



# Full wwPDB X-ray Structure Validation Report ⓘ

Jun 24, 2024 – 04:39 PM EDT

PDB ID : 5O1N  
Title : Crystal structure of human aminoacidate semialdehyde synthase, saccharopine dehydrogenase domain with N-[(2S)-2-Pyrrolidinylmethyl]-trifluoromethanesulfonamide bound  
Authors : Kopec, J.; Rembeza, E.; Pena, I.A.; Strain-Damerell, C.; Goubin, S.; Sethi, R.; Velupillai, S.; Talon, R.; Collins, P.; Krojer, T.; McLaughlin, M.; Burgess-Brown, N.; Arrowsmith, C.; Edwards, A.; Bountra, C.; Brennan, P.; von Delft, F.; Arruda, P.; Yue, W.W.  
Deposited on : 2017-05-18  
Resolution : 2.28 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 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 : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.37.1  
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)

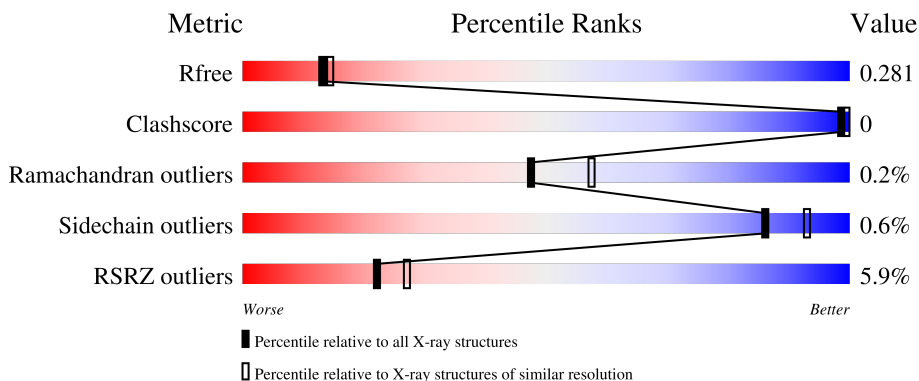
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.28 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	6980 (2.30-2.26)
Clashscore	141614	7711 (2.30-2.26)
Ramachandran outliers	138981	7597 (2.30-2.26)
Sidechain outliers	138945	7598 (2.30-2.26)
RSRZ outliers	127900	6849 (2.30-2.26)

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  $\geq 3$ , 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	497	

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:

Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
 Validation Pipeline (wwPDB-VP) : 2.37.1

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
5	PEG	A	1004	-	-	-	X

## 2 Entry composition i

There are 6 unique types of molecules in this entry. The entry contains 3338 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.

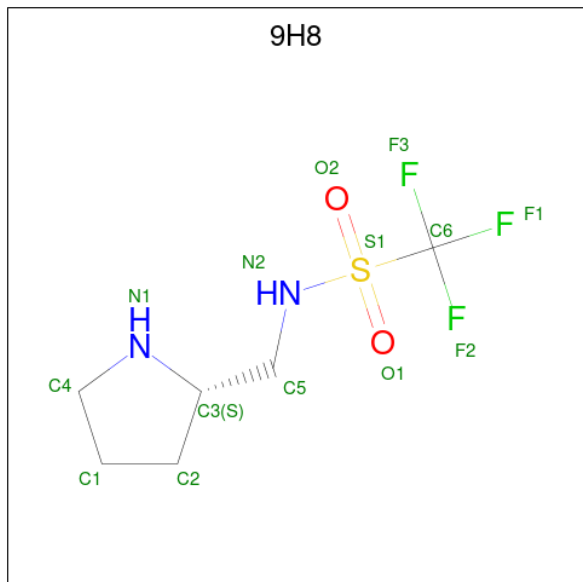
- Molecule 1 is a protein called Alpha-aminoadipic semialdehyde synthase, mitochondrial.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	442	3284	2103	536	625	20	0	0	0

There are 26 discrepancies between the modelled and reference sequences:

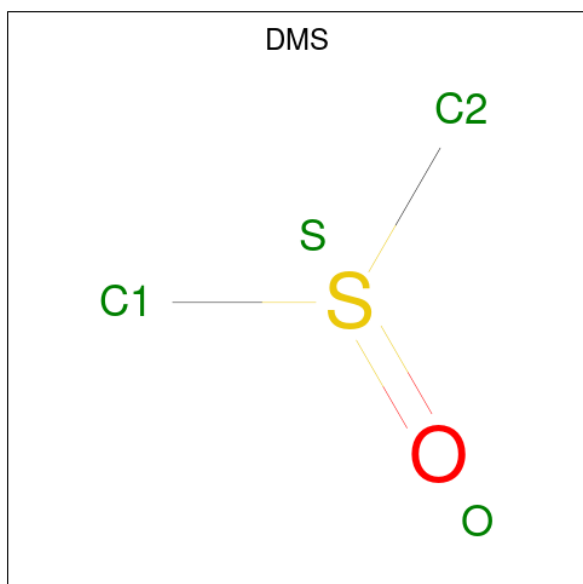
Chain	Residue	Modelled	Actual	Comment	Reference
A	430	MET	-	initiating methionine	UNP Q9UDR5
A	431	GLY	-	expression tag	UNP Q9UDR5
A	432	HIS	-	expression tag	UNP Q9UDR5
A	433	HIS	-	expression tag	UNP Q9UDR5
A	434	HIS	-	expression tag	UNP Q9UDR5
A	435	HIS	-	expression tag	UNP Q9UDR5
A	436	HIS	-	expression tag	UNP Q9UDR5
A	437	HIS	-	expression tag	UNP Q9UDR5
A	438	SER	-	expression tag	UNP Q9UDR5
A	439	SER	-	expression tag	UNP Q9UDR5
A	440	GLY	-	expression tag	UNP Q9UDR5
A	441	VAL	-	expression tag	UNP Q9UDR5
A	442	ASP	-	expression tag	UNP Q9UDR5
A	443	LEU	-	expression tag	UNP Q9UDR5
A	444	GLY	-	expression tag	UNP Q9UDR5
A	445	THR	-	expression tag	UNP Q9UDR5
A	446	GLU	-	expression tag	UNP Q9UDR5
A	447	ASN	-	expression tag	UNP Q9UDR5
A	448	LEU	-	expression tag	UNP Q9UDR5
A	449	TYR	-	expression tag	UNP Q9UDR5
A	450	PHE	-	expression tag	UNP Q9UDR5
A	451	GLN	-	expression tag	UNP Q9UDR5
A	452	SER	-	expression tag	UNP Q9UDR5
A	453	MET	-	expression tag	UNP Q9UDR5
A	454	ALA	-	expression tag	UNP Q9UDR5
A	615	SER	THR	conflict	UNP Q9UDR5

- Molecule 2 is 1,1,1-tris(fluoranyl)- {N}-[[2 {S}]-pyrrolidin-2-yl]methyl]methanesulfonamide (three-letter code: 9H8) (formula:  $C_6H_{11}F_3N_2O_2S$ ).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	
			Total	C	F	N	O			S
2	A	1	14	6	3	2	2	1	0	0

- Molecule 3 is DIMETHYL SULFOXIDE (three-letter code: DMS) (formula:  $C_2H_6OS$ ).



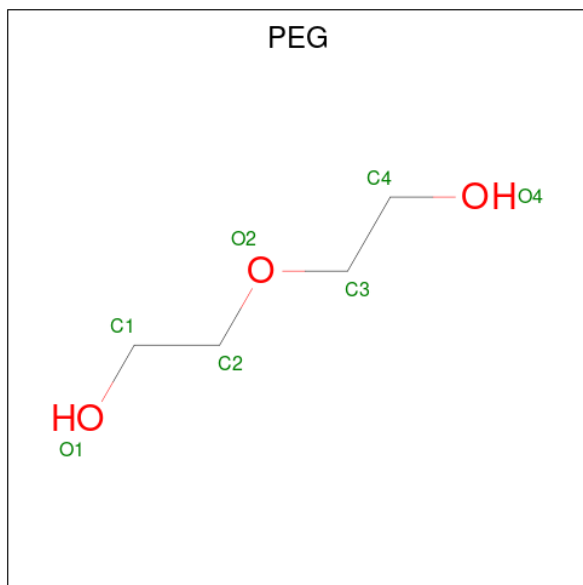
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	O	S		
3	A	1	4	2	1	1	0	0

- Molecule 4 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula:  $C_2H_6O_2$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
4	A	1	4	2	2	0	0

- Molecule 5 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula:  $C_4H_{10}O_3$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
5	A	1	7	4	3	0	0

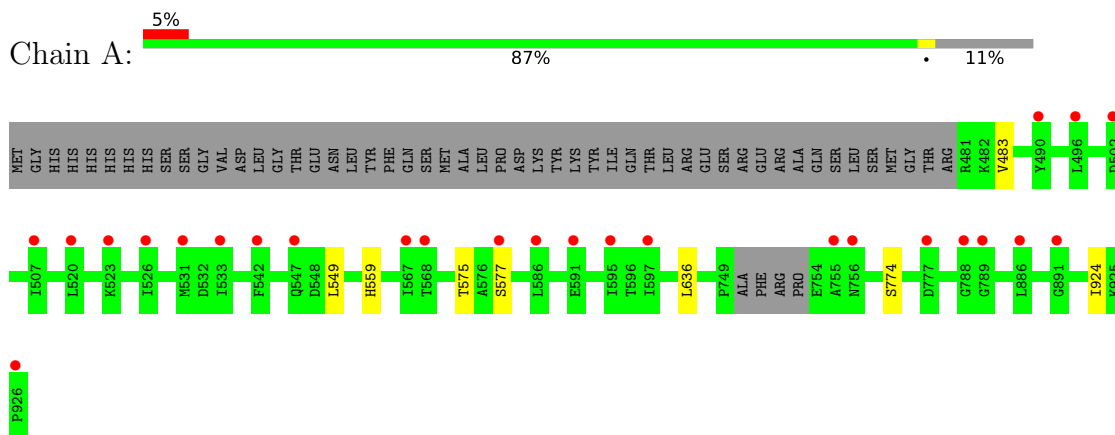
- Molecule 6 is water.

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
6	A	25	Total 25	O 25	0	0

### 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: Alpha-aminoadipic semialdehyde synthase, mitochondrial





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	89.40Å 89.40Å 147.89Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	44.70 – 2.28 44.70 – 2.28	Depositor EDS
% Data completeness (in resolution range)	99.1 (44.70-2.28) 99.1 (44.70-2.28)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.92 (at 2.29Å)	Xtrriage
Refinement program	REFMAC 5.8.0135	Depositor
R, $R_{free}$	0.265 , 0.286 0.264 , 0.281	Depositor DCC
$R_{free}$ test set	1371 reflections (4.92%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	52.0	Xtrriage
Anisotropy	0.488	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 36.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	3338	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	70.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.38% 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: PEG, DMS, EDO, 9H8

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	0.34	0/3346	0.50	0/4547

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	3284	0	3261	3	0
2	A	14	0	0	0	0
3	A	4	0	6	0	0
4	A	4	0	6	0	0
5	A	7	0	10	0	0
6	A	25	0	0	0	0
All	All	3338	0	3283	3	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 (3) 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:559:HIS:CE1	1:A:575:THR:HG23	2.53	0.43
1:A:483:VAL:HG12	1:A:549:LEU:HB3	2.03	0.40
1:A:924:ILE:N	1:A:924:ILE:HD12	2.37	0.40

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	438/497 (88%)	419 (96%)	18 (4%)	1 (0%)	47 57

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	774	SER

### 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	346/418 (83%)	344 (99%)	2 (1%)	86 93

All (2) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	577	SER

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	636	LEU

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

### 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](#)

There are no monosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

4 ligands 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
3	DMS	A	1002	-	3,3,3	0.64	0	3,3,3	0.48	0
5	PEG	A	1004	-	6,6,6	0.35	0	5,5,5	0.39	0
2	9H8	A	1001	-	13,14,14	0.55	0	17,21,21	0.40	0
4	EDO	A	1003	-	3,3,3	0.48	0	2,2,2	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
5	PEG	A	1004	-	-	0/4/4/4	-
2	9H8	A	1001	-	-	9/15/22/22	0/1/1/1
4	EDO	A	1003	-	-	1/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (10) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	1001	9H8	F1-C6-S1-N2
2	A	1001	9H8	F2-C6-S1-N2
2	A	1001	9H8	F3-C6-S1-N2
2	A	1001	9H8	F1-C6-S1-O1
2	A	1001	9H8	F2-C6-S1-O1
2	A	1001	9H8	F3-C6-S1-O1
2	A	1001	9H8	F2-C6-S1-O2
2	A	1001	9H8	F1-C6-S1-O2
2	A	1001	9H8	F3-C6-S1-O2
4	A	1003	EDO	O1-C1-C2-O2

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	442/497 (88%)	0.45	26 (5%) <span style="border: 1px solid red; padding: 2px;">22</span> <span style="border: 1px solid red; padding: 2px;">27</span>	43, 67, 106, 121	0

All (26) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	755	ALA	5.1
1	A	891	GLY	5.0
1	A	926	PRO	4.5
1	A	886	LEU	3.9
1	A	788	GLY	3.8
1	A	533	ILE	3.7
1	A	586	LEU	3.4
1	A	789	GLY	3.3
1	A	756	ASN	3.3
1	A	567	ILE	3.0
1	A	595	ILE	2.8
1	A	597	ILE	2.8
1	A	542	PHE	2.8
1	A	568	THR	2.7
1	A	777	ASP	2.6
1	A	547	GLN	2.6
1	A	523	LYS	2.5
1	A	496	LEU	2.5
1	A	531	MET	2.5
1	A	591	GLU	2.4
1	A	507	ILE	2.4
1	A	577	SER	2.3
1	A	502	ASP	2.3
1	A	526	ILE	2.1
1	A	520	LEU	2.0
1	A	490	TYR	2.0

## 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 monosaccharides 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
5	PEG	A	1004	7/7	0.60	0.58	86,88,90,93	0
4	EDO	A	1003	4/4	0.69	0.19	78,78,79,79	0
3	DMS	A	1002	4/4	0.82	0.25	71,73,73,73	0
2	9H8	A	1001	14/14	0.96	0.14	63,66,67,68	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.