wwPDB X-ray Structure Validation Summary Report (i)

Aug 27, 2023 – 02:10 PM EDT

PDB ID	:	3I8D
Title	:	The Pairing Geometry of the Hydrophobic Thymine Analog 2,4-Difluorotoluene in Duplex DNA as Analyzed by X-ray Crystallography
Authors	:	Egli, M.; Pallan, P.S.
Deposited on	:	2009-07-09
Resolution	:	1.61 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

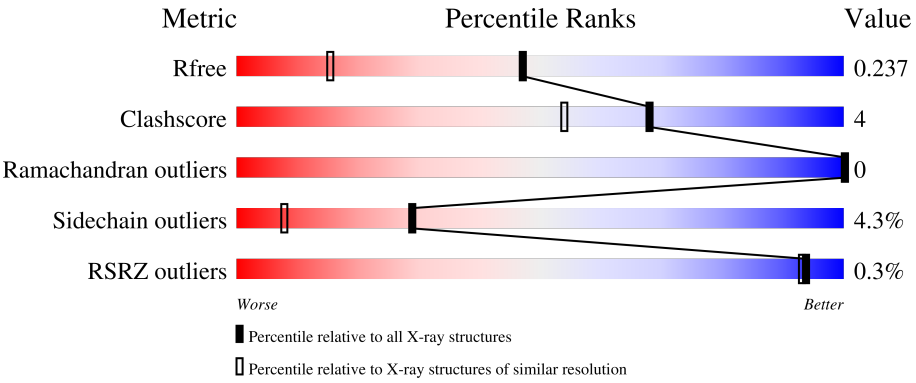
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

1 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 1.61 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	4693 (1.64-1.60)
Clashscore	141614	5002 (1.64-1.60)
Ramachandran outliers	138981	4888 (1.64-1.60)
Sidechain outliers	138945	4887 (1.64-1.60)
RSRZ outliers	127900	4609 (1.64-1.60)

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	132	% 85% 12% .
1	C	132	79% 16% 5%
2	B	12	92% 8%
2	D	12	75% 25%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 2895 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 Ribonuclease H.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	132	Total	C	N	O	S	11	0	0
			1077	689	180	207	1			
1	C	132	Total	C	N	O	S	10	0	0
			1077	689	180	207	1			

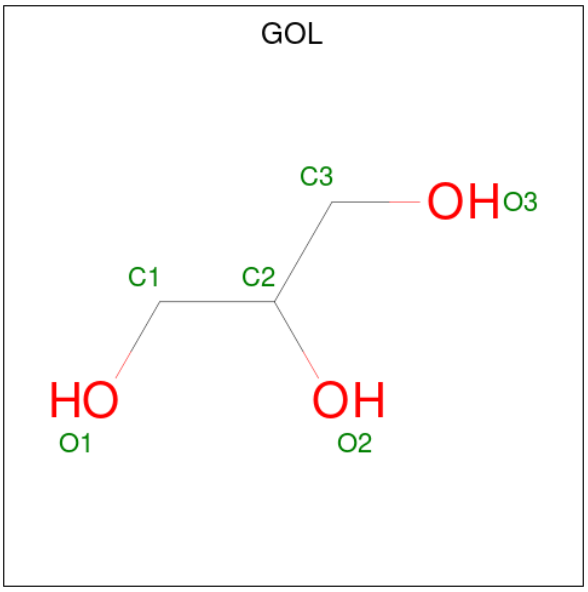
There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	132	ASN	ASP	engineered mutation	UNP Q9KEI9
C	132	ASN	ASP	engineered mutation	UNP Q9KEI9

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	12	Total	C	F	N	O	P	0	0	0
			243	118	2	44	68	11			
2	D	12	Total	C	F	N	O	P	0	0	0
			243	118	2	44	68	11			

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 6 3 3	0	0
3	C	1	Total C O 6 3 3	0	0
3	C	1	Total C O 6 3 3	0	0
3	C	1	Total C O 6 3 3	0	0

- Molecule 4 is water.

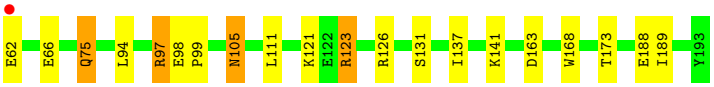
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	78	Total O 78 78	0	0
4	B	40	Total O 40 40	0	0
4	C	85	Total O 85 85	0	0
4	D	28	Total O 28 28	0	0

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.

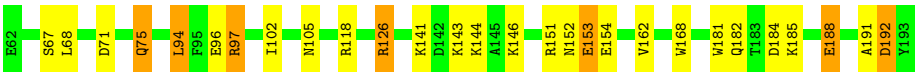
• Molecule 1: Ribonuclease H

Chain A:



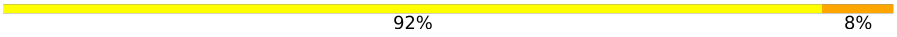
• Molecule 1: Ribonuclease H

Chain C:



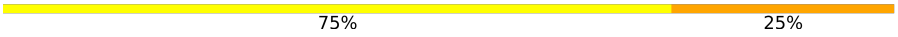
• Molecule 2: 5'-D(*CP*GP*CP*GP*AP*AP*TP*(DFT)P*CP*GP*CP*G)-3'

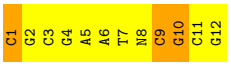
Chain B:



• Molecule 2: 5'-D(*CP*GP*CP*GP*AP*AP*TP*(DFT)P*CP*GP*CP*G)-3'

Chain D:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	96.18Å 66.71Å 77.57Å 90.00° 121.04° 90.00°	Depositor
Resolution (Å)	49.39 – 1.61 49.38 – 1.61	Depositor EDS
% Data completeness (in resolution range)	100.0 (49.39-1.61) 97.4 (49.38-1.61)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.21 (at 1.61Å)	Xtriage
Refinement program	REFMAC 5.5.0093	Depositor
R, R_{free}	0.197 , 0.238 0.195 , 0.237	Depositor DCC
R_{free} test set	2690 reflections (5.09%)	wwPDB-VP
Wilson B-factor (Å^2)	23.8	Xtriage
Anisotropy	0.260	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 45.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.52$, $< L^2 > = 0.36$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	2895	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 46.62 % 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 1.1158e-04. 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 (i)

5.1 Standard geometry (i)

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

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	1.62	6/1102 (0.5%)	1.13	5/1493 (0.3%)
1	C	1.70	12/1102 (1.1%)	1.28	13/1493 (0.9%)
2	B	2.52	14/249 (5.6%)	3.29	48/380 (12.6%)
2	D	2.63	14/249 (5.6%)	3.03	50/380 (13.2%)
All	All	1.86	46/2702 (1.7%)	1.79	116/3746 (3.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	C	0	1

The worst 5 of 46 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	75	GLN	CA-CB	-28.67	0.90	1.53
1	C	75	GLN	CA-CB	-27.48	0.93	1.53
1	A	105	ASN	CB-CG	13.22	1.81	1.51
1	C	188	GLU	CB-CG	-10.58	1.32	1.52
2	D	4	DG	P-O5'	-9.48	1.50	1.59

The worst 5 of 116 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	9	DC	O4'-C4'-C3'	-13.47	97.92	106.00
2	B	5	DA	N7-C8-N9	11.72	119.66	113.80
1	C	97	ARG	NE-CZ-NH1	-11.00	114.80	120.30
2	B	10	DG	N9-C4-C5	-9.74	101.50	105.40

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	3	DC	O4'-C1'-N1	-9.63	101.26	108.00

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	126	ARG	Sidechain

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	1077	0	1068	8	0
1	C	1077	0	1068	8	0
2	B	243	0	136	1	0
2	D	243	0	136	2	0
3	A	6	0	8	1	0
3	C	18	0	24	0	0
4	A	78	0	0	2	0
4	B	40	0	0	1	0
4	C	85	0	0	1	0
4	D	28	0	0	1	0
All	All	2895	0	2440	19	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 4.

The worst 5 of 19 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:105:ASN:CG	1:A:105:ASN:CB	1.81	1.47
1:C:97:ARG:NH1	4:C:205:HOH:O	2.06	0.87
1:A:97:ARG:NH1	4:A:223:HOH:O	2.06	0.86
2:D:1:DC:O5'	4:D:221:HOH:O	2.02	0.77
1:A:137:ILE:O	1:A:141:LYS:HG2	1.95	0.66

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	130/132 (98%)	127 (98%)	3 (2%)	0	100	100
1	C	130/132 (98%)	129 (99%)	1 (1%)	0	100	100
All	All	260/264 (98%)	256 (98%)	4 (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	117/117 (100%)	111 (95%)	6 (5%)	24	5
1	C	117/117 (100%)	113 (97%)	4 (3%)	37	12
All	All	234/234 (100%)	224 (96%)	10 (4%)	29	8

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	94	LEU
1	C	153	GLU
1	C	188	GLU
1	A	94	LEU
1	A	123	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	134	GLN
1	A	170	ASN
1	C	170	ASN
1	C	182	GLN

5.3.3 RNA

There are no RNA molecules 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	DFT	B	8	2	18,21,22	1.92	5 (27%)	26,30,33	2.51	7 (26%)
2	DFT	D	8	2	18,21,22	2.68	4 (22%)	26,30,33	3.41	6 (23%)

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	DFT	B	8	2	-	4/7/21/22	0/2/2/2
2	DFT	D	8	2	-	4/7/21/22	0/2/2/2

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	8	DFT	C6-C1	-6.90	1.28	1.39

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	8	DFT	C6-C5	-5.54	1.31	1.39
2	D	8	DFT	C3-C4	-5.21	1.28	1.37
2	B	8	DFT	C6-C1	-3.97	1.33	1.39
2	B	8	DFT	C3-C4	-3.93	1.30	1.37

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	8	DFT	C3-C4-C5	-11.35	117.12	124.39
2	D	8	DFT	C6-C5-C4	11.27	123.24	116.06
2	B	8	DFT	C3-C4-C5	-8.38	119.02	124.39
2	B	8	DFT	C6-C5-C4	5.48	119.55	116.06
2	B	8	DFT	O4'-C1'-C1	4.40	114.17	109.74

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	8	DFT	C6-C1-C1'-O4'
2	B	8	DFT	C6-C1-C1'-C2'
2	D	8	DFT	C6-C1-C1'-O4'
2	D	8	DFT	C6-C1-C1'-C2'
2	B	8	DFT	C2-C1-C1'-O4'

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)

4 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	$\#\|Z\| > 2$
3	GOL	C	604	-	5,5,5	0.35	0	5,5,5	1.18	0
3	GOL	C	603	-	5,5,5	0.37	0	5,5,5	0.55	0
3	GOL	C	602	-	5,5,5	1.49	1 (20%)	5,5,5	1.62	2 (40%)
3	GOL	A	601	-	5,5,5	1.89	2 (40%)	5,5,5	1.87	1 (20%)

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	GOL	C	604	-	-	3/4/4/4	-
3	GOL	C	603	-	-	2/4/4/4	-
3	GOL	C	602	-	-	0/4/4/4	-
3	GOL	A	601	-	-	1/4/4/4	-

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	601	GOL	O3-C3	2.99	1.55	1.42
3	C	602	GOL	O1-C1	2.38	1.52	1.42
3	A	601	GOL	C3-C2	2.13	1.60	1.51

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	601	GOL	O2-C2-C3	3.43	124.21	109.12
3	C	602	GOL	O2-C2-C1	2.57	120.44	109.12
3	C	602	GOL	O3-C3-C2	-2.37	98.83	110.20

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	601	GOL	C1-C2-C3-O3
3	C	604	GOL	O1-C1-C2-C3
3	C	603	GOL	C1-C2-C3-O3
3	C	604	GOL	O1-C1-C2-O2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
3	C	604	GOL	C1-C2-C3-O3

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	601	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, 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	132/132 (100%)	-0.31	1 (0%) 86 86	11, 15, 24, 36	2 (1%)
1	C	132/132 (100%)	-0.34	0 100 100	12, 16, 24, 32	2 (1%)
2	B	11/12 (91%)	-0.68	0 100 100	10, 12, 16, 16	0
2	D	11/12 (91%)	-0.48	0 100 100	7, 12, 16, 19	0
All	All	286/288 (99%)	-0.35	1 (0%) 94 93	7, 15, 24, 36	4 (1%)

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	62	GLU	2.9

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. 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	DFT	B	8	20/21	0.95	0.08	12,17,23,26	0
2	DFT	D	8	20/21	0.97	0.07	12,16,21,23	0

6.3 Carbohydrates

There are no monosaccharides 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
3	GOL	A	601	6/6	0.53	0.18	26,36,44,45	0
3	GOL	C	604	6/6	0.70	0.17	55,59,62,63	0
3	GOL	C	602	6/6	0.86	0.15	25,28,33,41	0
3	GOL	C	603	6/6	0.90	0.11	38,40,43,46	0

6.5 Other polymers

There are no such residues in this entry.