



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jul 24, 2018 – 10:26 AM EDT

PDB ID : 5Y1K  
Title : Crystal structure of Plasmodium falciparum aminopeptidase N in complex with (S)-2-(3-(2-chlorobenzyl)ureido)-N-hydroxy-4-methylpentanamide  
Authors : Marapaka, A.K.; Zhang, Y.; Addlagatta, A.  
Deposited on : 2017-07-20  
Resolution : 1.81 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.7.3 (157068), CSD as539be (2018)  
Xtriage (Phenix) : 1.13  
EDS : rb-20031172  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Refmac : 5.8.0158  
CCP4 : 7.0 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20031172

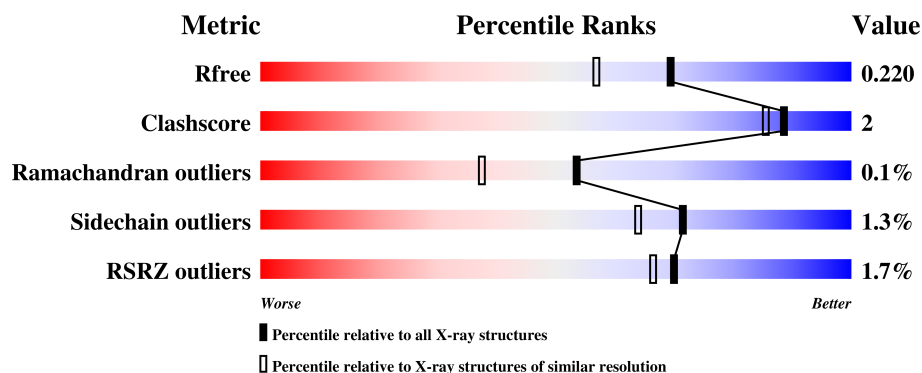
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.81 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))
$R_{free}$	111664	6455 (1.84-1.80)
Clashscore	122126	7349 (1.84-1.80)
Ramachandran outliers	120053	7272 (1.84-1.80)
Sidechain outliers	120020	7272 (1.84-1.80)
RSRZ outliers	108989	6347 (1.84-1.80)

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	914	<div> <div>2%</div> <div>90%</div> <div>7%</div> <div>.</div> </div>

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 7872 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 M1 family aminopeptidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	890	7350	4720	1197	1405	28	0	9	0

There are 23 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	172	MET	-	expression tag	UNP O96935
A	173	GLY	-	expression tag	UNP O96935
A	174	SER	-	expression tag	UNP O96935
A	175	SER	-	expression tag	UNP O96935
A	176	HIS	-	expression tag	UNP O96935
A	177	HIS	-	expression tag	UNP O96935
A	178	HIS	-	expression tag	UNP O96935
A	179	HIS	-	expression tag	UNP O96935
A	180	HIS	-	expression tag	UNP O96935
A	181	HIS	-	expression tag	UNP O96935
A	182	SER	-	expression tag	UNP O96935
A	183	SER	-	expression tag	UNP O96935
A	184	GLY	-	expression tag	UNP O96935
A	185	LEU	-	expression tag	UNP O96935
A	186	VAL	-	expression tag	UNP O96935
A	187	PRO	-	expression tag	UNP O96935
A	188	ARG	-	expression tag	UNP O96935
A	189	GLY	-	expression tag	UNP O96935
A	190	SER	-	expression tag	UNP O96935
A	191	HIS	-	expression tag	UNP O96935
A	192	MET	-	expression tag	UNP O96935
A	193	ALA	-	expression tag	UNP O96935
A	194	SER	-	expression tag	UNP O96935

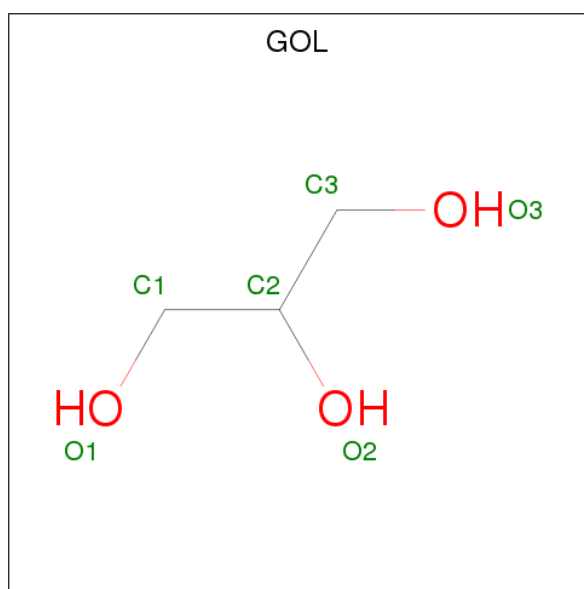
- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Zn	0	0
			1	1		

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

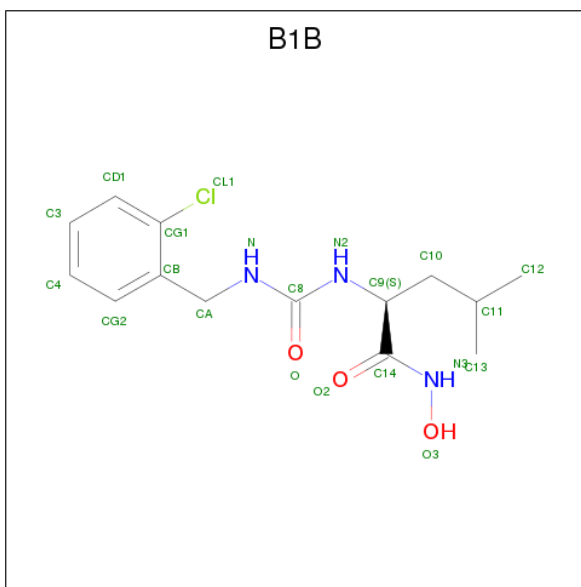
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	3	Total	Mg	0	0
			3	3		

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



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

- Molecule 5 is (2S)-2-[(2-chlorophenyl)methylcarbamoylamino]-4-methyl-N-oxidanyl-pentanamide (three-letter code: B1B) (formula: C<sub>14</sub>H<sub>20</sub>ClN<sub>3</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	A	1	Total	C	Cl	N	O	0	0
			21	14	1	3	3		

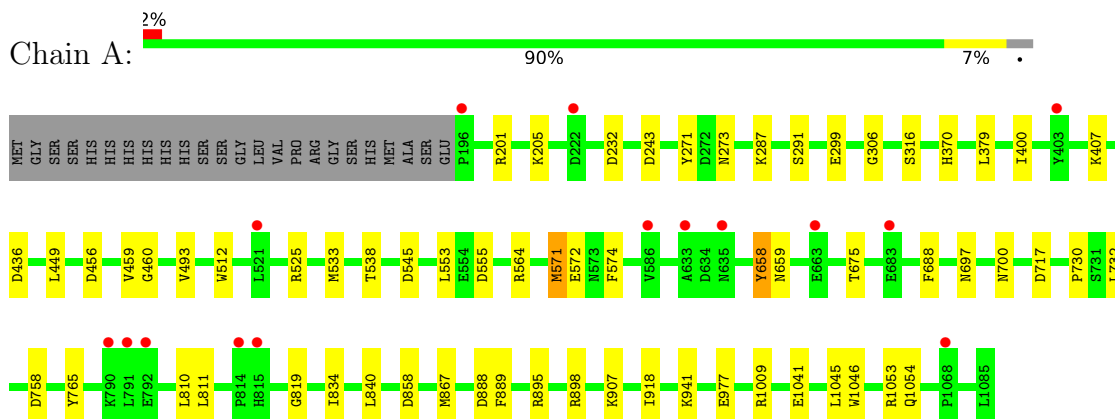
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	461	Total	O	0	0
			461	461		

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

These plots are drawn for all protein, RNA and DNA 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: M1 family aminopeptidase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	74.28Å 109.25Å 111.88Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	23.71 – 1.81 23.71 – 1.81	Depositor EDS
% Data completeness (in resolution range)	98.4 (23.71-1.81) 98.5 (23.71-1.81)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	10.18 (at 1.81Å)	Xtriage
Refinement program	REFMAC 5.8.0135 : 01/10/15	Depositor
R, $R_{free}$	0.170 , 0.213 0.180 , 0.220	Depositor DCC
$R_{free}$ test set	4002 reflections (4.86%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	25.3	Xtriage
Anisotropy	0.052	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 42.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.52$ , $\langle L^2 \rangle = 0.35$	Xtriage
Estimated twinning fraction	0.002 for -h,l,k	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	7872	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	27.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.09% 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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, ZN, B1B, MG

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	0.98	4/7540 (0.1%)	0.97	19/10190 (0.2%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	449	LEU	N-CA	7.79	1.61	1.46
1	A	564	ARG	CZ-NH1	7.16	1.42	1.33
1	A	299	GLU	CG-CD	6.23	1.61	1.51
1	A	658	TYR	CE1-CZ	-5.02	1.32	1.38

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	564	ARG	NE-CZ-NH2	-11.88	114.36	120.30
1	A	564	ARG	NE-CZ-NH1	11.81	126.21	120.30
1	A	898	ARG	NE-CZ-NH1	8.83	124.71	120.30
1	A	895	ARG	NE-CZ-NH1	6.75	123.67	120.30
1	A	898	ARG	NE-CZ-NH2	-6.65	116.97	120.30

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-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 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	7350	0	7294	31	0
2	A	1	0	0	0	0
3	A	3	0	0	0	0
4	A	36	0	46	11	0
5	A	21	0	0	2	0
6	A	461	0	0	4	0
All	All	7872	0	7340	31	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 2.

The worst 5 of 31 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:291:SER:OG	6:A:1201:HOH:O	1.76	1.02
1:A:545:ASP:CG	6:A:1203:HOH:O	2.03	0.95
1:A:232[B]:ASP:OD1	6:A:1202:HOH:O	1.90	0.86
1:A:533:MET:O	4:A:1105:GOL:H11	1.92	0.69
1:A:460:GLY:H	5:A:1111:B1B:C12	2.09	0.64

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	897/914 (98%)	885 (99%)	11 (1%)	1 (0%)	53 37

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	459	VAL

### 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	831/842 (99%)	820 (99%)	11 (1%)	71 64

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	555	ASP
1	A	688	PHE
1	A	1009	ARG
1	A	436	ASP
1	A	977	GLU

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

### 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 [i](#)

Of 11 ligands modelled in this entry, 4 are monoatomic - leaving 7 for Mogul analysis.

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	GOL	A	1105	-	5,5,5	1.03	0	5,5,5	1.25	1 (20%)
4	GOL	A	1106	-	5,5,5	0.22	0	5,5,5	0.49	0
4	GOL	A	1107	-	5,5,5	0.76	0	5,5,5	0.66	0
4	GOL	A	1108	-	5,5,5	0.47	0	5,5,5	1.27	1 (20%)
4	GOL	A	1109	-	5,5,5	0.58	0	5,5,5	0.62	0
4	GOL	A	1110	-	5,5,5	1.21	1 (20%)	5,5,5	2.60	2 (40%)
5	B1B	A	1111	2	21,21,21	1.02	1 (4%)	26,27,27	2.67	14 (53%)

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	GOL	A	1105	-	-	0/4/4/4	0/0/0/0
4	GOL	A	1106	-	-	0/4/4/4	0/0/0/0
4	GOL	A	1107	-	-	0/4/4/4	0/0/0/0
4	GOL	A	1108	-	-	0/4/4/4	0/0/0/0
4	GOL	A	1109	-	-	0/4/4/4	0/0/0/0
4	GOL	A	1110	-	-	0/4/4/4	0/0/0/0
5	B1B	A	1111	2	-	0/19/19/19	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	1110	GOL	O2-C2	-2.39	1.36	1.43
5	A	1111	B1B	C8-N	2.38	1.40	1.35

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	1111	B1B	C10-C9-N2	-5.82	97.01	110.57
4	A	1110	GOL	O2-C2-C3	-4.75	87.28	109.00
5	A	1111	B1B	CB-CA-N	-4.61	102.99	113.05
5	A	1111	B1B	C9-C14-N3	-4.27	107.34	116.76
5	A	1111	B1B	CG2-CB-CG1	-4.00	110.11	116.90

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

6 monomers are involved in 13 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1105	GOL	3	0
4	A	1107	GOL	1	0
4	A	1108	GOL	2	0
4	A	1109	GOL	2	0
4	A	1110	GOL	3	0
5	A	1111	B1B	2	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	890/914 (97%)	-0.10	15 (1%) 70 66	17, 25, 40, 89	0

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	196	PRO	4.7
1	A	815	HIS	3.5
1	A	790	LYS	3.3
1	A	791	LEU	3.2
1	A	403	TYR	3.0

### 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 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. 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
4	GOL	A	1108	6/6	0.86	0.22	37,41,45,54	0
4	GOL	A	1109	6/6	0.86	0.33	37,45,51,63	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
4	GOL	A	1107	6/6	0.88	0.23	24,40,47,49	0
4	GOL	A	1106	6/6	0.91	0.13	52,53,55,55	0
4	GOL	A	1110	6/6	0.91	0.20	30,45,49,52	0
5	B1B	A	1111	21/21	0.92	0.20	22,34,86,106	0
3	MG	A	1103	1/1	0.94	0.17	31,31,31,31	0
4	GOL	A	1105	6/6	0.94	0.28	30,38,47,49	0
3	MG	A	1102	1/1	0.97	0.24	38,38,38,38	0
3	MG	A	1104	1/1	0.97	0.14	43,43,43,43	0
2	ZN	A	1101	1/1	1.00	0.02	20,20,20,20	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.