Full wwPDB X-ray Structure Validation Report ⓘ

Dec 4, 2023 – 05:06 am GMT

PDB ID	:	1H5V
Title	:	Thiopentasaccharide complex of the endoglucanase Cel5A from Bacillus agaradharens at 1.1 A resolution in the tetragonal crystal form
Authors	:	Varrot, A.; Sulzenbacher, G.; Schulein, M.; Driguez, H.; Davies, G.J.
Deposited on	:	2001-05-28
Resolution	:	1.10 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*
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:

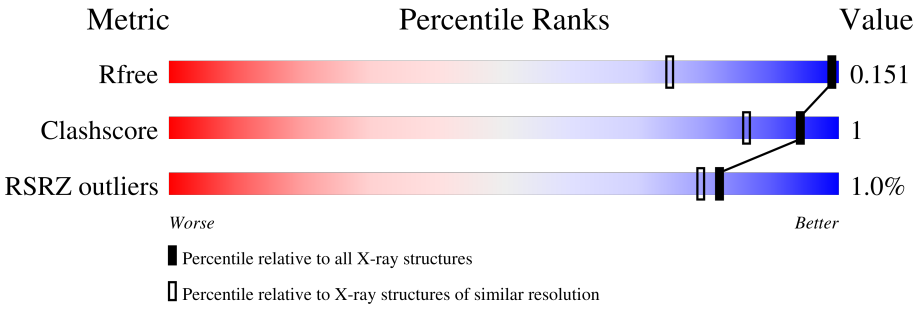
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

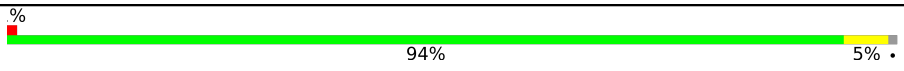
The reported resolution of this entry is 1.10 Å.

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	1619 (1.14-1.06)
Clashscore	141614	1671 (1.14-1.06)
RSRZ outliers	127900	1588 (1.14-1.06)

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 >=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 <=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	304	
2	B	5	

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
2	GLC	B	5	X	-	-	-

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 3035 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 ENDOGLUCANASE 5A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	301	Total	C	N	O	S	0	11	0
			2429	1531	399	491	8			

- Molecule 2 is an oligosaccharide called alpha-D-glucopyranose-(1-4)-1,4-dithio-beta-D-glucopyranose-(1-4)-4-thio-beta-D-glucopyranose-(1-4)-4-thio-beta-D-glucopyranose-(1-4)-methyl 4-thio-alpha-D-glucopyranoside.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	B	5	Total	C	O	S	0	0	0
			57	31	22	4			

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	10	Total	Ca	0	1
			10	10		

- Molecule 4 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	2	Total	Na	0	0
			2	2		

- Molecule 5 is water.

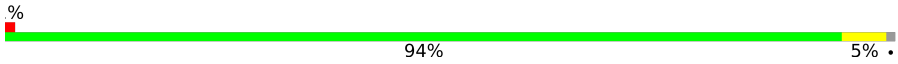
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	537	Total	O	0	88
			537	537		

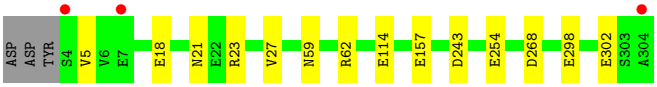
3 Residue-property plots (i)

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: ENDOGLUCANASE 5A

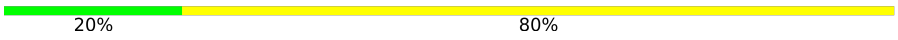
Chain A:





• Molecule 2: alpha-D-glucopyranose-(1-4)-1,4-dithio-beta-D-glucopyranose-(1-4)-4-thio-beta-D-glucopyranose-(1-4)-4-thio-beta-D-glucopyranose-(1-4)-methyl 4-thio-alpha-D-glucopyranoside

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	75.17Å 75.17Å 135.47Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 1.10 19.93 – 1.10	Depositor EDS
% Data completeness (in resolution range)	99.5 (20.00-1.10) 99.6 (19.93-1.10)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.34 (at 1.10Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.120 , 0.138 0.136 , 0.151	Depositor DCC
R_{free} test set	7867 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	11.0	Xtriage
Anisotropy	0.006	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 43.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.98	EDS
Total number of atoms	3035	wwPDB-VP
Average B, all atoms (Å^2)	14.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.15% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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: GLC, CA, NA, SSG, SGC, MA3

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	0.86	2/2533 (0.1%)	0.85	3/3447 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	114	GLU	CD-OE1	-6.43	1.18	1.25
1	A	157	GLU	CD-OE2	-5.90	1.19	1.25

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	243	ASP	CB-CG-OD1	5.94	123.65	118.30
1	A	23	ARG	NE-CZ-NH2	-5.64	117.48	120.30
1	A	268	ASP	CB-CG-OD1	5.12	122.91	118.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	62	ARG	Sidechain

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	2429	0	2236	6	0
2	B	57	0	50	0	0
3	A	10	0	0	0	0
4	A	2	0	0	0	0
5	A	537	0	0	2	0
All	All	3035	0	2286	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 1.

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:A:18:GLU:HB3	5:A:2045[A]:HOH:O	1.88	0.72
1:A:21:ASN:HB3	1:A:27[B]:VAL:CG2	2.37	0.55
1:A:254:GLU:HG3	5:A:2443[B]:HOH:O	2.06	0.53
1:A:21:ASN:HB3	1:A:27[B]:VAL:HG21	1.91	0.52
1:A:5:VAL:HG11	1:A:59:ASN:HB2	1.98	0.45
1:A:298:GLU:O	1:A:302:GLU:HG2	2.18	0.42

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

There are no protein backbone outliers to report in this entry.

5.3.2 Protein sidechains

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

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)

5 monosaccharides 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	$\#\|Z\| > 2$
2	MA3	B	1	2	12,13,13	0.91	1 (8%)	14,18,18	1.31	1 (7%)
2	SGC	B	2	2	10,11,12	1.05	1 (10%)	12,15,17	1.34	1 (8%)
2	SGC	B	3	2	10,11,12	1.35	2 (20%)	12,15,17	1.69	3 (25%)
2	SSG	B	4	2	10,11,12	0.85	0	12,15,17	1.26	1 (8%)
2	GLC	B	5	2	11,11,12	0.61	0	15,15,17	0.70	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
2	MA3	B	1	2	-	0/4/24/24	0/1/1/1
2	SGC	B	2	2	-	0/2/19/22	0/1/1/1
2	SGC	B	3	2	-	0/2/19/22	0/1/1/1
2	SSG	B	4	2	-	0/2/19/22	0/1/1/1
2	GLC	B	5	2	1/1/4/5	0/2/19/22	0/1/1/1

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	3	SGC	C3-C4	-2.87	1.51	1.53
2	B	3	SGC	C2-C3	2.24	1.55	1.52
2	B	1	MA3	C5-C4	2.17	1.54	1.53
2	B	2	SGC	O2-C2	2.00	1.47	1.43

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	1	MA3	O1-C1-C2	3.95	112.78	108.15
2	B	4	SSG	O5-C1-C2	-3.80	104.91	110.77
2	B	2	SGC	O5-C1-C2	-3.78	104.94	110.77
2	B	3	SGC	C1-C2-C3	3.39	113.83	109.67
2	B	3	SGC	C5-C4-S4	2.50	116.23	110.16
2	B	3	SGC	O5-C1-C2	2.05	113.94	110.77

All (1) chirality outliers are listed below:

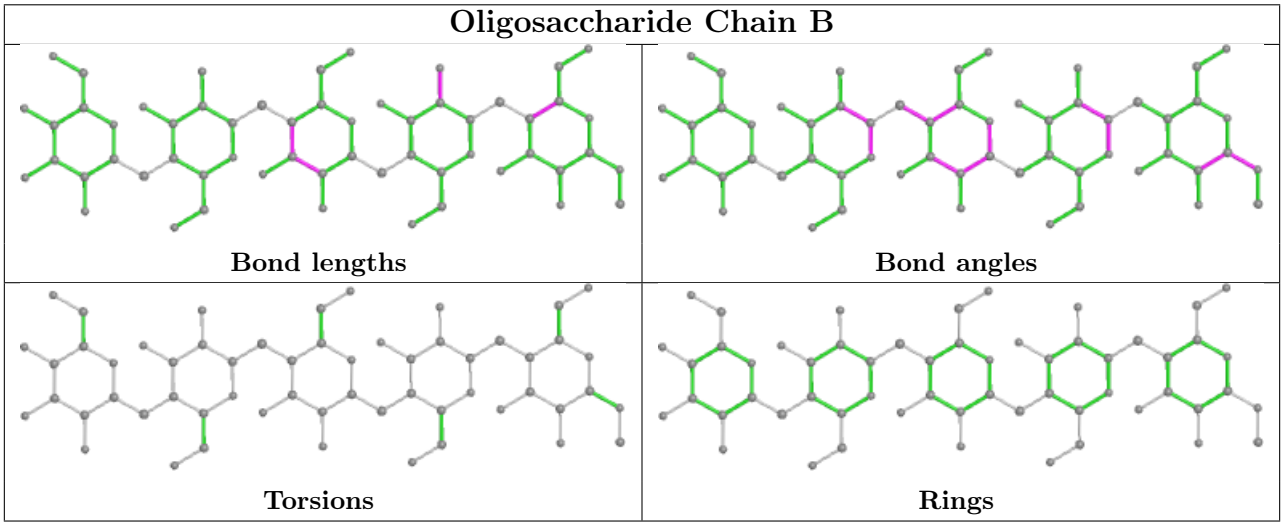
Mol	Chain	Res	Type	Atom
2	B	5	GLC	C1

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for oligosaccharide.



5.6 Ligand geometry

Of 12 ligands modelled in this entry, 12 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

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^{th} 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^2)	Q<0.9
1	A	301/304 (99%)	-0.20	3 (0%)	82	79	7, 11, 18, 27	1 (0%)

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	4	SER	3.6
1	A	304	ALA	3.3
1	A	7	GLU	2.4

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

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

6.3 Carbohydrates

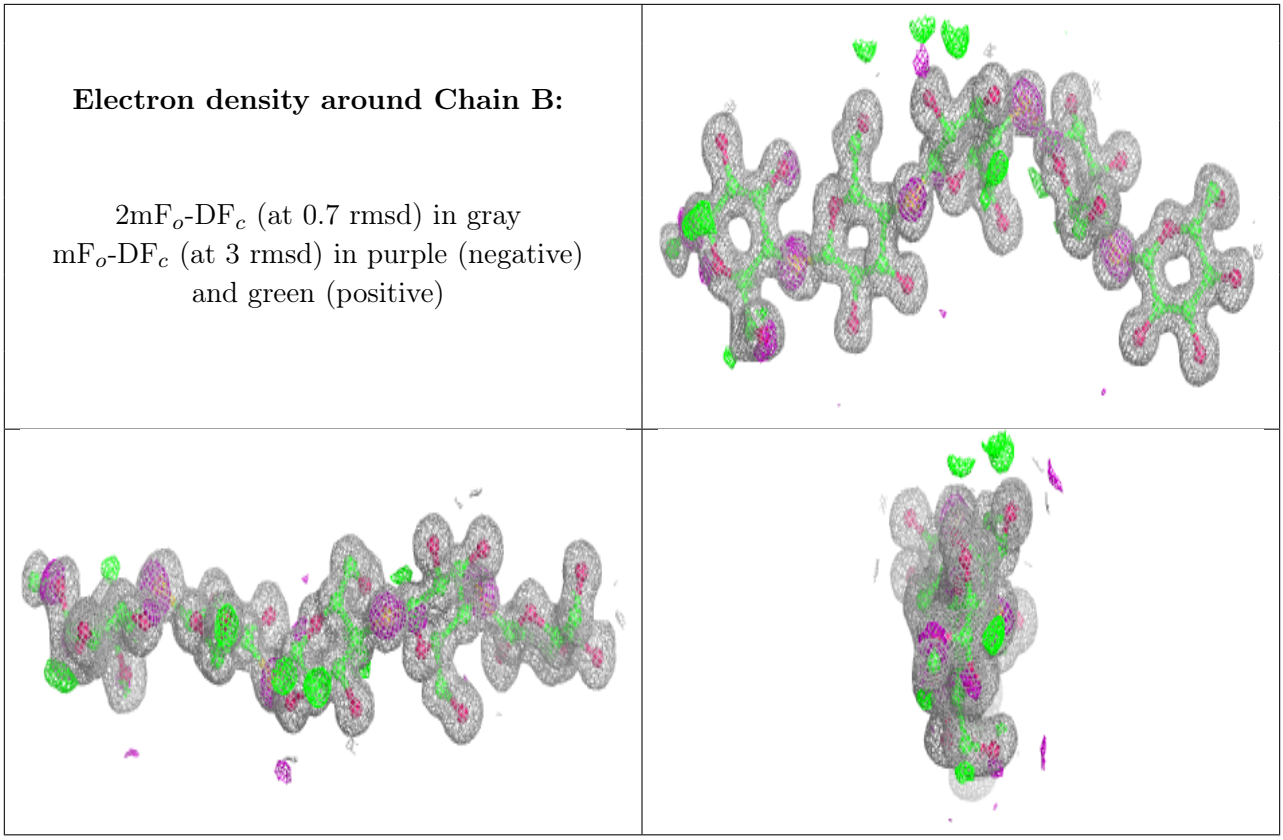
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	MA3	B	1	13/13	0.96	0.11	12,15,21,22	0
2	SGC	B	2	11/12	0.99	0.07	11,12,18,19	0
2	SGC	B	3	11/12	0.99	0.06	8,10,12,13	0
2	SSG	B	4	11/12	0.99	0.06	8,9,9,10	0
2	GLC	B	5	11/12	0.99	0.05	10,10,12,13	0

The following is a graphical depiction of the model fit to experimental electron density for oligosaccharide. Each fit is shown from different orientation to approximate a three-dimensional view.

Electron density around Chain B:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)



6.4 Ligands

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^{th} 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(Å^2)	Q<0.9
4	NA	A	316	1/1	0.91	0.65	30,30,30,30	0
3	CA	A	313	1/1	0.97	0.34	33,33,33,33	0
3	CA	A	314	1/1	0.98	0.52	38,38,38,38	0
4	NA	A	315	1/1	0.98	0.42	26,26,26,26	0
3	CA	A	312	1/1	0.98	0.48	27,27,27,27	0
3	CA	A	307	1/1	0.99	0.12	19,19,19,19	0
3	CA	A	308[A]	1/1	0.99	0.07	19,19,19,19	1
3	CA	A	309	1/1	0.99	0.37	24,24,24,24	0
3	CA	A	310	1/1	0.99	0.31	22,22,22,22	0
3	CA	A	311	1/1	0.99	0.31	21,21,21,21	0
3	CA	A	305	1/1	1.00	0.17	25,25,25,25	0
3	CA	A	306	1/1	1.00	0.13	17,17,17,17	0

6.5 Other polymers (i)

There are no such residues in this entry.