



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 02:27 am GMT

PDB ID : 4K62
Title : Structure of an avian influenza H5 hemagglutinin 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 : 2.50 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

<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

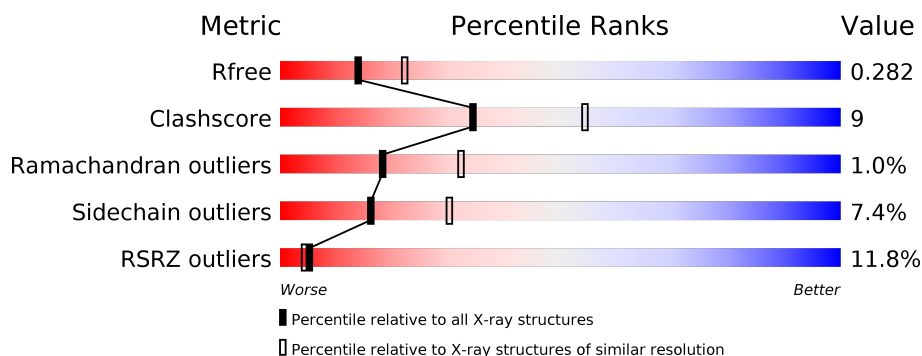
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.50 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	3846 (2.50-2.50)
Clashscore	112137	4554 (2.50-2.50)
Ramachandran outliers	110173	4463 (2.50-2.50)
Sidechain outliers	110143	4465 (2.50-2.50)
RSRZ outliers	101464	3876 (2.50-2.50)

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 ≥ 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>75%</div> <div>23%</div> <div>.</div> </div>
1	C	321	<div> <div>8%</div> <div>76%</div> <div>23%</div> <div>.</div> </div>
1	E	321	<div> <div>24%</div> <div>70%</div> <div>27%</div> <div>.</div> </div>
1	G	321	<div> <div>21%</div> <div>70%</div> <div>27%</div> <div>.</div> </div>
2	B	164	<div> <div>8%</div> <div>73%</div> <div>24%</div> <div>.</div> </div>
2	D	164	<div> <div>8%</div> <div>74%</div> <div>24%</div> <div>.</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	164	<div><div></div><div>17%</div><div></div><div>75%</div><div></div><div>22%</div><div></div><div>• •</div></div>
2	H	164	<div><div></div><div>16%</div><div></div><div>81%</div><div></div><div>15%</div><div></div><div>•</div></div>

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 15877 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
			2541	1605	436	485	15			
1	C	321	Total	C	N	O	S	0	0	0
			2541	1605	436	485	15			
1	E	321	Total	C	N	O	S	0	0	0
			2541	1605	436	485	15			
1	G	321	Total	C	N	O	S	0	0	0
			2541	1605	436	485	15			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	4	GLN	-	EXPRESSION TAG	UNP A8HWY8
C	4	GLN	-	EXPRESSION TAG	UNP A8HWY8
E	4	GLN	-	EXPRESSION TAG	UNP A8HWY8
G	4	GLN	-	EXPRESSION TAG	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 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	C	1	Total	C	N	O	0	0
			14	8	1	5		

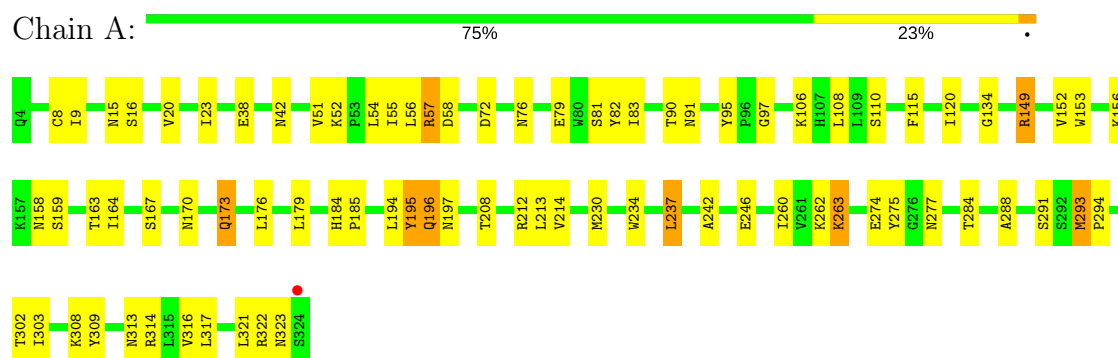
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	105	Total	O	0	0
			105	105		
4	B	23	Total	O	0	0
			23	23		
4	C	120	Total	O	0	0
			120	120		
4	D	19	Total	O	0	0
			19	19		
4	E	36	Total	O	0	0
			36	36		
4	F	24	Total	O	0	0
			24	24		
4	G	23	Total	O	0	0
			23	23		
4	H	23	Total	O	0	0
			23	23		

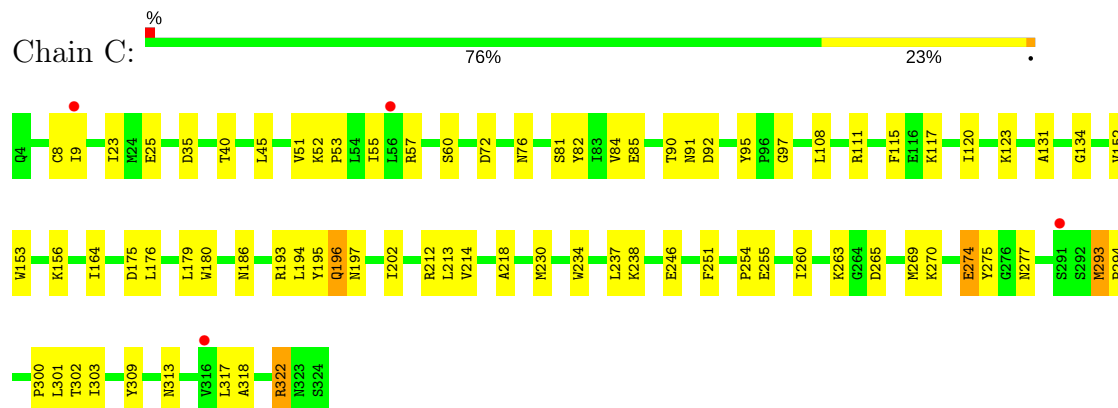
3 Residue-property plots [i](#)

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.

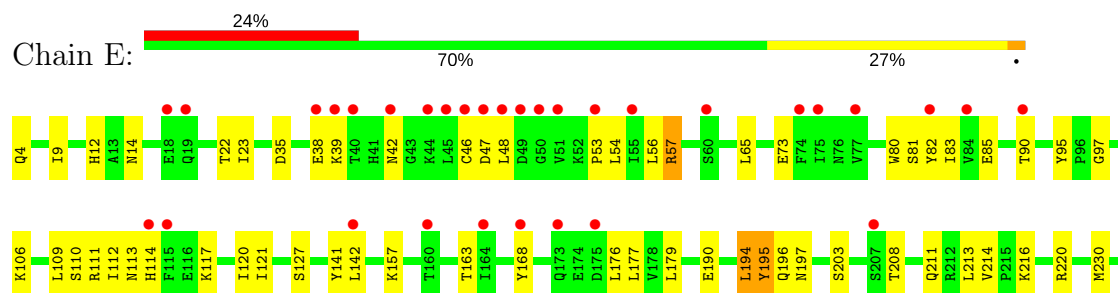
• Molecule 1: Hemagglutinin

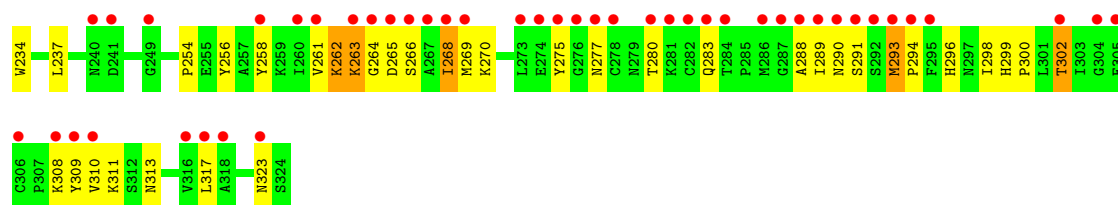


• Molecule 1: Hemagglutinin

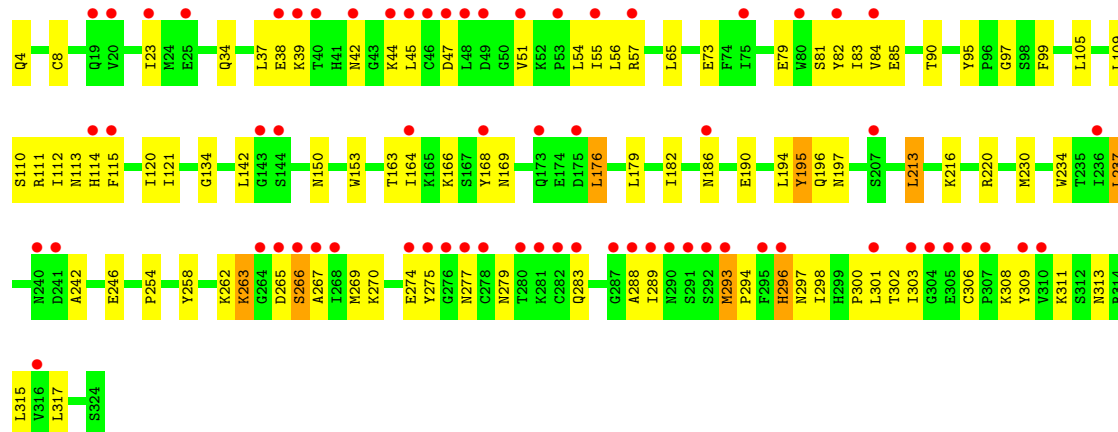


• Molecule 1: Hemagglutinin

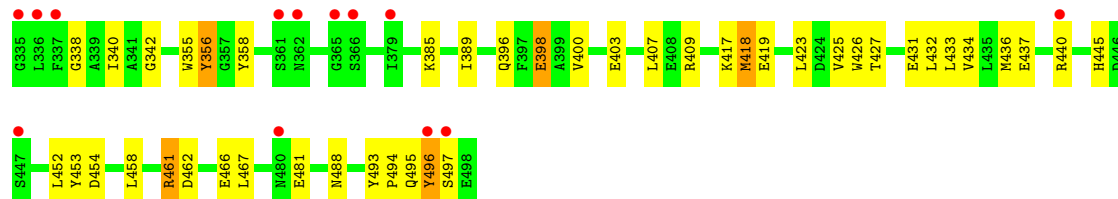
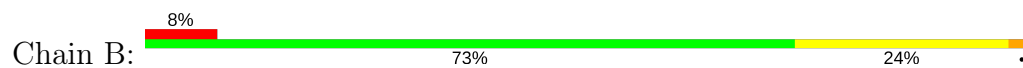




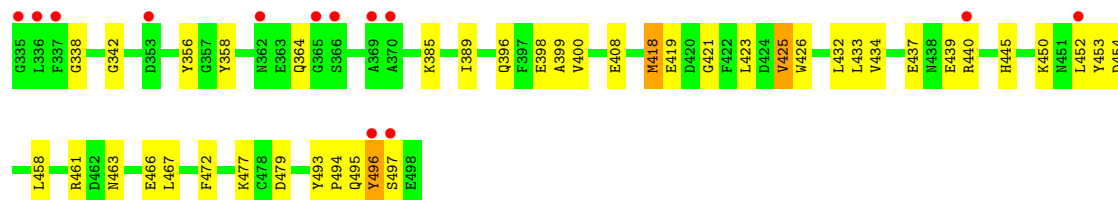
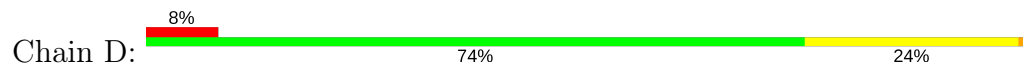
• Molecule 1: Hemagglutinin



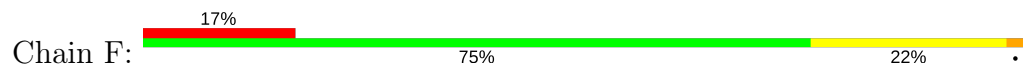
• Molecule 2: Hemagglutinin

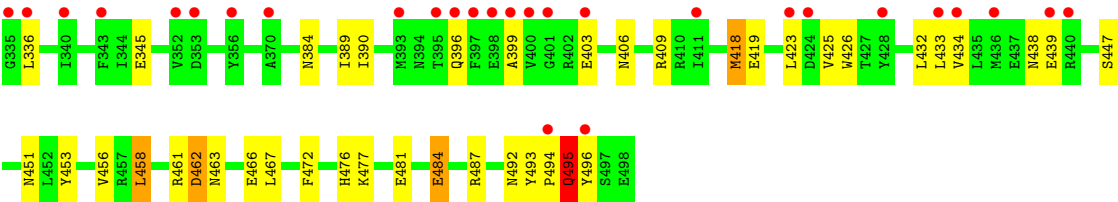


• Molecule 2: Hemagglutinin

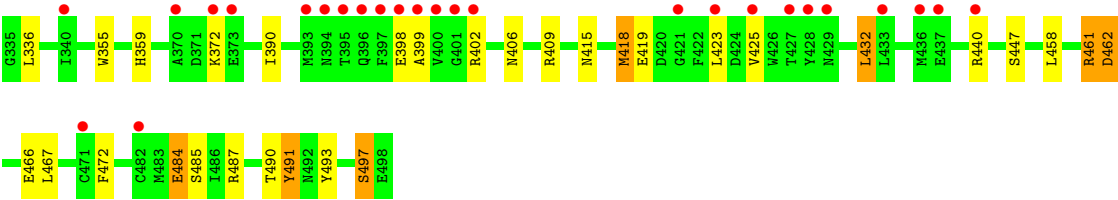
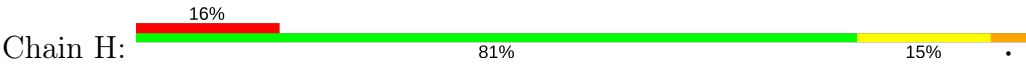


• Molecule 2: Hemagglutinin





● Molecule 2: Hemagglutinin



4 Data and refinement statistics

Property	Value	Source
Space group	P 3	Depositor
Cell constants a, b, c, α , β , γ	70.51Å 70.51Å 489.56Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	48.89 – 2.50 48.89 – 2.50	Depositor EDS
% Data completeness (in resolution range)	91.2 (48.89-2.50) 91.2 (48.89-2.50)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.33 (at 2.51Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, R_{free}	0.235 , 0.279 0.236 , 0.282	Depositor DCC
R_{free} test set	4281 reflections (4.99%)	DCC
Wilson B-factor (Å ²)	37.9	Xtriage
Anisotropy	0.056	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 41.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	0.428 for -h,-k,l 0.095 for h,-h-k,-l 0.095 for -k,-h,-l	Xtriage
Reported twinning fraction	0.410 for -h,-k,l	Depositor
Outliers	0 of 85815 reflections	Xtriage
F_o, F_c correlation	0.89	EDS
Total number of atoms	15877	wwPDB-VP
Average B, all atoms (Å ²)	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.94% 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	0.37	0/2603	0.57	0/3537
1	C	0.37	0/2603	0.58	0/3537
1	E	0.31	0/2603	0.55	0/3537
1	G	0.32	0/2603	0.56	0/3537
2	B	0.37	0/1355	0.54	0/1823
2	D	0.46	1/1355 (0.1%)	0.60	1/1823 (0.1%)
2	F	0.53	2/1355 (0.1%)	0.56	0/1823
2	H	0.35	0/1355	0.54	0/1823
All	All	0.38	3/15832 (0.0%)	0.56	1/21440 (0.0%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	496	TYR	CE1-CZ	-6.71	1.29	1.38
2	F	496	TYR	CE1-CZ	-5.81	1.30	1.38
2	F	496	TYR	CG-CD2	-5.05	1.32	1.39

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	496	TYR	N-CA-C	-5.01	97.46	111.00

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	2541	0	2478	44	0
1	C	2541	0	2478	45	0
1	E	2541	0	2479	67	0
1	G	2541	0	2479	60	0
2	B	1328	0	1231	25	2
2	D	1328	0	1231	25	1
2	F	1328	0	1231	35	1
2	H	1328	0	1231	19	1
3	A	14	0	13	1	0
3	C	14	0	13	0	0
4	A	105	0	0	12	0
4	B	23	0	0	5	0
4	C	120	0	0	12	1
4	D	19	0	0	4	1
4	E	36	0	0	6	0
4	F	24	0	0	4	0
4	G	23	0	0	4	0
4	H	23	0	0	3	0
All	All	15877	0	14864	288	6

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 9.

The worst 5 of 288 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:A:308:LYS:NZ	2:B:398:GLU:OE2	2.08	0.87
2:D:440:ARG:NH1	4:D:515:HOH:O	2.09	0.84
1:G:4:GLN:NE2	4:G:410:HOH:O	2.14	0.81
1:C:120:ILE:O	4:C:762:HOH:O	1.99	0.80
1:E:4:GLN:NE2	4:E:433:HOH:O	2.15	0.78

The worst 5 of 6 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)
2:B:398:GLU:O	2:B:417:LYS:NZ[3_675]	1.79	0.41
2:B:398:GLU:O	2:B:417:LYS:CE[3_675]	1.89	0.31

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:C:761:HOH:O	4:C:764:HOH:O[3_685]	2.00	0.20
2:F:336:LEU:O	2:F:447:SER:OG[2_865]	2.12	0.08
2:D:439:GLU:OE2	4:D:515:HOH:O[2_875]	2.13	0.07

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%)	306 (96%)	12 (4%)	1 (0%)	44	66
1	C	319/321 (99%)	305 (96%)	13 (4%)	1 (0%)	44	66
1	E	319/321 (99%)	300 (94%)	17 (5%)	2 (1%)	28	48
1	G	319/321 (99%)	301 (94%)	17 (5%)	1 (0%)	44	66
2	B	162/164 (99%)	149 (92%)	10 (6%)	3 (2%)	9	15
2	D	162/164 (99%)	151 (93%)	7 (4%)	4 (2%)	6	10
2	F	162/164 (99%)	153 (94%)	7 (4%)	2 (1%)	15	27
2	H	162/164 (99%)	149 (92%)	8 (5%)	5 (3%)	5	6
All	All	1924/1940 (99%)	1814 (94%)	91 (5%)	19 (1%)	18	32

5 of 19 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	196	GLN
2	B	461	ARG
2	B	496	TYR
1	C	196	GLN
2	D	461	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	288/288 (100%)	260 (90%)	28 (10%)	9	18
1	C	288/288 (100%)	269 (93%)	19 (7%)	19	36
1	E	288/288 (100%)	264 (92%)	24 (8%)	13	25
1	G	288/288 (100%)	266 (92%)	22 (8%)	15	29
2	B	140/140 (100%)	129 (92%)	11 (8%)	14	27
2	D	140/140 (100%)	135 (96%)	5 (4%)	40	67
2	F	140/140 (100%)	131 (94%)	9 (6%)	20	38
2	H	140/140 (100%)	131 (94%)	9 (6%)	20	38
All	All	1712/1712 (100%)	1585 (93%)	127 (7%)	16	30

5 of 127 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	322	ARG
1	E	194	LEU
1	G	313	ASN
2	D	425	VAL
1	E	48	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	186	ASN
1	G	4	GLN
1	E	114	HIS
1	C	76	ASN
1	E	4	GLN

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 carbohydrates 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	0.59	0	15,19,21	1.25	1 (6%)
3	NAG	C	601	1	14,14,15	0.53	0	15,19,21	0.92	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	601	1	-	0/6/23/26	0/1/1/1
3	NAG	C	601	1	-	0/6/23/26	0/1/1/1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	601	NAG	C1-O5-C5	3.27	116.67	112.17

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	601	NAG	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, 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%)	0.12	1 (0%) 93 94	2, 15, 41, 64	0
1	C	321/321 (100%)	0.10	4 (1%) 79 80	4, 16, 39, 57	0
1	E	321/321 (100%)	1.19	77 (23%) 1 0	15, 43, 84, 103	0
1	G	321/321 (100%)	1.21	67 (20%) 1 1	17, 44, 83, 100	0
2	B	164/164 (100%)	0.76	13 (7%) 13 13	8, 35, 57, 67	0
2	D	164/164 (100%)	0.74	13 (7%) 13 13	6, 36, 55, 72	0
2	F	164/164 (100%)	1.06	28 (17%) 2 1	19, 39, 71, 96	0
2	H	164/164 (100%)	1.07	26 (15%) 2 2	20, 39, 71, 98	0
All	All	1940/1940 (100%)	0.74	229 (11%) 5 4	2, 32, 73, 103	0

The worst 5 of 229 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	267	ALA	10.8
1	E	82	TYR	10.2
1	G	289	ILE	8.7
2	F	399	ALA	8.5
1	E	46	CYS	8.4

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. 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, 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	LLDF	B-factors(Å ²)	Q<0.9
3	NAG	C	601	14/15	0.85	0.22	0.85	15,20,25,28	0
3	NAG	A	601	14/15	0.85	0.22	-	11,19,28,35	0

6.5 Other polymers [i](#)

There are no such residues in this entry.