



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 12:35 AM GMT

PDB ID : 2AR0  
Title : Crystal structure of Type I restriction enzyme EcoKI M protein (EC 2.1.1.72) (M.EcoKI)  
Authors : Rajashankar, K.R.; Kniewel, R.; Lima, C.D.; Burley, S.K.; New York SGX Research Center for Structural Genomics (NYSGXRC)  
Deposited on : 2005-08-18  
Resolution : 2.80 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 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
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) : trunk26865

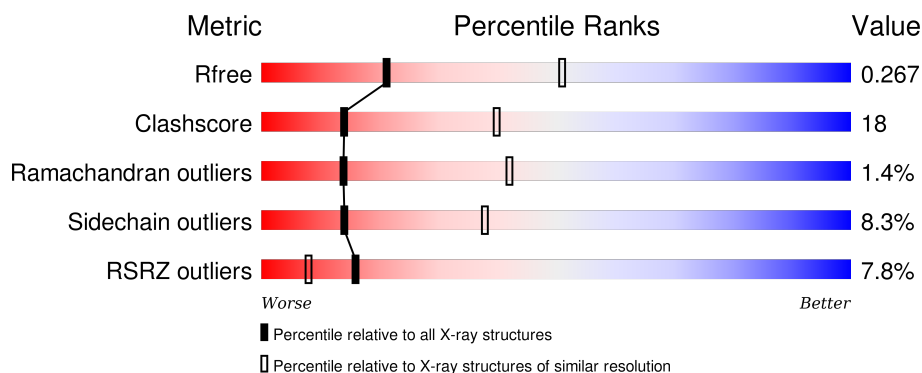
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.80 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))
$R_{free}$	91344	2393 (2.80-2.80)
Clashscore	102246	2827 (2.80-2.80)
Ramachandran outliers	100387	2782 (2.80-2.80)
Sidechain outliers	100360	2784 (2.80-2.80)
RSRZ outliers	91569	2404 (2.80-2.80)

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	541	<div> <div>6%</div> <div>58%</div> <div>28%</div> <div>•</div> <div>10%</div> </div>
1	B	541	<div> <div>8%</div> <div>57%</div> <div>29%</div> <div>•</div> <div>11%</div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 7828 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 Type I restriction enzyme EcoKI M protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	485	Total	C	N	O	S	Se	0	0	0
			3843	2415	671	742	6	9			
1	B	483	Total	C	N	O	S	Se	0	0	0
			3832	2408	669	740	6	9			

There are 50 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	-	CLONING ARTIFACT	UNP P08957
A	2	SER	-	CLONING ARTIFACT	UNP P08957
A	3	LEU	-	CLONING ARTIFACT	UNP P08957
A	41	MSE	MET	MODIFIED RESIDUE	UNP P08957
A	76	MSE	MET	MODIFIED RESIDUE	UNP P08957
A	112	MSE	MET	MODIFIED RESIDUE	UNP P08957
A	132	MSE	MET	MODIFIED RESIDUE	UNP P08957
A	229	MSE	MET	MODIFIED RESIDUE	UNP P08957
A	295	MSE	MET	MODIFIED RESIDUE	UNP P08957
A	331	MSE	MET	MODIFIED RESIDUE	UNP P08957
A	386	MSE	MET	MODIFIED RESIDUE	UNP P08957
A	492	MSE	MET	MODIFIED RESIDUE	UNP P08957
A	506	MSE	MET	MODIFIED RESIDUE	UNP P08957
A	530	LYS	-	EXPRESSION TAG	UNP P08957
A	531	GLU	-	EXPRESSION TAG	UNP P08957
A	532	GLU	-	EXPRESSION TAG	UNP P08957
A	533	GLY	-	EXPRESSION TAG	UNP P08957
A	534	GLY	-	EXPRESSION TAG	UNP P08957
A	535	SER	-	EXPRESSION TAG	UNP P08957
A	536	HIS	-	EXPRESSION TAG	UNP P08957
A	537	HIS	-	EXPRESSION TAG	UNP P08957
A	538	HIS	-	EXPRESSION TAG	UNP P08957
A	539	HIS	-	EXPRESSION TAG	UNP P08957
A	540	HIS	-	EXPRESSION TAG	UNP P08957
A	541	HIS	-	EXPRESSION TAG	UNP P08957

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	1	MSE	-	CLONING ARTIFACT	UNP P08957
B	2	SER	-	CLONING ARTIFACT	UNP P08957
B	3	LEU	-	CLONING ARTIFACT	UNP P08957
B	41	MSE	MET	MODIFIED RESIDUE	UNP P08957
B	76	MSE	MET	MODIFIED RESIDUE	UNP P08957
B	112	MSE	MET	MODIFIED RESIDUE	UNP P08957
B	132	MSE	MET	MODIFIED RESIDUE	UNP P08957
B	229	MSE	MET	MODIFIED RESIDUE	UNP P08957
B	295	MSE	MET	MODIFIED RESIDUE	UNP P08957
B	331	MSE	MET	MODIFIED RESIDUE	UNP P08957
B	386	MSE	MET	MODIFIED RESIDUE	UNP P08957
B	492	MSE	MET	MODIFIED RESIDUE	UNP P08957
B	506	MSE	MET	MODIFIED RESIDUE	UNP P08957
B	530	LYS	-	EXPRESSION TAG	UNP P08957
B	531	GLU	-	EXPRESSION TAG	UNP P08957
B	532	GLU	-	EXPRESSION TAG	UNP P08957
B	533	GLY	-	EXPRESSION TAG	UNP P08957
B	534	GLY	-	EXPRESSION TAG	UNP P08957
B	535	SER	-	EXPRESSION TAG	UNP P08957
B	536	HIS	-	EXPRESSION TAG	UNP P08957
B	537	HIS	-	EXPRESSION TAG	UNP P08957
B	538	HIS	-	EXPRESSION TAG	UNP P08957
B	539	HIS	-	EXPRESSION TAG	UNP P08957
B	540	HIS	-	EXPRESSION TAG	UNP P08957
B	541	HIS	-	EXPRESSION TAG	UNP P08957

- Molecule 2 is UNKNOWN ATOM OR ION (three-letter code: UNX) (formula: X).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	10	Total X 10 10	0	0

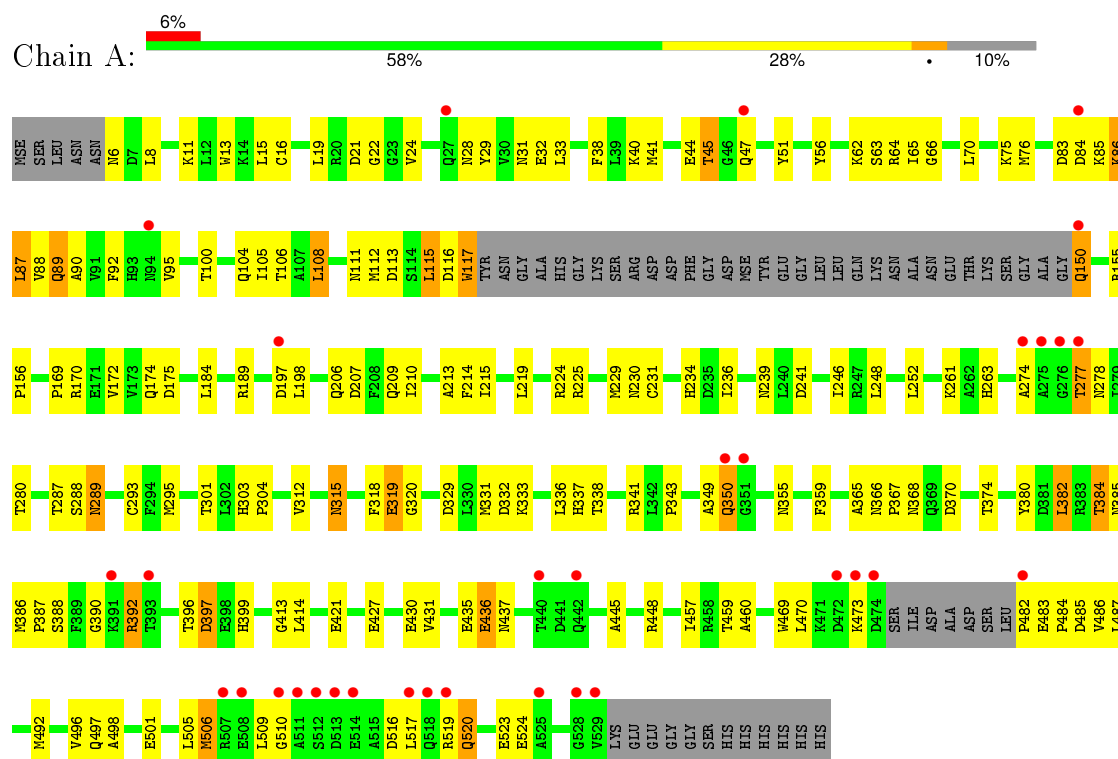
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	86	Total O 86 86	0	0
3	B	57	Total O 57 57	0	0

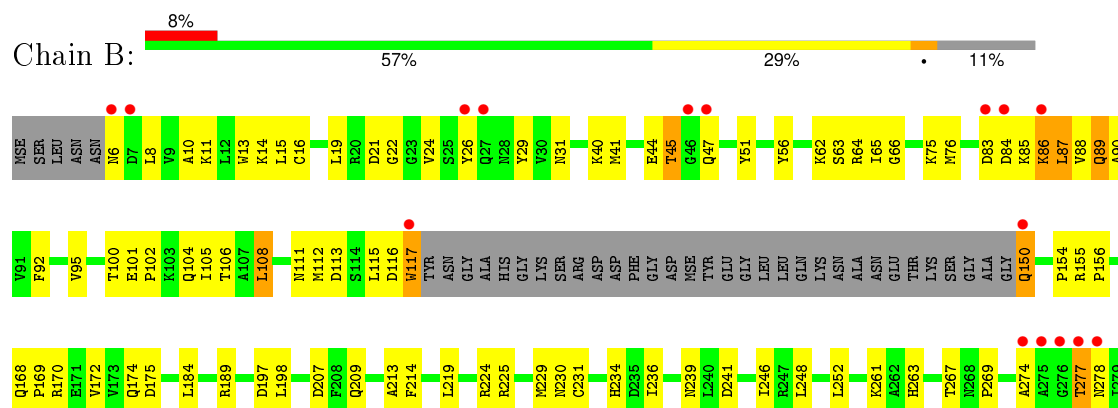
### 3 Residue-property plots [i](#)

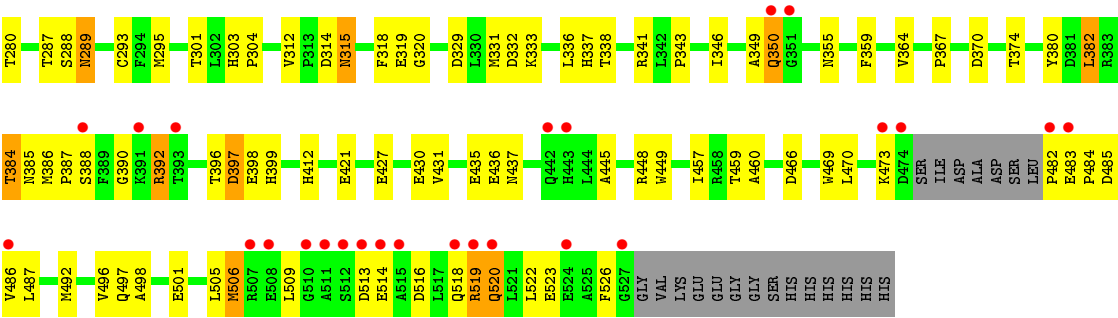
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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: Type I restriction enzyme EcoKI M protein



- Molecule 1: Type I restriction enzyme EcoKI M protein





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	105.43Å 105.43Å 138.28Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.81 – 2.80 30.03 – 2.80	Depositor EDS
% Data completeness (in resolution range)	95.3 (19.81-2.80) 98.9 (30.03-2.80)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.11	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.40 (at 2.80Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.224 , 0.261 0.230 , 0.267	Depositor DCC
$R_{free}$ test set	1810 reflections (5.31%)	DCC
Wilson B-factor (Å <sup>2</sup> )	38.9	Xtriage
Anisotropy	0.058	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 55.4	EDS
Estimated twinning fraction	0.044 for h,-k,-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 71685 reflections	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	7828	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.73% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: UNX

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	0.44	0/3915	0.59	0/5292
1	B	0.44	0/3904	0.59	0/5277
All	All	0.44	0/7819	0.59	0/10569

There are no bond length outliers.

There are no bond angle outliers.

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	3843	0	3743	145	0
1	B	3832	0	3732	140	0
2	A	10	0	0	0	0
3	A	86	0	0	4	0
3	B	57	0	0	2	0
All	All	7828	0	7475	278	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 18.

The worst 5 of 278 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:287:THR:HG22	1:B:289:ASN:H	1.23	1.03
1:A:287:THR:HG22	1:A:289:ASN:H	1.23	1.01
1:B:117:TRP:HB2	3:B:563:HOH:O	1.65	0.95
1:B:219:LEU:HD22	1:B:277:THR:HG23	1.53	0.90
1:A:219:LEU:HD22	1:A:277:THR:HG23	1.57	0.86

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	479/541 (88%)	445 (93%)	27 (6%)	7 (2%)	13	40
1	B	477/541 (88%)	439 (92%)	32 (7%)	6 (1%)	15	44
All	All	956/1082 (88%)	884 (92%)	59 (6%)	13 (1%)	14	42

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	86	LYS
1	A	349	ALA
1	B	86	LYS
1	B	349	ALA
1	A	350	GLN

### 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	418/451 (93%)	385 (92%)	33 (8%)	15	40
1	B	417/451 (92%)	381 (91%)	36 (9%)	13	36
All	All	835/902 (93%)	766 (92%)	69 (8%)	14	38

5 of 69 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	520	GLN
1	B	100	THR
1	B	506	MSE
1	A	524	GLU
1	B	64	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 47 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	442	GLN
1	B	67	GLN
1	B	399	HIS
1	B	6	ASN
1	B	71	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 10 ligands modelled in this entry, 10 are unknown - 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

There are no such residues in this entry.

## 5.8 Polymer linkage issues

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	476/541 (87%)	0.25	33 (6%)	20 11	12, 40, 78, 87	0
1	B	474/541 (87%)	0.25	41 (8%)	13 6	13, 40, 79, 88	0
All	All	950/1082 (87%)	0.25	74 (7%)	16 8	12, 40, 79, 88	0

The worst 5 of 74 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	276	GLY	7.7
1	B	511	ALA	7.1
1	A	513	ASP	6.7
1	A	511	ALA	6.2
1	A	474	ASP	5.6

### 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 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. 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( $\text{\AA}^2$ )	Q<0.9
2	UNX	A	1001	1/1	0.68	0.59	-	42,42,42,42	0
2	UNX	A	1008	1/1	0.90	0.83	-	24,24,24,24	0
2	UNX	A	1010	1/1	0.86	0.90	-	43,43,43,43	0
2	UNX	A	1005	1/1	0.73	0.44	-	27,27,27,27	0
2	UNX	A	1002	1/1	0.81	0.33	-	25,25,25,25	0
2	UNX	A	1009	1/1	-0.07	0.64	-	49,49,49,49	0
2	UNX	A	1006	1/1	0.65	0.88	-	47,47,47,47	0
2	UNX	A	1004	1/1	0.78	0.69	-	34,34,34,34	0
2	UNX	A	1003	1/1	0.73	1.05	-	52,52,52,52	0
2	UNX	A	1007	1/1	0.63	0.51	-	31,31,31,31	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.