



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 2, 2017 – 04:57 PM EDT

PDB ID : 5U73  
Title : Crystal structure of human Niemann-Pick C1 protein  
Authors : Li, X.; Wang, J.; Blobel, G.  
Deposited on : unknown  
Resolution : 3.35 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity	:	4.02b-467
Mogul	:	1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030345
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20030345

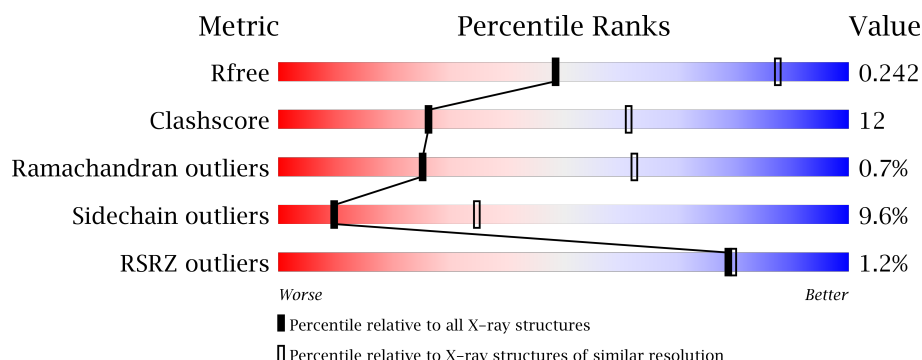
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.35 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	100719	1167 (3.40-3.28)
Clashscore	112137	1239 (3.40-3.28)
Ramachandran outliers	110173	1219 (3.40-3.28)
Sidechain outliers	110143	1218 (3.40-3.28)
RSRZ outliers	101464	1176 (3.40-3.28)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	1278	

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 7262 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.

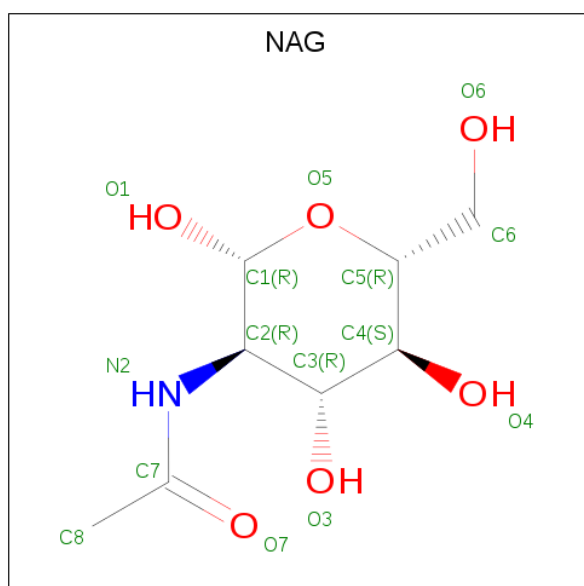
- Molecule 1 is a protein called Niemann-Pick C1 protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	898	7044	4581	1136	1281	46	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	642	ILE	MET	conflict	UNP O15118

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



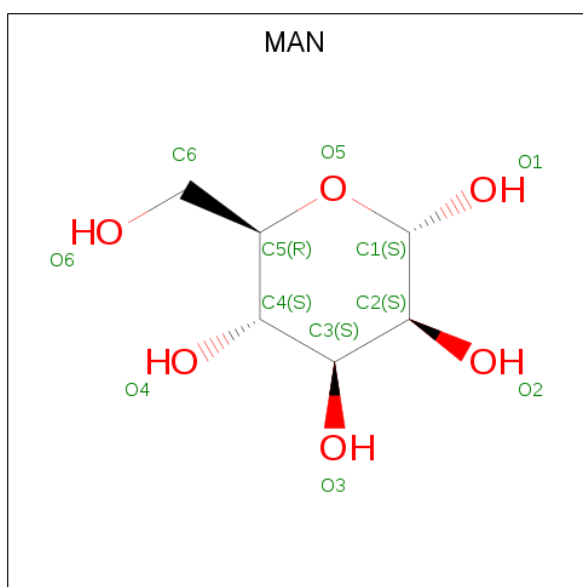
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	A	1	14	8	1	5	0	0
2	A	1	14	8	1	5	0	0
2	A	1	14	8	1	5	0	0

*Continued on next page...*

*Continued from previous page...*

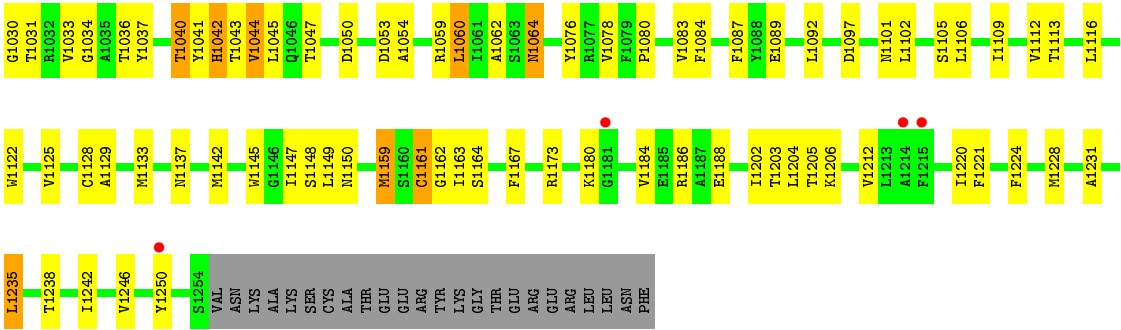
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			14	8	1	5		
2	A	1	Total	C	N	O	0	0
			14	8	1	5		
2	A	1	Total	C	N	O	0	0
			14	8	1	5		
2	A	1	Total	C	N	O	0	0
			14	8	1	5		
2	A	1	Total	C	N	O	0	0
			14	8	1	5		
2	A	1	Total	C	N	O	0	0
			14	8	1	5		
2	A	1	Total	C	N	O	0	0
			14	8	1	5		
2	A	1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

- Molecule 3 is ALPHA-D-MANNOSE (three-letter code: MAN) (formula: C<sub>6</sub>H<sub>12</sub>O<sub>6</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			11	6	5		
3	A	1	Total	C	O	0	0
			11	6	5		





## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	181.94Å 223.07Å 63.21Å 90.00° 105.31° 90.00°	Depositor
Resolution (Å)	41.15 – 3.35 41.15 – 3.35	Depositor EDS
% Data completeness (in resolution range)	51.5 (41.15-3.35) 51.5 (41.15-3.35)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.45 (at 3.32Å)	Xtriage
Refinement program	PHENIX (1.11.1 _2575: ???)	Depositor
R, $R_{free}$	0.233 , 0.258 0.227 , 0.242	Depositor DCC
$R_{free}$ test set	872 reflections (4.86%)	DCC
Wilson B-factor (Å <sup>2</sup> )	24.8	Xtriage
Anisotropy	0.617	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.26 , 31.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.85	EDS
Total number of atoms	7262	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	51.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.40% 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, 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	0.26	0/7221	0.44	0/9827

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
1	A	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	536	THR	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	7044	0	6960	168	0
2	A	196	0	176	3	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	22	0	20	1	0
All	All	7262	0	7156	169	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 12.

The worst 5 of 169 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:1231:ALA:O	1:A:1235:LEU:HB2	1.74	0.86
1:A:404:ARG:NH1	1:A:584:GLU:OE2	2.11	0.82
1:A:906:ASN:OD1	1:A:989:ARG:NH1	2.16	0.78
1:A:914:CYS:SG	1:A:989:ARG:NH2	2.56	0.77
1:A:635:ILE:HD11	1:A:1204:LEU:HD13	1.66	0.76

There are no symmetry-related clashes.

## 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	892/1278 (70%)	788 (88%)	98 (11%)	6 (1%)	<span style="border: 1px solid red; padding: 2px;">25</span> <span style="border: 1px solid blue; padding: 2px;">62</span>

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	934	ARG
1	A	529	LEU
1	A	536	THR
1	A	537	PHE
1	A	987	LYS

### 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	782/1109 (70%)	707 (90%)	75 (10%)	10	35

5 of 75 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	747	PHE
1	A	874	ASP
1	A	1159	MET
1	A	754	MET
1	A	823	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	447	GLN
1	A	490	ASN
1	A	578	GLN
1	A	775	GLN

### 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 carbohydrates in this entry.

## 5.6 Ligand geometry

16 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	NAG	A	1301	1	14,14,15	0.24	0	15,19,21	0.44	0
2	NAG	A	1302	1	14,14,15	0.25	0	15,19,21	0.52	0
2	NAG	A	1303	1,2	14,14,15	0.92	1 (7%)	15,19,21	0.87	1 (6%)
2	NAG	A	1304	2	14,14,15	0.32	0	15,19,21	0.66	0
2	NAG	A	1305	1	14,14,15	0.23	0	15,19,21	0.59	0
2	NAG	A	1306	1,2	14,14,15	0.86	1 (7%)	15,19,21	0.72	1 (6%)
2	NAG	A	1307	3,2	14,14,15	1.23	2 (14%)	15,19,21	1.33	1 (6%)
3	MAN	A	1308	2	11,11,12	1.67	3 (27%)	13,15,17	1.88	5 (38%)
2	NAG	A	1309	1	14,14,15	0.22	0	15,19,21	0.41	0
2	NAG	A	1310	1	14,14,15	0.50	0	15,19,21	0.79	1 (6%)
2	NAG	A	1311	1,2	14,14,15	0.95	2 (14%)	15,19,21	0.81	0
2	NAG	A	1312	3,2	14,14,15	1.11	2 (14%)	15,19,21	1.54	2 (13%)
3	MAN	A	1313	2	11,11,12	1.62	2 (18%)	13,15,17	2.00	4 (30%)
2	NAG	A	1314	1	14,14,15	0.38	0	15,19,21	0.45	0
2	NAG	A	1315	1	14,14,15	0.78	1 (7%)	15,19,21	1.16	1 (6%)
2	NAG	A	1316	1	14,14,15	1.02	1 (7%)	15,19,21	0.94	1 (6%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	NAG	A	1301	1	-	0/6/23/26	0/1/1/1
2	NAG	A	1302	1	-	0/6/23/26	0/1/1/1
2	NAG	A	1303	1,2	-	0/6/23/26	0/1/1/1
2	NAG	A	1304	2	-	0/6/23/26	0/1/1/1
2	NAG	A	1305	1	-	0/6/23/26	0/1/1/1
2	NAG	A	1306	1,2	-	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	A	1307	3,2	-	0/6/23/26	0/1/1/1
3	MAN	A	1308	2	-	0/2/19/22	0/1/1/1
2	NAG	A	1309	1	-	0/6/23/26	0/1/1/1
2	NAG	A	1310	1	-	0/6/23/26	0/1/1/1
2	NAG	A	1311	1,2	-	0/6/23/26	0/1/1/1
2	NAG	A	1312	3,2	-	0/6/23/26	0/1/1/1
3	MAN	A	1313	2	-	0/2/19/22	0/1/1/1
2	NAG	A	1314	1	-	0/6/23/26	0/1/1/1
2	NAG	A	1315	1	-	0/6/23/26	0/1/1/1
2	NAG	A	1316	1	-	0/6/23/26	0/1/1/1

The worst 5 of 15 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1316	NAG	O5-C1	-3.39	1.38	1.43
3	A	1308	MAN	C4-C5	2.18	1.57	1.53
2	A	1306	NAG	C1-C2	2.41	1.55	1.52
2	A	1311	NAG	C1-C2	2.42	1.55	1.52
3	A	1308	MAN	O5-C5	2.44	1.48	1.43

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1312	NAG	O4-C4-C5	-4.90	96.93	109.28
2	A	1307	NAG	O4-C4-C5	-4.79	97.21	109.28
3	A	1313	MAN	C1-C2-C3	-4.55	103.88	109.65
3	A	1308	MAN	C1-C2-C3	-2.96	105.90	109.65
3	A	1308	MAN	O2-C2-C3	-2.84	104.59	110.17

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1311	NAG	1	0
2	A	1312	NAG	2	0
3	A	1313	MAN	1	0
2	A	1314	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	898/1278 (70%)	-0.43	11 (1%) 79 80	8, 47, 93, 144	0

The worst 5 of 11 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1029	HIS	3.6
1	A	816	CYS	3.4
1	A	1215	PHE	3.0
1	A	1214	ALA	2.8
1	A	717	GLY	2.7

### 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 carbohydrates in this entry.

### 6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q< 0.9’ lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
2	NAG	A	1306	14/15	0.92	0.20	0.21	34,48,57,60	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	NAG	A	1309	14/15	0.94	0.20	-0.07	40,53,72,74	0
2	NAG	A	1314	14/15	0.74	0.49	-	70,76,93,96	0
2	NAG	A	1304	14/15	0.88	0.41	-	77,86,91,100	0
2	NAG	A	1301	14/15	0.87	0.27	-	72,83,89,90	0
2	NAG	A	1311	14/15	0.92	0.43	-	37,78,88,99	0
2	NAG	A	1312	14/15	0.83	0.62	-	107,113,118,121	0
3	MAN	A	1313	11/12	0.25	0.59	-	108,118,127,127	0
2	NAG	A	1307	14/15	0.94	0.32	-	35,58,65,68	0
2	NAG	A	1315	14/15	0.81	0.29	-	48,74,105,108	0
2	NAG	A	1303	14/15	0.87	0.25	-	28,50,62,72	0
2	NAG	A	1316	14/15	0.80	0.34	-	63,82,93,94	0
2	NAG	A	1302	14/15	0.45	0.58	-	114,124,126,128	0
3	MAN	A	1308	11/12	0.86	0.33	-	48,63,73,77	0
2	NAG	A	1310	14/15	0.90	0.38	-	65,91,94,94	0
2	NAG	A	1305	14/15	0.88	0.33	-	62,70,87,91	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.