Full wwPDB X-ray Structure Validation Report ⓘ

May 21, 2020 – 10:59 pm BST

PDB ID	:	5OLQ
Title	:	Rhamnogalacturonan lyase
Authors	:	Basle, A.; Luis, A.S.; Gilbert, H.J.
Deposited on	:	2017-07-28
Resolution	:	1.48 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 following versions of software and data (see references ⓘ) were used in the production of this report:

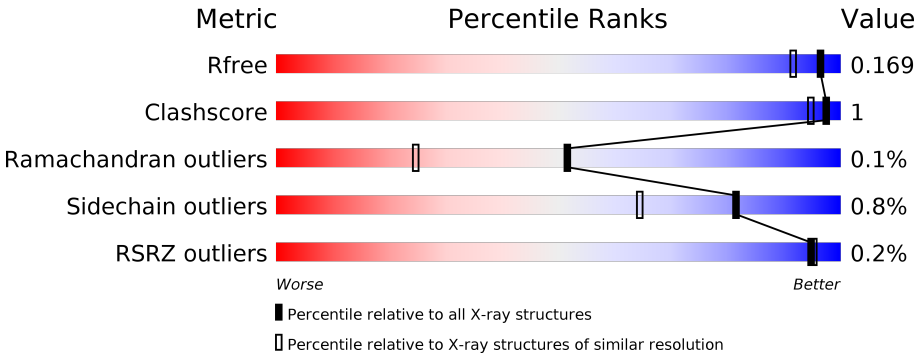
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 1.48 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	4690 (1.50-1.46)
Clashscore	141614	4955 (1.50-1.46)
Ramachandran outliers	138981	4846 (1.50-1.46)
Sidechain outliers	138945	4844 (1.50-1.46)
RSRZ outliers	127900	4614 (1.50-1.46)

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	522	78% · 18%
1	B	522	79% · 18%
1	C	522	80% · 18%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 11428 atoms, of which 0 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 Rhamnogalacturonan lyase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	426	Total	C	N	O	S	0	5	0
			3297	2067	569	644	17			
1	B	426	Total	C	N	O	S	0	4	0
			3289	2066	563	642	18			
1	C	427	Total	C	N	O	S	0	6	0
			3305	2076	565	647	17			

There are 30 discrepancies between the modelled and reference sequences:

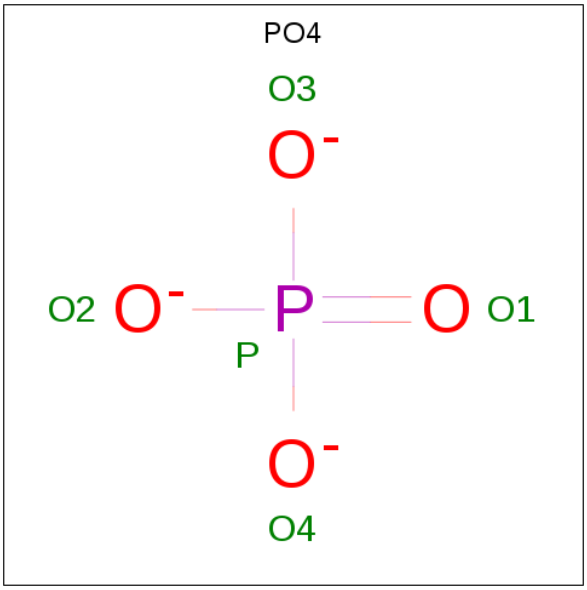
Chain	Residue	Modelled	Actual	Comment	Reference
A	22	MET	-	initiating methionine	UNP A0A139KMS2
A	23	GLY	-	expression tag	UNP A0A139KMS2
A	536	LEU	-	expression tag	UNP A0A139KMS2
A	537	GLU	-	expression tag	UNP A0A139KMS2
A	538	HIS	-	expression tag	UNP A0A139KMS2
A	539	HIS	-	expression tag	UNP A0A139KMS2
A	540	HIS	-	expression tag	UNP A0A139KMS2
A	541	HIS	-	expression tag	UNP A0A139KMS2
A	542	HIS	-	expression tag	UNP A0A139KMS2
A	543	HIS	-	expression tag	UNP A0A139KMS2
B	22	MET	-	initiating methionine	UNP A0A139KMS2
B	23	GLY	-	expression tag	UNP A0A139KMS2
B	536	LEU	-	expression tag	UNP A0A139KMS2
B	537	GLU	-	expression tag	UNP A0A139KMS2
B	538	HIS	-	expression tag	UNP A0A139KMS2
B	539	HIS	-	expression tag	UNP A0A139KMS2
B	540	HIS	-	expression tag	UNP A0A139KMS2
B	541	HIS	-	expression tag	UNP A0A139KMS2
B	542	HIS	-	expression tag	UNP A0A139KMS2
B	543	HIS	-	expression tag	UNP A0A139KMS2
C	22	MET	-	initiating methionine	UNP A0A139KMS2
C	23	GLY	-	expression tag	UNP A0A139KMS2
C	536	LEU	-	expression tag	UNP A0A139KMS2

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	537	GLU	-	expression tag	UNP A0A139KMS2
C	538	HIS	-	expression tag	UNP A0A139KMS2
C	539	HIS	-	expression tag	UNP A0A139KMS2
C	540	HIS	-	expression tag	UNP A0A139KMS2
C	541	HIS	-	expression tag	UNP A0A139KMS2
C	542	HIS	-	expression tag	UNP A0A139KMS2
C	543	HIS	-	expression tag	UNP A0A139KMS2

- Molecule 2 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).



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

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total Ca 1 1	0	0
3	A	1	Total Ca 1 1	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	C	1	Total Ca 1 1	0	0

- Molecule 4 is water.

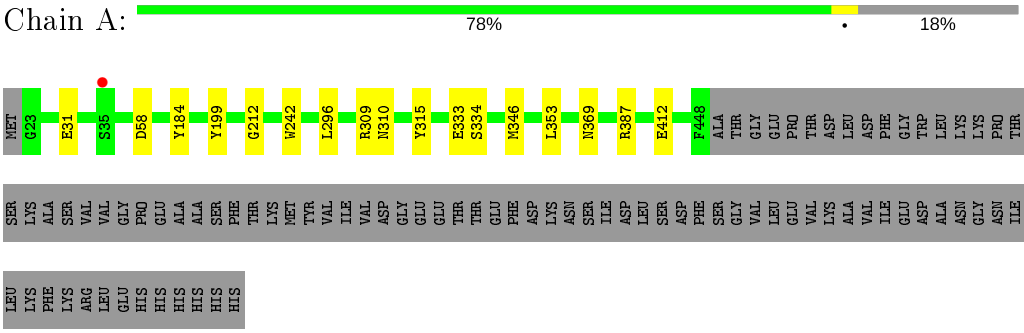
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	488	Total O 488 488	0	0
4	B	519	Total O 519 519	0	0
4	C	512	Total O 512 512	0	0

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.

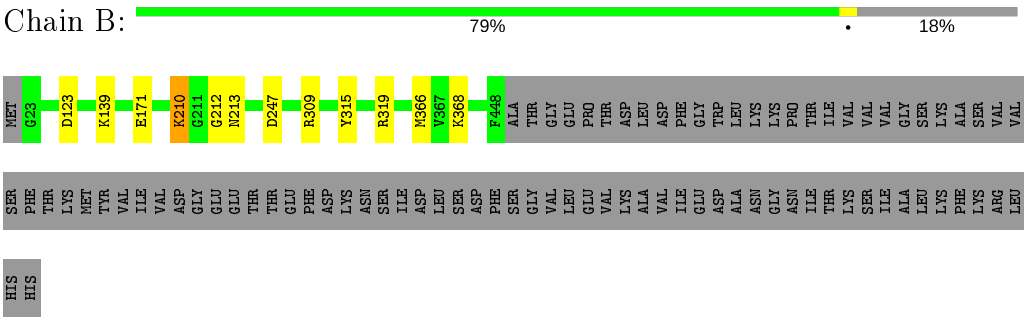
- Molecule 1: Rhamnogalacturonan lyase

Chain A:



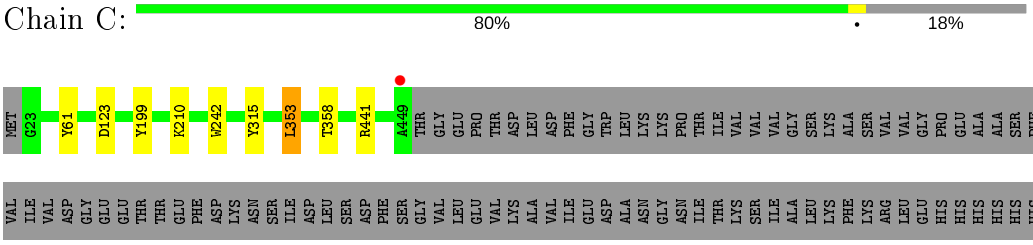
- Molecule 1: Rhamnogalacturonan lyase

Chain B:



- Molecule 1: Rhamnogalacturonan lyase

Chain C:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	77.55Å 123.82Å 138.42Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.65 – 1.48 39.65 – 1.48	Depositor EDS
% Data completeness (in resolution range)	99.9 (39.65-1.48) 99.9 (39.65-1.48)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.65 (at 1.48Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.127 , 0.169 0.127 , 0.169	Depositor DCC
R_{free} test set	10869 reflections (4.92%)	wwPDB-VP
Wilson B-factor (Å^2)	13.4	Xtriage
Anisotropy	0.200	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 42.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	11428	wwPDB-VP
Average B, all atoms (Å^2)	17.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 48.62 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 8.3870e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: PO4, CA

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	0.64	2/3387 (0.1%)	0.77	2/4587 (0.0%)
1	B	0.62	0/3377	0.76	1/4574 (0.0%)
1	C	0.60	0/3399	0.78	3/4605 (0.1%)
All	All	0.62	2/10163 (0.0%)	0.77	6/13766 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	31	GLU	CD-OE2	-5.23	1.19	1.25
1	A	184	TYR	CE1-CZ	5.08	1.45	1.38

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	346	MET	CG-SD-CE	6.98	111.37	100.20
1	C	441	ARG	NE-CZ-NH1	5.73	123.16	120.30
1	C	61	TYR	CB-CG-CD2	-5.45	117.73	121.00
1	A	58	ASP	CB-CG-OD2	-5.19	113.62	118.30
1	B	247	ASP	CB-CG-OD1	-5.10	113.71	118.30
1	C	353	LEU	CA-CB-CG	5.01	126.83	115.30

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-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 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	3297	0	3161	8	0
1	B	3289	0	3149	6	0
1	C	3305	0	3166	2	0
2	A	5	0	0	0	0
2	B	5	0	0	0	0
2	C	5	0	0	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
4	A	488	0	0	3	0
4	B	519	0	0	3	0
4	C	512	0	0	0	0
All	All	11428	0	9476	15	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 (15) 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:387[A]:ARG:NH1	4:A:701:HOH:O	2.06	0.88
1:B:212:GLY:HA3	4:B:978:HOH:O	1.74	0.86
1:A:212:GLY:HA3	4:A:755:HOH:O	1.82	0.80
1:A:333[B]:GLU:OE2	1:A:334:SER:N	2.21	0.73
1:B:309:ARG:NH1	1:C:123:ASP:OD2	2.27	0.67
1:A:309[A]:ARG:NH2	4:A:704:HOH:O	2.29	0.64
1:A:333[B]:GLU:OE2	1:A:369:ASN:HB2	1.98	0.64
1:B:366:MET:SD	1:B:368:LYS:NZ	2.76	0.57
1:B:210:LYS:NZ	4:B:702:HOH:O	2.27	0.57
1:B:319:ARG:NH1	4:B:701:HOH:O	2.40	0.53
1:A:296:LEU:HD22	1:A:353:LEU:HG	2.01	0.42
1:A:310:ASN:HA	1:A:334:SER:O	2.21	0.41
1:C:199:TYR:HA	1:C:242:TRP:O	2.21	0.40
1:B:139:LYS:HA	1:B:171:GLU:O	2.22	0.40
1:A:199:TYR:HA	1:A:242:TRP:O	2.21	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	429/522 (82%)	422 (98%)	7 (2%)	0	100 100
1	B	428/522 (82%)	418 (98%)	9 (2%)	1 (0%)	47 23
1	C	431/522 (83%)	422 (98%)	9 (2%)	0	100 100
All	All	1288/1566 (82%)	1262 (98%)	25 (2%)	1 (0%)	51 25

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	210	LYS

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	359/436 (82%)	357 (99%)	2 (1%)	86 72
1	B	358/436 (82%)	355 (99%)	3 (1%)	81 64
1	C	360/436 (83%)	355 (99%)	5 (1%)	67 40
All	All	1077/1308 (82%)	1067 (99%)	10 (1%)	81 59

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	315	TYR
1	A	412	GLU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	123	ASP
1	B	213	ASN
1	B	315	TYR
1	C	210	LYS
1	C	315	TYR
1	C	353	LEU
1	C	358[A]	THR
1	C	358[B]	THR

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)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 6 ligands modelled in this entry, 3 are monoatomic - leaving 3 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	$\#\|Z\| > 2$
2	PO4	A	601	3	4,4,4	1.26	1 (25%)	6,6,6	1.70	2 (33%)
2	PO4	B	601	3	4,4,4	1.23	1 (25%)	6,6,6	1.30	1 (16%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\|>2$	Counts	RMSZ	$\#\|Z\|>2$
2	PO4	C	601	3	4,4,4	1.29	1 (25%)	6,6,6	1.20	0

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	601	PO4	P-O1	2.57	1.56	1.50
2	A	601	PO4	P-O1	2.32	1.56	1.50
2	B	601	PO4	P-O1	2.18	1.56	1.50

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	601	PO4	O4-P-O3	2.58	116.25	107.97
2	A	601	PO4	O4-P-O3	2.41	115.72	107.97
2	A	601	PO4	O3-P-O1	-2.35	102.28	110.89

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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, 95^{th} 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	426/522 (81%)	-0.50	1 (0%) 95 95	9, 14, 28, 43	0
1	B	426/522 (81%)	-0.59	0 100 100	9, 14, 28, 45	0
1	C	427/522 (81%)	-0.58	1 (0%) 95 95	8, 14, 26, 41	0
All	All	1279/1566 (81%)	-0.56	2 (0%) 95 95	8, 14, 28, 45	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	449	ALA	3.4
1	A	35	SER	2.5

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

There are no non-standard protein/DNA/RNA residues in this entry.

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. The B-factors column lists the minimum, median, 95^{th} 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
2	PO4	A	601	5/5	0.81	0.21	18,29,32,35	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	PO4	B	601	5/5	0.88	0.15	16,21,27,28	0
2	PO4	C	601	5/5	0.92	0.13	16,22,23,26	0
3	CA	B	602	1/1	1.00	0.03	12,12,12,12	0
3	CA	C	602	1/1	1.00	0.04	12,12,12,12	0
3	CA	A	602	1/1	1.00	0.02	13,13,13,13	0

6.5 Other polymers

There are no such residues in this entry.