



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 01:40 PM GMT

PDB ID : 3UBQ  
Title : Influenza hemagglutinin from the 2009 pandemic in complex with ligand 3SLN  
Authors : Xu, R.; Wilson, I.A.  
Deposited on : 2011-10-24  
Resolution : 2.00 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

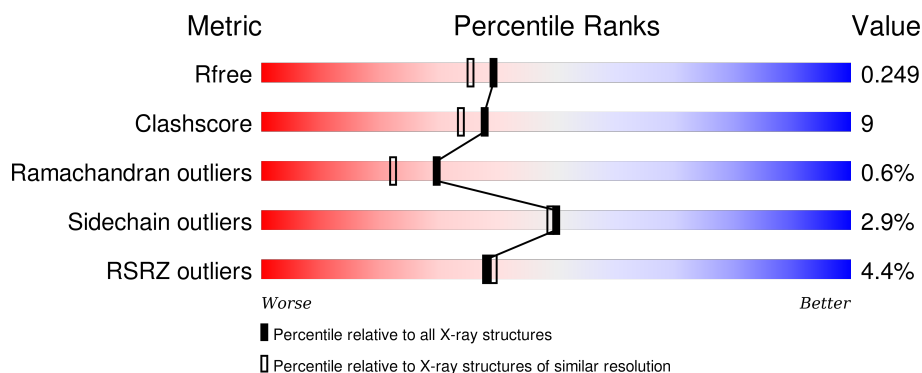
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.00 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	6249 (2.00-2.00)
Clashscore	102246	7340 (2.00-2.00)
Ramachandran outliers	100387	7248 (2.00-2.00)
Sidechain outliers	100360	7247 (2.00-2.00)
RSRZ outliers	91569	6262 (2.00-2.00)

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	329	<div> <div>2%</div> <div>75%</div> <div>21%</div> <div>..</div> </div>
1	C	329	<div> <div>2%</div> <div>74%</div> <div>22%</div> <div>..</div> </div>
1	E	329	<div> <div>2%</div> <div>76%</div> <div>20%</div> <div>..</div> </div>
1	G	329	<div> <div>%</div> <div>78%</div> <div>19%</div> <div>..</div> </div>
1	I	329	<div> <div>2%</div> <div>81%</div> <div>15%</div> <div>..</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	K	329	
2	B	177	
2	D	177	
2	F	177	
2	H	177	
2	J	177	
2	L	177	

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
4	NAG	K	651	-	-	-	X

## 2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 24675 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	323	Total	C	N	O	S	0	0	0
			2522	1595	432	482	13			
1	C	321	Total	C	N	O	S	0	0	0
			2509	1588	430	478	13			
1	E	321	Total	C	N	O	S	0	0	0
			2504	1585	430	476	13			
1	G	323	Total	C	N	O	S	0	0	0
			2522	1595	432	482	13			
1	I	319	Total	C	N	O	S	0	0	0
			2495	1579	428	475	13			
1	K	318	Total	C	N	O	S	0	0	0
			2490	1576	427	474	13			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	9	PRO	-	EXPRESSION TAG	UNP C3W5S1
A	10	GLY	-	EXPRESSION TAG	UNP C3W5S1
A	205	CYS	GLY	ENGINEERED MUTATION	UNP C3W5S1
A	220	CYS	ARG	ENGINEERED MUTATION	UNP C3W5S1
C	9	PRO	-	EXPRESSION TAG	UNP C3W5S1
C	10	GLY	-	EXPRESSION TAG	UNP C3W5S1
C	205	CYS	GLY	ENGINEERED MUTATION	UNP C3W5S1
C	220	CYS	ARG	ENGINEERED MUTATION	UNP C3W5S1
E	9	PRO	-	EXPRESSION TAG	UNP C3W5S1
E	10	GLY	-	EXPRESSION TAG	UNP C3W5S1
E	205	CYS	GLY	ENGINEERED MUTATION	UNP C3W5S1
E	220	CYS	ARG	ENGINEERED MUTATION	UNP C3W5S1
G	9	PRO	-	EXPRESSION TAG	UNP C3W5S1
G	10	GLY	-	EXPRESSION TAG	UNP C3W5S1
G	205	CYS	GLY	ENGINEERED MUTATION	UNP C3W5S1
G	220	CYS	ARG	ENGINEERED MUTATION	UNP C3W5S1
I	9	PRO	-	EXPRESSION TAG	UNP C3W5S1

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
I	10	GLY	-	EXPRESSION TAG	UNP C3W5S1
I	205	CYS	GLY	ENGINEERED MUTATION	UNP C3W5S1
I	220	CYS	ARG	ENGINEERED MUTATION	UNP C3W5S1
K	9	PRO	-	EXPRESSION TAG	UNP C3W5S1
K	10	GLY	-	EXPRESSION TAG	UNP C3W5S1
K	205	CYS	GLY	ENGINEERED MUTATION	UNP C3W5S1
K	220	CYS	ARG	ENGINEERED MUTATION	UNP C3W5S1

- Molecule 2 is a protein called hemagglutinin HA2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	169	Total	C	N	O	S	0	0	0
			1360	855	229	270	6			
2	D	170	Total	C	N	O	S	0	0	0
			1371	861	233	271	6			
2	F	170	Total	C	N	O	S	0	0	0
			1371	861	233	271	6			
2	H	172	Total	C	N	O	S	0	0	0
			1389	871	235	277	6			
2	J	170	Total	C	N	O	S	0	0	0
			1371	861	233	271	6			
2	L	169	Total	C	N	O	S	0	0	0
			1360	855	229	270	6			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	175	SER	-	EXPRESSION TAG	UNP C3W5S1
B	176	GLY	-	EXPRESSION TAG	UNP C3W5S1
B	177	ARG	-	EXPRESSION TAG	UNP C3W5S1
D	175	SER	-	EXPRESSION TAG	UNP C3W5S1
D	176	GLY	-	EXPRESSION TAG	UNP C3W5S1
D	177	ARG	-	EXPRESSION TAG	UNP C3W5S1
F	175	SER	-	EXPRESSION TAG	UNP C3W5S1
F	176	GLY	-	EXPRESSION TAG	UNP C3W5S1
F	177	ARG	-	EXPRESSION TAG	UNP C3W5S1
H	175	SER	-	EXPRESSION TAG	UNP C3W5S1
H	176	GLY	-	EXPRESSION TAG	UNP C3W5S1
H	177	ARG	-	EXPRESSION TAG	UNP C3W5S1
J	175	SER	-	EXPRESSION TAG	UNP C3W5S1
J	176	GLY	-	EXPRESSION TAG	UNP C3W5S1
J	177	ARG	-	EXPRESSION TAG	UNP C3W5S1

*Continued on next page...*

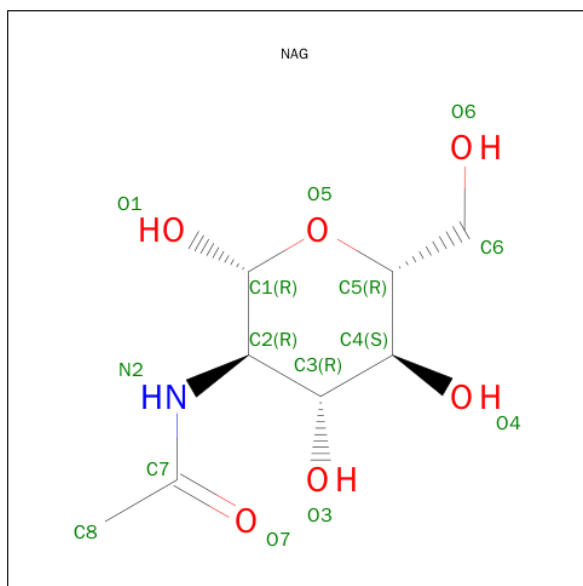
*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
L	175	SER	-	EXPRESSION TAG	UNP C3W5S1
L	176	GLY	-	EXPRESSION TAG	UNP C3W5S1
L	177	ARG	-	EXPRESSION TAG	UNP C3W5S1

- 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	E	2	Total	C	N	O	0	0
			28	16	2	10		
3	K	2	Total	C	N	O	0	0
			28	16	2	10		

- Molecule 4 is SUGAR (N-ACETYL-D-GLUCOSAMINE) (three-letter code: NAG) (formula:  $C_8H_{15}NO_6$ ).



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

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	G	1	Total	C	N	O	0	0
			14	8	1	5		
4	G	1	Total	C	N	O	0	0
			14	8	1	5		
4	I	1	Total	C	N	O	0	0
			14	8	1	5		
4	I	1	Total	C	N	O	0	0
			14	8	1	5		
4	I	1	Total	C	N	O	0	0
			14	8	1	5		
4	K	1	Total	C	N	O	0	0
			14	8	1	5		
4	K	1	Total	C	N	O	0	0
			14	8	1	5		
4	K	1	Total	C	N	O	0	0
			14	8	1	5		

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

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
5	C	3	Total	C	N	O	0	0
			46	25	2	19		

- Molecule 6 is a polymer of unknown type called SUGAR (2-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
6	I	2	Total	C	N	O	0	0
			31	17	1	13		
6	K	2	Total	C	N	O	0	0
			31	17	1	13		

- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	138	Total	O	0	0
			138	138		
7	B	54	Total	O	0	0
			54	54		
7	C	89	Total	O	0	0
			89	89		
7	D	66	Total	O	0	0
			66	66		

*Continued on next page...*

*Continued from previous page...*

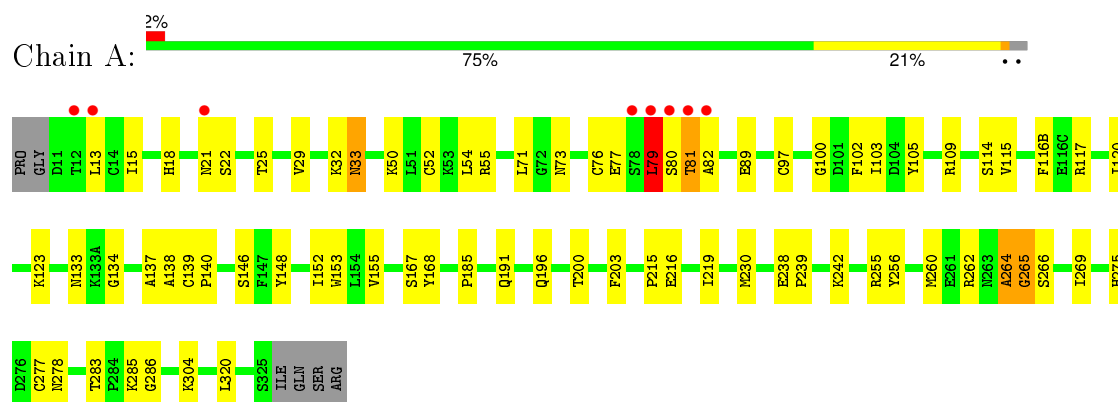
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	E	117	Total 117	O 117	0	0
7	F	62	Total 62	O 62	0	0
7	G	121	Total 121	O 121	0	0
7	H	69	Total 69	O 69	0	0
7	I	127	Total 127	O 127	0	0
7	J	57	Total 57	O 57	0	0
7	K	97	Total 97	O 97	0	0
7	L	54	Total 54	O 54	0	0



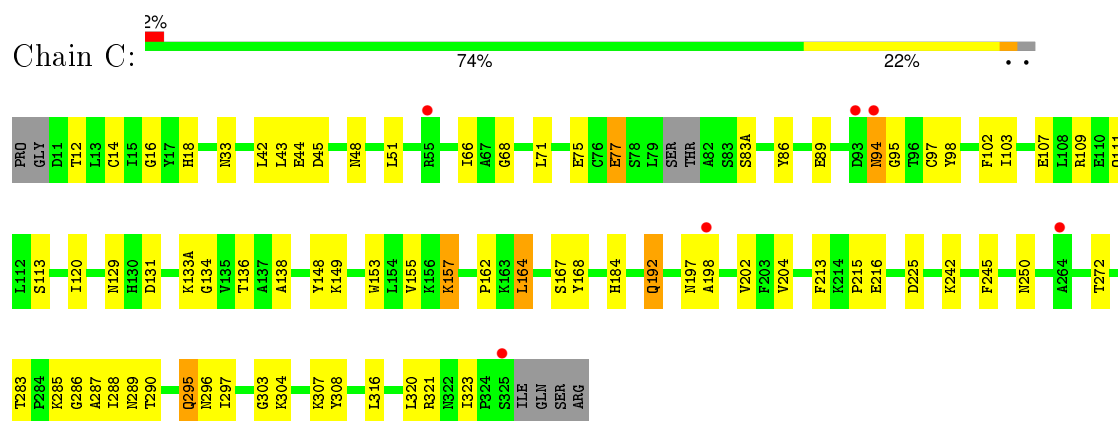
### 3 Residue-property plots [i](#)

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.

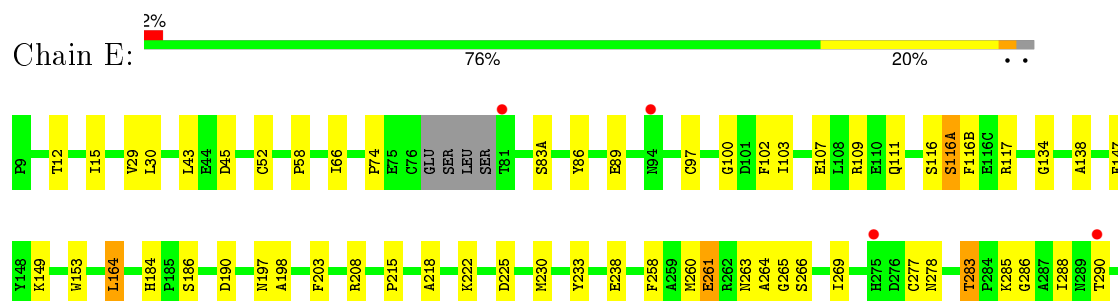
- Molecule 1: hemagglutinin HA1



- Molecule 1: hemagglutinin HA1

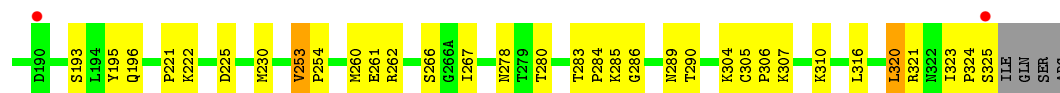
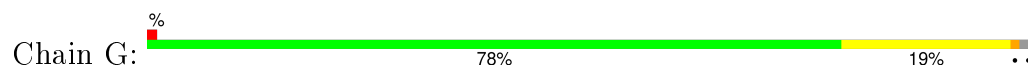


- Molecule 1: hemagglutinin HA1

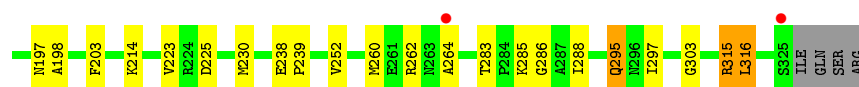
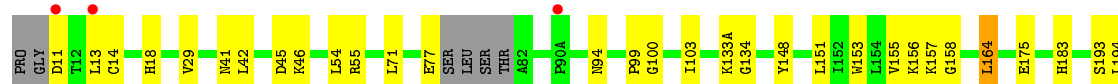
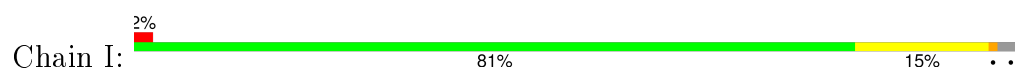




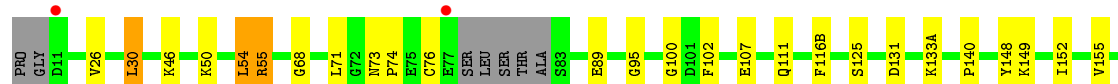
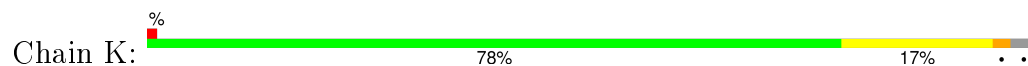
- Molecule 1: hemagglutinin HA1



- Molecule 1: hemagglutinin HA1



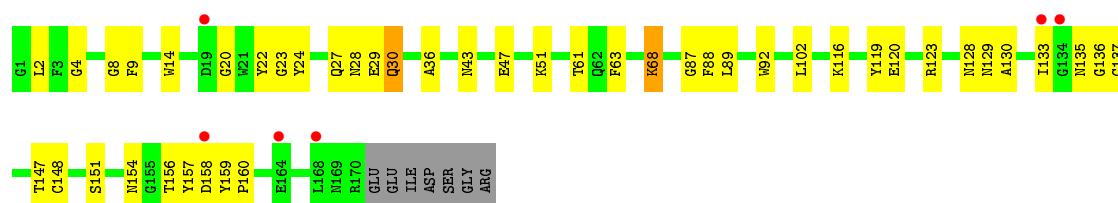
- Molecule 1: hemagglutinin HA1



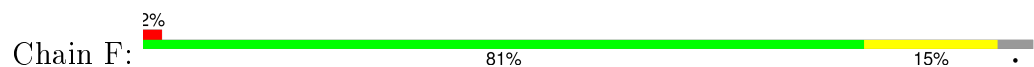
- Molecule 2: hemagglutinin HA2



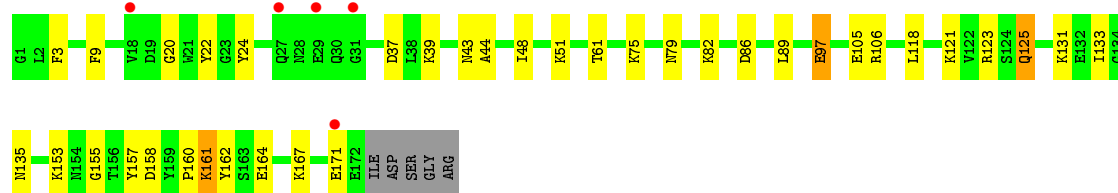
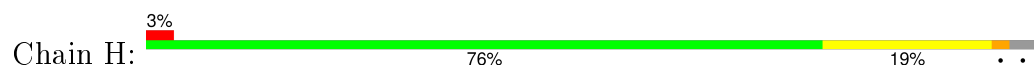
- Molecule 2: hemagglutinin HA2



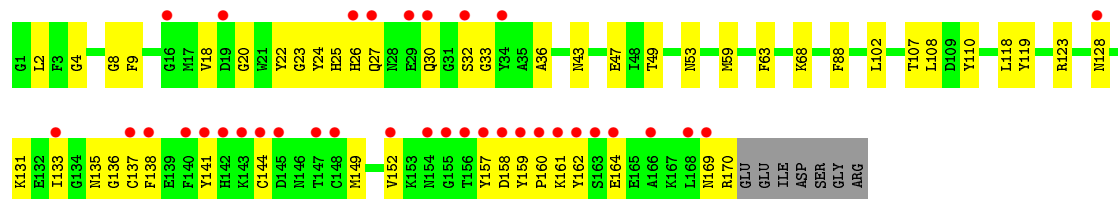
• Molecule 2: hemagglutinin HA2



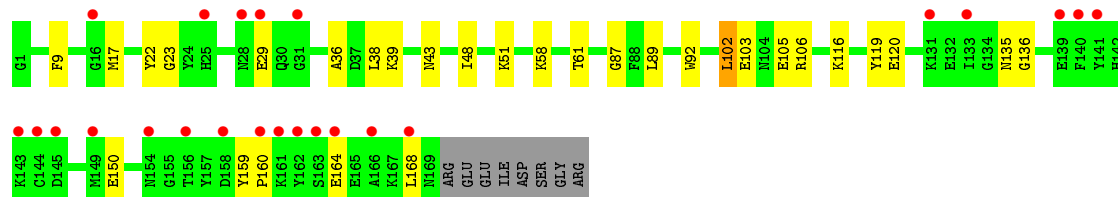
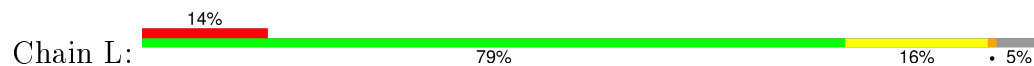
• Molecule 2: hemagglutinin HA2



• Molecule 2: hemagglutinin HA2



• Molecule 2: hemagglutinin HA2



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	66.88Å 116.26Å 118.37Å 60.96° 77.13° 80.40°	Depositor
Resolution (Å)	49.24 – 2.00 49.24 – 2.00	Depositor EDS
% Data completeness (in resolution range)	84.2 (49.24-2.00) 77.2 (49.24-2.00)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.67 (at 2.00Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7_650)	Depositor
R, $R_{free}$	0.199 , 0.249 0.199 , 0.249	Depositor DCC
$R_{free}$ test set	8650 reflections (5.28%)	DCC
Wilson B-factor (Å <sup>2</sup> )	29.5	Xtriage
Anisotropy	0.168	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 47.7	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 172543 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	24675	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	41.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.13% 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	0.30	0/2586	0.49	0/3516
1	C	0.27	0/2572	0.47	0/3495
1	E	0.30	0/2568	0.51	0/3490
1	G	0.31	0/2586	0.49	0/3516
1	I	0.30	0/2558	0.49	0/3476
1	K	0.28	0/2553	0.48	0/3469
2	B	0.27	0/1388	0.42	0/1871
2	D	0.28	0/1399	0.43	0/1885
2	F	0.29	0/1399	0.43	0/1885
2	H	0.29	0/1417	0.43	0/1909
2	J	0.27	0/1399	0.42	0/1885
2	L	0.27	0/1388	0.41	0/1871
All	All	0.29	0/23813	0.47	0/32268

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	2522	0	2466	61	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	2509	0	2455	56	0
1	E	2504	0	2449	54	0
1	G	2522	0	2465	52	0
1	I	2495	0	2436	39	0
1	K	2490	0	2430	49	0
2	B	1360	0	1284	31	0
2	D	1371	0	1296	36	0
2	F	1371	0	1297	21	0
2	H	1389	0	1309	33	0
2	J	1371	0	1297	35	0
2	L	1360	0	1284	20	0
3	A	28	0	25	0	0
3	E	28	0	25	0	0
3	K	28	0	25	2	0
4	A	14	0	13	1	0
4	D	14	0	13	0	0
4	E	14	0	13	0	0
4	G	42	0	39	3	0
4	I	42	0	39	0	0
4	K	42	0	39	0	0
5	C	46	0	40	1	0
6	I	31	0	26	2	0
6	K	31	0	26	1	0
7	A	138	0	0	5	0
7	B	54	0	0	1	0
7	C	89	0	0	2	0
7	D	66	0	0	1	0
7	E	117	0	0	3	0
7	F	62	0	0	0	0
7	G	121	0	0	2	0
7	H	69	0	0	2	0
7	I	127	0	0	1	0
7	J	57	0	0	0	0
7	K	97	0	0	1	0
7	L	54	0	0	2	0
All	All	24675	0	22791	426	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 9.

The worst 5 of 426 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)
2:H:97:GLU:HG2	2:L:58:LYS:HD2	1.37	1.06
1:C:283:THR:HG22	1:C:285:LYS:H	1.20	1.03
1:I:283:THR:HG22	1:I:285:LYS:H	1.20	1.02
1:A:283:THR:HG22	1:A:285:LYS:H	1.28	0.97
1:A:79:LEU:HD12	1:A:81:THR:HB	1.49	0.92

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	321/329 (98%)	303 (94%)	13 (4%)	5 (2%)	12	5
1	C	317/329 (96%)	301 (95%)	15 (5%)	1 (0%)	46	41
1	E	317/329 (96%)	306 (96%)	10 (3%)	1 (0%)	46	41
1	G	321/329 (98%)	305 (95%)	16 (5%)	0	100	100
1	I	315/329 (96%)	300 (95%)	12 (4%)	3 (1%)	19	11
1	K	314/329 (95%)	295 (94%)	17 (5%)	2 (1%)	30	22
2	B	167/177 (94%)	161 (96%)	5 (3%)	1 (1%)	30	22
2	D	168/177 (95%)	160 (95%)	6 (4%)	2 (1%)	16	8
2	F	168/177 (95%)	159 (95%)	9 (5%)	0	100	100
2	H	170/177 (96%)	164 (96%)	6 (4%)	0	100	100
2	J	168/177 (95%)	157 (94%)	9 (5%)	2 (1%)	16	8
2	L	167/177 (94%)	157 (94%)	10 (6%)	0	100	100
All	All	2913/3036 (96%)	2768 (95%)	128 (4%)	17 (1%)	30	22

5 of 17 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	81	THR

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	265	GLY
2	B	127	LYS
1	C	94	ASN
1	E	116(A)	SER

### 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	285/290 (98%)	280 (98%)	5 (2%)	66	69
1	C	283/290 (98%)	272 (96%)	11 (4%)	39	35
1	E	282/290 (97%)	273 (97%)	9 (3%)	46	44
1	G	285/290 (98%)	277 (97%)	8 (3%)	51	50
1	I	281/290 (97%)	274 (98%)	7 (2%)	55	55
1	K	281/290 (97%)	274 (98%)	7 (2%)	55	55
2	B	145/152 (95%)	139 (96%)	6 (4%)	37	32
2	D	146/152 (96%)	141 (97%)	5 (3%)	44	41
2	F	146/152 (96%)	144 (99%)	2 (1%)	74	77
2	H	148/152 (97%)	143 (97%)	5 (3%)	44	41
2	J	146/152 (96%)	140 (96%)	6 (4%)	37	32
2	L	145/152 (95%)	141 (97%)	4 (3%)	51	50
All	All	2573/2652 (97%)	2498 (97%)	75 (3%)	50	49

5 of 75 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	283	THR
1	G	115	VAL
1	K	283	THR
1	E	295	GLN
1	G	11	ASP



Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 18 such sidechains are listed below:

Mol	Chain	Res	Type
1	G	196	GLN
2	H	27	GLN
2	J	53	ASN
2	D	117	ASN
2	F	95	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 ⓘ

13 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	330	1,3	14,14,15	0.57	0	15,19,21	0.82	0
3	NAG	A	331	3	14,14,15	0.52	0	15,19,21	0.82	1 (6%)
5	SIA	C	330	5	16,20,21	0.23	0	18,28,31	1.04	1 (5%)
5	GAL	C	331	5	11,11,12	0.65	0	14,15,17	1.29	2 (14%)
5	NAG	C	332	5	15,15,15	0.49	0	17,21,21	0.60	0
3	NAG	E	331	1,3	14,14,15	0.49	0	15,19,21	0.60	0
3	NAG	E	332	3	14,14,15	0.51	0	15,19,21	0.90	1 (6%)
6	SIA	I	501	6	16,20,21	0.27	0	18,28,31	1.11	2 (11%)
6	GAL	I	502	6	11,11,12	0.64	0	14,15,17	1.06	0
6	SIA	K	601	6	16,20,21	0.24	0	18,28,31	1.03	2 (11%)
6	GAL	K	602	6	11,11,12	0.58	0	14,15,17	0.65	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
3	NAG	K	631	1,3	14,14,15	0.49	0	15,19,21	0.82	0
3	NAG	K	632	3	14,14,15	0.52	0	15,19,21	0.74	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	NAG	A	330	1,3	-	0/6/23/26	0/1/1/1
3	NAG	A	331	3	-	0/6/23/26	0/1/1/1
5	SIA	C	330	5	-	0/14/34/38	0/1/1/1
5	GAL	C	331	5	-	0/2/19/22	0/1/1/1
5	NAG	C	332	5	-	0/6/26/26	0/1/1/1
3	NAG	E	331	1,3	-	0/6/23/26	0/1/1/1
3	NAG	E	332	3	-	0/6/23/26	0/1/1/1
6	SIA	I	501	6	-	0/14/34/38	0/1/1/1
6	GAL	I	502	6	-	0/2/19/22	0/1/1/1
6	SIA	K	601	6	-	0/14/34/38	0/1/1/1
6	GAL	K	602	6	-	0/2/19/22	0/1/1/1
3	NAG	K	631	1,3	-	0/6/23/26	0/1/1/1
3	NAG	K	632	3	-	0/6/23/26	0/1/1/1

There are no bond length outliers.

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	C	331	GAL	O3-C3-C2	-3.62	103.47	110.00
3	A	331	NAG	C2-N2-C7	-2.41	119.94	123.04
3	E	332	NAG	C2-N2-C7	-2.34	120.03	123.04
6	I	501	SIA	C8-C7-C6	-2.22	108.55	113.01
6	K	601	SIA	C6-C5-N5	-2.21	107.22	111.07

There are no chirality outliers.

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
5	C	330	SIA	1	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	I	501	SIA	2	0
6	K	601	SIA	1	0
3	K	631	NAG	1	0
3	K	632	NAG	1	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	332	1	14,14,15	0.53	0	15,19,21	0.75	0
4	NAG	D	261	2	14,14,15	0.45	0	15,19,21	0.92	1 (6%)
4	NAG	E	341	1	14,14,15	0.49	0	15,19,21	0.99	1 (6%)
4	NAG	G	411	1	14,14,15	0.46	0	15,19,21	0.94	1 (6%)
4	NAG	G	431	1	14,14,15	0.51	0	15,19,21	1.09	2 (13%)
4	NAG	G	451	1	14,14,15	0.48	0	15,19,21	0.82	1 (6%)
4	NAG	I	521	1	14,14,15	0.44	0	15,19,21	0.74	0
4	NAG	I	531	1	14,14,15	0.47	0	15,19,21	1.09	1 (6%)
4	NAG	I	541	1	14,14,15	0.45	0	15,19,21	0.97	0
4	NAG	K	621	1	14,14,15	0.42	0	15,19,21	0.90	1 (6%)
4	NAG	K	641	1	14,14,15	0.49	0	15,19,21	0.80	1 (6%)
4	NAG	K	651	1	14,14,15	0.44	0	15,19,21	0.77	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	332	1	-	0/6/23/26	0/1/1/1
4	NAG	D	261	2	-	0/6/23/26	0/1/1/1
4	NAG	E	341	1	-	0/6/23/26	0/1/1/1

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	NAG	G	411	1	-	0/6/23/26	0/1/1/1
4	NAG	G	431	1	-	0/6/23/26	0/1/1/1
4	NAG	G	451	1	-	0/6/23/26	0/1/1/1
4	NAG	I	521	1	-	0/6/23/26	0/1/1/1
4	NAG	I	531	1	-	0/6/23/26	0/1/1/1
4	NAG	I	541	1	-	0/6/23/26	0/1/1/1
4	NAG	K	621	1	-	0/6/23/26	0/1/1/1
4	NAG	K	641	1	-	0/6/23/26	0/1/1/1
4	NAG	K	651	1	-	0/6/23/26	0/1/1/1

There are no bond length outliers.

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	G	431	NAG	C2-N2-C7	-2.56	119.74	123.04
4	E	341	NAG	C2-N2-C7	-2.15	120.28	123.04
4	D	261	NAG	C4-C3-C2	-2.10	107.96	111.23
4	K	641	NAG	C2-N2-C7	-2.04	120.42	123.04
4	K	621	NAG	C1-O5-C5	2.23	115.08	112.25

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	332	NAG	1	0
4	G	431	NAG	1	0
4	G	451	NAG	2	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	323/329 (98%)	-0.01	8 (2%) 61 61	22, 35, 56, 75	0
1	C	321/329 (97%)	-0.09	6 (1%) 70 70	27, 42, 58, 76	0
1	E	321/329 (97%)	-0.08	5 (1%) 74 75	24, 35, 51, 69	0
1	G	323/329 (98%)	-0.08	2 (0%) 90 90	24, 36, 50, 70	0
1	I	319/329 (96%)	-0.21	5 (1%) 74 75	24, 36, 54, 76	0
1	K	318/329 (96%)	-0.03	2 (0%) 90 90	27, 40, 58, 74	0
2	B	169/177 (95%)	1.00	29 (17%) 2 2	23, 45, 90, 106	0
2	D	170/177 (96%)	0.24	6 (3%) 48 49	25, 46, 63, 75	0
2	F	170/177 (96%)	0.11	4 (2%) 62 63	24, 40, 60, 67	0
2	H	172/177 (97%)	0.03	5 (2%) 55 56	24, 43, 58, 68	0
2	J	170/177 (96%)	0.80	35 (20%) 1 1	24, 46, 78, 83	0
2	L	169/177 (95%)	0.56	24 (14%) 4 4	26, 45, 81, 94	0
All	All	2945/3036 (97%)	0.10	131 (4%) 38 39	22, 39, 68, 106	0

The worst 5 of 131 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	L	168	LEU	9.1
2	B	166	ALA	7.7
2	B	160	PRO	7.6
2	B	168	LEU	7.5
2	J	147	THR	7.1

### 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(Å <sup>2</sup> )	Q<0.9
3	NAG	E	331	14/15	0.79	0.25	1.47	55,63,71,71	0
6	SIA	K	601	20/21	0.88	0.13	0.37	42,51,57,57	0
6	SIA	I	501	20/21	0.91	0.11	0.32	36,45,49,51	0
3	NAG	K	631	14/15	0.92	0.12	-0.15	51,58,66,69	0
5	SIA	C	330	20/21	0.94	0.10	-0.76	33,43,47,50	0
3	NAG	A	330	14/15	0.96	0.08	-1.60	32,39,46,48	0
6	GAL	K	602	11/12	0.85	0.20	-	59,70,77,77	0
3	NAG	A	331	14/15	0.89	0.14	-	44,53,62,67	0
5	GAL	C	331	11/12	0.82	0.13	-	53,65,70,76	0
6	GAL	I	502	11/12	0.85	0.13	-	55,60,68,69	0
5	NAG	C	332	15/15	0.83	0.26	-	74,83,91,94	0
3	NAG	K	632	14/15	0.81	0.18	-	67,75,84,85	0
3	NAG	E	332	14/15	0.77	0.36	-	62,77,84,85	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
4	NAG	K	651	14/15	0.72	0.23	5.59	53,74,82,85	0
4	NAG	I	531	14/15	0.84	0.15	1.57	41,53,63,64	0
4	NAG	I	541	14/15	0.86	0.15	1.04	57,67,75,80	0
4	NAG	G	431	14/15	0.92	0.10	-0.38	40,52,62,64	0
4	NAG	I	521	14/15	0.71	0.19	-	73,83,88,92	0
4	NAG	G	411	14/15	0.85	0.31	-	70,75,82,84	0
4	NAG	A	332	14/15	0.76	0.38	-	51,62,69,70	0
4	NAG	G	451	14/15	0.69	0.38	-	65,84,90,94	0
4	NAG	K	641	14/15	0.82	0.20	-	57,63,74,78	0
4	NAG	D	261	14/15	0.86	0.14	-	65,68,72,76	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	K	621	14/15	0.79	0.18	-	53,67,77,83	0
4	NAG	E	341	14/15	0.76	0.32	-	76,79,87,89	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.