



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 5, 2017 – 05:35 PM EDT

PDB ID : 4TPG  
Title : Selectivity mechanism of a bacterial homologue of the human drug peptide transporters PepT1 and PepT2  
Authors : Guettou, F.; Quistgaard, E.M.; Raba, M.; Moberg, P.; Low, C.; Nordlund, P.  
Deposited on : unknown  
Resolution : 3.91 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-20029824  
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-20029824

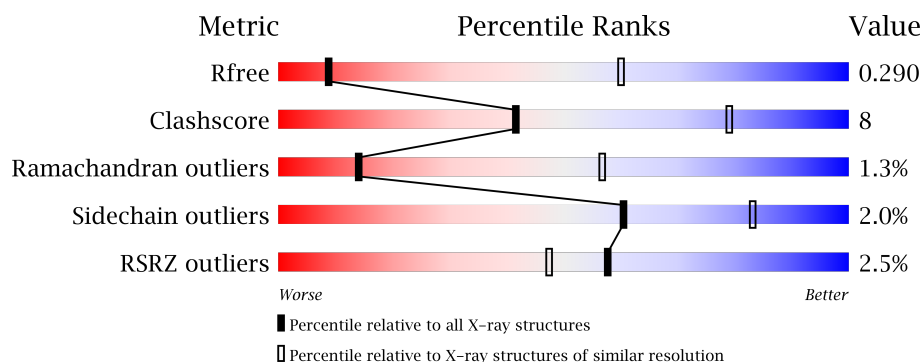
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.91 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	1012 (4.24-3.60)
Clashscore	112137	1108 (4.24-3.60)
Ramachandran outliers	110173	1067 (4.24-3.60)
Sidechain outliers	110143	1058 (4.24-3.60)
RSRZ outliers	101464	1025 (4.24-3.60)

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	523	<div> <div>2%</div> <div>67%</div> <div>20%</div> <div>12%</div> </div>
1	B	523	<div> <div>2%</div> <div>71%</div> <div>17%</div> <div>12%</div> </div>
2	E	3	<div> <div>33%</div> <div>67%</div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	LMT	A	602	-	-	-	X
4	LMT	B	601	-	-	-	X

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 7036 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 Proton:oligopeptide symporter POT family.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	459	Total	C	N	O	S	0	0	0
			3470	2308	555	589	18			
1	B	459	Total	C	N	O	S	0	0	0
			3470	2308	555	589	18			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	517	ALA	-	expression tag	UNP Q8EHE6
A	518	GLU	-	expression tag	UNP Q8EHE6
A	519	ASN	-	expression tag	UNP Q8EHE6
A	520	LEU	-	expression tag	UNP Q8EHE6
A	521	TYR	-	expression tag	UNP Q8EHE6
A	522	PHE	-	expression tag	UNP Q8EHE6
A	523	GLN	-	expression tag	UNP Q8EHE6
B	517	ALA	-	expression tag	UNP Q8EHE6
B	518	GLU	-	expression tag	UNP Q8EHE6
B	519	ASN	-	expression tag	UNP Q8EHE6
B	520	LEU	-	expression tag	UNP Q8EHE6
B	521	TYR	-	expression tag	UNP Q8EHE6
B	522	PHE	-	expression tag	UNP Q8EHE6
B	523	GLN	-	expression tag	UNP Q8EHE6

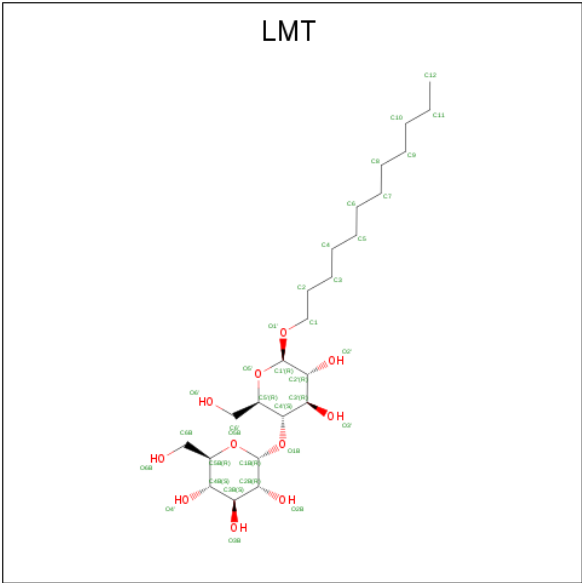
- Molecule 2 is a protein called Ala-L-3-Br-Tyr-Ala.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	3	Total	Br	C	N	O	0	0	0
			25	2	15	3	5			

- 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 DODECYL-BETA-D-MALTOSE (three-letter code: LMT) (formula: C<sub>24</sub>H<sub>46</sub>O<sub>11</sub>).

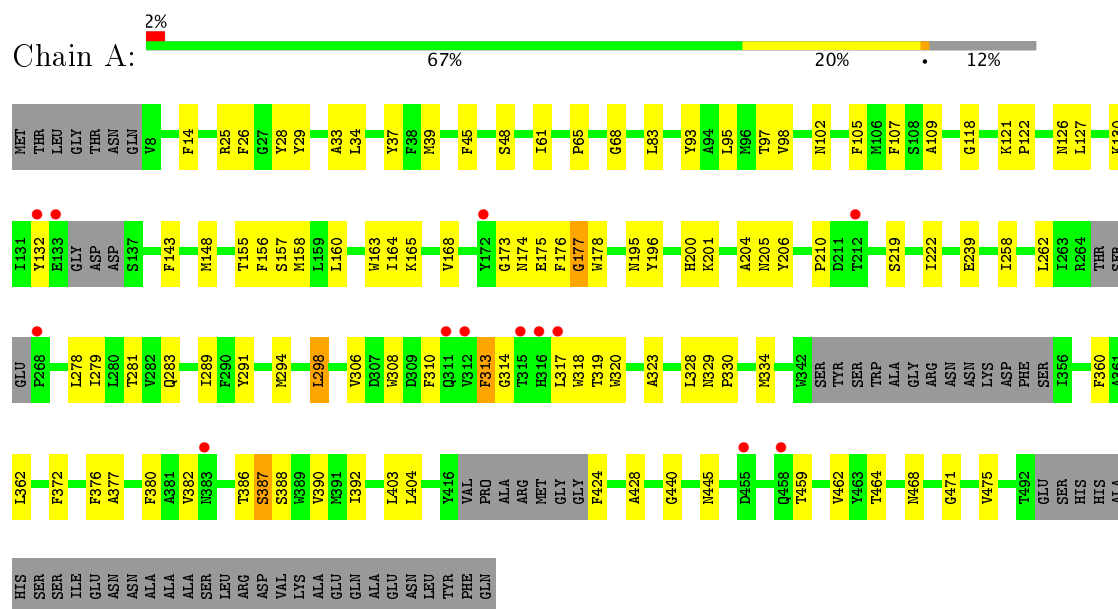


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			35	24	11		
4	B	1	Total	C	O	0	0
			35	24	11		

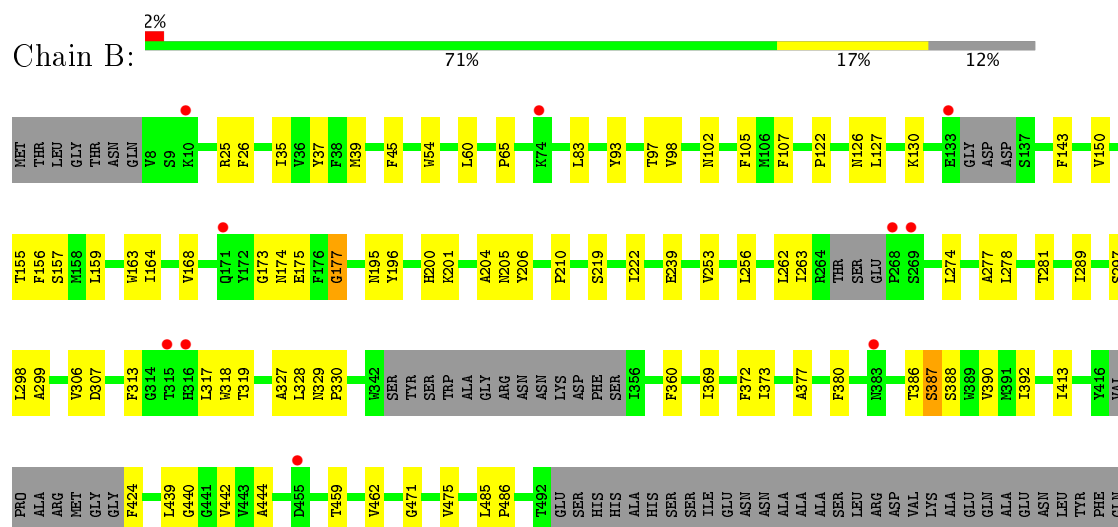
### 3 Residue-property plots [i](#)

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.

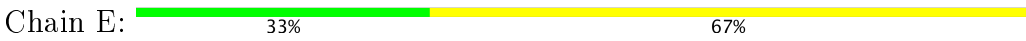
- Molecule 1: Proton:oligopeptide symporter POT family



- Molecule 1: Proton:oligopeptide symporter POT family



- Molecule 2: Ala-L-3-Br-Tyr-Ala



GLOBAL-STATISTICS INFOmissingINFO

## 4 Model quality [i](#)

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

Bond lengths and bond angles in the following residue types are not validated in this section: DBY, ZN, LMT

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	0/3559	0.42	0/4853
1	B	0.25	0/3559	0.42	0/4853
2	E	1.33	0/9	0.45	0/8
All	All	0.26	0/7127	0.42	0/9714

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 4.2 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	3470	0	3457	62	0
1	B	3470	0	3457	46	0
2	E	25	0	18	4	0
3	A	1	0	0	0	0
4	A	35	0	46	6	0
4	B	35	0	46	4	0
All	All	7036	0	7024	112	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 8.



The worst 5 of 112 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:289:ILE:HG23	1:B:440:GLY:HA2	1.71	0.72
1:A:200:HIS:CG	1:A:201:LYS:H	2.11	0.68
1:B:387:SER:HB3	1:B:390:VAL:HB	1.77	0.67
1:A:289:ILE:HG23	1:A:440:GLY:HA2	1.78	0.65
1:B:200:HIS:CG	1:B:201:LYS:H	2.14	0.65

There are no symmetry-related clashes.

## 4.3 Torsion angles [i](#)

### 4.3.1 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	449/523 (86%)	423 (94%)	20 (4%)	6 (1%)	14	57
1	B	449/523 (86%)	420 (94%)	23 (5%)	6 (1%)	14	57
All	All	898/1046 (86%)	843 (94%)	43 (5%)	12 (1%)	14	57

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	210	PRO
1	B	210	PRO
1	A	205	ASN
1	A	387	SER
1	B	205	ASN

### 4.3.2 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	344/417 (82%)	337 (98%)	7 (2%)	60	83
1	B	344/417 (82%)	337 (98%)	7 (2%)	60	83
All	All	688/834 (82%)	674 (98%)	14 (2%)	60	83

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	380	PHE
1	B	98	VAL
1	B	298	LEU
1	A	360	PHE
1	B	195	ASN

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

#### 4.3.3 RNA ⓘ

There are no RNA molecules in this entry.

### 4.4 Non-standard residues in protein, DNA, RNA chains ⓘ

1 non-standard protein/DNA/RNA residue is 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	DBY	E	2	2	14,14,15	1.11	2 (14%)	18,19,21	1.28	4 (22%)

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	DBY	E	2	2	-	0/4/6/8	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	E	2	DBY	OH-CZ	2.20	1.42	1.37
2	E	2	DBY	CA-C	2.47	1.53	1.50

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	E	2	DBY	CB-CA-C	-2.29	107.00	111.41
2	E	2	DBY	CD2-CE2-CZ	-2.15	119.49	121.92
2	E	2	DBY	O-C-CA	-2.14	119.11	125.02
2	E	2	DBY	CE2-CZ-CE1	2.46	119.84	116.56

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	E	2	DBY	2	0

## 4.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 4.6 Ligand geometry [i](#)

Of 3 ligands modelled in this entry, 1 is monoatomic - leaving 2 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	LMT	A	602	-	36,36,36	0.52	0	47,47,47	1.09	3 (6%)
4	LMT	B	601	-	36,36,36	0.51	0	47,47,47	1.18	3 (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
4	LMT	A	602	-	-	0/21/61/61	0/2/2/2
4	LMT	B	601	-	-	0/21/61/61	0/2/2/2

There are no bond length outliers.

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	B	601	LMT	C1B-O1B-C4'	-3.34	109.85	118.00
4	B	601	LMT	C1'-O5'-C5'	-3.19	107.70	113.72
4	A	602	LMT	C1'-O5'-C5'	-3.05	107.96	113.72
4	A	602	LMT	C1B-O1B-C4'	-2.98	110.74	118.00
4	A	602	LMT	C1B-O5B-C5B	-2.63	108.77	113.72

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 10 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	602	LMT	6	0
4	B	601	LMT	4	0

## 4.7 Other polymers [i](#)

There are no such residues in this entry.

## 4.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

## 5 Fit of model and data [i](#)

### 5.1 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	459/523 (87%)	-0.34	13 (2%) 53 43	51, 111, 191, 291	0
1	B	459/523 (87%)	-0.37	10 (2%) 62 53	52, 110, 200, 297	0
2	E	2/3 (66%)	0.41	0 100 100	137, 137, 137, 168	0
All	All	920/1049 (87%)	-0.35	23 (2%) 58 47	51, 111, 198, 297	0

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	269	SER	4.6
1	B	268	PRO	3.8
1	A	133	GLU	3.7
1	A	455	ASP	3.3
1	A	172	TYR	3.1

### 5.2 Non-standard residues in protein, DNA, RNA chains [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	DBY	E	2	14/15	0.58	0.46	-	200,235,439,505	0

### 5.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.4 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
4	LMT	B	601	35/35	0.79	0.52	2.82	128,170,199,199	0
4	LMT	A	602	35/35	0.85	0.49	2.52	83,155,174,224	0
3	ZN	A	601	1/1	0.85	0.29	-	197,197,197,197	0

## 5.5 Other polymers [i](#)

There are no such residues in this entry.