



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 23, 2019 – 10:30 AM EDT

PDB ID : 6D8B  
Title : The crystal structure of hemagglutinin from A/Hong Kong/125/2017 H7N9 influenza virus  
Authors : Yang, H.; Stevens, J.  
Deposited on : 2018-04-26  
Resolution : 2.95 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

<https://www.wwpdb.org/validation/2017/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:

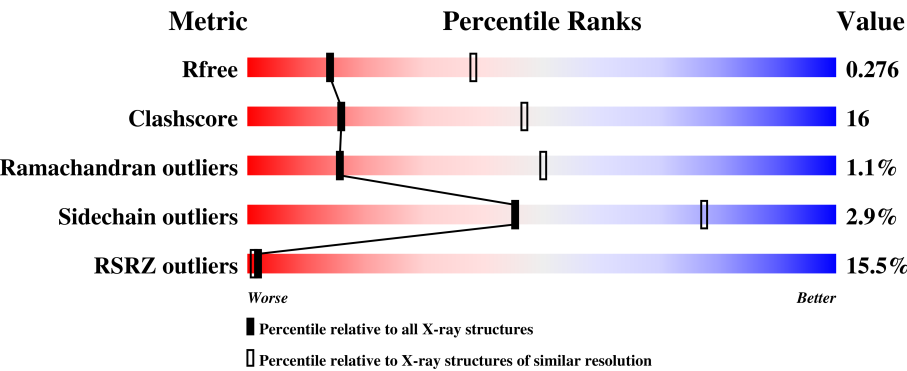
MolProbity : 4.02b-467  
Mogul : 1.8.0 (224370), CSD as540be (2019)  
Xtriage (Phenix) : 1.13  
EDS : rb-20031633  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Refmac : 5.8.0158  
CCP4 : 7.0 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20031633

# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:  
*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.95 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 <sub>free</sub>	111664	2641 (3.00-2.92)
Clashscore	122126	2988 (3.00-2.92)
Ramachandran outliers	120053	2892 (3.00-2.92)
Sidechain outliers	120020	2895 (3.00-2.92)
RSRZ outliers	108989	2527 (3.00-2.92)

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>16%</div><div>69%</div><div>27%</div><div>..</div></div>
1	C	321	<div><div>15%</div><div>71%</div><div>27%</div><div>..</div></div>
1	E	321	<div><div>13%</div><div>70%</div><div>26%</div><div>..</div></div>
2	B	221	<div><div>15%</div><div>52%</div><div>21%</div><div>•</div><div>23%</div></div>
2	D	221	<div><div>11%</div><div>56%</div><div>19%</div><div>•</div><div>23%</div></div>

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	F	221	

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	E	401	-	-	X	-

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 11499 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.

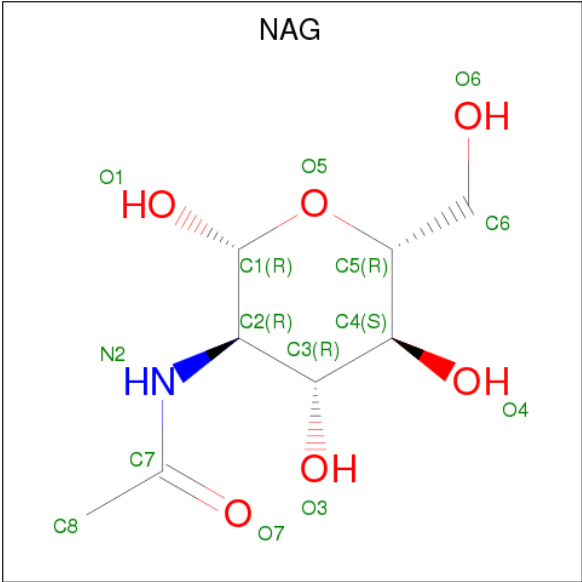
- Molecule 1 is a protein called Hemagglutinin HA1 chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	316	Total	C	N	O	S	1	0	0
			2416	1503	435	464	14			
1	C	316	Total	C	N	O	S	1	0	0
			2416	1503	435	464	14			
1	E	316	Total	C	N	O	S	1	0	0
			2416	1503	435	464	14			

- Molecule 2 is a protein called Hemagglutinin HA2 chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	171	Total	C	N	O	S	0	0	0
			1389	859	245	278	7			
2	D	171	Total	C	N	O	S	0	0	0
			1389	859	245	278	7			
2	F	171	Total	C	N	O	S	0	0	0
			1389	859	245	278	7			

- Molecule 3 is N-ACETYL-D-GLUCOSAMINE (three-letter code: NAG) (formula: C<sub>8</sub>H<sub>15</sub>NO<sub>6</sub>).

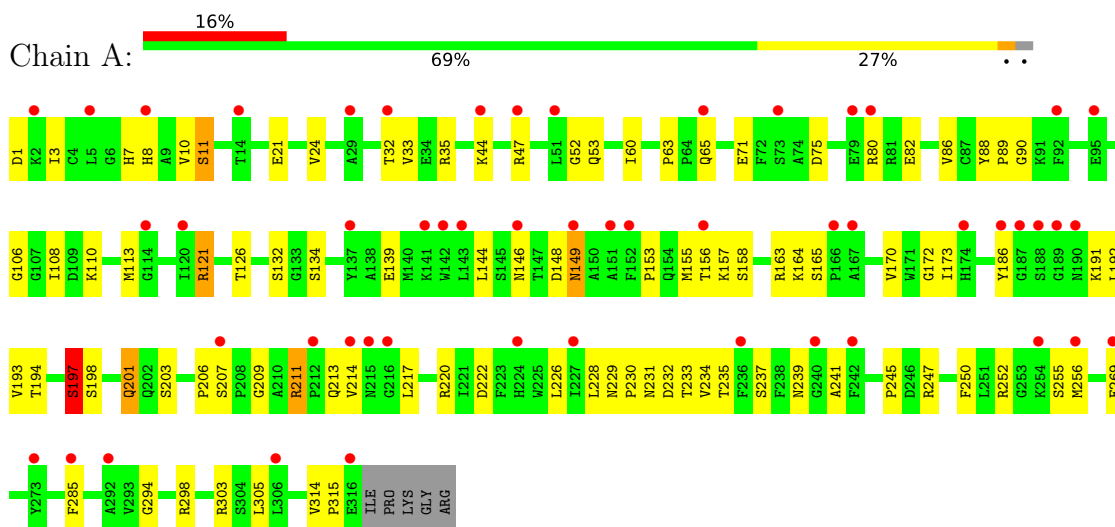


Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	B	1	Total	C	N	O	0	0
			14	8	1	5		
3	C	1	Total	C	N	O	0	0
			14	8	1	5		
3	D	1	Total	C	N	O	0	0
			14	8	1	5		
3	E	1	Total	C	N	O	0	0
			14	8	1	5		
3	F	1	Total	C	N	O	0	0
			14	8	1	5		

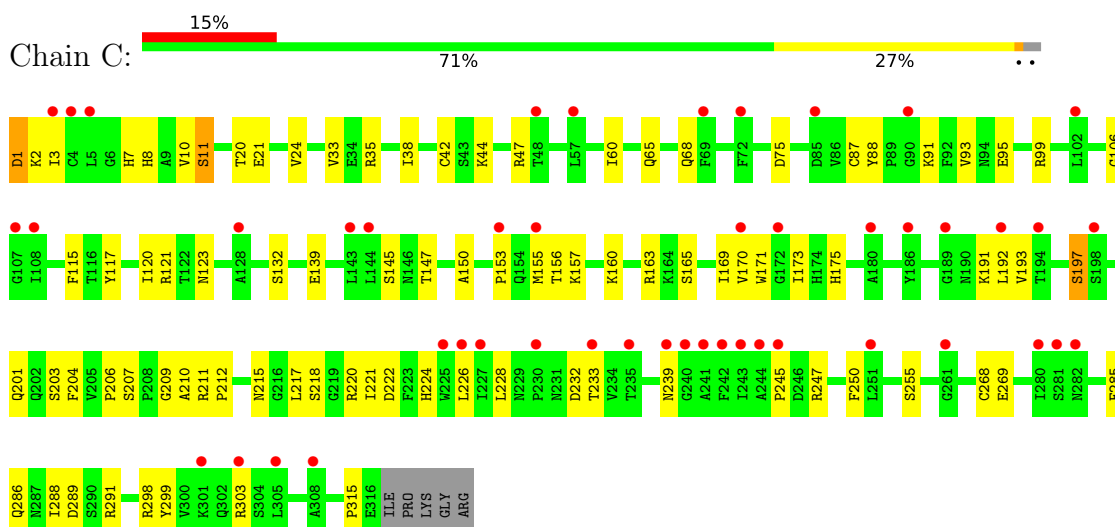
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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 HA1 chain

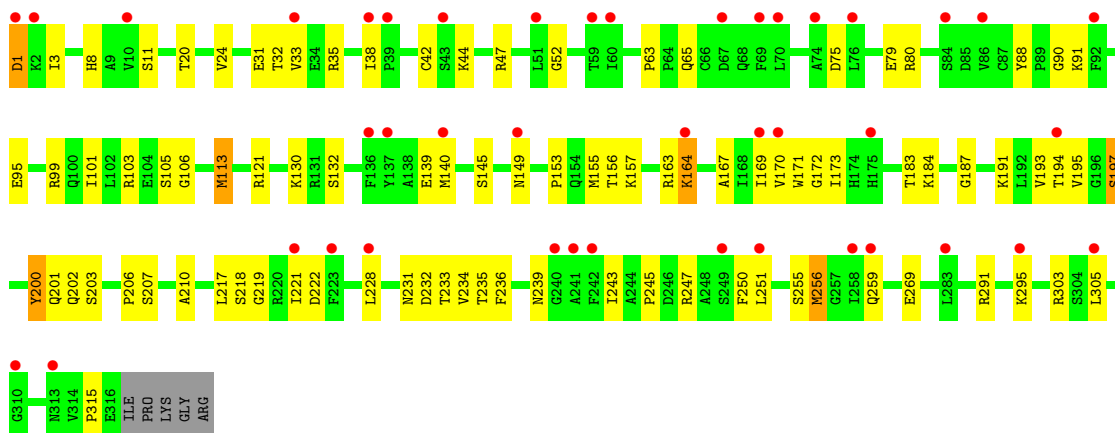


● Molecule 1: Hemagglutinin HA1 chain



● Molecule 1: Hemagglutinin HA1 chain





• Molecule 2: Hemagglutinin HA2 chain



• Molecule 2: Hemagglutinin HA2 chain



• Molecule 2: Hemagglutinin HA2 chain



SER	GLY	TYR	LYS	ASP	VAL	ILE	LEU	TRP	PHE	SER	PHE	GLY	ALA	SER	CYS	PHE	ILE	LEU	LEU	ALA	ILE	VAL	MET	GLY	LEU	VAL	PHE	ILE	CYS	VAL	LYS	ASN	GLY	ASN	MET	ARG	CYS	THR	ILE	CYS	ILE
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	202.64Å 116.89Å 119.63Å 90.00° 124.16° 90.00°	Depositor
Resolution (Å)	39.15 – 2.95 47.95 – 2.95	Depositor EDS
% Data completeness (in resolution range)	98.9 (39.15-2.95) 98.6 (47.95-2.95)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.65 (at 2.96Å)	Xtriage
Refinement program	PHENIX (dev_2733: ???)	Depositor
R, $R_{free}$	0.234 , 0.279 0.234 , 0.276	Depositor DCC
$R_{free}$ test set	2675 reflections (5.55%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	78.5	Xtriage
Anisotropy	0.101	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 38.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.000 for -1/2*h-3/2*k-l,-1/2*h+1/2*k-l,1/2* *h+1/2*k 0.000 for 1/2*h+1/2*k+2*l,1/2*h+1/2*k,-1 /2*h+1/2*k-l 0.000 for 1/2*h-1/2*k+2*l,-1/2*h+1/2*k,-1 /2*h-1/2*k-l 0.000 for -1/2*h+3/2*k-l,1/2*h+1/2*k+l,1 /2*h-1/2*k 0.000 for -h-k-l,l,k 0.003 for -h+k-l,-l,-k 0.000 for -1/2*h-1/2*k+l,-1/2*h-1/2*k-l,1/2* *h-1/2*k 0.000 for -1/2*h+1/2*k+l,1/2*h-1/2*k+l,1 /2*h+1/2*k 0.427 for 1/2*h-3/2*k,-1/2*h-1/2*k,-1/2*h +1/2*k-l 0.409 for 1/2*h+3/2*k,1/2*h-1/2*k,-1/2*h- 1/2*k-l 0.000 for -h-2*l,-k,l	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	11499	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	88.0	wwPDB-VP

<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.

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.22% of the height of the origin peak. No significant pseudotranslation is detected.*

## 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: 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	0.38	1/2462 (0.0%)	0.69	1/3329 (0.0%)
1	C	0.37	1/2462 (0.0%)	0.69	2/3329 (0.1%)
1	E	0.37	0/2462	0.71	7/3329 (0.2%)
2	B	0.37	0/1413	0.77	5/1903 (0.3%)
2	D	0.41	1/1413 (0.1%)	0.77	4/1903 (0.2%)
2	F	0.42	0/1413	0.83	3/1903 (0.2%)
All	All	0.38	3/11625 (0.0%)	0.73	22/15696 (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	A	0	2
1	C	0	1
1	E	0	1
All	All	0	4

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	268	CYS	CB-SG	-6.06	1.72	1.82
2	D	161	LYS	CE-NZ	5.37	1.62	1.49
1	A	164	LYS	CE-NZ	5.25	1.62	1.49

The worst 5 of 22 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	F	121	ARG	NE-CZ-NH1	10.68	125.64	120.30

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	164	LYS	CD-CE-NZ	-9.99	88.73	111.70
1	E	1	ASP	N-CA-C	-9.98	84.06	111.00
1	C	1	ASP	N-CA-C	-9.26	85.99	111.00
2	B	152	ILE	CG1-CB-CG2	-8.86	91.91	111.40

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	197	SER	Peptide
1	A	201	GLN	Sidechain
1	C	197	SER	Peptide
1	E	197	SER	Peptide

## 5.2 Too-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 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	2416	0	2382	76	0
1	C	2416	0	2382	86	0
1	E	2416	0	2382	74	0
2	B	1389	0	1300	63	0
2	D	1389	0	1300	35	0
2	F	1389	0	1300	69	0
3	A	14	0	13	0	0
3	B	14	0	13	0	0
3	C	14	0	13	0	0
3	D	14	0	13	0	0
3	E	14	0	13	8	0
3	F	14	0	13	0	0
All	All	11499	0	11124	353	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 16.

The worst 5 of 353 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:191:LYS:HA	1:C:239:ASN:HD21	1.15	1.01
2:B:9:PHE:HB2	2:F:121:ARG:CZ	1.90	1.01
2:F:121:ARG:HH11	2:F:124:ARG:HD2	1.24	1.00
1:A:191:LYS:HA	1:A:239:ASN:HD21	1.22	1.00
1:E:191:LYS:HA	1:E:239:ASN:HD21	1.24	1.00

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	314/321 (98%)	293 (93%)	19 (6%)	2 (1%)	27	64
1	C	314/321 (98%)	294 (94%)	18 (6%)	2 (1%)	27	64
1	E	314/321 (98%)	293 (93%)	18 (6%)	3 (1%)	17	53
2	B	169/221 (76%)	153 (90%)	13 (8%)	3 (2%)	9	37
2	D	169/221 (76%)	153 (90%)	13 (8%)	3 (2%)	9	37
2	F	169/221 (76%)	153 (90%)	13 (8%)	3 (2%)	9	37
All	All	1449/1626 (89%)	1339 (92%)	94 (6%)	16 (1%)	16	50

5 of 16 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	3	PHE
2	B	12	ASN
2	D	12	ASN
2	F	12	ASN
2	F	127	ARG

### 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	267/271 (98%)	258 (97%)	9 (3%)	40	74
1	C	267/271 (98%)	263 (98%)	4 (2%)	67	88
1	E	267/271 (98%)	259 (97%)	8 (3%)	44	77
2	B	145/189 (77%)	138 (95%)	7 (5%)	28	64
2	D	145/189 (77%)	140 (97%)	5 (3%)	40	74
2	F	145/189 (77%)	142 (98%)	3 (2%)	56	84
All	All	1236/1380 (90%)	1200 (97%)	36 (3%)	45	77

5 of 36 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	8	HIS
2	D	19	ASP
2	F	26	HIS
1	C	132	SER
2	D	26	HIS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 21 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	239	ASN
1	E	65	GLN
1	E	259	GLN
1	C	224	HIS
2	F	26	HIS

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

6 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	401	1	14,14,15	1.93	2 (14%)	17,19,21	2.41	3 (17%)
3	NAG	B	301	2	14,14,15	0.46	0	17,19,21	0.71	1 (5%)
3	NAG	C	401	1	14,14,15	0.28	0	17,19,21	1.17	3 (17%)
3	NAG	D	301	2	14,14,15	0.43	0	17,19,21	0.64	1 (5%)
3	NAG	E	401	1	14,14,15	0.96	1 (7%)	17,19,21	2.50	6 (35%)
3	NAG	F	301	2	14,14,15	0.38	0	17,19,21	0.60	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	401	1	-	0/6/23/26	0/1/1/1
3	NAG	B	301	2	-	0/6/23/26	0/1/1/1
3	NAG	C	401	1	-	0/6/23/26	0/1/1/1
3	NAG	D	301	2	-	0/6/23/26	0/1/1/1
3	NAG	E	401	1	-	0/6/23/26	0/1/1/1
3	NAG	F	301	2	-	0/6/23/26	0/1/1/1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	E	401	NAG	O5-C1	2.77	1.48	1.43
3	A	401	NAG	O5-C1	3.01	1.48	1.43
3	A	401	NAG	C1-C2	6.39	1.62	1.52

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	401	NAG	C4-C3-C2	-2.49	107.36	111.02
3	E	401	NAG	C4-C3-C2	-2.17	107.84	111.02
3	E	401	NAG	O3-C3-C4	2.09	115.20	110.34
3	D	301	NAG	C1-O5-C5	2.14	115.11	112.20
3	C	401	NAG	C1-C2-N2	2.22	114.29	110.49

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	E	401	NAG	8	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	316/321 (98%)	0.96	52 (16%) <b>1</b> <b>1</b>	42, 80, 122, 150	1 (0%)
1	C	316/321 (98%)	0.93	47 (14%) <b>2</b> <b>1</b>	45, 79, 119, 149	1 (0%)
1	E	316/321 (98%)	0.69	42 (13%) <b>3</b> <b>1</b>	43, 79, 123, 151	1 (0%)
2	B	171/221 (77%)	1.14	34 (19%) <b>1</b> <b>0</b>	41, 99, 139, 149	0
2	D	171/221 (77%)	0.78	24 (14%) <b>2</b> <b>1</b>	42, 100, 140, 156	0
2	F	171/221 (77%)	0.76	27 (15%) <b>2</b> <b>1</b>	40, 101, 139, 149	0
All	All	1461/1626 (89%)	0.87	226 (15%) <b>2</b> <b>1</b>	40, 87, 134, 156	3 (0%)

The worst 5 of 226 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	24	PHE	13.2
1	A	190	ASN	10.1
2	F	60	ASN	10.0
2	B	23	GLY	9.2
1	A	188	SER	8.7

### 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](#)

There are no carbohydrates in this entry.

## 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. 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	B-factors(Å <sup>2</sup> )	Q<0.9
3	NAG	A	401	14/15	0.43	0.36	124,154,169,169	0
3	NAG	C	401	14/15	0.72	0.33	138,158,163,169	0
3	NAG	F	301	14/15	0.78	0.18	81,107,124,128	0
3	NAG	D	301	14/15	0.80	0.36	79,100,123,142	0
3	NAG	E	401	14/15	0.82	0.22	112,158,164,165	0
3	NAG	B	301	14/15	0.85	0.22	84,107,127,136	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.