



wwPDB X-ray Structure Validation Summary Report

Feb 28, 2014 – 09:35 AM GMT

PDB ID : 4DFP
Title : Crystal structure of the large fragment of DNA Polymerase I from *Thermus aquaticus* in a ternary complex with 7-(aminopentynyl)-7-deaza-dGTP
Authors : Bergen, K.; Steck, A.; Struett, S.; Baccaro, A.; Welte, W.; Diederichs, K.; Marx, A.
Deposited on : 2012-01-24
Resolution : 2.00 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>

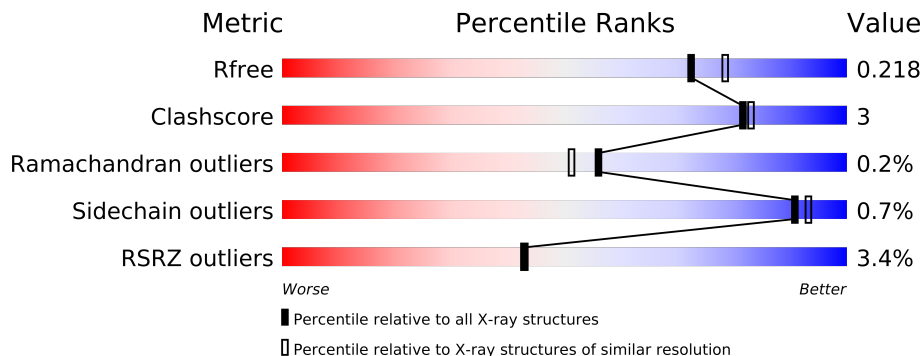
The following versions of software and data (see [references](#)) were used in the production of this report:

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 2.00 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))
R_{free}	66092	4888 (2.00-2.00)
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 ≥ 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	540	
2	B	12	
3	C	16	

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

Mol	Type	Chain	Res	Geometry	Electron density
5	EDO	A	905	-	X
5	EDO	A	909	-	X
5	EDO	C	302	-	X
6	FMT	A	912	-	X
6	FMT	C	301	-	X
6	FMT	C	303	-	X

2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 5219 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 DNA polymerase I, thermostable.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	539	Total	C	N	O	S	0	3	0
			4291	2725	778	775	13			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	12	Total	C	N	O	P	0	0	0
			240	114	48	67	11			

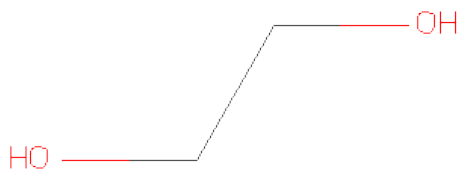
- Molecule 3 is a DNA chain called 5'-d(AAACGGCGCCGTGGTC)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	16	Total	C	N	O	P	0	0	0
			327	155	64	93	15			

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

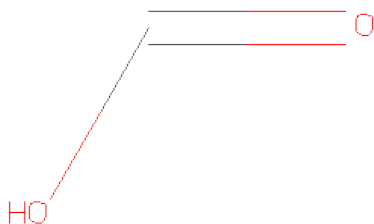
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Mg	0	0
			1	1		
4	A	4	Total	Mg	0	0
			4	4		

- Molecule 5 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C₂H₆O₂).



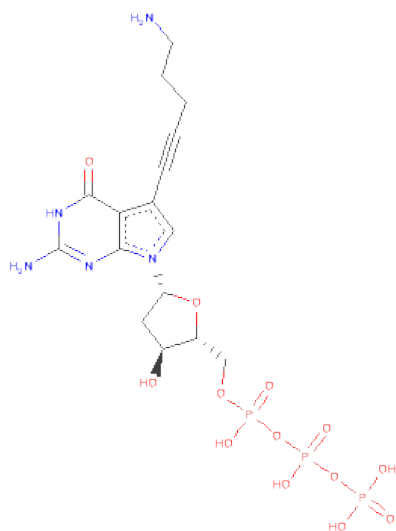
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	C	O	0	0
			4	2	2		
5	A	1	Total	C	O	0	0
			4	2	2		
5	A	1	Total	C	O	0	0
			4	2	2		
5	A	1	Total	C	O	0	0
			4	2	2		
5	C	1	Total	C	O	0	0
			4	2	2		

- Molecule 6 is FORMIC ACID (three-letter code: FMT) (formula: CH_2O_2).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	A	1	Total	C	O	0	0
			3	1	2		
6	A	1	Total	C	O	0	0
			3	1	2		
6	A	1	Total	C	O	0	0
			3	1	2		
6	A	1	Total	C	O	0	0
			3	1	2		
6	B	1	Total	C	O	0	0
			3	1	2		
6	C	1	Total	C	O	0	0
			3	1	2		
6	C	1	Total	C	O	0	0
			3	1	2		

- Molecule 7 is 2-AMINO-5-(5-AMINOPENT-1-YN-1-YL)-7-{2-DEOXY-5-O-[(S)-HYDROXY{[(S)-HYDROXY(PHOSPHONOOXY)PHOSPHORYL]OXY}PHOSPHORYL]-BETA-D-ERYTHRO-PENTOFURANOSYL}-3,7-DIHYDRO-4H-PYRROLO[2,3-D]PYRIMIDIN-4-ONE (three-letter code: 0L7) (formula: C₁₆H₂₄N₅O₁₃P₃).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
7	A	1	Total	C	N	O	P	0	0
			37	16	5	13	3		

- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	217	Total 217	O 217	0	0
8	B	27	Total 27	O 27	0	0
8	C	34	Total 34	O 34	0	0

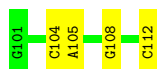
i

- Molecule 1: DNA polymerase I, thermostable

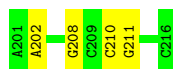
Government	Percentage
Current government	85%
Previous government	15%



- _____



- _____



4 Data and refinement statistics

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α , β , γ	108.45Å 108.45Å 90.44Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	46.96 – 2.00 46.96 – 2.00	Depositor EDS
% Data completeness (in resolution range)	99.8 (46.96-2.00) 99.8 (46.96-2.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.71 (at 2.00Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7.3_928)	Depositor
R, R_{free}	0.185 , 0.224 0.180 , 0.218	Depositor DCC
R_{free} test set	2076 reflections (4.97%)	DCC
Wilson B-factor (Å ²)	32.8	Xtriage
Anisotropy	0.134	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 40.6	EDS
Estimated twinning fraction	0.038 for -h,-k,l	Xtriage
L-test for twinning	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 41790 reflections	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	5219	wwPDB-VP
Average B, all atoms (Å ²)	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.89% 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: 0L7, DOC, MG, FMT, EDO

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	0.20	0/4383	0.39	0/5938
2	B	0.43	0/249	0.97	2/382 (0.5%)
3	C	0.42	0/367	0.94	0/565
All	All	0.24	0/4999	0.50	2/6885 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	108	DG	C1'-O4'-C4'	-5.35	104.75	110.10
2	B	108	DG	O4'-C1'-N9	5.07	111.55	108.00

There are no chirality outliers.

There are no planarity outliers.

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	4291	0	0	13	0
2	B	240	0	134	2	0
3	C	327	0	180	3	0
4	A	4	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	B	1	0	0	0	0
5	A	16	0	24	0	0
5	C	4	0	6	0	0
6	A	12	0	4	0	0
6	B	3	0	1	0	0
6	C	6	0	2	0	0
7	A	37	0	20	0	0
8	A	217	0	0	4	0
8	B	27	0	0	0	0
8	C	34	0	0	0	0
All	All	5219	0	371	15	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 3.

The worst 5 of 15 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:423:GLU:OE1	1:A:727:ARG:NH2	2.35	0.60
1:A:411:ARG:NH2	8:A:1021:HOH:O	2.37	0.56
1:A:473:GLU:OE2	1:A:476:ARG:NE	2.39	0.56
1:A:511:LYS:NZ	8:A:1100:HOH:O	2.39	0.55
1:A:804:LYS:NZ	1:A:808:GLU:OE2	2.41	0.54

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	539/540 (100%)	525 (97%)	13 (2%)	1 (0%)	56 51

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	586	VAL

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	439/441 (100%)	436 (99%)	3 (1%)	91 93

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	487	ARG
1	A	678	LEU
1	A	702	LYS

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 ⓘ

1 non-standard protein/DNA/RNA residue is 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	DOC	B	112	3,2	17,19,20	2.49	3 (17%)	20,26,29	2.10	4 (20%)

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	DOC	B	112	3,2	-	0/5/18/19	0/2/2/2

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	112	DOC	P-OP1	9.01	1.57	1.46
2	B	112	DOC	C4-N4	2.75	1.43	1.35
2	B	112	DOC	C3'-C2'	-2.09	1.47	1.54

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	112	DOC	C4'-O4'-C1'	-5.17	107.89	110.05
2	B	112	DOC	C2-N3-C4	4.34	121.85	115.57
2	B	112	DOC	C6-C5-C4	3.52	118.93	117.47
2	B	112	DOC	C3'-C2'-C1'	2.24	105.28	102.80

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

5.5 Carbohydrates

There are no carbohydrates in this entry.

5.6 Ligand geometry

Of 18 ligands modelled in this entry, 5 are monoatomic - leaving 13 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	EDO	A	903	-	3,3,3	0.53	0	2,2,2	0.37	0
6	FMT	A	904	-	2,2,2	0.64	0	1,1,1	0.23	0
5	EDO	A	905	-	3,3,3	0.54	0	2,2,2	0.39	0
6	FMT	A	906	-	2,2,2	0.64	0	1,1,1	0.23	0
6	FMT	A	907	-	2,2,2	0.63	0	1,1,1	0.23	0
7	OL7	A	908	4	39,39,39	1.14	3 (7%)	56,59,59	1.71	12 (21%)
5	EDO	A	909	-	3,3,3	0.53	0	2,2,2	0.42	0
5	EDO	A	911	-	3,3,3	0.54	0	2,2,2	0.39	0
6	FMT	A	912	-	2,2,2	0.64	0	1,1,1	0.24	0
6	FMT	B	201	-	2,2,2	0.64	0	1,1,1	0.24	0
6	FMT	C	301	-	2,2,2	0.63	0	1,1,1	0.24	0
5	EDO	C	302	-	3,3,3	0.54	0	2,2,2	0.38	0
6	FMT	C	303	-	2,2,2	0.64	0	1,1,1	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
5	EDO	A	903	-	-	0/1/1/1	0/0/0/0
6	FMT	A	904	-	-	0/0/0/0	0/0/0/0
5	EDO	A	905	-	-	0/1/1/1	0/0/0/0
6	FMT	A	906	-	-	0/0/0/0	0/0/0/0
6	FMT	A	907	-	-	0/0/0/0	0/0/0/0
7	OL7	A	908	4	-	0/24/40/40	0/1/3/3
5	EDO	A	909	-	-	0/1/1/1	0/0/0/0
5	EDO	A	911	-	-	0/1/1/1	0/0/0/0
6	FMT	A	912	-	-	0/0/0/0	0/0/0/0
6	FMT	B	201	-	-	0/0/0/0	0/0/0/0
6	FMT	C	301	-	-	0/0/0/0	0/0/0/0
5	EDO	C	302	-	-	0/1/1/1	0/0/0/0
6	FMT	C	303	-	-	0/0/0/0	0/0/0/0

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	A	908	OL7	C7-C32	-4.82	1.34	1.44
7	A	908	OL7	C2-N2	2.15	1.35	1.32
7	A	908	OL7	C8-N9	-2.02	1.34	1.38

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
7	A	908	OL7	C8-C7-C32	-5.30	121.12	126.35
7	A	908	OL7	C2-N3-C4	3.74	120.35	115.09
7	A	908	OL7	C7-C8-N9	-3.42	104.53	107.95
7	A	908	OL7	C2'-C1'-N9	-3.35	105.38	114.08
7	A	908	OL7	PA-O3A-PB	-3.32	121.94	131.68

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	539/540 (99%)	-0.13	20 (3%)	39 39	20, 34, 64, 81	0
2	B	12/12 (100%)	-0.58	0	100 100	25, 28, 55, 56	0
3	C	16/16 (100%)	-0.32	0	100 100	24, 32, 69, 85	0
All	All	567/568 (99%)	-0.14	20 (3%)	43 41	20, 33, 64, 85	0

The worst 5 of 20 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	682	LEU	5.2
1	A	692	PHE	5.0
1	A	686	TYR	4.1
1	A	691	ALA	3.4
1	A	683	ALA	3.0

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

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	DOC	B	112	18/19	0.10	-0.10	21,26,31,31	0

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
5	EDO	C	302	4/4	0.20	12.73	50,50,50,51	0
5	EDO	A	905	4/4	0.14	4.42	36,41,41,43	0
6	FMT	A	912	3/3	0.23	4.17	70,70,72,72	0
6	FMT	C	301	3/3	0.19	3.42	68,68,69,70	0
6	FMT	C	303	3/3	0.13	3.26	66,66,67,68	0
5	EDO	A	909	4/4	0.22	3.03	51,53,54,56	0
6	FMT	B	201	3/3	0.14	1.76	74,74,74,75	0
6	FMT	A	906	3/3	0.12	1.73	55,55,56,57	0
5	EDO	A	911	4/4	0.26	1.64	70,70,71,71	0
6	FMT	A	904	3/3	0.17	0.94	60,60,61,61	0
5	EDO	A	903	4/4	0.14	0.93	27,27,35,35	0
4	MG	A	902	1/1	0.13	0.19	34,34,34,34	0
7	OL7	A	908	37/37	0.10	-0.20	23,31,41,43	0
4	MG	A	901	1/1	0.13	-0.47	34,34,34,34	0
6	FMT	A	907	3/3	0.13	-0.61	56,56,57,58	0
4	MG	A	910	1/1	0.07	-0.92	48,48,48,48	0
4	MG	A	913	1/1	0.07	-1.04	73,73,73,73	0
4	MG	B	202	1/1	0.05	-34.07	45,45,45,45	0

6.5 Other polymers ⓘ

There are no such residues in this entry.