



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 06:36 PM GMT

PDB ID : 4M44  
Title : Crystal structure of hemagglutinin of influenza virus B/Yamanashi/166/1998  
in complex with avian-like receptor LSTa  
Authors : Ni, F.; Kondrashkina, E.; Wang, Q.  
Deposited on : 2013-08-06  
Resolution : 2.50 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

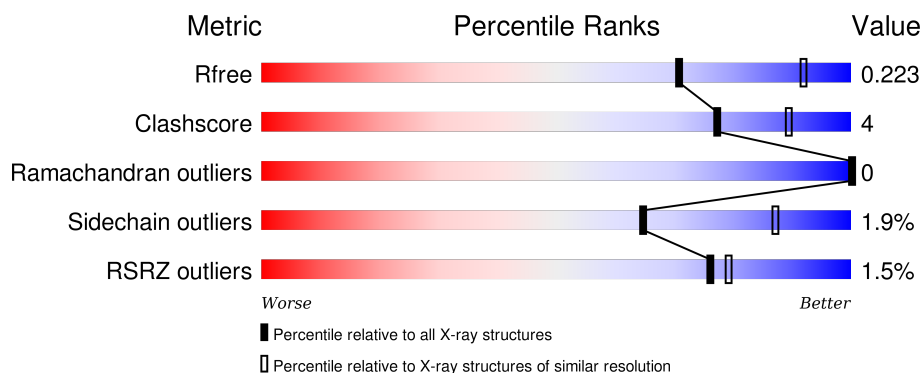
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 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))
$R_{free}$	91344	3553 (2.50-2.50)
Clashscore	102246	4242 (2.50-2.50)
Ramachandran outliers	100387	4156 (2.50-2.50)
Sidechain outliers	100360	4158 (2.50-2.50)
RSRZ outliers	91569	3562 (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	346	<div> <div>92%</div> <div>8%</div> </div>
1	C	346	<div> <div>91%</div> <div>8%</div> </div>
1	E	346	<div> <div>91%</div> <div>8%</div> </div>
2	B	182	<div> <div>2%</div> <div>84%</div> <div>9%</div> <div>6%</div> </div>
2	D	182	<div> <div>3%</div> <div>85%</div> <div>8%</div> <div>7%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	F	182	 A horizontal bar chart showing the quality of chain F. The bar is divided into segments: a small red segment at the beginning labeled '3%', a large green segment labeled '85%', a small yellow segment labeled '9%', and a small grey segment at the end labeled '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
5	NAG	E	412	-	-	-	X

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 12754 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 HA1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	343	Total	C	N	O	S	0	0	0
			2600	1633	464	487	16			
1	C	343	Total	C	N	O	S	0	0	0
			2600	1633	464	487	16			
1	E	343	Total	C	N	O	S	0	0	0
			2600	1633	464	487	16			

- Molecule 2 is a protein called Hemagglutinin HA2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	171	Total	C	N	O	S	0	0	0
			1301	814	222	259	6			
2	D	170	Total	C	N	O	S	0	0	0
			1293	808	221	258	6			
2	F	171	Total	C	N	O	S	0	0	0
			1301	814	222	259	6			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	177	GLY	-	EXPRESSION TAG	UNP A3DQM7
B	178	ALA	-	EXPRESSION TAG	UNP A3DQM7
B	179	LEU	-	EXPRESSION TAG	UNP A3DQM7
B	180	VAL	-	EXPRESSION TAG	UNP A3DQM7
B	181	PRO	-	EXPRESSION TAG	UNP A3DQM7
B	182	ARG	-	EXPRESSION TAG	UNP A3DQM7
D	177	GLY	-	EXPRESSION TAG	UNP A3DQM7
D	178	ALA	-	EXPRESSION TAG	UNP A3DQM7
D	179	LEU	-	EXPRESSION TAG	UNP A3DQM7
D	180	VAL	-	EXPRESSION TAG	UNP A3DQM7
D	181	PRO	-	EXPRESSION TAG	UNP A3DQM7
D	182	ARG	-	EXPRESSION TAG	UNP A3DQM7

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
F	177	GLY	-	EXPRESSION TAG	UNP A3DQM7
F	178	ALA	-	EXPRESSION TAG	UNP A3DQM7
F	179	LEU	-	EXPRESSION TAG	UNP A3DQM7
F	180	VAL	-	EXPRESSION TAG	UNP A3DQM7
F	181	PRO	-	EXPRESSION TAG	UNP A3DQM7
F	182	ARG	-	EXPRESSION TAG	UNP A3DQM7

- 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	A	2	Total	C	N	O	0	0
			28	16	2	10		
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 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	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	C	1	Total	C	N	O	0	0
			14	8	1	5		
4	C	1	Total	C	N	O	0	0
			14	8	1	5		
4	C	1	Total	C	N	O	0	0
			14	8	1	5		
4	D	1	Total	C	N	O	0	0
			14	8	1	5		
4	E	1	Total	C	N	O	0	0
			14	8	1	5		
4	E	1	Total	C	N	O	0	0
			14	8	1	5		
4	E	1	Total	C	N	O	0	0
			14	8	1	5		
4	F	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 5 is a polymer of unknown type called SUGAR (4-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	A	4	Total	C	N	O	0	0
			56	31	2	23		
5	C	4	Total	C	N	O	0	0
			56	31	2	23		
5	E	4	Total	C	N	O	0	0
			56	31	2	23		

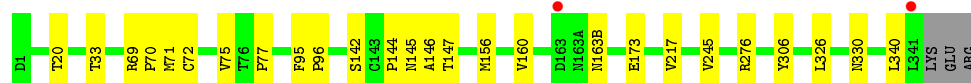
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	118	Total	O	0	0
			118	118		
6	B	45	Total	O	0	0
			45	45		
6	C	109	Total	O	0	0
			109	109		
6	D	47	Total	O	0	0
			47	47		
6	E	112	Total	O	0	0
			112	112		
6	F	40	Total	O	0	0
			40	40		

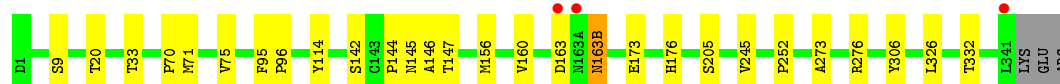
### 3 Residue-property plots [i](#)

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 ( $\text{RSRZ} > 2$ ). 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.

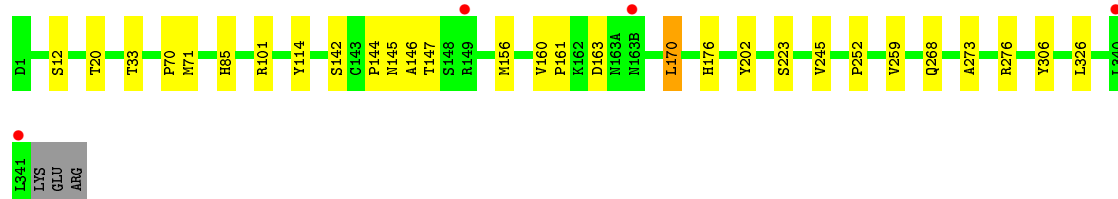
#### • Molecule 1: Hemagglutinin HA1



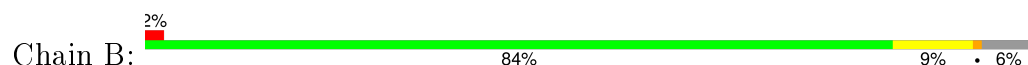
#### • Molecule 1: Hemagglutinin HA1



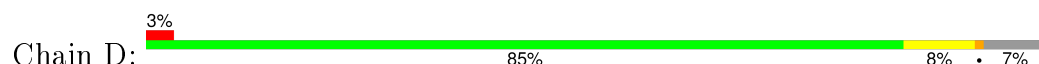
#### • Molecule 1: Hemagglutinin HA1



#### • Molecule 2: Hemagglutinin HA2



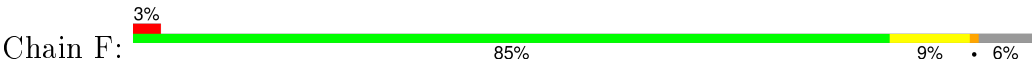
#### • Molecule 2: Hemagglutinin HA2







● Molecule 2: Hemagglutinin HA2



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	176.13Å 101.52Å 137.37Å 90.00° 115.22° 90.00°	Depositor
Resolution (Å)	39.36 – 2.50 39.36 – 2.50	Depositor EDS
% Data completeness (in resolution range)	99.9 (39.36-2.50) 99.5 (39.36-2.50)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.33 (at 2.51Å)	Xtriage
Refinement program	PHENIX (phenix.refine: dev_1452)	Depositor
R, $R_{free}$	0.186 , 0.223 0.186 , 0.223	Depositor DCC
$R_{free}$ test set	3789 reflections (5.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	46.2	Xtriage
Anisotropy	0.572	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 33.1	EDS
Estimated twinning fraction	0.457 for $1/2^*h-3/2^*k,-1/2^*h-1/2^*k,-1/2^*h+1/2^*k-l$ 0.457 for $1/2^*h+3/2^*k,1/2^*h-1/2^*k,-1/2^*h-1/2^*k-l$	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtriage
Outliers	0 of 75849 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	12754	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.88% 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 [i](#)

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

Bond lengths and bond angles in the following residue types are not validated in this section: SIA, GAL, NAG

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	0.25	0/2659	0.46	0/3613
1	C	0.29	0/2659	0.48	0/3613
1	E	0.28	0/2659	0.47	0/3613
2	B	0.22	0/1320	0.36	0/1780
2	D	0.36	0/1312	0.38	0/1769
2	F	0.22	0/1320	0.36	0/1780
All	All	0.28	0/11929	0.44	0/16168

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

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	2600	0	2621	17	0
1	C	2600	0	2622	16	0
1	E	2600	0	2621	19	0
2	B	1301	0	1273	12	0
2	D	1293	0	1262	9	0
2	F	1301	0	1273	10	0
3	A	84	0	75	1	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	C	84	0	75	1	0
3	E	84	0	75	4	0
4	A	42	0	39	2	0
4	B	14	0	13	0	0
4	C	42	0	39	1	0
4	D	14	0	13	0	0
4	E	42	0	39	2	0
4	F	14	0	13	0	0
5	A	56	0	47	1	0
5	C	56	0	47	3	0
5	E	56	0	47	3	0
6	A	118	0	0	6	0
6	B	45	0	0	2	0
6	C	109	0	0	6	0
6	D	47	0	0	2	0
6	E	112	0	0	6	0
6	F	40	0	0	0	0
All	All	12754	0	12194	90	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 4.

The worst 5 of 90 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)
5:E:412:NAG:O7	6:E:580:HOH:O	1.63	1.12
3:C:407:NAG:H61	6:C:539:HOH:O	1.62	0.99
5:A:413:GAL:O2	6:A:581:HOH:O	1.91	0.89
2:D:46:ASN:OD1	6:D:345:HOH:O	1.92	0.86
1:C:205:SER:OG	6:C:593:HOH:O	1.99	0.80

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	341/346 (99%)	327 (96%)	14 (4%)	0	100	100
1	C	341/346 (99%)	327 (96%)	14 (4%)	0	100	100
1	E	341/346 (99%)	327 (96%)	14 (4%)	0	100	100
2	B	169/182 (93%)	166 (98%)	3 (2%)	0	100	100
2	D	168/182 (92%)	165 (98%)	3 (2%)	0	100	100
2	F	169/182 (93%)	166 (98%)	3 (2%)	0	100	100
All	All	1529/1584 (96%)	1478 (97%)	51 (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	291/294 (99%)	287 (99%)	4 (1%)	74	91
1	C	291/294 (99%)	285 (98%)	6 (2%)	61	85
1	E	291/294 (99%)	287 (99%)	4 (1%)	74	91
2	B	139/145 (96%)	136 (98%)	3 (2%)	60	84
2	D	138/145 (95%)	134 (97%)	4 (3%)	50	77
2	F	139/145 (96%)	136 (98%)	3 (2%)	60	84
All	All	1289/1317 (98%)	1265 (98%)	24 (2%)	65	87

5 of 24 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	163(B)	ASN
2	D	22	HIS
2	F	51	ASN
1	C	276	ARG
1	C	326	LEU

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 ⓘ

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

## 5.5 Carbohydrates ⓘ

30 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	401	1,3	14,14,15	0.56	0	15,19,21	1.58	3 (20%)
3	NAG	A	402	3	14,14,15	0.93	2 (14%)	15,19,21	0.73	0
3	NAG	A	404	1,3	14,14,15	0.51	0	15,19,21	1.25	1 (6%)
3	NAG	A	405	3	14,14,15	1.05	1 (7%)	15,19,21	1.43	1 (6%)
3	NAG	A	407	1,3	14,14,15	0.33	0	15,19,21	1.06	1 (6%)
3	NAG	A	408	3	14,14,15	0.77	1 (7%)	15,19,21	1.51	1 (6%)
5	SIA	A	410	5	16,20,21	1.00	1 (6%)	18,28,31	1.43	2 (11%)
5	GAL	A	411	5	11,11,12	1.72	2 (18%)	14,15,17	1.63	4 (28%)
5	NAG	A	412	5	14,14,15	0.61	0	15,19,21	1.07	1 (6%)
5	GAL	A	413	5	11,11,12	1.07	1 (9%)	14,15,17	1.78	5 (35%)
3	NAG	C	401	1,3	14,14,15	0.63	0	15,19,21	1.27	1 (6%)
3	NAG	C	402	3	14,14,15	0.96	1 (7%)	15,19,21	1.48	1 (6%)
3	NAG	C	405	1,3	14,14,15	0.46	0	15,19,21	1.05	1 (6%)
3	NAG	C	406	3	14,14,15	1.02	1 (7%)	15,19,21	1.31	1 (6%)
3	NAG	C	407	1,3	14,14,15	1.89	1 (7%)	15,19,21	1.78	2 (13%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
3	NAG	C	408	3	14,14,15	0.98	1 (7%)	15,19,21	0.92	1 (6%)
5	SIA	C	410	5	16,20,21	1.12	1 (6%)	18,28,31	1.60	5 (27%)
5	GAL	C	411	5	11,11,12	2.03	3 (27%)	14,15,17	2.49	4 (28%)
5	NAG	C	412	5	14,14,15	0.66	1 (7%)	15,19,21	1.39	2 (13%)
5	GAL	C	413	5	11,11,12	1.28	2 (18%)	14,15,17	1.75	4 (28%)
3	NAG	E	401	1,3	14,14,15	0.72	1 (7%)	15,19,21	0.59	0
3	NAG	E	402	3	14,14,15	0.87	0	15,19,21	2.35	4 (26%)
3	NAG	E	405	1,3	14,14,15	0.41	0	15,19,21	1.11	1 (6%)
3	NAG	E	406	3	14,14,15	1.19	1 (7%)	15,19,21	1.71	1 (6%)
3	NAG	E	407	1,3	14,14,15	0.39	0	15,19,21	0.83	1 (6%)
3	NAG	E	408	3	14,14,15	1.73	2 (14%)	15,19,21	1.02	0
5	SIA	E	410	5	16,20,21	1.10	2 (12%)	18,28,31	1.64	5 (27%)
5	GAL	E	411	5	11,11,12	2.38	4 (36%)	14,15,17	2.58	6 (42%)
5	NAG	E	412	5	14,14,15	0.62	0	15,19,21	1.40	2 (13%)
5	GAL	E	413	5	11,11,12	1.72	2 (18%)	14,15,17	2.08	4 (28%)

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	401	1,3	-	0/6/23/26	0/1/1/1
3	NAG	A	402	3	-	0/6/23/26	0/1/1/1
3	NAG	A	404	1,3	-	0/6/23/26	0/1/1/1
3	NAG	A	405	3	-	0/6/23/26	0/1/1/1
3	NAG	A	407	1,3	-	0/6/23/26	0/1/1/1
3	NAG	A	408	3	-	0/6/23/26	0/1/1/1
5	SIA	A	410	5	-	0/14/34/38	0/1/1/1
5	GAL	A	411	5	-	0/2/19/22	0/1/1/1
5	NAG	A	412	5	-	0/6/23/26	0/1/1/1
5	GAL	A	413	5	-	0/2/19/22	0/1/1/1
3	NAG	C	401	1,3	-	0/6/23/26	0/1/1/1
3	NAG	C	402	3	-	0/6/23/26	0/1/1/1
3	NAG	C	405	1,3	-	0/6/23/26	0/1/1/1
3	NAG	C	406	3	-	0/6/23/26	0/1/1/1
3	NAG	C	407	1,3	-	0/6/23/26	0/1/1/1
3	NAG	C	408	3	-	0/6/23/26	0/1/1/1
5	SIA	C	410	5	-	0/14/34/38	0/1/1/1

Continued on next page...

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	GAL	C	411	5	-	0/2/19/22	0/1/1/1
5	NAG	C	412	5	-	0/6/23/26	0/1/1/1
5	GAL	C	413	5	-	0/2/19/22	0/1/1/1
3	NAG	E	401	1,3	-	0/6/23/26	0/1/1/1
3	NAG	E	402	3	-	0/6/23/26	0/1/1/1
3	NAG	E	405	1,3	-	0/6/23/26	0/1/1/1
3	NAG	E	406	3	-	0/6/23/26	0/1/1/1
3	NAG	E	407	1,3	-	0/6/23/26	0/1/1/1
3	NAG	E	408	3	-	0/6/23/26	0/1/1/1
5	SIA	E	410	5	-	0/14/34/38	0/1/1/1
5	GAL	E	411	5	-	0/2/19/22	0/1/1/1
5	NAG	E	412	5	-	0/6/23/26	0/1/1/1
5	GAL	E	413	5	-	0/2/19/22	0/1/1/1

The worst 5 of 31 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	407	NAG	O5-C1	-6.97	1.32	1.43
5	E	411	GAL	O5-C5	-4.02	1.34	1.43
5	C	411	GAL	O5-C5	-3.77	1.35	1.43
5	E	413	GAL	O5-C1	-3.23	1.38	1.43
3	E	401	NAG	O5-C1	-2.52	1.39	1.43

The worst 5 of 65 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	E	411	GAL	O5-C1-C2	-4.30	103.89	110.86
5	E	412	NAG	C1-O5-C5	-4.00	107.17	112.25
3	A	401	NAG	O4-C4-C3	-3.99	101.35	110.34
5	C	412	NAG	C1-O5-C5	-3.95	107.23	112.25
5	E	410	SIA	C7-C6-C5	-3.27	109.37	114.32

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

13 monomers are involved in 13 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	405	NAG	1	0
5	A	413	GAL	1	0
3	C	407	NAG	1	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	C	410	SIA	1	0
5	C	411	GAL	3	0
5	C	412	NAG	2	0
3	E	401	NAG	1	0
3	E	406	NAG	1	0
3	E	407	NAG	1	0
3	E	408	NAG	2	0
5	E	410	SIA	1	0
5	E	411	GAL	2	0
5	E	412	NAG	2	0

## 5.6 Ligand geometry

12 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	$\# Z  > 2$
4	NAG	A	403	1	14,14,15	0.43	0	15,19,21	0.46	0
4	NAG	A	406	1	14,14,15	0.30	0	15,19,21	0.80	1 (6%)
4	NAG	A	409	1	14,14,15	0.38	0	15,19,21	0.36	0
4	NAG	B	201	2	14,14,15	0.50	0	15,19,21	0.85	1 (6%)
4	NAG	C	403	1	14,14,15	0.44	0	15,19,21	0.83	1 (6%)
4	NAG	C	404	1	14,14,15	1.01	1 (7%)	15,19,21	1.30	2 (13%)
4	NAG	C	409	1	14,14,15	0.77	1 (7%)	15,19,21	0.96	1 (6%)
4	NAG	D	201	2	14,14,15	0.77	0	15,19,21	0.84	1 (6%)
4	NAG	E	403	1	14,14,15	1.51	2 (14%)	15,19,21	1.96	4 (26%)
4	NAG	E	404	1	14,14,15	0.45	0	15,19,21	1.07	1 (6%)
4	NAG	E	409	1	14,14,15	2.12	2 (14%)	15,19,21	1.45	2 (13%)
4	NAG	F	201	2	14,14,15	0.65	1 (7%)	15,19,21	0.56	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	403	1	-	0/6/23/26	0/1/1/1
4	NAG	A	406	1	-	0/6/23/26	0/1/1/1
4	NAG	A	409	1	-	0/6/23/26	0/1/1/1
4	NAG	B	201	2	-	0/6/23/26	0/1/1/1
4	NAG	C	403	1	-	0/6/23/26	0/1/1/1
4	NAG	C	404	1	-	0/6/23/26	0/1/1/1
4	NAG	C	409	1	-	0/6/23/26	0/1/1/1
4	NAG	D	201	2	-	0/6/23/26	0/1/1/1
4	NAG	E	403	1	-	0/6/23/26	0/1/1/1
4	NAG	E	404	1	-	0/6/23/26	0/1/1/1
4	NAG	E	409	1	-	0/6/23/26	0/1/1/1
4	NAG	F	201	2	-	0/6/23/26	0/1/1/1

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	E	409	NAG	O5-C1	-7.19	1.31	1.43
4	E	403	NAG	O5-C1	-4.01	1.37	1.43
4	C	404	NAG	C1-C2	-2.93	1.48	1.52
4	E	409	NAG	C1-C2	-2.81	1.48	1.52
4	F	201	NAG	O5-C1	2.25	1.47	1.43

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	E	403	NAG	C4-C3-C2	-3.31	106.09	111.23
4	C	403	NAG	C1-O5-C5	-2.66	108.87	112.25
4	D	201	NAG	C1-O5-C5	2.03	114.83	112.25
4	E	403	NAG	O3-C3-C4	2.04	114.92	110.34
4	C	404	NAG	C3-C4-C5	2.44	114.44	110.20

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	406	NAG	1	0
4	A	409	NAG	1	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	C	404	NAG	1	0
4	E	403	NAG	1	0
4	E	409	NAG	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	343/346 (99%)	0.07	2 (0%) 90 91	33, 46, 65, 100	0
1	C	343/346 (99%)	0.11	3 (0%) 85 88	35, 46, 65, 103	0
1	E	343/346 (99%)	0.09	4 (1%) 81 83	32, 46, 65, 102	0
2	B	171/182 (93%)	0.23	3 (1%) 71 75	31, 49, 73, 86	0
2	D	170/182 (93%)	0.24	5 (2%) 55 60	31, 49, 71, 86	0
2	F	171/182 (93%)	0.21	6 (3%) 48 53	30, 49, 74, 88	0
All	All	1541/1584 (97%)	0.14	23 (1%) 76 79	30, 47, 70, 103	0

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	341	LEU	4.1
1	C	163	ASP	3.8
1	E	340	LEU	3.6
2	B	170	LEU	3.4
2	B	172	ILE	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	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
5	NAG	E	412	14/15	0.77	0.22	2.30	55,72,87,90	0
5	NAG	C	412	14/15	0.76	0.22	1.11	48,70,85,86	0
5	SIA	C	410	20/21	0.94	0.16	0.24	44,49,55,59	0
5	SIA	A	410	20/21	0.96	0.15	0.06	42,48,53,54	0
3	NAG	A	401	14/15	0.94	0.14	-0.01	53,59,70,73	0
3	NAG	C	401	14/15	0.91	0.14	-0.03	48,62,68,72	0
5	SIA	E	410	20/21	0.96	0.14	-0.44	43,48,53,56	0
3	NAG	E	401	14/15	0.95	0.12	-0.63	48,56,62,67	0
3	NAG	C	405	14/15	0.90	0.13	-1.10	57,65,72,78	0
3	NAG	E	406	14/15	0.82	0.14	-	69,85,89,94	0
5	GAL	A	411	11/12	0.95	0.16	-	48,53,58,60	0
5	GAL	C	411	11/12	0.93	0.12	-	50,54,63,66	0
3	NAG	C	402	14/15	0.89	0.15	-	72,75,80,80	0
5	GAL	C	413	11/12	0.78	0.15	-	76,86,96,98	0
5	GAL	E	413	11/12	0.73	0.16	-	89,103,116,117	0
5	GAL	A	413	11/12	0.86	0.16	-	76,88,97,99	0
5	GAL	E	411	11/12	0.94	0.15	-	51,53,62,63	0
3	NAG	A	402	14/15	0.89	0.14	-	72,89,99,109	0
3	NAG	A	405	14/15	0.90	0.15	-	68,78,86,87	0
3	NAG	E	402	14/15	0.89	0.13	-	70,79,84,87	0
3	NAG	A	404	14/15	0.95	0.14	-	55,63,73,75	0
3	NAG	E	408	14/15	0.76	0.17	-	79,94,100,101	0
5	NAG	A	412	14/15	0.86	0.17	-	46,63,81,93	0
3	NAG	C	408	14/15	0.75	0.18	-	71,89,93,94	0
3	NAG	A	408	14/15	0.88	0.11	-	64,86,92,94	0
3	NAG	E	407	14/15	0.88	0.17	-	51,60,71,74	0
3	NAG	C	407	14/15	0.90	0.16	-	47,61,69,71	0
3	NAG	E	405	14/15	0.92	0.14	-	57,65,74,75	0
3	NAG	C	406	14/15	0.88	0.13	-	69,82,88,89	0
3	NAG	A	407	14/15	0.91	0.13	-	48,59,70,71	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
4	NAG	A	409	14/15	0.77	0.20	-	105,122,126,127	0
4	NAG	E	409	14/15	0.73	0.20	-	105,124,129,129	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	NAG	A	406	14/15	0.61	0.56	-	109,126,133,134	0
4	NAG	B	201	14/15	0.84	0.17	-	79,92,105,107	0
4	NAG	D	201	14/15	0.85	0.18	-	81,93,100,105	0
4	NAG	F	201	14/15	0.81	0.15	-	79,94,102,104	0
4	NAG	C	404	14/15	0.60	0.50	-	106,124,128,134	0
4	NAG	A	403	14/15	0.94	0.14	-	68,73,78,78	0
4	NAG	C	403	14/15	0.92	0.16	-	63,71,78,80	0
4	NAG	E	403	14/15	0.83	0.23	-	80,90,97,103	0
4	NAG	E	404	14/15	0.77	0.36	-	106,115,122,124	0
4	NAG	C	409	14/15	0.70	0.20	-	110,124,129,131	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.