



# Full wwPDB X-ray Structure Validation Report ⓘ

Mar 9, 2026 – 03:36 PM UTC

PDB ID : 9GEG / pdb\_00009geg  
Title : Entamoeba histolytica Gal/GalNAc lectin heavy chain residues 808-992 in the presence of GalNAc  
Authors : Gerard, S.F.; Higgins, M.K.  
Deposited on : 2024-08-07  
Resolution : 1.85 Å(reported)

This is a Full wwPDB X-ray Structure 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

<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	:	4-5-2 with Phenix2.0
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.49

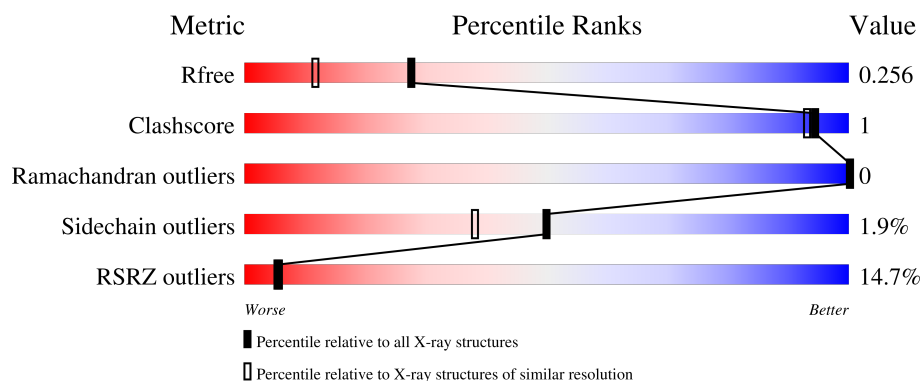
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.85 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))
$R_{free}$	180053	3428 (1.86-1.86)
Clashscore	190562	3579 (1.86-1.86)
Ramachandran outliers	187476	3553 (1.86-1.86)
Sidechain outliers	187428	3553 (1.86-1.86)
RSRZ outliers	180081	3429 (1.86-1.86)

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	189	<div> <div>18%</div> <div> <div></div> <div>90%</div> <div>6% • •</div> </div> </div>
1	B	189	<div> <div>13%</div> <div> <div></div> <div>89%</div> <div>6% • 5%</div> </div> </div>
1	C	189	<div> <div>11%</div> <div> <div></div> <div>92%</div> <div>• 5%</div> </div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 4749 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 Galactose/N-acetyl-D-galactosamine lectin heavy subunit 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	185	Total	C	N	O	S	0	0	0
			1491	933	241	302	15			
1	B	180	Total	C	N	O	S	0	0	0
			1457	914	235	293	15			
1	C	179	Total	C	N	O	S	0	0	0
			1450	909	234	292	15			

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	811	SER	CYS	conflict	UNP P32022
A	972	SER	CYS	conflict	UNP P32022
A	993	GLY	THR	conflict	UNP P32022
B	811	SER	CYS	conflict	UNP P32022
B	972	SER	CYS	conflict	UNP P32022
B	993	GLY	THR	conflict	UNP P32022
C	811	SER	CYS	conflict	UNP P32022
C	972	SER	CYS	conflict	UNP P32022
C	993	GLY	THR	conflict	UNP P32022

- Molecule 2 is 2-acetamido-2-deoxy-beta-D-glucopyranose (CCD ID: 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	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		

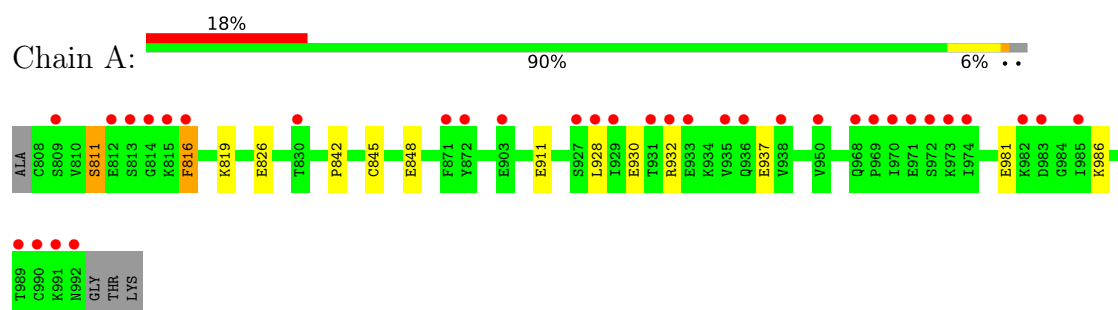
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	92	Total	O	0	0
			92	92		
3	B	123	Total	O	0	0
			123	123		
3	C	94	Total	O	0	0
			94	94		

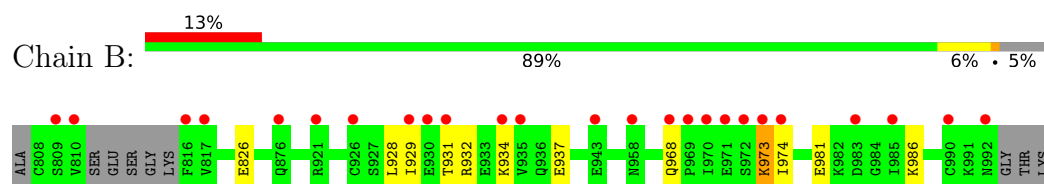
### 3 Residue-property plots [i](#)

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.

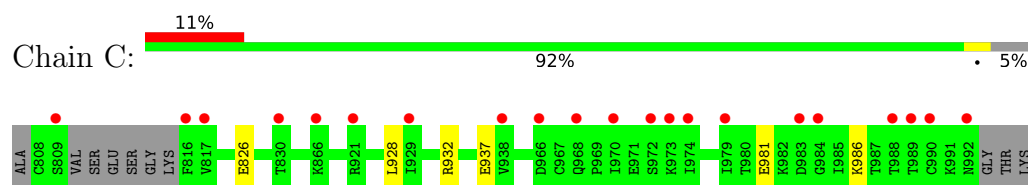
- Molecule 1: Galactose/N-acetyl-D-galactosamine lectin heavy subunit 1



- Molecule 1: Galactose/N-acetyl-D-galactosamine lectin heavy subunit 1



- Molecule 1: Galactose/N-acetyl-D-galactosamine lectin heavy subunit 1



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	83.02Å 146.31Å 85.09Å 90.00° 113.84° 90.00°	Depositor
Resolution (Å)	41.34 – 1.85 41.34 – 1.85	Depositor EDS
% Data completeness (in resolution range)	97.3 (41.34-1.85) 97.3 (41.34-1.85)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.99 (at 1.86Å)	Xtriage
Refinement program	BUSTER 2.10.4 (26-JUL-2023)	Depositor
R, $R_{free}$	0.247 , 0.269 0.236 , 0.256	Depositor DCC
$R_{free}$ test set	3814 reflections (4.83%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	41.6	Xtriage
Anisotropy	0.351	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 47.5	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	0.96	EDS
Total number of atoms	4749	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	62.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.99% 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

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	0.76	0/1517	1.03	5/2045 (0.2%)
1	B	0.78	0/1482	1.00	1/1998 (0.1%)
1	C	0.74	0/1475	0.97	0/1988
All	All	0.76	0/4474	1.00	6/6031 (0.1%)

There are no bond length outliers.

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	816	PHE	CA-CB-CG	7.49	121.29	113.80
1	A	911	GLU	CB-CG-CD	7.03	124.54	112.60
1	B	968	GLN	CB-CA-C	5.91	116.59	109.85
1	A	930	GLU	CA-C-N	5.71	128.21	120.38
1	A	930	GLU	C-N-CA	5.71	128.21	120.38
1	A	848	GLU	CB-CG-CD	5.17	121.40	112.60

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	1491	0	1437	4	0
1	B	1457	0	1404	5	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	1450	0	1395	2	0
2	A	14	0	13	0	0
2	B	14	0	13	0	0
2	C	14	0	13	0	0
3	A	92	0	0	0	0
3	B	123	0	0	0	0
3	C	94	0	0	0	0
All	All	4749	0	4275	11	0

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 1.

All (11) 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:A:928:LEU:HD13	1:A:932:ARG:HE	1.51	0.74
1:A:981:GLU:HG2	1:A:986:LYS:HG2	1.80	0.64
1:C:981:GLU:HG2	1:C:986:LYS:HG2	1.81	0.62
1:B:981:GLU:HG2	1:B:986:LYS:HG2	1.82	0.61
1:B:929:ILE:HD12	1:B:934:LYS:HB2	1.92	0.52
1:B:928:LEU:HD13	1:B:932:ARG:HE	1.77	0.49
1:B:929:ILE:HB	1:B:931:THR:HG22	1.94	0.49
1:C:928:LEU:HD13	1:C:932:ARG:HD2	2.02	0.42
1:A:819:LYS:HD2	1:A:842:PRO:HB3	2.03	0.41
1:B:973:LYS:HD2	1:B:973:LYS:HA	1.92	0.40
1:A:811:SER:OG	1:A:845:CYS:HB3	2.21	0.40

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	183/189 (97%)	180 (98%)	3 (2%)	0	100	100
1	B	176/189 (93%)	174 (99%)	2 (1%)	0	100	100
1	C	175/189 (93%)	174 (99%)	1 (1%)	0	100	100
All	All	534/567 (94%)	528 (99%)	6 (1%)	0	100	100

There are no Ramachandran outliers to report.

### 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	174/176 (99%)	170 (98%)	4 (2%)	44	30
1	B	170/176 (97%)	166 (98%)	4 (2%)	43	28
1	C	169/176 (96%)	167 (99%)	2 (1%)	63	55
All	All	513/528 (97%)	503 (98%)	10 (2%)	50	38

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	811	SER
1	A	816	PHE
1	A	826	GLU
1	A	937	GLU
1	B	826	GLU
1	B	937	GLU
1	B	973	LYS
1	B	974	ILE
1	C	826	GLU
1	C	937	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	964	ASN
1	B	964	ASN
1	C	964	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](#)

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

## 5.5 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

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	$\# Z  > 2$
2	NAG	B	1001	1	14,14,15	0.27	0	17,19,21	0.54	0
2	NAG	C	1001	1	14,14,15	0.26	0	17,19,21	0.56	0
2	NAG	A	1001	1	14,14,15	0.26	0	17,19,21	0.57	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	B	1001	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
2	NAG	C	1001	1	-	0/6/23/26	0/1/1/1
2	NAG	A	1001	1	-	0/6/23/26	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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

### 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	185/189 (97%)	1.27	34 (18%) <b>3</b> <b>3</b>	41, 62, 97, 116	0
1	B	180/189 (95%)	0.95	25 (13%) <b>6</b> <b>6</b>	43, 55, 85, 103	0
1	C	179/189 (94%)	0.99	21 (11%) <b>9</b> <b>8</b>	39, 62, 87, 103	0
All	All	544/567 (95%)	1.07	80 (14%) <b>6</b> <b>5</b>	39, 60, 92, 116	0

All (80) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	810	VAL	8.6
1	A	929	ILE	6.8
1	C	816	PHE	6.5
1	B	816	PHE	6.4
1	A	974	ILE	5.5
1	A	813	SER	5.2
1	B	931	THR	5.1
1	A	970	ILE	5.0
1	A	816	PHE	5.0
1	A	931	THR	4.9
1	B	974	ILE	4.7
1	B	970	ILE	4.7
1	B	935	VAL	4.4
1	A	815	LYS	4.4
1	A	812	GLU	4.3
1	A	972	SER	4.3
1	B	968	GLN	4.1
1	B	929	ILE	3.9
1	A	814	GLY	3.9
1	A	989	THR	3.7
1	A	928	LEU	3.6
1	C	970	ILE	3.6
1	B	973	LYS	3.6

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	B	972	SER	3.5
1	C	974	ILE	3.4
1	A	932	ARG	3.4
1	A	969	PRO	3.4
1	C	989	THR	3.3
1	B	926	CYS	3.2
1	B	969	PRO	3.2
1	C	968	GLN	3.1
1	A	938	VAL	3.1
1	C	817	VAL	3.0
1	B	817	VAL	3.0
1	C	984	GLY	2.9
1	B	934	LYS	2.9
1	A	933	GLU	2.9
1	A	991	LYS	2.9
1	B	809	SER	2.9
1	A	982	LYS	2.8
1	A	992	ASN	2.8
1	A	950	VAL	2.8
1	B	992	ASN	2.8
1	A	927	SER	2.7
1	C	972	SER	2.7
1	C	809	SER	2.7
1	A	936	GLN	2.7
1	A	935	VAL	2.6
1	C	938	VAL	2.5
1	C	988	THR	2.5
1	C	983	ASP	2.5
1	A	809	SER	2.5
1	A	973	LYS	2.4
1	C	830	THR	2.4
1	A	871	PHE	2.4
1	C	866	LYS	2.4
1	B	921	ARG	2.4
1	B	971	GLU	2.4
1	C	990	CYS	2.4
1	C	973	LYS	2.4
1	A	990	CYS	2.4
1	B	990	CYS	2.4
1	A	983	ASP	2.3
1	A	830	THR	2.3
1	B	983	ASP	2.3

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	903	GLU	2.3
1	C	929	ILE	2.2
1	C	979	ILE	2.2
1	B	930	GLU	2.2
1	A	985	ILE	2.2
1	B	876	GLN	2.2
1	A	971	GLU	2.2
1	C	966	ASP	2.2
1	C	992	ASN	2.1
1	B	943	GLU	2.1
1	A	968	GLN	2.1
1	B	958	ASN	2.1
1	B	985	ILE	2.1
1	C	921	ARG	2.1
1	A	872	TYR	2.0

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

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

## 6.3 Carbohydrates [i](#)

There are no oligosaccharides 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( $\text{\AA}^2$ )	Q<0.9
2	NAG	C	1001	14/15	0.89	0.12	47,50,55,56	0
2	NAG	B	1001	14/15	0.91	0.11	62,64,67,68	0
2	NAG	A	1001	14/15	0.91	0.13	54,57,59,61	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.