



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 27, 2025 – 09:14 am GMT

PDB ID : 9H6H / pdb_00009h6h
Title : Bacteroides ovatus GH98 endoxylanase with Seleno-methionine substituents
Authors : Tomlinson, C.W.E.; Cartmell, A.; Bolam, D.N.
Deposited on : 2024-10-24
Resolution : 2.00 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

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 : 4-5-2 with Phenix2.0
Mogul : 1.8.4, CSD as541be (2020)
Xtriage (Phenix) : 2.0
EDS : 3.0
Percentile statistics : 20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4 : 9.0.010 (Gargrove)
Density-Fitness : 1.0.12
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.46

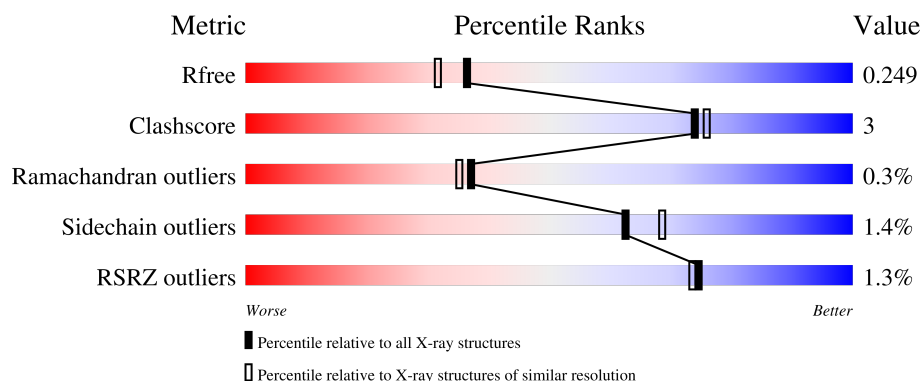
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.00 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	9409 (2.00-2.00)
Clashscore	180529	10737 (2.00-2.00)
Ramachandran outliers	177936	10628 (2.00-2.00)
Sidechain outliers	177891	10627 (2.00-2.00)
RSRZ outliers	164620	9409 (2.00-2.00)

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 $\leq 5\%$. 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	906	89% 9% ..
1	B	906	90% 8% ..

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 29357 atoms, of which 13772 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 Glycosyl hydrolase family 98.

Mol	Chain	Residues	Atoms								ZeroOcc	AltConf	Tra
1	A	897	Total	C	H	N	O	P	S	Se	217	6	0
			14013	4539	6861	1189	1399	2	14	9			
1	B	894	Total	C	H	N	O	P	S	Se	214	2	0
			13962	4524	6841	1187	1386	1	14	9			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	42	MSE	-	initiating methionine	UNP A0A6N3VET1
A	43	ALA	-	expression tag	UNP A0A6N3VET1
A	44	SER	-	expression tag	UNP A0A6N3VET1
A	353	ALA	GLU	engineered mutation	UNP A0A6N3VET1
A	940	LEU	-	expression tag	UNP A0A6N3VET1
A	941	GLU	-	expression tag	UNP A0A6N3VET1
A	942	HIS	-	expression tag	UNP A0A6N3VET1
A	943	HIS	-	expression tag	UNP A0A6N3VET1
A	944	HIS	-	expression tag	UNP A0A6N3VET1
A	945	HIS	-	expression tag	UNP A0A6N3VET1
A	946	HIS	-	expression tag	UNP A0A6N3VET1
A	947	HIS	-	expression tag	UNP A0A6N3VET1
B	42	MSE	-	initiating methionine	UNP A0A6N3VET1
B	43	ALA	-	expression tag	UNP A0A6N3VET1
B	44	SER	-	expression tag	UNP A0A6N3VET1
B	353	ALA	GLU	engineered mutation	UNP A0A6N3VET1
B	940	LEU	-	expression tag	UNP A0A6N3VET1
B	941	GLU	-	expression tag	UNP A0A6N3VET1
B	942	HIS	-	expression tag	UNP A0A6N3VET1
B	943	HIS	-	expression tag	UNP A0A6N3VET1
B	944	HIS	-	expression tag	UNP A0A6N3VET1
B	945	HIS	-	expression tag	UNP A0A6N3VET1
B	946	HIS	-	expression tag	UNP A0A6N3VET1
B	947	HIS	-	expression tag	UNP A0A6N3VET1

- Molecule 2 is DI(HYDROXYETHYL)ETHER (CCD ID: PEG) (formula: $C_4H_{10}O_3$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	H	O	2	0
			17	4	10	3		
2	A	1	Total	C	H	O	2	0
			17	4	10	3		
2	B	1	Total	C	H	O	2	0
			17	4	10	3		
2	B	1	Total	C	H	O	2	0
			17	4	10	3		
2	B	1	Total	C	H	O	2	0
			17	4	10	3		
2	B	1	Total	C	H	O	2	0
			17	4	10	3		

- Molecule 3 is CALCIUM ION (CCD ID: CA) (formula: Ca).

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

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	564	Total 564	O 564	0	0
4	B	697	Total 697	O 697	0	0

- Molecule 1: Glycosyl hydrolase family 98



4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	183.91Å 62.45Å 178.71Å 90.00° 118.06° 90.00°	Depositor
Resolution (Å)	58.40 – 2.00 58.40 – 2.00	Depositor EDS
% Data completeness (in resolution range)	99.6 (58.40-2.00) 99.6 (58.40-2.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.26 (at 2.00Å)	Xtriage
Refinement program	REFMAC 5.8.0430 (refmacat 0.4.82)	Depositor
R, R_{free}	0.192 , 0.242 0.199 , 0.249	Depositor DCC
R_{free} test set	6155 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å ²)	21.9	Xtriage
Anisotropy	0.164	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.40 , 46.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.003 for h,-k,-h-l	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	29357	wwPDB-VP
Average B, all atoms (Å ²)	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.20% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: PHD, PEG, CA

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	0.63	0/7318	1.07	15/9940 (0.2%)
1	B	0.66	0/7286	1.07	11/9897 (0.1%)
All	All	0.64	0/14604	1.07	26/19837 (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	4
1	B	0	4
All	All	0	8

There are no bond length outliers.

The worst 5 of 26 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	412	PHE	CA-CB-CG	-7.09	106.71	113.80
1	A	412	PHE	CA-CB-CG	-7.05	106.75	113.80
1	B	146	ASP	CA-CB-CG	7.00	119.60	112.60
1	A	800	GLN	CB-CA-C	6.99	121.32	109.72
1	A	590	ASP	CA-CB-CG	6.63	119.23	112.60

There are no chirality outliers.

5 of 8 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	248	ALA	Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
1	A	297	ARG	Sidechain
1	A	588	ARG	Sidechain
1	A	60	ARG	Sidechain
1	B	297	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	7152	6861	6825	42	2
1	B	7121	6841	6817	43	2
2	A	14	20	20	0	0
2	B	35	50	50	1	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	564	0	0	8	1
4	B	697	0	0	15	1
All	All	15585	13772	13712	85	4

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 3.

The worst 5 of 85 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:851[A]:PHD:OP1	4:A:1101:HOH:O	1.80	0.97
1:B:547[A]:ARG:NH1	4:B:1101:HOH:O	2.09	0.84
1:B:146:ASP:OD1	1:B:149:ASN:HB2	1.84	0.77
1:B:150:THR:O	1:B:151:VAL:HG13	1.91	0.71
1:B:938:ARG:NH1	1:B:939:LYS:O	2.25	0.70

All (4) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:378:LYS:HZ1	1:A:792:SER:HG[4_647]	1.25	0.35
1:B:378:LYS:HZ1	1:B:792:SER:HG[4_546]	1.28	0.32
1:A:156:ASP:OD1	1:B:511:ASN:ND2[3_545]	2.06	0.14
4:A:1627:HOH:O	4:B:1702:HOH:O[3_555]	2.17	0.03

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	899/906 (99%)	861 (96%)	35 (4%)	3 (0%)	37	35
1	B	893/906 (99%)	854 (96%)	36 (4%)	3 (0%)	37	35
All	All	1792/1812 (99%)	1715 (96%)	71 (4%)	6 (0%)	37	35

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	215	GLY
1	A	139	ASP
1	B	122	HIS
1	A	122	HIS
1	A	46	GLY

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	776/772 (100%)	764 (98%)	12 (2%)	60	66

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	773/772 (100%)	763 (99%)	10 (1%)	65	71
All	All	1549/1544 (100%)	1527 (99%)	22 (1%)	62	68

5 of 22 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	69	ASP
1	B	219	GLU
1	B	189	VAL
1	B	319	LYS
1	A	466	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	598	HIS
1	B	504	GLN
1	A	800	GLN
1	B	302	ASN
1	A	679	ASN

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	PHD	A	851[B]	1	9,11,12	1.24	1 (11%)	10,15,17	1.23	1 (10%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
1	PHD	A	851[A]	1	9,11,12	1.49	1 (11%)	10,15,17	1.35	1 (10%)
1	PHD	B	851	1	9,11,12	1.88	2 (22%)	10,15,17	1.05	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	PHD	A	851[B]	1	-	4/8/11/13	-
1	PHD	A	851[A]	1	-	5/8/11/13	-
1	PHD	B	851	1	-	5/8/11/13	-

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	851[A]	PHD	P-OD1	4.14	1.65	1.59
1	B	851	PHD	P-OD1	3.61	1.64	1.59
1	B	851	PHD	P-OP1	3.54	1.62	1.50
1	A	851[B]	PHD	P-OD1	3.39	1.64	1.59

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	851[B]	PHD	OP3-P-OD1	2.91	114.12	105.25
1	A	851[A]	PHD	CA-CB-CG	2.14	117.34	112.86

There are no chirality outliers.

5 of 14 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
1	A	851[A]	PHD	CG-OD1-P-OP1
1	A	851[B]	PHD	C-CA-CB-CG
1	B	851	PHD	C-CA-CB-CG
1	A	851[B]	PHD	CA-CB-CG-OD2
1	A	851[B]	PHD	N-CA-CB-CG

There are no ring outliers.

3 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	A	851[B]	PHD	2	0
1	A	851[A]	PHD	2	0
1	B	851	PHD	1	0

5.5 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 9 ligands modelled in this entry, 2 are monoatomic - leaving 7 for Mogul analysis.

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	PEG	A	1001	-	6,6,6	0.17	0	5,5,5	0.07	0
2	PEG	B	1004	-	6,6,6	0.20	0	5,5,5	0.14	0
2	PEG	B	1002	-	6,6,6	0.15	0	5,5,5	0.14	0
2	PEG	B	1001	-	6,6,6	0.26	0	5,5,5	0.19	0
2	PEG	B	1003	-	6,6,6	0.18	0	5,5,5	0.22	0
2	PEG	A	1002	-	6,6,6	0.22	0	5,5,5	0.22	0
2	PEG	B	1005	-	6,6,6	0.35	0	5,5,5	0.22	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	PEG	A	1001	-	-	0/4/4/4	-
2	PEG	B	1004	-	-	3/4/4/4	-
2	PEG	B	1002	-	-	3/4/4/4	-
2	PEG	B	1001	-	-	3/4/4/4	-
2	PEG	B	1003	-	-	3/4/4/4	-
2	PEG	A	1002	-	-	3/4/4/4	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	PEG	B	1005	-	-	3/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 18 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	1002	PEG	O1-C1-C2-O2
2	A	1002	PEG	O2-C3-C4-O4
2	B	1001	PEG	O2-C3-C4-O4
2	B	1002	PEG	O1-C1-C2-O2
2	B	1003	PEG	O1-C1-C2-O2

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	1004	PEG	1	0

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, 95th 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	887/906 (97%)	-0.07	12 (1%) 73 72	10, 28, 53, 104	2 (0%)
1	B	884/906 (97%)	-0.24	11 (1%) 76 75	12, 25, 47, 103	1 (0%)
All	All	1771/1812 (97%)	-0.15	23 (1%) 74 73	10, 26, 51, 104	3 (0%)

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	150	THR	5.0
1	B	46	GLY	3.7
1	A	215	GLY	3.1
1	B	148	ASN	3.0
1	B	149	ASN	3.0

6.2 Non-standard residues in protein, DNA, RNA chains [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, 95th 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
1	PHD	A	851[A]	12/13	0.83	0.10	23,27,27,28	16
1	PHD	A	851[B]	12/13	0.83	0.10	26,28,29,30	16
1	PHD	B	851	12/13	0.88	0.10	20,22,37,42	0

6.3 Carbohydrates [i](#)

There are no oligosaccharides 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, 95th 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	PEG	B	1005	7/7	0.77	0.19	30,58,64,65	2
2	PEG	A	1001	7/7	0.78	0.16	30,64,68,69	2
2	PEG	B	1004	7/7	0.80	0.14	30,54,56,56	2
2	PEG	B	1001	7/7	0.80	0.13	30,47,50,50	2
2	PEG	A	1002	7/7	0.81	0.14	30,61,65,65	2
2	PEG	B	1002	7/7	0.87	0.10	30,46,47,48	2
2	PEG	B	1003	7/7	0.88	0.11	30,49,55,55	2
3	CA	A	1003	1/1	0.98	0.04	22,22,22,22	0
3	CA	B	1006	1/1	0.98	0.02	17,17,17,17	0

6.5 Other polymers [i](#)

There are no such residues in this entry.