



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 05:05 am GMT

PDB ID : 4WSR  
Title : The crystal structure of hemagglutinin form A/chicken/New York/14677-13/1998  
Authors : Yang, H.; Carney, P.J.; Chang, J.C.; Villanueva, J.M.; Stevens, J.  
Deposited on : 2014-10-28  
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](mailto: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

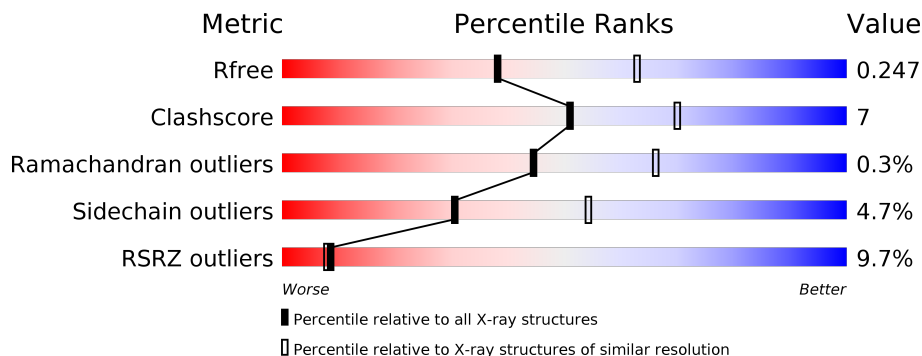
# 1 Overall quality at a glance

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  $\geq 3$ , 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	514	<div> <div>6%</div> <div>81% 11% • 6%</div> </div>
1	B	514	<div> <div>8%</div> <div>78% 13% • 6%</div> </div>
1	C	514	<div> <div>6%</div> <div>83% 9% • 6%</div> </div>
1	D	514	<div> <div>9%</div> <div>80% 16% • •</div> </div>
1	E	514	<div> <div>8%</div> <div>79% 14% • • 5%</div> </div>
1	F	514	<div> <div>17%</div> <div>79% 11% • 6%</div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	NAG	A	601	X	-	-	-
2	NAG	A	602	-	-	-	X
2	NAG	B	601	X	-	-	-
2	NAG	B	602	-	-	-	X
2	NAG	C	605	X	-	-	-
2	NAG	D	601	X	-	-	-
2	NAG	D	602	X	-	-	-
2	NAG	E	601	X	-	-	-

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 24430 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	483	Total	C	N	O	S	0	0	0
			3841	2410	669	743	19			
1	B	482	Total	C	N	O	S	0	0	0
			3832	2405	667	741	19			
1	C	482	Total	C	N	O	S	0	0	0
			3832	2405	667	741	19			
1	E	489	Total	C	N	O	S	0	0	0
			3886	2441	675	751	19			
1	D	498	Total	C	N	O	S	0	0	0
			3952	2484	688	761	19			
1	F	481	Total	C	N	O	S	0	0	0
			3825	2400	666	740	19			

There are 66 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	ALA	-	expression tag	UNP Q0A3A5
A	-2	ASP	-	expression tag	UNP Q0A3A5
A	-1	PRO	-	expression tag	UNP Q0A3A5
A	0	GLY	-	expression tag	UNP Q0A3A5
A	317	THR	UNK	conflict	UNP Q0A3A5
A	504	GLY	VAL	conflict	UNP Q0A3A5
A	505	ARG	MET	conflict	UNP Q0A3A5
A	507	VAL	-	expression tag	UNP Q0A3A5
A	508	PRO	-	expression tag	UNP Q0A3A5
A	509	ARG	-	expression tag	UNP Q0A3A5
A	510	GLY	-	expression tag	UNP Q0A3A5
B	-3	ALA	-	expression tag	UNP Q0A3A5
B	-2	ASP	-	expression tag	UNP Q0A3A5
B	-1	PRO	-	expression tag	UNP Q0A3A5
B	0	GLY	-	expression tag	UNP Q0A3A5
B	317	THR	UNK	conflict	UNP Q0A3A5
B	504	GLY	VAL	conflict	UNP Q0A3A5

*Continued on next page...*

*Continued from previous page...*

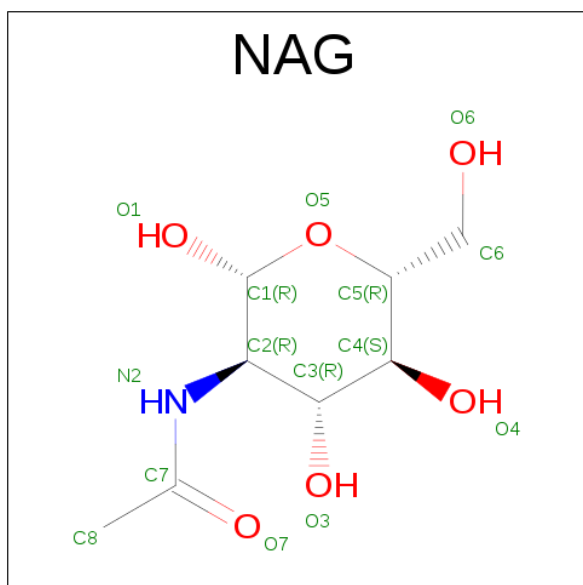
Chain	Residue	Modelled	Actual	Comment	Reference
B	505	ARG	MET	conflict	UNP Q0A3A5
B	507	VAL	-	expression tag	UNP Q0A3A5
B	508	PRO	-	expression tag	UNP Q0A3A5
B	509	ARG	-	expression tag	UNP Q0A3A5
B	510	GLY	-	expression tag	UNP Q0A3A5
C	-3	ALA	-	expression tag	UNP Q0A3A5
C	-2	ASP	-	expression tag	UNP Q0A3A5
C	-1	PRO	-	expression tag	UNP Q0A3A5
C	0	GLY	-	expression tag	UNP Q0A3A5
C	317	THR	UNK	conflict	UNP Q0A3A5
C	504	GLY	VAL	conflict	UNP Q0A3A5
C	505	ARG	MET	conflict	UNP Q0A3A5
C	507	VAL	-	expression tag	UNP Q0A3A5
C	508	PRO	-	expression tag	UNP Q0A3A5
C	509	ARG	-	expression tag	UNP Q0A3A5
C	510	GLY	-	expression tag	UNP Q0A3A5
E	-3	ALA	-	expression tag	UNP Q0A3A5
E	-2	ASP	-	expression tag	UNP Q0A3A5
E	-1	PRO	-	expression tag	UNP Q0A3A5
E	0	GLY	-	expression tag	UNP Q0A3A5
E	317	THR	UNK	conflict	UNP Q0A3A5
E	504	GLY	VAL	conflict	UNP Q0A3A5
E	505	ARG	MET	conflict	UNP Q0A3A5
E	507	VAL	-	expression tag	UNP Q0A3A5
E	508	PRO	-	expression tag	UNP Q0A3A5
E	509	ARG	-	expression tag	UNP Q0A3A5
E	510	GLY	-	expression tag	UNP Q0A3A5
D	-3	ALA	-	expression tag	UNP Q0A3A5
D	-2	ASP	-	expression tag	UNP Q0A3A5
D	-1	PRO	-	expression tag	UNP Q0A3A5
D	0	GLY	-	expression tag	UNP Q0A3A5
D	317	THR	UNK	conflict	UNP Q0A3A5
D	504	GLY	VAL	conflict	UNP Q0A3A5
D	505	ARG	MET	conflict	UNP Q0A3A5
D	507	VAL	-	expression tag	UNP Q0A3A5
D	508	PRO	-	expression tag	UNP Q0A3A5
D	509	ARG	-	expression tag	UNP Q0A3A5
D	510	GLY	-	expression tag	UNP Q0A3A5
F	-3	ALA	-	expression tag	UNP Q0A3A5
F	-2	ASP	-	expression tag	UNP Q0A3A5
F	-1	PRO	-	expression tag	UNP Q0A3A5
F	0	GLY	-	expression tag	UNP Q0A3A5

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	317	THR	UNK	conflict	UNP Q0A3A5
F	504	GLY	VAL	conflict	UNP Q0A3A5
F	505	ARG	MET	conflict	UNP Q0A3A5
F	507	VAL	-	expression tag	UNP Q0A3A5
F	508	PRO	-	expression tag	UNP Q0A3A5
F	509	ARG	-	expression tag	UNP Q0A3A5
F	510	GLY	-	expression tag	UNP Q0A3A5

- Molecule 2 is N-ACETYL-D-GLUCOSAMINE (three-letter code: NAG) (formula:  $C_8H_{15}NO_6$ ).



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

Continued on next page...

*Continued from previous page...*

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

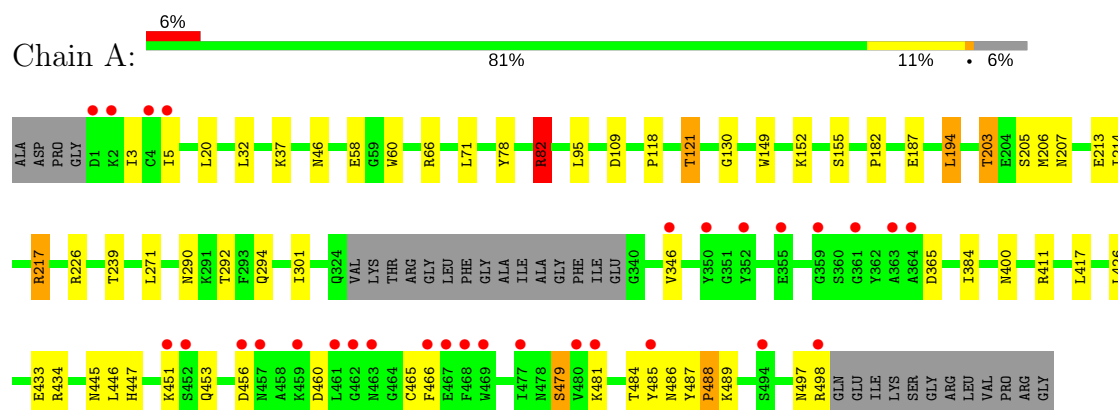
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	226	Total	O	0	0
			226	226		
3	B	219	Total	O	0	0
			219	219		
3	C	175	Total	O	0	0
			175	175		
3	E	85	Total	O	0	0
			85	85		
3	D	212	Total	O	0	0
			212	212		
3	F	107	Total	O	0	0
			107	107		

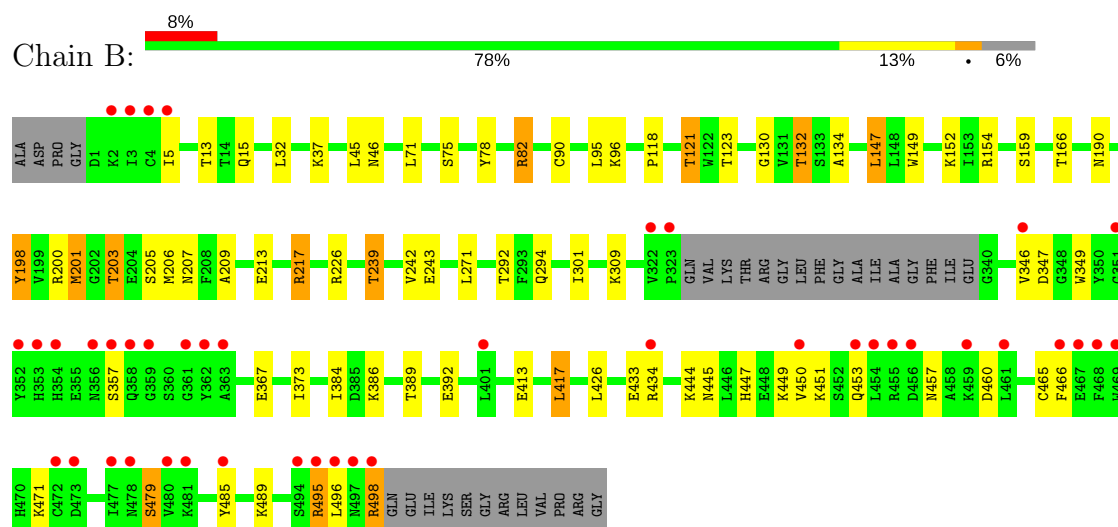
### 3 Residue-property plots [i](#)

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.

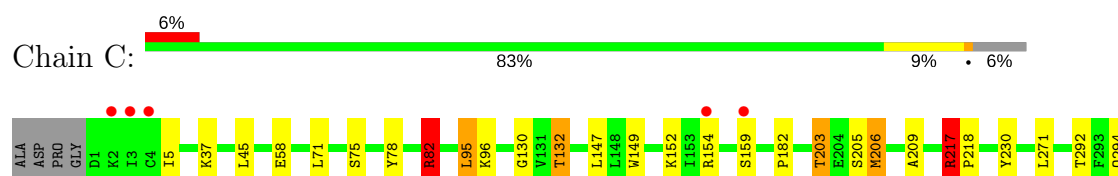
#### • Molecule 1: Hemagglutinin



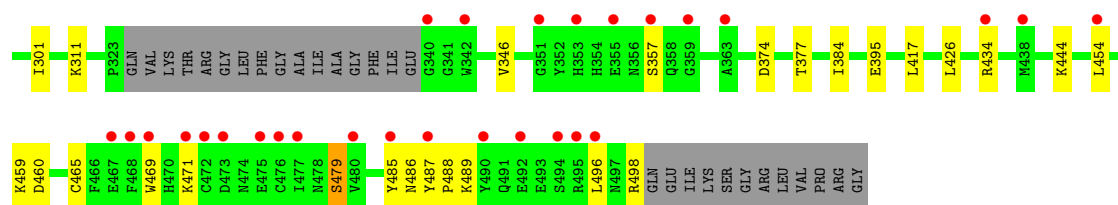
#### • Molecule 1: Hemagglutinin



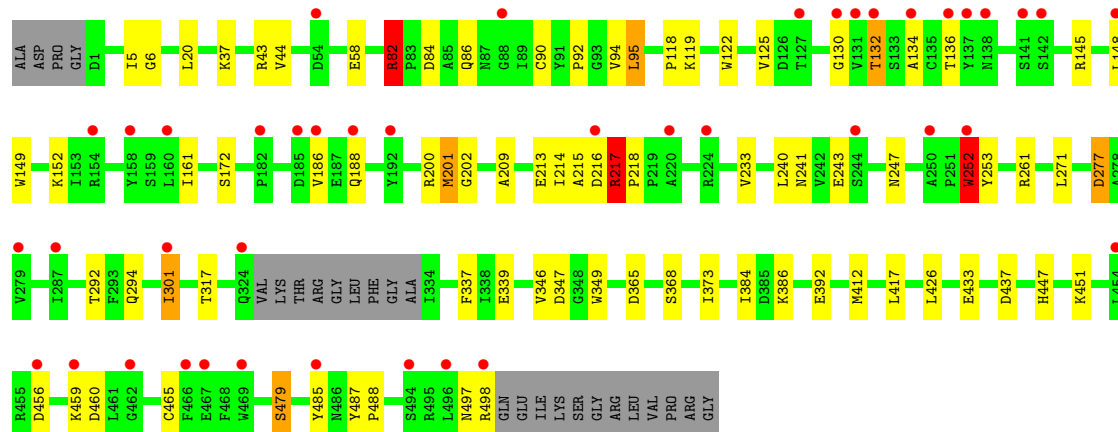
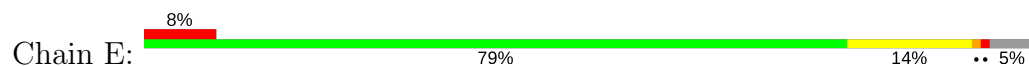
#### • Molecule 1: Hemagglutinin



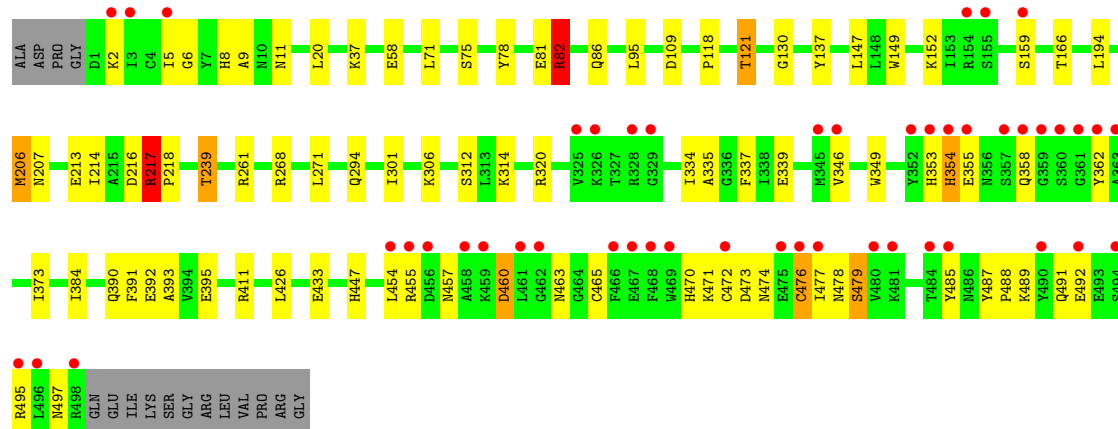
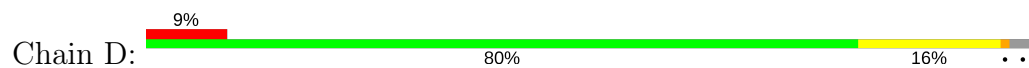




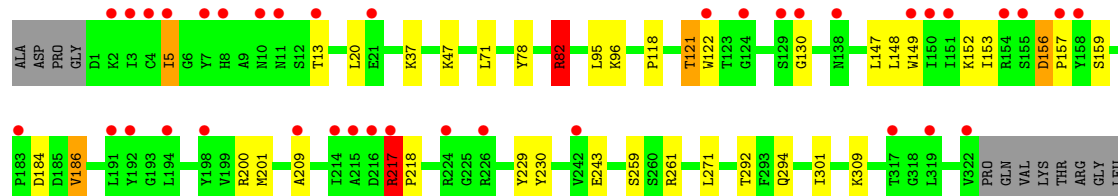
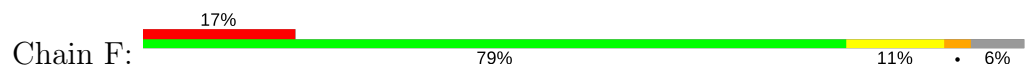
### • Molecule 1: Hemagglutinin

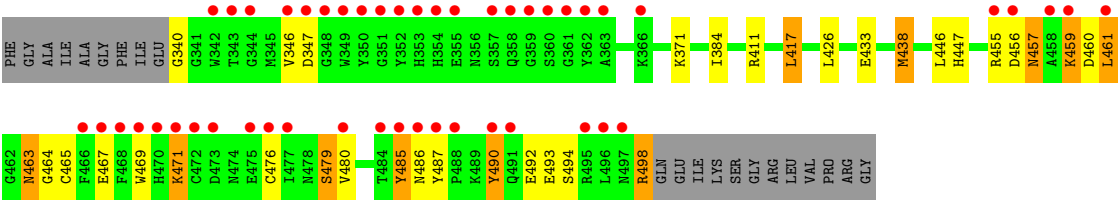


### • Molecule 1: Hemagglutinin



### • Molecule 1: Hemagglutinin





## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	252.54Å 134.82Å 123.09Å 90.00° 113.23° 90.00°	Depositor
Resolution (Å)	48.00 – 2.50 47.29 – 2.50	Depositor EDS
% Data completeness (in resolution range)	98.9 (48.00-2.50) 98.9 (47.29-2.50)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	5.65 (at 2.51Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, $R_{free}$	0.210 , 0.246 0.213 , 0.247	Depositor DCC
$R_{free}$ test set	6504 reflections (5.29%)	DCC
Wilson B-factor (Å <sup>2</sup> )	49.1	Xtriage
Anisotropy	0.338	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 39.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	24430	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	67.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.95% 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

### 5.1 Standard geometry

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	0.73	2/3928 (0.1%)	0.80	5/5323 (0.1%)
1	B	0.70	0/3919	0.83	9/5311 (0.2%)
1	C	0.70	0/3919	0.81	6/5311 (0.1%)
1	D	0.68	1/4042 (0.0%)	0.82	8/5477 (0.1%)
1	E	0.64	0/3974	0.82	11/5385 (0.2%)
1	F	0.67	1/3911 (0.0%)	0.86	12/5299 (0.2%)
All	All	0.69	4/23693 (0.0%)	0.82	51/32106 (0.2%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	60	TRP	CB-CG	-5.61	1.40	1.50
1	F	485	TYR	N-CA	5.58	1.57	1.46
1	A	109	ASP	CB-CG	-5.22	1.40	1.51
1	D	109	ASP	CB-CG	-5.13	1.41	1.51

The worst 5 of 51 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	201	MET	CG-SD-CE	9.91	116.06	100.20
1	F	156	ASP	CB-CG-OD2	9.56	126.91	118.30
1	F	156	ASP	CB-CG-OD1	-9.17	110.05	118.30
1	F	201	MET	CG-SD-CE	8.19	113.30	100.20
1	F	461	LEU	CA-CB-CG	7.27	132.03	115.30

There are no chirality outliers.

There are no planarity outliers.

## 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	3841	0	3701	51	0
1	B	3832	0	3693	51	0
1	C	3832	0	3691	32	0
1	D	3952	0	3819	62	1
1	E	3886	0	3746	72	0
1	F	3825	0	3687	63	0
2	A	42	0	38	0	0
2	B	42	0	38	1	0
2	C	70	0	64	2	0
2	D	28	0	26	0	0
2	E	28	0	26	0	0
2	F	28	0	25	0	0
3	A	226	0	0	4	0
3	B	219	0	0	8	0
3	C	175	0	0	4	0
3	D	212	0	0	14	0
3	E	85	0	0	15	0
3	F	107	0	0	1	0
All	All	24430	0	22554	299	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 7.

The worst 5 of 299 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:145:ARG:O	1:E:252:TRP:NE1	1.57	1.35
1:F:480:VAL:HG23	1:F:485:TYR:OH	1.47	1.13
1:F:476:CYS:O	1:F:485:TYR:OH	1.75	1.04
1:A:447:HIS:CE1	1:A:451:LYS:NZ	2.28	1.01
1:E:241:ASN:O	3:E:701:HOH:O	1.86	0.93

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)
1:D:137:TYR:OH	1:D:137:TYR:OH[2_555]	2.00	0.20

## 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	479/514 (93%)	464 (97%)	14 (3%)	1 (0%)	51	73
1	B	478/514 (93%)	464 (97%)	13 (3%)	1 (0%)	51	73
1	C	478/514 (93%)	465 (97%)	12 (2%)	1 (0%)	51	73
1	D	496/514 (96%)	479 (97%)	15 (3%)	2 (0%)	38	59
1	E	485/514 (94%)	472 (97%)	12 (2%)	1 (0%)	51	73
1	F	477/514 (93%)	461 (97%)	14 (3%)	2 (0%)	38	59
All	All	2893/3084 (94%)	2805 (97%)	80 (3%)	8 (0%)	44	66

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	F	486	ASN
1	D	497	ASN
1	D	346	VAL
1	A	346	VAL
1	E	346	VAL

### 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	421/443 (95%)	401 (95%)	20 (5%)	30	53
1	B	420/443 (95%)	396 (94%)	24 (6%)	24	44
1	C	420/443 (95%)	402 (96%)	18 (4%)	33	58
1	D	431/443 (97%)	409 (95%)	22 (5%)	28	50
1	E	425/443 (96%)	410 (96%)	15 (4%)	41	68
1	F	419/443 (95%)	399 (95%)	20 (5%)	30	53
All	All	2536/2658 (95%)	2417 (95%)	119 (5%)	30	54

5 of 119 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	159	SER
1	E	132	THR
1	F	217	ARG
1	C	206	MET
1	C	454	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	86	GLN
1	E	171	GLN
1	D	8	HIS
1	C	497	ASN
1	E	445	ASN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

17 non-standard protein/DNA/RNA residues 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	$\# Z  > 2$
2	NAG	A	601	1	14,14,15	0.68	0	15,19,21	1.55	5 (33%)
2	NAG	A	602	1,2	14,14,15	0.78	1 (7%)	15,19,21	2.03	5 (33%)
2	NAG	A	603	2	14,14,15	0.62	0	15,19,21	1.26	2 (13%)
2	NAG	B	601	1	14,14,15	0.64	0	15,19,21	2.67	6 (40%)
2	NAG	B	602	1,2	14,14,15	0.72	1 (7%)	15,19,21	1.84	4 (26%)
2	NAG	B	603	2	14,14,15	0.73	0	15,19,21	1.58	2 (13%)
2	NAG	C	601	1	14,14,15	1.00	1 (7%)	15,19,21	2.92	4 (26%)
2	NAG	C	602	1	14,14,15	0.75	1 (7%)	15,19,21	1.70	3 (20%)
2	NAG	C	603	1,2	14,14,15	0.81	1 (7%)	15,19,21	1.22	1 (6%)
2	NAG	C	604	2	14,14,15	0.69	0	15,19,21	2.04	3 (20%)
2	NAG	C	605	1	14,14,15	1.00	1 (7%)	15,19,21	1.68	3 (20%)
2	NAG	D	601	1	14,14,15	0.74	0	15,19,21	2.38	5 (33%)
2	NAG	D	602	1	14,14,15	0.63	0	15,19,21	1.97	1 (6%)
2	NAG	E	601	1	14,14,15	0.79	0	15,19,21	2.40	3 (20%)
2	NAG	E	602	1	14,14,15	0.49	0	15,19,21	1.63	1 (6%)
2	NAG	F	2000	1,2	14,14,15	0.91	1 (7%)	15,19,21	1.44	2 (13%)
2	NAG	F	2001	2	14,14,15	0.66	0	15,19,21	2.18	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
2	NAG	A	601	1	1/1/5/7	0/6/23/26	0/1/1/1
2	NAG	A	602	1,2	-	0/6/23/26	0/1/1/1
2	NAG	A	603	2	-	0/6/23/26	0/1/1/1
2	NAG	B	601	1	1/1/5/7	0/6/23/26	0/1/1/1
2	NAG	B	602	1,2	-	0/6/23/26	0/1/1/1
2	NAG	B	603	2	-	0/6/23/26	0/1/1/1
2	NAG	C	601	1	-	0/6/23/26	0/1/1/1
2	NAG	C	602	1	-	0/6/23/26	0/1/1/1
2	NAG	C	603	1,2	-	0/6/23/26	0/1/1/1
2	NAG	C	604	2	-	0/6/23/26	0/1/1/1
2	NAG	C	605	1	1/1/5/7	0/6/23/26	0/1/1/1
2	NAG	D	601	1	1/1/5/7	0/6/23/26	0/1/1/1

*Continued on next page...*



*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	NAG	D	602	1	1/1/5/7	0/6/23/26	0/1/1/1
2	NAG	E	601	1	1/1/5/7	0/6/23/26	0/1/1/1
2	NAG	E	602	1	-	0/6/23/26	0/1/1/1
2	NAG	F	2000	1,2	-	0/6/23/26	0/1/1/1
2	NAG	F	2001	2	-	0/6/23/26	0/1/1/1

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	603	NAG	O5-C1	-2.17	1.40	1.43
2	B	602	NAG	C1-C2	2.07	1.55	1.52
2	F	2000	NAG	C1-C2	2.14	1.55	1.52
2	C	605	NAG	C1-C2	2.18	1.55	1.52
2	A	602	NAG	C1-C2	2.27	1.55	1.52

The worst 5 of 54 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	602	NAG	O5-C1-C2	-7.27	101.36	111.47
2	D	601	NAG	O5-C1-C2	-6.72	102.12	111.47
2	E	601	NAG	O5-C1-C2	-5.99	103.13	111.47
2	B	602	NAG	O5-C1-C2	-4.19	105.65	111.47
2	C	605	NAG	O5-C1-C2	-3.54	106.55	111.47

5 of 6 chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	B	601	NAG	C1
2	A	601	NAG	C1
2	D	601	NAG	C1
2	E	601	NAG	C1
2	D	602	NAG	C1

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	601	NAG	1	0
2	C	601	NAG	1	0
2	C	605	NAG	1	0

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

17 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$
2	NAG	A	601	1	14,14,15	0.68	0	15,19,21	1.55	5 (33%)
2	NAG	A	602	1,2	14,14,15	0.78	1 (7%)	15,19,21	2.03	5 (33%)
2	NAG	A	603	2	14,14,15	0.62	0	15,19,21	1.26	2 (13%)
2	NAG	B	601	1	14,14,15	0.64	0	15,19,21	2.67	6 (40%)
2	NAG	B	602	1,2	14,14,15	0.72	1 (7%)	15,19,21	1.84	4 (26%)
2	NAG	B	603	2	14,14,15	0.73	0	15,19,21	1.58	2 (13%)
2	NAG	C	601	1	14,14,15	1.00	1 (7%)	15,19,21	2.92	4 (26%)
2	NAG	C	602	1	14,14,15	0.75	1 (7%)	15,19,21	1.70	3 (20%)
2	NAG	C	603	1,2	14,14,15	0.81	1 (7%)	15,19,21	1.22	1 (6%)
2	NAG	C	604	2	14,14,15	0.69	0	15,19,21	2.04	3 (20%)
2	NAG	C	605	1	14,14,15	1.00	1 (7%)	15,19,21	1.68	3 (20%)
2	NAG	D	601	1	14,14,15	0.74	0	15,19,21	2.38	5 (33%)
2	NAG	D	602	1	14,14,15	0.63	0	15,19,21	1.97	1 (6%)
2	NAG	E	601	1	14,14,15	0.79	0	15,19,21	2.40	3 (20%)
2	NAG	E	602	1	14,14,15	0.49	0	15,19,21	1.63	1 (6%)
2	NAG	F	2000	1,2	14,14,15	0.91	1 (7%)	15,19,21	1.44	2 (13%)
2	NAG	F	2001	2	14,14,15	0.66	0	15,19,21	2.18	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
2	NAG	A	601	1	1/1/5/7	0/6/23/26	0/1/1/1
2	NAG	A	602	1,2	-	0/6/23/26	0/1/1/1
2	NAG	A	603	2	-	0/6/23/26	0/1/1/1
2	NAG	B	601	1	1/1/5/7	0/6/23/26	0/1/1/1
2	NAG	B	602	1,2	-	0/6/23/26	0/1/1/1
2	NAG	B	603	2	-	0/6/23/26	0/1/1/1
2	NAG	C	601	1	-	0/6/23/26	0/1/1/1
2	NAG	C	602	1	-	0/6/23/26	0/1/1/1
2	NAG	C	603	1,2	-	0/6/23/26	0/1/1/1
2	NAG	C	604	2	-	0/6/23/26	0/1/1/1
2	NAG	C	605	1	1/1/5/7	0/6/23/26	0/1/1/1
2	NAG	D	601	1	1/1/5/7	0/6/23/26	0/1/1/1
2	NAG	D	602	1	1/1/5/7	0/6/23/26	0/1/1/1
2	NAG	E	601	1	1/1/5/7	0/6/23/26	0/1/1/1
2	NAG	E	602	1	-	0/6/23/26	0/1/1/1
2	NAG	F	2000	1,2	-	0/6/23/26	0/1/1/1
2	NAG	F	2001	2	-	0/6/23/26	0/1/1/1

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	603	NAG	O5-C1	-2.17	1.40	1.43
2	B	602	NAG	C1-C2	2.07	1.55	1.52
2	F	2000	NAG	C1-C2	2.14	1.55	1.52
2	C	605	NAG	C1-C2	2.18	1.55	1.52
2	A	602	NAG	C1-C2	2.27	1.55	1.52

The worst 5 of 54 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	602	NAG	O5-C1-C2	-7.27	101.36	111.47
2	D	601	NAG	O5-C1-C2	-6.72	102.12	111.47
2	E	601	NAG	O5-C1-C2	-5.99	103.13	111.47
2	B	602	NAG	O5-C1-C2	-4.19	105.65	111.47
2	C	605	NAG	O5-C1-C2	-3.54	106.55	111.47

5 of 6 chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
2	B	601	NAG	C1
2	A	601	NAG	C1
2	D	601	NAG	C1
2	E	601	NAG	C1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atom
2	D	602	NAG	C1

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	601	NAG	1	0
2	C	601	NAG	1	0
2	C	605	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, 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	483/514 (93%)	0.19	30 (6%) 21 22	29, 47, 101, 126	0
1	B	482/514 (93%)	0.45	43 (8%) 10 10	30, 51, 125, 173	0
1	C	482/514 (93%)	0.24	33 (6%) 18 18	31, 53, 104, 133	0
1	D	498/514 (96%)	0.55	48 (9%) 9 8	31, 59, 151, 182	0
1	E	489/514 (95%)	0.54	42 (8%) 11 11	39, 73, 112, 140	0
1	F	481/514 (93%)	0.79	86 (17%) 2 1	36, 72, 146, 195	0
All	All	2915/3084 (94%)	0.46	282 (9%) 8 8	29, 60, 126, 195	0

The worst 5 of 282 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	3	ILE	9.0
1	D	456	ASP	7.6
1	D	469	TRP	7.1
1	B	485	TYR	6.9
1	B	477	ILE	6.8

### 6.2 Non-standard residues in protein, DNA, RNA chains [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, 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	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
2	NAG	B	602	14/15	0.91	0.25	5.29	78,86,92,99	0
2	NAG	A	602	14/15	0.91	0.29	4.74	63,71,78,86	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	NAG	E	601	14/15	0.83	0.22	0.88	61,88,96,124	0
2	NAG	F	2000	14/15	0.91	0.14	0.20	60,72,76,81	0
2	NAG	C	602	14/15	0.81	0.26	-	106,118,123,131	0
2	NAG	A	601	14/15	0.83	0.24	-	87,105,115,115	0
2	NAG	F	2001	14/15	0.82	0.24	-	86,95,101,102	0
2	NAG	C	601	14/15	0.78	0.17	-	99,105,113,114	0
2	NAG	E	602	14/15	0.82	0.15	-	101,111,115,119	0
2	NAG	B	603	14/15	0.82	0.27	-	104,115,119,121	0
2	NAG	C	603	14/15	0.84	0.18	-	95,108,115,124	0
2	NAG	D	602	14/15	0.62	0.34	-	102,130,143,148	0
2	NAG	D	601	14/15	0.72	0.33	-	127,140,146,151	0
2	NAG	A	603	14/15	0.84	0.29	-	87,96,100,105	0
2	NAG	C	604	14/15	0.75	0.41	-	105,124,129,130	0
2	NAG	B	601	14/15	0.73	0.24	-	117,125,140,143	0
2	NAG	C	605	14/15	0.86	0.23	-	61,80,98,109	0

### 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, 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	NAG	B	602	14/15	0.91	0.25	5.29	78,86,92,99	0
2	NAG	A	602	14/15	0.91	0.29	4.74	63,71,78,86	0
2	NAG	E	601	14/15	0.83	0.22	0.88	61,88,96,124	0
2	NAG	F	2000	14/15	0.91	0.14	0.20	60,72,76,81	0
2	NAG	C	602	14/15	0.81	0.26	-	106,118,123,131	0
2	NAG	A	601	14/15	0.83	0.24	-	87,105,115,115	0
2	NAG	F	2001	14/15	0.82	0.24	-	86,95,101,102	0
2	NAG	C	601	14/15	0.78	0.17	-	99,105,113,114	0
2	NAG	E	602	14/15	0.82	0.15	-	101,111,115,119	0
2	NAG	B	603	14/15	0.82	0.27	-	104,115,119,121	0
2	NAG	C	603	14/15	0.84	0.18	-	95,108,115,124	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	NAG	D	602	14/15	0.62	0.34	-	102,130,143,148	0
2	NAG	D	601	14/15	0.72	0.33	-	127,140,146,151	0
2	NAG	A	603	14/15	0.84	0.29	-	87,96,100,105	0
2	NAG	C	604	14/15	0.75	0.41	-	105,124,129,130	0
2	NAG	B	601	14/15	0.73	0.24	-	117,125,140,143	0
2	NAG	C	605	14/15	0.86	0.23	-	61,80,98,109	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.