



wwPDB X-ray Structure Validation Summary Report ⓘ

Apr 20, 2021 – 08:37 AM JST

PDB ID : 7BWU
Title : Restructuring hemagglutinin-neuraminidase (HN) of Newcastle disease virus produced from *Oryza sativa*
Authors : Ma, F.S.; Zhang, G.P.
Deposited on : 2020-04-16
Resolution : 1.87 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

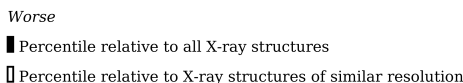
MolProbity : 4.02b-467
Mogul : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.18
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac : 5.8.0158
CCP4 : 7.0.044 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.18

i

X-RAY DIFFRACTION

A.

Metric	Percentile Banks	Value
--------	------------------	-------



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R _{free}	130704	9470 (1.90-1.86)
Clashscore	141614	10282 (1.90-1.86)
Ramachandran outliers	138981	10152 (1.90-1.86)
Sidechain outliers	138945	10152 (1.90-1.86)
RSRZ outliers	127900	9303 (1.90-1.86)

Mol	Chain	Length	Quality of chain
1	A	536	<div> <div>2%</div> <div>79%</div> <div>17%</div> </div>
1	B	536	<div> <div>3%</div> <div>79%</div> <div>17%</div> </div>
1	C	536	<div> <div>3%</div> <div>78%</div> <div>5%</div> <div>17%</div> </div>
1	D	536	<div> <div>6%</div> <div>77%</div> <div>6%</div> <div>17%</div> </div>

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 14942 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-neuraminidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	444	Total	C	N	O	S	0	0	0
			3420	2161	584	656	19			
1	D	447	Total	C	N	O	S	0	0	0
			3443	2172	592	659	20			
1	A	443	Total	C	N	O	S	0	0	0
			3428	2164	589	656	19			
1	C	445	Total	C	N	O	S	0	0	0
			3436	2168	589	660	19			

There are 64 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	443	THR	MET	engineered mutation	UNP I7EFR1
B	572	GLY	-	linker	UNP I7EFR1
B	573	GLY	-	linker	UNP I7EFR1
B	574	GLY	-	linker	UNP I7EFR1
B	575	GLY	-	linker	UNP I7EFR1
B	576	SER	-	linker	UNP I7EFR1
B	577	GLY	-	linker	UNP I7EFR1
B	578	GLY	-	linker	UNP I7EFR1
B	579	GLY	-	linker	UNP I7EFR1
B	580	GLY	-	linker	UNP I7EFR1
B	581	SER	-	linker	UNP I7EFR1
B	582	GLY	-	linker	UNP I7EFR1
B	583	GLY	-	linker	UNP I7EFR1
B	584	GLY	-	linker	UNP I7EFR1
B	585	GLY	-	linker	UNP I7EFR1
B	586	SER	-	linker	UNP I7EFR1
D	443	THR	MET	engineered mutation	UNP I7EFR1
D	572	GLY	-	linker	UNP I7EFR1
D	573	GLY	-	linker	UNP I7EFR1
D	574	GLY	-	linker	UNP I7EFR1
D	575	GLY	-	linker	UNP I7EFR1

Continued on next page...

Continued from previous page...

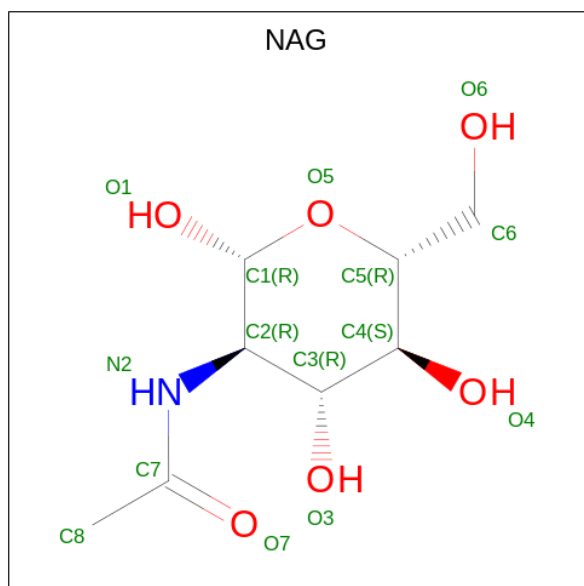
Chain	Residue	Modelled	Actual	Comment	Reference
D	576	SER	-	linker	UNP I7EFR1
D	577	GLY	-	linker	UNP I7EFR1
D	578	GLY	-	linker	UNP I7EFR1
D	579	GLY	-	linker	UNP I7EFR1
D	580	GLY	-	linker	UNP I7EFR1
D	581	SER	-	linker	UNP I7EFR1
D	582	GLY	-	linker	UNP I7EFR1
D	583	GLY	-	linker	UNP I7EFR1
D	584	GLY	-	linker	UNP I7EFR1
D	585	GLY	-	linker	UNP I7EFR1
D	586	SER	-	linker	UNP I7EFR1
A	443	THR	MET	engineered mutation	UNP I7EFR1
A	572	GLY	-	linker	UNP I7EFR1
A	573	GLY	-	linker	UNP I7EFR1
A	574	GLY	-	linker	UNP I7EFR1
A	575	GLY	-	linker	UNP I7EFR1
A	576	SER	-	linker	UNP I7EFR1
A	577	GLY	-	linker	UNP I7EFR1
A	578	GLY	-	linker	UNP I7EFR1
A	579	GLY	-	linker	UNP I7EFR1
A	580	GLY	-	linker	UNP I7EFR1
A	581	SER	-	linker	UNP I7EFR1
A	582	GLY	-	linker	UNP I7EFR1
A	583	GLY	-	linker	UNP I7EFR1
A	584	GLY	-	linker	UNP I7EFR1
A	585	GLY	-	linker	UNP I7EFR1
A	586	SER	-	linker	UNP I7EFR1
C	443	THR	MET	engineered mutation	UNP I7EFR1
C	572	GLY	-	linker	UNP I7EFR1
C	573	GLY	-	linker	UNP I7EFR1
C	574	GLY	-	linker	UNP I7EFR1
C	575	GLY	-	linker	UNP I7EFR1
C	576	SER	-	linker	UNP I7EFR1
C	577	GLY	-	linker	UNP I7EFR1
C	578	GLY	-	linker	UNP I7EFR1
C	579	GLY	-	linker	UNP I7EFR1
C	580	GLY	-	linker	UNP I7EFR1
C	581	SER	-	linker	UNP I7EFR1
C	582	GLY	-	linker	UNP I7EFR1
C	583	GLY	-	linker	UNP I7EFR1
C	584	GLY	-	linker	UNP I7EFR1
C	585	GLY	-	linker	UNP I7EFR1

Continued on next page...

Continued from previous page...

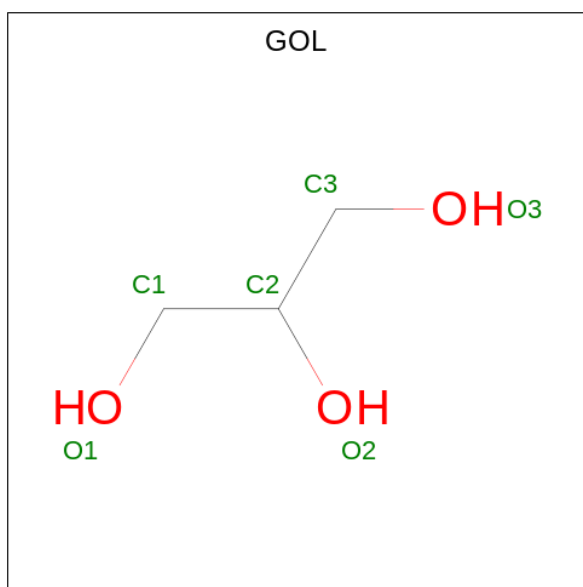
Chain	Residue	Modelled	Actual	Comment	Reference
C	586	SER	-	linker	UNP I7EFR1

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



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

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	B	1	Total	C	O	0	0
			6	3	3		
3	D	1	Total	C	O	0	0
			6	3	3		
3	A	1	Total	C	O	0	0
			6	3	3		
3	C	1	Total	C	O	0	0
			6	3	3		

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Ca	0	0
			1	1		
4	D	1	Total	Ca	0	0
			1	1		
4	A	1	Total	Ca	0	0
			1	1		
4	C	1	Total	Ca	0	0
			1	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	311	Total	O	0	0
			311	311		

Continued on next page...

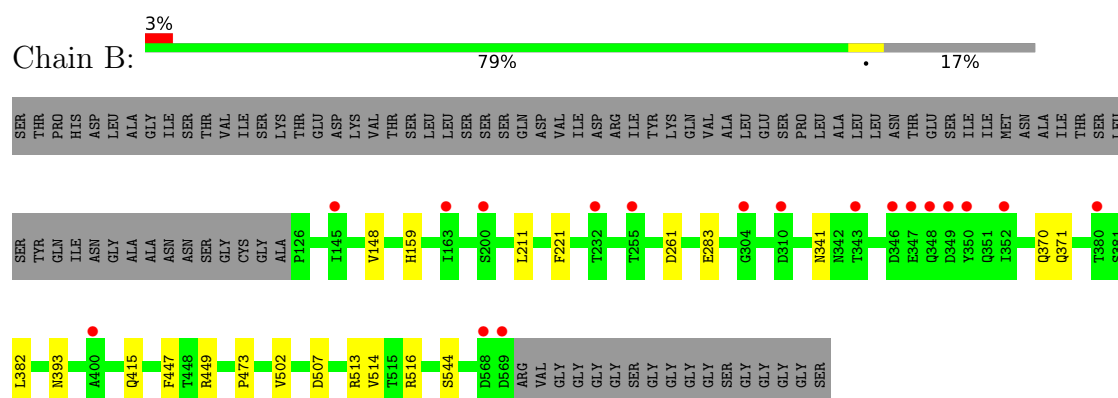
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	D	262	Total 262	O 262	0	0
5	A	311	Total 311	O 311	0	0
5	C	247	Total 247	O 247	0	0

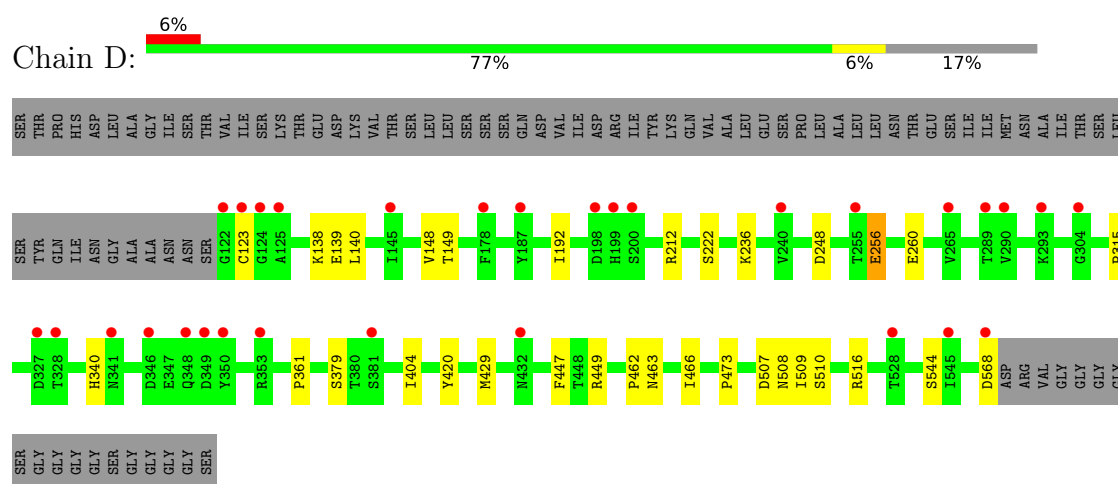
3 Residue-property plots [i](#)

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.

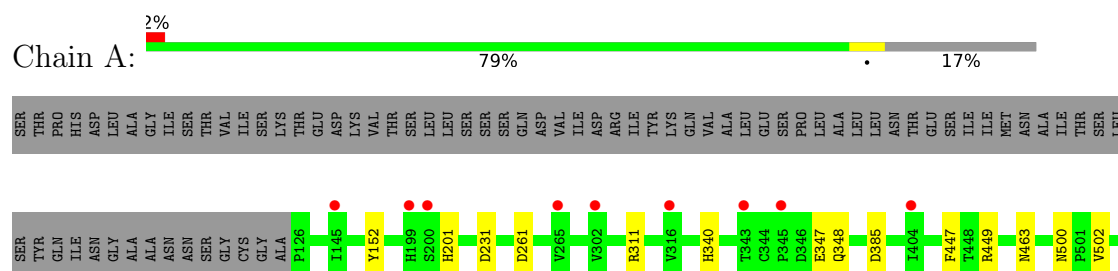
• Molecule 1: Hemagglutinin-neuraminidase



• Molecule 1: Hemagglutinin-neuraminidase



• Molecule 1: Hemagglutinin-neuraminidase





4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	83.88Å 142.07Å 191.21Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.86 – 1.87 39.86 – 1.87	Depositor EDS
% Data completeness (in resolution range)	100.0 (39.86-1.87) 99.9 (39.86-1.87)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.77 (at 1.87Å)	Xtriage
Refinement program	PHENIX 1.11.1 _2575	Depositor
R, R_{free}	0.167 , 0.195 0.167 , 0.195	Depositor DCC
R_{free} test set	9289 reflections (4.93%)	wwPDB-VP
Wilson B-factor (Å ²)	29.6	Xtriage
Anisotropy	0.303	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.38 , 51.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.44$, $\langle L^2 \rangle = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	14942	wwPDB-VP
Average B, all atoms (Å ²)	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 37.37 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 4.3324e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

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: CA, NAG, GOL

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	0.40	0/3514	0.59	0/4781
1	B	0.38	0/3505	0.58	0/4770
1	C	0.37	0/3521	0.57	0/4792
1	D	0.38	0/3529	0.58	0/4802
All	All	0.38	0/14069	0.58	0/19145

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	3428	0	3334	10	0
1	B	3420	0	3318	15	0
1	C	3436	0	3337	14	0
1	D	3443	0	3343	19	0
2	A	14	0	13	3	0
2	B	14	0	13	0	0
2	C	14	0	13	0	0
2	D	14	0	13	0	0
3	A	6	0	8	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	6	0	8	0	0
3	C	6	0	8	0	0
3	D	6	0	8	0	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
4	C	1	0	0	0	0
4	D	1	0	0	0	0
5	A	311	0	0	5	1
5	B	311	0	0	6	1
5	C	247	0	0	3	0
5	D	262	0	0	6	0
All	All	14942	0	13416	61	1

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 2.

The worst 5 of 61 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:B:513:ARG:NH1	5:B:1301:HOH:O	1.93	1.02
1:C:128:HIS:HD2	1:C:211:LEU:H	1.20	0.90
2:A:1201:NAG:O3	5:A:1301:HOH:O	1.93	0.84
1:C:333:LYS:HD2	1:C:394:THR:HG23	1.63	0.81
1:B:371:GLN:HE21	1:B:415:GLN:HE21	1.35	0.73

All (1) 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)
5:B:1549:HOH:O	5:A:1467:HOH:O[1_655]	2.07	0.13

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	441/536 (82%)	419 (95%)	22 (5%)	0	100	100
1	B	442/536 (82%)	419 (95%)	22 (5%)	1 (0%)	47	37
1	C	443/536 (83%)	418 (94%)	25 (6%)	0	100	100
1	D	445/536 (83%)	422 (95%)	22 (5%)	1 (0%)	47	37
All	All	1771/2144 (83%)	1678 (95%)	91 (5%)	2 (0%)	51	41

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	473	PRO
1	D	473	PRO

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	386/459 (84%)	381 (99%)	5 (1%)	69	64
1	B	383/459 (83%)	379 (99%)	4 (1%)	76	73
1	C	386/459 (84%)	381 (99%)	5 (1%)	69	64
1	D	386/459 (84%)	377 (98%)	9 (2%)	50	41
All	All	1541/1836 (84%)	1518 (98%)	23 (2%)	65	59

5 of 23 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	447	PHE
1	A	544	SER
1	A	463	ASN
1	C	248	ASP
1	D	256	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such

sidechains are listed below:

Mol	Chain	Res	Type
1	B	370	GLN
1	B	371	GLN
1	A	201	HIS
1	C	128	HIS
1	C	415	GLN

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](#)

Of 12 ligands modelled in this entry, 4 are monoatomic - leaving 8 for Mogul analysis.

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	GOL	A	1202	-	5,5,5	0.30	0	5,5,5	0.39	0
2	NAG	D	1201	1	14,14,15	0.33	0	17,19,21	0.74	1 (5%)
3	GOL	D	1202	-	5,5,5	0.38	0	5,5,5	0.49	0
3	GOL	B	1202	-	5,5,5	0.24	0	5,5,5	0.49	0
2	NAG	C	1201	1	14,14,15	0.36	0	17,19,21	1.18	1 (5%)
3	GOL	C	1202	-	5,5,5	0.32	0	5,5,5	0.51	0
2	NAG	B	1201	1	14,14,15	0.45	0	17,19,21	0.58	0
2	NAG	A	1201	1	14,14,15	0.44	0	17,19,21	0.70	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	GOL	A	1202	-	-	0/4/4/4	-
2	NAG	D	1201	1	-	0/6/23/26	0/1/1/1
3	GOL	D	1202	-	-	3/4/4/4	-
3	GOL	B	1202	-	-	0/4/4/4	-
2	NAG	C	1201	1	-	2/6/23/26	0/1/1/1
3	GOL	C	1202	-	-	2/4/4/4	-
2	NAG	B	1201	1	-	2/6/23/26	0/1/1/1
2	NAG	A	1201	1	-	2/6/23/26	0/1/1/1

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	1201	NAG	C1-O5-C5	4.45	118.22	112.19
2	D	1201	NAG	C1-O5-C5	2.29	115.30	112.19
2	A	1201	NAG	C1-O5-C5	2.05	114.97	112.19

There are no chirality outliers.

5 of 11 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	D	1202	GOL	O1-C1-C2-C3
3	C	1202	GOL	O1-C1-C2-C3
2	C	1201	NAG	O5-C5-C6-O6
2	B	1201	NAG	C4-C5-C6-O6
2	A	1201	NAG	O5-C5-C6-O6

There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1201	NAG	3	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	443/536 (82%)	0.08	11 (2%) 57 59	22, 33, 54, 88	0
1	B	444/536 (82%)	0.07	18 (4%) 37 39	22, 34, 53, 84	0
1	C	445/536 (83%)	0.20	18 (4%) 38 39	23, 37, 62, 80	0
1	D	447/536 (83%)	0.32	30 (6%) 17 19	22, 38, 60, 88	0
All	All	1779/2144 (82%)	0.17	77 (4%) 35 36	22, 35, 59, 88	0

The worst 5 of 77 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	123	CYS	9.2
1	D	124	GLY	6.9
1	D	145	ILE	6.2
1	C	569	ASP	5.6
1	C	145	ILE	5.3

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, 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(\AA^2)	Q<0.9
2	NAG	C	1201	14/15	0.83	0.27	49,65,69,71	0
3	GOL	D	1202	6/6	0.87	0.21	48,54,55,55	0
2	NAG	B	1201	14/15	0.88	0.20	44,59,66,71	0
2	NAG	D	1201	14/15	0.90	0.27	48,63,71,76	0
3	GOL	C	1202	6/6	0.91	0.17	43,55,60,62	0
4	CA	C	1203	1/1	0.91	0.14	77,77,77,77	0
2	NAG	A	1201	14/15	0.92	0.23	39,55,66,73	0
3	GOL	A	1202	6/6	0.92	0.12	32,36,42,42	0
3	GOL	B	1202	6/6	0.93	0.13	37,41,42,46	0
4	CA	A	1203	1/1	0.96	0.19	61,61,61,61	0
4	CA	D	1203	1/1	0.96	0.17	67,67,67,67	0
4	CA	B	1203	1/1	0.97	0.10	62,62,62,62	0

6.5 Other polymers [i](#)

There are no such residues in this entry.