



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 24, 2024 – 08:36 AM EDT

PDB ID : 3EFO  
Title : Crystal Structure of the mammalian COPII-coat protein Sec23/24 bound to the transport signal sequence of syntaxin 5  
Authors : Goldberg, J.; Mancias, J.D.  
Deposited on : 2008-09-09  
Resolution : 2.70 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.20.1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

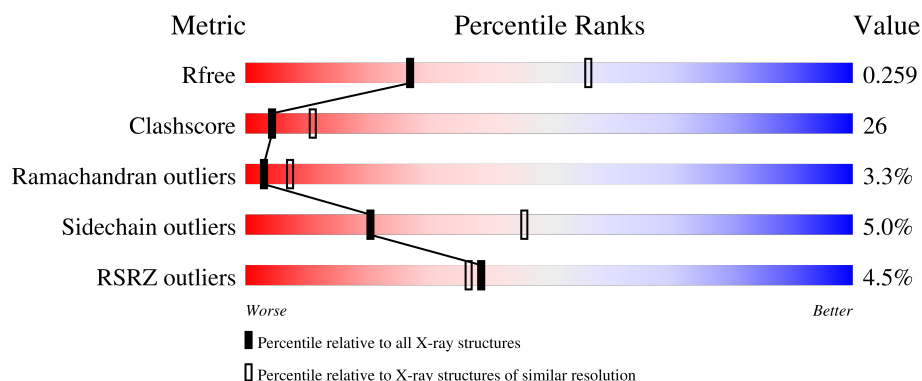
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.70 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))
$R_{free}$	164625	3333 (2.70-2.70)
Clashscore	180529	3684 (2.70-2.70)
Ramachandran outliers	177936	3633 (2.70-2.70)
Sidechain outliers	177891	3633 (2.70-2.70)
RSRZ outliers	164620	3333 (2.70-2.70)

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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria respectively. 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	765	<div> <div>3%</div> <div>57%</div> <div>32%</div> <div>6%</div> </div>
2	B	770	<div> <div>6%</div> <div>54%</div> <div>40%</div> <div>6%</div> </div>
3	C	7	<div> <div>14%</div> <div>71%</div> <div>29%</div> </div>

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 11927 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 Protein transport protein Sec23A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	716	Total	C	N	O	S	0	0	0
			5684	3623	975	1046	40			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	193	GLY	LYS	conflict	UNP Q15436

- Molecule 2 is a protein called SEC24 related gene family, member D.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	767	Total	C	N	O	S	0	0	0
			6038	3841	1024	1119	54			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1	ALA	-	expression tag	UNP Q8IYI7
B	2	MET	-	expression tag	UNP Q8IYI7
B	3	GLY	-	expression tag	UNP Q8IYI7

- Molecule 3 is a protein called Peptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	7	Total	C	N	O	S	0	0	0
			53	32	7	12	2			

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total 1	Zn 1	0	0

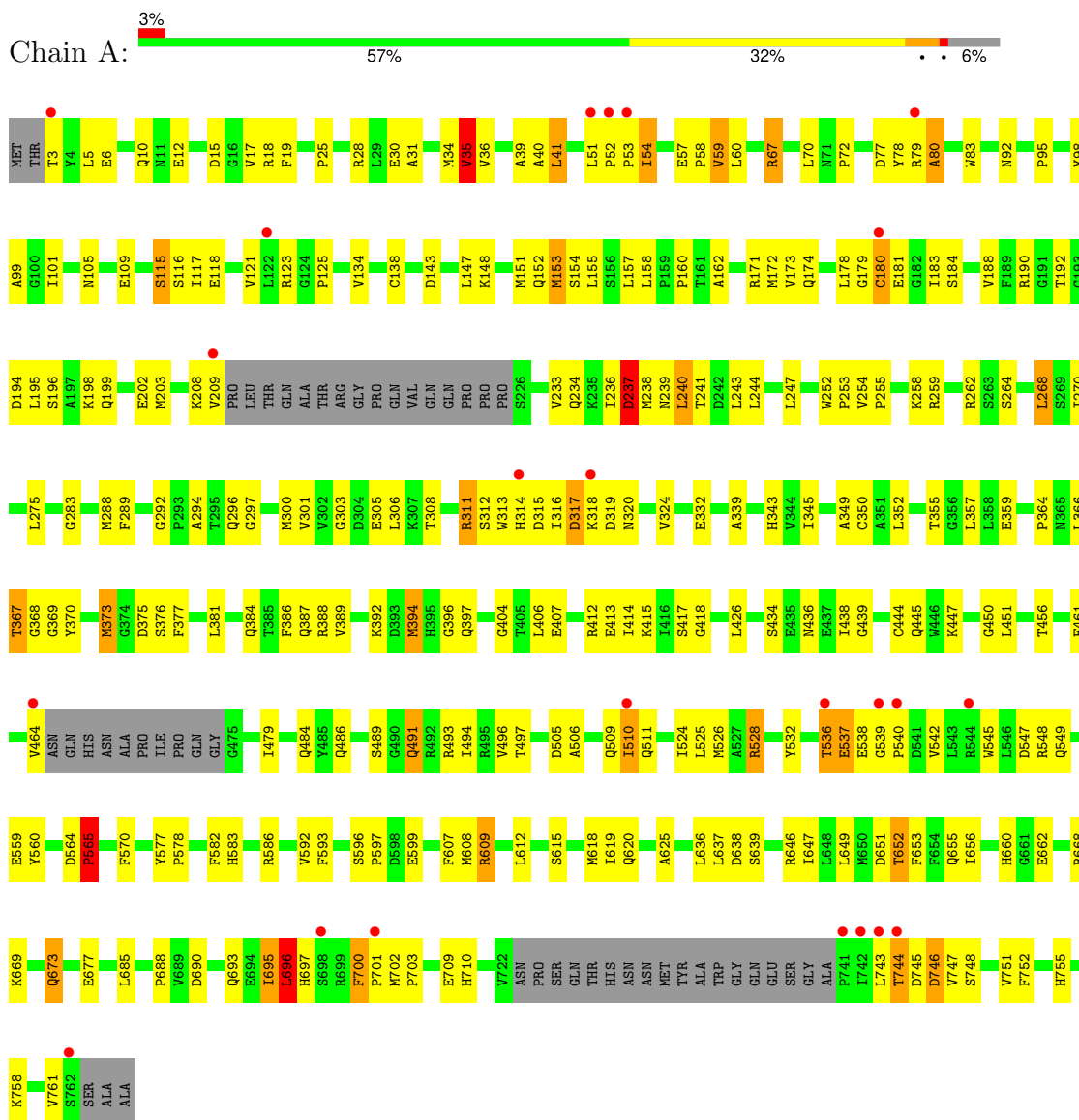
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	79	Total 79	O 79	0	0
5	B	71	Total 71	O 71	0	0

### 3 Residue-property plots

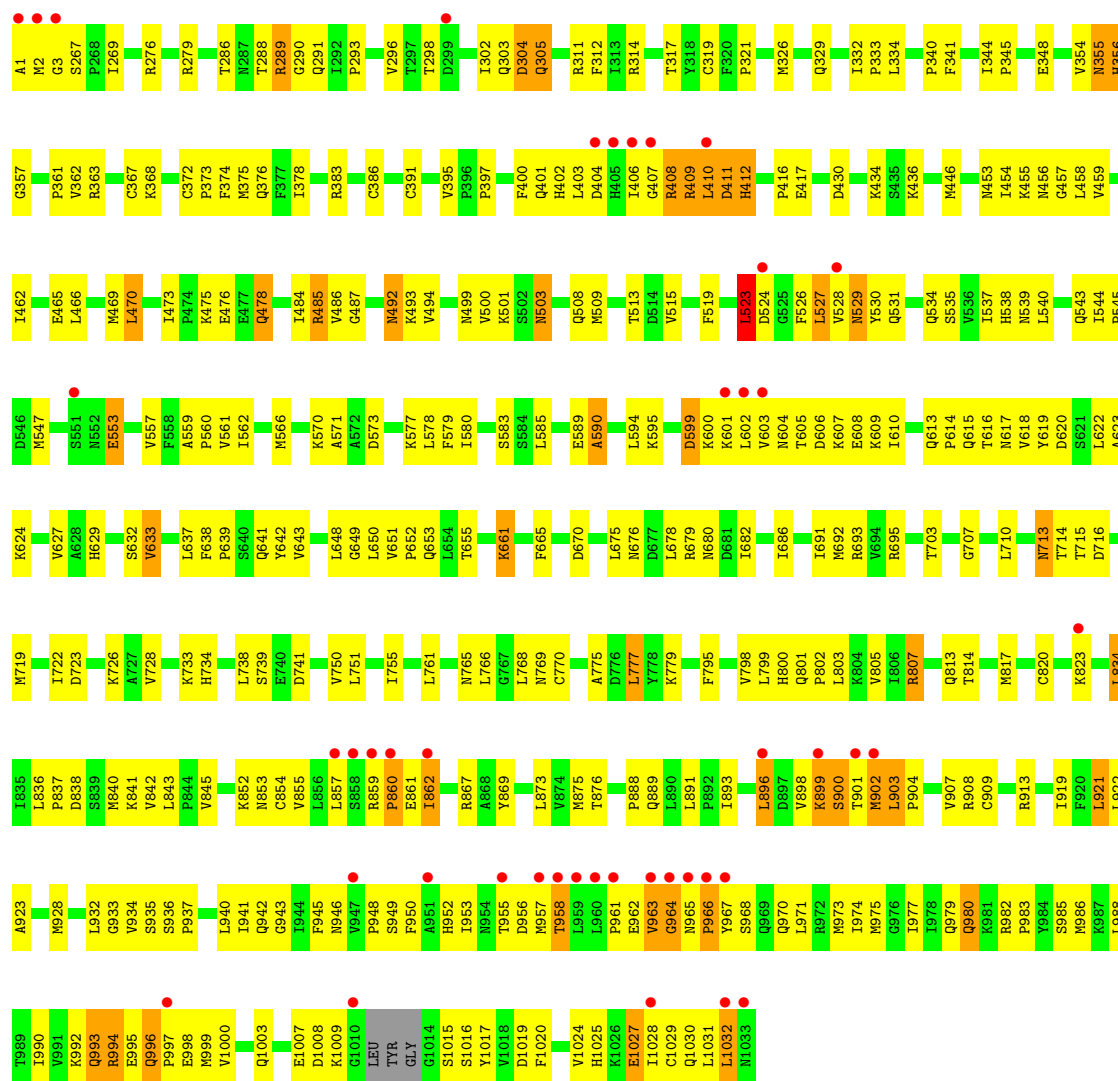
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: Protein transport protein Sec23A

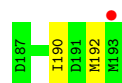


#### • Molecule 2: SEC24 related gene family, member D





### • Molecule 3: Peptide



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	103.06Å 140.83Å 152.25Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.70 50.00 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.4 (50.00-2.70) 94.9 (50.00-2.70)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.13 (at 2.69Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.211 , 0.261 0.210 , 0.259	Depositor DCC
$R_{free}$ test set	3050 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	51.1	Xtriage
Anisotropy	0.611	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 52.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	11927	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	58.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.81% 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: ZN

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	0.38	0/5817	0.66	3/7878 (0.0%)
2	B	0.38	0/6166	0.66	1/8351 (0.0%)
3	C	0.40	0/52	0.60	0/67
All	All	0.38	0/12035	0.66	4/16296 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	115	SER	N-CA-C	-7.63	90.40	111.00
1	A	652	THR	N-CA-C	-6.57	93.27	111.00
1	A	367	THR	N-CA-C	-5.82	95.28	111.00
2	B	523	LEU	N-CA-C	-5.21	96.93	111.00

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	5684	0	5636	246	0
2	B	6038	0	6038	376	0
3	C	53	0	50	5	0
4	A	1	0	0	0	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	B	1	0	0	0	0
5	A	79	0	0	10	0
5	B	71	0	0	7	0
All	All	11927	0	11724	617	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 26.

The worst 5 of 617 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)
2:B:2:MET:HE1	2:B:836:LEU:HD12	1.30	1.14
1:A:747:VAL:HG12	1:A:748:SER:H	1.25	1.01
1:A:123:ARG:HG3	1:A:125:PRO:HD2	1.39	1.01
2:B:639:PRO:HB3	2:B:643:VAL:HG21	1.40	1.00
2:B:601:LYS:HB3	2:B:605:THR:HG21	1.44	0.99

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	708/765 (92%)	635 (90%)	54 (8%)	19 (3%)	4	10
2	B	763/770 (99%)	659 (86%)	75 (10%)	29 (4%)	2	6
3	C	5/7 (71%)	5 (100%)	0	0	100	100
All	All	1476/1542 (96%)	1299 (88%)	129 (9%)	48 (3%)	3	7

5 of 48 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	35	VAL
1	A	746	ASP
2	B	289	ARG
2	B	304	ASP
2	B	408	ARG

### 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	626/665 (94%)	595 (95%)	31 (5%)	20	46
2	B	680/682 (100%)	645 (95%)	35 (5%)	20	45
3	C	6/6 (100%)	6 (100%)	0	100	100
All	All	1312/1353 (97%)	1246 (95%)	66 (5%)	20	46

5 of 66 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	807	ARG
2	B	843	LEU
2	B	1027	GLU
1	A	609	ARG
1	A	570	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 42 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	668	HIS
2	B	942	GLN
2	B	713	ASN
2	B	816	HIS
2	B	969	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 oligosaccharides in this entry.

### 5.6 Ligand geometry [i](#)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

### 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	716/765 (93%)	-0.09	23 (3%)	50 48	26, 53, 94, 131	0
2	B	767/770 (99%)	0.01	43 (5%)	31 29	26, 54, 109, 141	0
3	C	7/7 (100%)	0.26	1 (14%)	7 7	56, 60, 91, 119	0
All	All	1490/1542 (96%)	-0.04	67 (4%)	39 37	26, 54, 102, 141	0

The worst 5 of 67 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	540	PRO	6.8
2	B	960	LEU	5.2
2	B	1033	ASN	4.9
1	A	742	ILE	4.9
2	B	860	PRO	4.6

### 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 monosaccharides 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
4	ZN	A	766	1/1	1.00	0.04	46,46,46,46	0
4	ZN	B	1034	1/1	1.00	0.02	57,57,57,57	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.