



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 12:52 am GMT

PDB ID : 2IBX
Title : Influenza virus (VN1194) H5 HA
Authors : Yamada, S.; Russell, R.J.; Gamblin, S.J.; Skehel, J.J.; Kawaoka, Y.
Deposited on : 2006-09-12
Resolution : 2.80 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

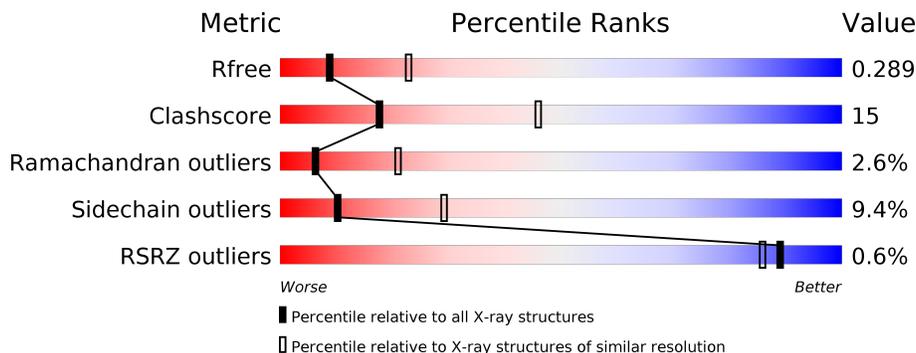
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.80 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	2583 (2.80-2.80)
Clashscore	112137	3033 (2.80-2.80)
Ramachandran outliers	110173	2983 (2.80-2.80)
Sidechain outliers	110143	2985 (2.80-2.80)
RSRZ outliers	101464	2610 (2.80-2.80)

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	340	
1	C	340	
1	E	340	
2	B	160	
2	D	160	
2	F	160	

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 11682 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.

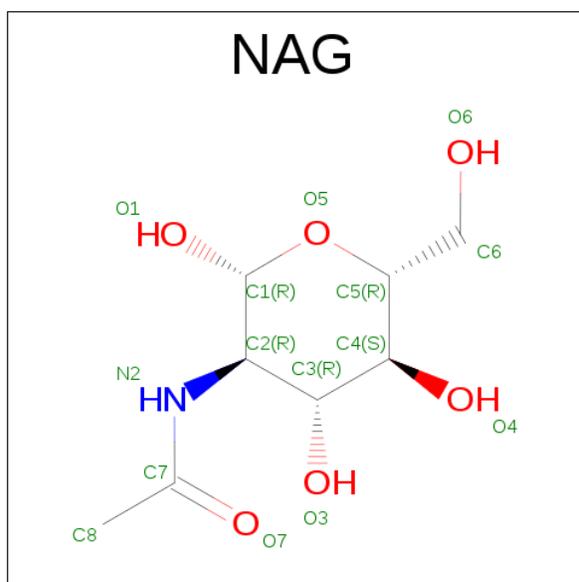
- Molecule 1 is a protein called Hemagglutinin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	321	2547	1609	439	484	15	0	0	0
1	C	321	2547	1609	439	484	15	0	0	0
1	E	321	2547	1609	439	484	15	0	0	0

- Molecule 2 is a protein called Hemagglutinin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	160	1291	805	223	255	8	0	0	0
2	D	160	1291	805	223	255	8	0	0	0
2	F	160	1291	805	223	255	8	0	0	0

- Molecule 3 is SUGAR (N-ACETYL-D-GLUCOSAMINE) (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



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

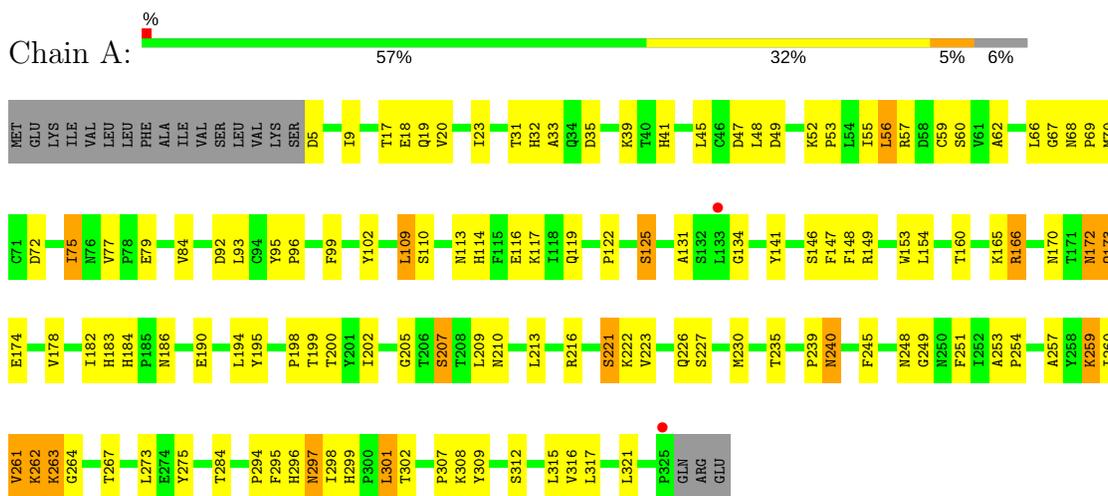
- Molecule 4 is a polymer of unknown type called SUGAR (2-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
4	A	2	Total 28	C 16	N 2	O 10	0	0
4	C	2	Total 28	C 16	N 2	O 10	0	0
4	E	2	Total 28	C 16	N 2	O 10	0	0

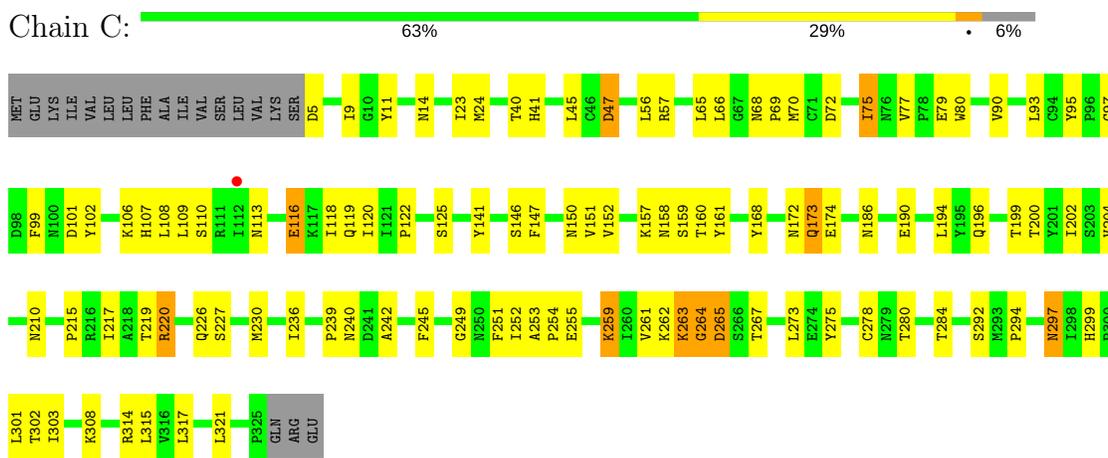
3 Residue-property plots [i](#)

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.

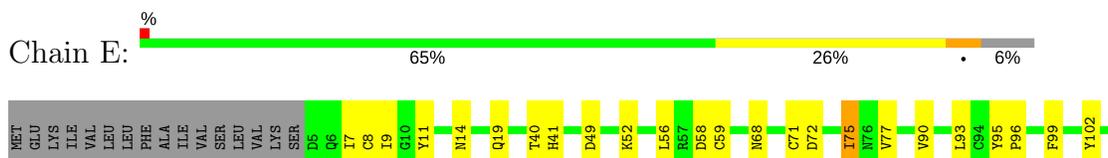
- Molecule 1: Hemagglutinin

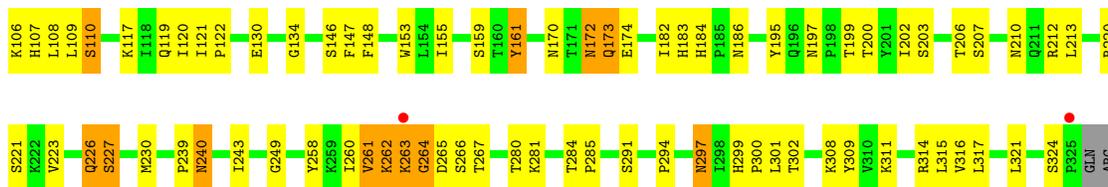


- Molecule 1: Hemagglutinin



- Molecule 1: Hemagglutinin





GLU

• Molecule 2: Hemagglutinin



• Molecule 2: Hemagglutinin



• Molecule 2: Hemagglutinin



4 Data and refinement statistics i

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	174.68Å 100.56Å 160.13Å 90.00° 111.39° 90.00°	Depositor
Resolution (Å)	30.00 – 2.80 19.92 – 2.80	Depositor EDS
% Data completeness (in resolution range)	76.8 (30.00-2.80) 76.4 (19.92-2.80)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.85 (at 2.79Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.241 , 0.302 0.234 , 0.289	Depositor DCC
R_{free} test set	2498 reflections (5.38%)	DCC
Wilson B-factor (Å ²)	60.5	Xtriage
Anisotropy	0.059	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 9.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	0.467 for $1/2^*h-3/2^*k,-1/2^*h-1/2^*k,-1/2^*h+1/2^*k-1$ 0.467 for $1/2^*h+3/2^*k,1/2^*h-1/2^*k,-1/2^*h-1/2^*k-1$	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	11682	wwPDB-VP
Average B, all atoms (Å ²)	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.45% 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	0.62	1/2609 (0.0%)	0.75	0/3545
1	C	0.63	0/2609	0.74	0/3545
1	E	0.61	0/2609	0.75	0/3545
2	B	0.70	0/1317	0.78	0/1772
2	D	0.71	0/1317	0.80	0/1772
2	F	0.72	0/1317	0.78	0/1772
All	All	0.65	1/11778 (0.0%)	0.76	0/15951

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	E	0	1
2	B	0	1
2	D	0	1
All	All	0	3

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	207	SER	CB-OG	5.47	1.49	1.42

There are no bond angle outliers.

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	159	TYR	Peptide
2	D	159	TYR	Peptide
1	E	261	VAL	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	2547	0	2487	101	1
1	C	2547	0	2487	69	1
1	E	2547	0	2487	80	2
2	B	1291	0	1202	45	0
2	D	1291	0	1202	48	0
2	F	1291	0	1202	51	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	0	0
3	F	14	0	13	1	0
4	A	28	0	25	0	0
4	C	28	0	25	0	0
4	E	28	0	25	0	0
All	All	11682	0	11220	351	3

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 15.

The worst 5 of 351 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:E:119:GLN:NE2	1:E:122:PRO:HA	1.46	1.27
1:A:119:GLN:NE2	1:A:122:PRO:HA	1.59	1.17
2:F:150:GLU:O	2:F:154:ASN:HB2	1.52	1.09
1:A:261:VAL:O	1:A:263:LYS:HG2	1.55	1.06
1:A:119:GLN:HE21	1:A:122:PRO:HA	1.10	1.05

All (3) 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)
1:E:173:GLN:NE2	1:E:260:ILE:O[2_556]	1.99	0.21
1:A:260:ILE:O	1:C:173:GLN:NE2[4_456]	2.17	0.03
1:E:172:ASN:OD1	1:E:263:LYS:NZ[2_556]	2.18	0.02

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	319/340 (94%)	277 (87%)	36 (11%)	6 (2%)	9	30
1	C	319/340 (94%)	289 (91%)	21 (7%)	9 (3%)	6	19
1	E	319/340 (94%)	287 (90%)	27 (8%)	5 (2%)	11	36
2	B	158/160 (99%)	142 (90%)	10 (6%)	6 (4%)	4	12
2	D	158/160 (99%)	142 (90%)	11 (7%)	5 (3%)	5	16
2	F	158/160 (99%)	140 (89%)	12 (8%)	6 (4%)	4	12
All	All	1431/1500 (95%)	1277 (89%)	117 (8%)	37 (3%)	6	21

5 of 37 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	240	ASN
1	A	262	LYS
1	A	263	LYS
2	B	62	GLN
2	B	127	ARG

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	289/307 (94%)	256 (89%)	33 (11%)	7	20
1	C	289/307 (94%)	258 (89%)	31 (11%)	8	22
1	E	289/307 (94%)	263 (91%)	26 (9%)	11	32
2	B	136/136 (100%)	124 (91%)	12 (9%)	12	33
2	D	136/136 (100%)	125 (92%)	11 (8%)	14	37
2	F	136/136 (100%)	129 (95%)	7 (5%)	28	61
All	All	1275/1329 (96%)	1155 (91%)	120 (9%)	10	29

5 of 120 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	108	LEU
1	C	267	THR
1	E	297	ASN
1	C	110	SER
1	C	199	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 34 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	119	GLN
1	C	299	HIS
2	F	62	GLN
1	C	186	ASN
1	A	186	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

6 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
4	NAG	A	503	1,4	14,14,15	0.53	0	15,19,21	1.15	1 (6%)
4	NAG	A	504	4	14,14,15	0.49	0	15,19,21	1.83	3 (20%)
4	NAG	C	510	1,4	14,14,15	0.46	0	15,19,21	1.43	3 (20%)
4	NAG	C	511	4	14,14,15	0.53	0	15,19,21	2.07	2 (13%)
4	NAG	E	517	1,4	14,14,15	0.58	0	15,19,21	1.32	1 (6%)
4	NAG	E	518	4	14,14,15	0.53	0	15,19,21	1.35	3 (20%)

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	503	1,4	-	0/6/23/26	0/1/1/1
4	NAG	A	504	4	-	0/6/23/26	0/1/1/1
4	NAG	C	510	1,4	-	0/6/23/26	0/1/1/1
4	NAG	C	511	4	-	0/6/23/26	0/1/1/1
4	NAG	E	517	1,4	-	0/6/23/26	0/1/1/1
4	NAG	E	518	4	-	0/6/23/26	0/1/1/1

There are no bond length outliers.

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	E	517	NAG	O4-C4-C3	-3.47	102.81	110.36
4	A	503	NAG	O4-C4-C3	-2.72	104.44	110.36
4	C	510	NAG	O5-C1-C2	-2.56	107.91	111.47
4	E	518	NAG	C1-O5-C5	2.06	115.01	112.17
4	C	510	NAG	O3-C3-C2	2.11	113.91	109.39

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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	502	1	14,14,15	0.57	0	15,19,21	0.92	0
3	NAG	B	507	-	14,14,15	0.59	0	15,19,21	1.79	3 (20%)
3	NAG	C	509	1	14,14,15	0.41	0	15,19,21	1.52	2 (13%)
3	NAG	D	514	-	14,14,15	0.65	0	15,19,21	1.22	2 (13%)
3	NAG	E	516	1	14,14,15	0.53	0	15,19,21	1.33	2 (13%)
3	NAG	F	521	-	14,14,15	0.45	0	15,19,21	1.83	4 (26%)

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	502	1	-	0/6/23/26	0/1/1/1
3	NAG	B	507	-	-	0/6/23/26	0/1/1/1
3	NAG	C	509	1	-	0/6/23/26	0/1/1/1
3	NAG	D	514	-	-	0/6/23/26	0/1/1/1
3	NAG	E	516	1	-	0/6/23/26	0/1/1/1
3	NAG	F	521	-	-	2/6/23/26	0/1/1/1

There are no bond length outliers.

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	F	521	NAG	O5-C1-C2	-5.22	104.21	111.47
3	B	507	NAG	O5-C1-C2	-4.00	105.90	111.47
3	D	514	NAG	O5-C1-C2	-3.07	107.20	111.47
3	F	521	NAG	C4-C3-C2	-2.55	107.28	111.02
3	C	509	NAG	C2-N2-C7	-2.27	119.63	122.94

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	F	521	NAG	C8-C7-N2-C2
3	F	521	NAG	O7-C7-N2-C2

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	F	521	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	321/340 (94%)	-0.06	2 (0%) 89 86	29, 58, 70, 74	0
1	C	321/340 (94%)	0.01	1 (0%) 93 92	31, 56, 72, 77	0
1	E	321/340 (94%)	-0.06	2 (0%) 89 86	32, 55, 70, 74	0
2	B	160/160 (100%)	0.07	1 (0%) 89 86	28, 48, 71, 93	0
2	D	160/160 (100%)	0.17	1 (0%) 89 86	27, 47, 72, 99	0
2	F	160/160 (100%)	0.09	2 (1%) 77 71	25, 48, 74, 97	0
All	All	1443/1500 (96%)	0.01	9 (0%) 89 86	25, 54, 71, 99	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	160	PRO	6.7
2	F	160	PRO	5.8
2	D	160	PRO	4.2
1	E	263	LYS	3.1
1	A	325	PRO	2.7

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](#)

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(\AA^2)	Q<0.9
4	NAG	C	510	14/15	0.96	0.15	-0.26	71,73,77,82	0
4	NAG	E	517	14/15	0.96	0.15	-0.54	72,74,77,82	0
4	NAG	A	503	14/15	0.94	0.14	-1.21	74,76,78,82	0
4	NAG	E	518	14/15	0.85	0.26	-	86,89,91,92	0
4	NAG	A	504	14/15	0.81	0.38	-	86,88,89,90	0
4	NAG	C	511	14/15	0.84	0.23	-	86,88,89,89	0

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(\AA^2)	Q<0.9
3	NAG	A	502	14/15	0.87	0.14	-	64,70,72,73	0
3	NAG	E	516	14/15	0.92	0.13	-	62,65,67,68	0
3	NAG	D	514	14/15	0.70	0.46	-	105,106,107,107	0
3	NAG	B	507	14/15	0.80	0.28	-	96,98,99,99	0
3	NAG	F	521	14/15	0.79	0.37	-	99,100,102,102	0
3	NAG	C	509	14/15	0.91	0.19	-	65,70,73,74	0

6.5 Other polymers [i](#)

There are no such residues in this entry.