



# wwPDB X-ray Structure Validation Summary Report

Aug 17, 2022 – 04:22 PM EDT


PDB ID : 4K65  
Title : Structure of an airborne transmissible avian influenza H5 hemagglutinin mutant from the influenza virus A/Indonesia/5/2005  
Authors : Zhang, W.; Shi, Y.; Lu, X.; Shu, Y.; Qi, J.; Gao, G.F.  
Deposited on : 2013-04-15  
Resolution : 2.90 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 types of validation reports are described at <http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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

MolProbity : 4.02b-467  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.29  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0158  
CCP4 : 7.0.044 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.29

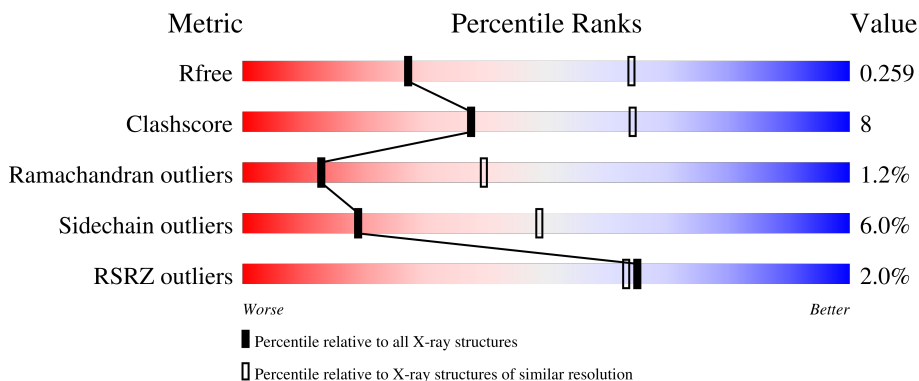
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.90 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.



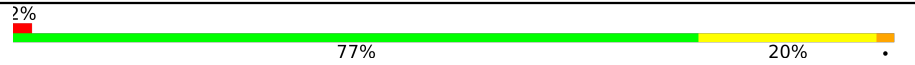


Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria respectively. 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	 79% 21% .
1	C	321	 81% 17% .
1	E	321	 72% 26% .
1	G	321	 73% 26% .
2	B	164	 81% 16% .

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	D	164	 <p>2% 77% 20%</p>
2	F	164	 <p>3% 73% 24%</p>
2	H	164	 <p>4% 71% 26%</p>

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 15508 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
			Total	C	N	O	S			
1	A	321	2542	1609	433	485	15	0	0	0
1	C	321	2542	1609	433	485	15	0	0	0
1	E	321	2542	1609	433	485	15	0	0	0
1	G	321	2542	1609	433	485	15	0	0	0

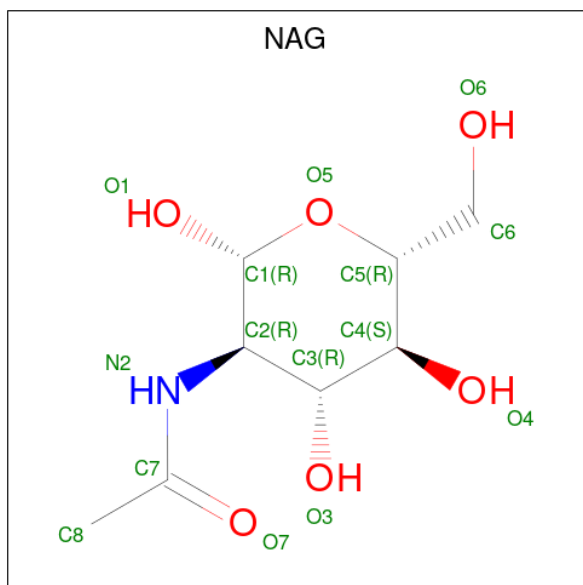
There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	4	GLN	-	expression tag	UNP A8HWY8
A	107	TYR	HIS	engineered mutation	UNP A8HWY8
A	160	ALA	THR	engineered mutation	UNP A8HWY8
A	226	LEU	GLN	engineered mutation	UNP A8HWY8
A	228	SER	GLY	engineered mutation	UNP A8HWY8
C	4	GLN	-	expression tag	UNP A8HWY8
C	107	TYR	HIS	engineered mutation	UNP A8HWY8
C	160	ALA	THR	engineered mutation	UNP A8HWY8
C	226	LEU	GLN	engineered mutation	UNP A8HWY8
C	228	SER	GLY	engineered mutation	UNP A8HWY8
E	4	GLN	-	expression tag	UNP A8HWY8
E	107	TYR	HIS	engineered mutation	UNP A8HWY8
E	160	ALA	THR	engineered mutation	UNP A8HWY8
E	226	LEU	GLN	engineered mutation	UNP A8HWY8
E	228	SER	GLY	engineered mutation	UNP A8HWY8
G	4	GLN	-	expression tag	UNP A8HWY8
G	107	TYR	HIS	engineered mutation	UNP A8HWY8
G	160	ALA	THR	engineered mutation	UNP A8HWY8
G	226	LEU	GLN	engineered mutation	UNP A8HWY8
G	228	SER	GLY	engineered mutation	UNP A8HWY8

- Molecule 2 is a protein called Hemagglutinin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	164	Total	C	N	O	S	0	0	0
			1328	828	229	263	8			
2	D	164	Total	C	N	O	S	0	0	0
			1328	828	229	263	8			
2	F	164	Total	C	N	O	S	0	0	0
			1328	828	229	263	8			
2	H	164	Total	C	N	O	S	0	0	0
			1328	828	229	263	8			

- Molecule 3 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula:  $C_8H_{15}NO_6$ ).

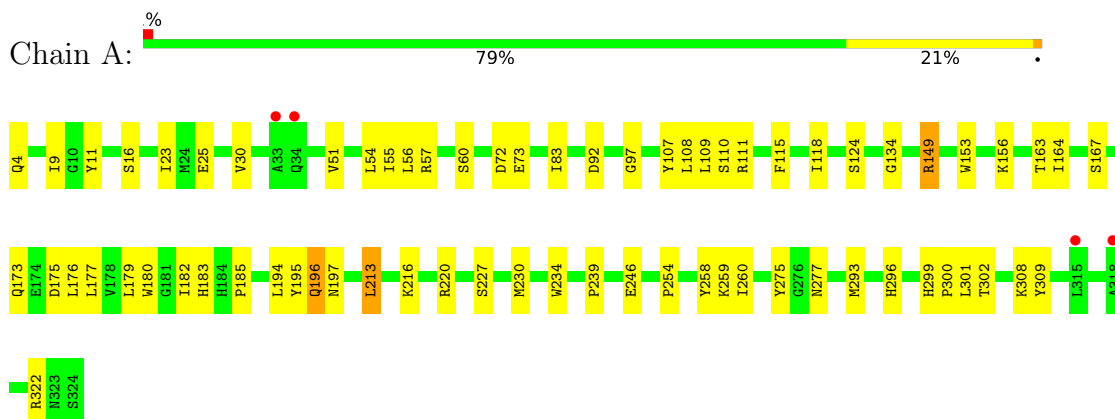


Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	C	1	Total	C	N	O	0	0
			14	8	1	5		

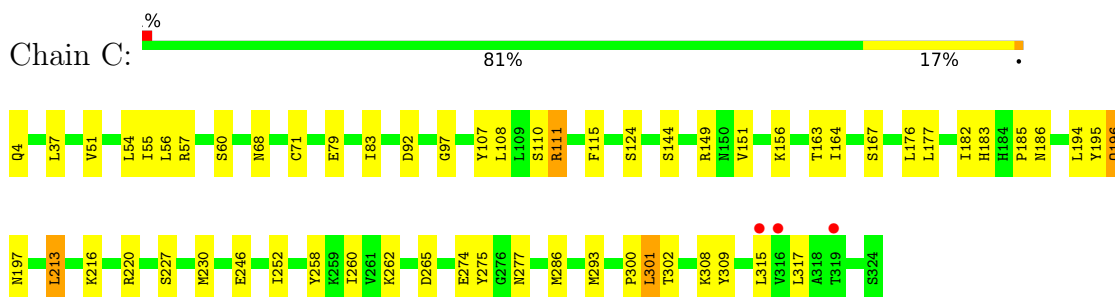
### 3 Residue-property plots

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

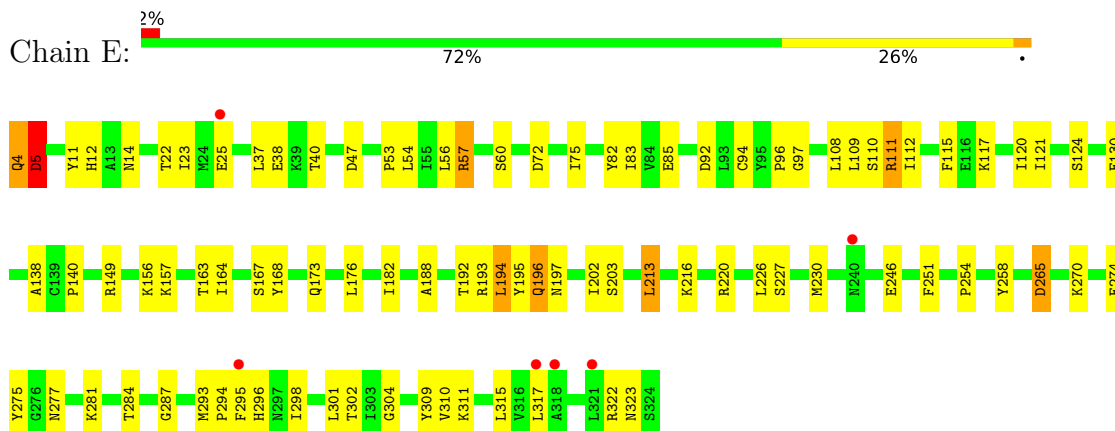
- Molecule 1: Hemagglutinin



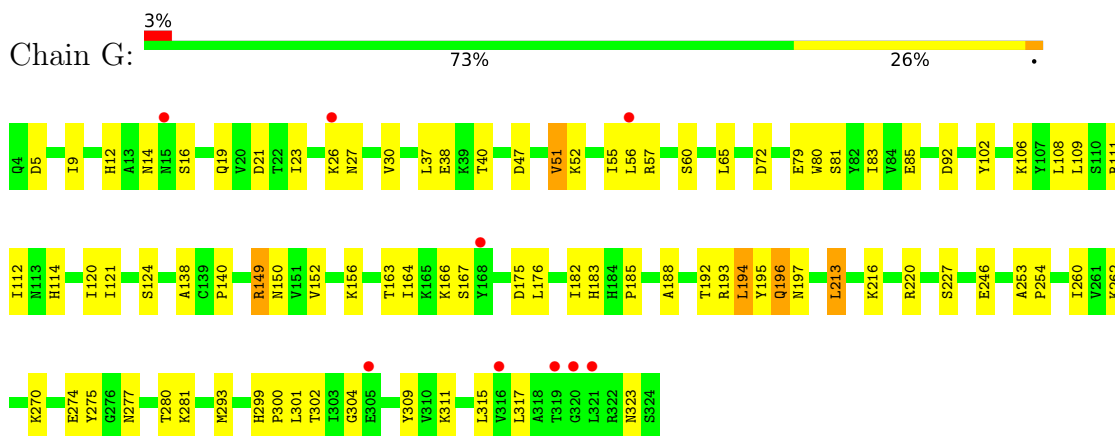
- Molecule 1: Hemagglutinin



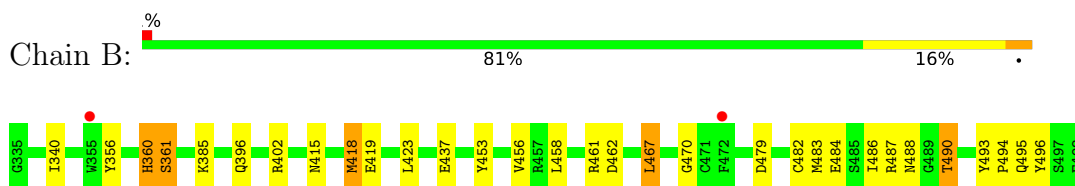
- Molecule 1: Hemagglutinin



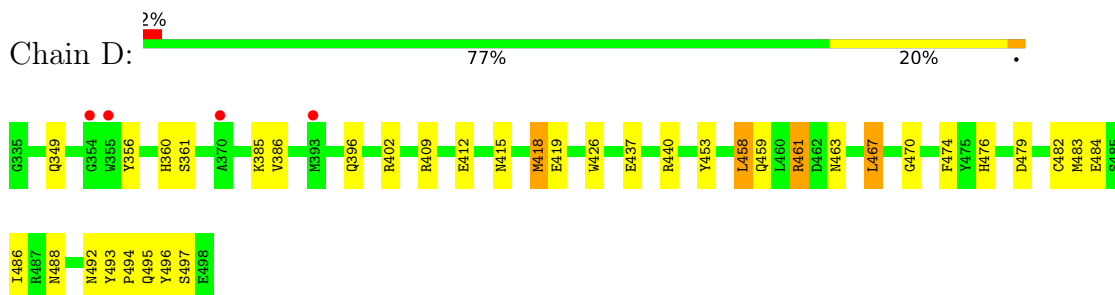
- Molecule 1: Hemagglutinin



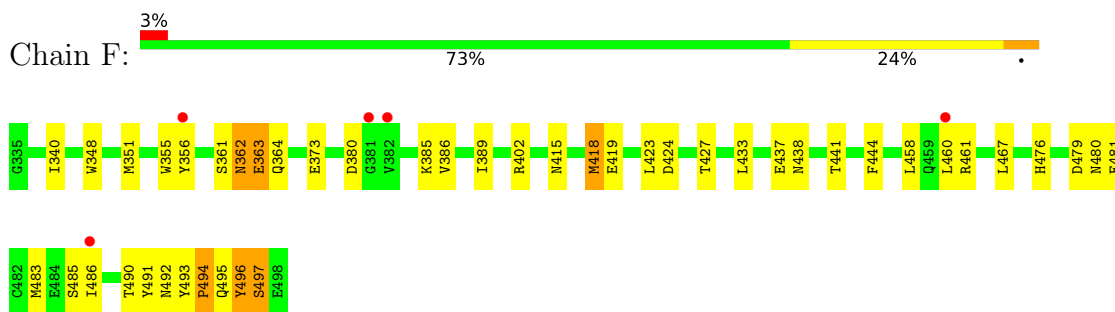
- Molecule 2: Hemagglutinin



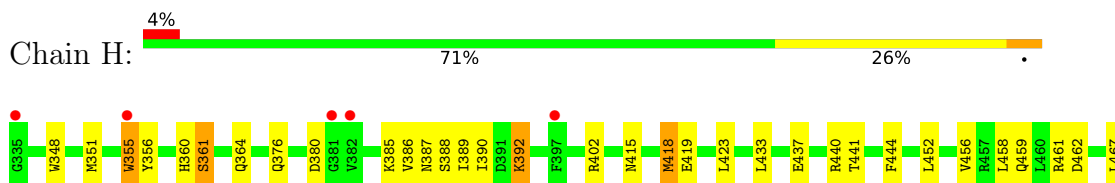
- Molecule 2: Hemagglutinin

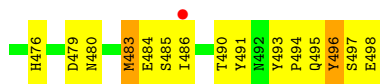


- Molecule 2: Hemagglutinin



- Molecule 2: Hemagglutinin







## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 3	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	70.64Å 70.64Å 487.66Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	48.88 – 2.90 48.88 – 2.89	Depositor EDS
% Data completeness (in resolution range)	92.3 (48.88-2.90) 92.4 (48.88-2.89)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.78 (at 2.91Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, $R_{free}$	0.208 , 0.256 0.210 , 0.259	Depositor DCC
$R_{free}$ test set	2834 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	51.9	Xtriage
Anisotropy	0.091	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 6.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.447 for -h,-k,l 0.083 for h,-h-k,-l 0.078 for -k,-h,-l	Xtriage
Reported twinning fraction	0.431 for -h,-k,l	Depositor
Outliers	0 of 56175 reflections	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	15508	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	61.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.36% 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: 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	0.29	0/2604	0.52	0/3539
1	C	0.30	0/2604	0.51	0/3539
1	E	0.27	0/2604	0.50	0/3539
1	G	0.28	0/2604	0.49	0/3539
2	B	0.31	0/1355	0.48	0/1823
2	D	0.32	0/1355	0.49	0/1823
2	F	0.35	0/1355	0.50	0/1823
2	H	0.33	0/1355	0.53	0/1823
All	All	0.30	0/15836	0.50	0/21448

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	2542	0	2483	33	0
1	C	2542	0	2483	26	0
1	E	2542	0	2484	52	0
1	G	2542	0	2484	50	0
2	B	1328	0	1231	17	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	D	1328	0	1231	20	0
2	F	1328	0	1231	27	0
2	H	1328	0	1231	27	0
3	A	14	0	13	0	0
3	C	14	0	13	0	0
All	All	15508	0	14884	233	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 8.

The worst 5 of 233 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:111:ARG:NH2	1:C:265:ASP:OD1	2.13	0.81
1:G:120:ILE:HD11	1:G:254:PRO:HB2	1.70	0.74
1:G:14:ASN:H	1:G:323:ASN:HD21	1.39	0.71
1:G:14:ASN:OD1	1:G:323:ASN:ND2	2.25	0.70
2:F:402:ARG:NH1	2:F:415:ASN:OD1	2.25	0.69

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	319/321 (99%)	308 (97%)	10 (3%)	1 (0%)	41	71
1	C	319/321 (99%)	307 (96%)	10 (3%)	2 (1%)	25	58
1	E	319/321 (99%)	305 (96%)	11 (3%)	3 (1%)	17	48
1	G	319/321 (99%)	306 (96%)	11 (3%)	2 (1%)	25	58
2	B	162/164 (99%)	150 (93%)	9 (6%)	3 (2%)	8	28

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	D	162/164 (99%)	149 (92%)	9 (6%)	4 (2%)	5	21
2	F	162/164 (99%)	142 (88%)	16 (10%)	4 (2%)	5	21
2	H	162/164 (99%)	148 (91%)	10 (6%)	4 (2%)	5	21
All	All	1924/1940 (99%)	1815 (94%)	86 (4%)	23 (1%)	13	40

5 of 23 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	196	GLN
1	C	196	GLN
2	D	461	ARG
2	D	496	TYR
1	E	196	GLN

### 5.3.2 Protein sidechains [i](#)

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	288/288 (100%)	271 (94%)	17 (6%)	19	49
1	C	288/288 (100%)	270 (94%)	18 (6%)	18	46
1	E	288/288 (100%)	270 (94%)	18 (6%)	18	46
1	G	288/288 (100%)	272 (94%)	16 (6%)	21	52
2	B	140/140 (100%)	133 (95%)	7 (5%)	24	57
2	D	140/140 (100%)	132 (94%)	8 (6%)	20	51
2	F	140/140 (100%)	131 (94%)	9 (6%)	17	45
2	H	140/140 (100%)	130 (93%)	10 (7%)	14	40
All	All	1712/1712 (100%)	1609 (94%)	103 (6%)	19	49

5 of 103 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	124	SER

Continued on next page...

*Continued from previous page...*

Mol	Chain	Res	Type
2	F	380	ASP
2	H	440	ARG
1	E	167	SER
1	E	293	MET

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	E	296	HIS
2	F	492	ASN
1	G	323	ASN

### 5.3.3 RNA [i](#)

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 monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

2 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	C	601	1	14,14,15	0.48	0	17,19,21	1.09	1 (5%)
3	NAG	A	601	1	14,14,15	0.50	0	17,19,21	0.97	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	C	601	1	-	4/6/23/26	0/1/1/1
3	NAG	A	601	1	-	4/6/23/26	0/1/1/1

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	601	NAG	C1-O5-C5	3.86	117.42	112.19
3	A	601	NAG	C1-O5-C5	2.72	115.88	112.19

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	601	NAG	O5-C5-C6-O6
3	C	601	NAG	O5-C5-C6-O6
3	C	601	NAG	C4-C5-C6-O6
3	A	601	NAG	C4-C5-C6-O6
3	A	601	NAG	C8-C7-N2-C2

There are no ring outliers.

No monomer is involved in short contacts.

## 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%)	-0.19	4 (1%) 79 79	9, 34, 74, 117	0
1	C	321/321 (100%)	-0.17	3 (0%) 84 84	9, 33, 72, 119	0
1	E	321/321 (100%)	-0.04	6 (1%) 66 65	28, 70, 114, 155	0
1	G	321/321 (100%)	0.01	9 (2%) 53 49	31, 68, 113, 163	0
2	B	164/164 (100%)	-0.03	2 (1%) 79 79	17, 68, 100, 121	0
2	D	164/164 (100%)	0.07	4 (2%) 59 56	19, 67, 99, 114	0
2	F	164/164 (100%)	0.23	5 (3%) 50 45	34, 78, 141, 173	0
2	H	164/164 (100%)	0.23	6 (3%) 41 37	28, 77, 139, 174	0
All	All	1940/1940 (100%)	-0.02	39 (2%) 65 63	9, 59, 115, 174	0

The worst 5 of 39 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	H	382	VAL	4.5
1	G	168	TYR	4.3
1	G	319	THR	3.8
2	H	381	GLY	3.6
2	H	355	TRP	3.6

### 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 monosaccharides 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	601	14/15	0.87	0.17	42,60,69,78	0
3	NAG	C	601	14/15	0.89	0.15	39,59,72,72	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.