Full wwPDB X-ray Structure Validation Report (i)

Feb 1, 2016 – 03:54 PM GMT

PDB ID	:	4DTO
Title	:	RB69 DNA Polymerase Ternary Complex with dCTP Opposite an Abasic Site and ddA/dT as the Penultimate Base-pair
Authors	:	Xia, S.; Wang, J.; Konigsberg, W.H.
Deposited on	:	2012-02-21
Resolution	:	2.05 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

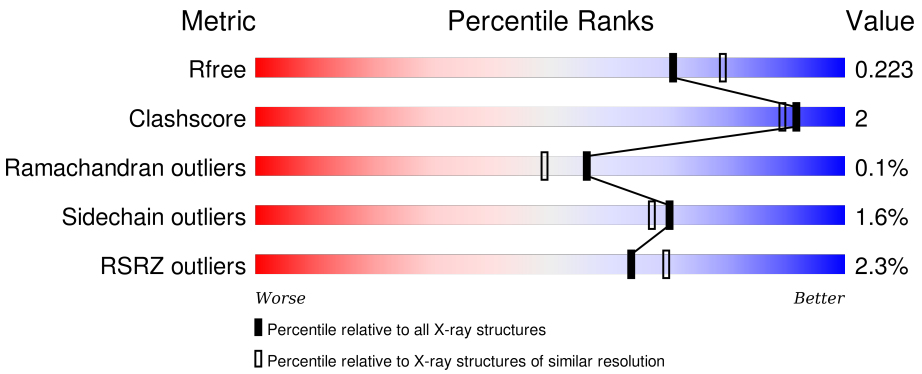
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.05 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	1192 (2.04-2.04)
Clashscore	102246	1269 (2.04-2.04)
Ramachandran outliers	100387	1258 (2.04-2.04)
Sidechain outliers	100360	1258 (2.04-2.04)
RSRZ outliers	91569	1194 (2.04-2.04)

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 <=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	903	2% 94% 6%
2	T	17	88% 6% 6%
3	P	13	8% 62% 15% 23%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 8580 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.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	903	7375	4735	1230	1377	33	0	2	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	222	ALA	ASP	CONFLICT	UNP Q38087
A	327	ALA	ASP	CONFLICT	UNP Q38087
A	561	ALA	LEU	CONFLICT	UNP Q38087
A	565	GLY	SER	CONFLICT	UNP Q38087
A	567	ALA	TYR	CONFLICT	UNP Q38087

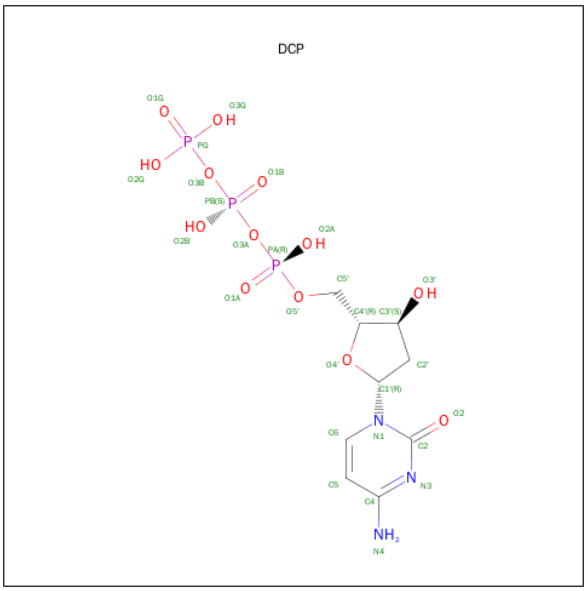
- Molecule 2 is a DNA chain called DNA template.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	T	17	338	160	61	100	17	0	0	0

- Molecule 3 is a DNA chain called DNA primer.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	P	13	264	127	50	75	12	0	0	0

- Molecule 4 is 2'-DEOXYCYTIDINE-5'-TRIPHOSPHATE (three-letter code: DCP) (formula: $C_9H_{16}N_3O_{13}P_3$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
4	A	1	28	9	3	13	3	0	0

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

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

- Molecule 6 is water.

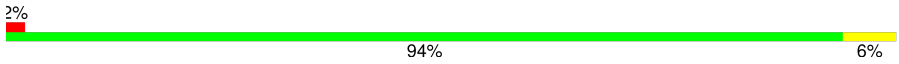
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
6	A	512	520	520	0	8
			Total	O		
6	T	36	36	36	0	0
			Total	O		
6	P	14	14	14	0	0

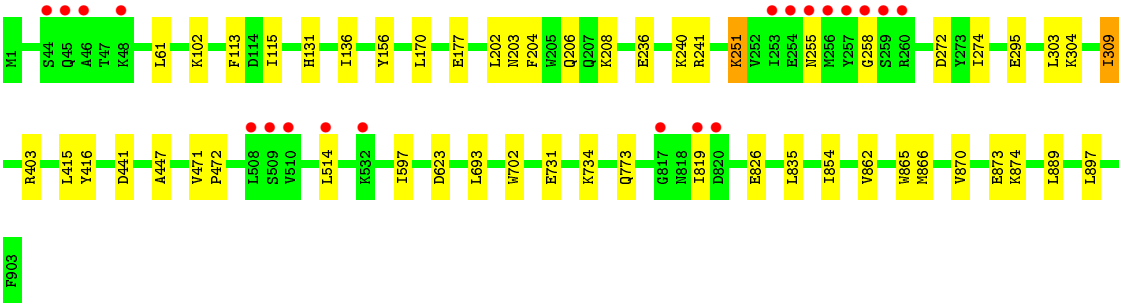
3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: DNA polymerase

Chain A:



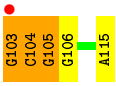


- Molecule 2: DNA template

Chain T:

- Molecule 3: DNA primer

Chain P:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	78.24Å 119.26Å 130.92Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.05 47.42 – 2.03	Depositor EDS
% Data completeness (in resolution range)	98.2 (50.00-2.05) 98.3 (47.42-2.03)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.30 (at 2.03Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, R_{free}	0.185 , 0.225 0.184 , 0.223	Depositor DCC
R_{free} test set	3892 reflections (5.31%)	DCC
Wilson B-factor (Å^2)	31.0	Xtriage
Anisotropy	0.077	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 46.7	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Outliers	0 of 78596 reflections	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	8580	wwPDB-VP
Average B, all atoms (Å^2)	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.22% 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, 0.375 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: CA, 3DR, DCP

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	0.45	2/7561 (0.0%)	0.51	0/10216
2	T	0.30	0/365	0.83	1/558 (0.2%)
3	P	0.32	0/296	1.00	4/455 (0.9%)
All	All	0.44	2/8222 (0.0%)	0.56	5/11229 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	702	TRP	CD2-CE2	5.27	1.47	1.41
1	A	865	TRP	CD2-CE2	5.01	1.47	1.41

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	T	2	DC	P-O3'-C3'	8.01	129.31	119.70
3	P	105	DG	P-O3'-C3'	6.14	127.06	119.70
3	P	103	DG	P-O3'-C3'	5.37	126.15	119.70
3	P	104	DC	P-O3'-C3'	5.21	125.95	119.70
3	P	115	DA	C1'-O4'-C4'	-5.15	104.95	110.10

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	7375	0	7274	24	0
2	T	338	0	188	3	0
3	P	264	0	146	6	0
4	A	28	0	12	1	0
5	A	5	0	0	0	0
6	A	520	0	0	4	0
6	P	14	0	0	0	0
6	T	36	0	0	0	0
All	All	8580	0	7620	33	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 2.

All (33) 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:170:LEU:HA	1:A:177:GLU:HG3	1.76	0.66
1:A:773:GLN:HG2	6:A:1586:HOH:O	2.01	0.59
3:P:105:DG:H2"	3:P:106:DG:C8	2.40	0.57
1:A:862:VAL:O	1:A:866:MET:HG3	2.05	0.57
3:P:105:DG:C2'	3:P:106:DG:C8	2.90	0.54
1:A:441:ASP:HB3	1:A:447:ALA:HB2	1.93	0.51
1:A:403:ARG:NH2	1:A:889:LEU:HD13	2.26	0.50
3:P:105:DG:H2"	3:P:106:DG:O5'	2.13	0.49
1:A:203:ASN:ND2	1:A:241:ARG:HH22	2.10	0.49
1:A:131:HIS:HD2	1:A:156:TYR:OH	1.96	0.49
3:P:104:DC:H2"	3:P:105:DG:C8	2.47	0.48
1:A:854:ILE:HD13	1:A:862:VAL:HG21	1.95	0.48
1:A:202:LEU:O	1:A:206:GLN:HG2	2.13	0.47
3:P:103:DG:H2"	3:P:104:DC:O5'	2.15	0.46
3:P:105:DG:H2"	3:P:106:DG:H8	1.78	0.46
1:A:773:GLN:HG3	6:A:1587:HOH:O	2.16	0.45
1:A:255:ASN:HD21	1:A:826:GLU:HG2	1.82	0.44
2:T:3:DA:H2'	2:T:3:DA:N3	2.32	0.44
1:A:416:TYR:CG	4:A:1001:DCP:H2'2	2.52	0.44
2:T:2:DC:H2"	2:T:3:DA:OP1	2.18	0.44
1:A:236:GLU:HG2	1:A:240:LYS:HE2	2.00	0.43
1:A:731:GLU:HA	1:A:734:LYS:HD3	1.99	0.43
1:A:415:LEU:HD22	1:A:623:ASP:HB3	2.00	0.43
1:A:272:ASP:OD1	1:A:274:ILE:HG22	2.19	0.42
1:A:115:ILE:HG22	1:A:136:ILE:HG12	2.01	0.42
1:A:870:VAL:HG13	1:A:874:LYS:HD3	2.02	0.42

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:309:ILE:O	1:A:309:ILE:HD13	2.20	0.41
1:A:204:PHE:CE1	1:A:208:LYS:HD2	2.56	0.41
1:A:597:ILE:HD12	1:A:597:ILE:HA	1.96	0.41
1:A:873:GLU:HG3	6:A:1420:HOH:O	2.20	0.41
6:A:1407:HOH:O	2:T:3:DA:H5'	2.20	0.40
1:A:471:VAL:HB	1:A:472:PRO:HD3	2.02	0.40
1:A:251:LYS:NZ	1:A:251:LYS:HB2	2.36	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	903/903 (100%)	885 (98%)	17 (2%)	1 (0%)	56 49

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	258	GLY

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	799/797 (100%)	786 (98%)	13 (2%)	70 67

All (13) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	61	LEU
1	A	102	LYS
1	A	113	PHE
1	A	251	LYS
1	A	295	GLU
1	A	303	LEU
1	A	304	LYS
1	A	309	ILE
1	A	514	LEU
1	A	693	LEU
1	A	819	ILE
1	A	835	LEU
1	A	897	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (9) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	45	GLN
1	A	131	HIS
1	A	203	ASN
1	A	206	GLN
1	A	255	ASN
1	A	333	GLN
1	A	339	GLN
1	A	761	GLN
1	A	823	GLN

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	$\#\|Z\| > 2$
2	3DR	T	4	2	7,11,12	0.47	0	8,14,17	0.85	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	3DR	T	4	2	-	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 carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 6 ligands modelled in this entry, 5 are monoatomic - leaving 1 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	DCP	A	1001	5	21,29,29	0.58	0	33,45,45	1.33	4 (12%)

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	DCP	A	1001	5	-	0/18/34/34	0/2/2/2

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	1001	DCP	C2'-C1'-N1	-3.20	106.37	114.16
4	A	1001	DCP	O2B-PB-O3A	2.35	115.75	105.09
4	A	1001	DCP	O4'-C1'-N1	2.43	111.92	107.72
4	A	1001	DCP	C2-N3-C4	3.47	120.51	115.61

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
4	A	1001	DCP	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 (i)

6.1 Protein, DNA and RNA chains (i)

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	903/903 (100%)	-0.11	20 (2%) 65 71	17, 30, 56, 148	0
2	T	16/17 (94%)	0.14	0 100 100	23, 41, 79, 102	0
3	P	13/13 (100%)	0.50	1 (7%) 16 19	23, 44, 84, 92	0
All	All	932/933 (99%)	-0.10	21 (2%) 64 70	17, 31, 59, 148	0

All (21) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	257	TYR	12.1
1	A	258	GLY	10.7
1	A	255	ASN	9.9
1	A	256	MET	9.7
1	A	254	GLU	6.6
1	A	259	SER	5.8
1	A	819	ILE	5.7
1	A	510	VAL	4.2
1	A	260	ARG	3.7
1	A	44	SER	3.7
1	A	817	GLY	3.6
1	A	508	LEU	3.5
1	A	46	ALA	3.3
1	A	45	GLN	3.0
1	A	253	ILE	2.9
3	P	103	DG	2.3
1	A	532	LYS	2.3
1	A	514	LEU	2.1
1	A	48	LYS	2.1
1	A	820	ASP	2.1
1	A	509	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. 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
2	3DR	T	4	11/12	0.98	0.15	-	26,30,40,40	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(Å^2)	Q<0.9
4	DCP	A	1001	28/28	0.98	0.14	0.43	21,25,31,39	0
5	CA	A	1005	1/1	0.83	0.11	-1.65	69,69,69,69	0
5	CA	A	1002	1/1	0.99	0.10	-2.06	27,27,27,27	0
5	CA	A	1004	1/1	0.96	0.05	-2.51	54,54,54,54	0
5	CA	A	1003	1/1	0.95	0.09	-	59,59,59,59	0
5	CA	A	1006	1/1	0.97	0.07	-	43,43,43,43	0

6.5 Other polymers (i)

There are no such residues in this entry.