



# Full wwPDB X-ray Structure Validation Report ⓘ

May 22, 2020 – 07:35 am BST

PDB ID : 5C2C  
Title : GWS1B RubisCO: Form II RubisCO derived from uncultivated Gallionellacea species (unliganded form)  
Authors : Arbing, M.A.; Varaljay, V.A.; Satagopan, S.; Tabita, F.R.  
Deposited on : 2015-06-15  
Resolution : 2.09 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
Mogul	:	1.8.5 (274361), CSD as541be (2020)
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

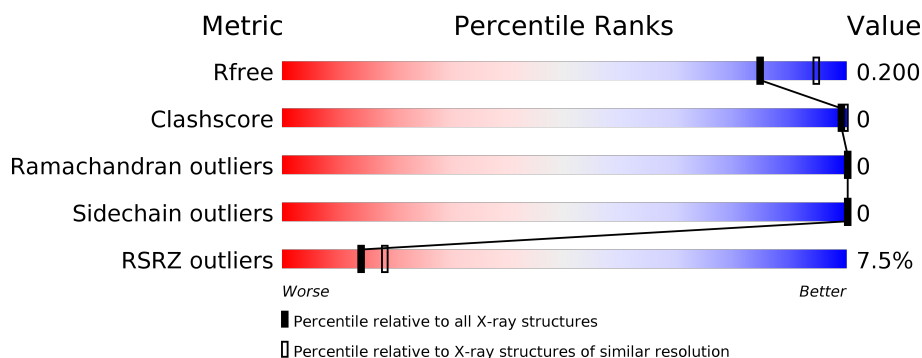
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.09 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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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	479	<div> <div style="width: 100%; height: 10px; background-color: red;"></div> <div style="width: 93%; height: 10px; background-color: green;"></div> <div style="width: 6%; height: 10px; background-color: grey;"></div> <div style="position: absolute; left: 0; top: 5px;">%</div> <div style="position: absolute; left: 48%; top: 5px;">93%</div> <div style="position: absolute; left: 95%; top: 5px;">6%</div> </div>
1	B	479	<div> <div style="width: 13%; height: 10px; background-color: red;"></div> <div style="width: 90%; height: 10px; background-color: green;"></div> <div style="width: 9%; height: 10px; background-color: grey;"></div> <div style="position: absolute; left: 13%; top: 5px;">13%</div> <div style="position: absolute; left: 48%; top: 5px;">90%</div> <div style="position: absolute; left: 95%; top: 5px;">9%</div> </div>
1	C	479	<div> <div style="width: 8%; height: 10px; background-color: red;"></div> <div style="width: 92%; height: 10px; background-color: green;"></div> <div style="width: 7%; height: 10px; background-color: grey;"></div> <div style="position: absolute; left: 8%; top: 5px;">8%</div> <div style="position: absolute; left: 48%; top: 5px;">92%</div> <div style="position: absolute; left: 95%; top: 5px;">7%</div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 20235 atoms, of which 9446 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 Form II RubisCO.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	448	Total	C	H	N	O	S	0	0	0
			6650	2185	3221	589	635	20			
1	B	435	Total	C	H	N	O	S	0	0	0
			6325	2068	3058	562	617	20			
1	C	447	Total	C	H	N	O	S	0	0	0
			6551	2155	3167	579	630	20			

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	2	Total	Mg	0	0
			2	2		
2	A	2	Total	Mg	0	0
			2	2		
2	C	2	Total	Mg	0	0
			2	2		

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total	Cl	0	0
			1	1		

- Molecule 4 is water.

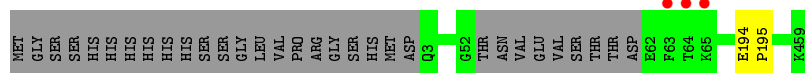
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	280	Total	O	0	0
			280	280		
4	B	211	Total	O	0	0
			211	211		
4	C	211	Total	O	0	0
			211	211		

### 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.

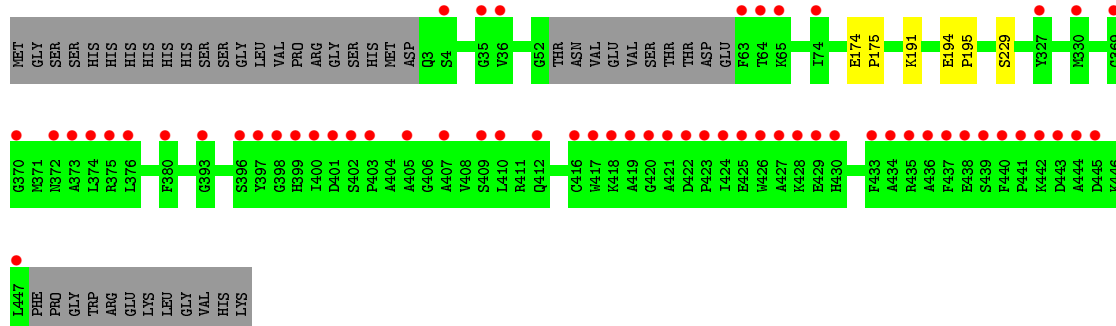
#### • Molecule 1: Form II RubisCO

Chain A: 




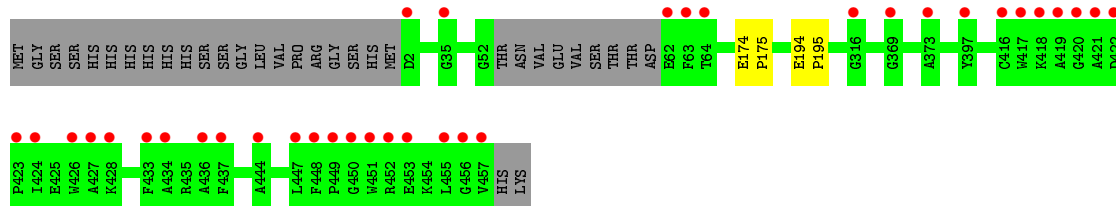
#### • Molecule 1: Form II RubisCO

Chain B: 



#### • Molecule 1: Form II RubisCO

Chain C: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	132.22Å 169.31Å 76.69Å 90.00° 119.20° 90.00°	Depositor
Resolution (Å)	49.36 – 2.09 49.36 – 2.09	Depositor EDS
% Data completeness (in resolution range)	99.0 (49.36-2.09) 99.1 (49.36-2.09)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.03 (at 2.08Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, $R_{free}$	0.167 , 0.199 0.170 , 0.200	Depositor DCC
$R_{free}$ test set	4299 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	25.4	Xtriage
Anisotropy	0.536	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.41 , 55.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	20235	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.82% 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: MG, KCX, CL

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	0.23	0/3505	0.40	0/4744
1	B	0.23	0/3335	0.39	0/4523
1	C	0.22	0/3458	0.39	0/4687
All	All	0.23	0/10298	0.39	0/13954

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

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	3429	3221	3287	1	0
1	B	3267	3058	3071	3	0
1	C	3384	3167	3209	2	0
2	A	2	0	0	0	0
2	B	2	0	0	0	0
2	C	2	0	0	0	0
3	B	1	0	0	0	0
4	A	280	0	0	0	0
4	B	211	0	0	0	0
4	C	211	0	0	0	0
All	All	10789	9446	9567	6	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 0.

All (6) 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:174:GLU:HB2	1:B:175:PRO:HD3	1.96	0.46
1:C:174:GLU:HB2	1:C:175:PRO:HD3	2.00	0.44
1:B:191:KCX:HB2	1:B:229:SER:HB3	2.02	0.42
1:B:194:GLU:N	1:B:195:PRO:CD	2.83	0.42
1:C:194:GLU:N	1:C:195:PRO:CD	2.82	0.42
1:A:194:GLU:N	1:A:195:PRO:CD	2.84	0.41

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	443/479 (92%)	424 (96%)	19 (4%)	0	100	100
1	B	430/479 (90%)	412 (96%)	18 (4%)	0	100	100
1	C	442/479 (92%)	426 (96%)	16 (4%)	0	100	100
All	All	1315/1437 (92%)	1262 (96%)	53 (4%)	0	100	100

There are no Ramachandran outliers to report.

### 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	342/381 (90%)	342 (100%)	0	100	100
1	B	319/381 (84%)	319 (100%)	0	100	100
1	C	332/381 (87%)	332 (100%)	0	100	100
All	All	993/1143 (87%)	993 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

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

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

3 non-standard protein/DNA/RNA residues 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	$\# Z  > 2$
1	KCX	C	191	1	7,11,12	0.63	0	4,12,14	0.31	0
1	KCX	B	191	1,2	7,11,12	0.63	0	4,12,14	0.51	0
1	KCX	A	191	1,2	7,11,12	0.73	0	4,12,14	1.19	0

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
1	KCX	C	191	1	-	3/7/10/12	-
1	KCX	B	191	1,2	-	5/7/10/12	-
1	KCX	A	191	1,2	-	0/7/10/12	-



There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (8) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	C	191	KCX	O-C-CA-CB
1	B	191	KCX	CA-CB-CG-CD
1	C	191	KCX	CA-CB-CG-CD
1	B	191	KCX	CD-CE-NZ-CX
1	B	191	KCX	CG-CD-CE-NZ
1	C	191	KCX	CE-CD-CG-CB
1	B	191	KCX	N-CA-CB-CG
1	B	191	KCX	C-CA-CB-CG

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	B	191	KCX	1	0

## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

Of 7 ligands modelled in this entry, 7 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 ⓘ

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	447/479 (93%)	0.02	3 (0%) 87 89	20, 29, 51, 96	0
1	B	434/479 (90%)	0.54	60 (13%) 2 4	21, 34, 81, 105	0
1	C	446/479 (93%)	0.42	36 (8%) 12 15	23, 33, 69, 102	0
All	All	1327/1437 (92%)	0.32	99 (7%) 14 18	20, 32, 72, 105	0

All (99) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	419	ALA	7.7
1	B	426	TRP	6.9
1	B	423	PRO	6.3
1	B	436	ALA	5.7
1	B	433	PHE	5.6
1	C	455	LEU	5.6
1	B	428	LYS	5.2
1	B	64	THR	5.2
1	B	422	ASP	5.2
1	A	63	PHE	5.0
1	B	444	ALA	4.6
1	B	443	ASP	4.6
1	B	424	ILE	4.5
1	B	441	PRO	4.4
1	B	427	ALA	4.3
1	C	451	TRP	4.2
1	B	420	GLY	4.1
1	B	440	PHE	4.1
1	B	376	LEU	4.1
1	B	437	PHE	4.0
1	C	420	GLY	4.0
1	C	433	PHE	4.0
1	B	445	ASP	4.0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	C	419	ALA	4.0
1	B	421	ALA	3.9
1	C	457	VAL	3.9
1	B	430	HIS	3.9
1	C	448	PHE	3.9
1	A	64	THR	3.8
1	B	327	TYR	3.8
1	B	369	GLY	3.8
1	B	405	ALA	3.7
1	B	398	GLY	3.7
1	B	397	TYR	3.6
1	C	444	ALA	3.5
1	B	416	CYS	3.5
1	C	421	ALA	3.5
1	B	425	GLU	3.5
1	B	429	GLU	3.5
1	C	426	TRP	3.5
1	C	447	LEU	3.4
1	C	456	GLY	3.4
1	C	423	PRO	3.3
1	C	427	ALA	3.3
1	B	418	LYS	3.2
1	B	396	SER	3.2
1	B	417	TRP	3.2
1	C	450	GLY	3.2
1	C	2	ASP	3.1
1	B	403	PRO	3.1
1	C	428	LYS	3.1
1	B	435	ARG	3.1
1	C	449	PRO	3.1
1	B	63	PHE	3.1
1	B	374	LEU	3.0
1	C	63	PHE	3.0
1	B	447	LEU	3.0
1	C	417	TRP	2.9
1	B	393	GLY	2.9
1	C	437	PHE	2.9
1	A	65	LYS	2.9
1	B	400	ILE	2.8
1	B	442	LYS	2.8
1	B	434	ALA	2.8
1	C	64	THR	2.7

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	C	418	LYS	2.7
1	B	4	SER	2.7
1	C	416	CYS	2.5
1	B	65	LYS	2.5
1	B	375	ARG	2.5
1	B	439	SER	2.5
1	B	438	GLU	2.4
1	B	35	GLY	2.4
1	B	36	VAL	2.4
1	B	372	ASN	2.4
1	C	422	ASP	2.4
1	C	452	ARG	2.4
1	B	401	ASP	2.4
1	B	412	GLN	2.3
1	B	380	PHE	2.3
1	C	434	ALA	2.3
1	C	373	ALA	2.2
1	B	373	ALA	2.2
1	B	399	HIS	2.2
1	C	397	TYR	2.2
1	C	369	GLY	2.2
1	C	453	GLU	2.1
1	B	330	MET	2.1
1	B	407	ALA	2.1
1	B	74	ILE	2.1
1	C	424	ILE	2.1
1	C	35	GLY	2.1
1	B	402	SER	2.1
1	B	370	GLY	2.1
1	C	62	GLU	2.0
1	C	316	GLY	2.0
1	C	436	ALA	2.0
1	B	409	SER	2.0
1	B	410	LEU	2.0

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

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
1	KCX	A	191	12/13	0.90	0.21	31,46,60,64	0
1	KCX	B	191	12/13	0.93	0.14	25,47,69,73	0
1	KCX	C	191	12/13	0.93	0.19	30,50,70,74	0

### 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	MG	C	501	1/1	0.82	0.11	52,52,52,52	0
2	MG	C	502	1/1	0.92	0.12	30,30,30,30	0
3	CL	B	501	1/1	0.93	0.10	59,59,59,59	0
2	MG	A	501	1/1	0.94	0.14	42,42,42,42	0
2	MG	B	503	1/1	0.95	0.07	41,41,41,41	0
2	MG	A	502	1/1	0.95	0.13	34,34,34,34	0
2	MG	B	502	1/1	0.97	0.12	51,51,51,51	0

### 6.5 Other polymers [i](#)

There are no such residues in this entry.