



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 04:34 am GMT

PDB ID : 4JTV  
Title : Crystal structure of 2009 pandemic influenza virus hemagglutinin complexed with human receptor analogue LSTc  
Authors : Zhang, W.; Shi, Y.; Qi, J.; Gao, F.; Li, Q.; Fan, Z.; Yan, J.; Gao, G.F.  
Deposited on : 2013-03-24  
Resolution : 3.00 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 : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : trunk28620  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : recalc28949

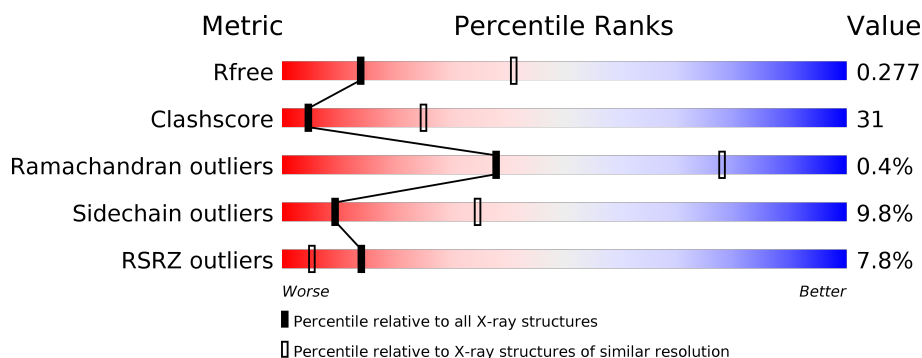
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.00 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))
$R_{free}$	100719	1692 (3.00-3.00)
Clashscore	112137	2037 (3.00-3.00)
Ramachandran outliers	110173	1973 (3.00-3.00)
Sidechain outliers	110143	1976 (3.00-3.00)
RSRZ outliers	101464	1716 (3.00-3.00)

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	321	<div> <div>3%</div> <div> <div></div> <div>48%</div> <div>45%</div> <div>7%</div> </div> </div>
1	C	321	<div> <div>6%</div> <div> <div></div> <div>52%</div> <div>40%</div> <div>7%</div> </div> </div>
1	E	321	<div> <div>2%</div> <div> <div></div> <div>56%</div> <div>38%</div> <div>5%</div> </div> </div>
1	G	321	<div> <div>5%</div> <div> <div></div> <div>60%</div> <div>36%</div> <div>• •</div> </div> </div>
1	I	321	<div> <div>3%</div> <div> <div></div> <div>52%</div> <div>40%</div> <div>7%</div> </div> </div>
1	K	321	<div> <div>7%</div> <div> <div></div> <div>55%</div> <div>40%</div> <div>• •</div> </div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	B	162	
2	D	162	
2	F	162	
2	H	162	
2	J	162	
2	L	162	

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	602	-	-	-	X

## 2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 23528 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	321	Total	C	N	O	S	0	0	0
			2505	1584	433	477	11			
1	C	321	Total	C	N	O	S	0	0	0
			2509	1586	433	479	11			
1	E	321	Total	C	N	O	S	0	0	0
			2509	1586	433	479	11			
1	G	321	Total	C	N	O	S	0	0	0
			2505	1584	433	477	11			
1	I	321	Total	C	N	O	S	0	0	0
			2510	1587	433	479	11			
1	K	321	Total	C	N	O	S	0	0	0
			2509	1586	433	479	11			

- Molecule 2 is a protein called Hemagglutinin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	162	Total	C	N	O	S	0	0	0
			1305	822	220	257	6			
2	D	162	Total	C	N	O	S	0	0	0
			1300	818	219	257	6			
2	F	161	Total	C	N	O	S	0	0	0
			1302	821	219	256	6			
2	H	162	Total	C	N	O	S	0	0	0
			1305	822	220	257	6			
2	J	162	Total	C	N	O	S	0	0	0
			1300	818	219	257	6			
2	L	161	Total	C	N	O	S	0	0	0
			1302	821	219	256	6			

- 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	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	C	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 (3-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	3	Total	C	N	O	0	0
			45	25	2	18		
4	I	3	Total	C	N	O	0	0
			45	25	2	18		
4	K	3	Total	C	N	O	0	0
			45	25	2	18		

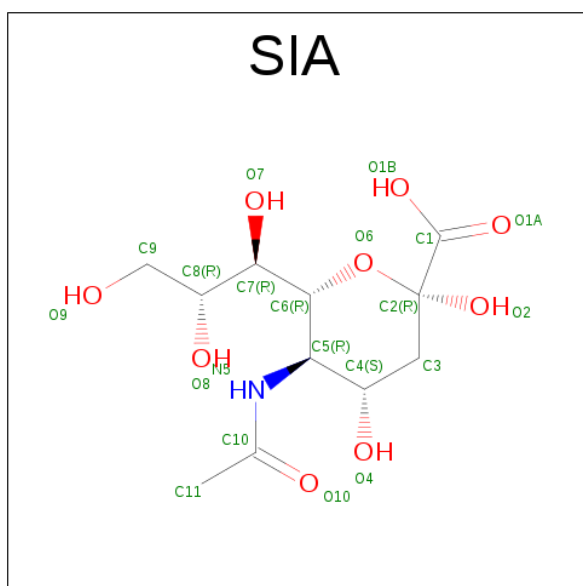
- Molecule 5 is a polymer of unknown type called SUGAR (4-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
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 a polymer of unknown type called SUGAR (2-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
6	E	2	Total	C	N	O	0	0
			28	16	2	10		

- Molecule 7 is SUGAR (O-SIALIC ACID) (three-letter code: SIA) (formula:  $C_{11}H_{19}NO_9$ ).



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

- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	40	Total	O	0	0
			40	40		
8	B	13	Total	O	0	0
			13	13		
8	C	29	Total	O	0	0
			29	29		

*Continued on next page...*

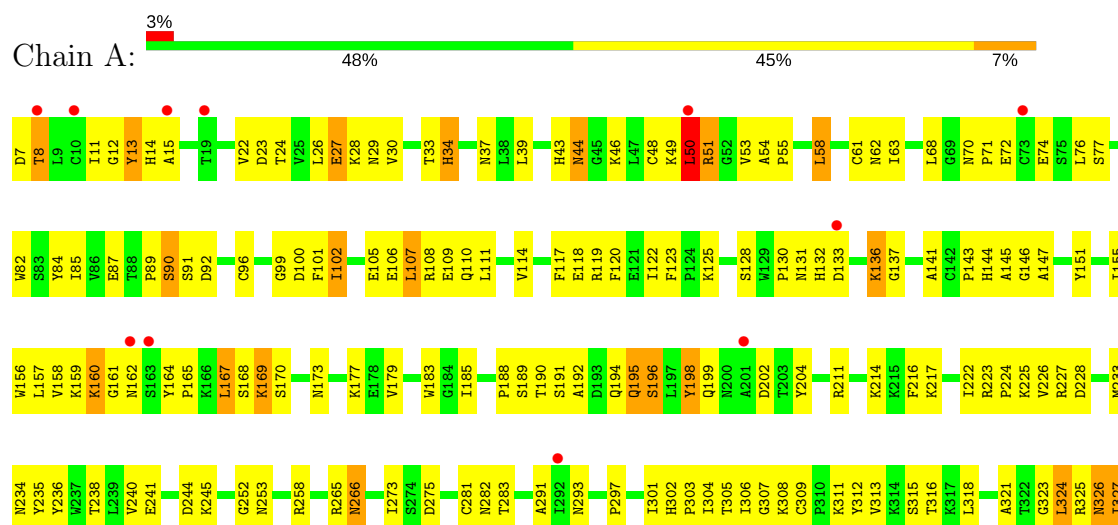
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	D	13	Total 13	O 13	0	0
8	E	31	Total 31	O 31	0	0
8	F	7	Total 7	O 7	0	0
8	G	19	Total 19	O 19	0	0
8	H	13	Total 13	O 13	0	0
8	I	35	Total 35	O 35	0	0
8	J	18	Total 18	O 18	0	0
8	K	25	Total 25	O 25	0	0
8	L	31	Total 31	O 31	0	0

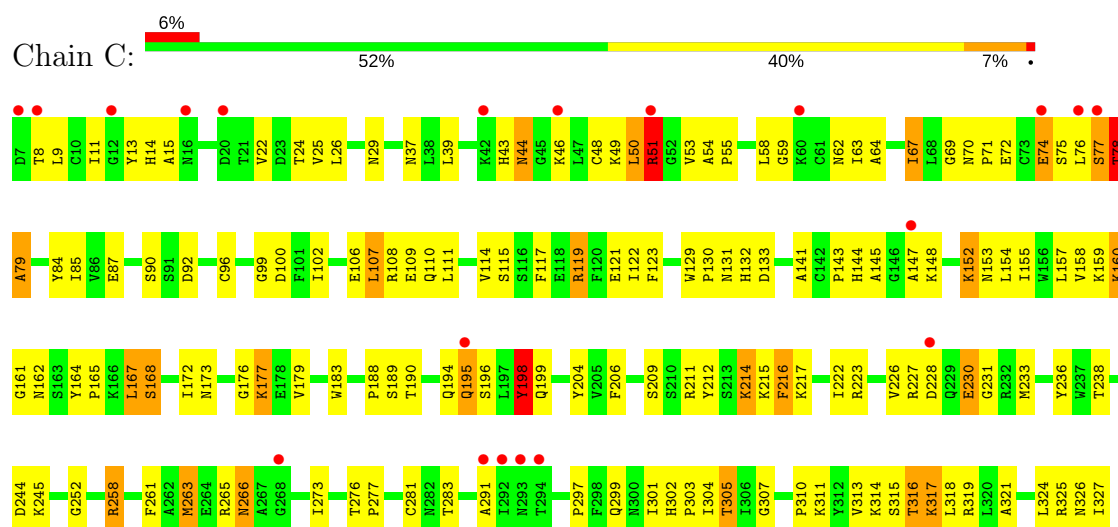
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

#### • Molecule 1: Hemagglutinin



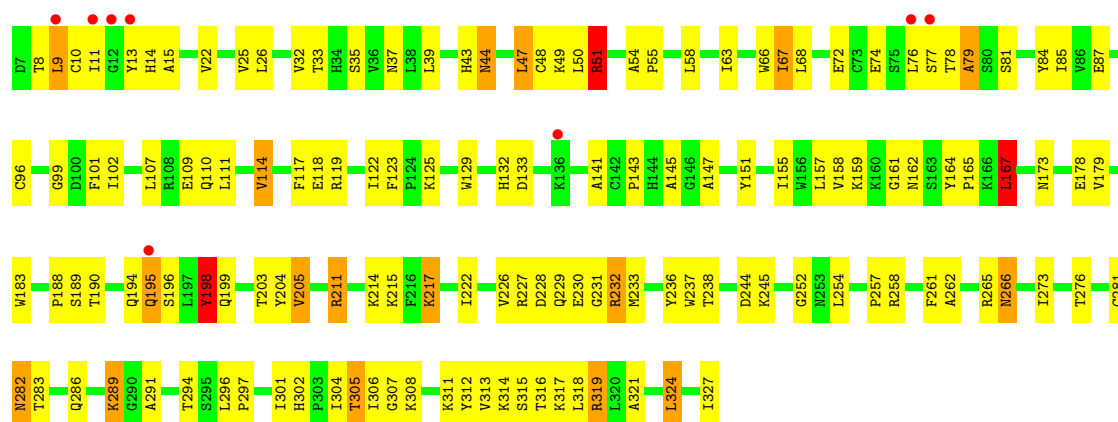
#### • Molecule 1: Hemagglutinin



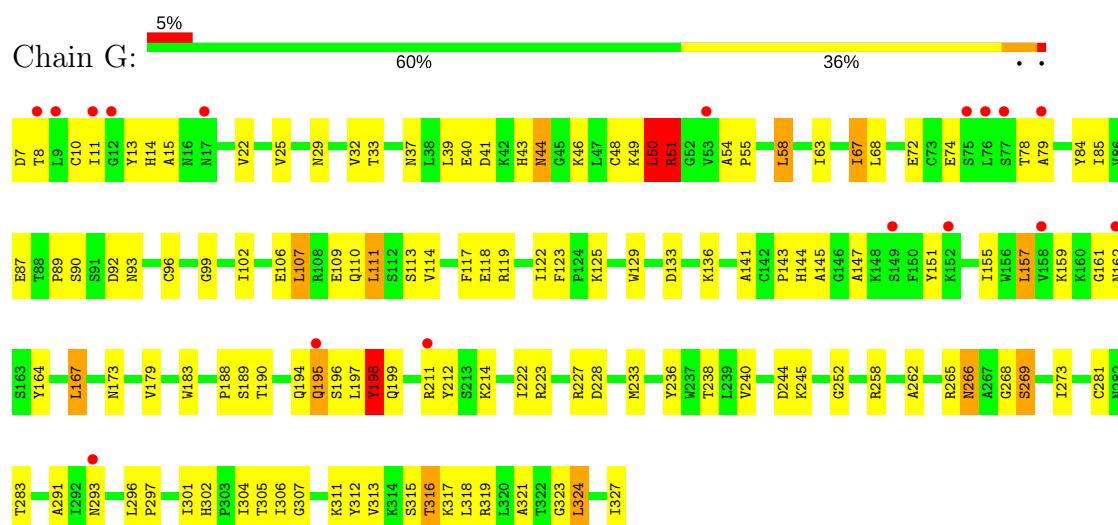
#### • Molecule 1: Hemagglutinin



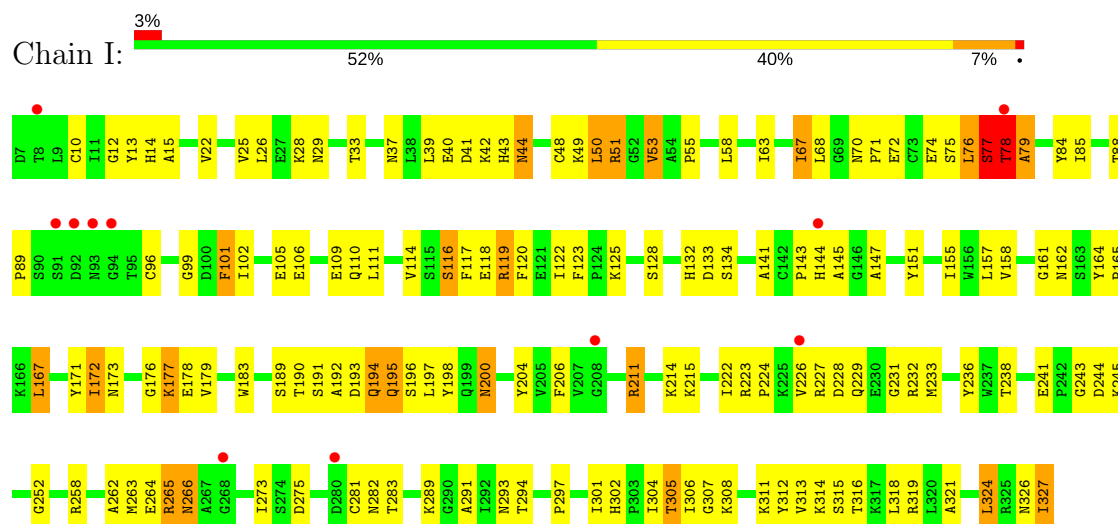




• Molecule 1: Hemagglutinin

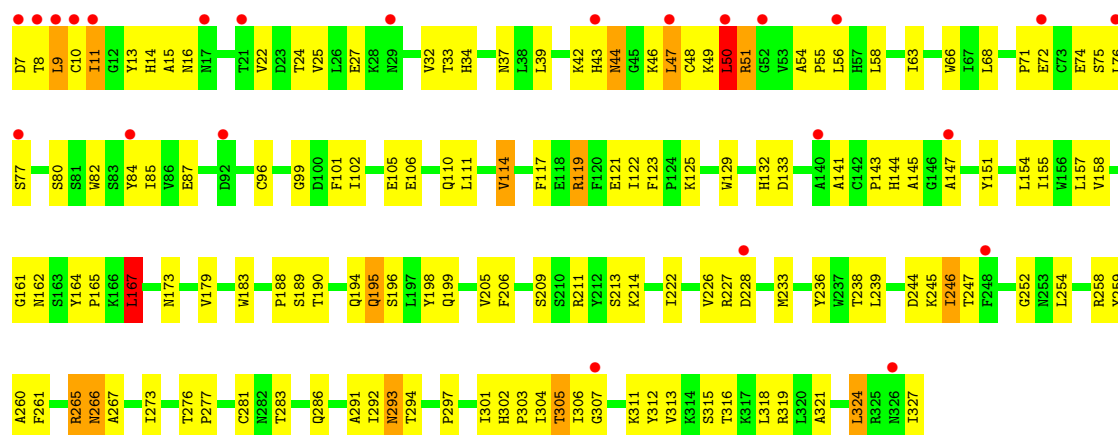


• Molecule 1: Hemagglutinin

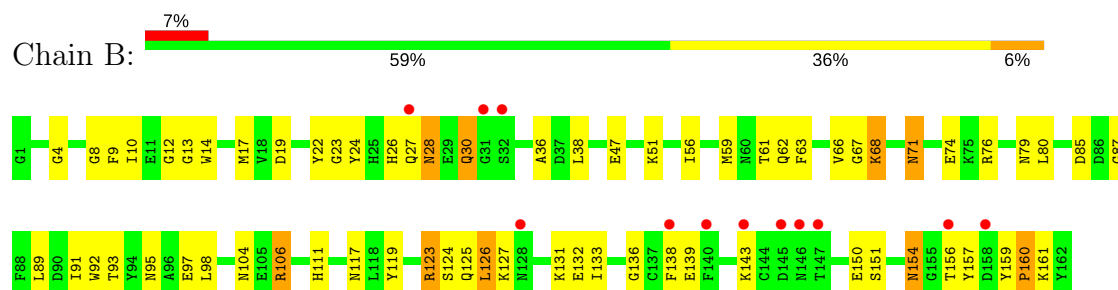


• Molecule 1: Hemagglutinin

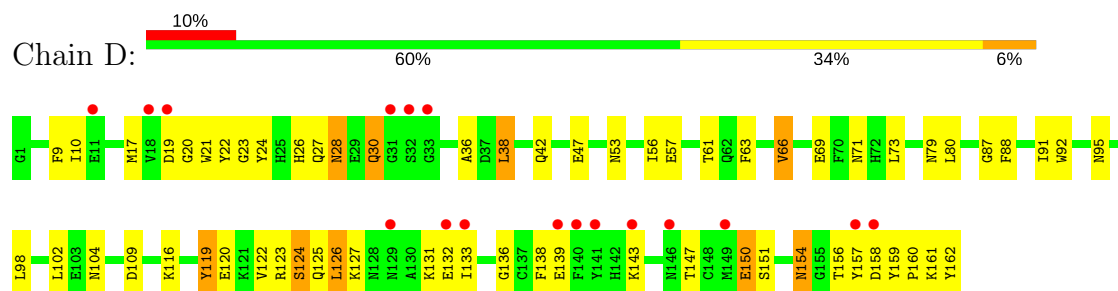




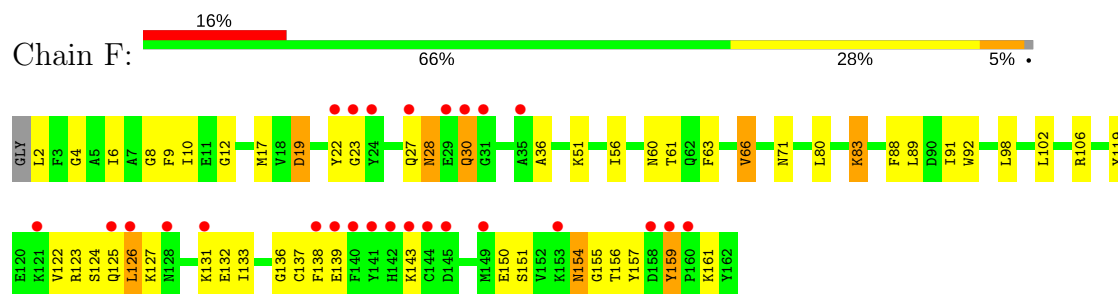
### • Molecule 2: Hemagglutinin



### • Molecule 2: Hemagglutinin

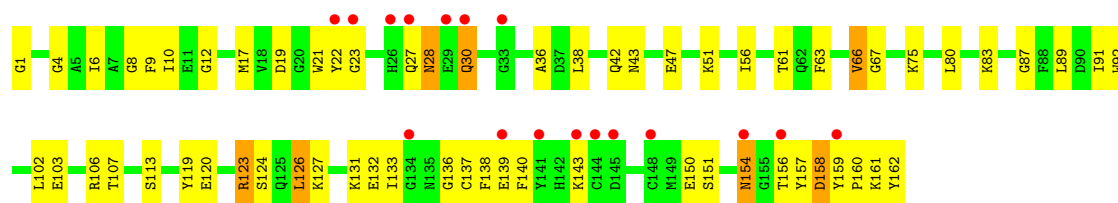


### • Molecule 2: Hemagglutinin

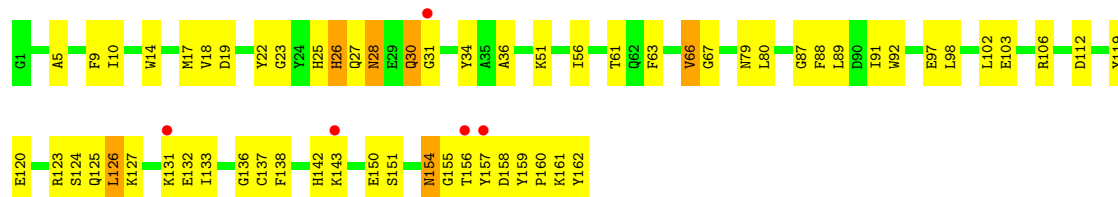


### • Molecule 2: Hemagglutinin

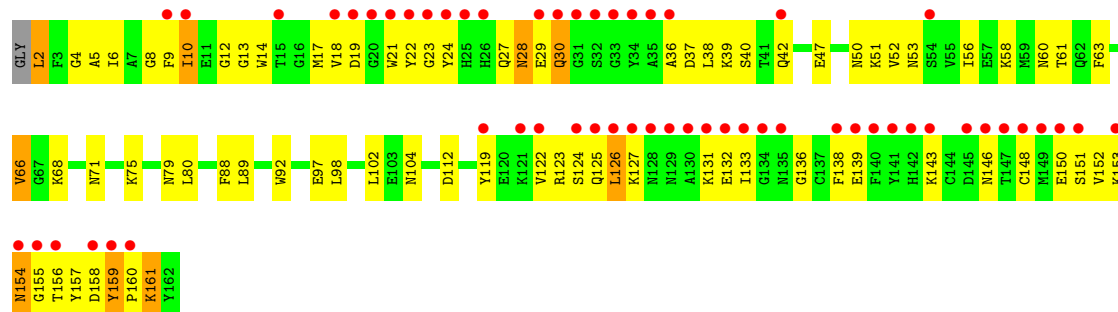




### • Molecule 2: Hemagglutinin



### • Molecule 2: Hemagglutinin



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	66.72Å 117.32Å 117.39Å 61.78° 81.82° 77.42°	Depositor
Resolution (Å)	38.39 – 3.00 49.68 – 3.00	Depositor EDS
% Data completeness (in resolution range)	81.5 (38.39-3.00) 80.6 (49.68-3.00)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.37 (at 3.01Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.5_2)	Depositor
R, $R_{free}$	0.224 , 0.275 0.227 , 0.277	Depositor DCC
$R_{free}$ test set	2707 reflections (5.04%)	DCC
Wilson B-factor (Å <sup>2</sup> )	55.4	Xtriage
Anisotropy	0.701	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 58.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	23528	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	84.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.66% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality [i](#)

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

Bond lengths and bond angles in the following residue types are not validated in this section: 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	0.57	0/2568	0.75	5/3488 (0.1%)
1	C	0.55	0/2572	0.69	8/3493 (0.2%)
1	E	0.39	0/2572	0.58	4/3493 (0.1%)
1	G	0.33	0/2568	0.77	9/3488 (0.3%)
1	I	0.47	0/2573	0.63	3/3495 (0.1%)
1	K	0.33	0/2572	0.63	5/3493 (0.1%)
2	B	0.48	0/1333	0.54	1/1797 (0.1%)
2	D	0.38	0/1328	0.45	0/1791
2	F	0.26	0/1330	0.46	0/1794
2	H	0.26	0/1333	0.49	1/1797 (0.1%)
2	J	0.43	0/1328	0.48	0/1791
2	L	0.26	0/1330	0.53	1/1794 (0.1%)
All	All	0.42	0/23407	0.62	37/31714 (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
1	I	0	1

There are no bond length outliers.

The worst 5 of 37 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	G	119	ARG	NE-CZ-NH1	-16.56	112.02	120.30
1	A	119	ARG	NE-CZ-NH2	16.50	128.55	120.30
1	A	119	ARG	NE-CZ-NH1	-16.32	112.14	120.30
1	G	119	ARG	NE-CZ-NH2	15.97	128.29	120.30

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	G	212	TYR	N-CA-CB	11.91	132.04	110.60

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	I	77	SER	Peptide

## 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	2505	0	2444	227	0
1	C	2509	0	2451	186	0
1	E	2509	0	2451	159	4
1	G	2505	0	2445	137	0
1	I	2510	0	2456	183	0
1	K	2509	0	2452	188	0
2	B	1305	0	1228	97	0
2	D	1300	0	1216	81	0
2	F	1302	0	1226	65	0
2	H	1305	0	1228	73	0
2	J	1300	0	1216	89	0
2	L	1302	0	1226	126	4
3	A	56	0	52	6	0
3	C	14	0	13	2	0
3	G	28	0	26	5	0
4	A	45	0	38	4	0
4	I	45	0	38	3	0
4	K	45	0	38	0	0
5	C	56	0	47	1	0
5	E	56	0	47	2	0
6	E	28	0	25	3	0
7	G	20	0	17	1	0
8	A	40	0	0	59	0
8	B	13	0	0	11	0
8	C	29	0	0	34	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
8	D	13	0	0	6	0
8	E	31	0	0	27	0
8	F	7	0	0	6	0
8	G	19	0	0	22	0
8	H	13	0	0	22	0
8	I	35	0	0	32	0
8	J	18	0	0	22	0
8	K	25	0	0	47	0
8	L	31	0	0	48	0
All	All	23528	0	22380	1414	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 31.

The worst 5 of 1414 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:C:11:ILE:HD13	2:D:24:TYR:CD2	1.53	1.42
1:I:190:THR:HG21	1:I:193:ASP:OD1	1.24	1.35
1:K:106:GLU:HB3	8:K:915:HOH:O	1.20	1.34
2:L:19:ASP:HA	8:L:228:HOH:O	1.33	1.28
1:A:8:THR:HG22	2:B:138:PHE:O	1.18	1.27

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)
1:E:119:ARG:NH2	2:L:38:LEU:CD2[1_565]	1.96	0.24
1:E:79:ALA:N	2:L:150:GLU:OE2[1_565]	2.01	0.19
1:E:258:ARG:NH2	2:L:42:GLN:OE1[1_565]	2.04	0.16
1:E:81:SER:OG	2:L:146:ASN:OD1[1_565]	2.12	0.08

## 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	319/321 (99%)	295 (92%)	23 (7%)	1 (0%)	44	81
1	C	319/321 (99%)	296 (93%)	19 (6%)	4 (1%)	14	51
1	E	319/321 (99%)	295 (92%)	23 (7%)	1 (0%)	44	81
1	G	319/321 (99%)	292 (92%)	27 (8%)	0	100	100
1	I	319/321 (99%)	291 (91%)	26 (8%)	2 (1%)	28	70
1	K	319/321 (99%)	290 (91%)	29 (9%)	0	100	100
2	B	160/162 (99%)	146 (91%)	13 (8%)	1 (1%)	28	70
2	D	160/162 (99%)	143 (89%)	16 (10%)	1 (1%)	28	70
2	F	159/162 (98%)	145 (91%)	14 (9%)	0	100	100
2	H	160/162 (99%)	146 (91%)	14 (9%)	0	100	100
2	J	160/162 (99%)	144 (90%)	16 (10%)	0	100	100
2	L	159/162 (98%)	145 (91%)	13 (8%)	1 (1%)	28	70
All	All	2872/2898 (99%)	2628 (92%)	233 (8%)	11 (0%)	38	78

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	160	PRO
1	A	146	GLY
1	C	77	SER
1	C	78	THR
1	C	161	GLY

### 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	281/282 (100%)	248 (88%)	33 (12%)	6	25
1	C	282/282 (100%)	248 (88%)	34 (12%)	6	24

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	E	282/282 (100%)	254 (90%)	28 (10%)	9	34
1	G	281/282 (100%)	258 (92%)	23 (8%)	13	44
1	I	282/282 (100%)	247 (88%)	35 (12%)	5	23
1	K	282/282 (100%)	255 (90%)	27 (10%)	10	36
2	B	139/139 (100%)	128 (92%)	11 (8%)	14	46
2	D	138/139 (99%)	125 (91%)	13 (9%)	10	37
2	F	139/139 (100%)	126 (91%)	13 (9%)	10	37
2	H	139/139 (100%)	130 (94%)	9 (6%)	20	56
2	J	138/139 (99%)	128 (93%)	10 (7%)	17	51
2	L	139/139 (100%)	128 (92%)	11 (8%)	14	46
All	All	2522/2526 (100%)	2275 (90%)	247 (10%)	9	34

5 of 247 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	232	ARG
1	G	107	LEU
1	K	228	ASP
1	E	305	THR
2	F	80	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 91 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	282	ASN
1	G	162	ASN
1	K	293	ASN
2	F	27	GLN
2	F	146	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 ⓘ

19 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	SIA	A	605	4	17,20,21	0.50	0	19,28,31	2.45	2 (10%)
4	GAL	A	606	4	11,11,12	0.63	0	13,15,17	0.71	0
4	NAG	A	607	4	14,14,15	0.60	0	15,19,21	0.97	0
5	SIA	C	602	5	17,20,21	0.36	0	19,28,31	2.61	1 (5%)
5	GAL	C	603	5	11,11,12	0.58	0	13,15,17	0.68	0
5	NAG	C	604	5	14,14,15	0.57	0	15,19,21	1.34	2 (13%)
5	GAL	C	605	5	11,11,12	0.70	0	13,15,17	1.04	1 (7%)
6	NAG	E	601	1,6	14,14,15	0.56	0	15,19,21	0.78	0
6	NAG	E	602	6	14,14,15	0.45	0	15,19,21	0.75	0
5	SIA	E	603	5	17,20,21	0.47	0	19,28,31	2.28	2 (10%)
5	GAL	E	604	5	11,11,12	0.68	0	13,15,17	0.76	0
5	NAG	E	605	5	14,14,15	0.61	0	15,19,21	0.78	0
5	GAL	E	606	5	11,11,12	0.63	0	13,15,17	0.52	0
4	SIA	I	801	4	17,20,21	0.40	0	19,28,31	2.38	2 (10%)
4	GAL	I	802	4	11,11,12	0.72	0	13,15,17	1.12	1 (7%)
4	NAG	I	803	4	14,14,15	0.49	0	15,19,21	0.64	0
4	SIA	K	801	4	17,20,21	0.39	0	19,28,31	2.40	3 (15%)
4	GAL	K	802	4	11,11,12	0.61	0	13,15,17	0.86	0
4	NAG	K	803	4	14,14,15	0.52	0	15,19,21	0.73	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	SIA	A	605	4	-	0/14/34/38	0/1/1/1
4	GAL	A	606	4	-	0/2/19/22	0/1/1/1
4	NAG	A	607	4	-	0/6/23/26	0/1/1/1
5	SIA	C	602	5	-	0/14/34/38	0/1/1/1
5	GAL	C	603	5	-	0/2/19/22	0/1/1/1
5	NAG	C	604	5	-	0/6/23/26	0/1/1/1
5	GAL	C	605	5	-	0/2/19/22	0/1/1/1
6	NAG	E	601	1,6	-	0/6/23/26	0/1/1/1
6	NAG	E	602	6	-	0/6/23/26	0/1/1/1
5	SIA	E	603	5	-	0/14/34/38	0/1/1/1
5	GAL	E	604	5	-	0/2/19/22	0/1/1/1
5	NAG	E	605	5	-	0/6/23/26	0/1/1/1
5	GAL	E	606	5	-	0/2/19/22	0/1/1/1
4	SIA	I	801	4	-	0/14/34/38	0/1/1/1
4	GAL	I	802	4	-	0/2/19/22	0/1/1/1
4	NAG	I	803	4	-	0/6/23/26	0/1/1/1
4	SIA	K	801	4	-	0/14/34/38	0/1/1/1
4	GAL	K	802	4	-	0/2/19/22	0/1/1/1
4	NAG	K	803	4	-	0/6/23/26	0/1/1/1

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	C	602	SIA	O6-C2-C3	-10.96	90.09	109.82
4	A	605	SIA	O6-C2-C3	-9.94	91.93	109.82
4	K	801	SIA	O6-C2-C3	-9.69	92.37	109.82
4	I	801	SIA	O6-C2-C3	-9.68	92.39	109.82
5	E	603	SIA	O6-C2-C3	-9.28	93.11	109.82

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

9 monomers are involved in 13 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	605	SIA	1	0
4	A	606	GAL	1	0
4	A	607	NAG	2	0
5	C	602	SIA	1	0
6	E	601	NAG	3	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	E	602	NAG	1	0
5	E	603	SIA	2	0
4	I	801	SIA	3	0
4	I	803	NAG	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	$\# Z  > 2$
3	NAG	A	601	1	14,14,15	0.51	0	15,19,21	0.85	0
3	NAG	A	602	1	14,14,15	0.55	0	15,19,21	0.62	0
3	NAG	A	603	1	14,14,15	0.53	0	15,19,21	0.52	0
3	NAG	A	604	1	14,14,15	0.56	0	15,19,21	0.62	0
3	NAG	C	601	1	14,14,15	0.46	0	15,19,21	1.08	1 (6%)
3	NAG	G	601	1	14,14,15	0.28	0	15,19,21	0.57	0
3	NAG	G	602	1	14,14,15	0.76	0	15,19,21	1.69	4 (26%)
7	SIA	G	603	-	17,20,21	0.39	0	19,28,31	1.10	1 (5%)

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	601	1	-	0/6/23/26	0/1/1/1
3	NAG	A	602	1	-	0/6/23/26	0/1/1/1
3	NAG	A	603	1	-	0/6/23/26	0/1/1/1
3	NAG	A	604	1	-	0/6/23/26	0/1/1/1
3	NAG	C	601	1	-	0/6/23/26	0/1/1/1
3	NAG	G	601	1	-	0/6/23/26	0/1/1/1
3	NAG	G	602	1	-	0/6/23/26	0/1/1/1
7	SIA	G	603	-	-	0/14/34/38	0/1/1/1

There are no bond length outliers.

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	G	602	NAG	O5-C1-C2	-2.33	108.22	111.47
3	G	602	NAG	C4-C3-C2	2.19	114.22	111.02
3	G	602	NAG	C1-C2-N2	2.42	114.62	110.49
3	C	601	NAG	C1-O5-C5	3.14	116.49	112.17
7	G	603	SIA	O6-C2-C3	3.42	115.98	109.82

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

6 monomers are involved in 14 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	602	NAG	4	0
3	A	603	NAG	2	0
3	C	601	NAG	2	0
3	G	601	NAG	1	0
3	G	602	NAG	4	0
7	G	603	SIA	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	321/321 (100%)	0.16	11 (3%) 46 20	35, 69, 111, 154	0
1	C	321/321 (100%)	0.26	20 (6%) 21 8	36, 68, 118, 223	0
1	E	321/321 (100%)	0.16	8 (2%) 58 29	41, 71, 121, 203	0
1	G	321/321 (100%)	0.31	17 (5%) 27 11	39, 76, 124, 154	0
1	I	321/321 (100%)	0.27	11 (3%) 46 20	39, 75, 117, 169	0
1	K	321/321 (100%)	0.53	24 (7%) 15 6	52, 91, 136, 187	0
2	B	162/162 (100%)	0.36	12 (7%) 15 6	35, 84, 142, 180	0
2	D	162/162 (100%)	0.54	17 (10%) 7 3	39, 86, 163, 220	0
2	F	161/162 (99%)	0.73	26 (16%) 2 1	43, 86, 160, 246	0
2	H	162/162 (100%)	0.44	17 (10%) 7 3	40, 85, 149, 181	0
2	J	162/162 (100%)	0.30	5 (3%) 49 22	47, 80, 129, 156	0
2	L	161/162 (99%)	1.83	57 (35%) 0 0	43, 101, 223, 362	0
All	All	2896/2898 (99%)	0.42	225 (7%) 14 5	35, 78, 142, 362	0

The worst 5 of 225 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	L	129	ASN	12.2
2	L	146	ASN	9.6
2	L	130	ALA	9.2
2	L	149	MET	9.2
2	L	147	THR	8.5

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

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

### 6.3 Carbohydrates [i](#)

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

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
5	SIA	E	603	20/21	0.93	0.23	0.09	42,65,87,99	0
4	SIA	K	801	20/21	0.94	0.22	-0.40	64,83,91,91	0
4	SIA	A	605	20/21	0.94	0.18	-0.54	37,53,76,80	0
5	SIA	C	602	20/21	0.94	0.19	-0.69	48,71,90,97	0
6	NAG	E	601	14/15	0.90	0.16	-0.71	74,82,98,102	0
4	SIA	I	801	20/21	0.94	0.19	-0.72	61,77,95,96	0
4	GAL	I	802	11/12	0.88	0.22	-0.90	61,95,108,114	0
5	NAG	E	605	14/15	0.86	0.34	-	120,126,146,147	0
6	NAG	E	602	14/15	0.76	0.36	-	88,113,132,144	0
5	GAL	C	603	11/12	0.93	0.17	-	56,84,92,100	0
5	GAL	C	605	11/12	0.60	0.25	-	104,129,151,153	0
4	GAL	A	606	11/12	0.94	0.10	-	53,93,104,109	0
5	NAG	C	604	14/15	0.83	0.25	-	75,106,127,128	0
4	NAG	I	803	14/15	0.89	0.27	-	111,129,140,142	0
5	GAL	E	606	11/12	0.77	0.26	-	110,139,162,174	0
4	NAG	K	803	14/15	0.80	0.23	-	110,146,159,172	0
4	GAL	K	802	11/12	0.79	0.22	-	96,131,149,169	0
5	GAL	E	604	11/12	0.92	0.14	-	78,95,110,122	0
4	NAG	A	607	14/15	0.85	0.23	-	65,80,89,96	0

### 6.4 Ligands [i](#)

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

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
3	NAG	G	602	14/15	0.82	0.28	2.09	100,121,147,149	0
7	SIA	G	603	20/21	0.89	0.28	0.54	65,86,119,122	0
3	NAG	A	603	14/15	0.88	0.28	0.52	70,102,144,146	0
3	NAG	C	601	14/15	0.79	0.23	0.39	70,86,97,103	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
3	NAG	A	604	14/15	0.71	0.26	-	99,145,158,161	0
3	NAG	A	602	14/15	0.70	0.31	-	146,167,182,187	0
3	NAG	G	601	14/15	0.55	0.48	-	119,147,162,173	0
3	NAG	A	601	14/15	0.68	0.34	-	142,168,177,178	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.