



Full wwPDB X-ray Structure Validation Report

Apr 28, 2024 – 05:39 am BST

PDB ID : 4C8M
Title : Binary complex of the large fragment of DNA polymerase I from *Thermus Aquaticus* with the artificial base pair d5SICS-dNaM at the postinsertion site (sequence context 2)
Authors : Betz, K.; Malyshev, D.A.; Lavergne, T.; Welte, W.; Diederichs, K.; Romesberg, F.E.; Marx, A.
Deposited on : 2013-10-01
Resolution : 1.57 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 : 4.02b-467
Mogul : 1.8.4, CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.36.2
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) : 2.36.2

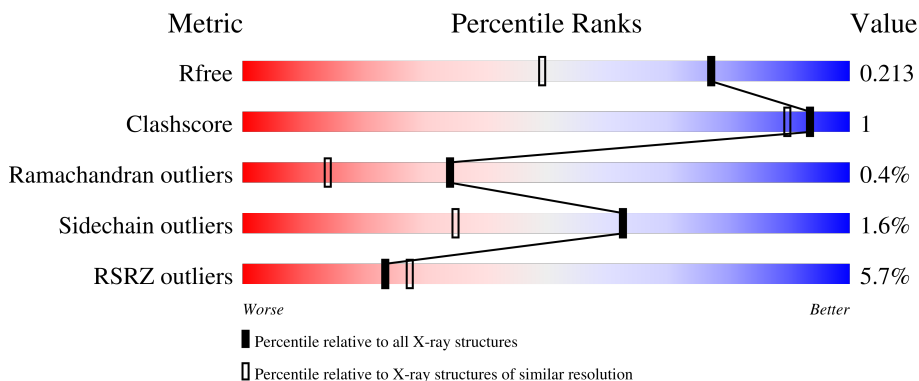
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.57 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	1483 (1.56-1.56)
Clashscore	141614	1529 (1.56-1.56)
Ramachandran outliers	138981	1498 (1.56-1.56)
Sidechain outliers	138945	1495 (1.56-1.56)
RSRZ outliers	127900	1465 (1.56-1.56)

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	 6% 95% 5%
2	B	11	 73% 27%
3	C	14	 14% 79% 14% 7%

2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 9935 atoms, of which 4706 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	8728	2758	4388	779	790	13	0	10	0

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	P			
2	B	11	352	111	128	40	63	10	0	0	0

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace	
			Total	C	H	N	O	P				S
3	C	13	418	130	150	45	79	13	1	0	0	0

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



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

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).



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

Continued on next page...

Continued from previous page...

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

- Molecule 6 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

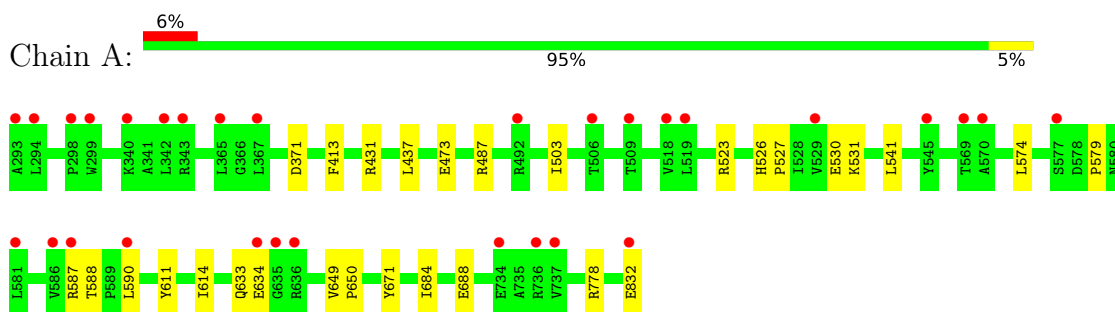
- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	318	Total	O	0	0
			318	318		
7	B	14	Total	O	0	0
			14	14		
7	C	18	Total	O	0	0
			18	18		

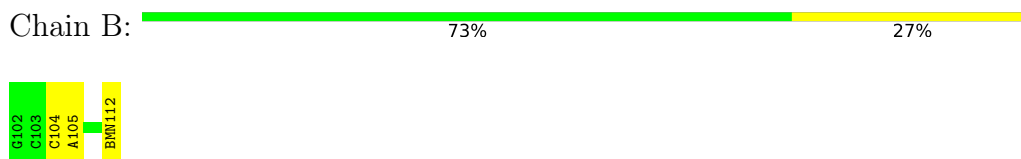
3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

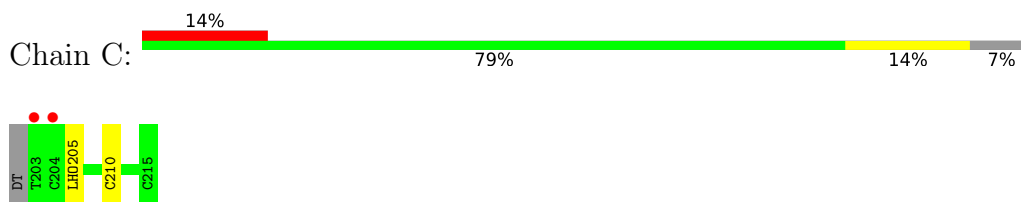
- Molecule 1: LARGE FRAGMENT OF TAQ DNA POLYMERASE I



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



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



4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α , β , γ	65.37Å 101.46Å 204.54Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.24 – 1.57 49.24 – 1.57	Depositor EDS
% Data completeness (in resolution range)	98.7 (49.24-1.57) 98.4 (49.24-1.57)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	0.96 (at 1.57Å)	Xtrriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.178 , 0.207 0.186 , 0.213	Depositor DCC
R_{free} test set	4653 reflections (4.96%)	wwPDB-VP
Wilson B-factor (Å ²)	30.9	Xtrriage
Anisotropy	0.486	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.39 , 47.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.97	EDS
Total number of atoms	9935	wwPDB-VP
Average B, all atoms (Å ²)	50.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.59% 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: SO4, MG, GOL, 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	0.56	0/4465	0.65	0/6046
2	B	0.81	0/225	0.89	0/345
3	C	0.89	0/272	1.08	0/415
All	All	0.59	0/4962	0.70	0/6806

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	4340	4388	4380	11	0
2	B	224	128	128	1	0
3	C	268	150	150	2	0
4	A	15	0	0	0	0
5	A	24	32	32	0	0
5	C	6	8	8	1	0
6	B	1	0	0	0	0
6	C	1	0	0	0	0
7	A	318	0	0	1	0
7	B	14	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
7	C	18	0	0	0	0
All	All	5229	4706	4698	13	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 1.

All (13) 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)
3:C:210:DC:H41	5:C:1217:GOL:C3	2.20	0.54
1:A:671:TYR:CE1	3:C:205:LHO:H310	2.44	0.53
1:A:541:LEU:HD22	1:A:590:LEU:HB3	1.95	0.49
1:A:523:ARG:NE	1:A:530:GLU:OE1	2.48	0.47
1:A:684:ILE:HD12	1:A:688:GLU:HG2	1.96	0.47
2:B:104:DC:H2'	2:B:105:DA:C8	2.53	0.44
1:A:611:TYR:HB3	1:A:614:ILE:HB	1.99	0.44
1:A:649:VAL:HB	1:A:650:PRO:HD2	1.99	0.43
1:A:371:ASP:HB3	1:A:413:PHE:CE1	2.53	0.43
1:A:437:LEU:HD21	1:A:574:LEU:HD21	2.01	0.42
1:A:431:ARG:HD2	7:A:2090:HOH:O	2.20	0.41
1:A:526:HIS:HA	1:A:527:PRO:HD3	1.96	0.41
1:A:473:GLU:HG2	1:A:531:LYS:HE3	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	548/540 (102%)	539 (98%)	7 (1%)	2 (0%)	34 14

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	503	ILE
1	A	588	THR

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	450/441 (102%)	443 (98%)	7 (2%)	62 35

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	487	ARG
1	A	579	PRO
1	A	587	ARG
1	A	633	GLN
1	A	634	GLU
1	A	778	ARG
1	A	832	GLU

Sometimes 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	$\# Z > 2$
2	BMN	B	112	2	22,25,26	0.16	0	31,35,38	0.78	1 (3%)
3	LHO	C	205	3	20,25,26	0.24	0	27,36,39	0.56	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	BMN	B	112	2	-	0/9/23/24	0/3/3/3
3	LHO	C	205	3	-	0/7/21/22	0/3/3/3

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	112	BMN	C2-C1-C6	3.43	121.93	117.06

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	C	205	LHO	1	0

5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

5.6 Ligand geometry [i](#)

Of 10 ligands modelled in this entry, 2 are monoatomic - leaving 8 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	GOL	A	1837	-	5,5,5	0.39	0	5,5,5	0.11	0
5	GOL	A	1839	-	5,5,5	0.28	0	5,5,5	0.36	0
4	SO4	A	1836	-	4,4,4	0.15	0	6,6,6	0.15	0
4	SO4	A	1834	-	4,4,4	0.16	0	6,6,6	0.18	0
5	GOL	A	1838	-	5,5,5	0.27	0	5,5,5	0.61	0
4	SO4	A	1833	-	4,4,4	0.14	0	6,6,6	0.24	0
5	GOL	C	1217	-	5,5,5	0.33	0	5,5,5	0.48	0
5	GOL	A	1835	-	5,5,5	0.40	0	5,5,5	0.42	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	GOL	A	1837	-	-	2/4/4/4	-
5	GOL	A	1839	-	-	4/4/4/4	-
5	GOL	A	1838	-	-	2/4/4/4	-
5	GOL	C	1217	-	-	2/4/4/4	-
5	GOL	A	1835	-	-	4/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (14) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	A	1835	GOL	C1-C2-C3-O3
5	A	1837	GOL	C1-C2-C3-O3
5	A	1837	GOL	O2-C2-C3-O3
5	A	1838	GOL	O1-C1-C2-O2
5	A	1838	GOL	O1-C1-C2-C3
5	A	1839	GOL	C1-C2-C3-O3
5	C	1217	GOL	C1-C2-C3-O3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
5	A	1835	GOL	O1-C1-C2-O2
5	C	1217	GOL	O2-C2-C3-O3
5	A	1835	GOL	O1-C1-C2-C3
5	A	1839	GOL	O1-C1-C2-C3
5	A	1835	GOL	O2-C2-C3-O3
5	A	1839	GOL	O2-C2-C3-O3
5	A	1839	GOL	O1-C1-C2-O2

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	C	1217	GOL	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

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	30 (5%) 24 28	26, 40, 71, 105	0
2	B	10/11 (90%)	-0.26	0 100 100	43, 48, 54, 55	0
3	C	12/14 (85%)	0.55	2 (16%) 1 1	40, 57, 118, 151	0
All	All	562/565 (99%)	0.21	32 (5%) 23 27	26, 40, 71, 151	0

All (32) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	832	GLU	6.5
1	A	635	GLY	5.9
1	A	298	PRO	4.4
3	C	203	DT	4.4
1	A	736	ARG	4.3
1	A	343	ARG	3.8
1	A	636	ARG	3.8
1	A	529	VAL	3.7
1	A	293	ALA	3.2
1	A	590	LEU	3.2
1	A	586	VAL	3.0
1	A	367	LEU	2.9
1	A	294	LEU	2.9
1	A	587	ARG	2.8
1	A	734	GLU	2.7
1	A	340	LYS	2.7
1	A	518	VAL	2.7
1	A	519	LEU	2.6
1	A	737	VAL	2.6
1	A	569	THR	2.6
1	A	545	TYR	2.6
1	A	365	LEU	2.5
1	A	581	LEU	2.5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
3	C	204	DC	2.4
1	A	342	LEU	2.4
1	A	509	THR	2.3
1	A	492	ARG	2.3
1	A	299	TRP	2.3
1	A	634	GLU	2.1
1	A	570	ALA	2.1
1	A	506	THR	2.0
1	A	577[A]	SER	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	LHO	C	205	23/24	0.70	0.16	67,90,110,110	0
2	BMN	B	112	23/24	0.92	0.14	60,77,108,108	0

6.3 Carbohydrates [i](#)

There are no monosaccharides 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	1112	1/1	0.02	0.10	66,66,66,66	1
5	GOL	A	1839	6/6	0.61	0.17	59,70,87,87	0
5	GOL	C	1217	6/6	0.82	0.12	53,70,80,85	0
5	GOL	A	1838	6/6	0.83	0.14	40,65,79,81	0
5	GOL	A	1835	6/6	0.88	0.15	33,51,64,77	14
5	GOL	A	1837	6/6	0.89	0.15	60,82,95,98	0
4	SO4	A	1834	5/5	0.93	0.13	68,80,90,91	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
4	SO4	A	1836	5/5	0.94	0.15	55,55,62,70	5
6	MG	C	1216	1/1	0.94	0.13	58,58,58,58	0
4	SO4	A	1833	5/5	0.99	0.04	42,47,51,60	5

6.5 Other polymers [i](#)

There are no such residues in this entry.