



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 09:02 PM GMT

PDB ID : 4UO7  
Title : Structure of the A\_Canine\_Colorado\_17864\_06 H3 haemagglutinin in complex with 6SO4 Sialyl Lewis X  
Authors : Vachieri, S.G.; Collins, P.J.; Haire, L.F.; Ogradowicz, R.W.; Martin, S.R.; Walker, P.A.; Xiong, X.; Gamblin, S.J.; Skehel, J.J.  
Deposited on : 2014-05-31  
Resolution : 3.00 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 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
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) : trunk26865

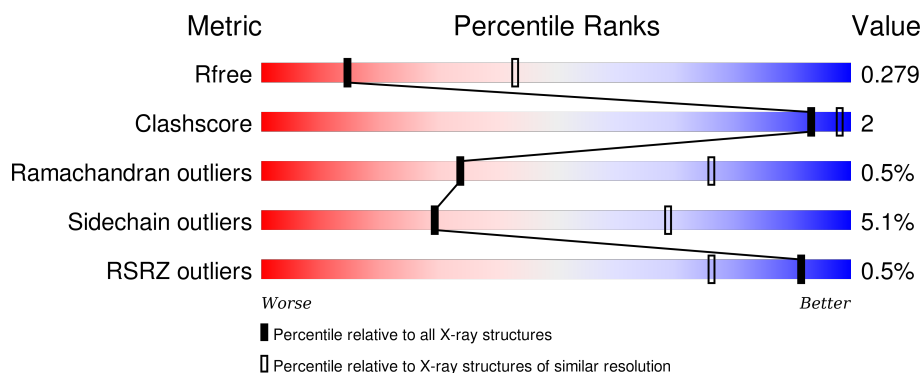
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.00 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))
$R_{free}$	91344	1578 (3.00-3.00)
Clashscore	102246	1912 (3.00-3.00)
Ramachandran outliers	100387	1853 (3.00-3.00)
Sidechain outliers	100360	1856 (3.00-3.00)
RSRZ outliers	91569	1592 (3.00-3.00)

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. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	328	<div> <div>%</div> <div> <div></div> <div>86%</div> <div>11%</div> <div>• •</div> </div> </div>
1	C	328	<div> <div></div> <div>88%</div> <div>9%</div> <div>• •</div> </div>
1	E	328	<div> <div></div> <div>86%</div> <div>10%</div> <div>• •</div> </div>
2	B	175	<div> <div>%</div> <div> <div></div> <div>97%</div> <div>• • •</div> </div> </div>
2	D	175	<div> <div>%</div> <div> <div></div> <div>94%</div> <div>5%</div> <div>•</div> </div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	F	175	 94% 6%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	NAG	A	612	X	-	-	-
3	NAG	E	641	-	-	-	X
6	GAL	A	702	X	-	-	-
6	FUC	A	704	X	-	-	-
6	FUC	C	704	X	-	-	-
6	FUC	E	704	X	-	-	-

## 2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 12430 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 HAEMAGGLUTININ HA1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	319	Total	C	N	O	S	0	0	0
			2458	1534	435	473	16			
1	C	319	Total	C	N	O	S	0	0	0
			2469	1541	437	475	16			
1	E	319	Total	C	N	O	S	0	0	0
			2471	1542	436	477	16			

- Molecule 2 is a protein called HAEMAGGLUTININ HA2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	174	Total	C	N	O	S	0	0	0
			1404	874	242	282	6			
2	D	175	Total	C	N	O	S	0	0	0
			1405	874	242	283	6			
2	F	174	Total	C	N	O	S	0	0	0
			1404	875	243	280	6			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	173	SER	-	EXPRESSION TAG	UNP E0UVR5
B	174	GLY	-	EXPRESSION TAG	UNP E0UVR5
B	175	ARG	-	EXPRESSION TAG	UNP E0UVR5
B	131	GLU	ASP	CONFLICT	UNP E0UVR5
D	173	SER	-	EXPRESSION TAG	UNP E0UVR5
D	174	GLY	-	EXPRESSION TAG	UNP E0UVR5
D	175	ARG	-	EXPRESSION TAG	UNP E0UVR5
D	131	GLU	ASP	CONFLICT	UNP E0UVR5
F	173	SER	-	EXPRESSION TAG	UNP E0UVR5
F	174	GLY	-	EXPRESSION TAG	UNP E0UVR5
F	175	ARG	-	EXPRESSION TAG	UNP E0UVR5
F	131	GLU	ASP	CONFLICT	UNP E0UVR5

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

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	2	Total	C	N	O	0	0
			28	16	2	10		
3	C	2	Total	C	N	O	0	0
			28	16	2	10		
3	C	2	Total	C	N	O	0	0
			28	16	2	10		
3	C	2	Total	C	N	O	0	0
			28	16	2	10		
3	E	2	Total	C	N	O	0	0
			28	16	2	10		
3	E	2	Total	C	N	O	0	0
			28	16	2	10		
3	E	2	Total	C	N	O	0	0
			28	16	2	10		

- Molecule 4 is a polymer of unknown type called SUGAR (3-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	3	Total	C	N	O	0	0
			39	22	2	15		
4	A	3	Total	C	N	O	0	0
			39	22	2	15		

- Molecule 5 is a polymer of unknown type called SUGAR (5-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	5	Total	C	N	O	0	0
			61	34	2	25		
5	C	5	Total	C	N	O	0	0
			61	34	2	25		
5	E	5	Total	C	N	O	0	0
			61	34	2	25		

- Molecule 6 is a polymer of unknown type called SUGAR (4-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
6	A	4	Total	C	N	O	S	0	0
			60	31	2	26	1		
6	C	4	Total	C	N	O	S	0	0
			60	31	2	26	1		

*Continued on next page...*

*Continued from previous page...*

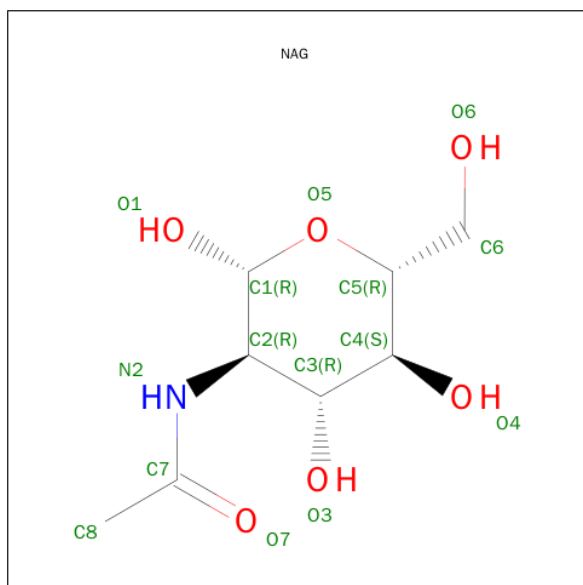
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
6	E	4	Total	C	N	O	S	0	0
			60	31	2	26	1		

- Molecule 7 is SULFATE ION (three-letter code: SO<sub>4</sub>) (formula: O<sub>4</sub>S).



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

- Molecule 8 is SUGAR (N-ACETYL-D-GLUCOSAMINE) (three-letter code: NAG) (formula: C<sub>8</sub>H<sub>15</sub>NO<sub>6</sub>).



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

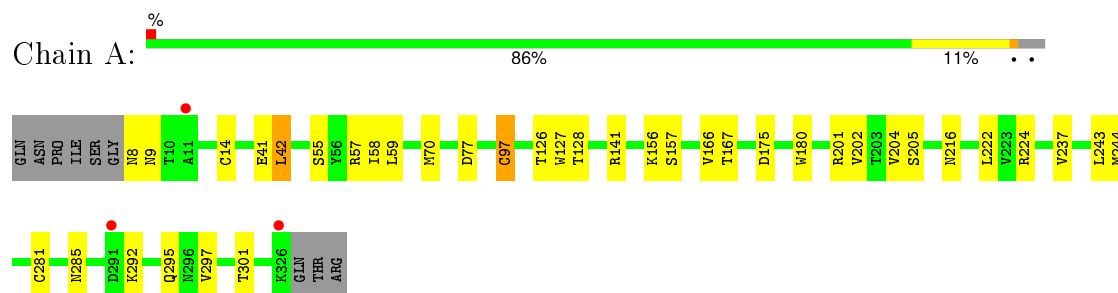
- Molecule 9 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	A	43	Total	O	0	0
			43	43		
9	B	12	Total	O	0	0
			12	12		
9	C	36	Total	O	0	0
			36	36		
9	D	14	Total	O	0	0
			14	14		
9	E	30	Total	O	0	0
			30	30		
9	F	3	Total	O	0	0
			3	3		

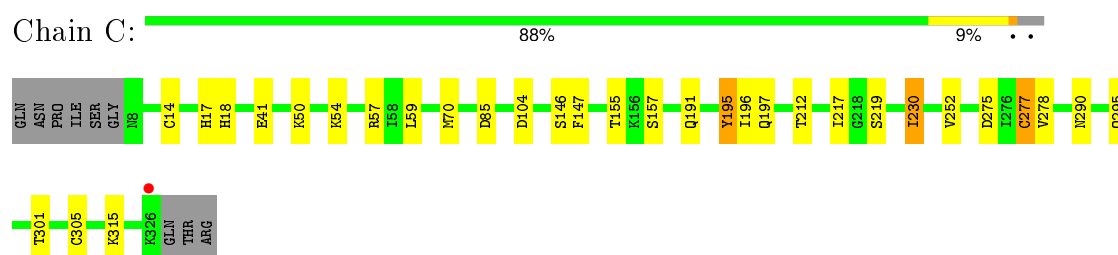
### 3 Residue-property plots [i](#)

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.

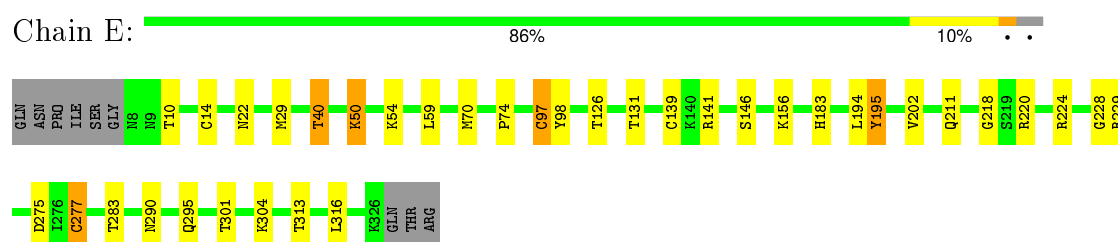
#### • Molecule 1: HAEMAGGLUTININ HA1



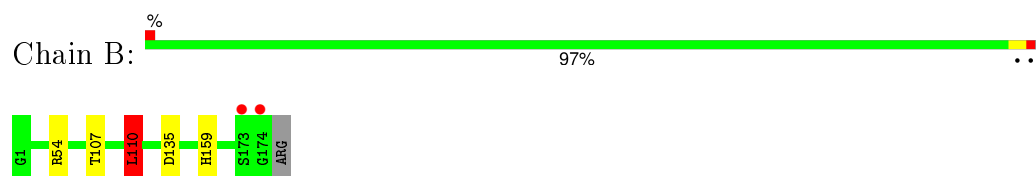
#### • Molecule 1: HAEMAGGLUTININ HA1



#### • Molecule 1: HAEMAGGLUTININ HA1

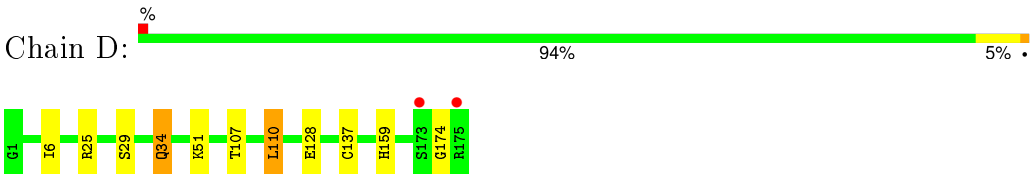


#### • Molecule 2: HAEMAGGLUTININ HA2

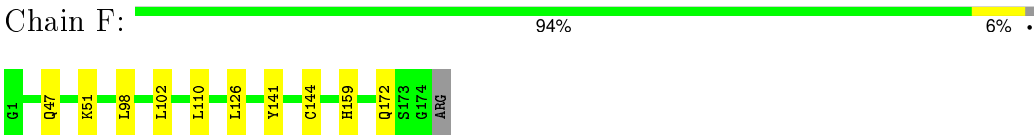


#### • Molecule 2: HAEMAGGLUTININ HA2





• Molecule 2: HAEMAGGLUTININ HA2



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	167.31Å 348.49Å 95.81Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	174.24 – 3.00 47.72 – 3.00	Depositor EDS
% Data completeness (in resolution range)	94.9 (174.24-3.00) 94.9 (47.72-3.00)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.07 (at 3.01Å)	Xtriage
Refinement program	REFMAC 5.8.0069	Depositor
R, $R_{free}$	0.233 , 0.276 0.234 , 0.279	Depositor DCC
$R_{free}$ test set	2728 reflections (5.36%)	DCC
Wilson B-factor (Å <sup>2</sup> )	58.1	Xtriage
Anisotropy	0.709	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 37.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	1 of 53624 reflections (0.002%)	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	12430	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	68.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.57% 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: BMA, NAG, SIA, GAL, FUC, NGS, MAN, 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	0.33	0/2509	0.56	0/3407
1	C	0.34	0/2520	0.56	0/3419
1	E	0.34	0/2522	0.57	0/3422
2	B	0.35	0/1429	0.53	1/1924 (0.1%)
2	D	0.35	0/1430	0.54	0/1927
2	F	0.33	0/1429	0.53	0/1923
All	All	0.34	0/11839	0.55	1/16022 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
3	A	1	0
6	A	2	0
6	C	1	0
6	E	1	0
All	All	5	0

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	110	LEU	CA-CB-CG	5.19	127.24	115.30

All (5) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
3	A	612	NAG	C1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atom
6	A	702	GAL	C1
6	A	704	FUC	C1
6	C	704	FUC	C1
6	E	704	FUC	C1

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	2458	0	2395	14	0
1	C	2469	0	2417	11	0
1	E	2471	0	2415	13	0
2	B	1404	0	1309	3	0
2	D	1405	0	1300	4	0
2	F	1404	0	1316	4	0
3	A	28	0	25	0	0
3	C	84	0	75	0	0
3	E	84	0	75	0	0
4	A	78	0	68	3	0
5	A	61	0	52	0	0
5	C	61	0	52	0	0
5	E	61	0	52	0	0
6	A	60	0	49	0	0
6	C	60	0	49	0	0
6	E	60	0	49	2	0
7	A	5	0	0	0	0
7	B	5	0	0	0	0
7	C	10	0	0	0	0
7	D	5	0	0	0	0
7	F	5	0	0	0	0
8	C	14	0	13	0	0
9	A	43	0	0	0	1
9	B	12	0	0	0	0
9	C	36	0	0	0	0
9	D	14	0	0	0	0
9	E	30	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
9	F	3	0	0	0	0
All	All	12430	0	11711	46	1

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 2.

The worst 5 of 46 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:C:230:ILE:HD11	1:C:252:VAL:HG21	1.44	0.99
1:A:97:CYS:O	1:A:224:ARG:NH1	2.14	0.80
1:A:285:ASN:HD21	4:A:641:NAG:C1	2.09	0.65
1:A:285:ASN:ND2	4:A:641:NAG:C1	2.62	0.63
1:A:180:TRP:CE2	1:A:204:VAL:HG21	2.35	0.60

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	Interatomic distance (Å)	Clash overlap (Å)
9:A:2011:HOH:O	9:A:2011:HOH:O[3_554]	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	317/328 (97%)	304 (96%)	12 (4%)	1 (0%)	46	84
1	C	317/328 (97%)	292 (92%)	24 (8%)	1 (0%)	46	84
1	E	317/328 (97%)	297 (94%)	18 (6%)	2 (1%)	30	72
2	B	172/175 (98%)	164 (95%)	8 (5%)	0	100	100
2	D	173/175 (99%)	163 (94%)	8 (5%)	2 (1%)	16	56

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	F	172/175 (98%)	158 (92%)	13 (8%)	1 (1%)	30	72
All	All	1468/1509 (97%)	1378 (94%)	83 (6%)	7 (0%)	34	76

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	196	ILE
1	A	55	SER
1	E	277	CYS
1	E	304	LYS
2	F	172	GLN

### 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	276/289 (96%)	258 (94%)	18 (6%)	21	58
1	C	279/289 (96%)	263 (94%)	16 (6%)	25	64
1	E	279/289 (96%)	260 (93%)	19 (7%)	20	56
2	B	145/148 (98%)	142 (98%)	3 (2%)	61	89
2	D	144/148 (97%)	139 (96%)	5 (4%)	43	80
2	F	145/148 (98%)	141 (97%)	4 (3%)	51	84
All	All	1268/1311 (97%)	1203 (95%)	65 (5%)	29	69

5 of 65 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	157	SER
1	C	315	LYS
1	E	313	THR
1	C	195	TYR
1	C	277	CYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	296	ASN
1	E	8	ASN
2	F	116	ASN
1	C	80	GLN
1	E	296	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules 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 ⓘ

47 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$
3	NAG	A	611	1,3	14,14,15	0.47	0	15,19,21	2.47	3 (20%)
3	NAG	A	612	3	14,14,15	0.91	1 (7%)	15,19,21	1.87	2 (13%)
4	NAG	A	621	1,4	14,14,15	0.50	0	15,19,21	1.02	0
4	NAG	A	622	4	14,14,15	0.66	0	15,19,21	1.17	1 (6%)
4	BMA	A	623	4	11,11,12	0.38	0	14,15,17	1.11	1 (7%)
5	NAG	A	631	1,5	14,14,15	0.57	0	15,19,21	1.36	3 (20%)
5	NAG	A	632	5	14,14,15	0.67	0	15,19,21	0.80	0
5	BMA	A	633	5	11,11,12	0.29	0	14,15,17	0.85	0
5	MAN	A	634	5	11,11,12	0.53	0	14,15,17	1.05	2 (14%)
5	MAN	A	637	5	11,11,12	0.71	0	14,15,17	1.78	3 (21%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
4	NAG	A	641	4	14,14,15	0.43	0	15,19,21	1.15	1 (6%)
4	NAG	A	642	4	14,14,15	0.75	0	15,19,21	2.45	5 (33%)
4	BMA	A	643	4	11,11,12	0.41	0	14,15,17	1.11	2 (14%)
6	SIA	A	701	6	16,20,21	0.32	0	18,28,31	1.11	2 (11%)
6	GAL	A	702	6	11,11,12	0.77	0	14,15,17	2.44	3 (21%)
6	NGS	A	703	6	19,19,19	0.91	1 (5%)	22,28,28	1.59	4 (18%)
6	FUC	A	704	6	10,10,11	0.76	0	14,14,16	1.90	3 (21%)
3	NAG	C	611	1,3	14,14,15	0.59	0	15,19,21	0.73	0
3	NAG	C	612	3	14,14,15	0.66	0	15,19,21	0.99	1 (6%)
3	NAG	C	621	1,3	14,14,15	0.52	0	15,19,21	0.95	1 (6%)
3	NAG	C	622	3	14,14,15	0.69	0	15,19,21	2.04	5 (33%)
5	NAG	C	631	1,5	14,14,15	0.49	0	15,19,21	1.22	2 (13%)
5	NAG	C	632	5	14,14,15	0.53	0	15,19,21	0.83	0
5	BMA	C	633	5	11,11,12	0.43	0	14,15,17	1.08	1 (7%)
5	MAN	C	634	5	11,11,12	0.60	0	14,15,17	1.03	1 (7%)
5	MAN	C	637	5	11,11,12	0.50	0	14,15,17	0.84	0
3	NAG	C	641	1,3	14,14,15	0.70	0	15,19,21	1.34	2 (13%)
3	NAG	C	642	3	14,14,15	0.48	0	15,19,21	0.98	0
6	SIA	C	701	6	16,20,21	0.35	0	18,28,31	1.01	1 (5%)
6	GAL	C	702	6	11,11,12	0.57	0	14,15,17	1.44	2 (14%)
6	NGS	C	703	6	19,19,19	0.88	1 (5%)	22,28,28	1.45	4 (18%)
6	FUC	C	704	6	10,10,11	0.85	0	14,14,16	1.44	2 (14%)
3	NAG	E	611	1,3	14,14,15	0.46	0	15,19,21	1.28	1 (6%)
3	NAG	E	612	3	14,14,15	0.52	0	15,19,21	0.96	1 (6%)
3	NAG	E	621	1,3	14,14,15	0.61	0	15,19,21	0.85	1 (6%)
3	NAG	E	622	3	14,14,15	0.67	0	15,19,21	1.36	2 (13%)
5	NAG	E	631	1,5	14,14,15	0.63	0	15,19,21	1.33	2 (13%)
5	NAG	E	632	5	14,14,15	0.57	0	15,19,21	0.95	0
5	BMA	E	633	5	11,11,12	0.29	0	14,15,17	2.01	3 (21%)
5	MAN	E	634	5	11,11,12	0.62	0	14,15,17	1.55	2 (14%)
5	MAN	E	637	5	11,11,12	0.58	0	14,15,17	0.88	1 (7%)
3	NAG	E	641	1,3	14,14,15	0.62	0	15,19,21	1.06	0
3	NAG	E	642	3	14,14,15	0.78	0	15,19,21	1.57	2 (13%)
6	SIA	E	701	6	16,20,21	0.36	0	18,28,31	1.14	1 (5%)
6	GAL	E	702	6	11,11,12	0.62	0	14,15,17	2.07	2 (14%)
6	NGS	E	703	6	19,19,19	0.89	1 (5%)	22,28,28	1.21	1 (4%)



Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
6	FUC	E	704	6	10,10,11	0.77	0	14,14,16	1.79	3 (21%)

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	NAG	A	611	1,3	-	0/6/23/26	0/1/1/1
3	NAG	A	612	3	1/1/5/7	0/6/23/26	0/1/1/1
4	NAG	A	621	1,4	-	0/6/23/26	0/1/1/1
4	NAG	A	622	4	-	0/6/23/26	0/1/1/1
4	BMA	A	623	4	-	0/2/19/22	0/1/1/1
5	NAG	A	631	1,5	-	0/6/23/26	0/1/1/1
5	NAG	A	632	5	-	0/6/23/26	0/1/1/1
5	BMA	A	633	5	-	0/2/19/22	0/1/1/1
5	MAN	A	634	5	-	0/2/19/22	0/1/1/1
5	MAN	A	637	5	-	0/2/19/22	0/1/1/1
4	NAG	A	641	4	-	0/6/23/26	0/1/1/1
4	NAG	A	642	4	-	0/6/23/26	0/1/1/1
4	BMA	A	643	4	-	0/2/19/22	0/1/1/1
6	SIA	A	701	6	-	0/14/34/38	0/1/1/1
6	GAL	A	702	6	1/1/4/5	0/2/19/22	0/1/1/1
6	NGS	A	703	6	-	0/10/30/30	0/1/1/1
6	FUC	A	704	6	1/1/4/5	0/0/17/20	0/1/1/1
3	NAG	C	611	1,3	-	0/6/23/26	0/1/1/1
3	NAG	C	612	3	-	0/6/23/26	0/1/1/1
3	NAG	C	621	1,3	-	0/6/23/26	0/1/1/1
3	NAG	C	622	3	-	0/6/23/26	0/1/1/1
5	NAG	C	631	1,5	-	0/6/23/26	0/1/1/1
5	NAG	C	632	5	-	0/6/23/26	0/1/1/1
5	BMA	C	633	5	-	0/2/19/22	0/1/1/1
5	MAN	C	634	5	-	0/2/19/22	1/1/1/1
5	MAN	C	637	5	-	0/2/19/22	1/1/1/1
3	NAG	C	641	1,3	-	0/6/23/26	0/1/1/1
3	NAG	C	642	3	-	0/6/23/26	0/1/1/1
6	SIA	C	701	6	-	0/14/34/38	0/1/1/1
6	GAL	C	702	6	-	0/2/19/22	0/1/1/1
6	NGS	C	703	6	-	0/10/30/30	0/1/1/1
6	FUC	C	704	6	1/1/4/5	0/0/17/20	0/1/1/1
3	NAG	E	611	1,3	-	0/6/23/26	0/1/1/1
3	NAG	E	612	3	-	0/6/23/26	0/1/1/1

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	NAG	E	621	1,3	-	0/6/23/26	0/1/1/1
3	NAG	E	622	3	-	0/6/23/26	0/1/1/1
5	NAG	E	631	1,5	-	0/6/23/26	0/1/1/1
5	NAG	E	632	5	-	0/6/23/26	0/1/1/1
5	BMA	E	633	5	-	0/2/19/22	1/1/1/1
5	MAN	E	634	5	-	0/2/19/22	0/1/1/1
5	MAN	E	637	5	-	0/2/19/22	0/1/1/1
3	NAG	E	641	1,3	-	0/6/23/26	0/1/1/1
3	NAG	E	642	3	-	0/6/23/26	0/1/1/1
6	SIA	E	701	6	-	0/14/34/38	0/1/1/1
6	GAL	E	702	6	-	0/2/19/22	0/1/1/1
6	NGS	E	703	6	-	0/10/30/30	0/1/1/1
6	FUC	E	704	6	1/1/4/5	0/0/17/20	0/1/1/1

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	612	NAG	O5-C1	-2.85	1.38	1.43
6	A	703	NGS	O8-S	2.65	1.64	1.50
6	C	703	NGS	O8-S	2.65	1.64	1.50
6	E	703	NGS	O8-S	2.66	1.64	1.50

The worst 5 of 79 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	E	704	FUC	C1-C2-C3	-5.11	103.49	109.54
3	A	612	NAG	C1-O5-C5	-5.04	105.85	112.25
6	A	704	FUC	C1-C2-C3	-4.50	104.22	109.54
3	A	611	NAG	C4-C3-C2	-3.79	105.34	111.23
3	A	612	NAG	C4-C3-C2	-3.14	106.35	111.23

All (5) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
6	A	704	FUC	C1
6	E	704	FUC	C1
6	A	702	GAL	C1
6	C	704	FUC	C1
3	A	612	NAG	C1

There are no torsion outliers.

All (3) ring outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	E	633	BMA	C1-C2-C3-C4-C5-O5
5	C	634	MAN	C1-C2-C3-C4-C5-O5
5	C	637	MAN	C1-C2-C3-C4-C5-O5

2 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	641	NAG	3	0
6	E	701	SIA	2	0

## 5.6 Ligand geometry [i](#)

7 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	$\# Z  > 2$
7	SO4	A	1327	-	4,4,4	0.40	0	6,6,6	0.13	0
7	SO4	B	1175	-	4,4,4	0.40	0	6,6,6	0.17	0
7	SO4	C	1327	-	4,4,4	0.41	0	6,6,6	0.17	0
7	SO4	C	1328	-	4,4,4	0.37	0	6,6,6	0.10	0
8	NAG	C	601	1	14,14,15	0.41	0	15,19,21	2.29	1 (6%)
7	SO4	D	1176	-	4,4,4	0.39	0	6,6,6	0.14	0
7	SO4	F	1175	-	4,4,4	0.40	0	6,6,6	0.12	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
7	SO4	A	1327	-	-	0/0/0/0	0/0/0/0
7	SO4	B	1175	-	-	0/0/0/0	0/0/0/0
7	SO4	C	1327	-	-	0/0/0/0	0/0/0/0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
7	SO4	C	1328	-	-	0/0/0/0	0/0/0/0
8	NAG	C	601	1	-	0/6/23/26	0/1/1/1
7	SO4	D	1176	-	-	0/0/0/0	0/0/0/0
7	SO4	F	1175	-	-	0/0/0/0	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(°)
8	C	601	NAG	C1-O5-C5	8.16	122.61	112.25

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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 [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

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	319/328 (97%)	-0.27	3 (0%) 85 64	41, 63, 84, 129	0
1	C	319/328 (97%)	-0.27	1 (0%) 94 84	40, 63, 87, 126	0
1	E	319/328 (97%)	-0.27	0 100 100	41, 63, 85, 127	0
2	B	174/175 (99%)	-0.09	2 (1%) 82 58	43, 79, 97, 116	0
2	D	175/175 (100%)	-0.16	2 (1%) 82 58	40, 74, 96, 130	0
2	F	174/175 (99%)	-0.20	0 100 100	40, 75, 99, 108	0
All	All	1480/1509 (98%)	-0.22	8 (0%) 91 76	40, 67, 95, 130	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	11	ALA	2.6
2	B	174	GLY	2.5
2	B	173	SER	2.3
2	D	173	SER	2.3
2	D	175	ARG	2.2

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates [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( $\text{\AA}^2$ )	Q<0.9
3	NAG	E	641	14/15	0.81	0.35	6.06	84,91,96,102	0
3	NAG	C	641	14/15	0.85	0.31	1.98	76,82,89,95	0
6	SIA	A	701	20/21	0.94	0.30	1.38	66,68,77,79	0
6	SIA	E	701	20/21	0.92	0.23	0.47	67,72,81,84	0
6	SIA	C	701	20/21	0.95	0.24	-0.07	68,72,78,79	0
5	NAG	A	632	14/15	0.92	0.20	-0.23	73,79,86,89	0
5	NAG	C	632	14/15	0.93	0.18	-1.27	69,79,86,90	0
5	NAG	E	632	14/15	0.94	0.17	-2.31	72,79,86,93	0
5	NAG	E	631	14/15	0.94	0.21	-	67,70,74,74	0
5	NAG	A	631	14/15	0.95	0.15	-	69,73,75,77	0
5	MAN	A	634	11/12	0.89	0.28	-	105,110,116,116	0
5	MAN	C	634	11/12	0.62	0.33	-	104,112,119,122	0
5	MAN	E	634	11/12	0.75	0.47	-	123,128,133,135	0
3	NAG	E	622	14/15	0.87	0.31	-	95,106,110,110	0
3	NAG	C	621	14/15	0.87	0.21	-	84,89,92,93	0
5	BMA	A	633	11/12	0.85	0.17	-	94,99,106,106	0
6	FUC	E	704	10/11	0.94	0.24	-	84,88,92,93	0
3	NAG	E	642	14/15	0.56	0.51	-	90,109,112,113	0
6	GAL	C	702	11/12	0.92	0.24	-	83,86,91,92	0
6	GAL	A	702	11/12	0.91	0.24	-	77,80,84,87	0
5	NAG	C	631	14/15	0.96	0.14	-	71,75,76,77	0
3	NAG	C	611	14/15	0.84	0.28	-	88,94,101,106	0
4	NAG	A	621	14/15	0.85	0.19	-	89,94,101,109	0
4	NAG	A	642	14/15	0.82	0.43	-	110,121,129,130	0
3	NAG	E	612	14/15	0.65	0.62	-	125,130,133,134	0
3	NAG	C	612	14/15	0.73	0.41	-	105,113,120,123	0
6	GAL	E	702	11/12	0.93	0.18	-	80,83,87,87	0
6	FUC	A	704	10/11	0.88	0.30	-	89,90,91,92	0
6	NGS	A	703	19/19	0.89	0.23	-	82,89,99,102	0
5	MAN	C	637	11/12	0.80	0.33	-	115,117,123,124	0
4	NAG	A	622	14/15	0.82	0.35	-	112,122,128,129	0
3	NAG	E	621	14/15	0.82	0.21	-	92,96,102,104	0
6	FUC	C	704	10/11	0.87	0.34	-	95,99,100,101	0
5	BMA	E	633	11/12	0.82	0.26	-	102,110,115,118	0
5	BMA	C	633	11/12	0.88	0.19	-	95,101,106,108	0
4	BMA	A	623	11/12	0.68	0.36	-	118,126,129,130	0
3	NAG	C	642	14/15	0.83	0.38	-	89,100,106,108	0
4	NAG	A	641	14/15	0.83	0.34	-	93,98,104,111	0
5	MAN	A	637	11/12	0.89	0.19	-	100,107,113,116	0
3	NAG	A	612	14/15	0.82	0.26	-	110,118,127,127	0
6	NGS	C	703	19/19	0.88	0.28	-	97,102,111,113	0
5	MAN	E	637	11/12	0.79	0.22	-	113,118,123,123	0
3	NAG	C	622	14/15	0.85	0.22	-	90,97,99,100	0

Continued on next page...

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
4	BMA	A	643	11/12	0.72	0.39	-	127,133,138,139	0
3	NAG	E	611	14/15	0.86	0.33	-	95,101,110,119	0
6	NGS	E	703	19/19	0.90	0.19	-	89,96,106,107	0
3	NAG	A	611	14/15	0.84	0.23	-	85,92,101,109	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	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
7	SO4	D	1176	5/5	0.95	0.18	-0.77	78,79,80,83	0
7	SO4	F	1175	5/5	0.95	0.15	-1.05	85,85,89,90	0
8	NAG	C	601	14/15	0.86	0.21	-	76,86,91,92	0
7	SO4	B	1175	5/5	0.89	0.21	-	85,89,90,96	0
7	SO4	A	1327	5/5	0.84	0.22	-	130,131,134,135	0
7	SO4	C	1327	5/5	0.81	0.27	-	120,121,124,125	0
7	SO4	C	1328	5/5	0.66	0.40	-	137,138,140,141	0

## 6.5 Other polymers

There are no such residues in this entry.