Full wwPDB X-ray Structure Validation Report ⓘ

Oct 5, 2023 – 10:30 PM EDT

PDB ID	:	6Q0C
Title	:	MutY adenine glycosylase bound to DNA containing a transition state analog (1N) paired with undamaged dG
Authors	:	O'Shea Murray, V.L.; Russelburg, L.P.; Horvath, M.P.; David, S.S.
Deposited on	:	2019-08-01
Resolution	:	2.00 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:

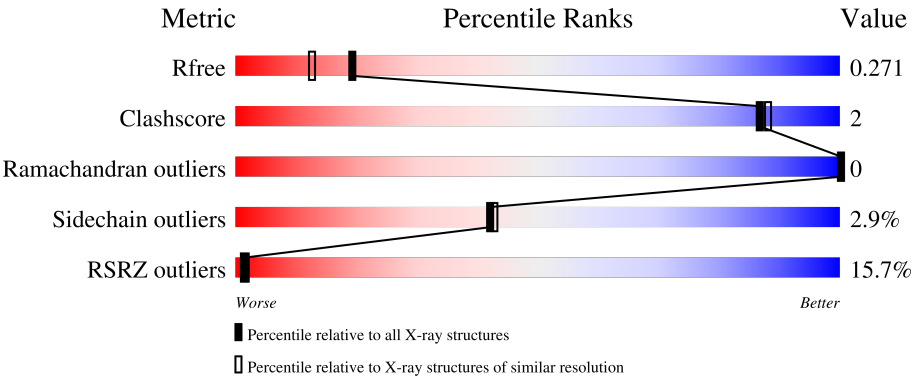
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 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}	130704	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (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 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 <=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	366	16%, 89%, 6% • 5%
2	B	11	91%, 9%
3	C	11	82%, 9%, 9%

2 Entry composition (i)

There are 7 unique types of molecules in this entry. The entry contains 6137 atoms, of which 2885 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 A/G-specific adenine glycosylase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	349	Total	C	H	N	O	S	0	4	0
			5360	1764	2630	463	492	11			

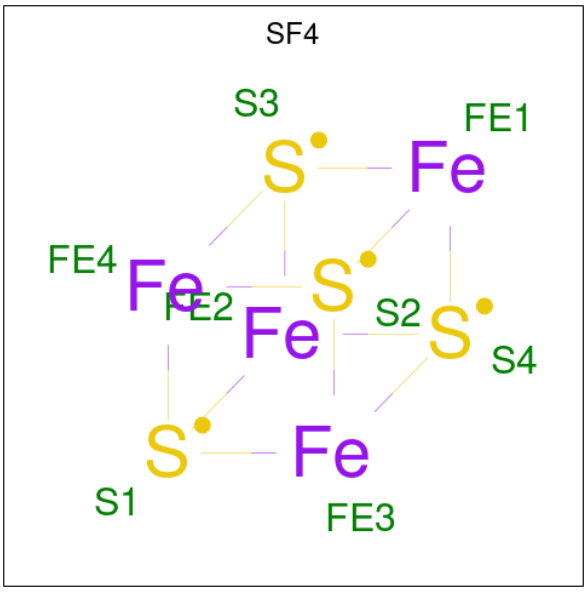
- Molecule 2 is a DNA chain called DNA (5'-D(*AP*AP*GP*AP*CP*GP*TP*GP*GP*AP*C)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	11	Total	C	H	N	O	P	0	1	0
			381	118	136	53	64	10			

- Molecule 3 is a DNA chain called DNA (5'-D(P*GP*TP*CP*CP*AP*(NR1)P*GP*TP*CP*T)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	C	10	Total	C	H	N	O	P	0	0	0
			306	92	113	31	60	10			

- Molecule 4 is IRON/SULFUR CLUSTER (three-letter code: SF4) (formula: Fe_4S_4).

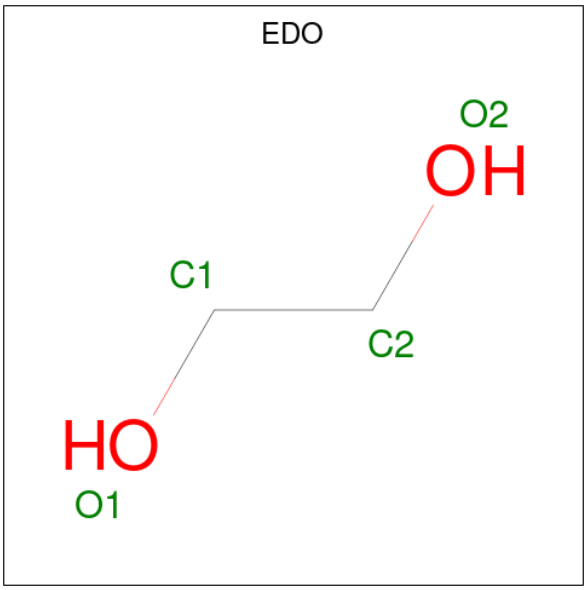


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total Fe S 8 4 4	0	0

- Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total Ca 1 1	0	0

- Molecule 6 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
6	A	1	Total	C	H	O	0	0
			10	2	6	2		

- Molecule 7 is water.

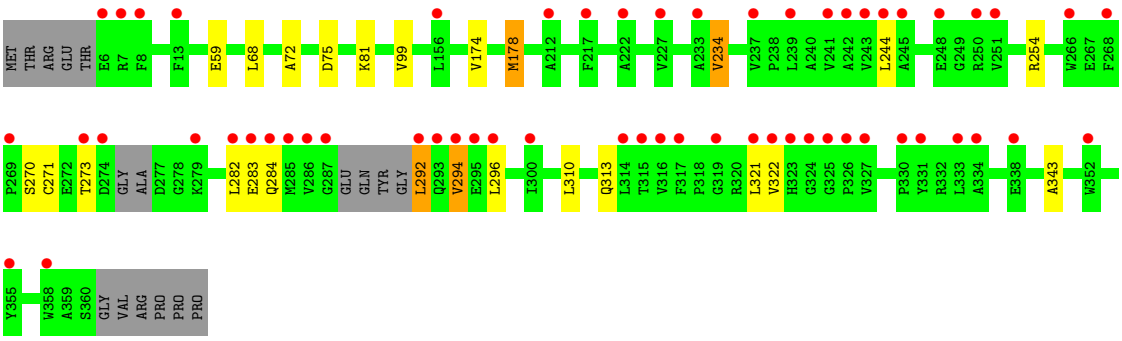
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	56	Total	O	0	0
			56	56		
7	B	7	Total	O	0	1
			8	8		
7	C	7	Total	O	0	0
			7	7		

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: A/G-specific adenine glycosylase

Chain A:



• Molecule 2: DNA (5'-D(*AP*AP*GP*AP*CP*GP*TP*GP*GP*AP*C)-3')

Chain B:

• Molecule 3: DNA (5'-D(P*GP*TP*CP*CP*AP*(NR1)P*GP*TP*CP*T)-3')

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	37.86Å 86.14Å 141.17Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	41.30 – 2.00 41.30 – 2.00	Depositor EDS
% Data completeness (in resolution range)	96.3 (41.30-2.00) 96.4 (41.30-2.00)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.23 (at 2.00Å)	Xtriage
Refinement program	PHENIX 1.15rc2_3428	Depositor
R, R_{free}	0.247 , 0.271 0.247 , 0.271	Depositor DCC
R_{free} test set	1498 reflections (4.84%)	wwPDB-VP
Wilson B-factor (Å^2)	36.1	Xtriage
Anisotropy	0.108	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 49.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.44$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	6137	wwPDB-VP
Average B, all atoms (Å^2)	62.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.64% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: EDO, SF4, CA, NR1

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	0.26	0/2812	0.46	0/3835
2	B	0.56	0/277	0.96	2/427 (0.5%)
3	C	0.51	0/201	1.01	0/305
All	All	0.32	0/3290	0.57	2/4567 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	10	DA	C1'-O4'-C4'	-5.28	104.82	110.10
2	B	10	DA	O4'-C1'-N9	5.19	111.63	108.00

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	2730	2630	2631	13	1
2	B	245	136	136	0	0
3	C	193	113	114	0	1
4	A	8	0	0	0	0
5	A	1	0	0	0	0
6	A	4	6	6	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
7	A	56	0	0	0	0
7	B	8	0	0	0	0
7	C	7	0	0	0	0
All	All	3252	2885	2887	13	1

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 (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)
1:A:244:LEU:HD13	1:A:294:VAL:HG11	1.46	0.97
1:A:283:GLU:HG2	1:A:296:LEU:HD12	1.59	0.84
1:A:68:LEU:HG	1:A:99:VAL:HG13	1.78	0.65
1:A:292:LEU:CD2	1:A:321:LEU:HD11	2.28	0.63
1:A:292:LEU:HD22	1:A:321:LEU:HD11	1.82	0.61
1:A:273:THR:O	1:A:273:THR:HG22	2.06	0.56
1:A:271[B]:CYS:SG	1:A:282:LEU:HD23	2.54	0.48
1:A:81:LYS:HA	1:A:343:ALA:HB2	1.97	0.47
1:A:174:VAL:O	1:A:178:MET:N	2.49	0.46
1:A:254:ARG:NH2	1:A:270:SER:O	2.49	0.46
1:A:68:LEU:CG	1:A:99:VAL:HG13	2.45	0.46
1:A:72:ALA:HB2	1:A:99:VAL:HG12	2.01	0.41
1:A:234:VAL:HG13	1:A:310:LEU:HA	2.03	0.41

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:59[B]:GLU:OE2	3:C:22:DT:O3'[1_455]	1.96	0.24

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	347/366 (95%)	341 (98%)	6 (2%)	0	100 100

There are no Ramachandran outliers to report.

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	276/313 (88%)	268 (97%)	8 (3%)	42 43

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	75	ASP
1	A	178	MET
1	A	234	VAL
1	A	284	GLN
1	A	292	LEU
1	A	294	VAL
1	A	313	GLN
1	A	322	VAL

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)

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	$\#\|Z\| > 2$
3	NR1	C	18	3	6,11,12	0.58	0	4,14,17	0.46	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	NR1	C	18	3	-	0/3/15/16	0/1/1/1

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.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 3 ligands modelled in this entry, 1 is monoatomic - leaving 2 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	SF4	A	401	1	0,12,12	-	-	-		
6	EDO	A	403	-	3,3,3	0.44	0	2,2,2	0.36	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	SF4	A	401	1	-	-	0/6/5/5
6	EDO	A	403	-	-	0/1/1/1	-

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.

No monomer is involved in short contacts.

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, 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	349/366 (95%)	1.27	58 (16%) 1 1	24, 57, 89, 100	0
2	B	11/11 (100%)	0.51	0 100 100	37, 47, 71, 73	0
3	C	9/11 (81%)	0.60	0 100 100	31, 41, 66, 75	0
All	All	369/388 (95%)	1.23	58 (15%) 2 1	24, 57, 89, 100	0

All (58) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	282	LEU	10.8
1	A	296	LEU	9.3
1	A	292	LEU	8.7
1	A	294	VAL	7.2
1	A	319	GLY	7.1
1	A	321	LEU	6.4
1	A	241	VAL	5.9
1	A	243	VAL	5.4
1	A	273	THR	5.3
1	A	325	GLY	4.9
1	A	355	TYR	4.8
1	A	293	GLN	4.6
1	A	244	LEU	4.5
1	A	245	ALA	4.5
1	A	279	LYS	4.4
1	A	286	VAL	4.4
1	A	284	GLN	4.2
1	A	285	MET	4.2
1	A	326	PRO	4.1
1	A	331	TYR	4.0
1	A	333	LEU	3.9
1	A	268	PHE	3.8
1	A	324	GLY	3.7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	317	PHE	3.6
1	A	237	VAL	3.5
1	A	7	ARG	3.5
1	A	283	GLU	3.5
1	A	295	GLU	3.3
1	A	212	ALA	3.3
1	A	8	PHE	3.2
1	A	300	ILE	3.1
1	A	352	TRP	3.1
1	A	358	TRP	3.1
1	A	314	LEU	3.0
1	A	269	PRO	3.0
1	A	327	VAL	2.9
1	A	322	VAL	2.9
1	A	250	ARG	2.8
1	A	323	HIS	2.8
1	A	330	PRO	2.7
1	A	334	ALA	2.7
1	A	266	TRP	2.7
1	A	156	LEU	2.7
1	A	287	GLY	2.6
1	A	217	PHE	2.6
1	A	6	GLU	2.6
1	A	248	GLU	2.6
1	A	239	LEU	2.6
1	A	242	ALA	2.4
1	A	315	THR	2.4
1	A	251	VAL	2.3
1	A	338	GLU	2.3
1	A	13	PHE	2.2
1	A	227	VAL	2.2
1	A	316	VAL	2.2
1	A	222	ALA	2.1
1	A	233	ALA	2.0
1	A	274	ASP	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, 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
3	NR1	C	18	11/12	0.99	0.18	21,28,32,34	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, 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
6	EDO	A	403	4/4	0.93	0.26	34,42,55,66	0
5	CA	A	402	1/1	0.94	0.17	47,47,47,47	0
4	SF4	A	401	8/8	0.96	0.15	40,45,49,50	0

6.5 Other polymers (i)

There are no such residues in this entry.