



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 17, 2026 – 08:27 PM UTC

PDB ID : 4K65 / pdb_00004k65
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

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 : 4-5-2 with Phenix2.0
Mogul : 2022.3.0, CSD as543be (2022)
Xtriage (Phenix) : 2.0
EDS : 3.0
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4 : 9.0.010 (Gargrove)
Density-Fitness : 1.0.12
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.49

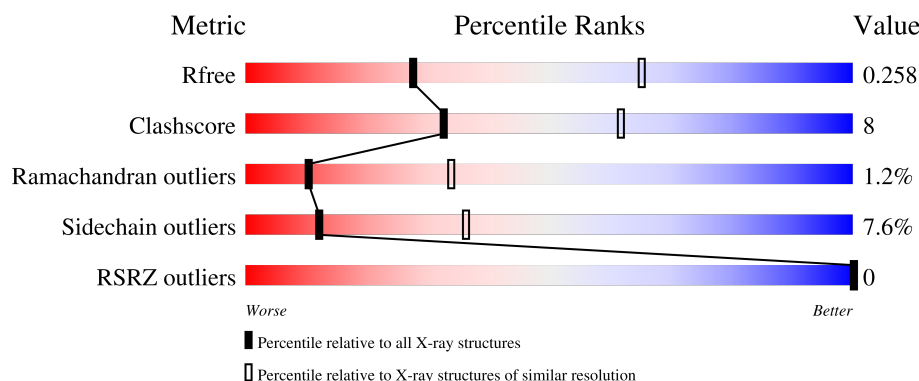
1 Overall quality at a glance

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}	180053	2481 (2.90-2.90)
Clashscore	190562	2690 (2.90-2.90)
Ramachandran outliers	187476	2623 (2.90-2.90)
Sidechain outliers	187428	2625 (2.90-2.90)
RSRZ outliers	180081	2481 (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 ≥ 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	 77% 22% .
1	C	321	 79% 19% .
1	E	321	 71% 26% .
1	G	321	 71% 27% .
2	B	164	 78% 20% .

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	D	164	 76% 21% .
2	F	164	 73% 23% .
2	H	164	 71% 26% .

2 Entry composition

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
1	A	321	Total	C	N	O	S	0	0	0
			2542	1609	433	485	15			
1	C	321	Total	C	N	O	S	0	0	0
			2542	1609	433	485	15			
1	E	321	Total	C	N	O	S	0	0	0
			2542	1609	433	485	15			
1	G	321	Total	C	N	O	S	0	0	0
			2542	1609	433	485	15			

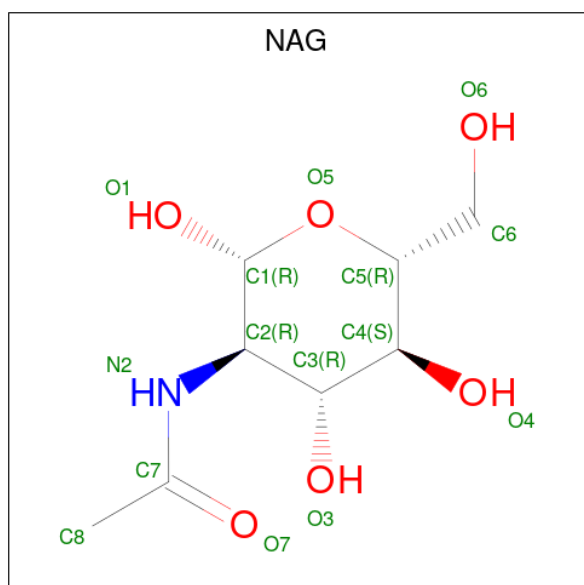
There are 20 discrepancies between the modelled and reference sequences:

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
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 (CCD ID: NAG) (formula: $C_8H_{15}NO_6$).

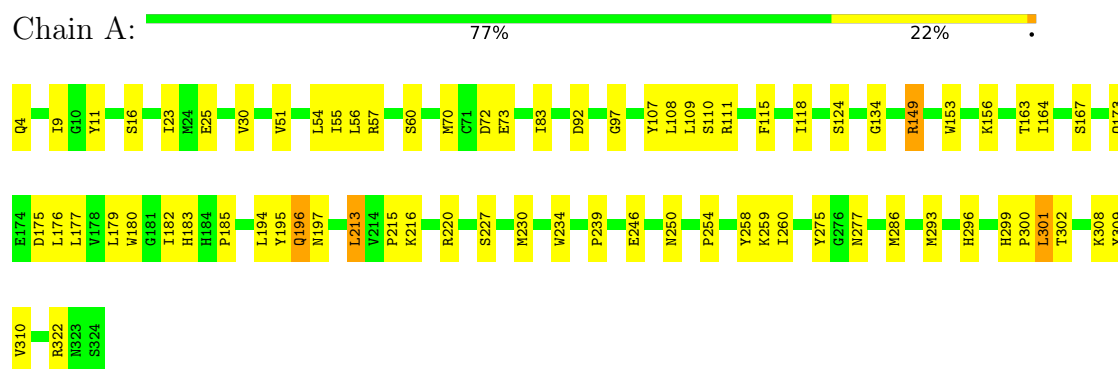


Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	C	1	Total	C	N	O	0	0
			14	8	1	5		

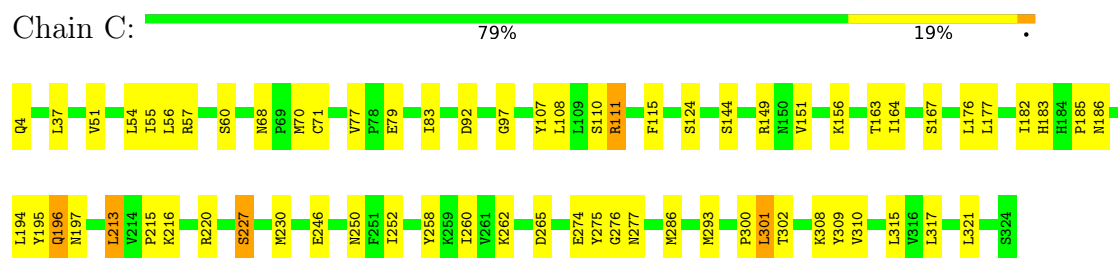
3 Residue-property plots

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.

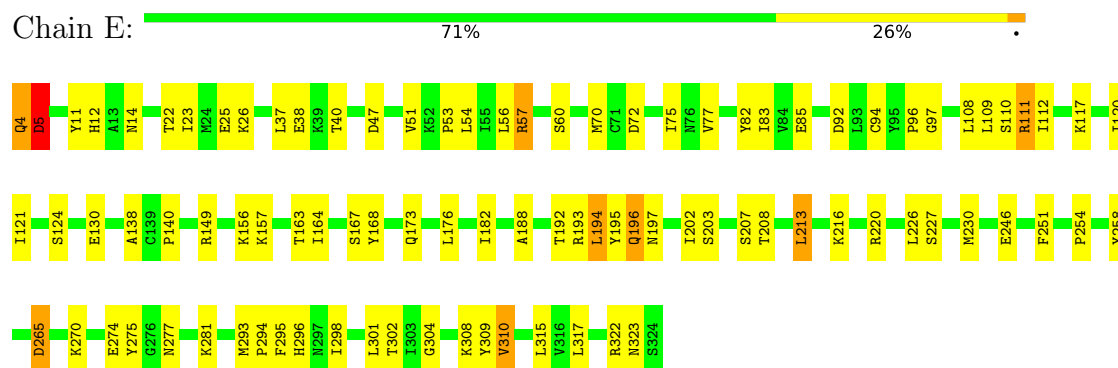
- Molecule 1: Hemagglutinin



- Molecule 1: Hemagglutinin

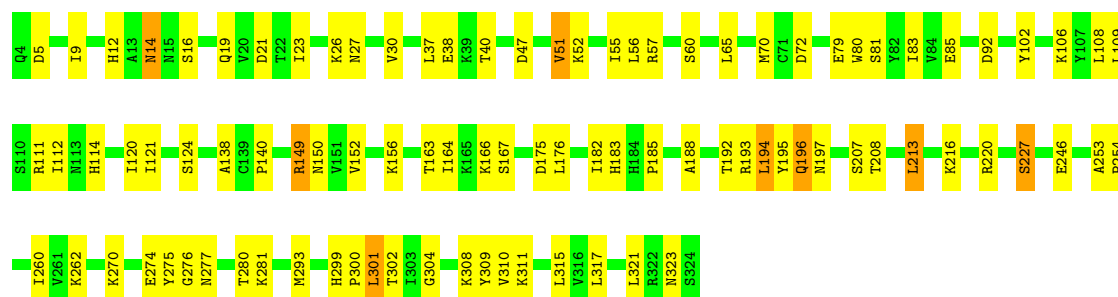


- Molecule 1: Hemagglutinin



- Molecule 1: Hemagglutinin

Chain G:  71% 27% .



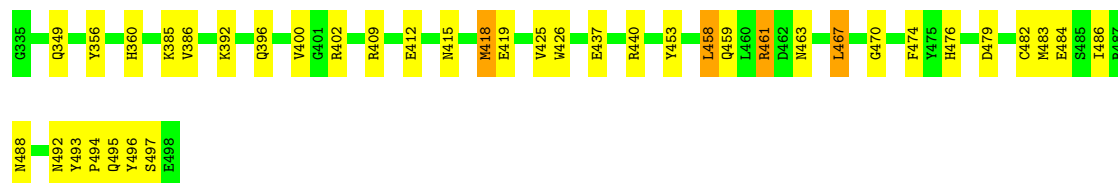
• Molecule 2: Hemagglutinin

Chain B:  78% 20% .



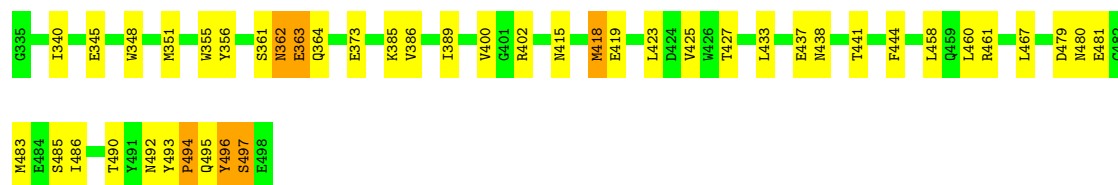
• Molecule 2: Hemagglutinin

Chain D:  76% 21% .



• Molecule 2: Hemagglutinin

Chain F:  73% 23% .



• Molecule 2: Hemagglutinin

Chain H:  71% 26% .



D479	M483	E484	S485	I486	T490	Y493	P494	Q495	Y496	S497	E498
------	------	------	------	------	------	------	------	------	------	------	------

4 Data and refinement statistics

Property	Value	Source
Space group	P 3	Depositor
Cell constants a, b, c, α , β , γ	70.64Å 70.64Å 487.66Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	48.88 – 2.90 48.88 – 2.90	Depositor EDS
% Data completeness (in resolution range)	92.3 (48.88-2.90) 92.6 (48.88-2.90)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.78 (at 2.91Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, R_{free}	0.208 , 0.256 0.211 , 0.258	Depositor DCC
R_{free} test set	2834 reflections (4.66%)	wwPDB-VP
Wilson B-factor (Å ²)	51.9	Xtriage
Anisotropy	0.091	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.26 , 12.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	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	0.96	EDS
Total number of atoms	15508	wwPDB-VP
Average B, all atoms (Å ²)	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.*

¹Intensities estimated from amplitudes.

²Theoretical 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	0.35	0/2604	0.81	0/3539
1	C	0.36	0/2604	0.80	1/3539 (0.0%)
1	E	0.33	0/2604	0.79	1/3539 (0.0%)
1	G	0.33	0/2604	0.79	3/3539 (0.1%)
2	B	0.35	0/1355	0.75	0/1823
2	D	0.36	0/1355	0.75	0/1823
2	F	0.44	0/1355	0.78	0/1823
2	H	0.37	0/1355	0.82	0/1823
All	All	0.36	0/15836	0.79	5/21448 (0.0%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	276	GLY	N-CA-C	-5.43	104.31	112.89
1	G	276	GLY	N-CA-C	-5.42	104.40	113.02
1	G	14	ASN	N-CA-C	5.34	115.17	108.45
1	G	207	SER	N-CA-C	-5.15	106.51	112.89
1	E	207	SER	N-CA-C	-5.01	106.67	112.89

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	35	0
1	C	2542	0	2483	28	0
1	E	2542	0	2484	51	0
1	G	2542	0	2484	52	0
2	B	1328	0	1231	19	0
2	D	1328	0	1231	20	0
2	F	1328	0	1231	25	0
2	H	1328	0	1231	28	0
3	A	14	0	13	0	0
3	C	14	0	13	0	0
All	All	15508	0	14884	240	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 240 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%)	36 65
1	C	319/321 (99%)	307 (96%)	10 (3%)	2 (1%)	21 51

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	E	319/321 (99%)	305 (96%)	11 (3%)	3 (1%)	14	41
1	G	319/321 (99%)	306 (96%)	11 (3%)	2 (1%)	21	51
2	B	162/164 (99%)	150 (93%)	9 (6%)	3 (2%)	6	23
2	D	162/164 (99%)	149 (92%)	9 (6%)	4 (2%)	4	17
2	F	162/164 (99%)	142 (88%)	16 (10%)	4 (2%)	4	17
2	H	162/164 (99%)	148 (91%)	10 (6%)	4 (2%)	4	17
All	All	1924/1940 (99%)	1815 (94%)	86 (4%)	23 (1%)	10	34

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 ⓘ

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%)	268 (93%)	20 (7%)	14	41
1	C	288/288 (100%)	266 (92%)	22 (8%)	12	36
1	E	288/288 (100%)	265 (92%)	23 (8%)	11	34
1	G	288/288 (100%)	266 (92%)	22 (8%)	12	36
2	B	140/140 (100%)	129 (92%)	11 (8%)	11	34
2	D	140/140 (100%)	130 (93%)	10 (7%)	13	40
2	F	140/140 (100%)	129 (92%)	11 (8%)	11	34
2	H	140/140 (100%)	129 (92%)	11 (8%)	11	34
All	All	1712/1712 (100%)	1582 (92%)	130 (8%)	12	36

5 of 130 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	302	THR
2	H	356	TYR
1	C	293	MET
1	C	286	MET
2	H	361	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
2	F	376	GLN
2	F	492	ASN
1	G	197	ASN
1	G	170	ASN
1	G	172	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 ⓘ

There are no oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

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	A	601	1	14,14,15	0.48	0	17,19,21	0.96	1 (5%)
3	NAG	C	601	1	14,14,15	0.47	0	17,19,21	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	-	4/6/23/26	0/1/1/1
3	NAG	C	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	3.90	117.42	112.19
3	A	601	NAG	C1-O5-C5	2.75	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 ⓘ

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 [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, 95th 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 ²)	Q<0.9
1	A	321/321 (100%)	-1.65	0 100 100	9, 34, 74, 117	0
1	C	321/321 (100%)	-1.69	0 100 100	9, 33, 72, 119	0
1	E	321/321 (100%)	-1.43	0 100 100	28, 70, 114, 155	0
1	G	321/321 (100%)	-1.40	0 100 100	31, 68, 113, 163	0
2	B	164/164 (100%)	-1.45	0 100 100	17, 68, 100, 121	0
2	D	164/164 (100%)	-1.40	0 100 100	19, 67, 99, 114	0
2	F	164/164 (100%)	-1.15	0 100 100	34, 78, 141, 173	0
2	H	164/164 (100%)	-1.21	0 100 100	28, 77, 139, 174	0
All	All	1940/1940 (100%)	-1.46	0 100 100	9, 59, 115, 174	0

There are no RSRZ outliers to report.

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 oligosaccharides 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, 95th 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($\text{\AA}^2$)	Q<0.9
3	NAG	A	601	14/15	0.98	0.04	42,60,69,78	0
3	NAG	C	601	14/15	0.98	0.03	39,59,72,72	0

6.5 Other polymers [i](#)

There are no such residues in this entry.