



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 7, 2019 – 10:49 PM EDT

PDB ID : 5N4D  
Title : Prolyl oligopeptidase B from *Galerina marginata* bound to 25mer macrocyclization substrate - D661A mutant  
Authors : Czekster, C.M.; McMahon, S.A.; Ludewig, H.; Naismith, J.H.  
Deposited on : 2017-02-10  
Resolution : 1.62 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	:	1.8.0 (224370), CSD as540be (2019)
Xtriage (Phenix)	:	1.13
EDS	:	2.4
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)	:	2.4

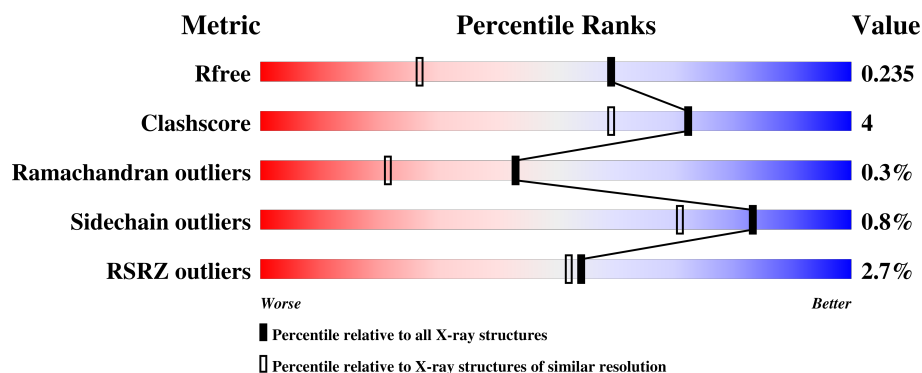
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.62 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	3975 (1.64-1.60)
Clashscore	122126	4258 (1.64-1.60)
Ramachandran outliers	120053	4162 (1.64-1.60)
Sidechain outliers	120020	4161 (1.64-1.60)
RSRZ outliers	108989	3894 (1.64-1.60)

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	730	<div> <div>2%</div> <div>91%</div> <div>7% ..</div> </div>
1	B	730	<div> <div>3%</div> <div>90%</div> <div>9% ..</div> </div>
2	C	25	<div> <div>12%</div> <div>64%</div> <div>8%</div> <div>28%</div> </div>
2	D	25	<div> <div>12%</div> <div>60%</div> <div>12%</div> <div>28%</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
3	GOL	D	101	-	-	X	-

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 13347 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 Prolyl oligopeptidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	720	Total	C	N	O	S	0	11	0
			5762	3696	962	1092	12			
1	B	722	Total	C	N	O	S	0	13	0
			5795	3719	972	1093	11			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	661	ALA	ASP	engineered mutation	UNP H2E7Q8
B	661	ALA	ASP	engineered mutation	UNP H2E7Q8

- Molecule 2 is a protein called Alpha-amanitin proprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	18	Total	C	N	O	S	0	0	0
			137	83	23	30	1			
2	D	18	Total	C	N	O	S	0	0	0
			137	83	23	30	1			

- Molecule 3 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
3	A	1	Total	C	O	0	0
			6	3	3		
3	A	1	Total	C	O	0	0
			6	3	3		
3	B	1	Total	C	O	0	0
			6	3	3		
3	D	1	Total	C	O	0	0
			6	3	3		

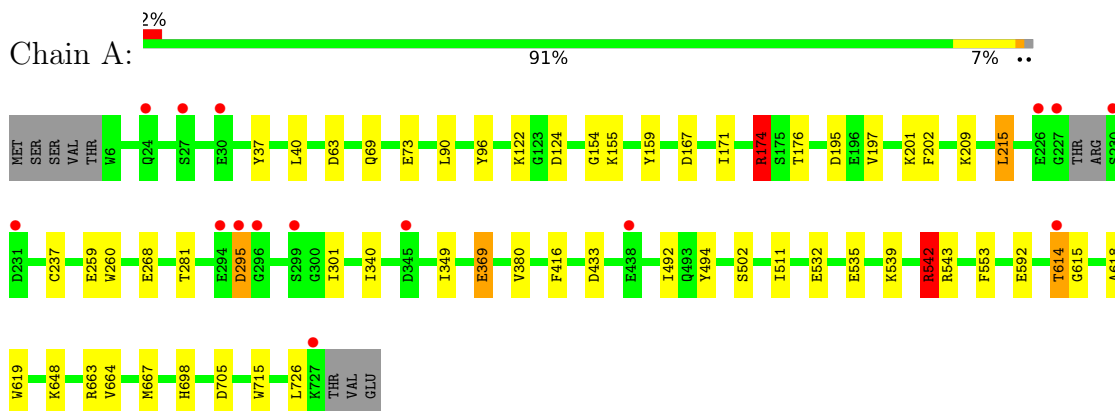
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	740	Total	O	0	0
			740	740		
4	B	721	Total	O	0	0
			721	721		
4	C	16	Total	O	0	0
			16	16		
4	D	15	Total	O	0	0
			15	15		

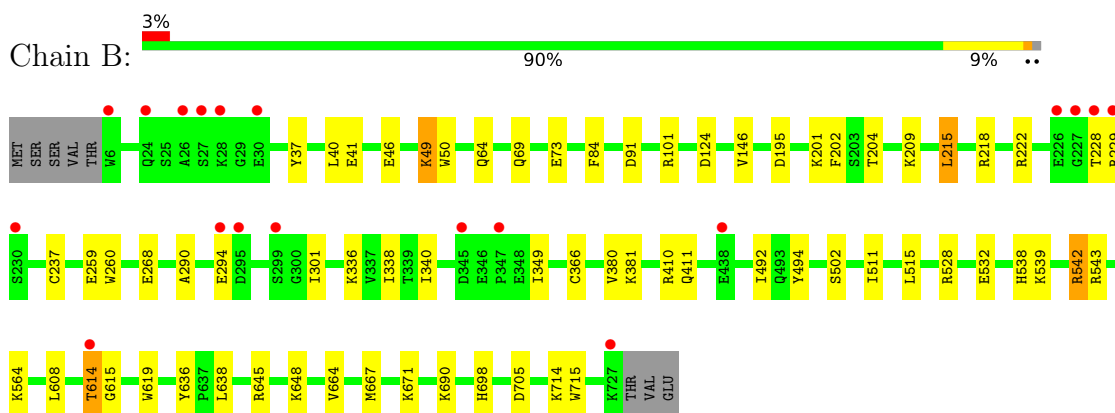
### 3 Residue-property plots

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.

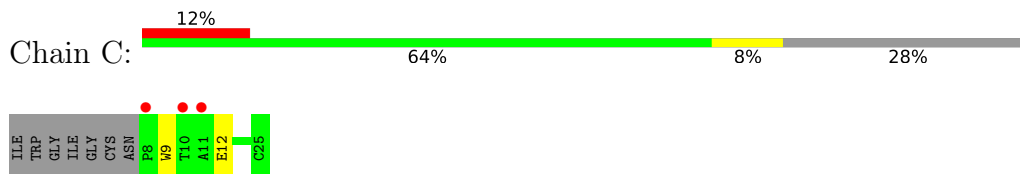
- Molecule 1: Prolyl oligopeptidase



- Molecule 1: Prolyl oligopeptidase

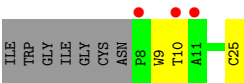


- Molecule 2: Alpha-amanitin proprotein



- Molecule 2: Alpha-amanitin proprotein





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	99.04Å 114.88Å 141.31Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.88 – 1.62 46.87 – 1.62	Depositor EDS
% Data completeness (in resolution range)	99.9 (46.88-1.62) 100.0 (46.87-1.62)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.65 (at 1.62Å)	Xtriage
Refinement program	PHENIX (1.11rc3_2553: ???)	Depositor
R, $R_{free}$	0.198 , 0.234 0.200 , 0.235	Depositor DCC
$R_{free}$ test set	10076 reflections (4.94%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	15.9	Xtriage
Anisotropy	0.259	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.38 , 47.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	13347	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	22.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 58.30 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 2.0872e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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

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	0.79	2/5952 (0.0%)	0.85	10/8084 (0.1%)
1	B	0.79	0/5992	0.84	10/8137 (0.1%)
2	C	0.64	0/140	0.70	0/190
2	D	0.70	0/140	0.72	0/190
All	All	0.79	2/12224 (0.0%)	0.84	20/16601 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	369	GLU	CG-CD	5.45	1.60	1.51
1	A	553	PHE	CE1-CZ	5.36	1.47	1.37

The worst 5 of 20 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	542	ARG	NE-CZ-NH1	-8.96	115.82	120.30
1	B	542	ARG	NE-CZ-NH2	6.87	123.73	120.30
1	A	155	LYS	CD-CE-NZ	6.68	127.07	111.70
1	B	124	ASP	CB-CG-OD2	-6.63	112.33	118.30
1	A	155	LYS	CB-CG-CD	-6.52	94.64	111.60

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	5762	0	5572	35	0
1	B	5795	0	5630	46	0
2	C	137	0	120	2	0
2	D	137	0	120	3	0
3	A	12	0	16	2	0
3	B	6	0	8	0	0
3	D	6	0	7	4	0
4	A	740	0	0	7	0
4	B	721	0	0	10	1
4	C	16	0	0	0	0
4	D	15	0	0	0	0
All	All	13347	0	11473	86	1

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 4.

The worst 5 of 86 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:290:ALA:HB1	1:B:301:ILE:HD11	1.52	0.89
1:B:222:ARG:NH1	1:B:228:THR:O	2.15	0.79
1:B:564:LYS:NZ	4:B:901:HOH:O	2.18	0.76
1:B:49:LYS:HD2	1:B:50:TRP:N	2.01	0.75
1:A:535:GLU:HG3	1:A:539:LYS:NZ	2.01	0.74

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:B:1454:HOH:O	4:B:1487:HOH:O[3_557]	2.10	0.10

## 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	727/730 (100%)	709 (98%)	16 (2%)	2 (0%)	43	21
1	B	733/730 (100%)	715 (98%)	16 (2%)	2 (0%)	43	21
2	C	16/25 (64%)	15 (94%)	1 (6%)	0	100	100
2	D	16/25 (64%)	15 (94%)	1 (6%)	0	100	100
All	All	1492/1510 (99%)	1454 (98%)	34 (2%)	4 (0%)	43	21

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	614	THR
1	B	614	THR
1	A	380	VAL
1	B	380	VAL

### 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	612/612 (100%)	606 (99%)	6 (1%)	78	62
1	B	616/612 (101%)	612 (99%)	4 (1%)	87	78
2	C	15/20 (75%)	15 (100%)	0	100	100
2	D	15/20 (75%)	15 (100%)	0	100	100
All	All	1258/1264 (100%)	1248 (99%)	10 (1%)	83	71

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

Mol	Chain	Res	Type
1	A	542	ARG
1	A	667	MET
1	B	195	ASP
1	A	295	ASP

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	49	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	64	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 [i](#)

4 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	$\# Z  > 2$
3	GOL	A	801	-	5,5,5	0.56	0	5,5,5	0.97	0
3	GOL	A	802	-	5,5,5	0.64	0	5,5,5	0.89	0
3	GOL	B	801	-	5,5,5	0.94	0	5,5,5	1.58	2 (40%)
3	GOL	D	101	-	5,5,5	1.40	1 (20%)	5,5,5	2.00	2 (40%)

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
3	GOL	A	801	-	-	1/4/4/4	-
3	GOL	A	802	-	-	3/4/4/4	-
3	GOL	B	801	-	-	2/4/4/4	-
3	GOL	D	101	-	-	2/4/4/4	-

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	101	GOL	O1-C1	-2.21	1.33	1.42

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	101	GOL	O3-C3-C2	3.23	125.70	110.12
3	D	101	GOL	O2-C2-C1	-2.63	97.52	109.12
3	B	801	GOL	C3-C2-C1	-2.30	102.79	111.75
3	B	801	GOL	O2-C2-C3	2.03	118.07	109.12

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	801	GOL	C1-C2-C3-O3
3	A	802	GOL	C1-C2-C3-O3
3	D	101	GOL	O1-C1-C2-C3
3	B	801	GOL	C1-C2-C3-O3
3	A	802	GOL	O2-C2-C3-O3

There are no ring outliers.

2 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	801	GOL	2	0
3	D	101	GOL	4	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	720/730 (98%)	-0.26	15 (2%) 63 62	12, 19, 35, 72	0
1	B	722/730 (98%)	-0.24	19 (2%) 56 54	12, 19, 38, 80	0
2	C	18/25 (72%)	0.66	3 (16%) 1 1	16, 27, 61, 76	0
2	D	18/25 (72%)	0.55	3 (16%) 1 1	15, 24, 62, 70	0
All	All	1478/1510 (97%)	-0.23	40 (2%) 54 52	12, 19, 38, 80	0

The worst 5 of 40 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	228	THR	11.5
1	B	227	GLY	9.3
1	A	227	GLY	7.2
1	B	229	ARG	7.1
2	C	8	PRO	4.2

### 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( $\text{\AA}^2$ )	Q<0.9
3	GOL	D	101	6/6	0.78	0.16	14,19,26,36	0
3	GOL	B	801	6/6	0.80	0.15	21,29,34,41	0
3	GOL	A	801	6/6	0.88	0.22	23,27,30,31	0
3	GOL	A	802	6/6	0.90	0.11	21,28,31,33	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.