



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 25, 2024 – 05:29 PM EST

PDB ID : 2WR2  
Title : structure of influenza H2 avian hemagglutinin with avian receptor  
Authors : Liu, J.; Stevens, D.J.; Haire, L.F.; Walker, P.A.; Coombs, P.J.; Russell, R.J.; Gamblin, S.J.; Skehel, J.J.  
Deposited on : 2009-08-29  
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	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.21
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.004 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.40

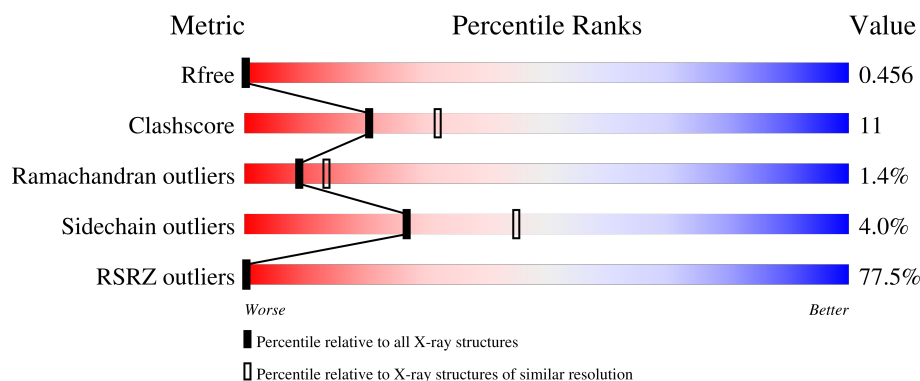
# 1 Overall quality at a glance

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))
$R_{free}$	164625	4642 (2.40-2.40)
Clashscore	180529	5218 (2.40-2.40)
Ramachandran outliers	177936	5158 (2.40-2.40)
Sidechain outliers	177891	5159 (2.40-2.40)
RSRZ outliers	164620	4642 (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  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	A	509	<div> <div>83%</div> <div>77%</div> <div>18%</div> <div>• •</div> </div>
1	B	509	<div> <div>59%</div> <div>72%</div> <div>22%</div> <div>• • •</div> </div>
1	C	509	<div> <div>81%</div> <div>76%</div> <div>17%</div> <div>• 5%</div> </div>
2	D	3	<div> <div>33%</div> <div>33%</div> <div>33%</div> </div>
2	E	3	<div> <div>100%</div> </div>

## 2 Entry composition [i](#)

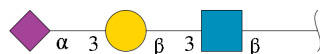
There are 4 unique types of molecules in this entry. The entry contains 12274 atoms, of which 0 are hydrogens and 0 are deuteriums.

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	490	Total	C	N	O	S	4	0	0
			3882	2436	671	753	22			
1	B	490	Total	C	N	O	S	0	1	0
			3887	2438	672	755	22			
1	C	485	Total	C	N	O	S	0	0	0
			3844	2413	664	745	22			

- Molecule 2 is an oligosaccharide called N-acetyl-alpha-neuraminic acid-(2-3)-beta-D-galactopyranose-(1-3)-2-acetamido-2-deoxy-beta-D-glucopyranose.



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	D	3	Total	C	N	O	0	0	0
			46	25	2	19			
2	E	3	Total	C	N	O	0	0	0
			46	25	2	19			

- Molecule 3 is 2-acetamido-2-deoxy-beta-D-glucopyranose (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	A	1	Total	C	N	O	0	0
			13	8	1	4		
3	C	1	Total	C	N	O	0	0
			14	8	1	5		

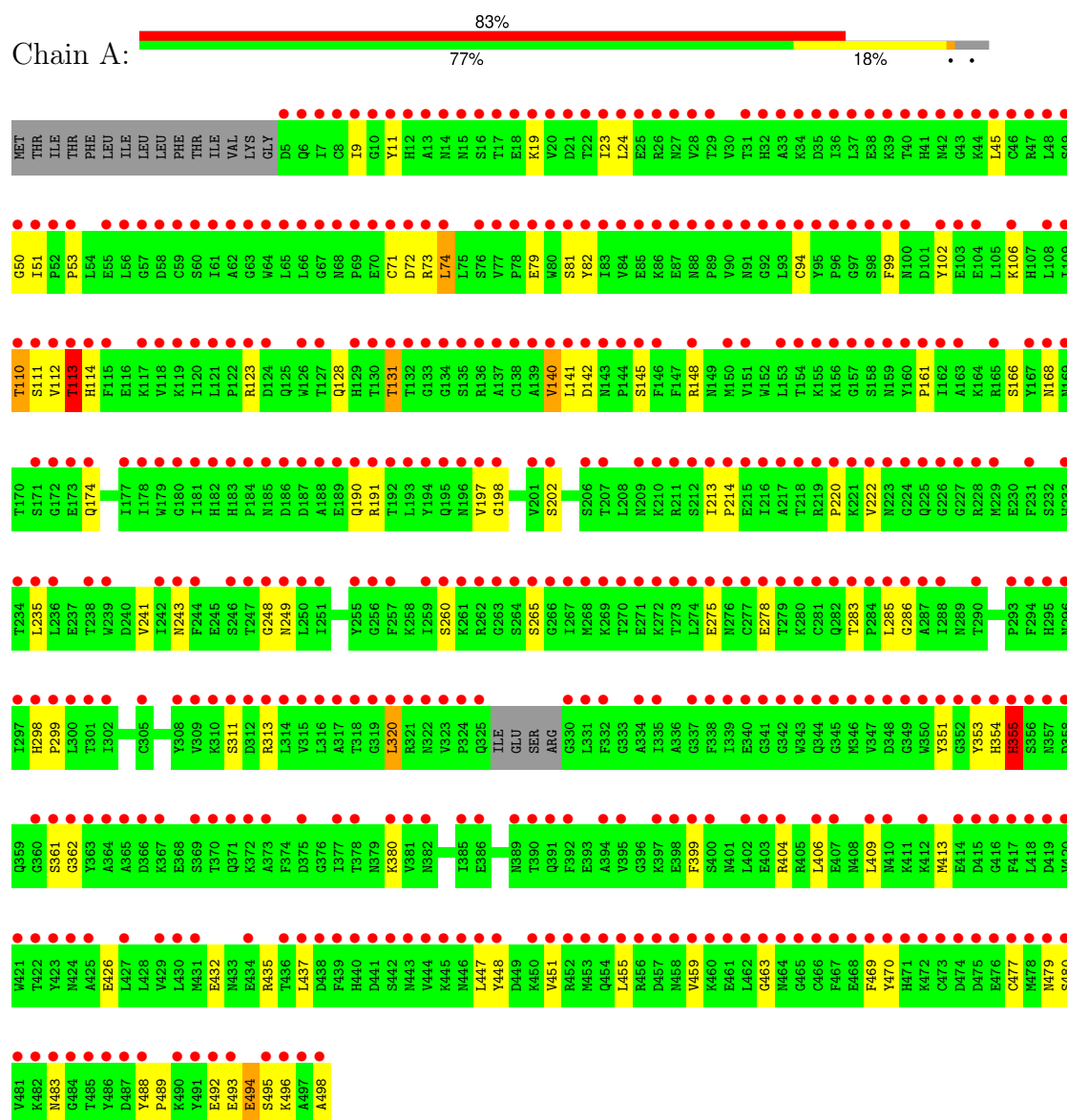
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	185	Total	O	0	0
			185	185		
4	B	169	Total	O	0	0
			169	169		
4	C	174	Total	O	0	0
			174	174		

### 3 Residue-property plots

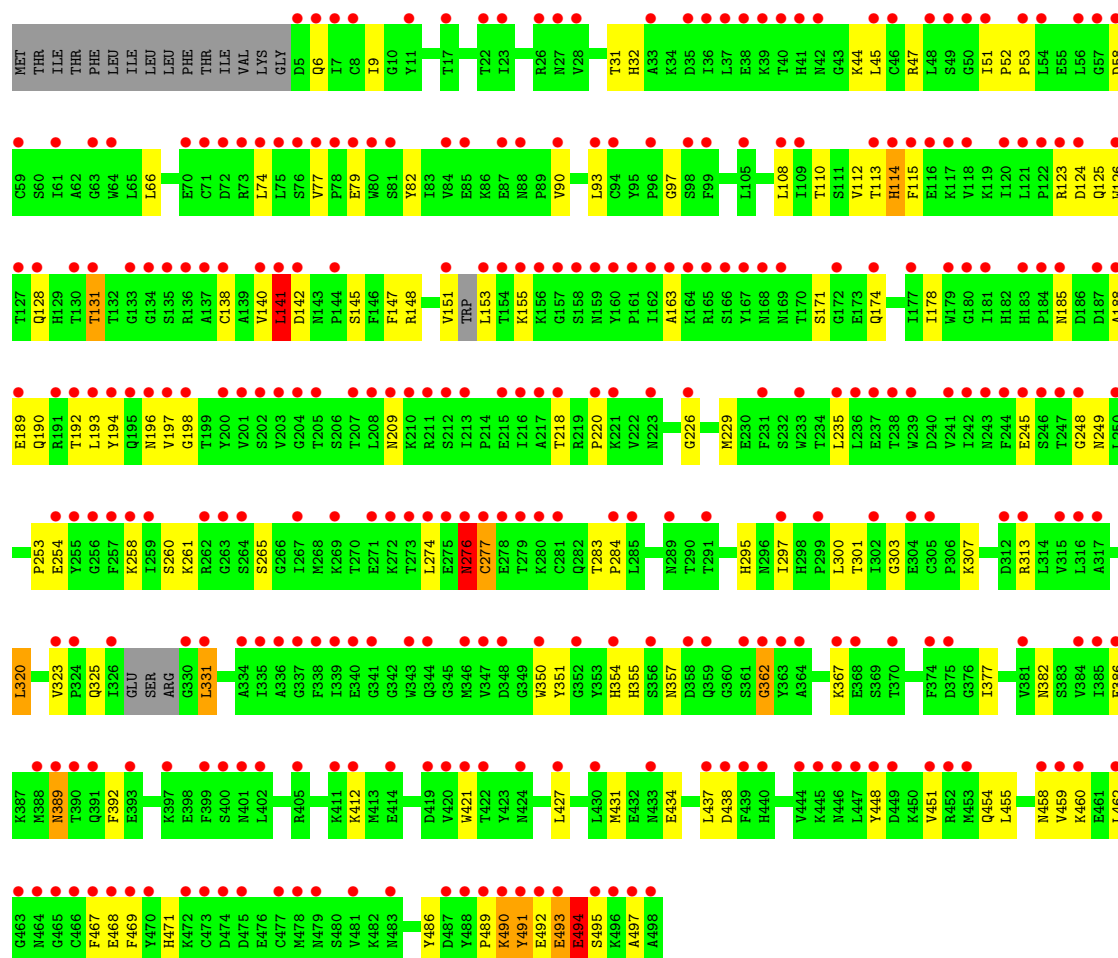
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. Residues are color-coded 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. 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

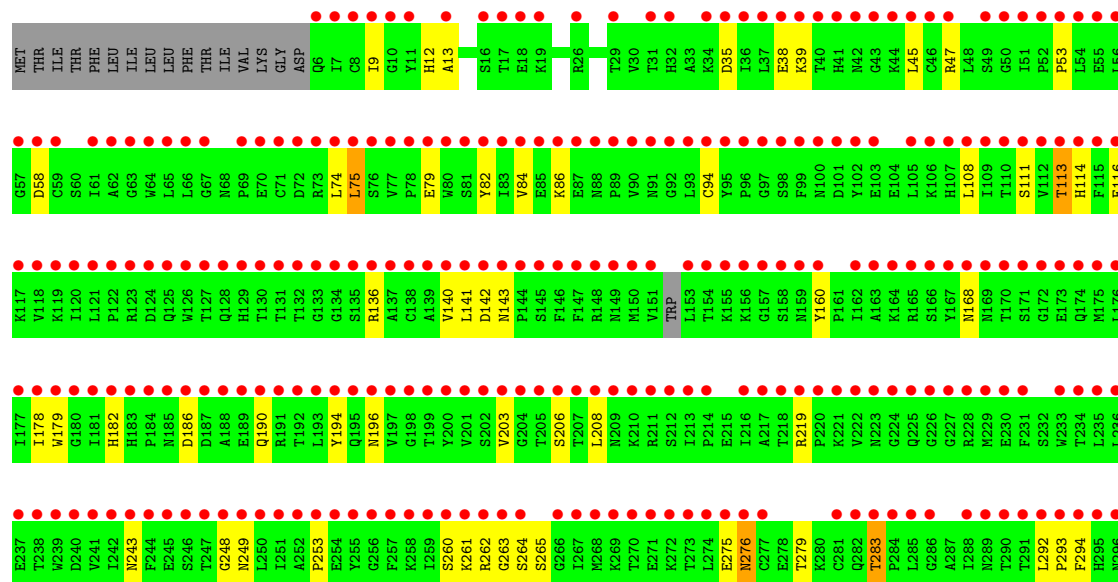
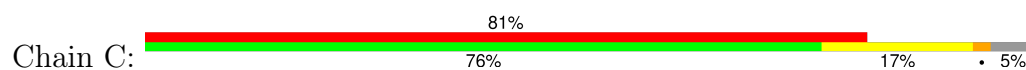


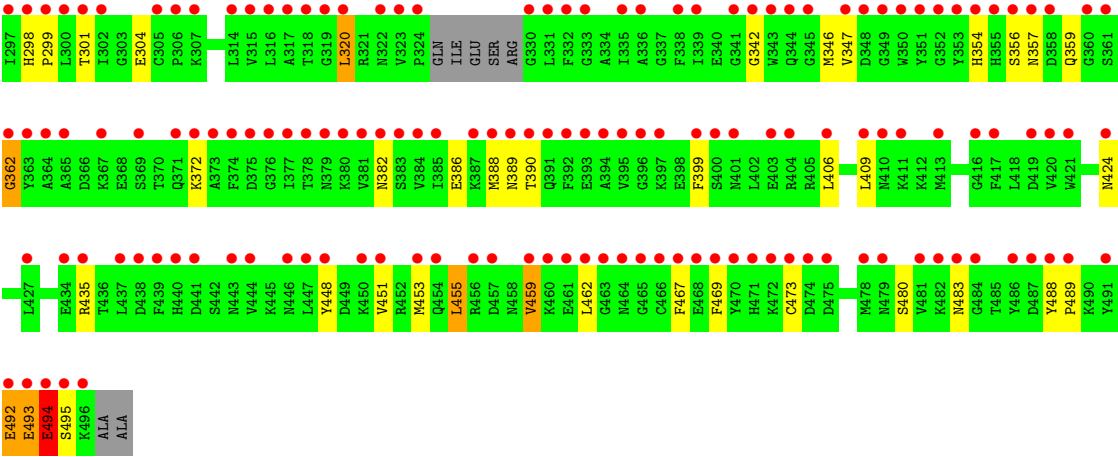
#### • Molecule 1: HEMAGGLUTININ



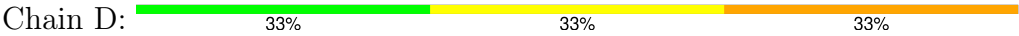


• Molecule 1: HEMAGGLUTININ

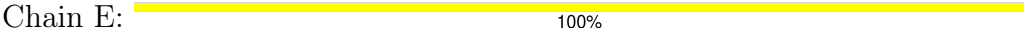




● Molecule 2: N-acetyl-alpha-neuraminic acid-(2-3)-beta-D-galactopyranose-(1-3)-2-acetamido-2-deoxy-beta-D-glucopyranose



● Molecule 2: N-acetyl-alpha-neuraminic acid-(2-3)-beta-D-galactopyranose-(1-3)-2-acetamido-2-deoxy-beta-D-glucopyranose



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	69.88Å 142.72Å 199.66Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.94 – 2.40 29.94 – 2.40	Depositor EDS
% Data completeness (in resolution range)	99.9 (29.94-2.40) 99.8 (29.94-2.40)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.33 (at 2.39Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, $R_{free}$	0.210 , 0.248 0.447 , 0.456	Depositor DCC
$R_{free}$ test set	3956 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	50.8	Xtriage
Anisotropy	0.428	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 31.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.77	EDS
Total number of atoms	12274	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	59.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.16% 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 [i](#)

### 5.1 Standard geometry [i](#)

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

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	0.33	0/3968	0.51	0/5372
1	B	0.33	0/3970	0.53	0/5372
1	C	0.33	0/3927	0.51	0/5313
All	All	0.33	0/11865	0.52	0/16057

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	3882	0	3742	85	2
1	B	3887	0	3756	109	2
1	C	3844	0	3718	67	0
2	D	46	0	40	2	0
2	E	46	0	40	0	0
3	A	27	0	25	7	0
3	C	14	0	13	3	0
4	A	185	0	0	5	0
4	B	169	0	0	9	0
4	C	174	0	0	2	0
All	All	12274	0	11334	254	2

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 11.

The worst 5 of 254 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:325:GLN:NE2	4:B:2109:HOH:O	1.62	1.27
1:A:283:THR:HG22	1:A:285:LEU:H	1.18	1.05
1:A:148:ARG:HG2	1:A:148:ARG:HH11	1.32	0.94
1:B:325:GLN:HB3	4:B:2110:HOH:O	1.65	0.93
1:A:361:SER:HB3	1:A:362:GLY:HA3	1.49	0.92

All (2) 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:A:141:LEU:CD2	1:B:128:GLN:OE1[1_655]	1.82	0.38
1:A:141:LEU:CD1	1:B:128:GLN:OE1[1_655]	2.03	0.17

## 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	486/509 (96%)	459 (94%)	21 (4%)	6 (1%)	11	16
1	B	485/509 (95%)	456 (94%)	20 (4%)	9 (2%)	6	8
1	C	479/509 (94%)	457 (95%)	16 (3%)	6 (1%)	10	15
All	All	1450/1527 (95%)	1372 (95%)	57 (4%)	21 (1%)	9	13

5 of 21 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	113	THR
1	A	494	GLU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	277	CYS
1	B	490	LYS
1	B	494	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	427/447 (96%)	413 (97%)	14 (3%)	33	53
1	B	429/447 (96%)	408 (95%)	21 (5%)	21	36
1	C	425/447 (95%)	409 (96%)	16 (4%)	28	47
All	All	1281/1341 (96%)	1230 (96%)	51 (4%)	27	45

5 of 51 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	389	ASN
1	C	94	CYS
1	C	493	GLU
1	B	434	GLU
1	B	491	TYR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 36 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	174	GLN
1	C	479	ASN
1	C	182	HIS
1	C	389	ASN
1	B	182	HIS

### 5.3.3 RNA [i](#)

There are no RNA molecules 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 ⓘ

6 monosaccharides 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	D	1	2	15,15,15	0.91	1 (6%)	21,21,21	0.74	0
2	GAL	D	2	2	11,11,12	0.74	0	15,15,17	0.67	0
2	SIA	D	3	2	20,20,21	0.90	0	21,28,31	0.92	1 (4%)
2	NAG	E	1	2	15,15,15	0.89	0	21,21,21	1.10	2 (9%)
2	GAL	E	2	2	11,11,12	0.60	0	15,15,17	0.99	1 (6%)
2	SIA	E	3	2	20,20,21	0.98	1 (5%)	21,28,31	0.90	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
2	NAG	D	1	2	-	2/6/26/26	0/1/1/1
2	GAL	D	2	2	-	2/2/19/22	0/1/1/1
2	SIA	D	3	2	-	0/18/34/38	0/1/1/1
2	NAG	E	1	2	-	1/6/26/26	0/1/1/1
2	GAL	E	2	2	-	0/2/19/22	0/1/1/1
2	SIA	E	3	2	-	0/18/34/38	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	E	3	SIA	C4-C5	2.17	1.55	1.53
2	D	1	NAG	C1-C2	2.04	1.55	1.52

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	E	2	GAL	C1-C2-C3	3.02	114.05	109.64
2	E	1	NAG	C3-C4-C5	2.72	115.16	110.23
2	D	3	SIA	O1B-C1-C2	2.28	118.64	112.71
2	E	1	NAG	O5-C5-C4	2.06	113.41	109.70

There are no chirality outliers.

All (5) torsion outliers are listed below:

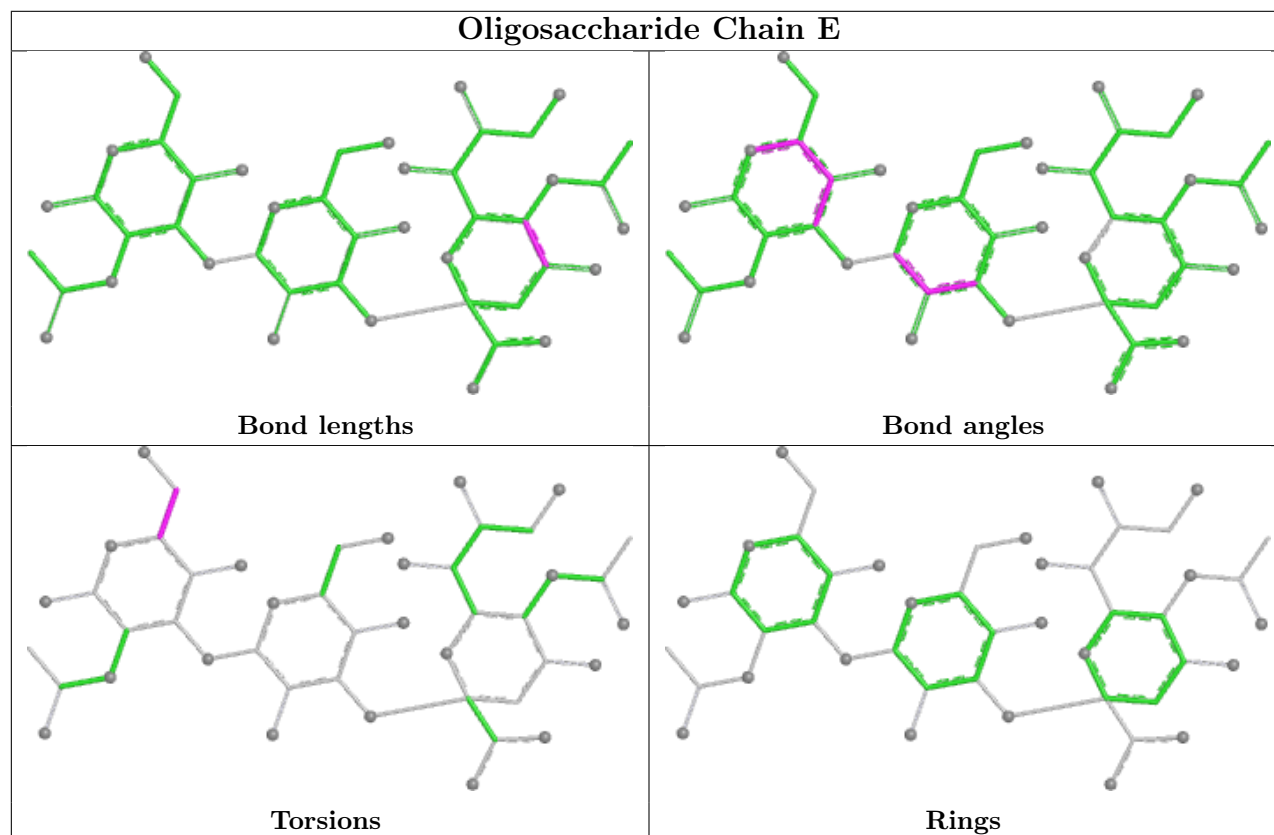
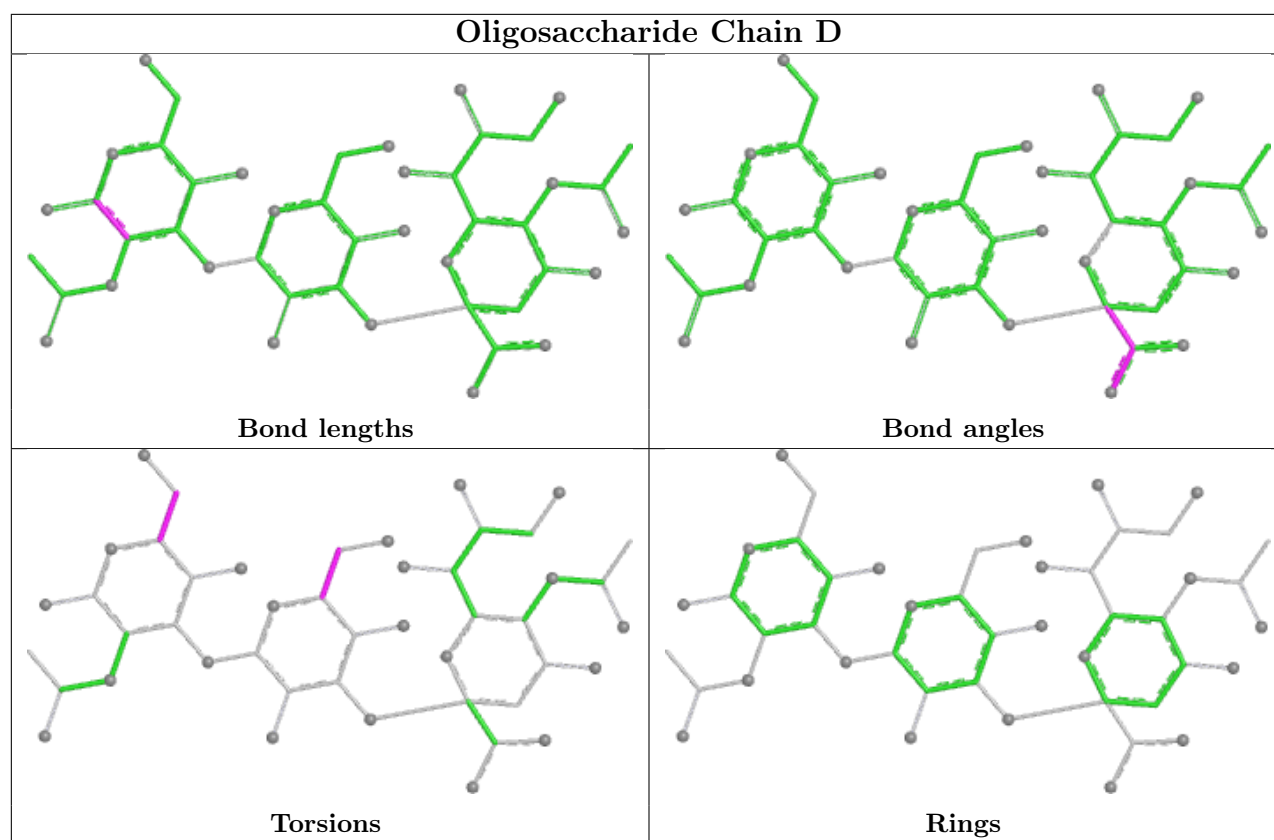
Mol	Chain	Res	Type	Atoms
2	D	2	GAL	C4-C5-C6-O6
2	D	2	GAL	O5-C5-C6-O6
2	D	1	NAG	O5-C5-C6-O6
2	E	1	NAG	O5-C5-C6-O6
2	D	1	NAG	C4-C5-C6-O6

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	D	3	SIA	2	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for oligosaccharide.



## 5.6 Ligand geometry

3 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	C	1500	-	14,14,15	0.56	0	17,19,21	0.90	0
3	NAG	A	1500	-	12,12,15	0.45	0	14,15,21	0.88	0
3	NAG	A	1499	-	14,14,15	0.46	0	17,19,21	0.75	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	C	1500	-	-	4/6/23/26	0/1/1/1
3	NAG	A	1500	-	-	6/15/15/26	-
3	NAG	A	1499	-	-	4/6/23/26	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 14 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	1499	NAG	C1-C2-N2-C7
3	A	1500	NAG	C1-C2-N2-C7
3	A	1500	NAG	O4-C4-C5-C6
3	A	1500	NAG	C8-C7-N2-C2
3	A	1500	NAG	O7-C7-N2-C2

There are no ring outliers.

3 monomers are involved in 10 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	C	1500	NAG	3	0
3	A	1500	NAG	6	0
3	A	1499	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](#)

**Warning:** The R factor obtained from EDS is 0.4472, which does not match the depositor's R factor of 0.2101. Please interpret the results in this section carefully.

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	490/509 (96%)	3.32	424 (86%) 0 0	29, 54, 94, 134	1 (0%)
1	B	490/509 (96%)	2.46	301 (61%) 0 0	30, 55, 95, 159	1 (0%)
1	C	485/509 (95%)	3.31	411 (84%) 0 0	36, 55, 87, 134	0
All	All	1465/1527 (95%)	3.03	1136 (77%) 0 0	29, 55, 92, 159	2 (0%)

The worst 5 of 1136 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	196	ASN	8.9
1	C	206	SER	7.9
1	A	9	ILE	7.7
1	C	66	LEU	7.5
1	C	120	ILE	7.4

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

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

### 6.3 Carbohydrates [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. 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.

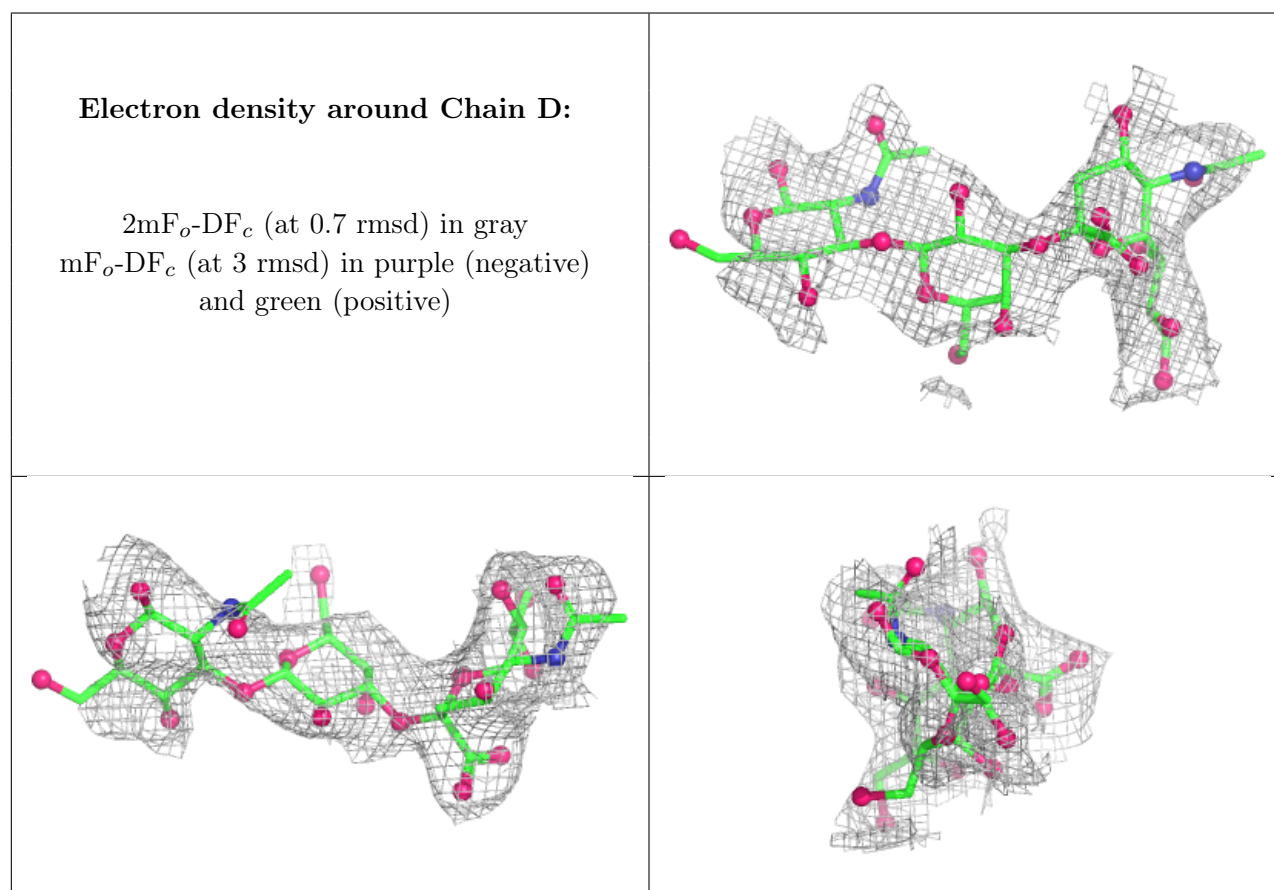
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
2	GAL	D	2	11/12	0.55	0.18	82,84,85,85	0

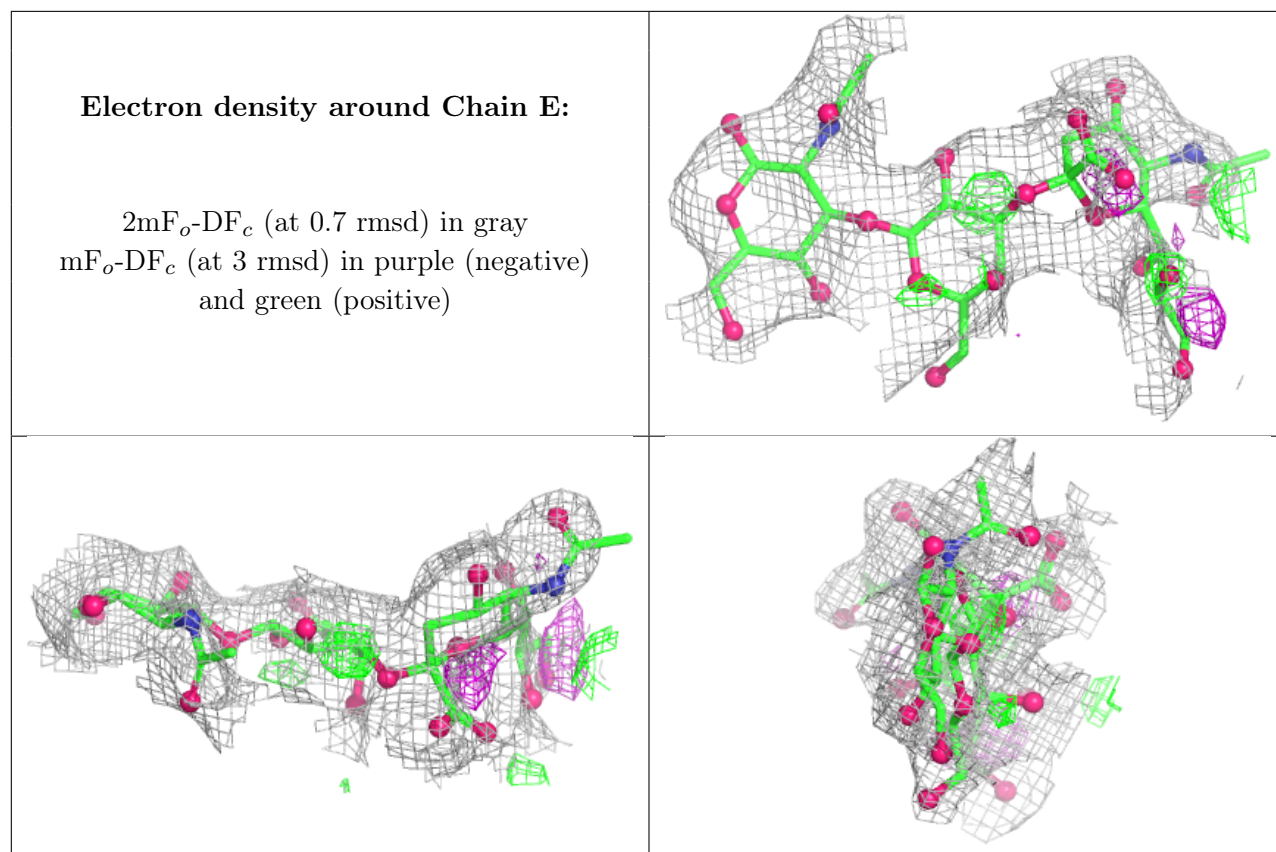
*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
2	NAG	E	1	15/15	0.56	0.20	70,74,74,75	0
2	NAG	D	1	15/15	0.60	0.19	86,89,89,90	0
2	SIA	E	3	20/21	0.63	0.24	54,58,60,61	0
2	GAL	E	2	11/12	0.73	0.25	63,66,67,68	0
2	SIA	D	3	20/21	0.76	0.15	77,79,81,81	0

The following is a graphical depiction of the model fit to experimental electron density for oligosaccharide. Each fit is shown from different orientation to approximate a three-dimensional view.





## 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( $\text{\AA}^2$ )	Q<0.9
3	NAG	A	1499	14/15	0.43	0.26	111,114,114,115	0
3	NAG	C	1500	14/15	0.47	0.26	78,79,80,80	0
3	NAG	A	1500	13/15	0.68	0.19	69,70,71,72	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.