WORLDWIDE wwPDB PROTEIN DATA BANK

Full wwPDB X-ray Structure Validation Report (i)

Feb 28, 2014 – 09:41 AM GMT

PDB ID : 4CCH

Title : Crystal structure of the large fragment of DNA polymerase I from Thermus Aquaticus in an open binary complex with d5SICS as templating nucleotide

Authors : Betz, K.; Malyshev, D.A.; Lavergne, T.; Welte, W.; Diederichs, K.; Romesberg, F.E.; Marx, A.

Deposited on : 2013-10-23

Resolution : 2.55 Å(reported)

Full wwPDB X-ray Structure Validation Report

1 Overall quality at a glance ⓘ

The reported resolution of this entry is 2.55 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.
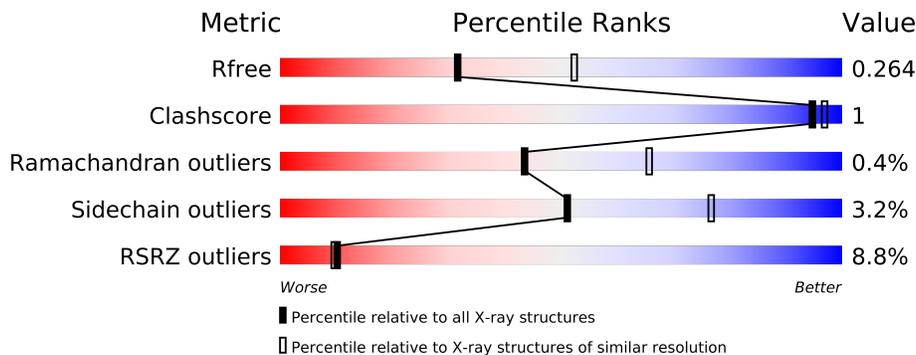
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	66092	3413 (2.58-2.50)
Clashscore	79885	4284 (2.58-2.50)
Ramachandran outliers	78287	4193 (2.58-2.50)
Sidechain outliers	78261	4195 (2.58-2.50)
RSRZ outliers	66119	3414 (2.58-2.50)

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
6	GOL	C	1217	-	X

2 Entry composition ⓘ

There are 7 unique types of molecules in this entry. The entry contains 9408 atoms, of which 4551 are hydrogens 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	528	Total 8458	C 2672	H 4253	N 757	O 764	S 12	0	0	0

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	12	Total 374	C 114	H 134	N 48	O 67	P 11	0	0	0

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

Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
3	C	13	Total 421	C 131	H 150	N 47	O 79	P 13	S 1	0	0	0

- Molecule 4 is FORMIC ACID (three-letter code: FMT) (formula: CH_2O_2).
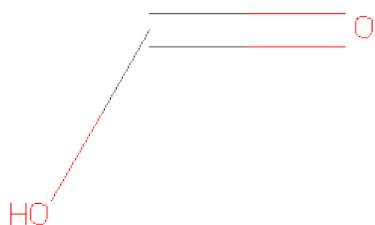
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total	C	H	O	0	0
			5	1	2	2		
4	A	1	Total	C	H	O	0	0
			5	1	2	2		
4	A	1	Total	C	H	O	0	0
			4	1	1	2		
4	A	1	Total	C	H	O	0	0
			4	1	1	2		

- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	1	Total	Mg	0	0
			1	1		

- Molecule 6 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).
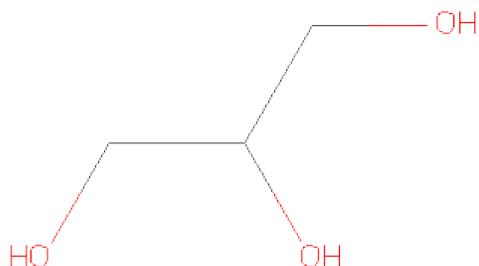
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
6	C	1	Total C H O 14 3 8 3				0	0

• Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	105	Total O 105 105		0	0
7	B	13	Total O 13 13		0	0
7	C	4	Total O 4 4		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: DNA POLYMERASE I, THERMOSTABLE
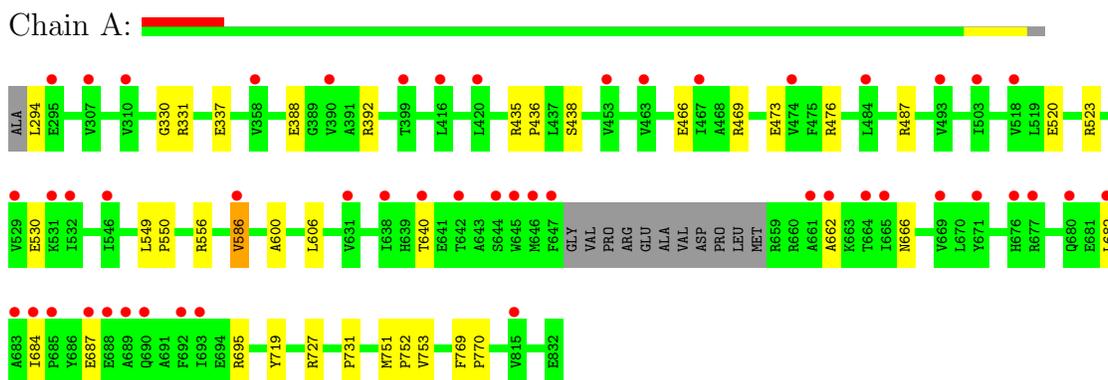
Chain A:

• Molecule 2: 5'-D(*GP*AP*CP*CP*AP*CP*GP*GP*CP*GP*CP*DOC)-3'

Chain B:

There are no outlier residues recorded for this chain.

• Molecule 3: 5'-D(*AP*AP*CP*LHOP*GP*GP*CP*GP*CP*CP*GP*TP*GP*GP*TP*C)-3'

Chain C:
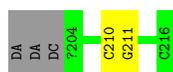
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α, β, γ	114.26Å　　114.26Å　　91.54Å 90.00°　　　90.00°　　　120.00°	Depositor
Resolution (Å)	43.53　–　2.55 48.47　–　2.55	Depositor EDS
% Data completeness (in resolution range)	99.9 (43.53-2.55) 99.9 (48.47-2.55)	Depositor EDS
R_{merge}	0.20	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	0.99 (at 2.54Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.207　,　0.261 0.213　,　0.264	Depositor DCC
R_{free} test set	1193 reflections (5.22%)	DCC
Wilson B-factor (Å^2)	46.8	Xtriage
Anisotropy	0.188	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 34.2	EDS
Estimated twinning fraction	0.037 for -h,-k,l	Xtriage
L-test for twinning	$< |L| > = 0.46, < L^2 > = 0.29$	Xtriage
Outliers	0 of 22869 reflections	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	9408	wwPDB-VP
Average B, all atoms (Å^2)	65.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.68% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]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: LHO, GOL, MG, FMT, DOC

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	0.27	0/4292	0.41	0/5810
2	B	0.61	0/249	0.82	0/382
3	C	0.57	0/277	0.80	0/426
All	All	0.32	0/4818	0.48	0/6618

There are no bond length outliers.

There are no bond angle outliers.

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	4205	4253	0	12	0
2	B	240	134	0	0	0
3	C	271	150	0	1	0
4	A	12	6	0	0	0
5	B	1	0	0	0	0
6	C	6	8	0	0	0
7	A	105	0	0	0	0
7	B	13	0	0	0	0
7	C	4	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	4857	4551	0	13	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 1.

All (13) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:294:LEU:N	1:A:330:GLY:O	2.34	0.60
1:A:388:GLU:OE2	1:A:392:ARG:NH2	2.35	0.59
1:A:523:ARG:NE	1:A:530:GLU:OE1	2.38	0.56
1:A:520:GLU:OE2	1:A:523:ARG:NH1	2.40	0.55
1:A:466:GLU:OE1	1:A:469:ARG:NH1	2.45	0.50
3:C:210:DC:H2'	3:C:211:DG:C8	2.49	0.47
1:A:662:ALA:O	1:A:666:ASN:ND2	2.46	0.47
1:A:600:ALA:HB3	1:A:606:LEU:HG	1.98	0.45
1:A:473:GLU:OE2	1:A:476:ARG:NE	2.51	0.44
1:A:435:ARG:N	1:A:436:PRO:HD2	2.35	0.42
1:A:549:LEU:HB2	1:A:550:PRO:HD3	2.03	0.41
1:A:751:MET:HB3	1:A:752:PRO:HD3	2.02	0.40
1:A:769:PHE:HB3	1:A:770:PRO:HD3	2.02	0.40

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	524/540 (97%)	498 (95%)	24 (5%)	2 (0%)	43	66

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	731	PRO
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	432/441 (98%)	418 (97%)	14 (3%)	51	77

All (14) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	331	ARG
1	A	337	GLU
1	A	438	SER
1	A	487	ARG
1	A	556	ARG
1	A	586	VAL
1	A	640	THR
1	A	682	LEU
1	A	684	ILE
1	A	687	GLU
1	A	695	ARG
1	A	719	TYR
1	A	727	ARG
1	A	753	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	566	GLN
1	A	627	ASN
1	A	666	ASN

5.3.3 RNA ⓘ

There are no RNA chains 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	DOC	B	112	3,2	17,19,20	3.19	8 (47%)	20,26,29	2.08	5 (25%)
3	LHO	C	204	3	23,25,26	1.78	2 (8%)	31,36,39	0.81	2 (6%)

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
3	LHO	C	204	3	-	0/5/21/22	0/1/3/3

All (10) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	204	LHO	C5-N1	7.28	1.41	1.36
2	B	112	DOC	O2-C2	6.39	1.36	1.23
2	B	112	DOC	C2-N1	6.28	1.45	1.38
2	B	112	DOC	C5-C4	5.07	1.52	1.40
2	B	112	DOC	C2-N3	5.05	1.48	1.35
2	B	112	DOC	P-OP1	3.32	1.50	1.46
2	B	112	DOC	C4-N4	3.00	1.44	1.35
2	B	112	DOC	C6-N1	2.60	1.40	1.35
3	C	204	LHO	P-OP1	2.49	1.49	1.46
2	B	112	DOC	O4'-C1'	-2.22	1.37	1.42

All (7) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	112	DOC	C6-C5-C4	5.64	119.81	117.47
2	B	112	DOC	C2-N3-C4	3.50	120.63	115.57
3	C	204	LHO	C9-C4-C5	-2.91	122.68	124.22
2	B	112	DOC	C4'-O4'-C1'	2.66	111.16	110.05
3	C	204	LHO	C5-N1-C1'	-2.65	117.22	120.82

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	112	DOC	C3'-C2'-C1'	2.45	105.51	102.80
2	B	112	DOC	O4'-C1'-C2'	2.17	108.96	106.37

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 6 ligands modelled in this entry, 1 is monoatomic - leaving 5 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	#$|Z| > 2$
4	FMT	A	1833	-	2,2,2	0.66	0	1,1,1	0.41	0
4	FMT	A	1834	-	2,2,2	0.61	0	1,1,1	0.38	0
4	FMT	A	1835	-	2,2,2	0.64	0	1,1,1	0.41	0
4	FMT	A	1836	-	2,2,2	0.63	0	1,1,1	0.42	0
6	GOL	C	1217	-	5,5,5	0.31	0	5,5,5	0.20	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	FMT	A	1833	-	-	0/0/0/0	0/0/0/0
4	FMT	A	1834	-	-	0/0/0/0	0/0/0/0
4	FMT	A	1835	-	-	0/0/0/0	0/0/0/0
4	FMT	A	1836	-	-	0/0/0/0	0/0/0/0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
6	GOL	C	1217	-	-	0/4/4/4	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 (i)

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	528/540 (97%)	0.66	49 (9%)	9	8	31, 56, 108, 153	0
2	B	12/12 (100%)	-0.05	0	100	100	31, 42, 71, 73	0
3	C	12/16 (75%)	0.01	0	100	100	30, 43, 59, 66	0
All	All	552/568 (97%)	0.63	49 (8%)	10	9	30, 56, 108, 153	0

All (49) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	684	ILE	5.0
1	A	689	ALA	4.8
1	A	642	THR	4.5
1	A	647	PHE	4.4
1	A	474	VAL	4.3
1	A	645	TRP	4.2
1	A	638	ILE	4.1
1	A	420	LEU	4.1
1	A	529	VAL	4.1
1	A	688	GLU	4.0
1	A	467	ILE	4.0
1	A	493	VAL	3.7
1	A	685	PRO	3.6
1	A	662	ALA	3.5
1	A	463	VAL	3.4
1	A	690	GLN	3.4
1	A	358	VAL	3.2
1	A	453	VAL	3.1
1	A	687	GLU	3.1
1	A	661	ALA	3.1
1	A	503	ILE	2.9
1	A	683	ALA	2.9
1	A	682	LEU	2.9

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	631	VAL	2.9
1	A	692	PHE	2.6
1	A	646	MET	2.5
1	A	307	VAL	2.5
1	A	815	VAL	2.4
1	A	677	ARG	2.4
1	A	680	GLN	2.4
1	A	676	HIS	2.4
1	A	546	ILE	2.3
1	A	310	VAL	2.3
1	A	532	ILE	2.3
1	A	295	GLU	2.3
1	A	693	ILE	2.3
1	A	484	LEU	2.2
1	A	518	VAL	2.2
1	A	586	VAL	2.2
1	A	671	TYR	2.2
1	A	644	SER	2.2
1	A	640	THR	2.2
1	A	531	LYS	2.2
1	A	664	THR	2.1
1	A	399	THR	2.1
1	A	665	ILE	2.1
1	A	390	VAL	2.0
1	A	669	VAL	2.0
1	A	416	LEU	2.0

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(Å²)	Q<0.9
3	LHO	C	204	23/24	0.28	0.34	52,77,97,98	0
2	DOC	B	112	18/19	0.14	-1.00	19,31,42,48	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
6	GOL	C	1217	6/6	0.23	5.10	47,63,82,94	0
4	FMT	A	1835	3/3	0.17	0.20	13,13,14,15	4
4	FMT	A	1833	3/3	0.17	0.05	20,20,24,24	5
4	FMT	A	1834	3/3	0.13	-1.12	19,19,23,23	5
4	FMT	A	1836	3/3	0.12	-2.18	53,60,64,72	0
5	MG	B	1112	1/1	0.04	-6.99	62,62,62,62	0

6.5 Other polymers ⓘ

There are no such residues in this entry.