



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jul 24, 2018 – 10:24 AM EDT

PDB ID : 5Y1H  
Title : Crystal structure of Plasmodium falciparum aminopeptidase N in complex with (S)-2-(3-(2,4-difluorobenzyl)ureido)-N-hydroxy-4-methylpentanamide  
Authors : Marapaka, A.K.; Zhang, Y.; Addlagatta, A.  
Deposited on : 2017-07-20  
Resolution : 1.73 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

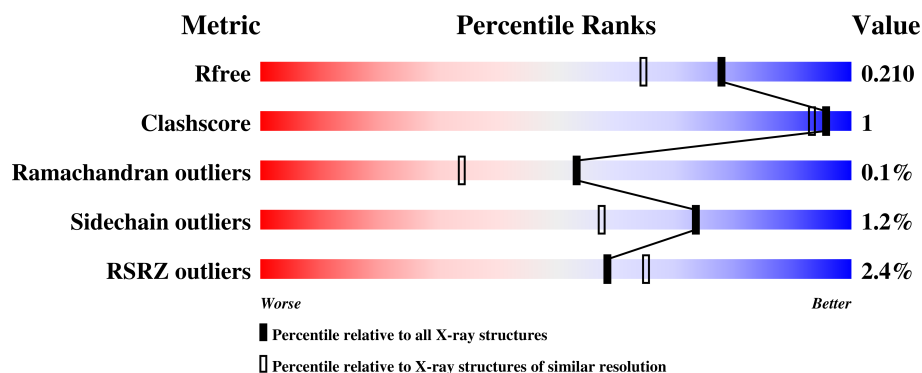
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

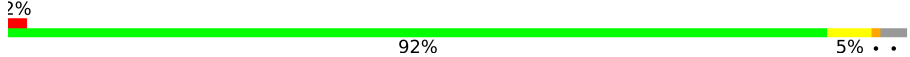
The reported resolution of this entry is 1.73 Å.

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	3053 (1.76-1.72)
Clashscore	122126	3201 (1.76-1.72)
Ramachandran outliers	120053	3169 (1.76-1.72)
Sidechain outliers	120020	3169 (1.76-1.72)
RSRZ outliers	108989	2999 (1.76-1.72)

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	

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 7955 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
1	A	890	Total	C	N	O	S	0	9	0
			7350	4719	1198	1406	27			

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

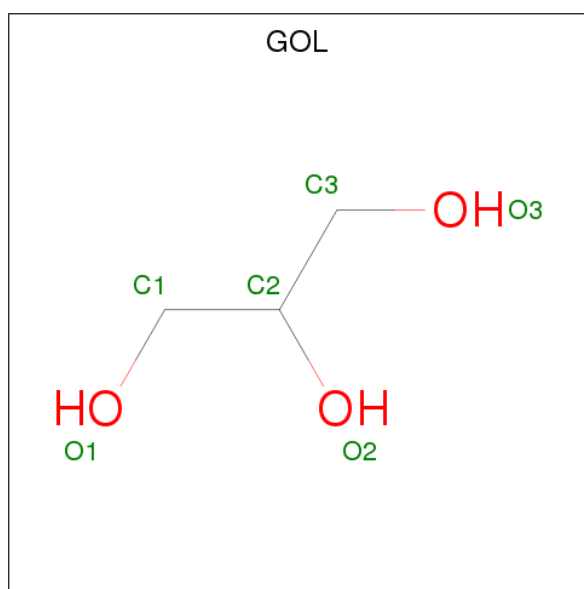
- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

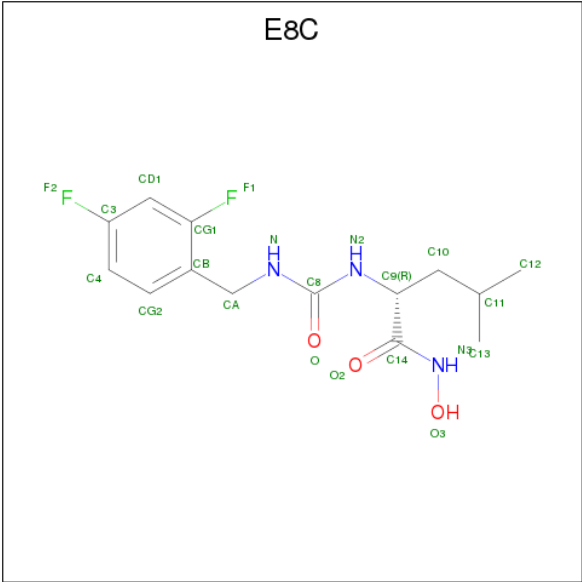
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Zn	0	0
			1	1		

- 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		

- Molecule 5 is (2R)-2-[[2,4-bis(fluoranyl)phenyl]methylcarbamoylamino]-4-methyl-N-oxidany l-pentanamide (three-letter code: E8C) (formula: C<sub>14</sub>H<sub>19</sub>F<sub>2</sub>N<sub>3</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	A	1	Total	C	F	N	O	0	0
			22	14	2	3	3		

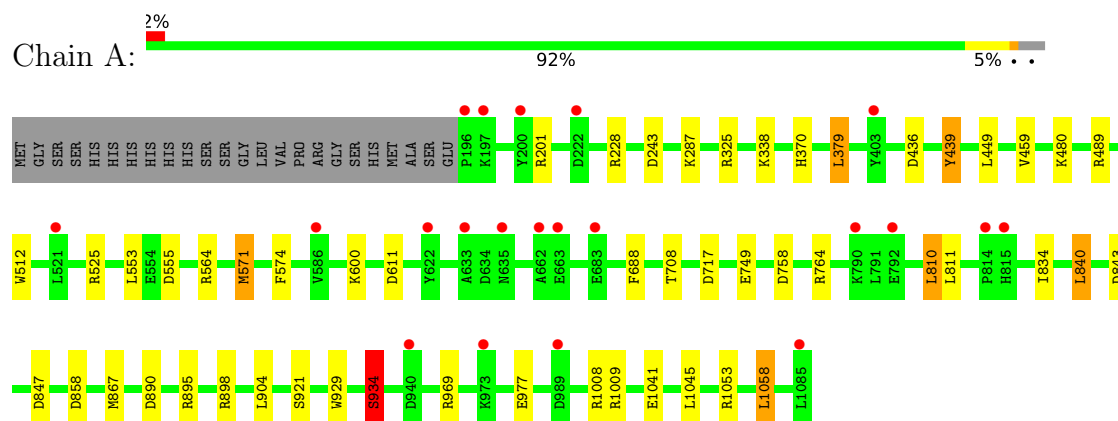
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	561	Total	O	0	0
			561	561		

### 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, $\alpha$ , $\beta$ , $\gamma$	74.56Å 109.24Å 112.52Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	24.66 – 1.73 24.66 – 1.73	Depositor EDS
% Data completeness (in resolution range)	99.0 (24.66-1.73) 99.0 (24.66-1.73)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	30.13 (at 1.73Å)	Xtriage
Refinement program	REFMAC 5.8.0135 : 01/10/15	Depositor
R, $R_{free}$	0.161 , 0.201 0.174 , 0.210	Depositor DCC
$R_{free}$ test set	4826 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	19.4	Xtriage
Anisotropy	0.016	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 43.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.35$	Xtriage
Estimated twinning fraction	0.004 for -h,l,k	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	7955	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	20.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.72% 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: GOL, MG, ZN, E8C

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	1.04	4/7540 (0.1%)	1.10	37/10191 (0.4%)

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	564	ARG	CZ-NH1	10.59	1.46	1.33
1	A	449	LEU	N-CA	9.28	1.65	1.46
1	A	934	SER	CB-OG	-7.81	1.32	1.42
1	A	921	SER	CB-OG	6.89	1.51	1.42

The worst 5 of 37 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	564	ARG	NE-CZ-NH2	-19.27	110.67	120.30
1	A	564	ARG	NE-CZ-NH1	18.66	129.63	120.30
1	A	898	ARG	NE-CZ-NH1	11.61	126.11	120.30
1	A	489[A]	ARG	NE-CZ-NH1	-9.10	115.75	120.30
1	A	489[B]	ARG	NE-CZ-NH1	-9.10	115.75	120.30

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	7350	0	7291	17	0
2	A	3	0	0	0	0
3	A	1	0	0	0	0
4	A	18	0	24	2	0
5	A	22	0	0	0	0
6	A	561	0	0	11	0
All	All	7955	0	7315	18	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.

The worst 5 of 18 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:934:SER:HB3	6:A:1283:HOH:O	1.83	0.77
1:A:890[A]:ASP:OD1	6:A:1202:HOH:O	2.08	0.69
1:A:708:THR:CG2	6:A:1685:HOH:O	2.50	0.60
1:A:969:ARG:HG3	6:A:1283:HOH:O	2.01	0.60
1:A:708:THR:HG21	6:A:1685:HOH:O	2.01	0.59

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%)	883 (98%)	13 (1%)	1 (0%)	53 34

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%)	821 (99%)	10 (1%)	74	58

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	810	LEU
1	A	934	SER
1	A	1041	GLU
1	A	688	PHE
1	A	1008	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	199	HIS
1	A	829	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

Of 8 ligands modelled in this entry, 4 are monoatomic - leaving 4 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	0.79	0	5,5,5	0.67	0
4	GOL	A	1106	-	5,5,5	0.87	0	5,5,5	0.62	0
4	GOL	A	1107	-	5,5,5	0.22	0	5,5,5	0.18	0
5	E8C	A	1108	3	22,22,22	1.93	5 (22%)	28,29,29	2.73	14 (50%)

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
5	E8C	A	1108	3	-	0/19/19/19	0/1/1/1

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	1108	E8C	C9-N2	-4.05	1.37	1.45
5	A	1108	E8C	C8-N2	2.38	1.40	1.35
5	A	1108	E8C	C8-N	2.41	1.40	1.35
5	A	1108	E8C	CB-CG1	4.11	1.45	1.38
5	A	1108	E8C	CA-N	4.87	1.55	1.46

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	1108	E8C	C10-C9-N2	-4.33	100.49	110.57
5	A	1108	E8C	CD1-CG1-CB	-4.27	118.08	123.97
5	A	1108	E8C	CA-CB-CG2	-4.13	111.71	121.00

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	1108	E8C	C9-N2-C8	-3.57	112.83	120.49
5	A	1108	E8C	O-C8-N	-3.26	116.74	122.51

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1105	GOL	1	0
4	A	1106	GOL	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	890/914 (97%)	-0.02	21 (2%) 59 65	12, 18, 33, 72	0

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	196	PRO	7.1
1	A	989	ASP	4.0
1	A	403	TYR	3.2
1	A	815	HIS	3.2
1	A	200	TYR	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. 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	1107	6/6	0.87	0.15	44,47,48,61	0
5	E8C	A	1108	22/22	0.87	0.18	14,24,39,42	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	1106	6/6	0.89	0.21	18,31,36,44	0
4	GOL	A	1105	6/6	0.94	0.17	30,36,39,44	0
2	MG	A	1102	1/1	0.95	0.13	22,22,22,22	0
2	MG	A	1101	1/1	0.95	0.33	30,30,30,30	0
2	MG	A	1103	1/1	0.96	0.17	27,27,27,27	0
3	ZN	A	1104	1/1	1.00	0.02	14,14,14,14	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.