



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 05:55 AM GMT

PDB ID : 2VA2
Title : Complex structure of Sulfolobus solfataricus DPO4 and DNA duplex containing a hydrophobic thymine isostere 2,4-difluorotoluene nucleotide in the template strand
Authors : Irimia, A.; Pallan, P.S.; Egli, M.
Deposited on : 2007-08-28
Resolution : 2.80 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
<http://wwpdb.org/validation/2016/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 : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk26865

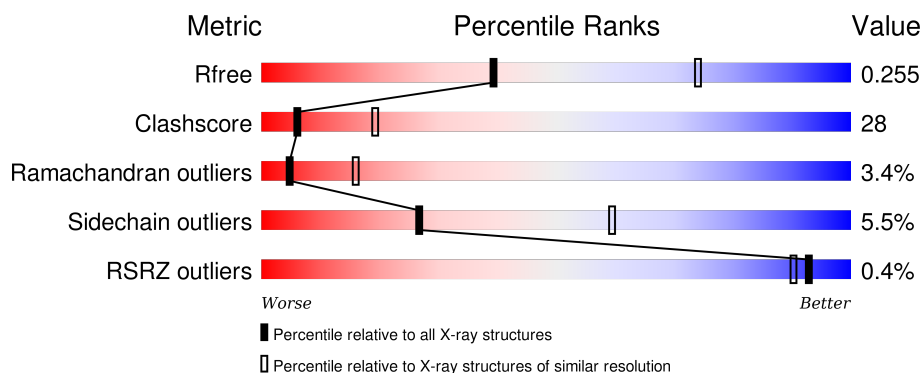
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.80 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}	91344	2393 (2.80-2.80)
Clashscore	102246	2827 (2.80-2.80)
Ramachandran outliers	100387	2782 (2.80-2.80)
Sidechain outliers	100360	2784 (2.80-2.80)
RSRZ outliers	91569	2404 (2.80-2.80)

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. 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	358	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red 1%, green 56%, yellow 35%, orange 6%, red 1%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> 56% 35% • • • </div> </div>
1	B	358	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, green 50%, yellow 40%, orange 6%, red 4%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> 50% 40% 6% • • </div> </div>
2	C	13	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, green 54%, yellow 46%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> 54% 46% </div> </div>
2	E	13	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, green 31%, yellow 69%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> 31% 69% </div> </div>
3	D	18	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, green 61%, yellow 22%, orange 6%, grey 11%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> 61% 22% 6% 11% </div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	18	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
5	DCT	A	1345	-	-	X	-
5	DCT	E	1014	-	-	X	-

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 7081 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 DNA POLYMERASE IV.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	343	Total	C	N	O	S	0	2	1
			2772	1775	479	511	7			
1	B	344	Total	C	N	O	S	0	5	1
			2810	1797	486	520	7			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	13	Total	C	N	O	P	0	0	0
			274	129	60	73	12			
2	E	13	Total	C	N	O	P	0	0	0
			274	129	60	73	12			

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

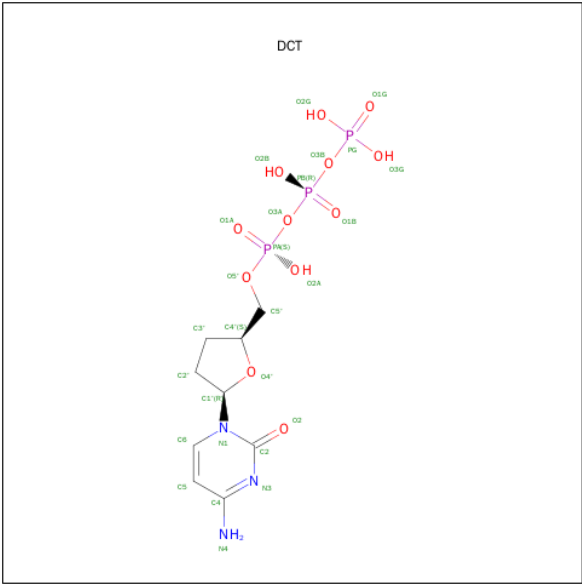
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	D	16	Total	C	F	N	O	P	0	0	0
			319	154	2	50	97	16			
3	F	18	Total	C	F	N	O	P	0	0	0
			355	174	2	54	108	17			

- Molecule 4 is CALCIUM ION (three-letter code: CA) (formula: Ca).

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

- Molecule 5 is 2',3'-DIDEOXYCYTIDINE 5'-TRIPHOSPHATE (three-letter code: DCT)

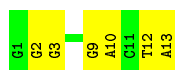
(formula: C₉H₁₆N₃O₁₂P₃).



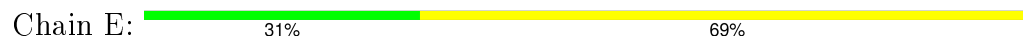
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	A	1	Total	C	N	O	P	0	0
			27	9	3	12	3		
5	E	1	Total	C	N	O	P	0	0
			27	9	3	12	3		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	90	Total	O	0	0
			90	90		
6	B	75	Total	O	0	0
			75	75		
6	C	17	Total	O	0	0
			17	17		
6	D	11	Total	O	0	0
			11	11		
6	E	11	Total	O	0	0
			11	11		
6	F	13	Total	O	0	0
			13	13		



- Molecule 2: 5'-D(*GP*GP*GP*GP*GP*AP*AP*GP*GP*AP *CP*TP*A)-3'



- Molecule 3: 5'-D(*TP*TP*CP*AP*GP*DFTP*AP*GP*TP*CP*CP *TP*TP*CP*CP*CP*CP*CP*CP)-3'



- Molecule 3: 5'-D(*TP*TP*CP*AP*GP*DFTP*AP*GP*TP*CP*CP *TP*TP*CP*CP*CP*CP*CP*CP)-3'



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	52.14Å 101.88Å 111.06Å 90.00° 94.92° 90.00°	Depositor
Resolution (Å)	22.57 – 2.80 22.56 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.7 (22.57-2.80) 99.8 (22.56-2.80)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.99 (at 2.80Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.211 , 0.250 0.209 , 0.255	Depositor DCC
R_{free} test set	1383 reflections (5.10%)	DCC
Wilson B-factor (Å ²)	53.3	Xtriage
Anisotropy	0.524	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 56.4	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 28484 reflections	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	7081	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 4.13% 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, 0.375 respectively for untwinned datasets, and 0.333, 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, CA, DCT

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	0.46	0/2811	0.60	0/3775
1	B	0.40	0/2850	0.56	0/3827
2	C	0.50	0/310	0.69	0/479
2	E	0.37	0/310	0.59	0/479
3	D	0.73	1/331 (0.3%)	0.74	0/502
3	F	0.51	0/371	0.68	0/566
All	All	0.46	1/6983 (0.0%)	0.60	0/9628

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	A	0	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	3	DC	OP3-P	-7.16	1.52	1.61

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	120	GLU	Mainchain

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	2772	0	2910	164	0
1	B	2810	0	2936	166	0
2	C	274	0	146	15	0
2	E	274	0	146	13	0
3	D	319	0	181	9	0
3	F	355	0	206	10	0
4	A	3	0	0	0	0
4	B	3	0	0	0	0
5	A	27	0	12	13	0
5	E	27	0	12	9	0
6	A	90	0	0	21	0
6	B	75	0	0	21	0
6	C	17	0	0	2	0
6	D	11	0	0	0	0
6	E	11	0	0	2	0
6	F	13	0	0	3	0
All	All	7081	0	6549	380	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 28.

The worst 5 of 380 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:B:52:LYS:HG2	6:B:2010:HOH:O	1.38	1.21
5:A:1345:DCT:H5''	2:C:13:DA:H2''	1.29	1.14
3:D:3:DC:H2''	3:D:4:DA:H5'	1.34	1.10
1:A:102:ALA:CB	1:A:240[A]:ARG:HH21	1.68	1.05
1:B:256:ARG:HH21	1:B:327:GLU:HA	1.15	1.03

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

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	343/358 (96%)	304 (89%)	28 (8%)	11 (3%)	5	17
1	B	347/358 (97%)	285 (82%)	49 (14%)	13 (4%)	4	14
All	All	690/716 (96%)	589 (85%)	77 (11%)	24 (4%)	5	15

5 of 24 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	118	TYR
1	A	163	ILE
1	A	235	GLU
1	B	38	GLU
1	B	163	ILE

5.3.2 Protein sidechains

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	303/315 (96%)	288 (95%)	15 (5%)	30	64
1	B	307/315 (98%)	286 (93%)	21 (7%)	20	49
All	All	610/630 (97%)	574 (94%)	36 (6%)	27	57

5 of 36 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	31	CYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	114	LYS
1	B	273	TYR
1	B	82[B]	GLN
1	B	115	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	14	GLN
1	B	234	ASN
1	B	83	GLN
1	A	304	HIS
1	B	20	ASN

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	$\# Z > 2$
3	DFT	D	6	3	17,21,22	0.72	0	23,30,33	2.33	7 (30%)
3	DFT	F	6	3	17,21,22	0.73	0	23,30,33	2.29	8 (34%)

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	DFT	D	6	3	-	0/7/21/22	0/2/2/2
3	DFT	F	6	3	-	0/7/21/22	0/2/2/2

There are no bond length outliers.

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	D	6	DFT	C4'-O4'-C1'	-5.14	100.34	109.14
3	D	6	DFT	C3-C2-C1	-3.32	120.24	123.78
3	F	6	DFT	C3-C2-C1	-3.20	120.37	123.78
3	F	6	DFT	C5M-C5-C4	-3.05	119.93	121.68
3	F	6	DFT	C3'-C2'-C1'	-3.05	99.79	102.69

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	F	6	DFT	1	0

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 8 ligands modelled in this entry, 6 are 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
5	DCT	A	1345	4	20,28,28	1.44	3 (15%)	29,43,43	2.00	5 (17%)
5	DCT	E	1014	4	20,28,28	1.47	3 (15%)	29,43,43	1.72	5 (17%)

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	DCT	A	1345	4	-	0/18/31/31	0/2/2/2
5	DCT	E	1014	4	-	0/18/31/31	0/2/2/2

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	A	1345	DCT	C4-N3	2.47	1.40	1.35
5	E	1014	DCT	C4-N3	2.65	1.40	1.35
5	A	1345	DCT	PG-O1G	3.07	1.61	1.51
5	E	1014	DCT	PG-O1G	3.11	1.61	1.51
5	A	1345	DCT	C6-N1	3.18	1.40	1.35

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	A	1345	DCT	O4'-C1'-C2'	-5.68	100.52	106.67
5	A	1345	DCT	C3'-C2'-C1'	-4.26	97.95	102.71
5	A	1345	DCT	PB-O3B-PG	-4.20	118.58	132.67
5	E	1014	DCT	PB-O3B-PG	-3.97	119.35	132.67
5	E	1014	DCT	PB-O3A-PA	-3.85	121.91	132.73

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 22 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	1345	DCT	13	0
5	E	1014	DCT	9	0

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, 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	343/358 (95%)	-0.20	2 (0%) 90 86	22, 43, 70, 80	12 (3%)
1	B	344/358 (96%)	-0.09	0 100 100	29, 53, 77, 90	15 (4%)
2	C	13/13 (100%)	0.20	0 100 100	37, 45, 56, 59	0
2	E	13/13 (100%)	0.46	0 100 100	52, 74, 88, 95	0
3	D	15/18 (83%)	-0.19	0 100 100	30, 44, 78, 92	0
3	F	17/18 (94%)	0.28	1 (5%) 26 16	41, 68, 90, 111	0
All	All	745/778 (95%)	-0.12	3 (0%) 93 90	22, 48, 76, 111	27 (3%)

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	F	1	DT	3.8
1	A	343	ALA	3.7
1	A	327	GLU	2.5

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å ²)	Q<0.9
3	DFT	D	6	20/21	0.96	0.16	-	35,41,45,47	0
3	DFT	F	6	20/21	0.97	0.14	-	34,42,50,52	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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(\AA^2)	Q<0.9
5	DCT	E	1014	27/27	0.90	0.22	0.71	62,79,94,96	0
5	DCT	A	1345	27/27	0.92	0.19	-0.08	49,64,74,75	0
4	CA	A	1343	1/1	0.97	0.14	-1.33	41,41,41,41	0
4	CA	B	1343	1/1	0.98	0.11	-2.90	42,42,42,42	0
4	CA	A	1344	1/1	0.86	0.07	-2.97	87,87,87,87	0
4	CA	B	1344	1/1	0.95	0.04	-3.44	83,83,83,83	0
4	CA	A	1346	1/1	0.99	0.06	-	76,76,76,76	0
4	CA	B	3002	1/1	0.95	0.07	-	92,92,92,92	0

6.5 Other polymers [i](#)

There are no such residues in this entry.