



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 10:32 AM GMT

PDB ID : 3M6S  
Title : Crystal structure of H1N1pdm Hemagglutinin  
Authors : Yang, H.; Carney, P.J.; Stevens, J.  
Deposited on : 2010-03-16  
Resolution : 2.80 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

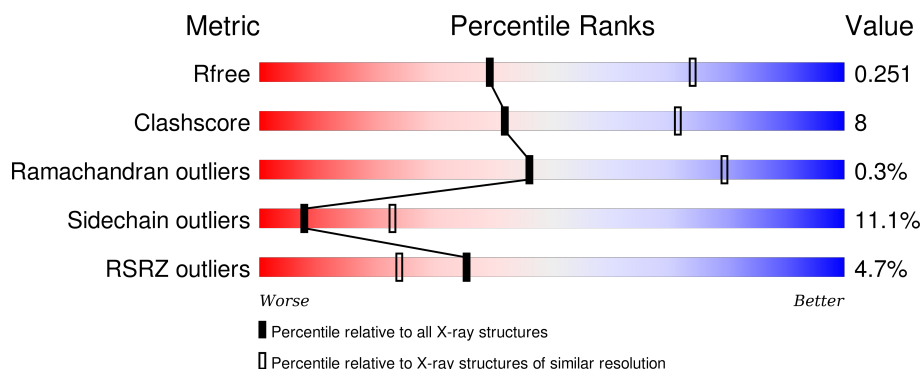
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.80 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	2393 (2.80-2.80)
Clashscore	102246	2827 (2.80-2.80)
Ramachandran outliers	100387	2782 (2.80-2.80)
Sidechain outliers	100360	2784 (2.80-2.80)
RSRZ outliers	91569	2404 (2.80-2.80)

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	331	<div> <div>3%</div> <div>79%</div> <div>16%</div> <div>..</div> </div>
1	C	331	<div> <div>%</div> <div>76%</div> <div>18%</div> <div>..</div> </div>
1	E	331	<div> <div>5%</div> <div>77%</div> <div>17%</div> <div>..</div> </div>
1	G	331	<div> <div>3%</div> <div>78%</div> <div>15%</div> <div>...</div> </div>
1	I	331	<div> <div>5%</div> <div>79%</div> <div>15%</div> <div>..</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	K	331	
2	B	181	
2	D	181	
2	F	181	
2	H	181	
2	J	181	
2	L	181	

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	G	331	X	-	-	-
5	MAN	G	330	X	-	-	-

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 23587 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.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	325	Total	C	N	O	S	0	0	0
			2533	1602	435	485	11			
1	C	323	Total	C	N	O	S	0	0	0
			2518	1593	433	481	11			
1	E	323	Total	C	N	O	S	0	0	0
			2518	1593	433	481	11			
1	G	322	Total	C	N	O	S	0	0	0
			2514	1591	432	480	11			
1	I	322	Total	C	N	O	S	0	0	0
			2511	1588	432	480	11			
1	K	321	Total	C	N	O	S	0	0	0
			2506	1587	431	477	11			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	ALA	-	EXPRESSION TAG	UNP C5MV42
A	-2	ASP	-	EXPRESSION TAG	UNP C5MV42
A	-1	PRO	-	EXPRESSION TAG	UNP C5MV42
A	0	GLY	-	EXPRESSION TAG	UNP C5MV42
C	-3	ALA	-	EXPRESSION TAG	UNP C5MV42
C	-2	ASP	-	EXPRESSION TAG	UNP C5MV42
C	-1	PRO	-	EXPRESSION TAG	UNP C5MV42
C	0	GLY	-	EXPRESSION TAG	UNP C5MV42
E	-3	ALA	-	EXPRESSION TAG	UNP C5MV42
E	-2	ASP	-	EXPRESSION TAG	UNP C5MV42
E	-1	PRO	-	EXPRESSION TAG	UNP C5MV42
E	0	GLY	-	EXPRESSION TAG	UNP C5MV42
G	-3	ALA	-	EXPRESSION TAG	UNP C5MV42
G	-2	ASP	-	EXPRESSION TAG	UNP C5MV42
G	-1	PRO	-	EXPRESSION TAG	UNP C5MV42
G	0	GLY	-	EXPRESSION TAG	UNP C5MV42
I	-3	ALA	-	EXPRESSION TAG	UNP C5MV42

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
I	-2	ASP	-	EXPRESSION TAG	UNP C5MV42
I	-1	PRO	-	EXPRESSION TAG	UNP C5MV42
I	0	GLY	-	EXPRESSION TAG	UNP C5MV42
K	-3	ALA	-	EXPRESSION TAG	UNP C5MV42
K	-2	ASP	-	EXPRESSION TAG	UNP C5MV42
K	-1	PRO	-	EXPRESSION TAG	UNP C5MV42
K	0	GLY	-	EXPRESSION TAG	UNP C5MV42

- Molecule 2 is a protein called Hemagglutinin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	169	Total	C	N	O	S	5	0	0
			1368	860	232	270	6			
2	D	171	Total	C	N	O	S	0	0	0
			1386	870	234	276	6			
2	F	171	Total	C	N	O	S	0	0	0
			1386	870	234	276	6			
2	H	170	Total	C	N	O	S	0	0	0
			1377	865	233	273	6			
2	J	170	Total	C	N	O	S	0	0	0
			1377	865	233	273	6			
2	L	171	Total	C	N	O	S	0	0	0
			1386	870	234	276	6			

There are 36 discrepancies between the modelled and reference sequences:

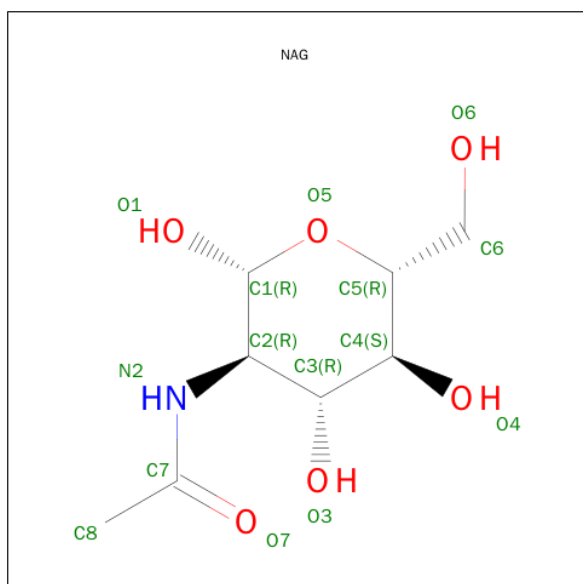
Chain	Residue	Modelled	Actual	Comment	Reference
B	175	SER	GLY	ENGINEERED	UNP C5MV42
B	176	GLY	VAL	ENGINEERED	UNP C5MV42
B	177	ARG	LYS	ENGINEERED	UNP C5MV42
B	179	VAL	-	EXPRESSION TAG	UNP C5MV42
B	180	PRO	-	EXPRESSION TAG	UNP C5MV42
B	181	ARG	-	EXPRESSION TAG	UNP C5MV42
D	175	SER	GLY	ENGINEERED	UNP C5MV42
D	176	GLY	VAL	ENGINEERED	UNP C5MV42
D	177	ARG	LYS	ENGINEERED	UNP C5MV42
D	179	VAL	-	EXPRESSION TAG	UNP C5MV42
D	180	PRO	-	EXPRESSION TAG	UNP C5MV42
D	181	ARG	-	EXPRESSION TAG	UNP C5MV42
F	175	SER	GLY	ENGINEERED	UNP C5MV42
F	176	GLY	VAL	ENGINEERED	UNP C5MV42
F	177	ARG	LYS	ENGINEERED	UNP C5MV42

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	179	VAL	-	EXPRESSION TAG	UNP C5MV42
F	180	PRO	-	EXPRESSION TAG	UNP C5MV42
F	181	ARG	-	EXPRESSION TAG	UNP C5MV42
H	175	SER	GLY	ENGINEERED	UNP C5MV42
H	176	GLY	VAL	ENGINEERED	UNP C5MV42
H	177	ARG	LYS	ENGINEERED	UNP C5MV42
H	179	VAL	-	EXPRESSION TAG	UNP C5MV42
H	180	PRO	-	EXPRESSION TAG	UNP C5MV42
H	181	ARG	-	EXPRESSION TAG	UNP C5MV42
J	175	SER	GLY	ENGINEERED	UNP C5MV42
J	176	GLY	VAL	ENGINEERED	UNP C5MV42
J	177	ARG	LYS	ENGINEERED	UNP C5MV42
J	179	VAL	-	EXPRESSION TAG	UNP C5MV42
J	180	PRO	-	EXPRESSION TAG	UNP C5MV42
J	181	ARG	-	EXPRESSION TAG	UNP C5MV42
L	175	SER	GLY	ENGINEERED	UNP C5MV42
L	176	GLY	VAL	ENGINEERED	UNP C5MV42
L	177	ARG	LYS	ENGINEERED	UNP C5MV42
L	179	VAL	-	EXPRESSION TAG	UNP C5MV42
L	180	PRO	-	EXPRESSION TAG	UNP C5MV42
L	181	ARG	-	EXPRESSION TAG	UNP C5MV42

- Molecule 3 is SUGAR (N-ACETYL-D-GLUCOSAMINE) (three-letter code: NAG) (formula:  $C_8H_{15}NO_6$ ).



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

- Molecule 4 is a polymer of unknown type called SUGAR (2-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	C	2	Total	C	N	O	0	0
			28	16	2	10		
4	I	2	Total	C	N	O	0	0
			28	16	2	10		
4	K	2	Total	C	N	O	0	0
			28	16	2	10		
4	K	2	Total	C	N	O	0	0
			28	16	2	10		

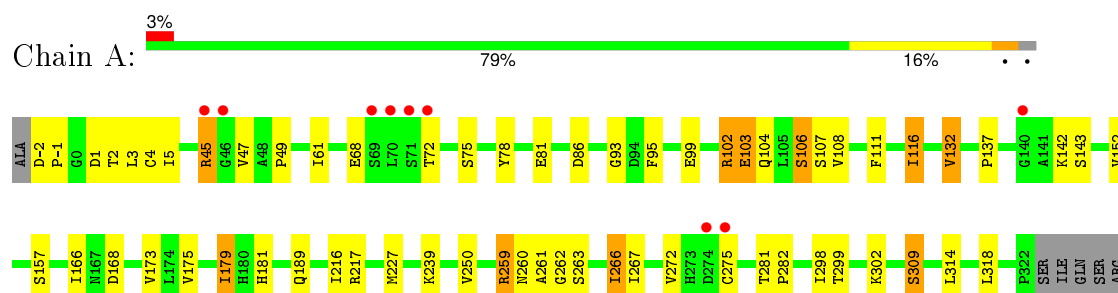
- Molecule 5 is a polymer of unknown type called SUGAR (3-MER).

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

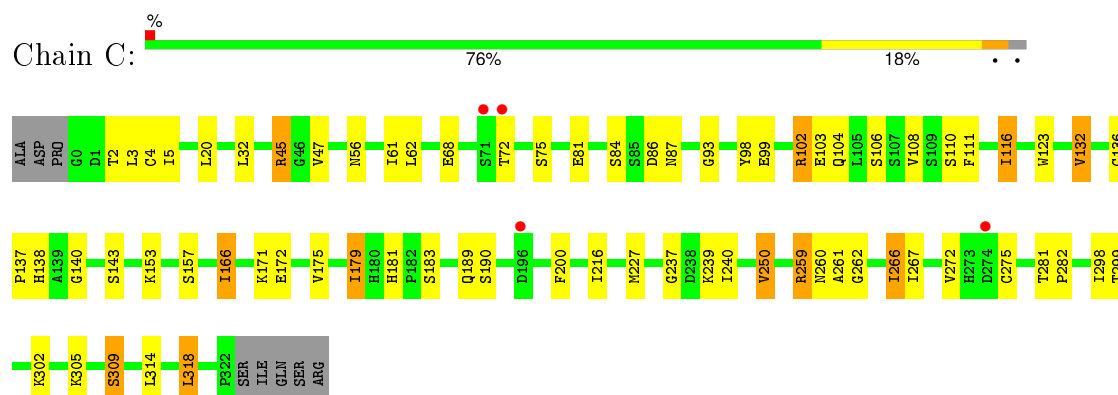
### 3 Residue-property plots [i](#)

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.

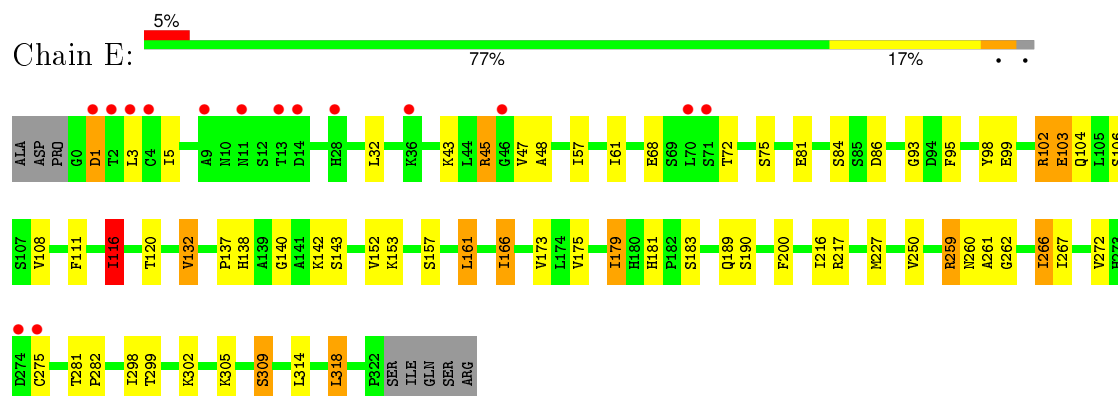
- Molecule 1: Hemagglutinin



- Molecule 1: Hemagglutinin

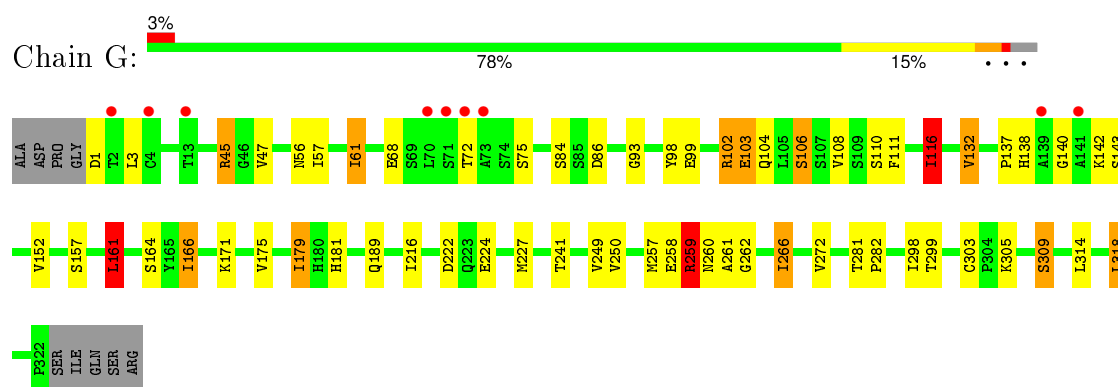


- Molecule 1: Hemagglutinin

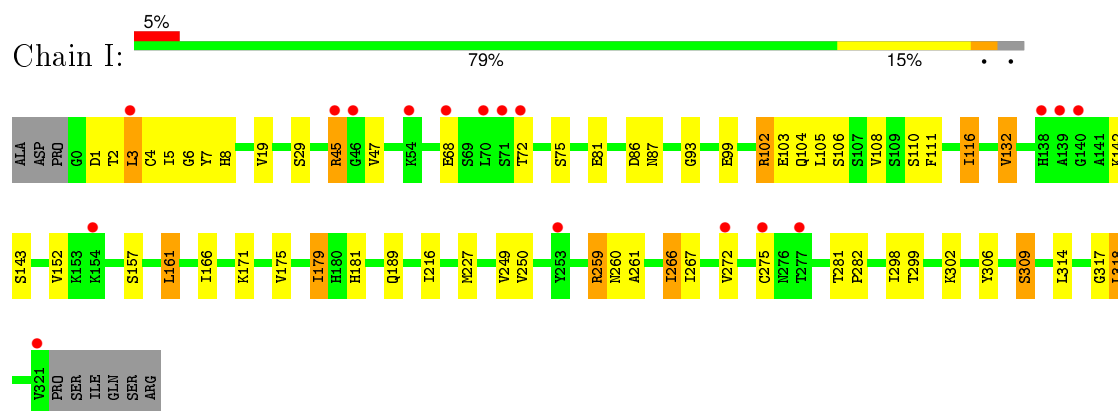


- Molecule 1: Hemagglutinin

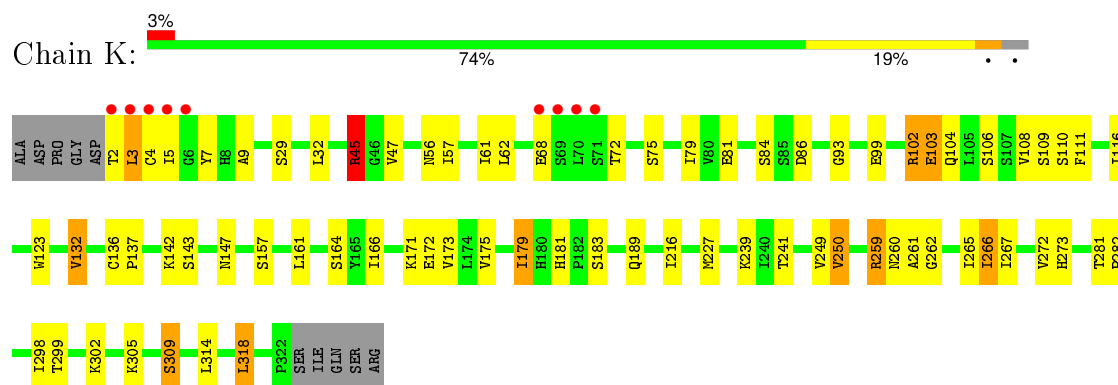




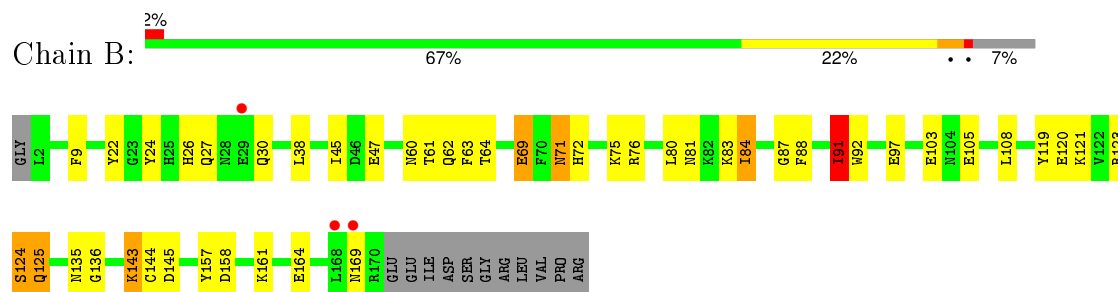
- Molecule 1: Hemagglutinin



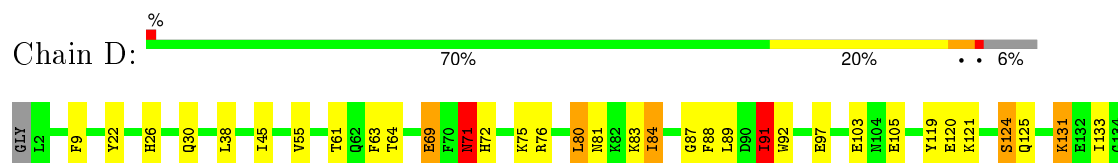
- Molecule 1: Hemagglutinin



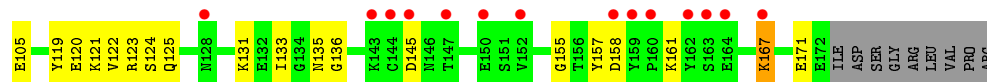
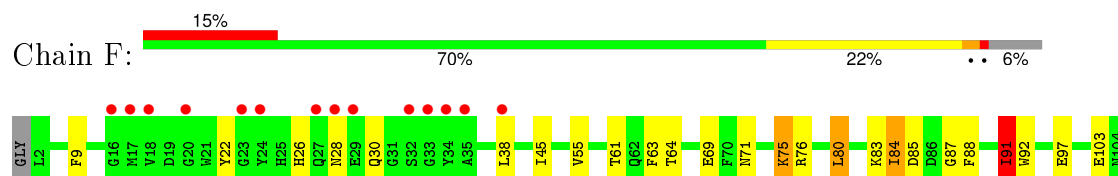
- Molecule 2: Hemagglutinin



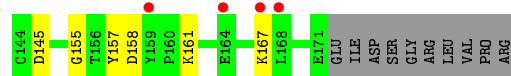
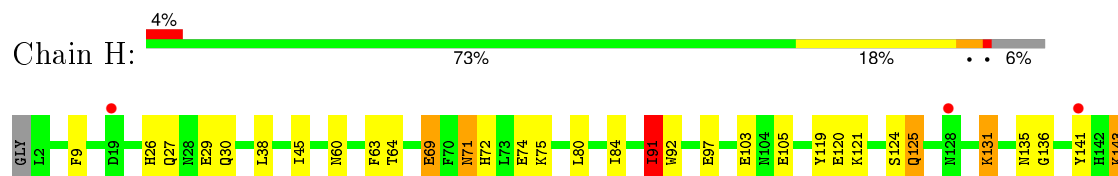
- Molecule 2: Hemagglutinin



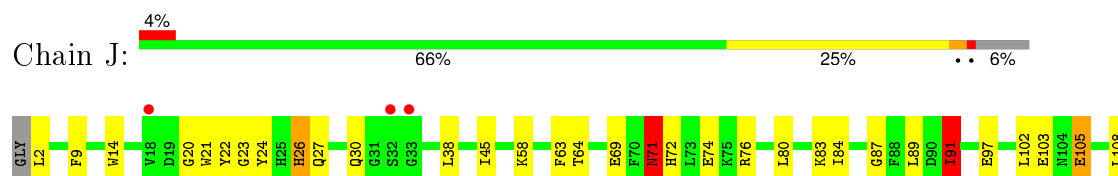
• Molecule 2: Hemagglutinin



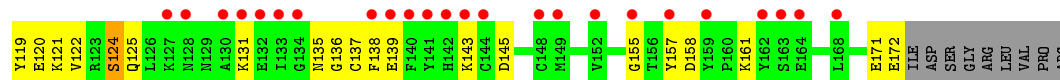
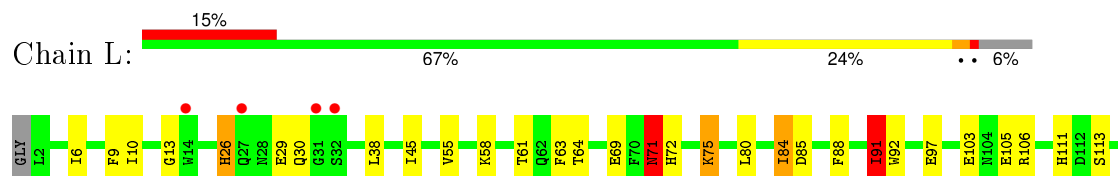
• Molecule 2: Hemagglutinin



• Molecule 2: Hemagglutinin



• Molecule 2: Hemagglutinin



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	73.98Å 109.71Å 129.90Å 86.25° 74.68° 75.10°	Depositor
Resolution (Å)	50.00 – 2.80 48.87 – 2.80	Depositor EDS
% Data completeness (in resolution range)	98.1 (50.00-2.80) 98.0 (48.87-2.80)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.08	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.32 (at 2.81Å)	Xtriage
Refinement program	REFMAC 5.5.0102	Depositor
R, $R_{free}$	0.230 , 0.256 0.227 , 0.251	Depositor DCC
$R_{free}$ test set	4608 reflections (5.28%)	DCC
Wilson B-factor (Å <sup>2</sup> )	50.1	Xtriage
Anisotropy	0.013	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 43.5	EDS
Estimated twinning fraction	0.016 for -h,-k,-h+l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 91952 reflections	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	23587	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.83% 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: NAG, 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	0.75	0/2597	0.83	1/3530 (0.0%)
1	C	0.79	1/2581 (0.0%)	0.84	2/3507 (0.1%)
1	E	0.71	0/2581	0.80	2/3507 (0.1%)
1	G	0.86	1/2577 (0.0%)	0.88	4/3502 (0.1%)
1	I	0.68	0/2573	0.78	1/3495 (0.0%)
1	K	0.75	0/2569	0.82	2/3491 (0.1%)
2	B	0.71	0/1396	0.79	3/1881 (0.2%)
2	D	0.77	0/1414	0.78	2/1905 (0.1%)
2	F	0.68	0/1414	0.73	1/1905 (0.1%)
2	H	0.75	1/1405 (0.1%)	0.72	1/1893 (0.1%)
2	J	0.77	1/1405 (0.1%)	0.75	2/1893 (0.1%)
2	L	0.73	0/1414	0.74	3/1905 (0.2%)
All	All	0.75	4/23926 (0.0%)	0.80	24/32414 (0.1%)

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	G	1	0

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	G	258	GLU	CG-CD	6.18	1.61	1.51
1	C	4	CYS	CB-SG	-5.94	1.72	1.81
2	J	137	CYS	CB-SG	5.78	1.92	1.82
2	H	74	GLU	CG-CD	5.42	1.60	1.51

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	143	LYS	CB-CA-C	9.40	129.19	110.40
2	L	91	ILE	CG1-CB-CG2	-8.03	93.73	111.40
2	B	91	ILE	CG1-CB-CG2	-7.20	95.55	111.40
2	D	91	ILE	CG1-CB-CG2	-6.26	97.62	111.40
2	H	91	ILE	CG1-CB-CG2	-6.16	97.85	111.40

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
5	G	330	MAN	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	2533	0	2483	40	1
1	C	2518	0	2472	52	0
1	E	2518	0	2472	43	1
1	G	2514	0	2466	36	2
1	I	2511	0	2464	62	0
1	K	2506	0	2460	59	1
2	B	1368	0	1293	36	3
2	D	1386	0	1305	41	0
2	F	1386	0	1305	38	0
2	H	1377	0	1299	18	0
2	J	1377	0	1299	56	0
2	L	1386	0	1305	45	0
3	A	14	0	13	1	0
3	E	14	0	13	1	0
3	G	28	0	26	0	0
4	C	28	0	25	2	0
4	I	28	0	24	2	0
4	K	56	0	50	1	0
5	G	39	0	34	1	0
All	All	23587	0	22808	390	4

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 8.

The worst 5 of 390 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:I:45:ARG:HH11	1:I:45:ARG:HG3	1.07	1.10
1:E:45:ARG:HG3	1:E:45:ARG:HH11	1.09	1.10
1:I:5:ILE:HD13	2:J:24:TYR:CD1	1.87	1.09
1:I:4:CYS:O	2:J:24:TYR:HA	1.52	1.09
1:A:45:ARG:HG3	1:A:45:ARG:HH11	1.14	1.08

All (4) 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)
2:B:143:LYS:CE	1:G:224:GLU:OE1[1_466]	1.91	0.29
1:A:168:ASP:O	1:K:273:HIS:CE1[1_655]	1.98	0.22
2:B:169:ASN:OD1	1:G:222:ASP:OD2[1_466]	2.13	0.07
2:B:164:GLU:OE1	1:E:120:THR:O[1_556]	2.19	0.01

## 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	323/331 (98%)	305 (94%)	16 (5%)	2 (1%)	30	65
1	C	321/331 (97%)	303 (94%)	16 (5%)	2 (1%)	30	65
1	E	321/331 (97%)	303 (94%)	16 (5%)	2 (1%)	30	65
1	G	320/331 (97%)	303 (95%)	16 (5%)	1 (0%)	46	79
1	I	320/331 (97%)	299 (93%)	19 (6%)	2 (1%)	30	65
1	K	319/331 (96%)	300 (94%)	18 (6%)	1 (0%)	46	79

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	167/181 (92%)	163 (98%)	4 (2%)	0	100	100
2	D	169/181 (93%)	163 (96%)	6 (4%)	0	100	100
2	F	169/181 (93%)	164 (97%)	5 (3%)	0	100	100
2	H	168/181 (93%)	164 (98%)	4 (2%)	0	100	100
2	J	168/181 (93%)	162 (96%)	6 (4%)	0	100	100
2	L	169/181 (93%)	162 (96%)	7 (4%)	0	100	100
All	All	2934/3072 (96%)	2791 (95%)	133 (4%)	10 (0%)	46	79

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	261	ALA
1	G	261	ALA
1	I	261	ALA
1	K	261	ALA
1	C	261	ALA

### 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	284/289 (98%)	254 (89%)	30 (11%)	8	24
1	C	282/289 (98%)	254 (90%)	28 (10%)	10	28
1	E	282/289 (98%)	249 (88%)	33 (12%)	7	20
1	G	282/289 (98%)	251 (89%)	31 (11%)	8	23
1	I	281/289 (97%)	250 (89%)	31 (11%)	8	23
1	K	281/289 (97%)	249 (89%)	32 (11%)	7	21
2	B	146/156 (94%)	128 (88%)	18 (12%)	6	18
2	D	148/156 (95%)	131 (88%)	17 (12%)	7	21
2	F	148/156 (95%)	133 (90%)	15 (10%)	9	27
2	H	147/156 (94%)	129 (88%)	18 (12%)	6	18

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	J	147/156 (94%)	131 (89%)	16 (11%)	8	23
2	L	148/156 (95%)	132 (89%)	16 (11%)	8	23
All	All	2576/2670 (96%)	2291 (89%)	285 (11%)	8	23

5 of 285 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	F	38	LEU
1	G	175	VAL
1	K	259	ARG
2	F	75	LYS
1	G	61	ILE

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 ⓘ

11 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	$\# Z  > 2$
4	NAG	C	328	1,4	14,14,15	0.73	0	15,19,21	1.36	2 (13%)
4	NAG	C	329	4	14,14,15	0.90	1 (7%)	15,19,21	1.76	4 (26%)



Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
5	NAG	G	328	1,5	14,14,15	0.78	0	15,19,21	1.52	2 (13%)
5	NAG	G	329	5	14,14,15	0.75	0	15,19,21	2.09	5 (33%)
5	MAN	G	330	5	11,11,12	0.84	0	14,15,17	2.85	7 (50%)
4	NAG	I	328	1,4	14,14,15	0.53	0	15,19,21	2.09	3 (20%)
4	NAG	I	329	4	14,14,15	1.01	1 (7%)	15,19,21	1.38	2 (13%)
4	NAG	K	328	1,4	14,14,15	1.14	1 (7%)	15,19,21	2.14	4 (26%)
4	NAG	K	329	4	14,14,15	0.60	0	15,19,21	1.13	1 (6%)
4	NAG	K	330	1,4	14,14,15	1.78	1 (7%)	15,19,21	1.93	4 (26%)
4	NAG	K	331	4	14,14,15	0.98	1 (7%)	15,19,21	2.18	4 (26%)

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	C	328	1,4	-	0/6/23/26	0/1/1/1
4	NAG	C	329	4	-	0/6/23/26	0/1/1/1
5	NAG	G	328	1,5	-	0/6/23/26	0/1/1/1
5	NAG	G	329	5	-	0/6/23/26	0/1/1/1
5	MAN	G	330	5	1/1/4/5	0/2/19/22	0/1/1/1
4	NAG	I	328	1,4	-	0/6/23/26	0/1/1/1
4	NAG	I	329	4	-	0/6/23/26	0/1/1/1
4	NAG	K	328	1,4	-	0/6/23/26	0/1/1/1
4	NAG	K	329	4	-	0/6/23/26	0/1/1/1
4	NAG	K	330	1,4	-	0/6/23/26	0/1/1/1
4	NAG	K	331	4	-	0/6/23/26	0/1/1/1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	C	329	NAG	C1-C2	2.22	1.55	1.52
4	K	331	NAG	C1-C2	2.26	1.55	1.52
4	I	329	NAG	C1-C2	2.68	1.56	1.52
4	K	328	NAG	C1-C2	4.01	1.58	1.52
4	K	330	NAG	C1-C2	5.41	1.60	1.52

The worst 5 of 38 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	G	330	MAN	C3-C4-C5	-4.28	102.74	110.20
4	K	328	NAG	O4-C4-C3	-4.16	100.97	110.34
5	G	330	MAN	C2-C3-C4	-3.81	104.56	111.04
5	G	328	NAG	C3-C4-C5	-3.38	104.31	110.20
4	C	328	NAG	C2-N2-C7	-3.16	118.98	123.04

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
5	G	330	MAN	C1

There are no torsion outliers.

There are no ring outliers.

5 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	C	328	NAG	2	0
5	G	328	NAG	1	0
4	I	328	NAG	2	0
4	I	329	NAG	1	0
4	K	328	NAG	1	0

## 5.6 Ligand geometry [i](#)

4 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	$\# Z  > 2$
3	NAG	A	328	1	14,14,15	0.85	1 (7%)	15,19,21	1.69	3 (20%)
3	NAG	E	328	1	14,14,15	0.52	0	15,19,21	1.40	2 (13%)
3	NAG	G	331	1	14,14,15	1.53	2 (14%)	15,19,21	2.54	5 (33%)
3	NAG	G	332	1	14,14,15	0.97	1 (7%)	15,19,21	2.95	2 (13%)

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	328	1	-	0/6/23/26	0/1/1/1
3	NAG	E	328	1	-	0/6/23/26	0/1/1/1
3	NAG	G	331	1	1/1/5/7	0/6/23/26	0/1/1/1
3	NAG	G	332	1	-	0/6/23/26	0/1/1/1

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	328	NAG	C1-C2	2.17	1.55	1.52
3	G	331	NAG	C3-C2	2.20	1.57	1.52
3	G	332	NAG	C1-C2	2.52	1.56	1.52
3	G	331	NAG	C1-C2	4.72	1.59	1.52

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	G	331	NAG	C3-C2-N2	-3.73	101.63	110.56
3	A	328	NAG	C2-N2-C7	-3.22	118.91	123.04
3	G	331	NAG	C3-C4-C5	-3.20	104.63	110.20
3	E	328	NAG	C3-C4-C5	-2.96	105.04	110.20
3	G	332	NAG	O4-C4-C3	2.14	115.16	110.34

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
3	G	331	NAG	C1

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	328	NAG	1	0
3	E	328	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 [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	325/331 (98%)	-0.11	9 (2%) 56 44	7, 20, 33, 49	0
1	C	323/331 (97%)	-0.26	4 (1%) 81 73	7, 20, 33, 49	0
1	E	323/331 (97%)	-0.06	15 (4%) 36 25	8, 20, 33, 48	0
1	G	322/331 (97%)	-0.20	9 (2%) 56 44	8, 20, 33, 49	0
1	I	322/331 (97%)	0.20	17 (5%) 30 20	8, 20, 33, 48	0
1	K	321/331 (96%)	-0.07	9 (2%) 56 44	8, 20, 33, 49	0
2	B	169/181 (93%)	0.08	3 (1%) 71 61	4, 31, 50, 69	1 (0%)
2	D	171/181 (94%)	-0.03	1 (0%) 90 86	4, 32, 53, 69	0
2	F	171/181 (94%)	0.65	28 (16%) 2 1	4, 32, 50, 69	0
2	H	170/181 (93%)	0.15	7 (4%) 41 29	4, 32, 50, 69	0
2	J	170/181 (93%)	0.34	8 (4%) 35 24	4, 32, 50, 69	0
2	L	171/181 (94%)	0.70	28 (16%) 2 1	4, 32, 50, 69	0
All	All	2958/3072 (96%)	0.06	138 (4%) 35 24	4, 22, 44, 69	1 (0%)

The worst 5 of 138 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	I	71	SER	8.5
1	G	71	SER	7.7
1	I	139	ALA	7.3
2	L	164	GLU	6.9
1	K	3	LEU	6.3

### 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 ⓘ

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
4	NAG	K	328	14/15	0.92	0.18	0.40	61,68,73,82	0
5	NAG	G	328	14/15	0.94	0.18	-0.00	48,53,55,57	0
4	NAG	C	328	14/15	0.93	0.17	-0.01	47,52,60,66	0
4	NAG	K	330	14/15	0.68	0.22	-0.32	56,67,74,78	0
4	NAG	I	328	14/15	0.83	0.19	-0.43	52,60,68,69	0
5	MAN	G	330	11/12	0.68	0.24	-	77,83,85,86	0
4	NAG	K	329	14/15	0.89	0.21	-	89,94,101,101	0
4	NAG	C	329	14/15	0.83	0.17	-	73,77,78,79	0
4	NAG	K	331	14/15	0.73	0.29	-	81,84,88,88	0
4	NAG	I	329	14/15	0.73	0.25	-	76,80,82,82	0
5	NAG	G	329	14/15	0.91	0.20	-	45,58,63,71	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(Å <sup>2</sup> )	Q<0.9
3	NAG	E	328	14/15	0.95	0.19	0.64	46,51,53,55	0
3	NAG	A	328	14/15	0.90	0.17	0.10	49,55,59,59	0
3	NAG	G	331	14/15	0.63	0.27	-	59,72,74,76	0
3	NAG	G	332	14/15	0.72	0.21	-	64,74,80,81	0

### 6.5 Other polymers ⓘ

There are no such residues in this entry.