wwPDB X-ray Structure Validation Summary Report (i)

Sep 19, 2023 – 12:30 PM EDT

PDB ID	:	5HOQ
Title	:	Apo structure of CalS11, TDP-rhamnose 3'-o-methyltransferase, an enzyme in Calicheamicin biosynthesis
Authors	:	Han, L.; Helmich, K.E.; Singh, S.; Thorson, J.S.; Bingman, C.A.; Phillips Jr., G.N.; Enzyme Discovery for Natural Product Biosynthesis
Deposited on	:	2016-01-19
Resolution	:	1.79 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 (i) 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 (i)) were used in the production of this report:

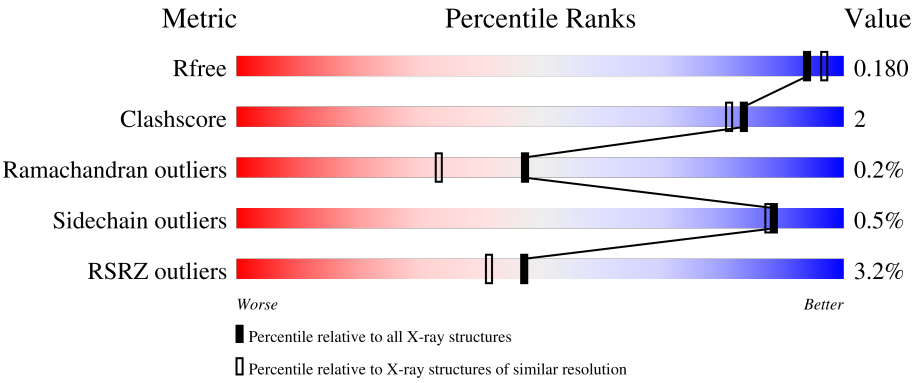
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 1.79 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	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)
RSRZ outliers	127900	5850 (1.80-1.80)

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 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	257	86% · 9%
1	B	257	4% 85% · 11%
1	C	257	3% 85% · 11%
1	D	257	3% 86% · 11%
1	E	257	4% 87% · 9%

2 Entry composition

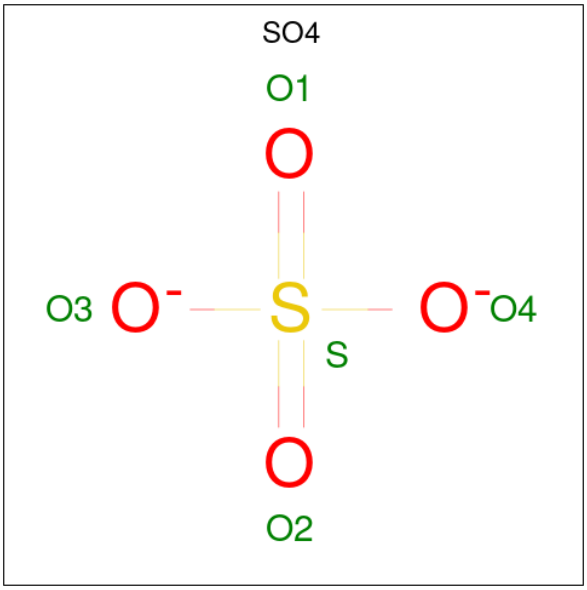
There are 3 unique types of molecules in this entry. The entry contains 20461 atoms, of which 9287 are hydrogens and 0 are deuteriums.

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 TDP-rhamnose 3'-O-methyltransferase (CalS11).

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	234	Total	C	H	N	O	S	0	8	0
			3817	1230	1897	335	350	5			
1	B	230	Total	C	H	N	O	S	0	3	0
			3696	1195	1835	319	342	5			
1	C	229	Total	C	H	N	O	S	0	3	0
			3690	1191	1834	321	339	5			
1	D	230	Total	C	H	N	O	S	0	3	0
			3705	1196	1842	322	340	5			
1	E	235	Total	C	H	N	O	S	0	2	0
			3782	1221	1879	331	346	5			

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	O	S		0	0
			5	4	1			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	1	Total O S 5 4 1	0	0
2	C	1	Total O S 5 4 1	0	0
2	D	1	Total O S 5 4 1	0	0
2	E	1	Total O S 5 4 1	0	0

- Molecule 3 is water.

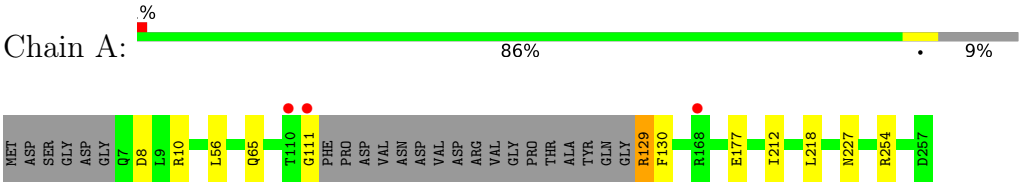
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	386	Total O 386 386	0	0
3	B	348	Total O 348 348	0	0
3	C	323	Total O 323 323	0	0
3	D	358	Total O 358 358	0	0
3	E	331	Total O 331 331	0	0

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.

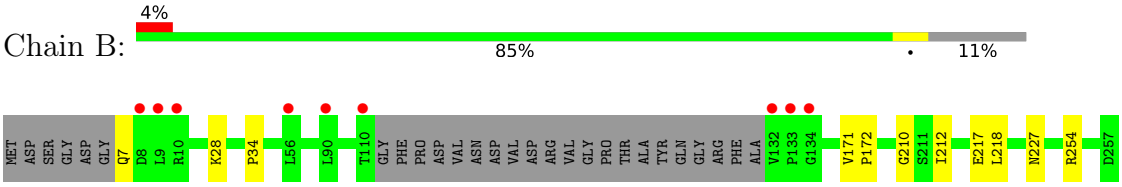
• Molecule 1: TDP-rhamnose 3'-O-methyltransferase (CalS11)

Chain A:



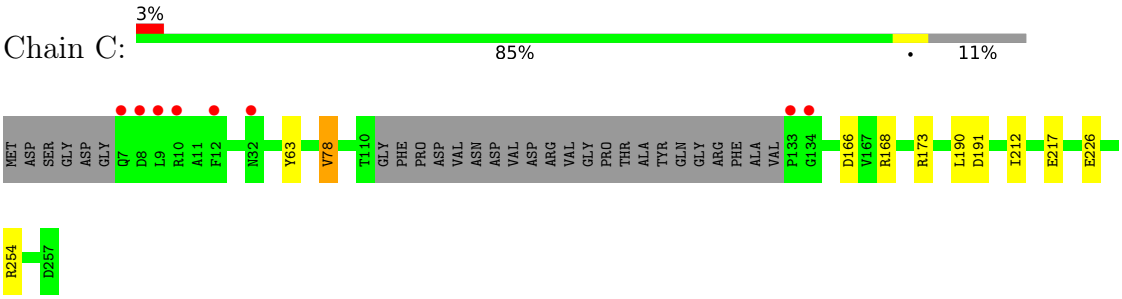
• Molecule 1: TDP-rhamnose 3'-O-methyltransferase (CalS11)

Chain B:



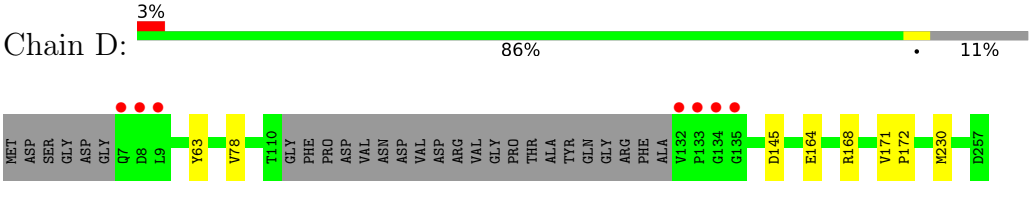
• Molecule 1: TDP-rhamnose 3'-O-methyltransferase (CalS11)

Chain C:



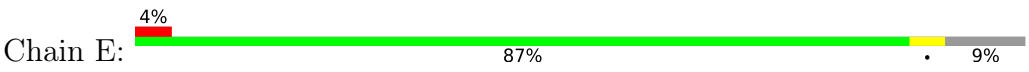
• Molecule 1: TDP-rhamnose 3'-O-methyltransferase (CalS11)

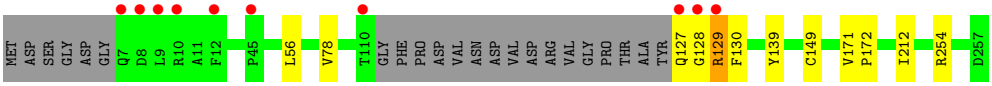
Chain D:



• Molecule 1: TDP-rhamnose 3'-O-methyltransferase (CalS11)

Chain E:




MET
ASP
SER
GLY
ASP
GLY
Q7
D8
L9
R10
A11
F12
P45
L56
V78
T110
GLY
PHE
PRO
ASP
VAL
ASN
ASP
VAL
ASP
ARG
VAL
GLY
PRO
THR
ALA
TYR
Q127
G128
R129
F130
Y139
C149
V171
P172
I212
R254
D257

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	148.26Å 125.14Å 107.15Å 90.00° 125.12° 90.00°	Depositor
Resolution (Å)	47.89 – 1.79 43.54 – 1.80	Depositor EDS
% Data completeness (in resolution range)	98.8 (47.89-1.79) 91.6 (43.54-1.80)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.42 (at 1.79Å)	Xtriage
Refinement program	PHENIX (1.10_2155: ???)	Depositor
R, R_{free}	0.146 , 0.179 0.147 , 0.180	Depositor DCC
R_{free} test set	7375 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	21.5	Xtriage
Anisotropy	0.205	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 66.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	20461	wwPDB-VP
Average B, all atoms (Å^2)	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.36% 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: SO4

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	0.47	0/1991	0.61	0/2708
1	B	0.46	0/1912	0.61	0/2604
1	C	0.46	0/1912	0.63	2/2602 (0.1%)
1	D	0.49	0/1919	0.66	3/2613 (0.1%)
1	E	0.47	1/1951 (0.1%)	0.62	2/2655 (0.1%)
All	All	0.47	1/9685 (0.0%)	0.63	7/13182 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	149	CYS	CB-SG	-5.26	1.73	1.81

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	E	78	VAL	CG1-CB-CG2	8.17	123.97	110.90
1	D	78	VAL	CG1-CB-CG2	7.97	123.65	110.90
1	C	78	VAL	CG1-CB-CG2	7.72	123.25	110.90
1	C	78	VAL	CA-CB-CG1	5.81	119.62	110.90
1	E	78	VAL	CA-CB-CG2	5.70	119.46	110.90

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	1920	1897	1878	12	0
1	B	1861	1835	1837	7	0
1	C	1856	1834	1833	7	0
1	D	1863	1842	1841	4	0
1	E	1903	1879	1884	10	0
2	A	5	0	0	1	0
2	B	5	0	0	1	0
2	C	5	0	0	0	0
2	D	5	0	0	1	0
2	E	5	0	0	1	0
3	A	386	0	0	8	0
3	B	348	0	0	4	1
3	C	323	0	0	4	1
3	D	358	0	0	4	2
3	E	331	0	0	3	2
All	All	11174	9287	9273	44	3

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 2.

The worst 5 of 44 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:7:GLN:N	3:B:402:HOH:O	2.13	0.82
1:C:168[A]:ARG:NH2	3:C:401:HOH:O	2.14	0.79
2:E:301:SO4:O1	3:E:401:HOH:O	2.00	0.78
2:A:301:SO4:O4	3:A:401:HOH:O	2.00	0.77
2:B:301:SO4:O3	3:B:401:HOH:O	2.03	0.74

All (3) 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)
3:B:607:HOH:O	3:C:632:HOH:O[2_555]	1.95	0.25
3:D:648:HOH:O	3:E:580:HOH:O[4_546]	2.01	0.19
3:D:653:HOH:O	3:E:632:HOH:O[4_546]	2.05	0.15

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	236/257 (92%)	228 (97%)	6 (2%)	2 (1%)	19 7
1	B	228/257 (89%)	220 (96%)	8 (4%)	0	100 100
1	C	227/257 (88%)	223 (98%)	4 (2%)	0	100 100
1	D	228/257 (89%)	224 (98%)	4 (2%)	0	100 100
1	E	232/257 (90%)	226 (97%)	5 (2%)	1 (0%)	34 21
All	All	1151/1285 (90%)	1121 (97%)	27 (2%)	3 (0%)	47 27

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	129	ARG
1	A	130[A]	PHE
1	A	130[B]	PHE

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	208/219 (95%)	206 (99%)	2 (1%)	76 71
1	B	201/219 (92%)	200 (100%)	1 (0%)	88 87
1	C	201/219 (92%)	199 (99%)	2 (1%)	76 71
1	D	202/219 (92%)	201 (100%)	1 (0%)	88 87
1	E	204/219 (93%)	204 (100%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	1016/1095 (93%)	1010 (99%)	6 (1%)	88 84

5 of 6 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	63	TYR
1	C	78	VAL
1	D	63	TYR
1	A	129[B]	ARG
1	A	129[A]	ARG

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)

5 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	$\#\|Z\| > 2$
2	SO4	A	301	-	4,4,4	0.17	0	6,6,6	0.18	0
2	SO4	E	301	-	4,4,4	0.16	0	6,6,6	0.16	0
2	SO4	C	301	-	4,4,4	0.18	0	6,6,6	0.23	0
2	SO4	D	301	-	4,4,4	0.19	0	6,6,6	0.32	0
2	SO4	B	301	-	4,4,4	0.18	0	6,6,6	0.22	0

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.

4 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	301	SO4	1	0
2	E	301	SO4	1	0
2	D	301	SO4	1	0
2	B	301	SO4	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

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	234/257 (91%)	-0.39	3 (1%)	77	74	17, 26, 49, 130	0
1	B	230/257 (89%)	-0.09	9 (3%)	39	33	19, 28, 59, 126	0
1	C	229/257 (89%)	-0.18	8 (3%)	44	38	21, 32, 58, 116	0
1	D	230/257 (89%)	-0.40	7 (3%)	50	44	16, 25, 52, 136	0
1	E	235/257 (91%)	-0.16	10 (4%)	35	29	20, 30, 59, 126	0
All	All	1158/1285 (90%)	-0.25	37 (3%)	47	41	16, 28, 57, 136	0

The worst 5 of 37 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	133	PRO	7.2
1	C	7	GLN	6.2
1	E	7	GLN	5.9
1	D	133	PRO	5.9
1	C	134	GLY	5.2

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

There are no monosaccharides in this entry.

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
2	SO4	C	301	5/5	0.95	0.12	61,71,76,92	0
2	SO4	D	301	5/5	0.96	0.11	53,61,71,72	0
2	SO4	A	301	5/5	0.97	0.10	61,61,67,81	0
2	SO4	B	301	5/5	0.97	0.09	53,63,70,85	0
2	SO4	E	301	5/5	0.97	0.09	66,68,79,81	0

6.5 Other polymers (i)

There are no such residues in this entry.