



Full wwPDB X-ray Structure Validation Report ⓘ

Mar 1, 2026 – 04:59 AM UTC

PDB ID : 4C8L / pdb_00004c8l
Title : Binary complex of the large fragment of DNA polymerase I from *Thermus Aquaticus* with the artificial base pair dNaM-d5SICS at the postinsertion site (sequence context 1)
Authors : Betz, K.; Malyshev, D.A.; Lavergne, T.; Welte, W.; Diederichs, K.; Romesberg, F.E.; Marx, A.
Deposited on : 2013-10-01
Resolution : 1.70 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

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	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
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)	:	2.49

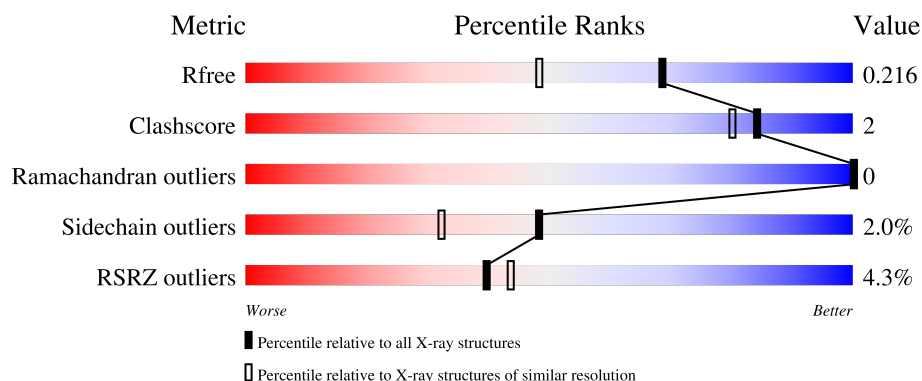
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.70 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}	180053	5551 (1.70-1.70)
Clashscore	190562	5924 (1.70-1.70)
Ramachandran outliers	187476	5846 (1.70-1.70)
Sidechain outliers	187428	5846 (1.70-1.70)
RSRZ outliers	180081	5554 (1.70-1.70)

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	540	4% 92% 8%
2	B	11	91% 9%
3	C	13	92% 8%

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 9840 atoms, of which 4694 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 DNA POLYMERASE I, THERMOSTABLE.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	540	Total	C	H	N	O	S	0	5	0
			8697	2743	4375	780	786	13			

- Molecule 2 is a DNA chain called 5'-D(*AP*CP*CP*AP*CP*GP*GP*CP*GP*CP*LHOP)-3'.

Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
2	B	11	Total	C	H	N	O	P	S	0	0	0
			350	110	127	41	61	10	1			

- Molecule 3 is a DNA chain called 5'-D(*AP*GP*BMNP*GP*CP*GP*CP*CP*GP*TP*GP*GP*TP)-3'.

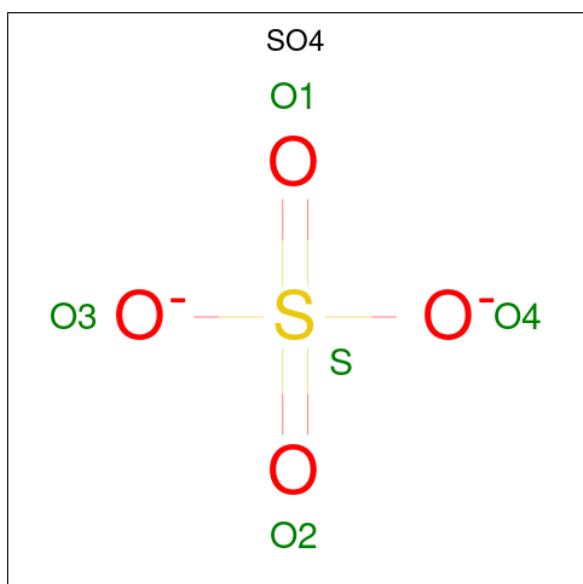
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	C	13	Total	C	H	N	O	P			
			422	133	152	48	77	12	0	0	0

- Molecule 4 is GLYCEROL (CCD ID: GOL) (formula: C₃H₈O₃).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total	C	H	O	0	0
			14	3	8	3		
4	A	1	Total	C	H	O	0	0
			14	3	8	3		
4	A	1	Total	C	H	O	0	0
			14	3	8	3		
4	A	1	Total	C	H	O	0	0
			14	3	8	3		
4	A	1	Total	C	H	O	0	0
			14	3	8	3		

- Molecule 5 is SULFATE ION (CCD ID: SO4) (formula: O₄S).



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

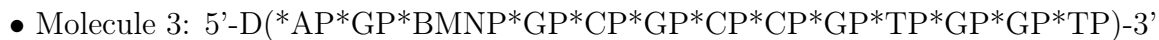
- Molecule 6 is MAGNESIUM ION (CCD ID: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	1	Total Mg 1 1	0	0
6	B	1	Total Mg 1 1	0	0

- Molecule 7 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	A	234	Total O 234 234	0	0
7	B	27	Total O 27 27	0	0
7	C	18	Total O 18 18	0	0

- Molecule 1: DNA POLYMERASE I, THERMOSTABLE



4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α , β , γ	65.61Å 100.54Å 204.34Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.39 – 1.70 48.39 – 1.70	Depositor EDS
% Data completeness (in resolution range)	99.8 (48.39-1.70) 99.7 (48.39-1.70)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	0.96 (at 1.70Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.189 , 0.214 0.192 , 0.216	Depositor DCC
R_{free} test set	3711 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å ²)	32.8	Xtriage
Anisotropy	0.460	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.39 , 35.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	9840	wwPDB-VP
Average B, all atoms (Å ²)	51.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.23% 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 [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, SO4, MG, LHO, BMN

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	0.68	0/4426	0.77	2/5993 (0.0%)
2	B	0.45	0/224	0.61	0/343
3	C	0.45	0/276	0.66	0/423
All	All	0.66	0/4926	0.75	2/6759 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	549	LEU	CA-C-N	-6.05	112.77	119.19
1	A	549	LEU	C-N-CA	-6.05	112.77	119.19

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	4322	4375	4358	22	0
2	B	223	127	127	0	0
3	C	270	152	152	0	0
4	A	30	40	40	0	0
5	A	15	0	0	0	0
5	C	5	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	A	1	0	0	0	0
6	B	1	0	0	0	0
7	A	234	0	0	1	0
7	B	27	0	0	0	0
7	C	18	0	0	0	0
All	All	5146	4694	4677	22	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 2.

All (22) 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:303:GLU:OE2	1:A:346:LYS:NZ	2.28	0.67
1:A:632:PHE:CE1	1:A:638:ILE:HG21	2.44	0.52
1:A:646:MET:O	1:A:695:ARG:NE	2.44	0.50
1:A:299:TRP:CG	1:A:300:PRO:HA	2.49	0.47
1:A:647:PHE:CZ	1:A:692:PHE:CE1	3.02	0.47
1:A:535:TYR:CZ	1:A:539:THR:HG21	2.50	0.47
1:A:638:ILE:HG13	1:A:639:HIS:N	2.29	0.46
1:A:674:SER:OG	1:A:677:ARG:HG2	2.16	0.46
1:A:769:PHE:HB3	1:A:770:PRO:HD3	1.99	0.44
1:A:780:LEU:HD11	1:A:790:GLU:HB2	2.00	0.43
1:A:437:LEU:HD21	1:A:574:LEU:HD21	2.00	0.43
1:A:595[B]:ARG:NE	7:A:2172:HOH:O	2.52	0.43
1:A:696:TYR:CD1	1:A:696:TYR:C	2.96	0.43
1:A:778:ARG:HD2	1:A:790:GLU:OE1	2.19	0.43
1:A:359:LEU:C	1:A:359:LEU:HD23	2.44	0.42
1:A:721:GLU:HA	1:A:726:ARG:O	2.19	0.42
1:A:811:TYR:CG	1:A:812:PRO:HD2	2.55	0.42
1:A:297:ALA:HB3	1:A:333:HIS:CE1	2.55	0.41
1:A:595[A]:ARG:NH2	1:A:831:LYS:HB3	2.35	0.41
1:A:526:HIS:ND1	1:A:528:ILE:HG22	2.36	0.41
1:A:611:TYR:HB3	1:A:614:ILE:HB	2.02	0.41
1:A:665:ILE:HG12	1:A:692:PHE:CD2	2.56	0.41

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	543/540 (101%)	528 (97%)	15 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	446/441 (101%)	437 (98%)	9 (2%)	48	32

All (9) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	294	LEU
1	A	487	ARG
1	A	556	ARG
1	A	579	PRO
1	A	587	ARG
1	A	734	GLU
1	A	778	ARG
1	A	802	LEU
1	A	831	LYS

Sometimes 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 molecules in this entry.

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

2 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	$\# Z > 2$
2	LHO	B	112	2	21,25,26	0.34	0	29,36,39	0.93	2 (6%)
3	BMN	C	205	3	22,25,26	0.24	0	30,35,38	0.84	1 (3%)

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	LHO	B	112	2	-	1/7/21/22	0/3/3/3
3	BMN	C	205	3	-	0/9/23/24	0/3/3/3

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	205	BMN	C2-C1-C6	3.44	121.91	117.09
2	B	112	LHO	C9-C4-C5	2.95	123.54	121.17
2	B	112	LHO	C3-C4-C5	-2.63	117.87	120.06

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	112	LHO	C4'-C5'-O5'-P

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 11 ligands modelled in this entry, 2 are monoatomic - leaving 9 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	$\# Z > 2$
5	SO4	A	1836	-	4,4,4	0.26	0	6,6,6	0.17	0
5	SO4	C	1216	-	4,4,4	0.32	0	6,6,6	0.12	0
5	SO4	A	1834	-	4,4,4	0.27	0	6,6,6	0.17	0
4	GOL	A	1838	-	5,5,5	0.39	0	5,5,5	0.38	0
5	SO4	A	1835	-	4,4,4	0.22	0	6,6,6	0.46	0
4	GOL	A	1837	-	5,5,5	0.44	0	5,5,5	0.11	0
4	GOL	A	1839	-	5,5,5	0.42	0	5,5,5	0.48	0
4	GOL	A	1833	-	5,5,5	0.43	0	5,5,5	1.19	0
4	GOL	A	1840	-	5,5,5	0.39	0	5,5,5	0.27	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
4	GOL	A	1838	-	-	0/4/4/4	-
4	GOL	A	1837	-	-	4/4/4/4	-
4	GOL	A	1839	-	-	3/4/4/4	-
4	GOL	A	1833	-	-	4/4/4/4	-
4	GOL	A	1840	-	-	0/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (11) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	A	1833	GOL	O1-C1-C2-C3
4	A	1833	GOL	C1-C2-C3-O3
4	A	1837	GOL	O1-C1-C2-C3
4	A	1839	GOL	O1-C1-C2-C3
4	A	1837	GOL	C1-C2-C3-O3
4	A	1839	GOL	C1-C2-C3-O3
4	A	1837	GOL	O2-C2-C3-O3
4	A	1839	GOL	O1-C1-C2-O2
4	A	1833	GOL	O1-C1-C2-O2
4	A	1833	GOL	O2-C2-C3-O3
4	A	1837	GOL	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, 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	540/540 (100%)	0.21	24 (4%) 39 42	14, 46, 84, 127	4 (0%)
2	B	10/11 (90%)	-0.20	0 100 100	33, 48, 64, 66	0
3	C	12/13 (92%)	-0.14	0 100 100	35, 41, 81, 103	0
All	All	562/564 (99%)	0.19	24 (4%) 40 44	14, 46, 83, 127	4 (0%)

All (24) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	294	LEU	4.7
1	A	647	PHE	4.4
1	A	692	PHE	4.3
1	A	293	ALA	4.1
1	A	645	TRP	4.1
1	A	638	ILE	3.7
1	A	525	ALA	3.6
1	A	503	ILE	3.3
1	A	737	VAL	3.2
1	A	683	ALA	2.9
1	A	665	ILE	2.9
1	A	832	GLU	2.8
1	A	586	VAL	2.6
1	A	588	THR	2.6
1	A	648	GLY	2.5
1	A	669	VAL	2.5
1	A	519	LEU	2.4
1	A	676	HIS	2.4
1	A	524	GLU	2.3
1	A	502	ALA	2.3
1	A	587	ARG	2.2
1	A	678	LEU	2.2
1	A	682	LEU	2.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	684	ILE	2.1

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
3	BMN	C	205	23/24	0.96	0.07	36,43,54,56	0
2	LHO	B	112	23/24	0.98	0.06	30,38,52,52	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
6	MG	B	1113	1/1	0.74	0.12	67,67,67,67	1
4	GOL	A	1833	6/6	0.76	0.20	40,48,51,54	14
5	SO4	C	1216	5/5	0.77	0.18	62,64,67,69	5
4	GOL	A	1840	6/6	0.81	0.16	72,87,92,93	0
4	GOL	A	1839	6/6	0.84	0.12	55,66,73,75	0
4	GOL	A	1838	6/6	0.86	0.12	58,71,80,85	0
5	SO4	A	1835	5/5	0.86	0.12	65,65,73,77	5
4	GOL	A	1837	6/6	0.87	0.19	72,92,110,110	0
6	MG	A	1841	1/1	0.87	0.17	72,72,72,72	0
5	SO4	A	1836	5/5	0.87	0.08	89,95,98,102	0
5	SO4	A	1834	5/5	0.88	0.09	62,64,69,71	5

6.5 Other polymers [i](#)

There are no such residues in this entry.