



# wwPDB X-ray Structure Validation Summary Report i

Feb 28, 2014 – 09:50 AM 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 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/ValidationPDFNotes.html>

---

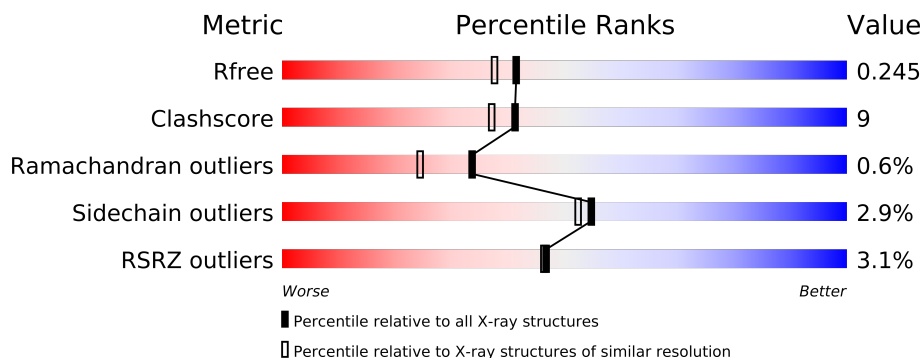
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22683

# 1 Overall quality at a glance

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))
$R_{free}$	66092	4888 (2.00-2.00)
Clashscore	79885	6188 (2.00-2.00)
Ramachandran outliers	78287	6102 (2.00-2.00)
Sidechain outliers	78261	6100 (2.00-2.00)
RSRZ outliers	66119	4890 (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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	329	
1	C	329	
1	E	329	
1	G	329	
1	I	329	
1	K	329	
2	B	177	
2	D	177	
2	F	177	
2	H	177	
2	J	177	
2	L	177	

## 2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 24675 atoms, of which 0 are hydrogen and 0 are deuterium.

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		

There are 12 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
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
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 4 is SUGAR (N-ACETYL-D-GLUCOSAMINE) (three-letter code: NAG) (formula: C<sub>8</sub>H<sub>15</sub>NO<sub>6</sub>).



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

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
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

- 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		

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
I	9	PRO	-	EXPRESSION TAG	UNP C3W5S1
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 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

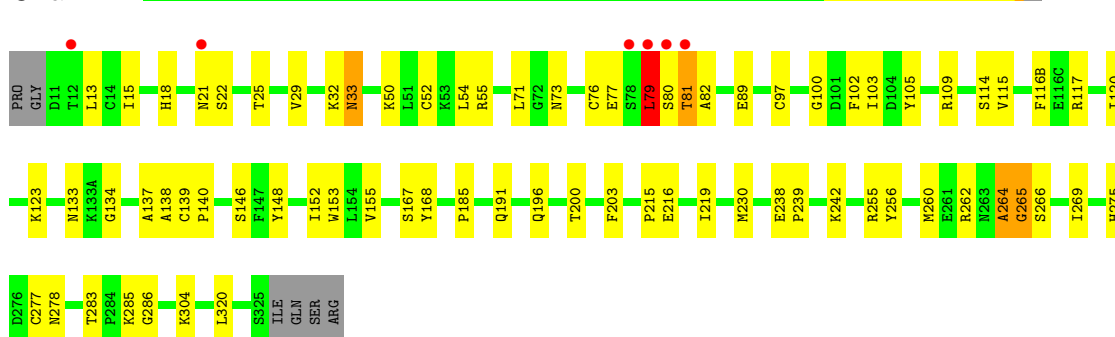


### 3 Residue-property plots

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.

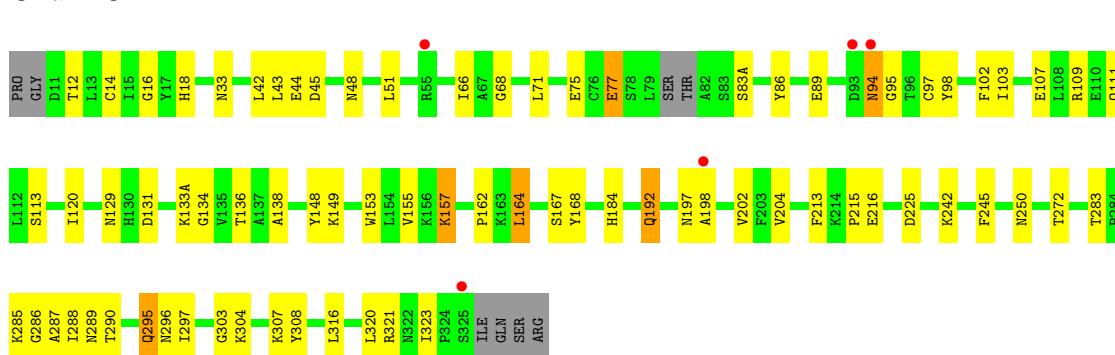
#### • Molecule 1: hemagglutinin HA1

Chain A:



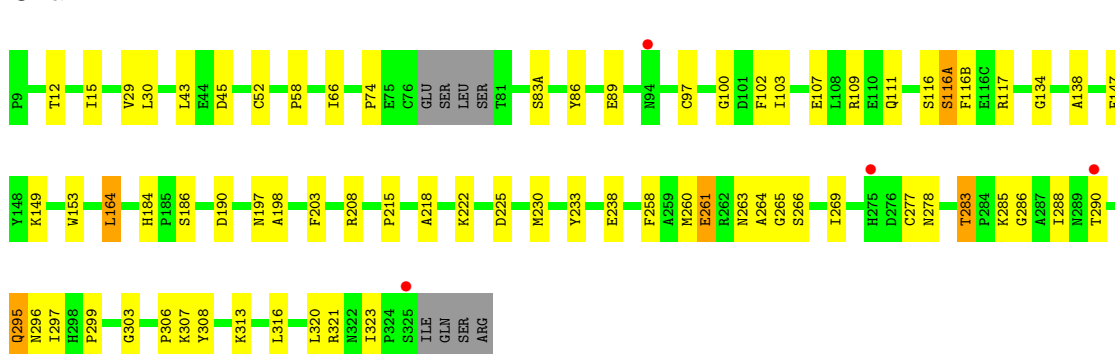
#### • Molecule 1: hemagglutinin HA1

Chain C:



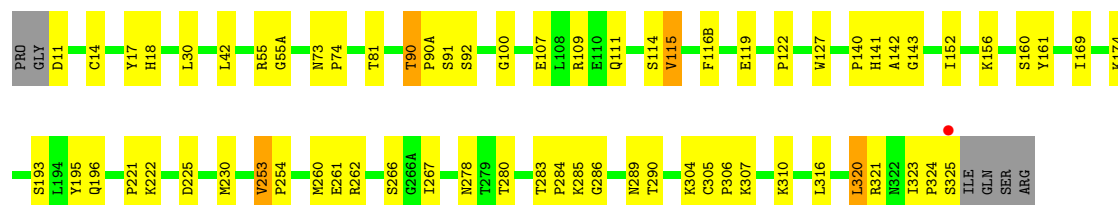
#### • Molecule 1: hemagglutinin HA1

Chain E:



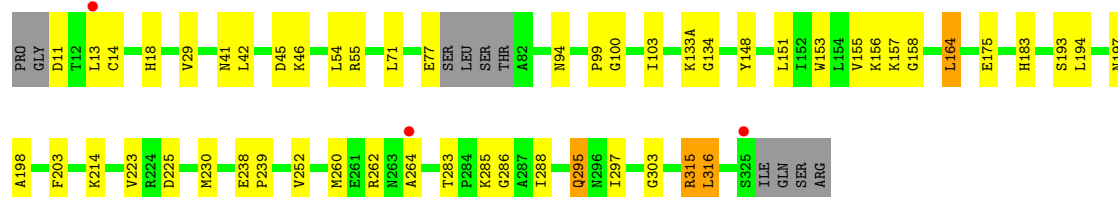
- Molecule 1: hemagglutinin HA1

Chain G:



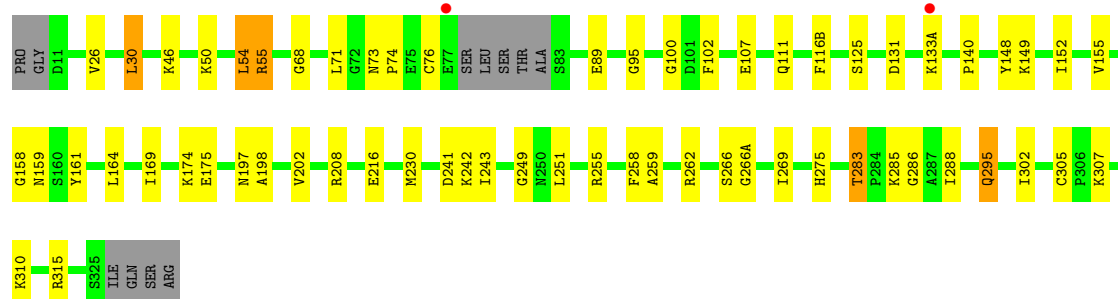
- Molecule 1: hemagglutinin HA1

Chain I:



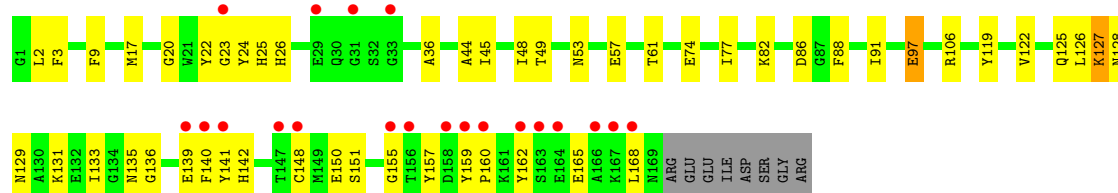
- Molecule 1: hemagglutinin HA1

Chain K:



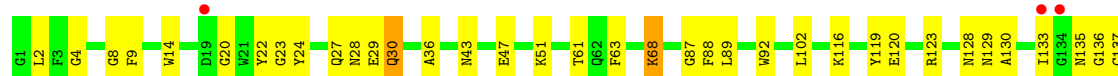
- Molecule 2: hemagglutinin HA2

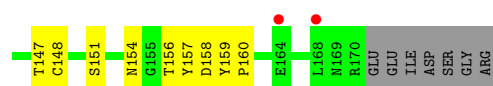
Chain B:



- Molecule 2: hemagglutinin HA2

Chain D:





• Molecule 2: hemagglutinin HA2

Chain F:



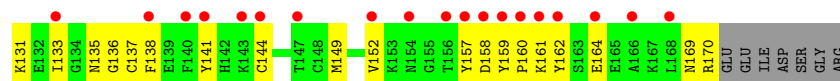
• Molecule 2: hemagglutinin HA2

Chain H:



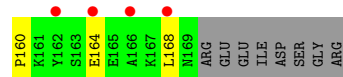
• Molecule 2: hemagglutinin HA2

Chain J:



• Molecule 2: hemagglutinin HA2

Chain L:



## 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) 84.2 (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 0.197 , 0.245	Depositor DCC
$R_{free}$ test set	8642 reflections (5.01%)	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 , 42.1	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	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.

## 5 Model quality

### 5.1 Standard geometry

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 Close contacts

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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2522	0	2466	61	0
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

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 9.

The worst 5 of 426 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(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

### 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	321/329 (98%)	303 (94%)	13 (4%)	5 (2%)	14	5
1	C	317/329 (96%)	301 (95%)	15 (5%)	1 (0%)	50	44
1	E	317/329 (96%)	306 (96%)	10 (3%)	1 (0%)	50	44
1	G	321/329 (98%)	305 (95%)	16 (5%)	0	100	100
1	I	315/329 (96%)	300 (95%)	12 (4%)	3 (1%)	22	12
1	K	314/329 (95%)	295 (94%)	17 (5%)	2 (1%)	33	24
2	B	167/177 (94%)	161 (96%)	5 (3%)	1 (1%)	33	24
2	D	168/177 (95%)	160 (95%)	6 (4%)	2 (1%)	19	9
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%)	19	9
2	L	167/177 (94%)	157 (94%)	10 (6%)	0	100	100
All	All	2913/3036 (96%)	2768 (95%)	128 (4%)	17 (1%)	33	24

5 of 17 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	81	THR
1	A	265	GLY
2	B	127	LYS

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
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%)	71	73
1	C	283/290 (98%)	272 (96%)	11 (4%)	43	38
1	E	282/290 (97%)	273 (97%)	9 (3%)	51	47
1	G	285/290 (98%)	277 (97%)	8 (3%)	56	54
1	I	281/290 (97%)	274 (98%)	7 (2%)	60	59
1	K	281/290 (97%)	274 (98%)	7 (2%)	60	59
2	B	145/152 (95%)	139 (96%)	6 (4%)	41	35
2	D	146/152 (96%)	141 (97%)	5 (3%)	49	45
2	F	146/152 (96%)	144 (99%)	2 (1%)	78	81
2	H	148/152 (97%)	143 (97%)	5 (3%)	49	45
2	J	146/152 (96%)	140 (96%)	6 (4%)	41	35
2	L	145/152 (95%)	141 (97%)	4 (3%)	56	54
All	All	2573/2652 (97%)	2498 (97%)	75 (3%)	55	52

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 chains 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	12,14,15	0.70	1 (8%)	15,19,21	1.09	1 (6%)
3	NAG	A	331	3	12,14,15	0.69	0	15,19,21	0.86	0
5	SIA	C	330	5	20,20,21	3.79	1 (5%)	23,28,31	2.46	3 (13%)
5	GAL	C	331	5	10,11,12	0.87	1 (10%)	11,15,17	1.44	2 (18%)
5	NAG	C	332	5	15,15,15	0.49	0	21,21,21	0.77	1 (4%)
3	NAG	E	331	1,3	12,14,15	0.69	1 (8%)	15,19,21	0.86	0
3	NAG	E	332	3	12,14,15	0.66	0	15,19,21	0.98	1 (6%)
6	SIA	I	501	6	20,20,21	3.84	1 (5%)	23,28,31	2.49	4 (17%)
6	GAL	I	502	6	10,11,12	0.80	0	11,15,17	1.15	1 (9%)
6	SIA	K	601	6	20,20,21	3.84	1 (5%)	23,28,31	2.60	5 (21%)
6	GAL	K	602	6	10,11,12	0.78	0	11,15,17	0.82	1 (9%)
3	NAG	K	631	1,3	12,14,15	0.66	0	15,19,21	0.91	1 (6%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
3	NAG	K	632	3	12,14,15	0.66	0	15,19,21	0.83	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	-	0/15/34/38	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	-	0/15/34/38	0/1/1/1
6	GAL	I	502	6	-	0/2/19/22	0/1/1/1
6	SIA	K	601	6	-	0/15/34/38	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

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
6	K	601	SIA	O6-C2	16.91	1.43	1.28
6	I	501	SIA	O6-C2	16.91	1.43	1.28
5	C	330	SIA	O6-C2	16.72	1.43	1.28
3	E	331	NAG	O5-C5	-2.12	1.41	1.45
5	C	331	GAL	O5-C5	-2.07	1.41	1.45

The worst 5 of 20 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
6	K	601	SIA	O6-C2-C3	-11.15	110.34	124.91
6	I	501	SIA	O6-C2-C3	-10.77	110.84	124.91
5	C	330	SIA	O6-C2-C3	-10.66	110.98	124.91
5	C	331	GAL	O3-C3-C2	-3.54	103.47	109.94
6	I	502	GAL	O5-C5-C6	2.99	110.12	106.98

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 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	12,14,15	0.67	0	15,19,21	0.82	0
4	NAG	D	261	2	12,14,15	0.63	0	15,19,21	0.96	1 (6%)
4	NAG	E	341	1	12,14,15	0.63	0	15,19,21	1.03	0
4	NAG	G	411	1	12,14,15	0.63	0	15,19,21	0.90	1 (6%)
4	NAG	G	431	1	12,14,15	0.69	0	15,19,21	0.94	0
4	NAG	G	451	1	12,14,15	0.64	0	15,19,21	0.68	0
4	NAG	I	521	1	12,14,15	0.61	0	15,19,21	1.14	1 (6%)
4	NAG	I	531	1	12,14,15	0.53	0	15,19,21	0.91	1 (6%)
4	NAG	I	541	1	12,14,15	0.61	0	15,19,21	1.03	1 (6%)
4	NAG	K	621	1	12,14,15	0.59	0	15,19,21	1.05	1 (6%)
4	NAG	K	641	1	12,14,15	0.63	0	15,19,21	0.88	1 (6%)
4	NAG	K	651	1	12,14,15	0.62	0	15,19,21	0.90	1 (6%)

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
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

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
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 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	I	521	NAG	O5-C5-C6	3.76	110.93	106.98
4	K	621	NAG	O5-C5-C6	3.21	110.34	106.98
4	D	261	NAG	O5-C5-C6	2.70	109.81	106.98
4	G	411	NAG	O5-C5-C6	2.61	109.72	106.98
4	I	541	NAG	O5-C5-C6	2.36	109.46	106.98

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 5.7 Other polymers ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues

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%)	-0.08	6 (1%) 64 64	22, 35, 56, 75	0
1	C	321/329 (97%)	-0.12	5 (1%) 68 69	27, 42, 58, 76	0
1	E	321/329 (97%)	-0.17	4 (1%) 75 76	24, 35, 51, 69	0
1	G	323/329 (98%)	-0.20	1 (0%) 91 93	24, 36, 50, 70	0
1	I	319/329 (96%)	-0.25	3 (0%) 81 82	24, 36, 54, 76	0
1	K	318/329 (96%)	-0.12	2 (0%) 86 88	27, 40, 58, 74	0
2	B	169/177 (95%)	0.78	20 (11%) 5 5	23, 45, 90, 106	0
2	D	170/177 (96%)	0.18	5 (2%) 49 49	25, 46, 63, 75	0
2	F	170/177 (96%)	0.08	3 (1%) 65 66	24, 40, 60, 67	0
2	H	172/177 (97%)	0.04	5 (2%) 49 49	24, 43, 58, 68	0
2	J	170/177 (96%)	0.52	26 (15%) 3 3	24, 46, 78, 83	0
2	L	169/177 (95%)	0.39	12 (7%) 16 15	26, 45, 81, 94	0
All	All	2945/3036 (97%)	0.01	92 (3%) 47 46	22, 39, 68, 106	0

The worst 5 of 92 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	L	168	LEU	7.2
2	B	160	PRO	6.1
2	J	147	THR	5.8
2	B	166	ALA	5.7
2	B	168	LEU	5.7

### 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	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
6	SIA	I	501	20/21	0.10	-	36,45,49,51	0
3	NAG	E	331	14/15	0.26	-	55,63,71,71	0
3	NAG	A	330	14/15	0.08	-	32,39,46,48	0
5	GAL	C	331	11/12	0.12	-	53,65,70,76	0
6	GAL	I	502	11/12	0.14	-	55,60,68,69	0
3	NAG	E	332	14/15	0.34	-	62,77,84,85	0
6	GAL	K	602	11/12	0.20	-	59,70,77,77	0
3	NAG	A	331	14/15	0.15	-	44,53,62,67	0
6	SIA	K	601	20/21	0.12	-	42,51,57,57	0
5	SIA	C	330	20/21	0.09	-	33,43,47,50	0
5	NAG	C	332	15/15	0.25	-	74,83,91,94	0
3	NAG	K	632	14/15	0.21	-	67,75,84,85	0
3	NAG	K	631	14/15	0.12	-	51,58,66,69	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	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
4	NAG	G	451	14/15	0.44	-	65,84,90,94	0
4	NAG	K	651	14/15	0.23	-	53,74,82,85	0
4	NAG	I	531	14/15	0.15	-	41,53,63,64	0
4	NAG	E	341	14/15	0.37	-	76,79,87,89	0
4	NAG	I	521	14/15	0.20	-	73,83,88,92	0
4	NAG	G	411	14/15	0.32	-	70,75,82,84	0
4	NAG	A	332	14/15	0.38	-	51,62,69,70	0
4	NAG	I	541	14/15	0.14	-	57,67,75,80	0
4	NAG	K	641	14/15	0.23	-	57,63,74,78	0
4	NAG	D	261	14/15	0.13	-	65,68,72,76	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
4	NAG	K	621	14/15	0.20	-	53,67,77,83	0
4	NAG	G	431	14/15	0.10	-	40,52,62,64	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.