Full wwPDB X-ray Structure Validation Report (i)

May 18, 2020 – 07:44 am BST

PDB ID	:	4C8N
Title	:	Binary complex of the large fragment of DNA polymerase I from Thermus Aquaticus with the aritificial base pair dNaM-d5SICS at the postinsertion site (sequence context 3)
Authors	:	Betz, K.; Malyshev, D.A.; Lavergne, T.; Welte, W.; Diederichs, K.; Romesberg, F.E.; Marx, A.
Deposited on	:	2013-10-01
Resolution	:	1.88 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

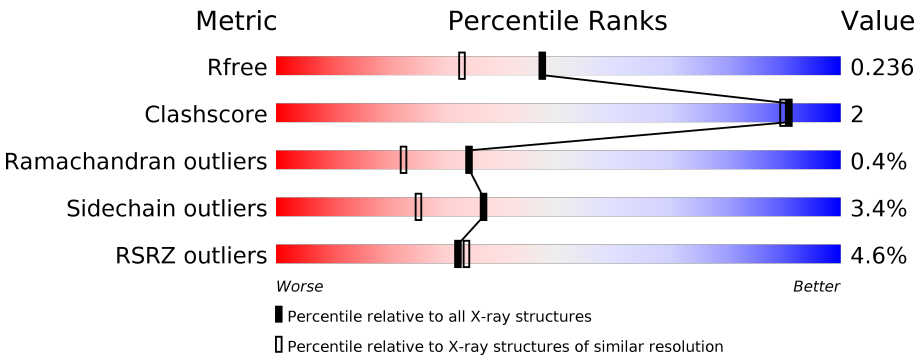
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 1.88 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}	130704	9470 (1.90-1.86)
Clashscore	141614	10282 (1.90-1.86)
Ramachandran outliers	138981	10152 (1.90-1.86)
Sidechain outliers	138945	10152 (1.90-1.86)
RSRZ outliers	127900	9303 (1.90-1.86)

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 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 <=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	5% 93% 7%
2	B	11	91% 9%
3	C	14	93% 7%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 9683 atoms, of which 4648 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 LARGE FRAGMENT OF TAQ DNA POLYMERASE I.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	540	8657	2732	4355	776	781	13	0	1	0

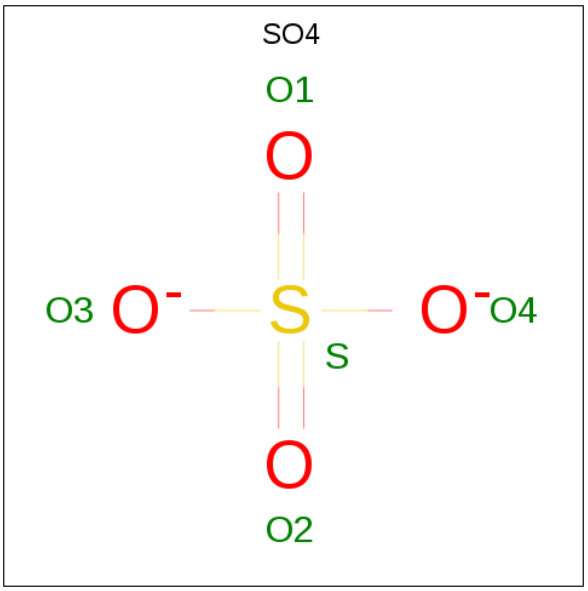
- Molecule 2 is a DNA chain called PRIMER, 5'-D(*AP*CP*CP*AP*CP*GP*GP*CP*GP*CP*LHOP)-3'.

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

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

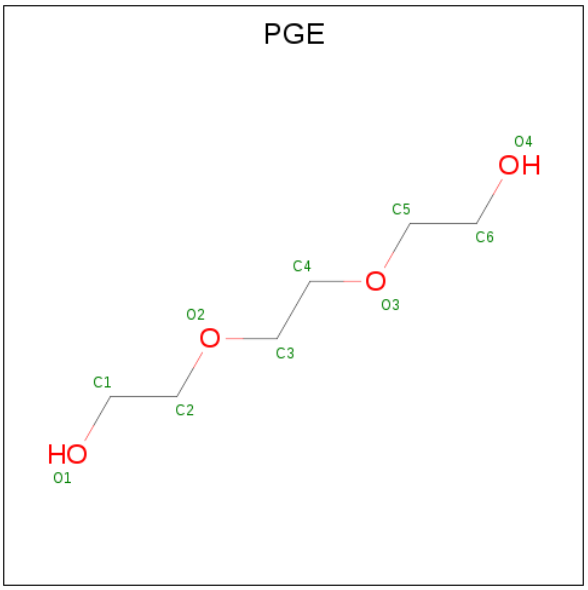
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P			
3	C	13	424	133	152	45	81	13	0	0	0

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



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

- Molecule 5 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula: $C_6H_{14}O_4$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	H	O		
5	A	1	24	6	14	4	0	0

- Molecule 6 is water.

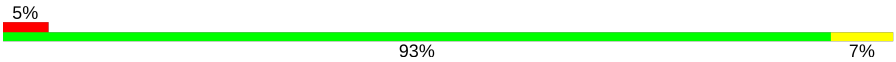
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
6	A	166	166	166	0	0
			Total	O		
6	B	24	24	24	0	0
			Total	O		
6	C	13	13	13	0	0

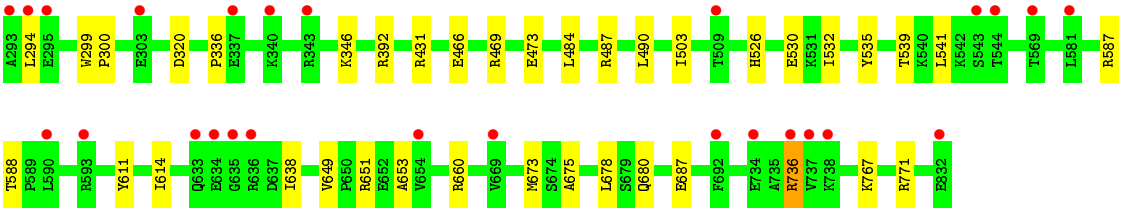
3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes 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: LARGE FRAGMENT OF TAQ DNA POLYMERASE I

Chain A:



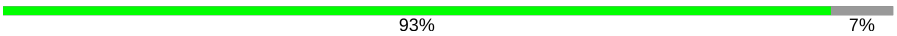


• Molecule 2: PRIMER, 5'-D(*AP*CP*CP*AP*CP*GP*GP*CP*GP*CP*LHOP)-3'

Chain B:

• Molecule 3: TEMPLATE, 5'-D(*TP*TP*GP*BMNP*GP*CP*GP*CP*CP*GP*TP*GP*GP*TP)-3'

Chain C:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	65.62Å 101.10Å 204.32Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.07 – 1.88 49.07 – 1.88	Depositor EDS
% Data completeness (in resolution range)	99.5 (49.07-1.88) 99.5 (49.07-1.88)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.02 (at 1.88Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.196 , 0.232 0.200 , 0.236	Depositor DCC
R_{free} test set	2776 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	37.2	Xtriage
Anisotropy	0.667	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.40 , 40.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	9683	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 4.15% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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: LHO, PGE, BMN, 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	0.37	0/4392	0.53	0/5948
2	B	0.68	0/224	0.81	0/343
3	C	0.75	0/277	1.03	0/424
All	All	0.42	0/4893	0.59	0/6715

There are no bond length outliers.

There are no bond angle outliers.

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	4302	4355	4353	16	0
2	B	223	127	127	1	0
3	C	272	152	152	0	0
4	A	15	0	0	1	0
4	B	5	0	0	0	0
4	C	5	0	0	0	0
5	A	10	14	14	1	0
6	A	166	0	0	3	0
6	B	24	0	0	0	0
6	C	13	0	0	0	0
All	All	5035	4648	4646	16	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 (16) 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:431[A]:ARG:NH2	6:A:2051:HOH:O	2.25	0.70
1:A:466:GLU:OE2	1:A:469:ARG:NH2	2.28	0.65
1:A:767:LYS:NZ	6:A:2158:HOH:O	2.30	0.64
1:A:660:ARG:NH1	5:A:1836:PGE:O1	2.38	0.57
1:A:675:ALA:N	4:A:1834:SO4:O3	2.35	0.56
1:A:673:MET:HE3	1:A:678:LEU:HD13	1.86	0.56
1:A:587:ARG:NH2	2:B:112:LHO:OP2	2.44	0.49
1:A:611:TYR:HB3	1:A:614:ILE:HB	1.94	0.48
1:A:649:VAL:CG1	1:A:653:ALA:HB3	2.47	0.44
1:A:736:ARG:HE	1:A:736:ARG:H	1.65	0.44
1:A:649:VAL:HG12	1:A:653:ALA:HB3	1.98	0.43
1:A:299:TRP:CG	1:A:300:PRO:HA	2.54	0.43
1:A:526:HIS:O	6:A:2087:HOH:O	2.22	0.42
1:A:687:GLU:OE1	1:A:687:GLU:N	2.49	0.41
1:A:535:TYR:CZ	1:A:539:THR:HG21	2.56	0.41
1:A:490:LEU:HD11	1:A:532:ILE:HD13	2.03	0.40

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

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%)	529 (98%)	8 (2%)	2 (0%)	34 22

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	336	PRO
1	A	503	ILE

5.3.2 Protein sidechains (i)

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	442/441 (100%)	427 (97%)	15 (3%)	37 25

All (15) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	294	LEU
1	A	320	ASP
1	A	346	LYS
1	A	392	ARG
1	A	473	GLU
1	A	484	LEU
1	A	487	ARG
1	A	530	GLU
1	A	541	LEU
1	A	588	THR
1	A	638	ILE
1	A	651	ARG
1	A	680	GLN
1	A	736	ARG
1	A	771	ARG

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

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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	$\#\|Z\| > 2$
3	BMN	C	205	3	22,25,26	0.20	0	31,35,38	0.88	1 (3%)
2	LHO	B	112	2	21,25,26	0.87	1 (4%)	26,36,39	0.40	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
3	BMN	C	205	3	-	0/9/23/24	0/3/3/3
2	LHO	B	112	2	-	1/4/21/22	0/3/3/3

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	112	LHO	C5-C4	2.21	1.45	1.41

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	205	BMN	C2-C1-C6	3.61	122.18	117.06

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.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	112	LHO	1	0

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

6 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	$\#\|Z\| > 2$
4	SO4	A	1833	-	4,4,4	0.14	0	6,6,6	0.13	0
4	SO4	A	1834	-	4,4,4	0.15	0	6,6,6	0.12	0
4	SO4	B	1112	-	4,4,4	0.17	0	6,6,6	0.09	0
4	SO4	C	1216	-	4,4,4	0.17	0	6,6,6	0.29	0
4	SO4	A	1835	-	4,4,4	0.15	0	6,6,6	0.05	0
5	PGE	A	1836	-	9,9,9	0.31	0	8,8,8	0.34	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	PGE	A	1836	-	-	5/7/7/7	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (5) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	A	1836	PGE	O3-C5-C6-O4
5	A	1836	PGE	O1-C1-C2-O2
5	A	1836	PGE	C1-C2-O2-C3
5	A	1836	PGE	O2-C3-C4-O3
5	A	1836	PGE	C6-C5-O3-C4

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	1834	SO4	1	0
5	A	1836	PGE	1	0

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 (i)

6.1 Protein, DNA and RNA chains (i)

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%)	0.48	26 (4%) 30 32	29, 43, 76, 151	0
2	B	10/11 (90%)	0.12	0 100 100	33, 41, 49, 51	0
3	C	12/14 (85%)	0.27	0 100 100	38, 41, 73, 100	0
All	All	562/565 (99%)	0.47	26 (4%) 32 34	29, 43, 76, 151	0

All (26) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	293	ALA	7.0
1	A	294	LEU	5.0
1	A	832	GLU	4.6
1	A	692	PHE	3.6
1	A	734	GLU	3.5
1	A	340	LYS	3.2
1	A	636	ARG	3.1
1	A	738	LYS	3.1
1	A	736	ARG	3.0
1	A	343	ARG	2.7
1	A	295	GLU	2.7
1	A	593	ARG	2.6
1	A	581	LEU	2.4
1	A	590	LEU	2.4
1	A	654	VAL	2.4
1	A	737	VAL	2.4
1	A	633	GLN	2.4
1	A	634	GLU	2.4
1	A	569	THR	2.4
1	A	635	GLY	2.2
1	A	543	SER	2.2
1	A	669	VAL	2.1
1	A	337	GLU	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	509	THR	2.1
1	A	544	THR	2.0
1	A	303	GLU	2.0

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	0.93	0.13	38,51,69,77	0
2	LHO	B	112	23/24	0.98	0.18	36,45,57,57	0

6.3 Carbohydrates (i)

There are no carbohydrates 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
5	PGE	A	1836	10/10	0.75	0.17	29,42,61,74	0
4	SO4	C	1216	5/5	0.84	0.15	48,50,73,81	0
4	SO4	B	1112	5/5	0.86	0.18	48,61,67,68	5
4	SO4	A	1833	5/5	0.89	0.18	61,67,75,77	0
4	SO4	A	1834	5/5	0.90	0.13	81,81,93,94	0
4	SO4	A	1835	5/5	0.96	0.10	74,82,88,91	0

6.5 Other polymers (i)

There are no such residues in this entry.