



# wwPDB X-ray Structure Validation Summary Report ⓘ

Dec 17, 2023 – 07:00 PM EST

PDB ID : 4WST  
Title : The crystal structure of hemagglutinin from A/Taiwan/1/2013 influenza virus  
Authors : Yang, H.; Carney, P.J.; Chang, J.; Villanueva, J.M.; Stevens, J.  
Deposited on : 2014-10-28  
Resolution : 2.40 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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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 : 2.36  
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) : 2.36

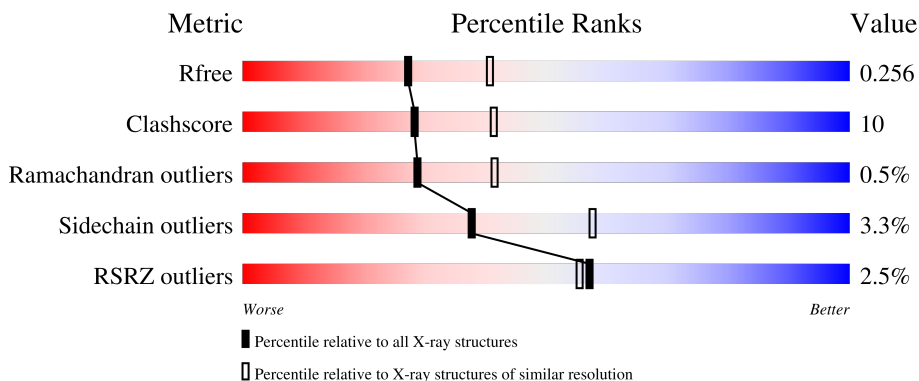
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.40 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))
$R_{free}$	130704	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria respectively. 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	334	 83% 13% ..
1	C	334	 82% 14% ..
1	E	334	 82% 14% ..
1	G	334	 84% 13% ..
1	I	334	 84% 12% ..

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	K	334	
2	B	181	
2	D	181	
2	F	181	
2	H	181	
2	J	181	
2	L	181	

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
3	NAG	A	401	X	-	-	-
3	NAG	C	401	X	-	-	-
3	NAG	E	401	X	-	-	-
3	NAG	G	401	X	-	-	-
3	NAG	I	401	X	-	-	-
3	NAG	K	401	X	-	-	-

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 24510 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 HA1 chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	325	2566	1626	437	490	13	0	0	0
1	C	325	2566	1626	437	490	13	0	0	0
1	E	325	2566	1626	437	490	13	0	0	0
1	G	325	2566	1626	437	490	13	0	0	0
1	I	325	2566	1626	437	490	13	0	0	0
1	K	325	2566	1626	437	490	13	0	0	0

- Molecule 2 is a protein called Hemagglutinin HA2 chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	168	1353	845	234	267	7	0	0	0
2	D	168	1353	845	234	267	7	0	0	0
2	F	168	1353	845	234	267	7	0	0	0
2	H	168	1353	845	234	267	7	0	0	0
2	J	168	1353	845	234	267	7	0	0	0
2	L	168	1353	845	234	267	7	0	0	0

- Molecule 3 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: C<sub>8</sub>H<sub>15</sub>NO<sub>6</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
3	A	1	14	8	1	5	0	0
3	A	1	14	8	1	5	0	0
3	A	1	14	8	1	5	0	0
3	C	1	14	8	1	5	0	0
3	C	1	14	8	1	5	0	0
3	C	1	14	8	1	5	0	0
3	E	1	14	8	1	5	0	0
3	E	1	14	8	1	5	0	0
3	E	1	14	8	1	5	0	0
3	G	1	14	8	1	5	0	0
3	G	1	14	8	1	5	0	0
3	G	1	14	8	1	5	0	0
3	I	1	14	8	1	5	0	0
3	I	1	14	8	1	5	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	I	1	Total	C	N	O	0	0
			14	8	1	5		
3	K	1	Total	C	N	O	0	0
			14	8	1	5		
3	K	1	Total	C	N	O	0	0
			14	8	1	5		
3	K	1	Total	C	N	O	0	0
			14	8	1	5		

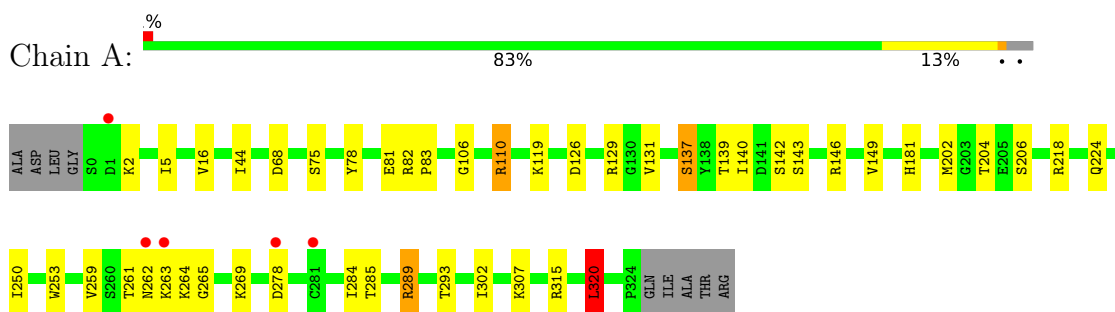
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	105	Total	O	0	0
			105	105		
4	B	33	Total	O	0	0
			33	33		
4	C	91	Total	O	0	0
			91	91		
4	D	43	Total	O	0	0
			43	43		
4	E	94	Total	O	0	0
			94	94		
4	F	46	Total	O	0	0
			46	46		
4	G	99	Total	O	0	0
			99	99		
4	H	36	Total	O	0	0
			36	36		
4	I	65	Total	O	0	0
			65	65		
4	J	28	Total	O	0	0
			28	28		
4	K	66	Total	O	0	0
			66	66		
4	L	38	Total	O	0	0
			38	38		

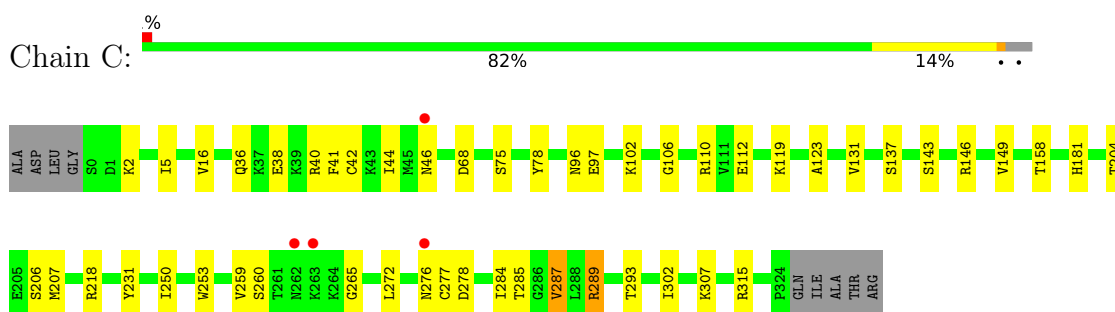
### 3 Residue-property plots [i](#)

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.

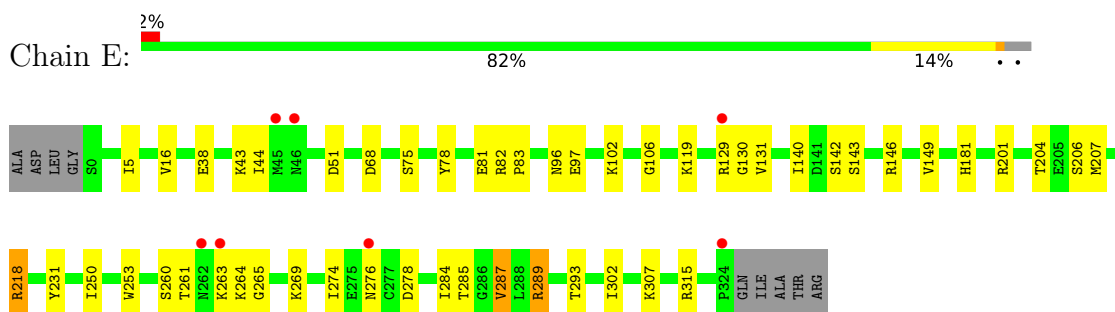
- Molecule 1: Hemagglutinin HA1 chain



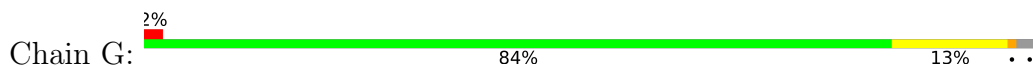
- Molecule 1: Hemagglutinin HA1 chain

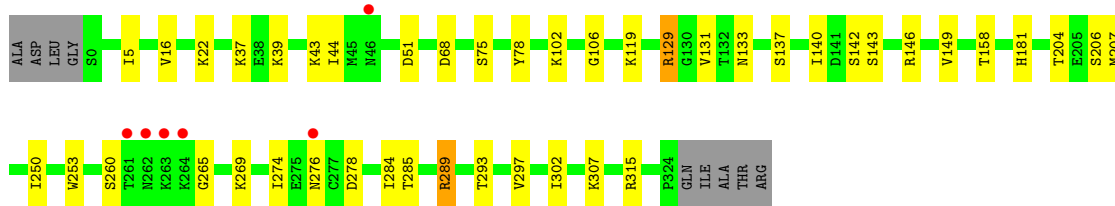


- Molecule 1: Hemagglutinin HA1 chain

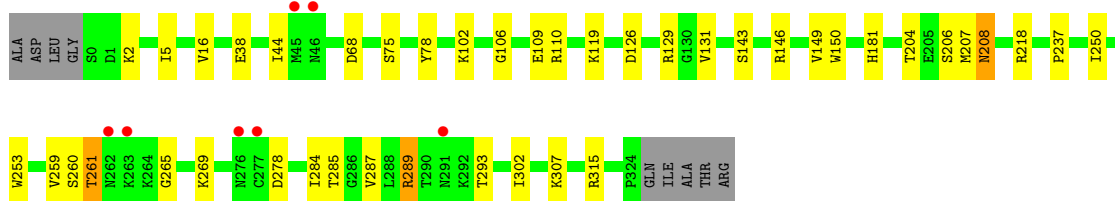
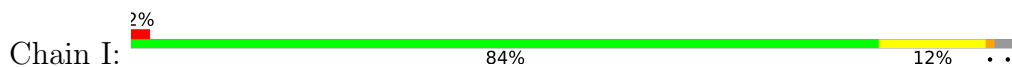


- Molecule 1: Hemagglutinin HA1 chain

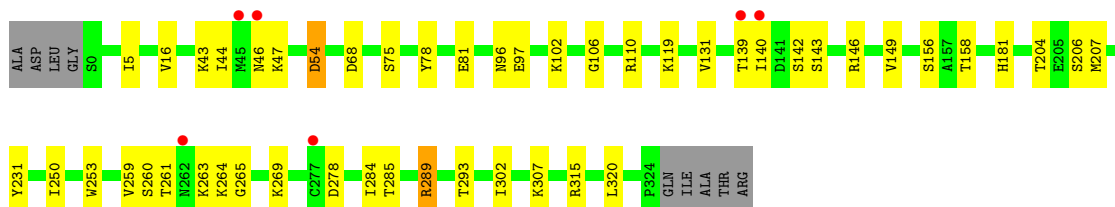
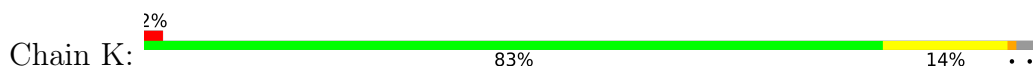




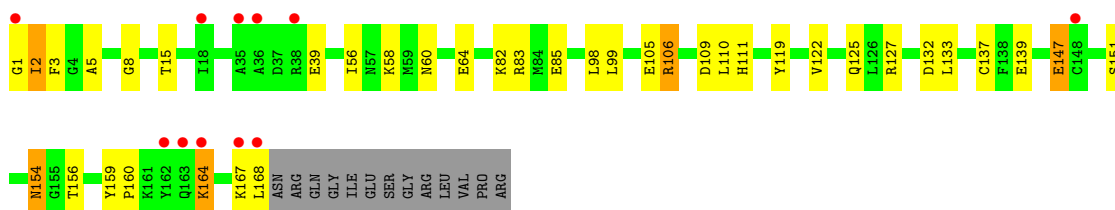
• Molecule 1: Hemagglutinin HA1 chain



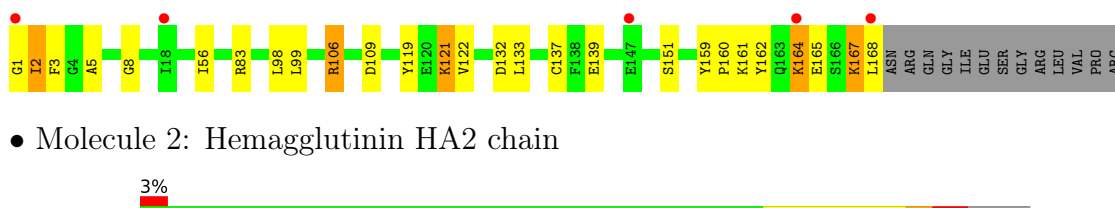
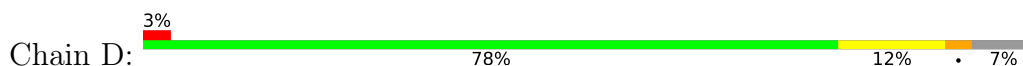
• Molecule 1: Hemagglutinin HA1 chain



• Molecule 2: Hemagglutinin HA2 chain



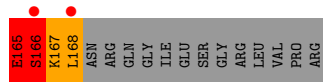
• Molecule 2: Hemagglutinin HA2 chain



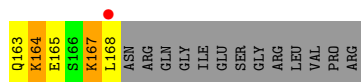
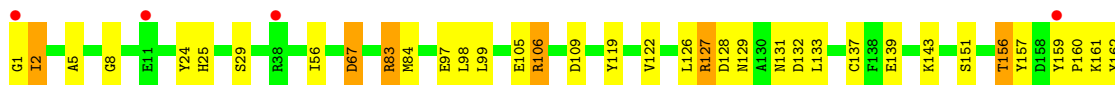
• Molecule 2: Hemagglutinin HA2 chain



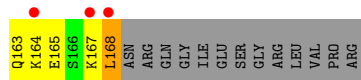
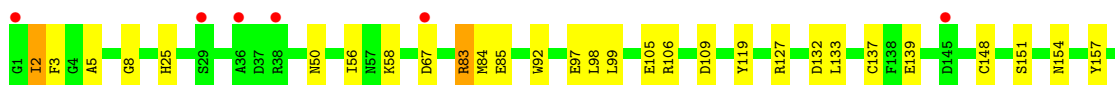
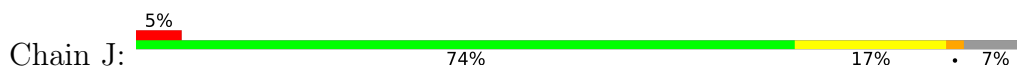




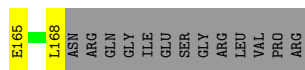
• Molecule 2: Hemagglutinin HA2 chain



• Molecule 2: Hemagglutinin HA2 chain



• Molecule 2: Hemagglutinin HA2 chain



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	226.63Å 100.61Å 175.48Å 90.00° 99.66° 90.00°	Depositor
Resolution (Å)	50.00 – 2.40 48.30 – 2.40	Depositor EDS
% Data completeness (in resolution range)	96.5 (50.00-2.40) 96.6 (48.30-2.40)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.92 (at 2.39Å)	Xtrriage
Refinement program	REFMAC 5.8.0049	Depositor
R, $R_{free}$	0.222 , 0.253 0.226 , 0.256	Depositor DCC
$R_{free}$ test set	7372 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	39.6	Xtrriage
Anisotropy	0.288	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 58.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	24510	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	47.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 46.40 % 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 1.1515e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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 [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	0.77	0/2627	0.83	3/3574 (0.1%)
1	C	0.75	0/2627	0.77	2/3574 (0.1%)
1	E	0.75	0/2627	0.78	2/3574 (0.1%)
1	G	0.74	0/2627	0.79	2/3574 (0.1%)
1	I	0.69	1/2627 (0.0%)	0.76	3/3574 (0.1%)
1	K	0.69	0/2627	0.77	1/3574 (0.0%)
2	B	0.80	0/1381	0.91	3/1860 (0.2%)
2	D	0.81	0/1381	0.90	3/1860 (0.2%)
2	F	0.86	0/1381	1.07	11/1860 (0.6%)
2	H	0.77	0/1381	0.91	4/1860 (0.2%)
2	J	0.79	1/1381 (0.1%)	0.90	5/1860 (0.3%)
2	L	0.80	0/1381	0.90	3/1860 (0.2%)
All	All	0.76	2/24048 (0.0%)	0.84	42/32604 (0.1%)

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
2	F	0	2

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	I	150	TRP	CB-CG	-5.88	1.39	1.50
2	J	67	ASP	CB-CG	5.65	1.63	1.51

The worst 5 of 42 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	F	106	ARG	NE-CZ-NH2	-9.53	115.53	120.30
2	J	106	ARG	NE-CZ-NH2	-9.19	115.70	120.30
2	F	127	ARG	CB-CG-CD	9.16	135.41	111.60
2	L	106	ARG	NE-CZ-NH2	-9.12	115.74	120.30
2	H	106	ARG	NE-CZ-NH2	-9.03	115.78	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	F	164	LYS	Peptide
2	F	166	SER	Peptide

## 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	2566	0	2514	28	0
1	C	2566	0	2514	29	0
1	E	2566	0	2514	33	0
1	G	2566	0	2514	26	0
1	I	2566	0	2514	24	0
1	K	2566	0	2514	26	1
2	B	1353	0	1268	33	1
2	D	1353	0	1268	33	0
2	F	1353	0	1267	83	0
2	H	1353	0	1268	75	0
2	J	1353	0	1268	34	0
2	L	1353	0	1268	72	0
3	A	42	0	39	1	0
3	C	42	0	39	1	0
3	E	42	0	39	0	0
3	G	42	0	39	0	0
3	I	42	0	39	1	0
3	K	42	0	39	0	0
4	A	105	0	0	3	0
4	B	33	0	0	2	0
4	C	91	0	0	8	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	D	43	0	0	2	0
4	E	94	0	0	3	0
4	F	46	0	0	2	0
4	G	99	0	0	6	0
4	H	36	0	0	3	0
4	I	65	0	0	6	0
4	J	28	0	0	2	0
4	K	66	0	0	1	0
4	L	38	0	0	3	0
All	All	24510	0	22925	457	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 10.

The worst 5 of 457 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)
2:H:127:ARG:HD3	2:H:159:TYR:CE1	1.30	1.66
2:L:127:ARG:HD3	2:L:159:TYR:CE1	1.31	1.64
2:H:127:ARG:CD	2:H:159:TYR:CE1	2.21	1.24
2:L:127:ARG:CD	2:L:159:TYR:CE1	2.23	1.20
1:C:277:CYS:SG	4:C:519:HOH:O	2.00	1.17

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)
2:B:15:THR:OG1	1:K:139:THR:O[4_546]	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	323/334 (97%)	314 (97%)	9 (3%)	0	100	100
1	C	323/334 (97%)	314 (97%)	9 (3%)	0	100	100
1	E	323/334 (97%)	315 (98%)	8 (2%)	0	100	100
1	G	323/334 (97%)	315 (98%)	8 (2%)	0	100	100
1	I	323/334 (97%)	314 (97%)	9 (3%)	0	100	100
1	K	323/334 (97%)	315 (98%)	8 (2%)	0	100	100
2	B	166/181 (92%)	157 (95%)	8 (5%)	1 (1%)	25	36
2	D	166/181 (92%)	155 (93%)	7 (4%)	4 (2%)	6	6
2	F	166/181 (92%)	154 (93%)	8 (5%)	4 (2%)	6	6
2	H	166/181 (92%)	156 (94%)	7 (4%)	3 (2%)	8	10
2	J	166/181 (92%)	157 (95%)	8 (5%)	1 (1%)	25	36
2	L	166/181 (92%)	157 (95%)	7 (4%)	2 (1%)	13	19
All	All	2934/3090 (95%)	2823 (96%)	96 (3%)	15 (0%)	29	41

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	2	ILE
2	D	2	ILE
2	F	2	ILE
2	F	164	LYS
2	F	165	GLU

### 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	286/292 (98%)	277 (97%)	9 (3%)	40	60
1	C	286/292 (98%)	275 (96%)	11 (4%)	33	51
1	E	286/292 (98%)	280 (98%)	6 (2%)	53	72
1	G	286/292 (98%)	277 (97%)	9 (3%)	40	60

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	I	286/292 (98%)	279 (98%)	7 (2%)	49 68
1	K	286/292 (98%)	276 (96%)	10 (4%)	36 55
2	B	143/154 (93%)	136 (95%)	7 (5%)	25 40
2	D	143/154 (93%)	141 (99%)	2 (1%)	67 82
2	F	143/154 (93%)	130 (91%)	13 (9%)	9 14
2	H	143/154 (93%)	138 (96%)	5 (4%)	36 55
2	J	143/154 (93%)	143 (100%)	0	100 100
2	L	143/154 (93%)	138 (96%)	5 (4%)	36 55
All	All	2574/2676 (96%)	2490 (97%)	84 (3%)	38 57

5 of 84 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	H	29	SER
1	K	75	SER
2	H	127	ARG
1	I	260	SER
1	K	264	LYS

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

Mol	Chain	Res	Type
2	J	50	ASN
1	G	224	GLN
2	F	142	HIS
1	C	46	ASN
1	G	133	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](#)

18 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
3	NAG	G	401	1	14,14,15	1.08	1 (7%)	17,19,21	1.44	3 (17%)
3	NAG	A	402	1	14,14,15	0.52	0	17,19,21	2.17	4 (23%)
3	NAG	E	400	1	14,14,15	0.50	0	17,19,21	1.61	3 (17%)
3	NAG	I	402	1	14,14,15	0.44	0	17,19,21	1.63	3 (17%)
3	NAG	G	402	1	14,14,15	0.79	1 (7%)	17,19,21	2.02	4 (23%)
3	NAG	A	401	1	14,14,15	1.13	2 (14%)	17,19,21	2.06	8 (47%)
3	NAG	I	400	1	14,14,15	0.54	0	17,19,21	1.16	1 (5%)
3	NAG	C	400	1	14,14,15	0.57	0	17,19,21	1.70	1 (5%)
3	NAG	K	400	1	14,14,15	0.80	0	17,19,21	1.84	4 (23%)
3	NAG	A	400	1	14,14,15	0.55	0	17,19,21	1.11	1 (5%)
3	NAG	C	401	1	14,14,15	1.09	2 (14%)	17,19,21	2.01	7 (41%)
3	NAG	G	400	1	14,14,15	0.51	0	17,19,21	2.03	4 (23%)
3	NAG	K	401	1	14,14,15	1.01	1 (7%)	17,19,21	1.78	4 (23%)
3	NAG	I	401	1	14,14,15	1.05	1 (7%)	17,19,21	1.74	2 (11%)
3	NAG	K	402	1	14,14,15	0.49	0	17,19,21	2.11	3 (17%)
3	NAG	E	402	1	14,14,15	0.90	1 (7%)	17,19,21	1.97	5 (29%)
3	NAG	E	401	1	14,14,15	0.84	1 (7%)	17,19,21	1.41	3 (17%)
3	NAG	C	402	1	14,14,15	0.77	1 (7%)	17,19,21	1.84	4 (23%)

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	G	401	1	1/1/5/7	0/6/23/26	0/1/1/1
3	NAG	A	402	1	-	2/6/23/26	0/1/1/1
3	NAG	E	400	1	-	0/6/23/26	0/1/1/1
3	NAG	I	402	1	-	0/6/23/26	0/1/1/1
3	NAG	G	402	1	-	2/6/23/26	0/1/1/1
3	NAG	A	401	1	1/1/5/7	0/6/23/26	0/1/1/1
3	NAG	I	400	1	-	0/6/23/26	0/1/1/1

*Continued on next page...*



Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	NAG	C	400	1	-	1/6/23/26	0/1/1/1
3	NAG	K	400	1	-	0/6/23/26	0/1/1/1
3	NAG	A	400	1	-	1/6/23/26	0/1/1/1
3	NAG	C	401	1	1/1/5/7	2/6/23/26	0/1/1/1
3	NAG	G	400	1	-	2/6/23/26	0/1/1/1
3	NAG	K	401	1	1/1/5/7	0/6/23/26	0/1/1/1
3	NAG	I	401	1	1/1/5/7	2/6/23/26	0/1/1/1
3	NAG	K	402	1	-	1/6/23/26	0/1/1/1
3	NAG	E	402	1	-	0/6/23/26	0/1/1/1
3	NAG	E	401	1	1/1/5/7	0/6/23/26	0/1/1/1
3	NAG	C	402	1	-	0/6/23/26	0/1/1/1

The worst 5 of 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	G	401	NAG	C1-C2	2.86	1.56	1.52
3	A	401	NAG	C1-C2	2.75	1.56	1.52
3	C	401	NAG	C1-C2	2.68	1.56	1.52
3	E	402	NAG	C1-C2	2.62	1.56	1.52
3	K	401	NAG	C3-C2	2.49	1.57	1.52

The worst 5 of 64 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	K	402	NAG	C1-O5-C5	6.95	121.60	112.19
3	G	402	NAG	C1-O5-C5	5.79	120.04	112.19
3	G	400	NAG	C1-O5-C5	5.75	119.98	112.19
3	C	400	NAG	C1-O5-C5	5.49	119.64	112.19
3	C	402	NAG	C1-O5-C5	5.33	119.41	112.19

5 of 6 chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
3	A	401	NAG	C1
3	C	401	NAG	C1
3	E	401	NAG	C1
3	G	401	NAG	C1
3	I	401	NAG	C1

5 of 13 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	G	402	NAG	O5-C5-C6-O6
3	G	400	NAG	O5-C5-C6-O6
3	G	402	NAG	C4-C5-C6-O6
3	G	400	NAG	C4-C5-C6-O6
3	I	401	NAG	C4-C5-C6-O6

There are no ring outliers.

3 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	C	400	NAG	1	0
3	A	400	NAG	1	0
3	I	401	NAG	1	0

## 5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

18 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	G	401	1	14,14,15	1.08	1 (7%)	17,19,21	1.44	3 (17%)
3	NAG	A	402	1	14,14,15	0.52	0	17,19,21	2.17	4 (23%)
3	NAG	E	400	1	14,14,15	0.50	0	17,19,21	1.61	3 (17%)
3	NAG	I	402	1	14,14,15	0.44	0	17,19,21	1.63	3 (17%)
3	NAG	G	402	1	14,14,15	0.79	1 (7%)	17,19,21	2.02	4 (23%)
3	NAG	A	401	1	14,14,15	1.13	2 (14%)	17,19,21	2.06	8 (47%)
3	NAG	I	400	1	14,14,15	0.54	0	17,19,21	1.16	1 (5%)
3	NAG	C	400	1	14,14,15	0.57	0	17,19,21	1.70	1 (5%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
3	NAG	K	400	1	14,14,15	0.80	0	17,19,21	1.84	4 (23%)
3	NAG	A	400	1	14,14,15	0.55	0	17,19,21	1.11	1 (5%)
3	NAG	C	401	1	14,14,15	1.09	2 (14%)	17,19,21	2.01	7 (41%)
3	NAG	G	400	1	14,14,15	0.51	0	17,19,21	2.03	4 (23%)
3	NAG	K	401	1	14,14,15	1.01	1 (7%)	17,19,21	1.78	4 (23%)
3	NAG	I	401	1	14,14,15	1.05	1 (7%)	17,19,21	1.74	2 (11%)
3	NAG	K	402	1	14,14,15	0.49	0	17,19,21	2.11	3 (17%)
3	NAG	E	402	1	14,14,15	0.90	1 (7%)	17,19,21	1.97	5 (29%)
3	NAG	E	401	1	14,14,15	0.84	1 (7%)	17,19,21	1.41	3 (17%)
3	NAG	C	402	1	14,14,15	0.77	1 (7%)	17,19,21	1.84	4 (23%)

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	G	401	1	1/1/5/7	0/6/23/26	0/1/1/1
3	NAG	A	402	1	-	2/6/23/26	0/1/1/1
3	NAG	E	400	1	-	0/6/23/26	0/1/1/1
3	NAG	I	402	1	-	0/6/23/26	0/1/1/1
3	NAG	G	402	1	-	2/6/23/26	0/1/1/1
3	NAG	A	401	1	1/1/5/7	0/6/23/26	0/1/1/1
3	NAG	I	400	1	-	0/6/23/26	0/1/1/1
3	NAG	C	400	1	-	1/6/23/26	0/1/1/1
3	NAG	K	400	1	-	0/6/23/26	0/1/1/1
3	NAG	A	400	1	-	1/6/23/26	0/1/1/1
3	NAG	C	401	1	1/1/5/7	2/6/23/26	0/1/1/1
3	NAG	G	400	1	-	2/6/23/26	0/1/1/1
3	NAG	K	401	1	1/1/5/7	0/6/23/26	0/1/1/1
3	NAG	I	401	1	1/1/5/7	2/6/23/26	0/1/1/1
3	NAG	K	402	1	-	1/6/23/26	0/1/1/1
3	NAG	E	402	1	-	0/6/23/26	0/1/1/1
3	NAG	E	401	1	1/1/5/7	0/6/23/26	0/1/1/1
3	NAG	C	402	1	-	0/6/23/26	0/1/1/1

The worst 5 of 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	G	401	NAG	C1-C2	2.86	1.56	1.52
3	A	401	NAG	C1-C2	2.75	1.56	1.52
3	C	401	NAG	C1-C2	2.68	1.56	1.52
3	E	402	NAG	C1-C2	2.62	1.56	1.52
3	K	401	NAG	C3-C2	2.49	1.57	1.52

The worst 5 of 64 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	K	402	NAG	C1-O5-C5	6.95	121.60	112.19
3	G	402	NAG	C1-O5-C5	5.79	120.04	112.19
3	G	400	NAG	C1-O5-C5	5.75	119.98	112.19
3	C	400	NAG	C1-O5-C5	5.49	119.64	112.19
3	C	402	NAG	C1-O5-C5	5.33	119.41	112.19

5 of 6 chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
3	A	401	NAG	C1
3	C	401	NAG	C1
3	E	401	NAG	C1
3	G	401	NAG	C1
3	I	401	NAG	C1

5 of 13 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	G	402	NAG	O5-C5-C6-O6
3	G	400	NAG	O5-C5-C6-O6
3	G	402	NAG	C4-C5-C6-O6
3	G	400	NAG	C4-C5-C6-O6
3	I	401	NAG	C4-C5-C6-O6

There are no ring outliers.

3 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	C	400	NAG	1	0
3	A	400	NAG	1	0
3	I	401	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	325/334 (97%)	-0.28	5 (1%) 73 72	24, 37, 61, 128	0
1	C	325/334 (97%)	-0.33	4 (1%) 79 77	28, 40, 68, 119	0
1	E	325/334 (97%)	-0.35	7 (2%) 62 60	27, 39, 62, 110	0
1	G	325/334 (97%)	-0.28	6 (1%) 68 66	29, 43, 64, 115	0
1	I	325/334 (97%)	-0.17	7 (2%) 62 60	32, 49, 75, 128	0
1	K	325/334 (97%)	-0.15	6 (1%) 68 66	33, 49, 76, 123	0
2	B	168/181 (92%)	0.25	11 (6%) 18 17	26, 44, 83, 100	0
2	D	168/181 (92%)	0.11	5 (2%) 50 49	25, 41, 74, 99	0
2	F	168/181 (92%)	0.23	6 (3%) 42 42	26, 43, 98, 149	0
2	H	168/181 (92%)	0.10	5 (2%) 50 49	29, 47, 84, 116	0
2	J	168/181 (92%)	0.18	9 (5%) 25 24	27, 48, 74, 102	0
2	L	168/181 (92%)	0.08	3 (1%) 68 66	27, 45, 71, 102	0
All	All	2958/3090 (95%)	-0.12	74 (2%) 57 55	24, 44, 75, 149	0

The worst 5 of 74 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	263	LYS	6.6
1	A	262	ASN	6.4
1	E	263	LYS	5.5
2	B	1	GLY	5.3
2	H	1	GLY	5.1

### 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. 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	B-factors(Å <sup>2</sup> )	Q<0.9
3	NAG	C	401	14/15	0.68	0.22	74,105,115,116	0
3	NAG	I	401	14/15	0.68	0.19	83,90,98,100	0
3	NAG	G	401	14/15	0.71	0.20	80,98,105,107	0
3	NAG	K	401	14/15	0.73	0.22	75,89,93,95	0
3	NAG	A	401	14/15	0.75	0.15	78,88,94,94	0
3	NAG	E	401	14/15	0.76	0.18	83,94,101,106	0
3	NAG	K	400	14/15	0.79	0.24	101,108,125,125	0
3	NAG	I	400	14/15	0.81	0.24	103,112,116,116	0
3	NAG	E	402	14/15	0.82	0.32	75,85,96,97	0
3	NAG	I	402	14/15	0.83	0.25	84,96,100,107	0
3	NAG	A	400	14/15	0.84	0.14	89,100,106,108	0
3	NAG	E	400	14/15	0.84	0.23	88,104,110,113	0
3	NAG	G	400	14/15	0.85	0.18	79,96,111,114	0
3	NAG	G	402	14/15	0.86	0.32	82,89,95,97	0
3	NAG	C	400	14/15	0.86	0.17	96,107,116,118	0
3	NAG	K	402	14/15	0.87	0.24	72,84,95,100	0
3	NAG	C	402	14/15	0.88	0.23	72,80,87,89	0
3	NAG	A	402	14/15	0.92	0.22	56,67,72,72	0

### 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, 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	B-factors(Å <sup>2</sup> )	Q<0.9
3	NAG	C	401	14/15	0.68	0.22	74,105,115,116	0
3	NAG	I	401	14/15	0.68	0.19	83,90,98,100	0
3	NAG	G	401	14/15	0.71	0.20	80,98,105,107	0
3	NAG	K	401	14/15	0.73	0.22	75,89,93,95	0
3	NAG	A	401	14/15	0.75	0.15	78,88,94,94	0
3	NAG	E	401	14/15	0.76	0.18	83,94,101,106	0
3	NAG	K	400	14/15	0.79	0.24	101,108,125,125	0
3	NAG	I	400	14/15	0.81	0.24	103,112,116,116	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
3	NAG	E	402	14/15	0.82	0.32	75,85,96,97	0
3	NAG	I	402	14/15	0.83	0.25	84,96,100,107	0
3	NAG	A	400	14/15	0.84	0.14	89,100,106,108	0
3	NAG	E	400	14/15	0.84	0.23	88,104,110,113	0
3	NAG	G	400	14/15	0.85	0.18	79,96,111,114	0
3	NAG	G	402	14/15	0.86	0.32	82,89,95,97	0
3	NAG	C	400	14/15	0.86	0.17	96,107,116,118	0
3	NAG	K	402	14/15	0.87	0.24	72,84,95,100	0
3	NAG	C	402	14/15	0.88	0.23	72,80,87,89	0
3	NAG	A	402	14/15	0.92	0.22	56,67,72,72	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.