



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 14, 2017 – 03:02 am GMT

PDB ID : 4C7W  
Title : Crystal structure of Mouse Hepatitis virus strain S Hemagglutinin- esterase in complex with 4-O-acetylated sialic acid  
Authors : Zeng, Q.H.; Huizinga, E.G.  
Deposited on : 2013-09-26  
Resolution : 2.50 Å(reported)

This is a Full wwPDB X-ray Structure 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/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 : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : trunk28620  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
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) : recalc28949

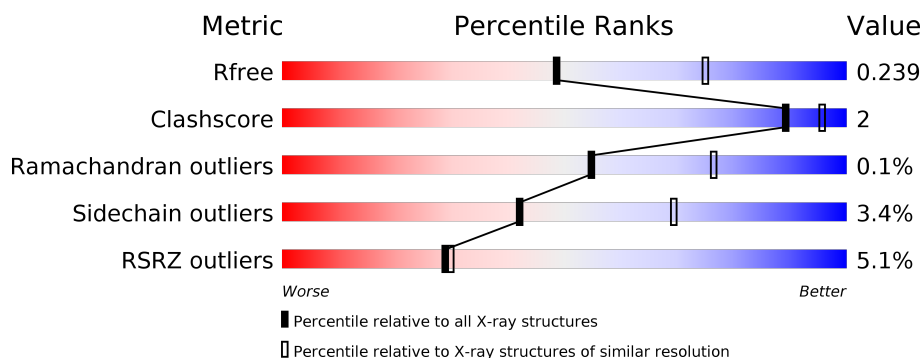
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.50 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))
$R_{free}$	100719	3846 (2.50-2.50)
Clashscore	112137	4554 (2.50-2.50)
Ramachandran outliers	110173	4463 (2.50-2.50)
Sidechain outliers	110143	4465 (2.50-2.50)
RSRZ outliers	101464	3876 (2.50-2.50)

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	386	<div> <div>3%</div> <div>85%</div> <div>6%</div> <div>9%</div> </div>
1	B	386	<div> <div>6%</div> <div>79%</div> <div>7%</div> <div>13%</div> </div>

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
5	NAG	A	1962	X	-	-	-
7	GOL	B	803	-	-	-	X

## 2 Entry composition

There are 9 unique types of molecules in this entry. The entry contains 5958 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.

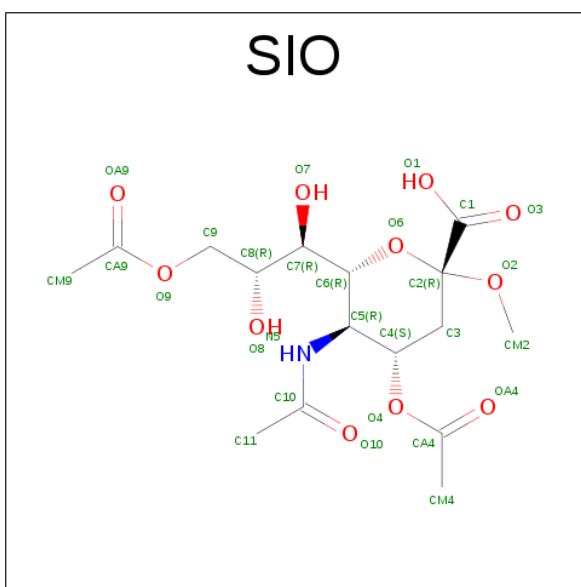
- Molecule 1 is a protein called HEMAGGLUTININ-ESTERASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	353	Total	C	N	O	S	0	1	0
			2808	1797	460	531	20			
1	B	334	Total	C	N	O	S	0	1	0
			2649	1693	432	505	19			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	45	ALA	SER	ENGINEERED MUTATION	UNP O55252
A	404	SER	-	EXPRESSION TAG	UNP O55252
A	405	ASP	-	EXPRESSION TAG	UNP O55252
A	406	PRO	-	EXPRESSION TAG	UNP O55252
A	407	LEU	-	EXPRESSION TAG	UNP O55252
A	408	VAL	-	EXPRESSION TAG	UNP O55252
A	409	PRO	-	EXPRESSION TAG	UNP O55252
A	410	ARG	-	EXPRESSION TAG	UNP O55252
B	45	ALA	SER	ENGINEERED MUTATION	UNP O55252
B	404	SER	-	EXPRESSION TAG	UNP O55252
B	405	ASP	-	EXPRESSION TAG	UNP O55252
B	406	PRO	-	EXPRESSION TAG	UNP O55252
B	407	LEU	-	EXPRESSION TAG	UNP O55252
B	408	VAL	-	EXPRESSION TAG	UNP O55252
B	409	PRO	-	EXPRESSION TAG	UNP O55252
B	410	ARG	-	EXPRESSION TAG	UNP O55252

- Molecule 2 is SUGAR (METHYL 4,9-DI-O-ACETYL-5-(ACETYLAMINO)-3,5-DIDEOXY-D-GLYCERO-ALPHA-D-GALACTO-NON-2-ULOPYRANOSIDONIC ACID) (three-letter code: SIO) (formula: C<sub>16</sub>H<sub>25</sub>NO<sub>11</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			25	14	1	10		
2	B	1	Total	C	N	O	0	0
			25	14	1	10		

- Molecule 3 is POTASSIUM ION (three-letter code: K) (formula: K).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total	K	0	0
			1	1		
3	A	1	Total	K	0	0
			1	1		

- Molecule 4 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
4	A	1	Total	C	N	O	0	0
			14	8	1	5		
4	A	1	Total	C	N	O	0	0
			14	8	1	5		
4	B	1	Total	C	N	O	0	0
			14	8	1	5		
4	B	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 5 is a polymer of unknown type called SUGAR (2-MER).

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

- Molecule 6 is a polymer of unknown type called SUGAR (3-MER).

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

*Continued on next page...*

Continued from previous page...

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

- Molecule 7 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



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

- Molecule 8 is a polymer of unknown type called SUGAR (7-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
8	B	7	Total	C	N	O	0	0
			83	46	2	35		

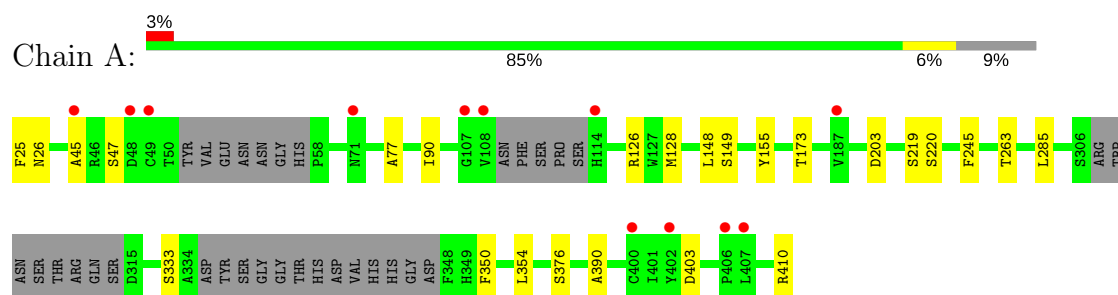
- Molecule 9 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
9	A	51	Total	O	0	0
			51	51		
9	B	24	Total	O	0	0
			24	24		

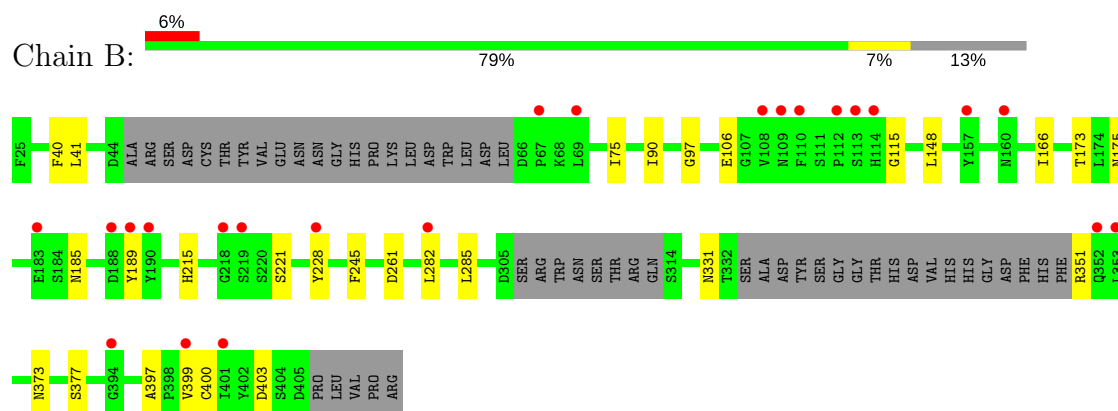
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes 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: HEMAGGLUTININ-ESTERASE



#### • Molecule 1: HEMAGGLUTININ-ESTERASE





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	91.57Å 106.59Å 135.61Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	83.92 – 2.50 54.49 – 2.50	Depositor EDS
% Data completeness (in resolution range)	100.0 (83.92-2.50) 100.0 (54.49-2.50)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.00 (at 2.51Å)	Xtriage
Refinement program	REFMAC 5.5.0110	Depositor
R, $R_{free}$	0.214 , 0.249 0.210 , 0.239	Depositor DCC
$R_{free}$ test set	2356 reflections (5.32%)	DCC
Wilson B-factor (Å <sup>2</sup> )	49.7	Xtriage
Anisotropy	0.056	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 43.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	5958	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	52.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.49% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, BMA, NAG, K, SIO, MAN

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	0/2885	0.59	0/3912
1	B	0.46	0/2721	0.55	0/3691
All	All	0.49	0/5606	0.57	0/7603

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
5	A	1	0

There are no bond length outliers.

There are no bond angle outliers.

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
5	A	1962	NAG	C1

There are no planarity outliers.

### 5.2 Too-close contacts [i](#)

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	2808	0	2639	14	0
1	B	2649	0	2481	11	0
2	A	25	0	21	0	0
2	B	25	0	21	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	28	0	26	0	0
4	B	28	0	26	0	0
5	A	56	0	50	0	0
5	B	56	0	50	0	0
6	A	78	0	68	0	0
6	B	39	0	34	0	0
7	B	6	0	8	0	0
8	B	83	0	70	0	0
9	A	51	0	0	0	0
9	B	24	0	0	0	0
All	All	5958	0	5494	25	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 2.

All (25) 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:45:ALA:HB3	1:A:350:PHE:CG	2.27	0.69
1:A:128[B]:MET:CE	1:A:128[B]:MET:HA	2.24	0.67
1:B:90:ILE:HD13	1:B:173:THR:O	1.95	0.66
1:A:47:SER:HB3	1:A:77:ALA:HB3	1.80	0.64
1:A:155:TYR:HB3	1:A:285:LEU:HD11	1.82	0.62
1:A:128[B]:MET:HA	1:A:128[B]:MET:HE2	1.84	0.58
1:B:90:ILE:HD11	1:B:175:ASN:HB3	1.85	0.57
1:B:41:LEU:HD23	1:B:75:ILE:HD12	1.87	0.55
1:A:45:ALA:CB	1:A:350:PHE:CG	2.92	0.52
1:B:90:ILE:HD12	1:B:90:ILE:N	2.27	0.49
1:A:155:TYR:CB	1:A:285:LEU:HD11	2.43	0.49
1:B:148:LEU:HD23	1:B:148:LEU:C	2.33	0.48
1:B:189:TYR:HB2	1:B:228:TYR:CE1	2.48	0.48
1:B:40:PHE:CE2	1:B:97:GLY:HA3	2.49	0.48
1:A:45:ALA:HB3	1:A:350:PHE:CD1	2.49	0.47
1:A:90:ILE:HD12	1:A:173:THR:HB	1.97	0.47
1:A:148:LEU:HD23	1:A:148:LEU:C	2.35	0.46
1:B:90:ILE:CD1	1:B:90:ILE:N	2.79	0.45

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:376:SER:O	1:A:390:ALA:HA	2.17	0.43
1:B:373:ASN:C	1:B:373:ASN:OD1	2.57	0.42
1:A:25:PHE:CE2	1:A:26:ASN:HB2	2.54	0.42
1:B:397:ALA:HB3	1:B:400:CYS:SG	2.60	0.42
1:B:166:ILE:HD12	1:B:282:LEU:CD2	2.50	0.41
1:A:148:LEU:HD23	1:A:149:SER:N	2.36	0.41
1:A:126:ARG:HD2	1:A:263:THR:HG22	2.03	0.40

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	344/386 (89%)	333 (97%)	11 (3%)	0	100	100
1	B	327/386 (85%)	306 (94%)	20 (6%)	1 (0%)	44	66
All	All	671/772 (87%)	639 (95%)	31 (5%)	1 (0%)	55	76

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	115	GLY

### 5.3.2 Protein sidechains [i](#)

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	308/336 (92%)	300 (97%)	8 (3%)	51	78
1	B	291/336 (87%)	279 (96%)	12 (4%)	35	61
All	All	599/672 (89%)	579 (97%)	20 (3%)	42	70

All (20) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	203	ASP
1	A	219	SER
1	A	220	SER
1	A	245	PHE
1	A	333	SER
1	A	354	LEU
1	A	403	ASP
1	A	410	ARG
1	B	106	GLU
1	B	185	ASN
1	B	215	HIS
1	B	221	SER
1	B	245	PHE
1	B	261	ASP
1	B	285	LEU
1	B	331	ASN
1	B	351	ARG
1	B	377	SER
1	B	399	VAL
1	B	403	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	71	ASN
1	A	366	GLN
1	B	185	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 ⓘ

24 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	$\# Z  > 2$
5	NAG	A	1961	1,5	14,14,15	0.53	0	15,19,21	1.44	2 (13%)
5	NAG	A	1962	5	14,14,15	0.71	0	15,19,21	2.23	3 (20%)
6	NAG	A	2461	1,6	14,14,15	0.60	0	15,19,21	1.47	1 (6%)
6	NAG	A	2462	6	14,14,15	0.44	0	15,19,21	1.32	1 (6%)
6	BMA	A	2463	6	11,11,12	0.49	0	13,15,17	0.99	1 (7%)
5	NAG	A	3601	1,5	14,14,15	0.49	0	15,19,21	0.89	1 (6%)
5	NAG	A	3602	5	14,14,15	0.53	0	15,19,21	1.43	2 (13%)
6	NAG	A	3741	1,6	14,14,15	0.51	0	15,19,21	1.25	2 (13%)
6	NAG	A	3742	6	14,14,15	0.59	0	15,19,21	0.70	0
6	BMA	A	3743	6	11,11,12	0.48	0	13,15,17	0.76	0
6	NAG	B	2461	1,6	14,14,15	0.47	0	15,19,21	1.45	1 (6%)
6	NAG	B	2462	6	14,14,15	0.57	0	15,19,21	0.85	0
6	BMA	B	2463	6	11,11,12	0.51	0	13,15,17	0.58	0
5	NAG	B	3601	1,5	14,14,15	0.57	0	15,19,21	0.85	0
5	NAG	B	3602	5	14,14,15	0.48	0	15,19,21	1.05	2 (13%)
8	NAG	B	3741	1,8	14,14,15	0.48	0	15,19,21	0.92	0
8	NAG	B	3742	8	14,14,15	0.86	1 (7%)	15,19,21	1.08	1 (6%)
8	BMA	B	3743	8	11,11,12	0.37	0	13,15,17	0.82	0
8	MAN	B	3744	8	11,11,12	0.79	0	13,15,17	0.91	1 (7%)
8	MAN	B	3745	8	11,11,12	0.77	0	13,15,17	0.90	0
8	MAN	B	3746	8	11,11,12	0.71	0	13,15,17	0.62	0
8	MAN	B	3747	8	11,11,12	0.65	0	13,15,17	1.14	3 (23%)
5	NAG	B	941	1,5	14,14,15	0.50	0	15,19,21	0.78	0
5	NAG	B	942	5	14,14,15	0.45	0	15,19,21	0.82	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
5	NAG	A	1961	1,5	-	0/6/23/26	0/1/1/1
5	NAG	A	1962	5	1/1/5/7	0/6/23/26	0/1/1/1
6	NAG	A	2461	1,6	-	0/6/23/26	0/1/1/1
6	NAG	A	2462	6	-	0/6/23/26	0/1/1/1
6	BMA	A	2463	6	-	0/2/19/22	0/1/1/1
5	NAG	A	3601	1,5	-	0/6/23/26	0/1/1/1
5	NAG	A	3602	5	-	0/6/23/26	0/1/1/1
6	NAG	A	3741	1,6	-	0/6/23/26	0/1/1/1
6	NAG	A	3742	6	-	0/6/23/26	0/1/1/1
6	BMA	A	3743	6	-	0/2/19/22	0/1/1/1
6	NAG	B	2461	1,6	-	0/6/23/26	0/1/1/1
6	NAG	B	2462	6	-	0/6/23/26	0/1/1/1
6	BMA	B	2463	6	-	0/2/19/22	0/1/1/1
5	NAG	B	3601	1,5	-	0/6/23/26	0/1/1/1
5	NAG	B	3602	5	-	0/6/23/26	0/1/1/1
8	NAG	B	3741	1,8	-	0/6/23/26	0/1/1/1
8	NAG	B	3742	8	-	0/6/23/26	0/1/1/1
8	BMA	B	3743	8	-	0/2/19/22	0/1/1/1
8	MAN	B	3744	8	-	0/2/19/22	0/1/1/1
8	MAN	B	3745	8	-	0/2/19/22	0/1/1/1
8	MAN	B	3746	8	-	0/2/19/22	0/1/1/1
8	MAN	B	3747	8	-	0/2/19/22	0/1/1/1
5	NAG	B	941	1,5	-	0/6/23/26	0/1/1/1
5	NAG	B	942	5	-	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)
8	B	3742	NAG	O5-C1	-2.22	1.40	1.43

All (21) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	3602	NAG	O5-C1-C2	-4.04	105.84	111.47
5	A	1962	NAG	C1-O5-C5	-3.77	106.97	112.17
5	A	1961	NAG	O4-C4-C3	-3.13	103.56	110.36
8	B	3742	NAG	O5-C1-C2	-2.87	107.48	111.47
6	A	2462	NAG	O5-C1-C2	-2.70	107.71	111.47

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	A	3741	NAG	O4-C4-C3	-2.67	104.55	110.36
5	B	3602	NAG	C4-C3-C2	-2.34	107.59	111.02
8	B	3747	MAN	O5-C1-C2	-2.28	107.22	110.79
8	B	3747	MAN	C6-C5-C4	-2.17	107.92	113.00
8	B	3744	MAN	O5-C1-C2	-2.07	107.55	110.79
6	A	2463	BMA	O5-C1-C2	-2.05	107.58	110.79
5	A	3601	NAG	O4-C4-C5	2.13	114.66	109.28
8	B	3747	MAN	C1-O5-C5	2.24	115.25	112.17
5	B	3602	NAG	C1-O5-C5	2.31	115.35	112.17
6	A	3741	NAG	C1-C2-N2	2.70	115.10	110.49
5	A	3602	NAG	C3-C4-C5	3.03	115.55	110.22
5	A	1962	NAG	C4-C3-C2	3.43	116.05	111.02
5	A	1961	NAG	C1-O5-C5	3.87	117.50	112.17
6	B	2461	NAG	C1-O5-C5	4.87	118.88	112.17
6	A	2461	NAG	C1-O5-C5	5.02	119.09	112.17
5	A	1962	NAG	O5-C1-C2	6.11	119.97	111.47

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
5	A	1962	NAG	C1

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.6 Ligand geometry ⓘ

Of 9 ligands modelled in this entry, 2 are monoatomic - leaving 7 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	$\# Z  > 2$
4	NAG	A	1521	1	14,14,15	0.42	0	15,19,21	1.32	2 (13%)



Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	SIO	A	801	-	22,25,28	1.18	1 (4%)	23,36,40	1.58	4 (17%)
4	NAG	A	941	1	14,14,15	0.70	0	15,19,21	0.98	1 (6%)
4	NAG	B	1521	1	14,14,15	0.44	0	15,19,21	1.00	0
4	NAG	B	1961	1	14,14,15	0.56	0	15,19,21	1.02	1 (6%)
2	SIO	B	801	-	22,25,28	1.35	2 (9%)	23,36,40	1.40	4 (17%)
7	GOL	B	803	-	5,5,5	0.42	0	5,5,5	0.16	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
4	NAG	A	1521	1	-	0/6/23/26	0/1/1/1
2	SIO	A	801	-	-	0/21/45/48	0/1/1/1
4	NAG	A	941	1	-	0/6/23/26	0/1/1/1
4	NAG	B	1521	1	-	0/6/23/26	0/1/1/1
4	NAG	B	1961	1	-	0/6/23/26	0/1/1/1
2	SIO	B	801	-	-	0/21/45/48	0/1/1/1
7	GOL	B	803	-	-	0/4/4/4	0/0/0/0

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	801	SIO	C3-C2	2.78	1.54	1.52
2	A	801	SIO	O4-CA4	4.74	1.46	1.35
2	B	801	SIO	O4-CA4	4.89	1.46	1.35

All (12) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	801	SIO	O6-C2-C3	-3.49	106.07	111.14
2	B	801	SIO	CM2-O2-C2	-3.00	112.33	115.39
4	A	1521	NAG	C4-C3-C2	-2.74	107.01	111.02
2	B	801	SIO	C6-C5-N5	-2.29	106.98	111.00
2	A	801	SIO	C4-C5-N5	-2.11	105.87	110.62
2	A	801	SIO	C8-C7-C6	-2.09	108.98	113.04
4	A	941	NAG	C4-C3-C2	2.19	114.23	111.02
4	A	1521	NAG	C1-O5-C5	2.32	115.36	112.17
4	B	1961	NAG	C1-O5-C5	2.55	115.68	112.17
2	B	801	SIO	C2-O6-C6	2.56	118.10	113.91

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
2	B	801	SIO	O4-CA4-CM4	3.53	117.72	111.10
2	A	801	SIO	O4-CA4-CM4	4.51	119.57	111.10

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

### 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	353/386 (91%)	0.13	12 (3%) 46 48	21, 43, 78, 112	0
1	B	334/386 (86%)	0.41	23 (6%) 18 18	26, 54, 98, 115	0
All	All	687/772 (88%)	0.26	35 (5%) 29 30	21, 48, 92, 115	0

All (35) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	108	VAL	6.4
1	A	187	VAL	5.0
1	A	402	TYR	4.9
1	B	353	LEU	4.8
1	A	45	ALA	4.2
1	B	219	SER	3.4
1	B	352	GLN	3.4
1	A	114	HIS	3.4
1	A	49	CYS	3.3
1	B	108	VAL	3.3
1	A	71	ASN	3.3
1	A	107	GLY	3.3
1	B	69	LEU	3.1
1	B	110	PHE	3.0
1	B	109	ASN	3.0
1	B	114	HIS	2.9
1	B	401	ILE	2.9
1	B	157	TYR	2.8
1	B	190	TYR	2.8
1	B	218	GLY	2.7
1	A	48	ASP	2.7
1	B	188	ASP	2.7
1	B	399	VAL	2.7
1	A	400	CYS	2.7

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	B	113	SER	2.7
1	B	189	TYR	2.7
1	B	282	LEU	2.6
1	B	67	PRO	2.6
1	A	407	LEU	2.5
1	B	183	GLU	2.5
1	A	406	PRO	2.4
1	B	160	ASN	2.3
1	B	228	TYR	2.1
1	B	394	GLY	2.1
1	B	112	PRO	2.0

## 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(Å <sup>2</sup> )	Q<0.9
5	NAG	B	942	14/15	0.86	0.23	1.63	74,77,78,79	0
5	NAG	B	3601	14/15	0.94	0.18	1.03	52,59,62,69	0
8	MAN	B	3745	11/12	0.98	0.13	-0.76	36,37,38,42	0
6	NAG	A	2461	14/15	0.96	0.11	-1.06	44,49,53,57	0
6	NAG	A	3741	14/15	0.95	0.12	-1.13	39,43,48,56	0
5	NAG	A	3601	14/15	0.94	0.12	-1.15	47,52,59,67	0
6	NAG	B	2461	14/15	0.93	0.13	-1.22	48,53,58,64	0
8	NAG	B	3741	14/15	0.97	0.09	-2.05	35,38,41,41	0
8	MAN	B	3746	11/12	0.97	0.11	-2.74	36,38,41,44	0
6	NAG	A	3742	14/15	0.92	0.10	-	59,63,67,72	0
6	NAG	A	2462	14/15	0.86	0.14	-	54,62,67,72	0
5	NAG	A	1962	14/15	0.85	0.33	-	80,83,84,85	0
8	MAN	B	3747	11/12	0.93	0.16	-	44,47,49,51	0
6	BMA	B	2463	11/12	0.76	0.29	-	83,85,85,86	0
8	BMA	B	3743	11/12	0.97	0.12	-	35,38,40,43	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
6	BMA	A	2463	11/12	0.79	0.22	-	76,78,79,79	0
8	MAN	B	3744	11/12	0.94	0.11	-	47,49,53,53	0
6	BMA	A	3743	11/12	0.86	0.13	-	75,76,77,78	0
5	NAG	A	1961	14/15	0.91	0.23	-	57,67,69,75	0
5	NAG	B	941	14/15	0.92	0.23	-	55,62,66,69	0
5	NAG	A	3602	14/15	0.83	0.30	-	74,77,80,81	0
6	NAG	B	2462	14/15	0.83	0.23	-	66,71,75,79	0
5	NAG	B	3602	14/15	0.90	0.41	-	75,79,80,81	0
8	NAG	B	3742	14/15	0.97	0.13	-	37,39,41,43	0

## 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( $\text{\AA}^2$ )	Q<0.9
7	GOL	B	803	6/6	0.91	0.23	2.21	64,65,65,65	0
2	SIO	B	801	25/28	0.91	0.30	0.61	87,90,91,92	0
4	NAG	B	1961	14/15	0.90	0.23	0.38	57,62,63,63	0
2	SIO	A	801	25/28	0.95	0.17	0.09	51,59,63,64	0
3	K	B	802	1/1	0.97	0.09	-1.42	66,66,66,66	0
3	K	A	802	1/1	0.99	0.10	-2.48	51,51,51,51	0
4	NAG	B	1521	14/15	0.90	0.18	-	57,65,66,67	0
4	NAG	A	941	14/15	0.85	0.31	-	47,56,61,61	0
4	NAG	A	1521	14/15	0.85	0.22	-	58,65,67,68	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.