



# Full wwPDB X-ray Structure Validation Report

Feb 28, 2014 – 10:48 AM GMT

PDB ID : 3VUN  
Title : Crystal structure of a influenza A virus (A/Aichi/2/1968 H3N2) hemagglutinin in C2 space group.  
Authors : Yasutake, Y.; Suzuki, T.; Kawaguchi, A.; Nobusawa, E.  
Deposited on : 2012-07-02  
Resolution : 3.00 Å(reported)

This is a full wwPDB validation 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/ValidationPDFNotes.html>

---

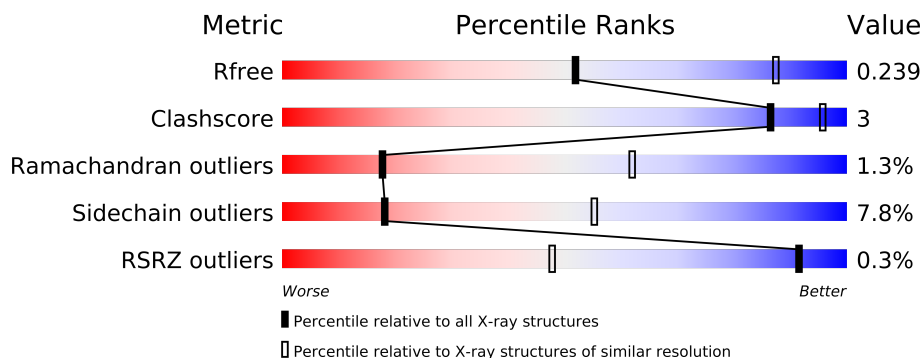
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.15 2013  
Xtriage (Phenix) : dev-1323  
EDS : stable22639  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable22683

# 1 Overall quality at a glance

The reported resolution of this entry is 3.00 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))
$R_{free}$	66092	1216 (3.00-3.00)
Clashscore	79885	1594 (3.00-3.00)
Ramachandran outliers	78287	1537 (3.00-3.00)
Sidechain outliers	78261	1540 (3.00-3.00)
RSRZ outliers	66119	1217 (3.00-3.00)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	329	
1	C	329	
1	E	329	
2	B	175	
2	D	175	
2	F	175	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
3	NAG	D	401	-	X
3	NAG	E	401	-	X

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 12255 atoms, of which 0 are hydrogen and 0 are deuterium.

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 HA1 chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	319	Total	C	N	O	S	0	0	0
			2463	1540	432	478	13			
1	C	319	Total	C	N	O	S	0	0	0
			2463	1540	432	478	13			
1	E	319	Total	C	N	O	S	0	0	0
			2463	1540	432	478	13			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	144	SER	GLY	SEE REMARK 999	UNP P03437
A	182	VAL	ILE	SEE REMARK 999	UNP P03437
C	144	SER	GLY	SEE REMARK 999	UNP P03437
C	182	VAL	ILE	SEE REMARK 999	UNP P03437
E	144	SER	GLY	SEE REMARK 999	UNP P03437
E	182	VAL	ILE	SEE REMARK 999	UNP P03437

- Molecule 2 is a protein called Hemagglutinin HA2 chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	173	Total	C	N	O	S	0	0	0
			1406	873	247	280	6			
2	D	173	Total	C	N	O	S	0	0	0
			1406	873	247	280	6			
2	F	173	Total	C	N	O	S	0	0	0
			1406	873	247	280	6			

There are 3 discrepancies between the modelled and reference sequences:

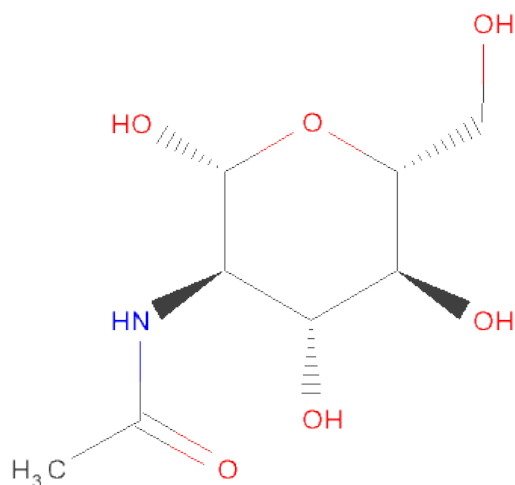
Chain	Residue	Modelled	Actual	Comment	Reference
B	132	ASP	GLU	SEE REMARK 999	UNP P03437
D	132	ASP	GLU	SEE REMARK 999	UNP P03437

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
F	132	ASP	GLU	SEE REMARK 999	UNP P03437

- 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	B	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	D	1	Total	C	N	O	0	0
			14	8	1	5		
3	E	1	Total	C	N	O	0	0
			14	8	1	5		
3	F	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 4 is a polymer of unknown type called SUGAR (9-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	9	Total	C	N	O	0	0
			110	62	4	44		
4	C	9	Total	C	N	O	0	0
			110	62	4	44		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	E	9	Total	C	N	O	0	0
			110	62	4	44		

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	144	SER	GLY	SEE REMARK 999	UNP P03437
A	182	VAL	ILE	SEE REMARK 999	UNP P03437
C	144	SER	GLY	SEE REMARK 999	UNP P03437
C	182	VAL	ILE	SEE REMARK 999	UNP P03437
E	144	SER	GLY	SEE REMARK 999	UNP P03437
E	182	VAL	ILE	SEE REMARK 999	UNP P03437

- Molecule 5 is a polymer of unknown type called SUGAR (2-MER).

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

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	144	SER	GLY	SEE REMARK 999	UNP P03437
A	182	VAL	ILE	SEE REMARK 999	UNP P03437
C	144	SER	GLY	SEE REMARK 999	UNP P03437
C	182	VAL	ILE	SEE REMARK 999	UNP P03437
E	144	SER	GLY	SEE REMARK 999	UNP P03437
E	182	VAL	ILE	SEE REMARK 999	UNP P03437

- Molecule 6 is a polymer of unknown type called SUGAR (4-MER).

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
6	A	4	Total	C	N	O	0	0
			50	28	2	20		
6	C	4	Total	C	N	O	0	0
			50	28	2	20		
6	E	4	Total	C	N	O	0	0
			50	28	2	20		

There are 6 discrepancies between the modelled and reference sequences:

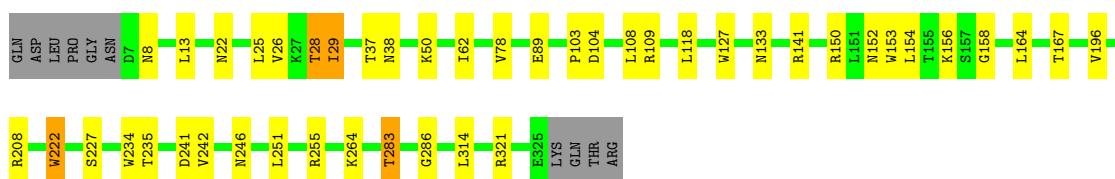
Chain	Residue	Modelled	Actual	Comment	Reference
A	144	SER	GLY	SEE REMARK 999	UNP P03437
A	182	VAL	ILE	SEE REMARK 999	UNP P03437
C	144	SER	GLY	SEE REMARK 999	UNP P03437
C	182	VAL	ILE	SEE REMARK 999	UNP P03437
E	144	SER	GLY	SEE REMARK 999	UNP P03437
E	182	VAL	ILE	SEE REMARK 999	UNP P03437

### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

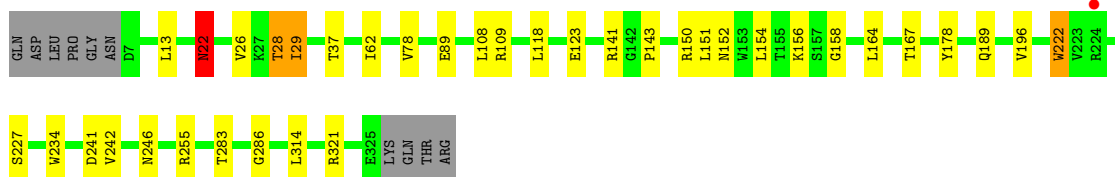
- Molecule 1: Hemagglutinin HA1 chain

Chain A: 



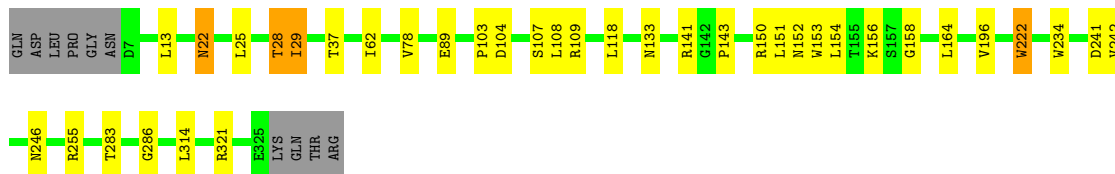
- Molecule 1: Hemagglutinin HA1 chain

Chain C: 



- Molecule 1: Hemagglutinin HA1 chain

Chain E: 



- Molecule 2: Hemagglutinin HA2 chain

Chain B: 



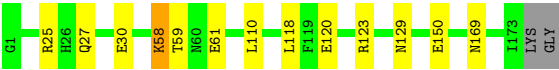
- Molecule 2: Hemagglutinin HA2 chain

Chain D: 



● Molecule 2: Hemagglutinin HA2 chain

Chain F:





## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	170.02Å 98.13Å 144.78Å 90.00° 113.11° 90.00°	Depositor
Resolution (Å)	30.59 – 3.00 30.59 – 3.00	Depositor EDS
% Data completeness (in resolution range)	99.1 (30.59-3.00) 99.3 (30.59-3.00)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.95 (at 3.00Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, $R_{free}$	0.216 , 0.275 0.212 , 0.239	Depositor DCC
$R_{free}$ test set	2197 reflections (5.31%)	DCC
Wilson B-factor (Å <sup>2</sup> )	73.4	Xtriage
Anisotropy	0.158	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 31.4	EDS
Estimated twinning fraction	0.457 for 1/2*h-3/2*k,-1/2*h-1/2*k,-1/2*h +1/2*k-l 0.448 for 1/2*h+3/2*k,1/2*h-1/2*k,-1/2*h- 1/2*k-l	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.52$ , $\langle L^2 \rangle = 0.36$	Xtriage
Outliers	0 of 43586 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	12255	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	84.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.71% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: FUC, GAL, NAG, BMA, MAN

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	0.53	5/2519 (0.2%)	0.58	0/3434
1	C	0.48	2/2519 (0.1%)	0.58	0/3434
1	E	0.50	3/2519 (0.1%)	0.58	0/3434
2	B	0.48	0/1430	0.54	0/1922
2	D	0.48	2/1430 (0.1%)	0.54	0/1922
2	F	0.48	0/1430	0.54	0/1922
All	All	0.50	12/11847 (0.1%)	0.57	0/16068

All (12) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	153	TRP	CD2-CE2	9.31	1.52	1.41
1	A	153	TRP	CG-CD1	7.02	1.46	1.36
1	E	153	TRP	CD2-CE2	6.58	1.49	1.41
1	A	222	TRP	CD2-CE2	5.46	1.48	1.41
1	E	222	TRP	CD2-CE2	5.38	1.47	1.41
1	A	234	TRP	CD2-CE2	5.31	1.47	1.41
1	E	234	TRP	CD2-CE2	5.27	1.47	1.41
1	C	222	TRP	CD2-CE2	5.27	1.47	1.41
1	C	234	TRP	CD2-CE2	5.18	1.47	1.41
2	D	21	TRP	CD2-CE2	5.14	1.47	1.41
1	A	127	TRP	CD2-CE2	5.11	1.47	1.41
2	D	14	TRP	CD2-CE2	5.10	1.47	1.41

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

## 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2463	0	0	7	0
1	C	2463	0	0	7	0
1	E	2463	0	0	8	0
2	B	1406	0	0	2	0
2	D	1406	0	0	2	0
2	F	1406	0	0	1	0
3	A	14	0	13	2	0
3	B	14	0	13	0	0
3	C	14	0	13	1	0
3	D	14	0	13	0	0
3	E	14	0	13	1	0
3	F	14	0	13	0	0
4	A	110	0	94	1	0
4	C	110	0	94	0	0
4	E	110	0	94	2	0
5	A	28	0	25	1	0
5	C	28	0	25	0	0
5	E	28	0	25	0	0
6	A	50	0	43	0	0
6	C	50	0	43	0	0
6	E	50	0	43	0	0
All	All	12255	0	564	33	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 3.

All (33) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:E:89:GLU:OE2	1:E:109:ARG:NH2	2.21	0.72
2:F:120:GLU:OE1	2:F:123:ARG:NH2	2.23	0.72
1:C:89:GLU:OE2	1:C:109:ARG:NH2	2.22	0.71
1:A:89:GLU:OE2	1:A:109:ARG:NH2	2.24	0.71
2:D:120:GLU:OE1	2:D:123:ARG:NH2	2.26	0.68
4:E:404:BMA:H61	4:E:405:MAN:H3	1.78	0.66

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Distance(Å)	Clash(Å)
4:A:404:BMA:H61	4:A:405:MAN:H3	1.78	0.65
2:B:120:GLU:OE1	2:B:123:ARG:NH2	2.32	0.62
1:C:283:THR:CG2	1:C:286:GLY:O	2.51	0.58
3:E:401:NAG:H3	3:E:401:NAG:H83	1.90	0.54
1:A:28:THR:CG2	1:A:29:ILE:N	2.73	0.51
3:C:401:NAG:H83	3:C:401:NAG:H3	1.91	0.51
3:A:401:NAG:H83	3:A:401:NAG:H3	1.92	0.51
1:E:22:ASN:OD1	1:E:22:ASN:N	2.44	0.49
1:C:22:ASN:N	1:C:22:ASN:OD1	2.46	0.49
1:E:283:THR:CG2	1:E:286:GLY:O	2.62	0.47
1:E:133:ASN:ND2	1:E:255:ARG:NH1	2.63	0.47
1:E:151:LEU:O	1:E:255:ARG:NH2	2.48	0.47
1:E:103:PRO:O	1:E:104:ASP:OD1	2.34	0.46
1:A:133:ASN:ND2	1:A:255:ARG:NH1	2.64	0.46
1:A:283:THR:CG2	1:A:286:GLY:O	2.64	0.45
5:A:411:NAG:H61	5:A:412:NAG:N2	2.30	0.45
1:E:241:ASP:OD1	1:E:242:VAL:N	2.50	0.45
1:A:103:PRO:O	1:A:104:ASP:OD1	2.36	0.43
1:A:241:ASP:OD1	1:A:242:VAL:N	2.51	0.43
1:C:151:LEU:O	1:C:255:ARG:NH2	2.50	0.43
1:A:38:ASN:ND2	3:A:401:NAG:C4	2.82	0.43
1:C:28:THR:CG2	1:C:29:ILE:N	2.81	0.43
1:E:28:THR:CG2	1:E:29:ILE:N	2.81	0.43
1:C:241:ASP:OD1	1:C:242:VAL:N	2.52	0.43
2:B:90:ASP:OD2	2:D:60:ASN:ND2	2.53	0.42
4:E:404:BMA:H61	4:E:405:MAN:C3	2.46	0.41
1:C:123:GLU:OE2	1:C:178:TYR:OH	2.39	0.41

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone ⓘ

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	317/329 (96%)	293 (92%)	20 (6%)	4 (1%)	18 62

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	C	317/329 (96%)	291 (92%)	21 (7%)	5 (2%)	14 56
1	E	317/329 (96%)	295 (93%)	17 (5%)	5 (2%)	14 56
2	B	171/175 (98%)	156 (91%)	12 (7%)	3 (2%)	13 53
2	D	171/175 (98%)	158 (92%)	12 (7%)	1 (1%)	33 81
2	F	171/175 (98%)	157 (92%)	13 (8%)	1 (1%)	33 81
All	All	1464/1512 (97%)	1350 (92%)	95 (6%)	19 (1%)	18 62

All (19) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	22	ASN
1	C	62	ILE
1	A	22	ASN
1	E	22	ASN
1	E	158	GLY
1	A	62	ILE
1	A	196	VAL
2	B	34	GLN
2	D	58	LYS
1	E	62	ILE
1	E	143	PRO
2	F	58	LYS
1	A	158	GLY
2	B	58	LYS
1	C	158	GLY
1	C	196	VAL
1	E	196	VAL
2	B	17	MET
1	C	143	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	282/291 (97%)	254 (90%)	28 (10%)	11 40

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	C	282/291 (97%)	260 (92%)	22 (8%)	18	55
1	E	282/291 (97%)	263 (93%)	19 (7%)	23	64
2	B	148/149 (99%)	138 (93%)	10 (7%)	22	63
2	D	148/149 (99%)	137 (93%)	11 (7%)	20	58
2	F	148/149 (99%)	137 (93%)	11 (7%)	20	58
All	All	1290/1320 (98%)	1189 (92%)	101 (8%)	18	55

All (101) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	8	ASN
1	A	13	LEU
1	A	25	LEU
1	A	26	VAL
1	A	28	THR
1	A	29	ILE
1	A	37	THR
1	A	50	LYS
1	A	78	VAL
1	A	108	LEU
1	A	118	LEU
1	A	141	ARG
1	A	150	ARG
1	A	152	ASN
1	A	154	LEU
1	A	156	LYS
1	A	164	LEU
1	A	167	THR
1	A	208	ARG
1	A	222	TRP
1	A	227	SER
1	A	235	THR
1	A	246	ASN
1	A	251	LEU
1	A	264	LYS
1	A	283	THR
1	A	314	LEU
1	A	321	ARG
2	B	25	ARG
2	B	27	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
2	B	58	LYS
2	B	59	THR
2	B	60	ASN
2	B	61	GLU
2	B	118	LEU
2	B	143	LYS
2	B	150	GLU
2	B	169	ASN
1	C	13	LEU
1	C	22	ASN
1	C	26	VAL
1	C	28	THR
1	C	29	ILE
1	C	37	THR
1	C	78	VAL
1	C	108	LEU
1	C	118	LEU
1	C	141	ARG
1	C	150	ARG
1	C	152	ASN
1	C	154	LEU
1	C	156	LYS
1	C	164	LEU
1	C	167	THR
1	C	189	GLN
1	C	222	TRP
1	C	227	SER
1	C	246	ASN
1	C	314	LEU
1	C	321	ARG
2	D	25	ARG
2	D	27	GLN
2	D	30	GLU
2	D	58	LYS
2	D	59	THR
2	D	61	GLU
2	D	87	THR
2	D	118	LEU
2	D	143	LYS
2	D	150	GLU
2	D	169	ASN
1	E	13	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	E	25	LEU
1	E	28	THR
1	E	29	ILE
1	E	37	THR
1	E	78	VAL
1	E	107	SER
1	E	108	LEU
1	E	118	LEU
1	E	141	ARG
1	E	150	ARG
1	E	152	ASN
1	E	154	LEU
1	E	156	LYS
1	E	164	LEU
1	E	222	TRP
1	E	246	ASN
1	E	314	LEU
1	E	321	ARG
2	F	25	ARG
2	F	27	GLN
2	F	30	GLU
2	F	58	LYS
2	F	59	THR
2	F	61	GLU
2	F	110	LEU
2	F	118	LEU
2	F	129	ASN
2	F	150	GLU
2	F	169	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

There are no RNA chains 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 ⓘ

45 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	402	1,4	12,14,15	0.88	1 (8%)	15,19,21	0.84	0
4	NAG	A	403	4	12,14,15	0.66	0	15,19,21	0.98	0
4	BMA	A	404	4	10,11,12	0.45	0	11,15,17	1.40	1 (9%)
4	MAN	A	405	4	10,11,12	0.79	0	11,15,17	0.89	1 (9%)
4	NAG	A	406	4	12,14,15	0.59	0	15,19,21	0.93	0
4	GAL	A	407	4	10,11,12	0.69	0	11,15,17	0.76	1 (9%)
4	FUC	A	408	4	9,10,11	1.43	1 (11%)	10,14,16	1.02	1 (10%)
4	MAN	A	409	4	10,11,12	0.64	0	11,15,17	1.08	1 (9%)
4	NAG	A	410	4	12,14,15	0.66	0	15,19,21	0.83	0
5	NAG	A	411	1,5	12,14,15	0.70	1 (8%)	15,19,21	0.81	0
5	NAG	A	412	5	12,14,15	0.44	0	15,19,21	1.03	1 (6%)
6	NAG	A	413	1,6	12,14,15	0.61	0	15,19,21	0.92	0
6	NAG	A	414	6	12,14,15	0.61	0	15,19,21	1.12	1 (6%)
6	BMA	A	415	6	10,11,12	0.47	0	11,15,17	1.16	1 (9%)
6	MAN	A	416	6	10,11,12	0.70	0	11,15,17	1.06	1 (9%)
4	NAG	C	402	1,4	12,14,15	0.95	1 (8%)	15,19,21	1.01	0
4	NAG	C	403	4	12,14,15	0.58	0	15,19,21	1.17	1 (6%)
4	BMA	C	404	4	10,11,12	0.45	0	11,15,17	1.30	2 (18%)
4	MAN	C	405	4	10,11,12	0.78	0	11,15,17	0.98	0
4	NAG	C	406	4	12,14,15	0.65	0	15,19,21	0.83	0
4	GAL	C	407	4	10,11,12	0.69	0	11,15,17	1.13	1 (9%)
4	FUC	C	408	4	9,10,11	1.02	1 (11%)	10,14,16	0.65	0
4	MAN	C	409	4	10,11,12	0.68	0	11,15,17	0.84	0
4	NAG	C	410	4	12,14,15	0.52	0	15,19,21	0.84	0
5	NAG	C	411	1,5	12,14,15	0.70	0	15,19,21	0.99	1 (6%)
5	NAG	C	412	5	12,14,15	0.44	0	15,19,21	1.07	1 (6%)
6	NAG	C	413	1,6	12,14,15	0.55	0	15,19,21	1.12	2 (13%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
6	NAG	C	414	6	12,14,15	0.52	0	15,19,21	1.08	1 (6%)
6	BMA	C	415	6	10,11,12	0.46	0	11,15,17	1.20	1 (9%)
6	MAN	C	416	6	10,11,12	0.71	0	11,15,17	1.25	1 (9%)
4	NAG	E	402	1,4	12,14,15	0.94	1 (8%)	15,19,21	0.92	1 (6%)
4	NAG	E	403	4	12,14,15	0.60	0	15,19,21	1.20	1 (6%)
4	BMA	E	404	4	10,11,12	0.44	0	11,15,17	1.40	1 (9%)
4	MAN	E	405	4	10,11,12	0.75	0	11,15,17	0.93	1 (9%)
4	NAG	E	406	4	12,14,15	0.59	0	15,19,21	1.00	1 (6%)
4	GAL	E	407	4	10,11,12	0.69	0	11,15,17	1.12	1 (9%)
4	FUC	E	408	4	9,10,11	0.81	0	10,14,16	0.84	0
4	MAN	E	409	4	10,11,12	0.68	0	11,15,17	0.84	0
4	NAG	E	410	4	12,14,15	0.57	0	15,19,21	0.73	0
5	NAG	E	411	1,5	12,14,15	0.70	0	15,19,21	1.01	2 (13%)
5	NAG	E	412	5	12,14,15	0.48	0	15,19,21	1.26	1 (6%)
6	NAG	E	413	1,6	12,14,15	0.58	0	15,19,21	1.10	0
6	NAG	E	414	6	12,14,15	0.50	0	15,19,21	1.22	1 (6%)
6	BMA	E	415	6	10,11,12	0.46	0	11,15,17	1.18	1 (9%)
6	MAN	E	416	6	10,11,12	0.69	0	11,15,17	1.08	1 (9%)

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	402	1,4	-	0/6/23/26	0/1/1/1
4	NAG	A	403	4	-	0/6/23/26	0/1/1/1
4	BMA	A	404	4	-	0/2/19/22	0/1/1/1
4	MAN	A	405	4	-	0/2/19/22	0/1/1/1
4	NAG	A	406	4	-	0/6/23/26	0/1/1/1
4	GAL	A	407	4	-	0/2/19/22	0/1/1/1
4	FUC	A	408	4	-	0/0/17/20	0/1/1/1
4	MAN	A	409	4	-	0/2/19/22	0/1/1/1
4	NAG	A	410	4	-	0/6/23/26	0/1/1/1
5	NAG	A	411	1,5	-	0/6/23/26	0/1/1/1
5	NAG	A	412	5	-	0/6/23/26	0/1/1/1
6	NAG	A	413	1,6	-	0/6/23/26	0/1/1/1
6	NAG	A	414	6	-	0/6/23/26	0/1/1/1
6	BMA	A	415	6	-	0/2/19/22	0/1/1/1

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
6	MAN	A	416	6	-	0/2/19/22	0/1/1/1
4	NAG	C	402	1,4	-	0/6/23/26	0/1/1/1
4	NAG	C	403	4	-	0/6/23/26	0/1/1/1
4	BMA	C	404	4	-	0/2/19/22	0/1/1/1
4	MAN	C	405	4	-	0/2/19/22	0/1/1/1
4	NAG	C	406	4	-	0/6/23/26	0/1/1/1
4	GAL	C	407	4	-	0/2/19/22	0/1/1/1
4	FUC	C	408	4	-	0/0/17/20	0/1/1/1
4	MAN	C	409	4	-	0/2/19/22	0/1/1/1
4	NAG	C	410	4	-	0/6/23/26	0/1/1/1
5	NAG	C	411	1,5	-	0/6/23/26	0/1/1/1
5	NAG	C	412	5	-	0/6/23/26	0/1/1/1
6	NAG	C	413	1,6	-	0/6/23/26	0/1/1/1
6	NAG	C	414	6	-	0/6/23/26	0/1/1/1
6	BMA	C	415	6	-	0/2/19/22	0/1/1/1
6	MAN	C	416	6	-	0/2/19/22	0/1/1/1
4	NAG	E	402	1,4	-	0/6/23/26	0/1/1/1
4	NAG	E	403	4	-	0/6/23/26	0/1/1/1
4	BMA	E	404	4	-	0/2/19/22	0/1/1/1
4	MAN	E	405	4	-	0/2/19/22	0/1/1/1
4	NAG	E	406	4	-	0/6/23/26	0/1/1/1
4	GAL	E	407	4	-	0/2/19/22	0/1/1/1
4	FUC	E	408	4	-	0/0/17/20	0/1/1/1
4	MAN	E	409	4	-	0/2/19/22	0/1/1/1
4	NAG	E	410	4	-	0/6/23/26	0/1/1/1
5	NAG	E	411	1,5	-	0/6/23/26	0/1/1/1
5	NAG	E	412	5	-	0/6/23/26	0/1/1/1
6	NAG	E	413	1,6	-	0/6/23/26	0/1/1/1
6	NAG	E	414	6	-	0/6/23/26	0/1/1/1
6	BMA	E	415	6	-	0/2/19/22	0/1/1/1
6	MAN	E	416	6	-	0/2/19/22	0/1/1/1

All (6) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	408	FUC	O5-C5	-2.67	1.40	1.45
4	E	402	NAG	O5-C5	-2.17	1.41	1.45
4	C	402	NAG	O5-C5	-2.16	1.41	1.45
4	A	402	NAG	O5-C5	-2.13	1.41	1.45
5	A	411	NAG	O5-C5	-2.03	1.41	1.45
4	C	408	FUC	O5-C5	-2.03	1.41	1.45

All (32) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	E	412	NAG	O5-C5-C6	4.14	111.32	106.98
4	E	403	NAG	O5-C5-C6	3.52	110.67	106.98
6	E	415	BMA	O5-C5-C6	3.48	110.64	106.98
4	C	403	NAG	O5-C5-C6	3.45	110.60	106.98
6	C	414	NAG	O5-C5-C4	3.36	114.92	110.65
4	E	407	GAL	O5-C5-C6	3.24	110.39	106.98
6	C	416	MAN	O5-C5-C6	3.18	110.31	106.98
5	C	412	NAG	O5-C5-C6	3.11	110.25	106.98
4	C	407	GAL	O5-C5-C6	3.11	110.24	106.98
6	E	414	NAG	O5-C5-C4	3.09	114.58	110.65
5	A	412	NAG	O5-C5-C6	2.98	110.10	106.98
6	E	416	MAN	O5-C5-C6	2.97	110.09	106.98
6	A	416	MAN	O5-C5-C6	2.92	110.05	106.98
6	C	415	BMA	O5-C5-C6	2.81	109.93	106.98
6	A	414	NAG	O5-C5-C4	2.80	114.21	110.65
6	A	415	BMA	O5-C5-C6	2.80	109.92	106.98
5	C	411	NAG	C3-C2-N2	-2.68	107.68	111.76
4	C	404	BMA	O3-C3-C4	-2.55	104.63	110.35
5	E	411	NAG	C3-C2-N2	-2.25	108.33	111.76
5	E	411	NAG	C4-C3-C2	2.23	116.77	111.32
6	C	413	NAG	C3-C4-C5	-2.22	106.24	110.20
4	A	407	GAL	O5-C5-C6	2.21	109.30	106.98
4	A	409	MAN	O5-C5-C4	2.18	113.42	110.65
4	E	404	BMA	O3-C3-C4	-2.17	105.49	110.35
4	E	405	MAN	O5-C5-C6	2.17	109.25	106.98
4	A	404	BMA	C4-C3-C2	2.14	113.38	110.50
4	E	402	NAG	O5-C5-C6	-2.12	104.75	106.98
4	E	406	NAG	O5-C5-C4	2.12	113.34	110.65
4	A	405	MAN	O5-C5-C6	2.06	109.14	106.98
4	C	404	BMA	O5-C5-C4	-2.06	108.04	110.65
4	A	408	FUC	O5-C5-C4	2.03	113.29	110.22
6	C	413	NAG	O4-C4-C5	2.01	114.58	109.28

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 5.6 Ligand geometry ⓘ

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	401	1	12,14,15	0.69	0	15,19,21	2.37	5 (33%)
3	NAG	B	401	2	12,14,15	0.68	0	15,19,21	1.46	3 (20%)
3	NAG	C	401	1	12,14,15	0.70	0	15,19,21	2.42	6 (40%)
3	NAG	D	401	2	12,14,15	0.66	0	15,19,21	1.33	3 (20%)
3	NAG	E	401	1	12,14,15	0.65	0	15,19,21	2.35	6 (40%)
3	NAG	F	401	2	12,14,15	0.60	0	15,19,21	1.12	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	401	1	-	0/6/23/26	0/1/1/1
3	NAG	B	401	2	-	0/6/23/26	0/1/1/1
3	NAG	C	401	1	-	0/6/23/26	0/1/1/1
3	NAG	D	401	2	-	0/6/23/26	0/1/1/1
3	NAG	E	401	1	-	0/6/23/26	0/1/1/1
3	NAG	F	401	2	-	0/6/23/26	0/1/1/1

There are no bond length outliers.

All (23) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	401	NAG	C2-N2-C7	6.17	133.46	123.09
3	A	401	NAG	C2-N2-C7	6.06	133.26	123.09
3	E	401	NAG	C2-N2-C7	5.90	133.00	123.09
3	E	401	NAG	O5-C5-C6	4.34	111.54	106.98
3	A	401	NAG	O5-C5-C6	4.02	111.20	106.98
3	C	401	NAG	O5-C5-C6	3.74	110.90	106.98
3	C	401	NAG	C8-C7-N2	3.34	122.63	116.11
3	A	401	NAG	C8-C7-N2	3.22	122.40	116.11
3	B	401	NAG	C3-C4-C5	3.18	115.88	110.20

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	E	401	NAG	C8-C7-N2	3.15	122.27	116.11
3	B	401	NAG	C4-C3-C2	2.80	118.18	111.32
3	D	401	NAG	C3-C4-C5	2.71	115.04	110.20
3	C	401	NAG	O7-C7-C8	-2.71	116.76	122.04
3	D	401	NAG	C4-C3-C2	2.63	117.75	111.32
3	C	401	NAG	C3-C4-C5	2.60	114.85	110.20
3	A	401	NAG	O7-C7-C8	-2.58	117.01	122.04
3	E	401	NAG	O7-C7-C8	-2.46	117.25	122.04
3	C	401	NAG	O5-C5-C4	2.33	113.61	110.65
3	A	401	NAG	C3-C4-C5	2.28	114.27	110.20
3	B	401	NAG	O5-C5-C6	2.19	109.27	106.98
3	E	401	NAG	C3-C4-C5	2.18	114.09	110.20
3	D	401	NAG	O5-C5-C6	2.09	109.17	106.98
3	E	401	NAG	O5-C5-C4	2.02	113.22	110.65

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 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 ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

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	319/329 (96%)	-0.04	0 100 100	54, 79, 116, 139	0
1	C	319/329 (96%)	-0.04	1 (0%) 91 48	55, 79, 116, 151	0
1	E	319/329 (96%)	-0.05	0 100 100	57, 79, 116, 142	0
2	B	173/175 (98%)	0.01	2 (1%) 75 20	53, 83, 114, 132	0
2	D	173/175 (98%)	-0.04	1 (0%) 86 32	51, 85, 113, 136	0
2	F	173/175 (98%)	-0.05	0 100 100	54, 84, 113, 136	0
All	All	1476/1512 (97%)	-0.04	4 (0%) 91 48	51, 81, 115, 151	0

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	60	ASN	2.3
1	C	224	ARG	2.2
2	B	59	THR	2.2
2	D	59	THR	2.0

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

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates ⓘ

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	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
6	NAG	C	414	14/15	0.29	31.29	115,124,129,130	0
4	GAL	C	407	11/12	0.35	1.81	92,106,108,110	0
6	MAN	A	416	11/12	0.21	1.50	133,139,142,145	0
5	NAG	A	411	14/15	0.25	0.27	94,101,110,111	0
5	NAG	E	412	14/15	0.31	0.23	111,120,122,123	0
4	NAG	E	402	14/15	0.21	0.04	63,65,73,76	0
4	MAN	A	409	11/12	0.22	-0.06	82,85,88,92	0
4	FUC	E	408	10/11	0.23	-0.15	71,74,76,77	0
5	NAG	E	411	14/15	0.20	-0.28	94,103,108,116	0
6	NAG	E	414	14/15	0.23	-0.28	112,121,129,132	0
4	NAG	C	406	14/15	0.21	-0.34	84,100,104,107	0
4	NAG	E	406	14/15	0.21	-0.40	93,108,110,115	0
4	FUC	A	408	10/11	0.23	-0.42	71,74,75,81	0
4	GAL	E	407	11/12	0.25	-0.45	106,111,116,118	0
6	NAG	C	413	14/15	0.16	-0.54	73,82,92,104	0
4	FUC	C	408	10/11	0.18	-0.73	71,74,78,79	0
6	MAN	E	416	11/12	0.20	-0.83	131,141,143,147	0
4	NAG	A	406	14/15	0.13	-0.96	94,107,111,116	0
4	NAG	C	402	14/15	0.18	-1.12	60,64,71,74	0
5	NAG	C	411	14/15	0.17	-1.25	92,102,111,112	0
4	NAG	C	403	14/15	0.19	-1.27	62,68,70,75	0
4	NAG	A	403	14/15	0.16	-1.41	63,70,72,75	0
6	MAN	C	416	11/12	0.10	-1.45	118,129,131,134	0
4	GAL	A	407	11/12	0.19	-1.52	107,114,118,121	0
4	NAG	A	402	14/15	0.15	-1.82	61,64,71,75	0
6	NAG	A	413	14/15	0.14	-1.99	75,81,93,106	0
4	NAG	E	403	14/15	0.17	-2.21	66,72,74,75	0
4	MAN	A	405	11/12	0.18	-2.36	95,104,110,113	0
6	NAG	A	414	14/15	0.21	-2.64	109,122,130,133	0
4	MAN	C	405	11/12	0.21	-2.85	95,100,107,110	0
6	NAG	E	413	14/15	0.14	-3.01	78,86,97,107	0
5	NAG	A	412	14/15	0.20	-3.65	107,114,116,121	0
4	MAN	E	405	11/12	0.16	-3.82	91,104,114,116	0
4	MAN	C	409	11/12	0.13	-3.86	77,80,84,91	0
4	MAN	E	409	11/12	0.13	-3.87	75,79,85,95	0
5	NAG	C	412	14/15	0.12	-4.58	98,113,117,120	0
4	BMA	A	404	11/12	0.15	-	78,82,86,95	0
4	NAG	E	410	14/15	0.28	-	105,113,121,125	0
6	BMA	C	415	11/12	0.26	-	124,129,134,137	0
4	NAG	A	410	14/15	0.30	-	102,106,114,117	0
4	BMA	E	404	11/12	0.17	-	75,83,85,96	0
4	NAG	C	410	14/15	0.19	-	94,105,109,115	0
6	BMA	E	415	11/12	0.17	-	128,132,135,135	0

Continued on next page...



*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
4	BMA	C	404	11/12	0.15	-	74,83,92,97	0
6	BMA	A	415	11/12	0.20	-	130,134,136,136	0

## 6.4 Ligands

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	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	NAG	E	401	14/15	0.42	8.39	114,126,130,132	0
3	NAG	D	401	14/15	0.34	7.14	110,116,121,125	0
3	NAG	B	401	14/15	0.28	1.29	110,118,121,122	0
3	NAG	F	401	14/15	0.19	0.04	108,115,120,121	0
3	NAG	C	401	14/15	0.20	-2.29	112,123,128,130	0
3	NAG	A	401	14/15	0.17	-3.14	109,119,123,124	0

## 6.5 Other polymers

There are no such residues in this entry.