



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 03:53 PM GMT

PDB ID : 4DQP
Title : Ternary complex of Bacillus DNA Polymerase I Large Fragment, DNA duplex,
and ddCTP (paired with dG of template)
Authors : Wang, W.; Beese, L.S.
Deposited on : 2012-02-16
Resolution : 1.74 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

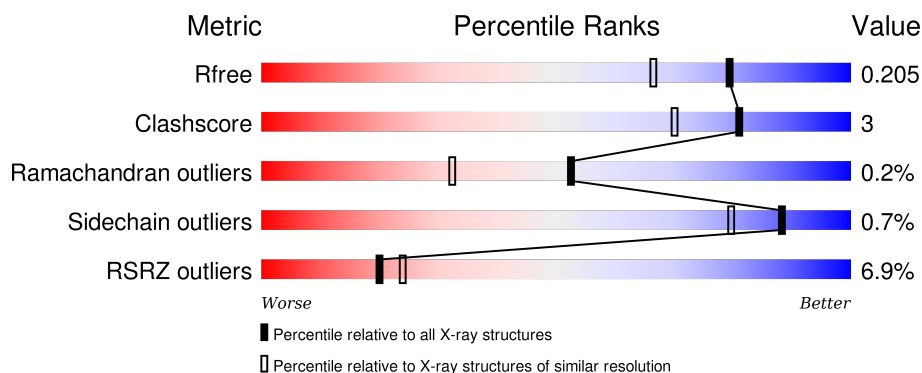
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.74 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	2417 (1.76-1.72)
Clashscore	102246	2570 (1.76-1.72)
Ramachandran outliers	100387	2544 (1.76-1.72)
Sidechain outliers	100360	2544 (1.76-1.72)
RSRZ outliers	91569	2420 (1.76-1.72)

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	592	<div> <div>13%</div> <div>89%</div> <div>8%</div> <div>•</div> </div>
1	D	592	<div> <div>%</div> <div>90%</div> <div>8%</div> <div>•</div> </div>
2	B	9	<div> <div>56%</div> <div>33%</div> <div>11%</div> </div>
2	E	9	<div> <div>44%</div> <div>56%</div> </div>
3	C	13	<div> <div>8%</div> <div>62%</div> <div>31%</div> <div>8%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	13	<div><div></div><div></div><div></div><div></div></div> <div><div>8%</div><div>54%</div><div>38%</div><div>8%</div></div>

2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 21755 atoms, of which 10084 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
1	A	576	Total	C	H	N	O	S	0	17	0
			9444	2979	4762	810	876	17			
1	D	579	Total	C	H	N	O	S	0	26	0
			9574	3014	4830	824	889	17			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	598	ALA	ASP	ENGINEERED MUTATION	UNP Q5KWC1
A	823	HIS	ARG	SEE REMARK 999	UNP Q5KWC1
D	598	ALA	ASP	ENGINEERED MUTATION	UNP Q5KWC1
D	823	HIS	ARG	SEE REMARK 999	UNP Q5KWC1

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

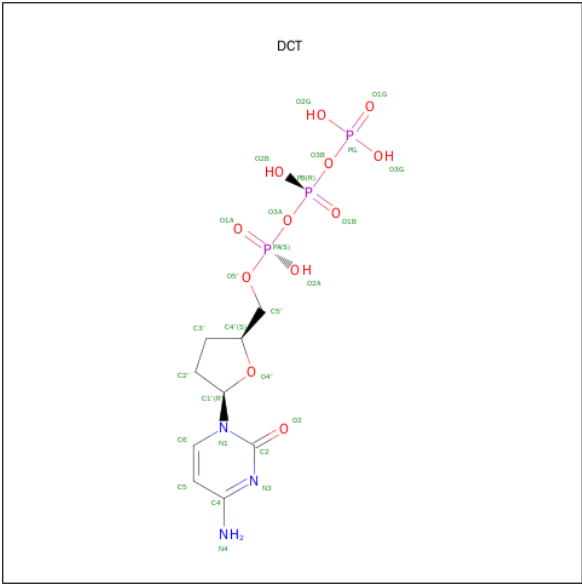
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	9	Total	C	H	N	O	P	0	0	0
			274	85	100	29	52	8			
2	E	9	Total	C	H	N	O	P	0	0	0
			274	85	100	29	52	8			

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	C	12	Total	C	H	N	O	P	0	0	0
			385	119	134	52	69	11			
3	F	12	Total	C	H	N	O	P	0	0	0
			385	119	134	52	69	11			

- Molecule 4 is 2',3'-DIDEOXYCYTIDINE 5'-TRIPHOSPHATE (three-letter code: DCT)

(formula: C₉H₁₆N₃O₁₂P₃).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
4	A	1	Total	C	H	N	O	P	0	0
			35	9	8	3	12	3		
4	A	1	Total	C	H	N	O	P	0	0
			35	9	8	3	12	3		
4	D	1	Total	C	H	N	O	P	0	0
			35	9	8	3	12	3		

- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	1	Total	Mg	0	0
			1	1		
5	D	1	Total	Mg	0	0
			1	1		

- Molecule 6 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	D	1	Total	O	S	0	0
			5	4	1		
6	D	1	Total	O	S	0	0
			5	4	1		

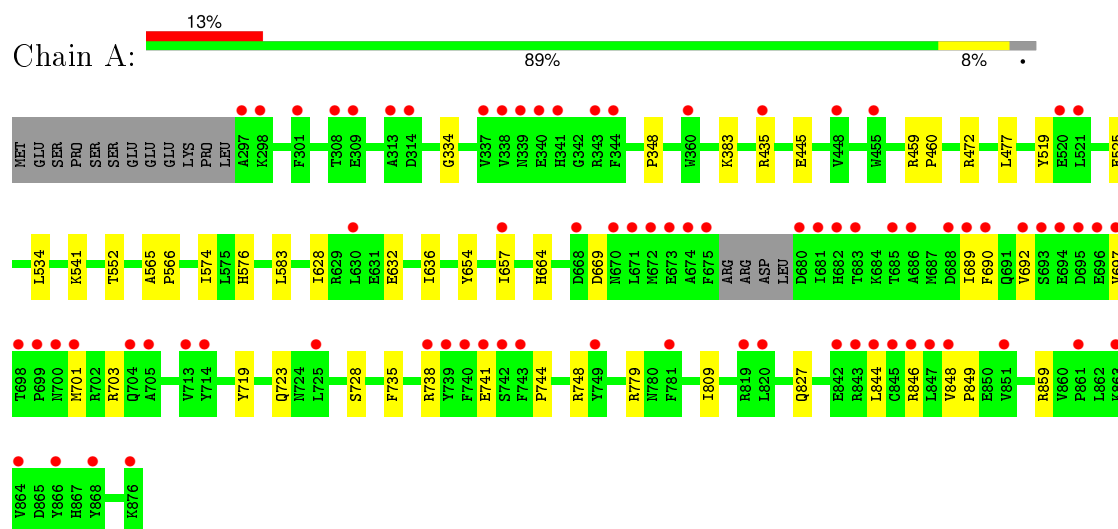
- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	388	Total	O	0	0
			388	388		
7	D	710	Total	O	0	0
			710	710		
7	B	32	Total	O	0	0
			32	32		
7	C	57	Total	O	0	0
			57	57		
7	E	37	Total	O	0	0
			37	37		
7	F	78	Total	O	0	0
			78	78		

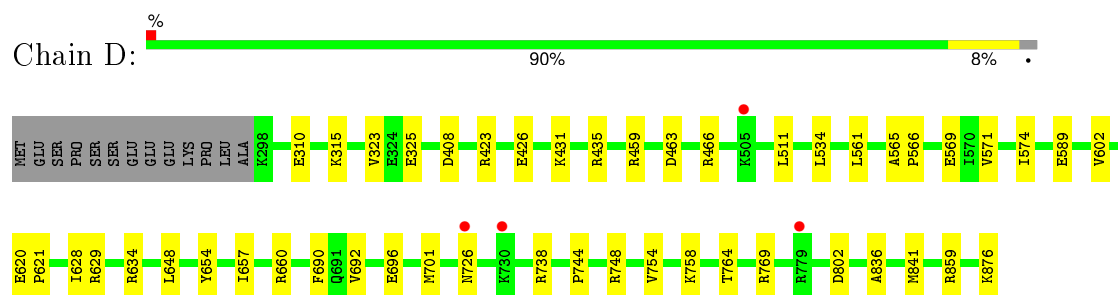
3 Residue-property plots [i](#)

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



- Molecule 1: DNA polymerase



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



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

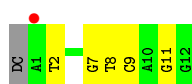




- Molecule 3: DNA (5'-D(*C*AP*TP*GP*GP*GP*AP*GP*TP*CP*AP*GP*G)-3')



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



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	94.22Å 109.34Å 149.93Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.45 – 1.74 47.11 – 1.74	Depositor EDS
% Data completeness (in resolution range)	87.0 (29.45-1.74) 87.0 (47.11-1.74)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.99 (at 1.74Å)	Xtriage
Refinement program	PHENIX (phenix.refine: dev_1026)	Depositor
R, R_{free}	0.167 , 0.195 0.179 , 0.205	Depositor DCC
R_{free} test set	6153 reflections (4.66%)	DCC
Wilson B-factor (Å ²)	19.5	Xtriage
Anisotropy	0.139	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.39 , 53.4	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 138108 reflections	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	21755	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 3.35% 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: DOC, MG, DCT, SO4

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	0/4824	0.56	0/6520
1	D	0.64	0/4944	0.73	5/6681 (0.1%)
2	B	1.11	0/173	1.99	8/264 (3.0%)
2	E	1.27	1/173 (0.6%)	1.88	5/264 (1.9%)
3	C	0.98	0/283	1.65	6/437 (1.4%)
3	F	1.16	0/283	1.91	6/437 (1.4%)
All	All	0.62	1/10680 (0.0%)	0.84	30/14603 (0.2%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	E	27	DT	C5-C7	5.17	1.53	1.50

The worst 5 of 30 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	E	23	DT	O4'-C1'-N1	-9.95	101.04	108.00
3	F	8	DT	O4'-C1'-N1	-9.71	101.20	108.00
2	B	26	DC	O4'-C1'-N1	8.99	114.30	108.00
3	F	2	DT	O4'-C1'-N1	8.70	114.09	108.00
3	C	8	DT	N3-C4-O4	7.52	124.41	119.90

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	4682	4762	4687	32	0
1	D	4744	4830	4704	36	0
2	B	174	100	103	1	0
2	E	174	100	103	0	0
3	C	251	134	136	0	0
3	F	251	134	136	0	0
4	A	54	16	24	0	0
4	D	27	8	12	0	0
5	A	1	0	0	0	0
5	D	1	0	0	0	0
6	D	10	0	0	0	0
7	A	388	0	0	13	0
7	B	32	0	0	0	0
7	C	57	0	0	0	0
7	D	710	0	0	19	0
7	E	37	0	0	0	0
7	F	78	0	0	0	0
All	All	11671	10084	9905	68	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 3.

The worst 5 of 68 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:D:423:ARG:NE	7:D:1668:HOH:O	1.97	0.94
1:D:629:ARG:NE	7:D:1666:HOH:O	2.08	0.82
1:D:629:ARG:NH2	7:D:1666:HOH:O	2.11	0.81
1:D:435:ARG:NH1	7:D:1356:HOH:O	2.15	0.80
1:A:445:GLU:OE1	7:A:1368:HOH:O	1.99	0.79

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	589/592 (100%)	572 (97%)	16 (3%)	1 (0%)	52	32
1	D	603/592 (102%)	589 (98%)	13 (2%)	1 (0%)	52	32
All	All	1192/1184 (101%)	1161 (97%)	29 (2%)	2 (0%)	52	32

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	628	ILE
1	D	628	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	507/507 (100%)	501 (99%)	6 (1%)	78	62
1	D	521/507 (103%)	520 (100%)	1 (0%)	95	92
All	All	1028/1014 (101%)	1021 (99%)	7 (1%)	88	80

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	735	PHE
1	D	511	LEU
1	A	827	GLN
1	A	669	ASP
1	A	844	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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	DOC	B	29	3,2	11,19,20	0.66	0	14,26,29	1.20	1 (7%)
2	DOC	E	29	3,2	11,19,20	0.96	0	14,26,29	1.08	2 (14%)

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	DOC	B	29	3,2	-	0/3/18/19	0/2/2/2
2	DOC	E	29	3,2	-	0/3/18/19	0/2/2/2

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	E	29	DOC	O4'-C1'-C2'	2.15	109.00	106.67
2	E	29	DOC	C2-N3-C4	2.56	119.22	115.61
2	B	29	DOC	C2-N3-C4	3.56	120.63	115.61

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

There are no carbohydrates in this entry.

5.6 Ligand geometry

Of 7 ligands modelled in this entry, 2 are monoatomic - leaving 5 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	DCT	A	901	5	20,28,28	0.81	0	29,43,43	1.97	7 (24%)
4	DCT	A	902	-	20,28,28	0.79	0	29,43,43	1.92	8 (27%)
4	DCT	D	901	5	20,28,28	0.85	1 (5%)	29,43,43	1.22	2 (6%)
6	SO4	D	903	-	4,4,4	0.42	0	6,6,6	0.61	0
6	SO4	D	904	-	4,4,4	0.13	0	6,6,6	0.17	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	DCT	A	901	5	-	0/18/31/31	0/2/2/2
4	DCT	A	902	-	-	0/18/31/31	0/2/2/2
4	DCT	D	901	5	-	0/18/31/31	0/2/2/2
6	SO4	D	903	-	-	0/0/0/0	0/0/0/0
6	SO4	D	904	-	-	0/0/0/0	0/0/0/0

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	D	901	DCT	C4-N3	2.31	1.39	1.35

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
4	A	901	DCT	C2'-C1'-N1	-5.72	100.91	112.49
4	A	902	DCT	C2'-C1'-N1	-4.47	103.43	112.49
4	A	902	DCT	PB-O3B-PG	-3.56	120.73	132.67
4	A	901	DCT	PB-O3A-PA	-3.32	123.40	132.73
4	A	902	DCT	PB-O3A-PA	-3.23	123.66	132.73

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 [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	576/592 (97%)	0.72	77 (13%) 4 5	18, 39, 79, 106	0
1	D	579/592 (97%)	0.07	4 (0%) 89 92	8, 19, 41, 59	0
2	B	8/9 (88%)	-0.31	0 100 100	22, 25, 45, 57	0
2	E	8/9 (88%)	-0.37	0 100 100	14, 19, 33, 48	0
3	C	12/13 (92%)	0.05	1 (8%) 14 17	20, 27, 77, 100	0
3	F	12/13 (92%)	-0.01	1 (8%) 14 17	11, 18, 56, 76	0
All	All	1195/1228 (97%)	0.38	83 (6%) 20 24	8, 29, 68, 106	0

The worst 5 of 83 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	297	ALA	7.0
1	A	657	ILE	6.1
1	A	866	TYR	5.8
1	A	842	GLU	5.7
1	A	671	LEU	5.6

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	DOC	B	29	18/19	0.97	0.08	-	24,29,33,39	0
2	DOC	E	29	18/19	0.99	0.11	-	11,14,18,21	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
4	DCT	A	901	27/27	0.95	0.13	-0.15	20,31,50,52	0
4	DCT	D	901	27/27	0.99	0.09	-0.18	8,13,19,20	0
4	DCT	A	902	27/27	0.91	0.10	-0.70	30,39,87,90	0
5	MG	A	903	1/1	0.84	0.09	-1.07	40,40,40,40	0
5	MG	D	902	1/1	0.99	0.07	-1.98	18,18,18,18	0
6	SO4	D	904	5/5	0.94	0.20	-	57,59,65,67	0
6	SO4	D	903	5/5	0.98	0.09	-	25,28,45,49	0

6.5 Other polymers [i](#)

There are no such residues in this entry.