



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 01:39 PM GMT

PDB ID : 3UBE  
Title : Influenza hemagglutinin from the 2009 pandemic in complex with ligand LSTc  
Authors : Xu, R.; Wilson, I.A.  
Deposited on : 2011-10-24  
Resolution : 2.15 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

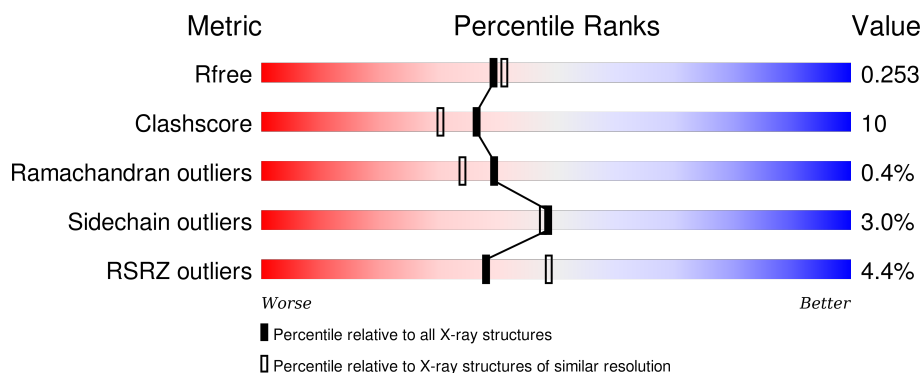
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.15 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	1045 (2.16-2.16)
Clashscore	102246	1152 (2.16-2.16)
Ramachandran outliers	100387	1131 (2.16-2.16)
Sidechain outliers	100360	1131 (2.16-2.16)
RSRZ outliers	91569	1050 (2.16-2.16)

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>81% 16% ..</div> </div>
1	C	329	<div> <div>2%</div> <div>83% 13% ..</div> </div>
1	E	329	<div> <div>2%</div> <div>76% 21% ..</div> </div>
1	G	329	<div> <div>3%</div> <div>76% 21% ..</div> </div>
1	I	329	<div> <div>2%</div> <div>73% 21% ..</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	A	336	-	-	-	X

## 2 Entry composition [i](#)

There are 8 unique types of molecules in this entry. The entry contains 24588 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	320	Total	C	N	O	S	0	0	0
			2501	1582	429	477	13			
1	E	323	Total	C	N	O	S	0	0	0
			2522	1595	432	482	13			
1	G	321	Total	C	N	O	S	0	0	0
			2508	1586	430	479	13			
1	I	319	Total	C	N	O	S	0	0	0
			2495	1579	428	475	13			
1	K	320	Total	C	N	O	S	0	0	0
			2497	1581	429	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	174	Total	C	N	O	S	0	0	0
			1405	881	237	281	6			
2	D	171	Total	C	N	O	S	0	0	0
			1380	866	234	274	6			
2	F	170	Total	C	N	O	S	0	0	0
			1371	861	233	271	6			
2	H	170	Total	C	N	O	S	0	0	0
			1371	861	233	271	6			
2	J	169	Total	C	N	O	S	0	0	0
			1360	855	229	270	6			
2	L	170	Total	C	N	O	S	0	0	0
			1371	861	233	271	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...*

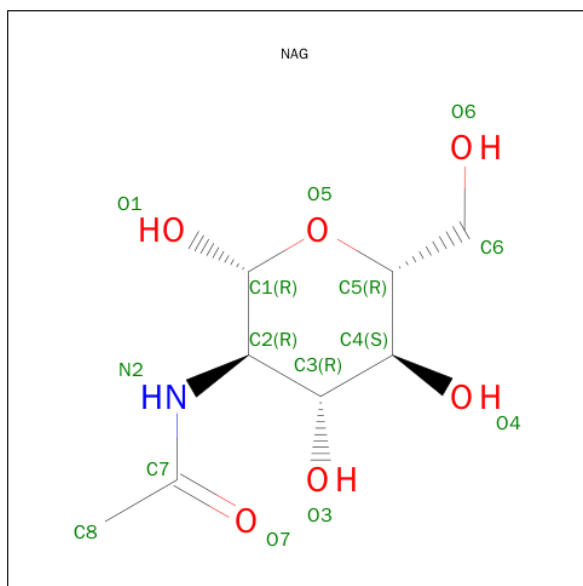
Continued from previous page...

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 (4-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	4	Total	C	N	O	0	0
			57	31	2	24		
3	C	4	Total	C	N	O	0	0
			57	31	2	24		
3	K	4	Total	C	N	O	0	0
			57	31	2	24		

- 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	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	C	1	Total	C	N	O	0	0
			14	8	1	5		

Continued on next page...

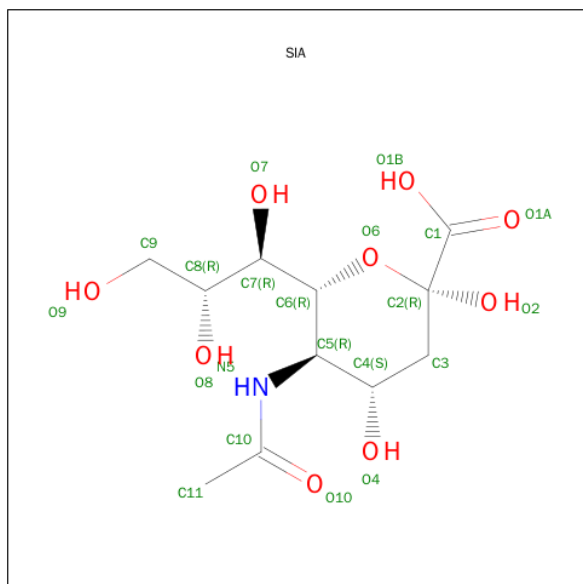
Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	E	1	Total	C	N	O	0	0
			14	8	1	5		
4	J	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	E	3	Total	C	N	O	0	0
			45	25	2	18		
5	I	3	Total	C	N	O	0	0
			45	25	2	18		

- Molecule 6 is SUGAR (O-SIALIC ACID) (three-letter code: SIA) (formula:  $C_{11}H_{19}NO_9$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
6	G	1	Total	C	N	O	0	0
			20	11	1	8		

- Molecule 7 is a polymer of unknown type called SUGAR (2-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
7	G	2	Total	C	N	O	0	0
			28	16	2	10		

- Molecule 8 is water.

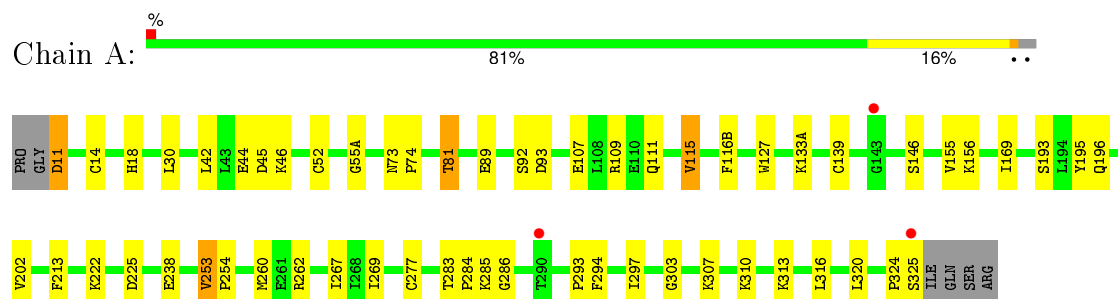
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	99	Total 99	O 99	0	0
8	B	54	Total 54	O 54	0	0
8	C	124	Total 124	O 124	0	0
8	D	48	Total 48	O 48	0	0
8	E	85	Total 85	O 85	0	0
8	F	49	Total 49	O 49	0	0
8	G	105	Total 105	O 105	0	0
8	H	45	Total 45	O 45	0	0
8	I	90	Total 90	O 90	0	0
8	J	50	Total 50	O 50	0	0
8	K	71	Total 71	O 71	0	0
8	L	58	Total 58	O 58	0	0



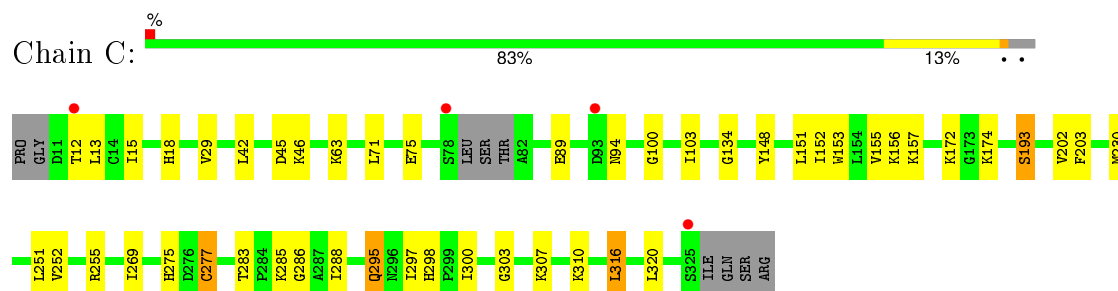
### 3 Residue-property plots [i](#)

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.

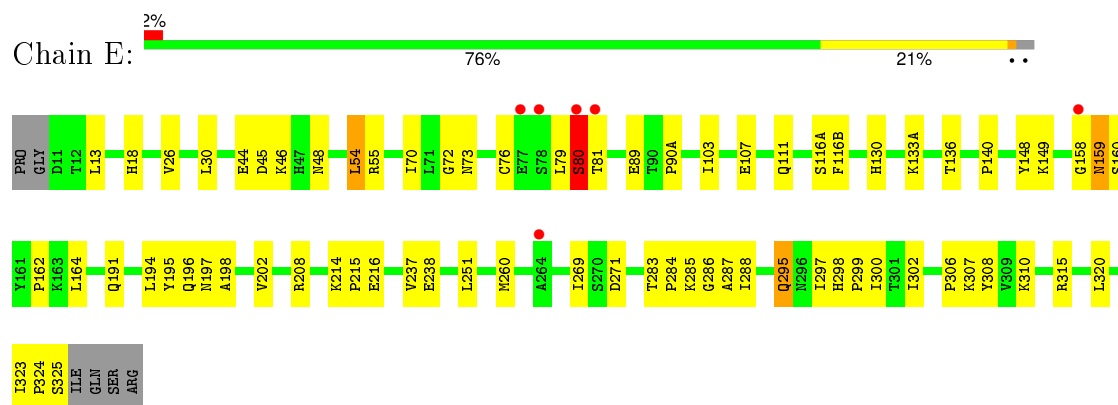
#### • Molecule 1: Hemagglutinin HA1



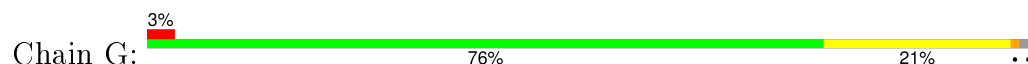
#### • Molecule 1: Hemagglutinin HA1

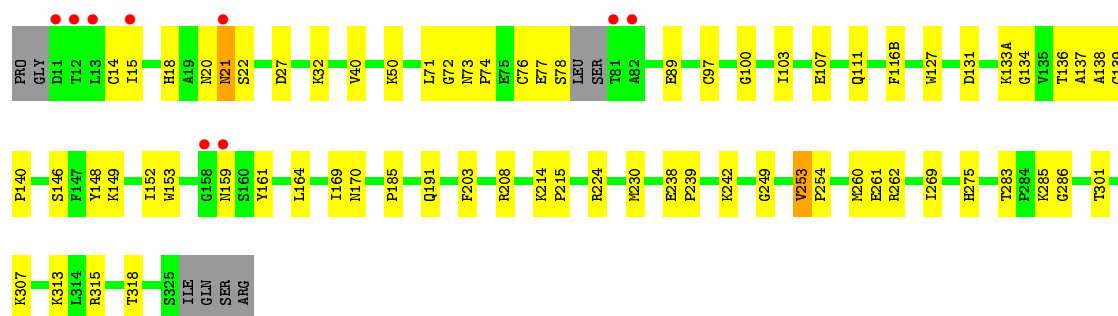


#### • Molecule 1: Hemagglutinin HA1

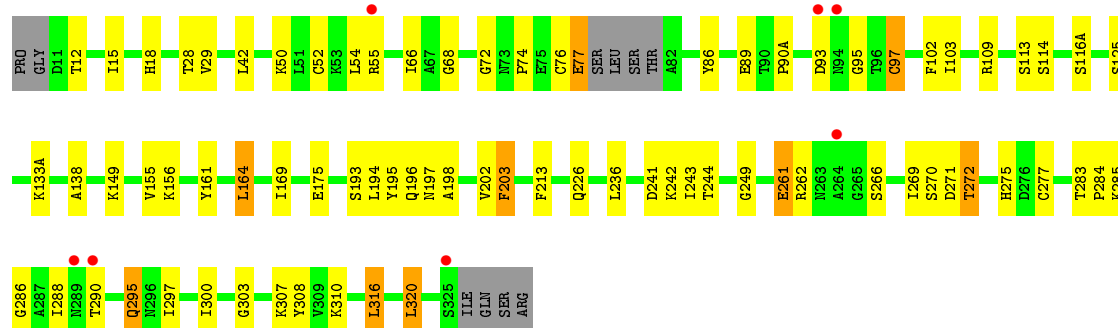


#### • Molecule 1: Hemagglutinin HA1

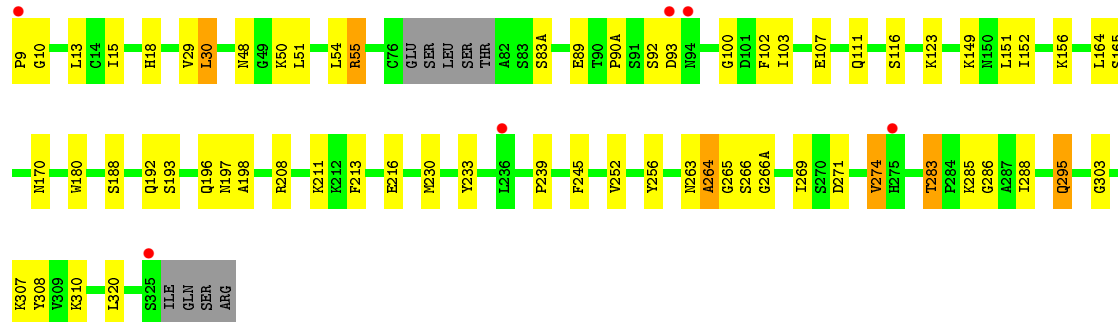
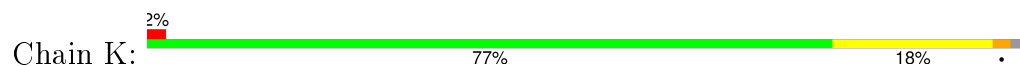




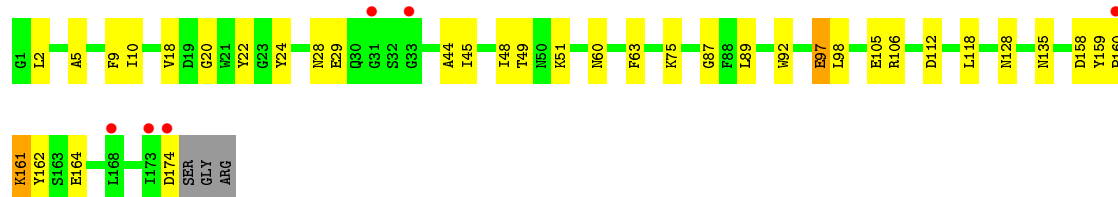
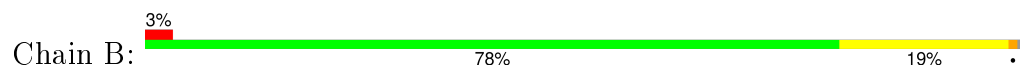
● Molecule 1: Hemagglutinin HA1



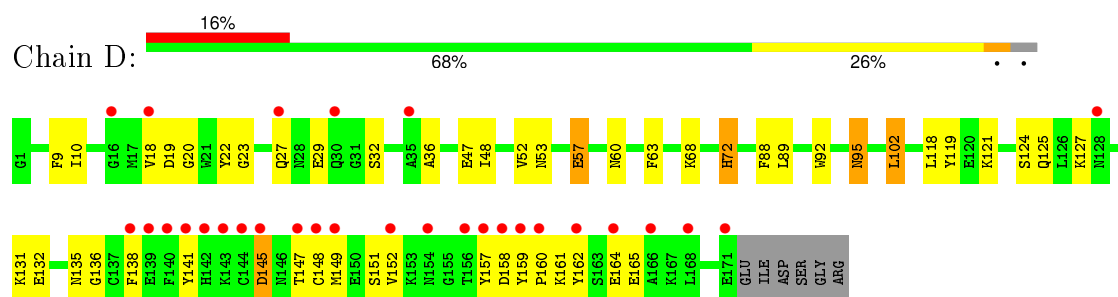
● Molecule 1: Hemagglutinin HA1



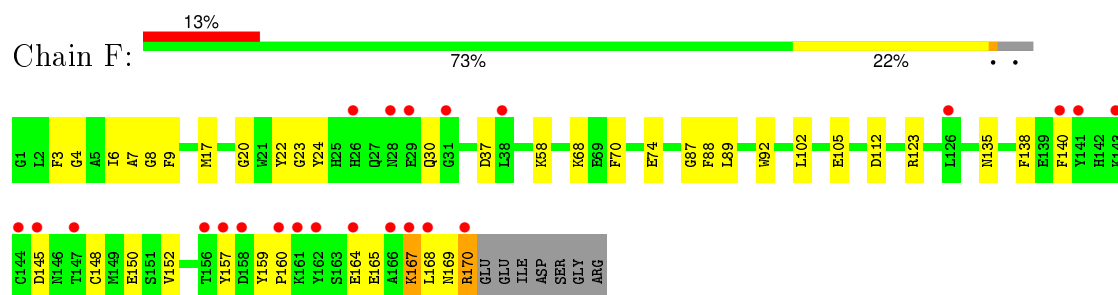
● Molecule 2: Hemagglutinin HA2



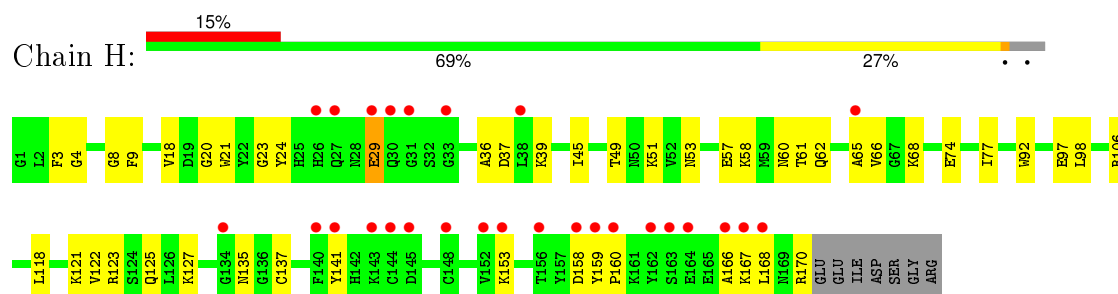
● Molecule 2: Hemagglutinin HA2



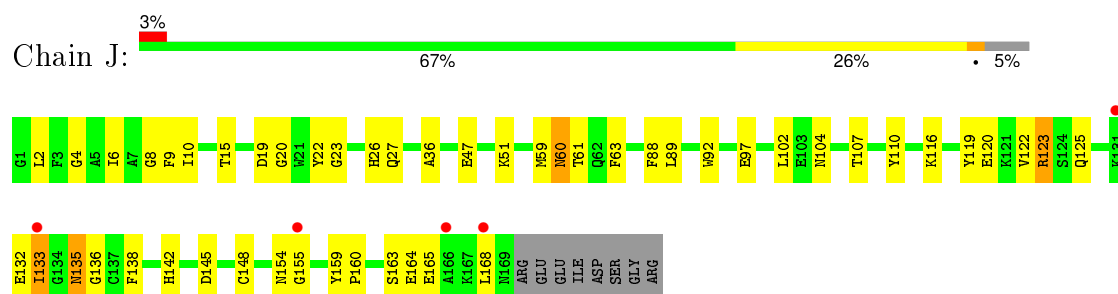
• Molecule 2: Hemagglutinin HA2



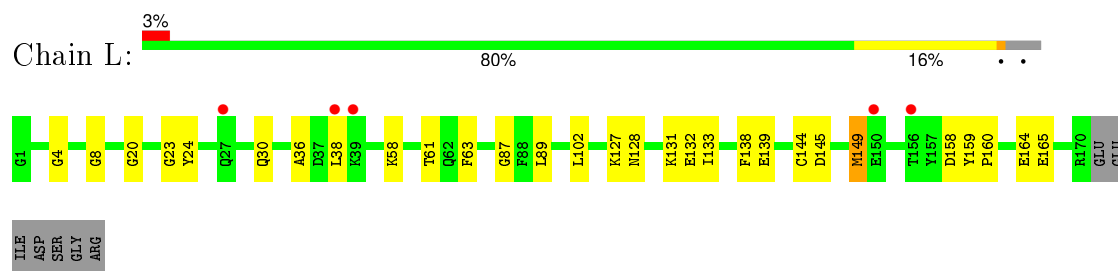
• Molecule 2: Hemagglutinin HA2



• 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, $\alpha$ , $\beta$ , $\gamma$	66.78Å 116.59Å 118.72Å 60.86° 77.29° 80.71°	Depositor
Resolution (Å)	49.54 – 2.15 49.54 – 2.15	Depositor EDS
% Data completeness (in resolution range)	92.4 (49.54-2.15) 79.4 (49.54-2.15)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.50 (at 2.14Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7_650)	Depositor
R, $R_{free}$	0.193 , 0.252 0.192 , 0.253	Depositor DCC
$R_{free}$ test set	7691 reflections (5.29%)	DCC
Wilson B-factor (Å <sup>2</sup> )	33.5	Xtriage
Anisotropy	0.169	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 50.0	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 153138 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	24588	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.73% 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	0.27	0/2586	0.46	0/3516
1	C	0.26	0/2564	0.46	0/3484
1	E	0.26	0/2586	0.47	0/3516
1	G	0.26	0/2571	0.45	0/3494
1	I	0.25	0/2558	0.46	0/3476
1	K	0.26	0/2561	0.47	0/3480
2	B	0.25	0/1433	0.40	0/1931
2	D	0.26	0/1408	0.41	0/1897
2	F	0.25	0/1399	0.40	0/1885
2	H	0.25	0/1399	0.41	0/1885
2	J	0.25	0/1388	0.42	0/1871
2	L	0.26	0/1399	0.40	0/1885
All	All	0.26	0/23852	0.44	0/32320

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	2465	39	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	2501	0	2443	32	0
1	E	2522	0	2467	63	0
1	G	2508	0	2450	49	0
1	I	2495	0	2439	67	0
1	K	2497	0	2443	50	0
2	B	1405	0	1324	28	0
2	D	1380	0	1303	40	0
2	F	1371	0	1297	28	0
2	H	1371	0	1297	50	0
2	J	1360	0	1283	42	0
2	L	1371	0	1297	23	0
3	A	57	0	49	1	0
3	C	57	0	49	0	0
3	K	57	0	49	0	0
4	A	42	0	39	1	0
4	C	14	0	13	1	0
4	E	14	0	13	1	0
4	J	14	0	13	0	0
4	K	14	0	13	0	0
5	E	45	0	38	1	0
5	I	45	0	38	0	0
6	G	20	0	17	0	0
7	G	28	0	25	0	0
8	A	99	0	0	2	0
8	B	54	0	0	3	0
8	C	124	0	0	2	0
8	D	48	0	0	0	0
8	E	85	0	0	1	0
8	F	49	0	0	0	0
8	G	105	0	0	0	0
8	H	45	0	0	1	0
8	I	90	0	0	0	0
8	J	50	0	0	1	0
8	K	71	0	0	3	0
8	L	58	0	0	1	0
All	All	24588	0	22864	443	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 10.

The worst 5 of 443 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:J:59:MET:HA	2:J:60:ASN:CB	1.77	1.11
2:J:59:MET:CA	2:J:60:ASN:HB2	1.82	1.09
1:E:79:LEU:HA	1:E:80:SER:HB3	1.13	1.08
1:E:79:LEU:CA	1:E:80:SER:HB3	1.88	1.03
1:C:283:THR:HG22	1:C:285:LYS:H	1.27	0.99

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%)	17 (5%)	1 (0%)	46	42
1	C	316/329 (96%)	304 (96%)	11 (4%)	1 (0%)	46	42
1	E	321/329 (98%)	303 (94%)	15 (5%)	3 (1%)	21	13
1	G	317/329 (96%)	299 (94%)	18 (6%)	0	100	100
1	I	315/329 (96%)	297 (94%)	17 (5%)	1 (0%)	46	42
1	K	316/329 (96%)	304 (96%)	10 (3%)	2 (1%)	30	21
2	B	172/177 (97%)	168 (98%)	4 (2%)	0	100	100
2	D	169/177 (96%)	159 (94%)	8 (5%)	2 (1%)	16	9
2	F	168/177 (95%)	155 (92%)	13 (8%)	0	100	100
2	H	168/177 (95%)	155 (92%)	13 (8%)	0	100	100
2	J	167/177 (94%)	157 (94%)	8 (5%)	2 (1%)	16	9
2	L	168/177 (95%)	163 (97%)	5 (3%)	0	100	100
All	All	2918/3036 (96%)	2767 (95%)	139 (5%)	12 (0%)	39	34

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	116(A)	SER

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	E	159	ASN
1	A	92	SER
1	E	80	SER
1	I	93	ASP

### 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%)	279 (98%)	6 (2%)	61	65
1	C	282/290 (97%)	276 (98%)	6 (2%)	61	65
1	E	285/290 (98%)	278 (98%)	7 (2%)	55	58
1	G	283/290 (98%)	276 (98%)	7 (2%)	55	58
1	I	281/290 (97%)	267 (95%)	14 (5%)	30	25
1	K	281/290 (97%)	274 (98%)	7 (2%)	55	58
2	B	150/152 (99%)	143 (95%)	7 (5%)	32	28
2	D	147/152 (97%)	141 (96%)	6 (4%)	37	35
2	F	146/152 (96%)	141 (97%)	5 (3%)	44	42
2	H	146/152 (96%)	143 (98%)	3 (2%)	61	65
2	J	145/152 (95%)	139 (96%)	6 (4%)	37	35
2	L	146/152 (96%)	143 (98%)	3 (2%)	61	65
All	All	2577/2652 (97%)	2500 (97%)	77 (3%)	48	48

5 of 77 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	F	150	GLU
1	G	315	ARG
1	K	283	THR
2	F	167	LYS
1	G	159	ASN



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	159	ASN
2	H	43	ASN
2	J	95	ASN
1	E	298	HIS
2	F	30	GLN

### 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 ⓘ

20 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	SIA	A	330	3	16,20,21	0.28	0	18,28,31	0.97	1 (5%)
3	GAL	A	331	3	12,12,12	0.53	0	17,17,17	0.48	0
3	NAG	A	332	3	14,14,15	0.55	0	15,19,21	1.13	1 (6%)
3	GAL	A	333	3	11,11,12	0.60	0	14,15,17	0.79	1 (7%)
3	SIA	C	330	3	16,20,21	0.26	0	18,28,31	1.26	1 (5%)
3	GAL	C	331	3	11,11,12	0.65	0	14,15,17	0.76	0
3	NAG	C	332	3	14,14,15	0.59	0	15,19,21	0.98	1 (6%)
3	GAL	C	333	3	12,12,12	0.59	0	17,17,17	0.58	0
5	SIA	E	330	5	16,20,21	0.23	0	18,28,31	0.98	2 (11%)
5	NAG	E	331	5	14,14,15	0.54	0	15,19,21	0.76	0
5	GAL	E	332	5	11,11,12	0.57	0	14,15,17	0.81	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
7	NAG	G	431	1,7	14,14,15	0.52	0	15,19,21	0.85	0
7	NAG	G	432	7	14,14,15	0.62	0	15,19,21	1.40	3 (20%)
5	SIA	I	501	5	16,20,21	0.21	0	18,28,31	1.26	2 (11%)
5	GAL	I	502	5	11,11,12	0.64	0	14,15,17	0.71	0
5	NAG	I	503	5	14,14,15	0.54	0	15,19,21	0.67	0
3	SIA	K	601	3	16,20,21	0.31	0	18,28,31	1.33	1 (5%)
3	GAL	K	602	3	11,11,12	0.64	0	14,15,17	0.70	0
3	NAG	K	603	3	14,14,15	0.64	0	15,19,21	0.85	0
3	GAL	K	604	3	12,12,12	0.53	0	17,17,17	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
3	SIA	A	330	3	-	0/14/34/38	0/1/1/1
3	GAL	A	331	3	-	0/2/22/22	0/1/1/1
3	NAG	A	332	3	-	0/6/23/26	0/1/1/1
3	GAL	A	333	3	-	0/2/19/22	0/1/1/1
3	SIA	C	330	3	-	0/14/34/38	0/1/1/1
3	GAL	C	331	3	-	0/2/19/22	0/1/1/1
3	NAG	C	332	3	-	0/6/23/26	0/1/1/1
3	GAL	C	333	3	-	0/2/22/22	0/1/1/1
5	SIA	E	330	5	-	0/14/34/38	0/1/1/1
5	NAG	E	331	5	-	0/6/23/26	0/1/1/1
5	GAL	E	332	5	-	0/2/19/22	0/1/1/1
7	NAG	G	431	1,7	-	0/6/23/26	0/1/1/1
7	NAG	G	432	7	-	0/6/23/26	0/1/1/1
5	SIA	I	501	5	-	0/14/34/38	0/1/1/1
5	GAL	I	502	5	-	0/2/19/22	0/1/1/1
5	NAG	I	503	5	-	0/6/23/26	0/1/1/1
3	SIA	K	601	3	-	0/14/34/38	0/1/1/1
3	GAL	K	602	3	-	0/2/19/22	0/1/1/1
3	NAG	K	603	3	-	0/6/23/26	0/1/1/1
3	GAL	K	604	3	-	0/2/22/22	0/1/1/1

There are no bond length outliers.

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
7	G	432	NAG	C2-N2-C7	-3.06	119.11	123.04
5	I	501	SIA	C7-C6-C5	-2.77	110.14	114.32
5	E	330	SIA	C7-C6-C5	-2.26	110.91	114.32
3	A	333	GAL	O5-C1-C2	-2.04	107.54	110.86
3	A	330	SIA	O6-C6-C5	2.44	112.49	108.48

There are no chirality outliers.

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	330	SIA	1	0
5	E	330	SIA	1	0

## 5.6 Ligand geometry

8 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	334	1	14,14,15	0.56	0	15,19,21	0.88	0
4	NAG	A	335	1	14,14,15	0.41	0	15,19,21	1.13	2 (13%)
4	NAG	A	336	1	14,14,15	0.58	0	15,19,21	0.84	0
4	NAG	C	334	1	14,14,15	0.48	0	15,19,21	0.80	0
4	NAG	E	333	1	14,14,15	0.47	0	15,19,21	0.92	0
6	SIA	G	401	-	16,20,21	0.26	0	18,28,31	0.95	2 (11%)
4	NAG	J	561	2	14,14,15	0.47	0	15,19,21	0.79	0
4	NAG	K	641	1	14,14,15	0.40	0	15,19,21	1.56	3 (20%)

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	334	1	-	0/6/23/26	0/1/1/1
4	NAG	A	335	1	-	0/6/23/26	0/1/1/1
4	NAG	A	336	1	-	0/6/23/26	0/1/1/1
4	NAG	C	334	1	-	0/6/23/26	0/1/1/1
4	NAG	E	333	1	-	0/6/23/26	0/1/1/1
6	SIA	G	401	-	-	0/14/34/38	0/1/1/1
4	NAG	J	561	2	-	0/6/23/26	0/1/1/1
4	NAG	K	641	1	-	0/6/23/26	0/1/1/1

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	K	641	NAG	C4-C3-C2	-2.85	106.80	111.23
4	K	641	NAG	C2-N2-C7	-2.77	119.49	123.04
4	A	335	NAG	C2-N2-C7	-2.21	120.20	123.04
6	G	401	SIA	C7-C6-C5	-2.03	111.25	114.32
6	G	401	SIA	O6-C6-C5	2.45	112.50	108.48

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	336	NAG	1	0
4	C	334	NAG	1	0
4	E	333	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	323/329 (98%)	0.08	3 (0%) 85 89	21, 34, 52, 71	0
1	C	320/329 (97%)	-0.16	4 (1%) 79 84	20, 33, 51, 76	0
1	E	323/329 (98%)	0.07	6 (1%) 70 78	24, 38, 58, 80	0
1	G	321/329 (97%)	0.04	9 (2%) 56 66	22, 35, 56, 71	0
1	I	319/329 (96%)	-0.00	7 (2%) 65 73	28, 41, 57, 77	0
1	K	320/329 (97%)	0.07	6 (1%) 70 78	22, 35, 53, 71	0
2	B	174/177 (98%)	0.17	6 (3%) 49 59	24, 43, 66, 82	0
2	D	171/177 (96%)	0.69	29 (16%) 2 4	23, 43, 76, 96	0
2	F	170/177 (96%)	0.70	23 (13%) 4 7	22, 44, 80, 95	0
2	H	170/177 (96%)	0.86	27 (15%) 3 4	23, 47, 84, 95	0
2	J	169/177 (95%)	0.31	5 (2%) 54 64	22, 46, 68, 77	0
2	L	170/177 (96%)	0.24	5 (2%) 55 65	22, 41, 61, 69	0
All	All	2950/3036 (97%)	0.18	130 (4%) 38 49	20, 38, 68, 96	0

The worst 5 of 130 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	168	LEU	7.6
2	H	144	CYS	6.0
2	D	160	PRO	5.9
2	D	152	VAL	5.5
2	H	164	GLU	5.3

### 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
3	SIA	A	330	20/21	0.95	0.11	-0.08	31,37,42,46	0
3	GAL	K	602	11/12	0.90	0.16	-0.10	40,56,61,69	0
3	SIA	K	601	20/21	0.96	0.11	-0.35	33,40,45,46	0
3	GAL	A	333	11/12	0.93	0.13	-0.37	39,52,58,62	0
5	SIA	I	501	20/21	0.96	0.10	-0.49	29,38,42,42	0
5	SIA	E	330	20/21	0.94	0.10	-0.55	32,45,50,53	0
7	NAG	G	431	14/15	0.95	0.10	-0.98	32,39,46,46	0
3	SIA	C	330	20/21	0.96	0.08	-1.44	27,33,39,42	0
3	GAL	C	333	12/12	0.80	0.21	-	55,70,76,89	0
5	NAG	I	503	14/15	0.89	0.11	-	35,44,51,51	0
3	NAG	K	603	14/15	0.79	0.19	-	46,62,67,67	0
5	GAL	I	502	11/12	0.95	0.08	-	38,43,48,54	0
7	NAG	G	432	14/15	0.80	0.23	-	48,55,66,75	0
3	GAL	A	331	12/12	0.84	0.32	-	63,71,82,83	0
5	NAG	E	331	14/15	0.89	0.15	-	47,51,56,60	0
3	NAG	C	332	14/15	0.93	0.13	-	39,49,55,57	0
5	GAL	E	332	11/12	0.92	0.10	-	41,50,54,62	0
3	GAL	K	604	12/12	0.84	0.23	-	67,76,84,88	0
3	NAG	A	332	14/15	0.85	0.21	-	49,60,65,65	0
3	GAL	C	331	11/12	0.95	0.13	-	39,44,53,59	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	A	336	14/15	0.79	0.30	2.72	62,69,71,76	0
4	NAG	A	335	14/15	0.88	0.13	0.56	34,46,56,58	0
4	NAG	E	333	14/15	0.89	0.12	-0.06	44,51,62,65	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	SIA	G	401	20/21	0.93	0.11	-0.10	29,38,45,46	0
4	NAG	J	561	14/15	0.87	0.15	-0.41	62,69,76,78	0
4	NAG	C	334	14/15	0.84	0.17	-	50,63,74,75	0
4	NAG	A	334	14/15	0.70	0.38	-	78,80,93,93	0
4	NAG	K	641	14/15	0.90	0.27	-	65,75,84,88	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.