121	GLY	-	EXPRESSION TAG	UNP P15291
B	122	ARG	-	EXPRESSION TAG	UNP P15291
B	123	GLY	-	EXPRESSION TAG	UNP P15291
B	124	SER	-	EXPRESSION TAG	UNP P15291
B	125	ALA	-	EXPRESSION TAG	UNP P15291
B	337	THR	ARG	ENGINEERED MUTATION	UNP P15291
B	338	THR	CYS	ENGINEERED MUTATION	UNP P15291
B	340	HIS	MET	ENGINEERED MUTATION	UNP P15291
C	112	ALA	-	EXPRESSION TAG	UNP P15291
C	113	SER	-	EXPRESSION TAG	UNP P15291
C	114	MET	-	EXPRESSION TAG	UNP P15291
C	115	THR	-	EXPRESSION TAG	UNP P15291
C	116	GLY	-	EXPRESSION TAG	UNP P15291
C	117	GLY	-	EXPRESSION TAG	UNP P15291
C	118	GLN	-	EXPRESSION TAG	UNP P15291
C	119	GLN	-	EXPRESSION TAG	UNP P15291
C	120	MET	-	EXPRESSION TAG	UNP P15291
C	121	GLY	-	EXPRESSION TAG	UNP P15291
C	122	ARG	-	EXPRESSION TAG	UNP P15291
C	123	GLY	-	EXPRESSION TAG	UNP P15291
C	124	SER	-	EXPRESSION TAG	UNP P15291
C	125	ALA	-	EXPRESSION TAG	UNP P15291
C	337	THR	ARG	ENGINEERED MUTATION	UNP P15291
C	338	THR	CYS	ENGINEERED MUTATION	UNP P15291
C	340	HIS	MET	ENGINEERED MUTATION	UNP P15291

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

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	2	Total	C	N	O	0	0
			26	14	1	11		
2	B	2	Total	C	N	O	0	0
			26	14	1	11		
2	C	2	Total	C	N	O	0	0
			26	14	1	11		

There are 51 discrepancies between the modelled and reference sequences:

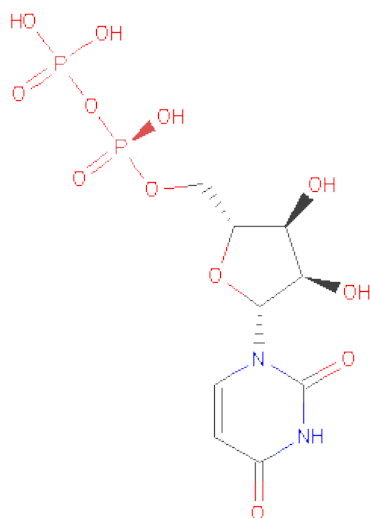
Chain	Residue	Modelled	Actual	Comment	Reference
A	112	ALA	-	EXPRESSION TAG	UNP P15291
A	113	SER	-	EXPRESSION TAG	UNP P15291
A	114	MET	-	EXPRESSION TAG	UNP P15291
A	115	THR	-	EXPRESSION TAG	UNP P15291
A	116	GLY	-	EXPRESSION TAG	UNP P15291
A	117	GLY	-	EXPRESSION TAG	UNP P15291
A	118	GLN	-	EXPRESSION TAG	UNP P15291
A	119	GLN	-	EXPRESSION TAG	UNP P15291
A	120	MET	-	EXPRESSION TAG	UNP P15291
A	121	GLY	-	EXPRESSION TAG	UNP P15291
A	122	ARG	-	EXPRESSION TAG	UNP P15291
A	123	GLY	-	EXPRESSION TAG	UNP P15291
A	124	SER	-	EXPRESSION TAG	UNP P15291
A	125	ALA	-	EXPRESSION TAG	UNP P15291
A	337	THR	ARG	ENGINEERED MUTATION	UNP P15291
A	338	THR	CYS	ENGINEERED MUTATION	UNP P15291
A	340	HIS	MET	ENGINEERED MUTATION	UNP P15291
B	112	ALA	-	EXPRESSION TAG	UNP P15291
B	113	SER	-	EXPRESSION TAG	UNP P15291
B	114	MET	-	EXPRESSION TAG	UNP P15291
B	115	THR	-	EXPRESSION TAG	UNP P15291
B	116	GLY	-	EXPRESSION TAG	UNP P15291
B	117	GLY	-	EXPRESSION TAG	UNP P15291
B	118	GLN	-	EXPRESSION TAG	UNP P15291
B	119	GLN	-	EXPRESSION TAG	UNP P15291
B	120	MET	-	EXPRESSION TAG	UNP P15291
B	121	GLY	-	EXPRESSION TAG	UNP P15291
B	122	ARG	-	EXPRESSION TAG	UNP P15291
B	123	GLY	-	EXPRESSION TAG	UNP P15291
B	124	SER	-	EXPRESSION TAG	UNP P15291
B	125	ALA	-	EXPRESSION TAG	UNP P15291
B	337	THR	ARG	ENGINEERED MUTATION	UNP P15291
B	338	THR	CYS	ENGINEERED MUTATION	UNP P15291
B	340	HIS	MET	ENGINEERED MUTATION	UNP P15291
C	112	ALA	-	EXPRESSION TAG	UNP P15291
C	113	SER	-	EXPRESSION TAG	UNP P15291
C	114	MET	-	EXPRESSION TAG	UNP P15291
C	115	THR	-	EXPRESSION TAG	UNP P15291
C	116	GLY	-	EXPRESSION TAG	UNP P15291
C	117	GLY	-	EXPRESSION TAG	UNP P15291
C	118	GLN	-	EXPRESSION TAG	UNP P15291
C	119	GLN	-	EXPRESSION TAG	UNP P15291
C	120	MET	-	EXPRESSION TAG	UNP P15291

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	121	GLY	-	EXPRESSION TAG	UNP P15291
C	122	ARG	-	EXPRESSION TAG	UNP P15291
C	123	GLY	-	EXPRESSION TAG	UNP P15291
C	124	SER	-	EXPRESSION TAG	UNP P15291
C	125	ALA	-	EXPRESSION TAG	UNP P15291
C	337	THR	ARG	ENGINEERED MUTATION	UNP P15291
C	338	THR	CYS	ENGINEERED MUTATION	UNP P15291
C	340	HIS	MET	ENGINEERED MUTATION	UNP P15291

- Molecule 3 is URIDINE-5'-DIPHOSPHATE (three-letter code: UDP) (formula:  $C_9H_{14}N_2O_{12}P_2$ ).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	A	1	Total	C	N	O	P	0	0
			25	9	2	12	2		
3	B	1	Total	C	N	O	P	0	0
			25	9	2	12	2		
3	C	1	Total	C	N	O	P	0	0
			25	9	2	12	2		

- Molecule 4 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

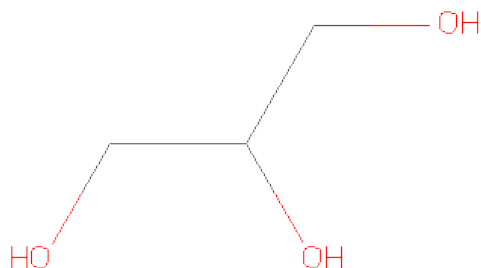
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Mn	0	0
			1	1		
4	A	1	Total	Mn	0	0
			1	1		

Continued on next page...

*Continued from previous page...*

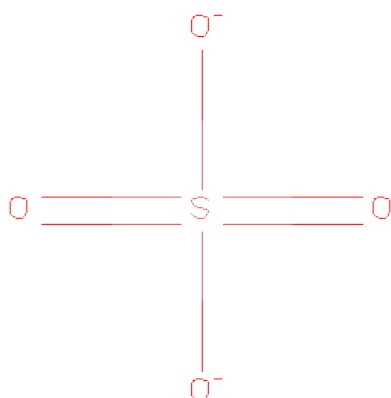
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	C	1	Total	Mn	0	0
			1	1		

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



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

- Molecule 6 is SULFATE ION (three-letter code: SO4) (formula:  $O_4S$ ).



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

- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	189	Total	O	0	0
			189	189		
7	B	202	Total	O	0	0
			202	202		

*Continued on next page...*

*Continued from previous page...*

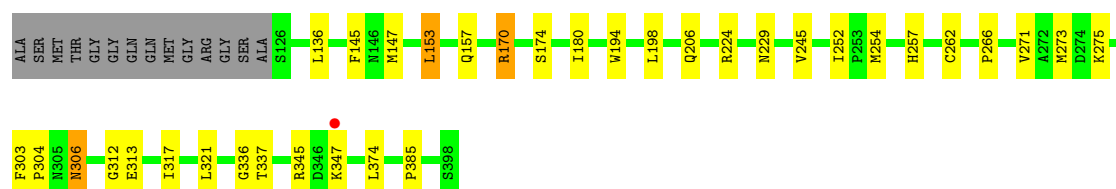
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	C	126	Total 126	O 126	0	0

### 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.

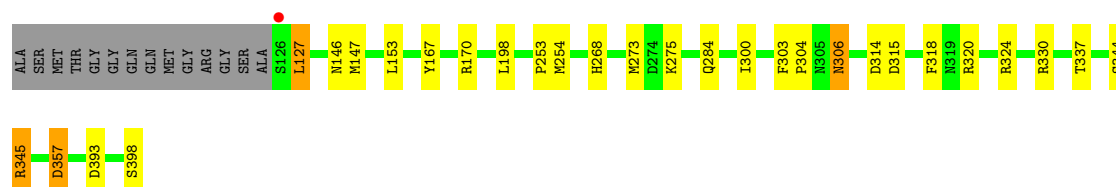
- Molecule 1: Beta-1,4-galactosyltransferase1

Chain A: 



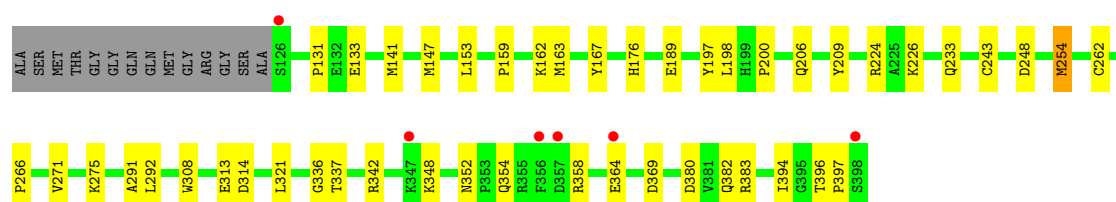
- Molecule 1: Beta-1,4-galactosyltransferase1

Chain B: 



- Molecule 1: Beta-1,4-galactosyltransferase1

Chain C: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	107.36Å 195.46Å 143.73Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.05 – 2.20 47.05 – 2.20	Depositor EDS
% Data completeness (in resolution range)	96.2 (47.05-2.20) 96.2 (47.05-2.20)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.08	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.42 (at 2.20Å)	Xtriage
Refinement program	REFMAC 5.5.0110	Depositor
R, $R_{free}$	0.199 , 0.244 0.202 , 0.244	Depositor DCC
$R_{free}$ test set	3705 reflections (5.28%)	DCC
Wilson B-factor (Å <sup>2</sup> )	38.5	Xtriage
Anisotropy	0.020	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 34.9	EDS
Estimated twinning fraction	0.000 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.014 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.35$	Xtriage
Outliers	0 of 73889 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	7417	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.90% 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: GOL, UDP, NAG, MN, GAL, 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	0.83	0/2297	0.84	1/3120 (0.0%)
1	B	0.85	0/2288	0.88	4/3108 (0.1%)
1	C	0.67	0/2280	0.75	0/3097
All	All	0.79	0/6865	0.83	5/9325 (0.1%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	357	ASP	CB-CG-OD2	-6.47	112.47	118.30
1	A	262	CYS	CA-CB-SG	-5.86	103.46	114.00
1	B	393	ASP	CB-CG-OD2	5.84	123.56	118.30
1	B	315	ASP	CB-CG-OD1	5.50	123.25	118.30
1	B	127	LEU	CA-CB-CG	5.05	126.92	115.30

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	2229	0	2174	24	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	2223	0	2166	17	0
1	C	2218	0	2162	29	0
2	A	26	0	23	1	0
2	B	26	0	24	1	0
2	C	26	0	24	0	0
3	A	25	0	11	0	0
3	B	25	0	11	1	0
3	C	25	0	11	1	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
4	C	1	0	0	0	0
5	A	6	0	8	0	0
5	B	12	0	16	2	0
5	C	6	0	8	0	0
6	A	15	0	0	0	0
6	B	25	0	0	2	0
6	C	10	0	0	0	0
7	A	189	0	0	2	0
7	B	202	0	0	4	0
7	C	126	0	0	1	0
All	All	7417	0	6638	69	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 69 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
5:B:405:GOL:H12	7:B:695:HOH:O	1.83	0.78
1:A:170:ARG:HD2	1:C:342:ARG:HE	1.57	0.69
1:C:266:PRO:HG3	1:C:321:LEU:HD22	1.81	0.63
1:A:312:GLY:HA2	2:A:402:NAG:H83	1.83	0.59
1:B:306:ASN:HD22	1:B:306:ASN:H	1.51	0.58

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	273/287 (95%)	269 (98%)	4 (2%)	0	100	100
1	B	272/287 (95%)	268 (98%)	4 (2%)	0	100	100
1	C	271/287 (94%)	265 (98%)	6 (2%)	0	100	100
All	All	816/861 (95%)	802 (98%)	14 (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	246/252 (98%)	240 (98%)	6 (2%)	61	73
1	B	245/252 (97%)	238 (97%)	7 (3%)	55	66
1	C	244/252 (97%)	234 (96%)	10 (4%)	41	49
All	All	735/756 (97%)	712 (97%)	23 (3%)	52	63

5 of 23 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	306	ASN
1	B	357	ASP
1	C	348	LYS
1	B	345	ARG
1	C	133	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	206	GLN
1	C	257	HIS
1	C	157	GLN
1	B	186	ASN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	306	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 ⓘ

6 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	GAL	A	401	2	12,12,12	0.46	0	17,17,17	2.38	7 (41%)
2	NAG	A	402	2	12,14,15	0.83	0	15,19,21	2.01	3 (20%)
2	GAL	B	401	2	12,12,12	0.59	0	17,17,17	1.32	1 (5%)
2	NAG	B	402	2	12,14,15	0.82	1 (8%)	15,19,21	1.67	2 (13%)
2	GAL	C	401	2	12,12,12	0.63	0	17,17,17	1.62	2 (11%)
2	NAG	C	402	2	12,14,15	0.71	0	15,19,21	1.21	3 (20%)

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	GAL	A	401	2	-	0/2/22/22	0/1/1/1
2	NAG	A	402	2	-	0/6/23/26	0/1/1/1
2	GAL	B	401	2	-	0/2/22/22	0/1/1/1

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	NAG	B	402	2	-	0/6/23/26	0/1/1/1
2	GAL	C	401	2	-	0/2/22/22	0/1/1/1
2	NAG	C	402	2	-	0/6/23/26	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	402	NAG	O5-C5	-2.29	1.41	1.45

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	401	GAL	O5-C5-C6	6.01	121.12	106.34
2	A	402	NAG	C3-C2-N2	-5.70	103.09	111.76
2	A	401	GAL	C1-O5-C5	-4.86	104.70	113.40
2	B	402	NAG	O5-C5-C6	4.55	111.75	106.98
2	C	401	GAL	O6-C6-C5	-4.22	96.84	111.36

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 5.6 Ligand geometry ⓘ

Of 20 ligands modelled in this entry, 3 are monoatomic - leaving 17 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
3	UDP	A	403	4	26,26,26	1.44	2 (7%)	36,40,40	1.40	5 (13%)
5	GOL	A	405	-	5,5,5	0.39	0	5,5,5	0.91	0
6	SO4	A	406	-	4,4,4	0.12	0	6,6,6	0.41	0
6	SO4	A	407	-	4,4,4	0.24	0	6,6,6	0.66	0
6	SO4	A	408	-	4,4,4	0.13	0	6,6,6	0.53	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
3	UDP	B	403	4	26,26,26	1.58	3 (11%)	36,40,40	2.10	8 (22%)
5	GOL	B	405	-	5,5,5	0.15	0	5,5,5	2.12	2 (40%)
5	GOL	B	406	-	5,5,5	0.26	0	5,5,5	1.43	1 (20%)
6	SO4	B	407	-	4,4,4	0.33	0	6,6,6	0.75	0
6	SO4	B	408	-	4,4,4	0.21	0	6,6,6	0.27	0
6	SO4	B	409	-	4,4,4	0.19	0	6,6,6	0.15	0
6	SO4	B	410	-	4,4,4	0.25	0	6,6,6	0.51	0
6	SO4	B	411	-	4,4,4	0.33	0	6,6,6	0.41	0
3	UDP	C	403	4	26,26,26	1.01	1 (3%)	36,40,40	1.77	5 (13%)
5	GOL	C	405	-	5,5,5	0.42	0	5,5,5	0.81	0
6	SO4	C	406	-	4,4,4	0.29	0	6,6,6	0.44	0
6	SO4	C	407	-	4,4,4	0.12	0	6,6,6	0.27	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	UDP	A	403	4	-	0/14/32/32	0/2/2/2
5	GOL	A	405	-	-	0/4/4/4	0/0/0/0
6	SO4	A	406	-	-	0/0/0/0	0/0/0/0
6	SO4	A	407	-	-	0/0/0/0	0/0/0/0
6	SO4	A	408	-	-	0/0/0/0	0/0/0/0
3	UDP	B	403	4	-	0/14/32/32	0/2/2/2
5	GOL	B	405	-	-	0/4/4/4	0/0/0/0
5	GOL	B	406	-	-	0/4/4/4	0/0/0/0
6	SO4	B	407	-	-	0/0/0/0	0/0/0/0
6	SO4	B	408	-	-	0/0/0/0	0/0/0/0
6	SO4	B	409	-	-	0/0/0/0	0/0/0/0
6	SO4	B	410	-	-	0/0/0/0	0/0/0/0
6	SO4	B	411	-	-	0/0/0/0	0/0/0/0
3	UDP	C	403	4	-	0/14/32/32	0/2/2/2
5	GOL	C	405	-	-	0/4/4/4	0/0/0/0
6	SO4	C	406	-	-	0/0/0/0	0/0/0/0
6	SO4	C	407	-	-	0/0/0/0	0/0/0/0

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	403	UDP	C2-N1	4.68	1.43	1.38

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	403	UDP	C6-C5	3.44	1.41	1.36
3	B	403	UDP	C6-C5	3.10	1.41	1.36
3	A	403	UDP	C6-N1	2.93	1.40	1.35
3	B	403	UDP	PB-O3A	2.59	1.64	1.60

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	403	UDP	N3-C2-N1	7.39	122.14	115.97
3	C	403	UDP	N3-C2-N1	6.63	121.50	115.97
3	C	403	UDP	C2-N1-C1'	4.70	121.16	118.21
3	B	403	UDP	O4'-C1'-N1	4.66	117.89	108.06
3	A	403	UDP	PA-O3A-PB	-4.33	118.98	131.68

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	273/287 (95%)	-0.46	1 (0%) 90 92	24, 33, 53, 68	0
1	B	273/287 (95%)	-0.48	1 (0%) 90 92	22, 30, 46, 63	0
1	C	273/287 (95%)	-0.11	6 (2%) 59 59	30, 42, 68, 79	1 (0%)
All	All	819/861 (95%)	-0.35	8 (0%) 81 80	22, 35, 57, 79	1 (0%)

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	347	LYS	3.7
1	C	126	SER	3.6
1	C	357	ASP	3.0
1	C	356	PHE	2.9
1	B	126	SER	2.9

### 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	GAL	B	401	12/12	0.35	40.93	49,81,88,89	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	GAL	C	401	12/12	0.32	24.02	69,91,92,93	0
2	GAL	A	401	12/12	0.36	11.71	56,87,90,91	0
2	NAG	A	402	14/15	0.11	0.61	37,41,44,47	0
2	NAG	B	402	14/15	0.09	0.36	28,32,35,37	0
2	NAG	C	402	14/15	0.09	-0.71	52,56,59,60	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( $\text{\AA}^2$ )	Q<0.9
6	SO4	A	408	5/5	0.24	19.69	79,82,83,84	0
6	SO4	B	411	5/5	0.16	6.62	76,77,79,80	0
5	GOL	A	405	6/6	0.15	5.62	41,44,52,58	0
5	GOL	B	405	6/6	0.13	4.41	31,41,47,51	0
5	GOL	B	406	6/6	0.18	2.76	34,36,38,43	0
5	GOL	C	405	6/6	0.18	2.05	68,70,71,71	0
6	SO4	B	407	5/5	0.19	1.93	80,83,85,86	0
3	UDP	B	403	25/25	0.14	1.39	23,30,86,89	0
6	SO4	B	410	5/5	0.15	1.33	58,63,64,66	0
6	SO4	C	406	5/5	0.14	0.96	54,56,60,61	0
6	SO4	B	409	5/5	0.13	0.69	85,85,86,86	0
6	SO4	A	406	5/5	0.15	0.67	70,70,75,78	0
3	UDP	A	403	25/25	0.13	0.28	27,34,87,88	0
6	SO4	B	408	5/5	0.13	0.27	68,68,71,71	0
3	UDP	C	403	25/25	0.16	0.12	41,49,98,99	0
6	SO4	C	407	5/5	0.14	-0.57	79,80,82,82	0
6	SO4	A	407	5/5	0.10	-0.75	69,70,72,73	0
4	MN	C	404	1/1	0.05	-3.08	61,61,61,61	0
4	MN	B	404	1/1	0.03	-4.10	40,40,40,40	0
4	MN	A	404	1/1	0.02	-7.86	50,50,50,50	0

## 6.5 Other polymers

There are no such residues in this entry.