



# Full wwPDB X-ray Structure Validation Report ⓘ

May 29, 2020 – 04:03 am BST

PDB ID : 3WA8  
Title : Crystal structure of M. ruber CasB  
Authors : Yuan, Y.A.; Yuan, Z.  
Deposited on : 2013-04-28  
Resolution : 2.80 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

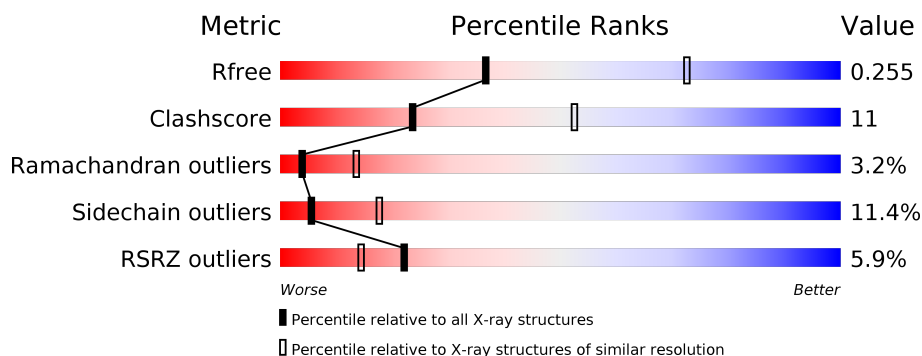
# 1 Overall quality at a glance

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))
$R_{free}$	130704	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (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 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	205	<div> <div>5%</div> <div> <div></div> <div>63%</div> <div>20%</div> <div>•</div> <div>13%</div> </div> </div>
1	B	205	<div> <div>5%</div> <div> <div></div> <div>59%</div> <div>21%</div> <div>• •</div> <div>15%</div> </div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 2972 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 CRISPR-associated protein, Cse2 family.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	179	Total	C	N	O	S	0	0	0
			1497	943	278	272	4			
1	B	175	Total	C	N	O	S	0	0	0
			1457	917	269	267	4			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	EXPRESSION TAG	UNP D3PQC9
A	-1	SER	-	EXPRESSION TAG	UNP D3PQC9
A	0	HIS	-	EXPRESSION TAG	UNP D3PQC9
B	-2	GLY	-	EXPRESSION TAG	UNP D3PQC9
B	-1	SER	-	EXPRESSION TAG	UNP D3PQC9
B	0	HIS	-	EXPRESSION TAG	UNP D3PQC9

- Molecule 2 is MERCURY (II) ION (three-letter code: HG) (formula: Hg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	2	Total	Hg	0	0
			2	2		
2	A	2	Total	Hg	0	0
			2	2		

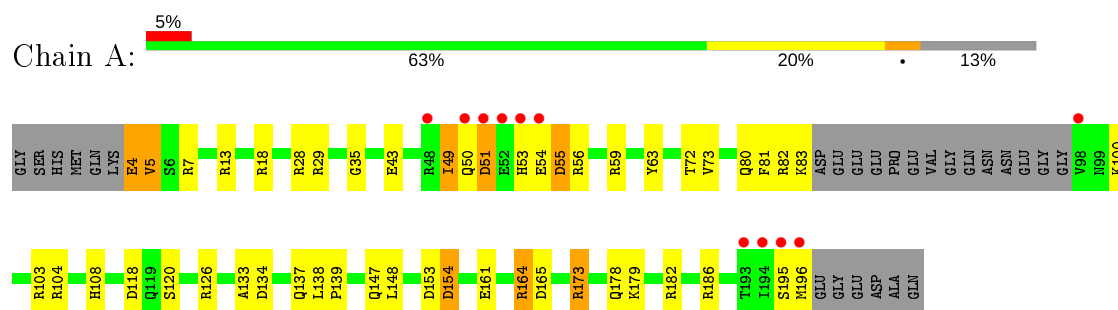
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	5	Total	O	0	0
			5	5		
3	B	9	Total	O	0	0
			9	9		

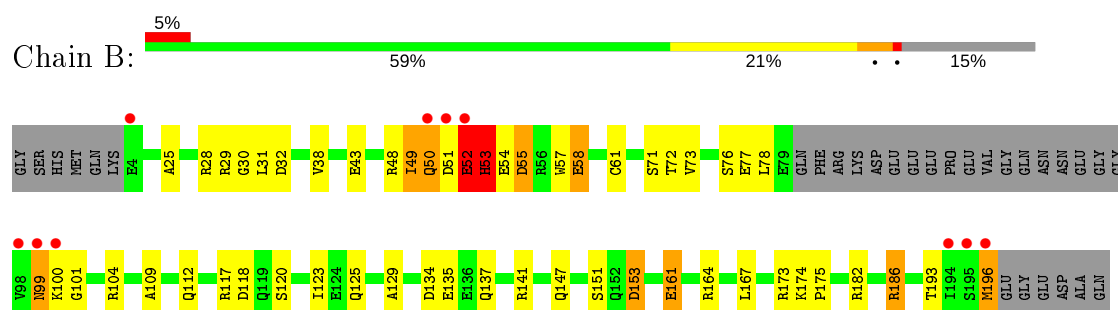
### 3 Residue-property plots [i](#)

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: CRISPR-associated protein, Cse2 family



- Molecule 1: CRISPR-associated protein, Cse2 family



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	71.82Å 75.38Å 112.67Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.13 – 2.80 45.13 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.8 (45.13-2.80) 99.9 (45.13-2.80)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.52 (at 2.81Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.208 , 0.259 0.206 , 0.255	Depositor DCC
$R_{free}$ test set	780 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	69.9	Xtriage
Anisotropy	0.019	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 42.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.024 for k,h,-l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	2972	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.70% 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: HG

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	0.70	0/1525	0.74	0/2055
1	B	0.76	2/1484 (0.1%)	0.77	0/2002
All	All	0.73	2/3009 (0.1%)	0.76	0/4057

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	52	GLU	CD-OE1	6.05	1.32	1.25
1	B	196	MET	SD-CE	5.60	2.09	1.77

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	1497	0	1489	38	0
1	B	1457	0	1446	33	0
2	A	2	0	0	0	0
2	B	2	0	0	0	0
3	A	5	0	0	0	0
3	B	9	0	0	0	0
All	All	2972	0	2935	66	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 11.

All (66) 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:196:MET:SD	1:B:196:MET:CE	2.09	1.41
1:A:173:ARG:NH1	1:A:173:ARG:HB2	1.59	1.17
1:A:173:ARG:HH11	1:A:173:ARG:HB2	1.22	0.93
1:A:173:ARG:HH11	1:A:173:ARG:CB	1.87	0.87
1:B:53:HIS:O	1:B:55:ASP:N	2.07	0.86
1:A:72:THR:O	1:B:28:ARG:NH1	2.16	0.78
1:A:28:ARG:NH1	1:B:72:THR:O	2.17	0.78
1:B:99:ASN:C	1:B:101:GLY:H	1.90	0.75
1:A:53:HIS:O	1:A:55:ASP:N	2.26	0.68
1:B:53:HIS:HE1	1:B:58:GLU:HB3	1.60	0.67
1:B:134:ASP:H	1:B:137:GLN:NE2	1.92	0.67
1:B:58:GLU:HG3	1:B:167:LEU:HD22	1.76	0.66
1:A:173:ARG:HB2	1:A:173:ARG:CZ	2.24	0.65
1:B:99:ASN:O	1:B:101:GLY:N	2.31	0.63
1:B:109:ALA:HA	1:B:112:GLN:HE21	1.65	0.59
1:A:29:ARG:HD2	1:B:73:VAL:O	2.02	0.59
1:B:49:ILE:O	1:B:51:ASP:N	2.35	0.59
1:B:49:ILE:C	1:B:51:ASP:H	2.07	0.58
1:A:164:ARG:HH21	1:A:179:LYS:HZ1	1.50	0.57
1:A:153:ASP:HB3	1:A:154:ASP:OD1	2.06	0.55
1:A:35:GLY:HA2	1:A:63:TYR:CD1	2.43	0.54
1:A:165:ASP:OD1	1:A:179:LYS:NZ	2.40	0.54
1:A:43:GLU:OE2	1:A:59:ARG:NH2	2.39	0.54
1:A:7:ARG:HD3	1:A:49:ILE:HD13	1.91	0.52
1:B:109:ALA:HA	1:B:112:GLN:NE2	2.23	0.52
1:B:99:ASN:C	1:B:101:GLY:N	2.61	0.52
1:A:173:ARG:NH1	1:A:173:ARG:CB	2.46	0.51
1:B:174:LYS:N	1:B:175:PRO:HD3	2.26	0.51
1:A:164:ARG:HH21	1:A:179:LYS:NZ	2.09	0.50
1:A:53:HIS:C	1:A:55:ASP:N	2.65	0.50
1:A:29:ARG:NH2	1:B:153:ASP:HA	2.27	0.49
1:B:48:ARG:C	1:B:50:GLN:H	2.16	0.49
1:A:154:ASP:OD1	1:A:154:ASP:N	2.47	0.47
1:A:4:GLU:HB2	1:A:5:VAL:H	1.43	0.46
1:B:57:TRP:CE2	1:B:61:CYS:SG	3.09	0.46
1:A:13:ARG:HB2	1:A:13:ARG:HE	1.55	0.46
1:A:182:ARG:O	1:A:186:ARG:HB2	2.16	0.46

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:29:ARG:HH22	1:B:153:ASP:HA	1.82	0.45
1:B:134:ASP:H	1:B:137:GLN:HE21	1.60	0.45
1:B:38:VAL:HG13	1:B:43:GLU:HG3	1.98	0.45
1:A:126:ARG:HH11	1:A:148:LEU:HD11	1.82	0.45
1:A:103:ARG:HA	1:A:108:HIS:CG	2.53	0.44
1:B:30:GLY:O	1:B:32:ASP:N	2.51	0.44
1:A:18:ARG:HH11	1:A:18:ARG:HG3	1.82	0.44
1:A:104:ARG:HD3	1:A:161:GLU:OE2	2.18	0.43
1:A:178:GLN:O	1:A:182:ARG:HG3	2.18	0.43
1:B:57:TRP:CZ3	1:B:135:GLU:HG2	2.52	0.43
1:A:43:GLU:CD	1:A:59:ARG:HH22	2.22	0.43
1:B:25:ALA:O	1:B:29:ARG:HG3	2.18	0.43
1:B:52:GLU:O	1:B:53:HIS:CB	2.67	0.43
1:B:76:SER:O	1:B:78:LEU:N	2.52	0.43
1:A:100:LYS:HB2	1:A:100:LYS:HE3	1.65	0.42
1:A:133:ALA:HA	1:A:137:GLN:HE21	1.84	0.42
1:B:196:MET:CE	1:B:196:MET:CG	2.96	0.42
1:A:134:ASP:H	1:A:137:GLN:NE2	2.18	0.42
1:A:83:LYS:HE3	1:A:83:LYS:HB2	1.77	0.42
1:B:182:ARG:O	1:B:186:ARG:HB2	2.19	0.41
1:B:104:ARG:HB3	1:B:161:GLU:HG3	2.02	0.41
1:A:148:LEU:HD23	1:A:148:LEU:HA	1.97	0.41
1:B:137:GLN:HE21	1:B:137:GLN:HB2	1.48	0.41
1:A:138:LEU:O	1:A:139:PRO:C	2.58	0.41
1:A:18:ARG:HG3	1:A:18:ARG:NH1	2.35	0.41
1:A:7:ARG:HE	1:A:49:ILE:HG21	1.86	0.41
1:A:82:ARG:CB	1:A:82:ARG:HH11	2.35	0.40
1:B:129:ALA:O	1:B:141:ARG:NH1	2.53	0.40
1:B:49:ILE:C	1:B:51:ASP:N	2.72	0.40

There are no symmetry-related clashes.

## 5.3 Torsion angles ⓘ

### 5.3.1 Protein backbone ⓘ

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	175/205 (85%)	160 (91%)	12 (7%)	3 (2%)	9	29
1	B	171/205 (83%)	161 (94%)	2 (1%)	8 (5%)	2	7
All	All	346/410 (84%)	321 (93%)	14 (4%)	11 (3%)	4	13

All (11) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	54	GLU
1	B	31	LEU
1	B	50	GLN
1	B	54	GLU
1	B	99	ASN
1	B	100	LYS
1	A	51	ASP
1	B	49	ILE
1	B	77	GLU
1	B	53	HIS
1	A	80	GLN

### 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	156/176 (89%)	139 (89%)	17 (11%)	6	19
1	B	152/176 (86%)	134 (88%)	18 (12%)	5	16
All	All	308/352 (88%)	273 (89%)	35 (11%)	5	18

All (35) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	4	GLU
1	A	5	VAL
1	A	49	ILE
1	A	50	GLN
1	A	51	ASP

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	55	ASP
1	A	56	ARG
1	A	73	VAL
1	A	81	PHE
1	A	118	ASP
1	A	120	SER
1	A	147	GLN
1	A	154	ASP
1	A	164	ARG
1	A	173	ARG
1	A	195	SER
1	A	196	MET
1	B	52	GLU
1	B	53	HIS
1	B	55	ASP
1	B	58	GLU
1	B	71	SER
1	B	117	ARG
1	B	118	ASP
1	B	120	SER
1	B	123	ILE
1	B	125	GLN
1	B	147	GLN
1	B	151	SER
1	B	153	ASP
1	B	161	GLU
1	B	164	ARG
1	B	173	ARG
1	B	186	ARG
1	B	193	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	137	GLN
1	B	112	GLN
1	B	137	GLN

### 5.3.3 RNA ⓘ

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](#)

Of 4 ligands modelled in this entry, 4 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 ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

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	179/205 (87%)	0.16	11 (6%)	21 13	41, 64, 123, 186	0
1	B	175/205 (85%)	0.11	10 (5%)	23 15	41, 63, 127, 159	0
All	All	354/410 (86%)	0.13	21 (5%)	22 14	41, 64, 127, 186	0

All (21) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	99	ASN	6.4
1	B	196	MET	6.0
1	A	194	ILE	5.9
1	A	52	GLU	5.0
1	A	53	HIS	4.6
1	A	51	ASP	4.5
1	A	50	GLN	4.2
1	B	4	GLU	3.6
1	A	193	THR	3.6
1	A	54	GLU	3.6
1	B	52	GLU	3.2
1	B	98	VAL	3.2
1	A	196	MET	3.0
1	A	195	SER	3.0
1	B	194	ILE	2.8
1	B	50	GLN	2.8
1	B	100	LYS	2.7
1	B	195	SER	2.7
1	B	51	ASP	2.6
1	A	98	VAL	2.4
1	A	48	ARG	2.1

## 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
2	HG	A	302	1/1	0.96	0.07	141,141,141,141	0
2	HG	B	301	1/1	0.97	0.09	103,103,103,103	0
2	HG	B	302	1/1	0.98	0.10	162,162,162,162	0
2	HG	A	301	1/1	0.99	0.10	93,93,93,93	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.