



wwPDB X-ray Structure Validation Summary Report

Feb 28, 2014 – 10:58 PM GMT

PDB ID : 2EC2
Title : Crystal structure of transposase from Sulfolobus tokodaii
Authors : Kawai, K.; Suzuki, A.; Kuramitsu, S.; Masui, R.; Yamane, T.
Deposited on : 2007-02-09
Resolution : 2.80 Å(reported)

This is a wwPDB validation summary report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

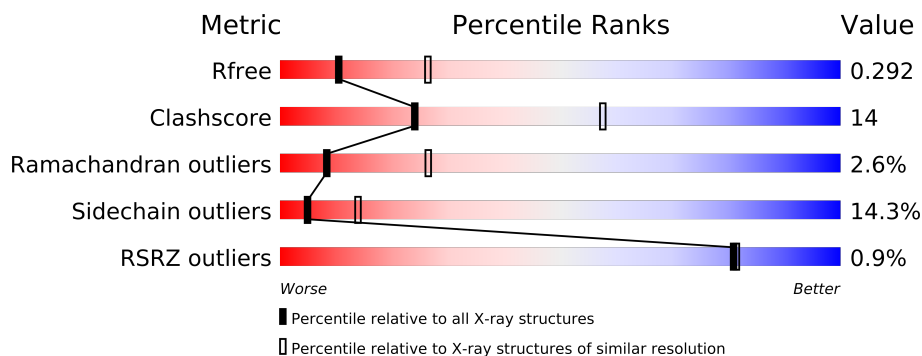
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.15 2013
Xtriage (Phenix) : dev-1323
EDS : stable22639
Percentile statistics : 21963
Refmac : 5.8.0049
CCP4 : 6.3.0 (Settle)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP) : stable22683

1 Overall quality at a glance

The reported resolution of this entry is 2.80 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}	66092	1799 (2.80-2.80)
Clashscore	79885	2295 (2.80-2.80)
Ramachandran outliers	78287	2252 (2.80-2.80)
Sidechain outliers	78261	2254 (2.80-2.80)
RSRZ outliers	66119	1802 (2.80-2.80)

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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	136	
1	B	136	
1	C	136	
1	D	136	
1	E	136	
1	F	136	

2 Entry composition

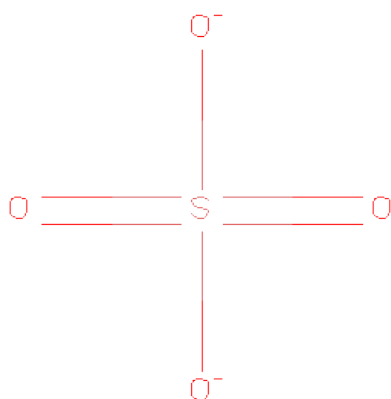
There are 2 unique types of molecules in this entry. The entry contains 6497 atoms, of which 0 are hydrogen and 0 are deuterium.

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 136aa long hypothetical transposase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	130	Total	C	N	O	S	0	0	0
			1080	702	183	190	5			
1	B	131	Total	C	N	O	S	0	0	0
			1089	708	185	191	5			
1	C	130	Total	C	N	O	S	0	0	0
			1080	702	183	190	5			
1	D	128	Total	C	N	O	S	0	0	0
			1061	688	180	188	5			
1	E	129	Total	C	N	O	S	0	0	0
			1072	697	182	189	4			
1	F	129	Total	C	N	O	S	0	0	0
			1075	699	182	189	5			

- Molecule 2 is SULFATE ION (three-letter code: SO₄) (formula: O₄S).

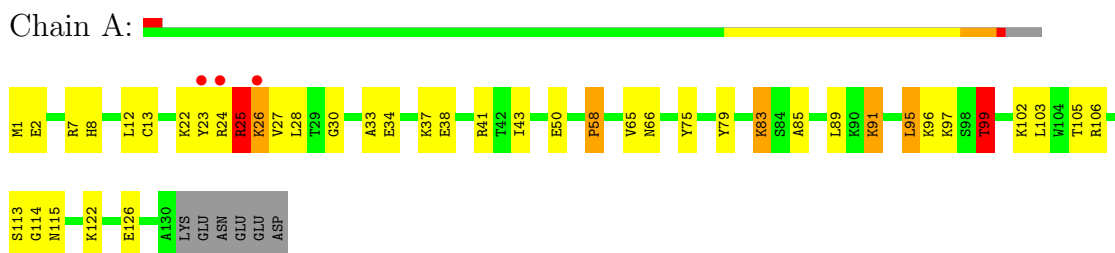


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	F	1	Total	O	S	0	0
			5	4	1		
2	E	1	Total	O	S	0	0
			5	4	1		
2	A	1	Total	O	S	0	0
			5	4	1		
2	D	1	Total	O	S	0	0
			5	4	1		
2	B	1	Total	O	S	0	0
			5	4	1		
2	A	1	Total	O	S	0	0
			5	4	1		
2	A	1	Total	O	S	0	0
			5	4	1		
2	B	1	Total	O	S	0	0
			5	4	1		

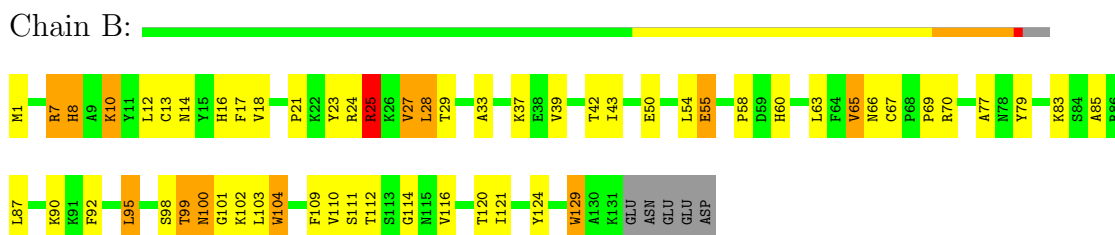
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

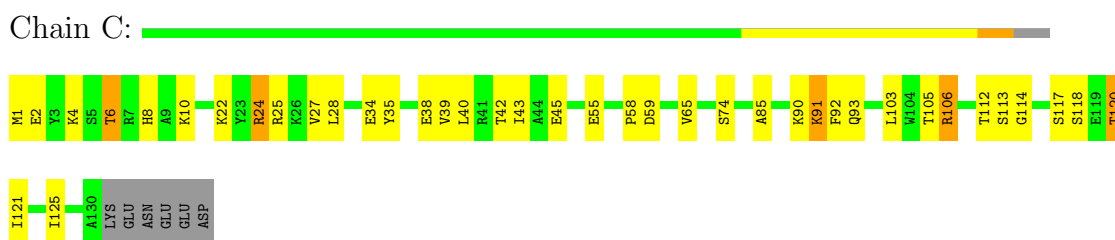
- Molecule 1: 136aa long hypothetical transposase



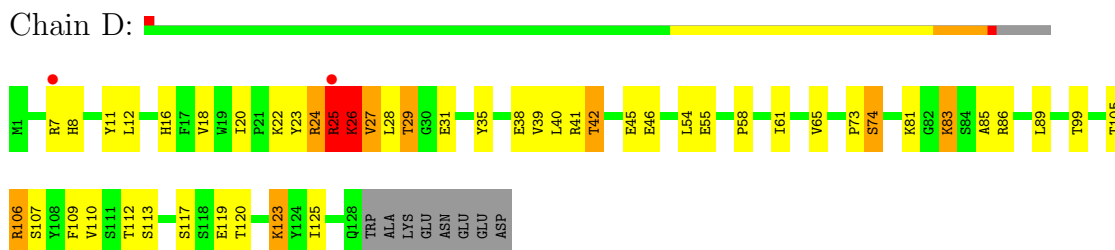
- Molecule 1: 136aa long hypothetical transposase



- Molecule 1: 136aa long hypothetical transposase



- Molecule 1: 136aa long hypothetical transposase



- Molecule 1: 136aa long hypothetical transposase

Figure 1: Distribution of 120 amino acids across 10 categories. The chart shows the relative frequency of each amino acid, with the total summing to 120. The categories are represented by colored blocks: MET (grey), E2 (green), H8 (yellow), L12 (light blue), C13 (orange), V18 (light yellow), K22 (dark orange), Y23 (light green), R24 (yellow), R25 (orange), K26 (green), V27 (light blue), L28 (light yellow), E34 (light green), Y35 (light yellow), T36 (light blue), K37 (yellow), F38 (orange), V39 (light blue), T42 (orange), E45 (light green), E50 (light yellow), P68 (light blue), V65 (light green), N66 (light yellow), S84 (light blue), A85 (orange), I88 (light yellow), K91 (light blue), E94 (light green), L95 (light yellow), K96 (green), S98 (light blue), T99 (orange), L103 (light blue), W104 (green), T105 (orange), R106 (light yellow), F109 (light blue), V110 (light green), S111 (light yellow), and W115 (light blue). The bars are connected by lines, and some bars have red dots above them.

- Chain F:

Q93	E94	L95	K96	K97	S98	T99	K100	L101	K102	L103	T104	T105	R106	S107	F108	F109	V110	S111	T112	S118	E119	T120	L121	K122	K123	V124	L125	F126	E127	T128	W129	ALA	LYS	GLU	ASN	GLU	GLU	ASP				
M1	E2	Y3	K4	R7	K10	Y11	L12	H16	F17	V18	W19	I20	P21	R23	K25	K26	V27	L28	T29	G30	E31	V32	A33	T36	K37	L40	I43	E46	L47	A53	L54	P58	D59	F64	P73	S74	Y79	G82	A85	R86	L89	T92

4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	55.12Å 66.57Å 69.47Å 107.96° 104.39° 110.72°	Depositor
Resolution (Å)	25.78 – 2.80 25.77 – 2.80	Depositor EDS
% Data completeness (in resolution range)	93.1 (25.78-2.80) 93.0 (25.77-2.80)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	6.36 (at 2.80Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.201 , 0.299 0.201 , 0.292	Depositor DCC
R_{free} test set	945 reflections (5.39%)	DCC
Wilson B-factor (Å ²)	45.5	Xtriage
Anisotropy	0.190	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 14.4	EDS
Estimated twinning fraction	0.043 for -h,-k,h+k+l	Xtriage
L-test for twinning	$\langle L \rangle = 0.45$, $\langle L^2 \rangle = 0.28$	Xtriage
Outliers	0 of 18474 reflections	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	6497	wwPDB-VP
Average B, all atoms (Å ²)	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.63% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

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	0.60	0/1109	0.68	0/1497
1	B	0.57	0/1118	0.68	0/1508
1	C	0.58	0/1109	0.68	0/1497
1	D	0.56	0/1088	0.72	0/1467
1	E	0.61	1/1101 (0.1%)	0.64	0/1487
1	F	0.63	0/1104	0.73	0/1490
All	All	0.59	1/6629 (0.0%)	0.69	0/8946

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	D	0	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	13	CYS	CB-SG	-5.18	1.73	1.81

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	25	ARG	Peptide

5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1080	0	1091	28	0
1	B	1089	0	1104	44	0
1	C	1080	0	1091	24	0
1	D	1061	0	1076	38	0
1	E	1072	0	1079	32	0
1	F	1075	0	1086	47	0
2	A	15	0	0	0	0
2	B	10	0	0	0	0
2	D	5	0	0	1	0
2	E	5	0	0	0	0
2	F	5	0	0	2	0
All	All	6497	0	6527	185	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 14.

The worst 5 of 185 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:E:106:ARG:HG3	1:E:106:ARG:HH11	0.91	1.05
1:D:23:TYR:HB3	1:D:25:ARG:HH22	1.18	1.04
1:C:25:ARG:HE	1:C:59:ASP:HB3	1.36	0.91
1:E:106:ARG:NH1	1:E:106:ARG:HG3	1.72	0.89
1:B:23:TYR:HB3	1:B:25:ARG:HD3	1.56	0.87

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	128/136 (94%)	112 (88%)	10 (8%)	6 (5%)	4	11
1	B	129/136 (95%)	117 (91%)	8 (6%)	4 (3%)	7	21
1	C	128/136 (94%)	111 (87%)	15 (12%)	2 (2%)	14	44
1	D	126/136 (93%)	113 (90%)	10 (8%)	3 (2%)	9	29
1	E	127/136 (93%)	115 (91%)	8 (6%)	4 (3%)	7	21
1	F	127/136 (93%)	116 (91%)	10 (8%)	1 (1%)	27	65
All	All	765/816 (94%)	684 (89%)	61 (8%)	20 (3%)	8	26

5 of 20 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	25	ARG
1	B	100	ASN
1	C	24	ARG
1	D	25	ARG
1	D	26	LYS

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	117/123 (95%)	101 (86%)	16 (14%)	5	15
1	B	118/123 (96%)	100 (85%)	18 (15%)	4	12
1	C	117/123 (95%)	104 (89%)	13 (11%)	9	25
1	D	116/123 (94%)	96 (83%)	20 (17%)	3	8
1	E	116/123 (94%)	102 (88%)	14 (12%)	7	21
1	F	117/123 (95%)	98 (84%)	19 (16%)	3	10
All	All	701/738 (95%)	601 (86%)	100 (14%)	5	14

5 of 100 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	106	ARG
1	D	42	THR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	F	40	LEU
1	C	120	THR
1	D	24	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	66	ASN
1	C	93	GLN
1	E	60	HIS
1	C	62	HIS
1	D	93	GLN

5.3.3 RNA ⓘ

There are no RNA chains in this entry.

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

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

5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry ⓘ

8 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	137	-	4,4,4	0.12	0	6,6,6	0.31	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z > 2	Counts	RMSZ	# Z > 2
2	SO4	A	138	-	4,4,4	0.13	0	6,6,6	0.15	0
2	SO4	A	139	-	4,4,4	0.17	0	6,6,6	0.09	0
2	SO4	B	137	-	4,4,4	0.19	0	6,6,6	0.10	0
2	SO4	B	138	-	4,4,4	0.17	0	6,6,6	0.14	0
2	SO4	D	137	-	4,4,4	0.16	0	6,6,6	0.15	0
2	SO4	E	137	-	4,4,4	0.15	0	6,6,6	0.07	0
2	SO4	F	137	-	4,4,4	0.15	0	6,6,6	0.35	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	SO4	A	137	-	-	0/0/0/0	0/0/0/0
2	SO4	A	138	-	-	0/0/0/0	0/0/0/0
2	SO4	A	139	-	-	0/0/0/0	0/0/0/0
2	SO4	B	137	-	-	0/0/0/0	0/0/0/0
2	SO4	B	138	-	-	0/0/0/0	0/0/0/0
2	SO4	D	137	-	-	0/0/0/0	0/0/0/0
2	SO4	E	137	-	-	0/0/0/0	0/0/0/0
2	SO4	F	137	-	-	0/0/0/0	0/0/0/0

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.

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, 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	130/136 (95%)	-0.51	3 (2%) 57 58	27, 46, 66, 76	0
1	B	131/136 (96%)	-0.49	0 100 100	24, 48, 66, 73	0
1	C	130/136 (95%)	-0.52	0 100 100	25, 41, 63, 68	0
1	D	128/136 (94%)	-0.39	2 (1%) 68 69	22, 45, 66, 72	0
1	E	129/136 (94%)	-0.54	2 (1%) 68 69	21, 39, 59, 76	0
1	F	129/136 (94%)	-0.60	0 100 100	19, 40, 60, 71	0
All	All	777/816 (95%)	-0.51	7 (0%) 81 81	19, 43, 65, 76	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	26	LYS	4.1
1	E	25	ARG	3.4
1	A	23	TYR	3.0
1	A	24	ARG	2.7
1	E	24	ARG	2.7

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 carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	RSR	LLDF	B-factors(Å ²)	Q<0.9
2	SO4	B	137	5/5	0.15	1.68	101,101,101,101	0
2	SO4	A	137	5/5	0.15	0.44	72,73,73,73	0
2	SO4	A	138	5/5	0.13	-0.05	107,108,108,108	0
2	SO4	B	138	5/5	0.12	-0.87	108,109,109,109	0
2	SO4	E	137	5/5	0.10	-1.71	70,70,71,71	0
2	SO4	A	139	5/5	0.09	-3.15	110,110,111,111	0
2	SO4	F	137	5/5	0.08	-4.44	48,49,50,51	0
2	SO4	D	137	5/5	0.06	-8.20	76,76,77,77	0

6.5 Other polymers

There are no such residues in this entry.