



# Full wwPDB X-ray Structure Validation Report

Feb 28, 2014 – 05:20 AM GMT

PDB ID : 3ELQ  
Title : Crystal structure of a bacterial arylsulfate sulfotransferase  
Authors : Malojcic, G.; Owen, R.L.; Grimshaw, J.P.; Glockshuber, R.  
Deposited on : 2008-09-23  
Resolution : 2.00 Å(reported)

This is a full wwPDB 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 <http://wwpdb.org/ValidationPDFNotes.html>

---

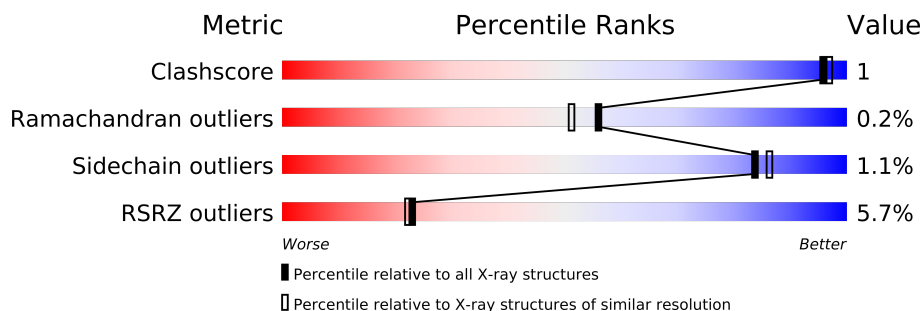
The following versions of software and data (see [references](#)) were used in the production of this report:

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 2.00 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))
Clashscore	79885	6188 (2.00-2.00)
Ramachandran outliers	78287	6102 (2.00-2.00)
Sidechain outliers	78261	6100 (2.00-2.00)
RSRZ outliers	66119	4890 (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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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	571	
1	B	571	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
2	SO4	A	2004	-	X
2	SO4	A	2005	-	X
2	SO4	A	2007	-	X
2	SO4	A	2009	-	X
2	SO4	B	2008	-	X
2	SO4	B	2010	-	X

## 2 Entry composition i

There are 4 unique types of molecules in this entry. The entry contains 9997 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 Arylsulfate sulfotransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	564	Total	C	N	O	S	0	3	0
			4457	2833	762	852	10			
1	B	566	Total	C	N	O	S	0	2	0
			4467	2840	764	853	10			

- Molecule 2 is SULFATE ION (three-letter code: SO<sub>4</sub>) (formula: O<sub>4</sub>S).



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

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	5	Total Cl 5 5	0	0
3	A	3	Total Cl 3 3	0	0

- Molecule 4 is water.

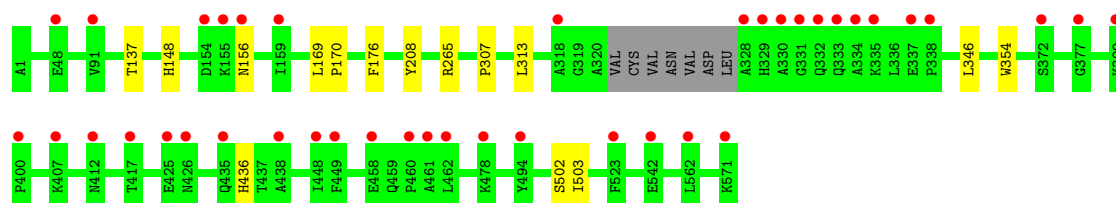
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	509	Total O 509 509	0	0
4	B	501	Total O 501 501	0	0

### 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.

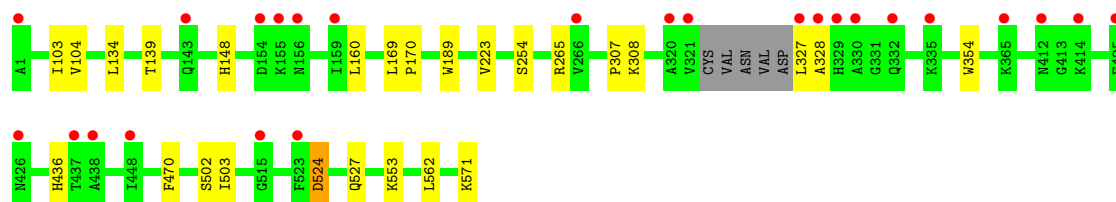
- Molecule 1: Arylsulfate sulfotransferase

Chain A: 



- Molecule 1: Arylsulfate sulfotransferase

Chain B: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 32 1 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	181.73Å 181.73Å 100.43Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	33.90 – 2.00 33.69 – 2.00	Depositor EDS
% Data completeness (in resolution range)	100.0 (33.90-2.00) 97.1 (33.69-2.00)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	0.09	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.84 (at 2.00Å)	Xtriage
Refinement program	REFMAC 5.3	Depositor
R, $R_{free}$	0.207 , 0.248 0.220 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	DCC
Wilson B-factor (Å <sup>2</sup> )	34.7	Xtriage
Anisotropy	0.307	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 44.0	EDS
Estimated twinning fraction	0.024 for -h,-k,l	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtriage
Outliers	1 of 122878 reflections (0.001%)	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	9997	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.95% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities 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, CL

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	0/4576	0.61	0/6210
1	B	0.46	0/4583	0.61	0/6221
All	All	0.47	0/9159	0.61	0/12431

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	B	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	328	ALA	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	4457	0	4305	6	0
1	B	4467	0	4324	14	1
2	A	30	0	0	2	0
2	B	25	0	0	1	0
3	A	3	0	0	0	0
3	B	5	0	0	2	0
4	A	509	0	0	0	0
4	B	501	0	0	3	0
All	All	9997	0	8629	21	1

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 1.

All (21) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
3:B:5010:CL:CL	4:B:5444:HOH:O	2.52	0.64
1:B:308:LYS:NZ	3:B:5010:CL:CL	2.69	0.63
1:B:327:LEU:N	4:B:5510:HOH:O	2.38	0.56
1:A:313:LEU:HD22	1:A:346:LEU:HD22	1.91	0.51
1:B:436:HIS:NE2	2:B:2002:SO4:S	2.68	0.51
1:A:169:LEU:N	1:A:170:PRO:CD	2.74	0.51
1:B:169:LEU:N	1:B:170:PRO:CD	2.75	0.49
1:A:307:PRO:HA	1:A:354:TRP:CE3	2.48	0.49
1:B:307:PRO:HA	1:B:354:TRP:CE3	2.48	0.48
1:A:436:HIS:NE2	2:A:2001:SO4:S	2.72	0.48
1:A:436:HIS:NE2	2:A:2001:SO4:O2	2.47	0.47
1:B:524:ASP:HB3	1:B:527:GLN:HB2	1.98	0.45
1:B:571:LYS:NZ	4:B:5465:HOH:O	2.49	0.44
1:A:137:THR:HG23	1:A:176:PHE:HB3	1.99	0.44
1:B:134:LEU:HD22	1:B:562:LEU:HD13	2.00	0.44
1:B:470:PHE:N	1:B:470:PHE:CD1	2.86	0.43
1:B:223:VAL:HG13	1:B:254:SER:HB3	1.99	0.43
1:B:134:LEU:C	1:B:134:LEU:HD23	2.38	0.43
1:B:103:ILE:HG23	1:B:189:TRP:CE2	2.54	0.41
1:B:103:ILE:HG23	1:B:189:TRP:CD2	2.56	0.41
1:B:139:THR:OG1	1:B:553:LYS:NZ	2.53	0.40

All (1) 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	Distance(Å)	Clash(Å)
1:B:104:VAL:CG2	1:B:104:VAL:CG2[5_555]	2.19	0.01



## 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	563/571 (99%)	545 (97%)	17 (3%)	1 (0%)	56	51
1	B	564/571 (99%)	543 (96%)	20 (4%)	1 (0%)	56	51
All	All	1127/1142 (99%)	1088 (96%)	37 (3%)	2 (0%)	56	51

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	502	SER
1	A	502	SER

### 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	471/483 (98%)	465 (99%)	6 (1%)	80	82
1	B	472/483 (98%)	467 (99%)	5 (1%)	84	86
All	All	943/966 (98%)	932 (99%)	11 (1%)	84	84

All (11) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	148	HIS
1	A	156[A]	ASN
1	A	156[B]	ASN
1	A	208	TYR
1	A	265	ARG
1	A	503	ILE
1	B	148	HIS

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	160	LEU
1	B	265	ARG
1	B	503	ILE
1	B	524	ASP

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

### 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 ⓘ

Of 19 ligands modelled in this entry, 8 are monoatomic - leaving 11 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	SO4	A	2001	-	4,4,4	0.32	0	6,6,6	0.22	0
2	SO4	A	2004	-	4,4,4	0.13	0	6,6,6	0.06	0
2	SO4	A	2005	-	4,4,4	0.12	0	6,6,6	0.30	0
2	SO4	A	2006	-	4,4,4	0.13	0	6,6,6	0.07	0
2	SO4	A	2007	-	4,4,4	0.12	0	6,6,6	0.15	0
2	SO4	A	2009	-	4,4,4	0.12	0	6,6,6	0.17	0
2	SO4	B	2002	-	4,4,4	0.37	0	6,6,6	0.40	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	SO4	B	2003	-	4,4,4	0.12	0	6,6,6	0.15	0
2	SO4	B	2008	-	4,4,4	0.05	0	6,6,6	0.16	0
2	SO4	B	2010	-	4,4,4	0.03	0	6,6,6	0.12	0
2	SO4	B	2011	-	4,4,4	0.10	0	6,6,6	0.23	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	2001	-	-	0/0/0/0	0/0/0/0
2	SO4	A	2004	-	-	0/0/0/0	0/0/0/0
2	SO4	A	2005	-	-	0/0/0/0	0/0/0/0
2	SO4	A	2006	-	-	0/0/0/0	0/0/0/0
2	SO4	A	2007	-	-	0/0/0/0	0/0/0/0
2	SO4	A	2009	-	-	0/0/0/0	0/0/0/0
2	SO4	B	2002	-	-	0/0/0/0	0/0/0/0
2	SO4	B	2003	-	-	0/0/0/0	0/0/0/0
2	SO4	B	2008	-	-	0/0/0/0	0/0/0/0
2	SO4	B	2010	-	-	0/0/0/0	0/0/0/0
2	SO4	B	2011	-	-	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, 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	564/571 (98%)	0.26	40 (7%) 16 15	29, 41, 61, 80	0
1	B	566/571 (99%)	0.25	25 (4%) 33 32	28, 41, 61, 81	0
All	All	1130/1142 (98%)	0.26	65 (5%) 23 22	28, 41, 61, 81	0

All (65) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	329	HIS	6.3
1	B	329	HIS	5.2
1	B	425	GLU	4.8
1	A	331	GLY	4.8
1	B	327	LEU	4.8
1	A	155	LYS	4.1
1	B	426	ASN	3.9
1	A	426	ASN	3.9
1	B	155	LYS	3.7
1	A	159	ILE	3.7
1	A	462	LEU	3.7
1	B	328	ALA	3.7
1	B	159	ILE	3.6
1	B	156	ASN	3.5
1	B	154	ASP	3.4
1	A	338	PRO	3.2
1	A	334	ALA	3.2
1	A	318	ALA	3.1
1	A	425	GLU	3.0
1	B	1	ALA	3.0
1	A	156[A]	ASN	2.8
1	B	143	GLN	2.8
1	B	321	VAL	2.8
1	A	330	ALA	2.7

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	460	PRO	2.7
1	B	448	ILE	2.6
1	A	448	ILE	2.6
1	A	332	GLN	2.6
1	A	461	ALA	2.6
1	B	437	THR	2.6
1	A	438	ALA	2.5
1	A	542	GLU	2.5
1	A	412	ASN	2.5
1	B	332	GLN	2.5
1	A	400	PRO	2.5
1	B	320	ALA	2.5
1	A	523	PHE	2.5
1	A	377	GLY	2.5
1	A	48	GLU	2.4
1	B	266	VAL	2.4
1	B	414	LYS	2.4
1	A	417	THR	2.4
1	A	399	LYS	2.4
1	A	407	LYS	2.3
1	A	449	PHE	2.3
1	A	91	VAL	2.2
1	B	438	ALA	2.2
1	A	478	LYS	2.2
1	A	372	SER	2.2
1	B	412	ASN	2.2
1	A	494	TYR	2.2
1	A	562	LEU	2.2
1	A	337	GLU	2.2
1	A	335	LYS	2.1
1	B	335	LYS	2.1
1	A	328	ALA	2.1
1	A	435	GLN	2.1
1	A	458	GLU	2.1
1	B	515	GLY	2.1
1	B	523	PHE	2.1
1	B	330	ALA	2.1
1	A	571	LYS	2.0
1	A	154	ASP	2.0
1	B	365	LYS	2.0
1	A	333	GLN	2.0

## 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, 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	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
2	SO4	A	2005	5/5	0.22	14.10	83,83,83,84	0
2	SO4	A	2007	5/5	0.26	7.29	87,88,88,89	0
2	SO4	B	2008	5/5	0.16	6.11	80,80,80,80	0
2	SO4	B	2010	5/5	0.26	3.42	87,88,88,88	0
2	SO4	A	2009	5/5	0.26	2.68	84,85,86,87	0
2	SO4	A	2004	5/5	0.24	2.58	106,106,106,107	0
3	CL	A	5004	1/1	0.16	0.87	66,66,66,66	0
3	CL	B	5009	1/1	0.15	0.84	81,81,81,81	0
2	SO4	B	2011	5/5	0.24	0.82	93,93,94,94	0
2	SO4	A	2006	5/5	0.17	0.67	63,63,63,64	0
2	SO4	B	2003	5/5	0.11	-0.43	66,67,67,67	0
2	SO4	B	2002	5/5	0.13	-0.83	42,43,44,46	0
2	SO4	A	2001	5/5	0.13	-1.14	43,43,44,45	0
3	CL	B	5010	1/1	0.07	-1.54	50,50,50,50	0
3	CL	A	5008	1/1	0.07	-2.50	58,58,58,58	0
3	CL	B	5007	1/1	0.10	-2.54	69,69,69,69	0
3	CL	B	5001	1/1	0.04	-3.18	45,45,45,45	0
3	CL	A	5002	1/1	0.04	-3.26	43,43,43,43	0
3	CL	B	5006	1/1	0.07	-8.03	40,40,40,40	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.