



# wwPDB X-ray Structure Validation Summary Report

Feb 28, 2014 – 11:30 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 wwPDB validation summary report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

---

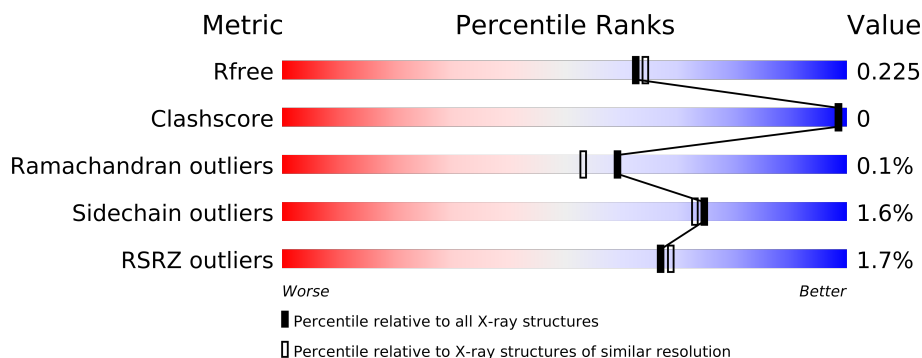
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22683

# 1 Overall quality at a glance

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))
$R_{free}$	66092	1380 (2.06-2.02)
Clashscore	79885	1577 (2.06-2.02)
Ramachandran outliers	78287	1565 (2.06-2.02)
Sidechain outliers	78261	1565 (2.06-2.02)
RSRZ outliers	66119	1381 (2.06-2.02)

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  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	903	
2	T	17	
3	P	13	

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 8580 atoms, of which 0 are hydrogen and 0 are deuterium.

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	903	Total	C	N	O	S	0	2	0
			7375	4735	1230	1377	33			

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

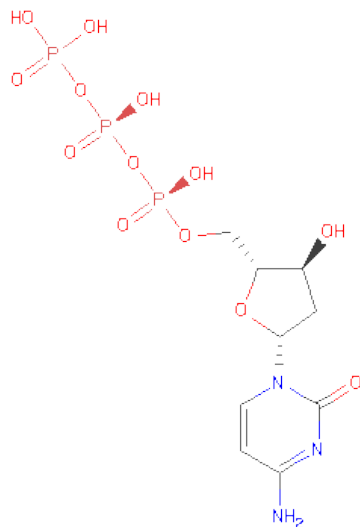
- Molecule 2 is a DNA chain called DNA template.

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

- Molecule 3 is a DNA chain called DNA primer.

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

- Molecule 4 is 2'-DEOXYCYTIDINE-5'-TRIPHOSPHATE (three-letter code: DCP) (formula: C<sub>9</sub>H<sub>16</sub>N<sub>3</sub>O<sub>13</sub>P<sub>3</sub>).



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

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

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

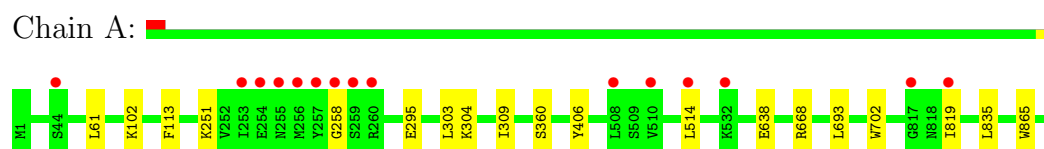
- Molecule 6 is water.

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

### 3 Residue-property plots

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 2: DNA template



- Molecule 3: DNA primer



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	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
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.30 (at 2.03Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, $R_{free}$	0.185 , 0.225 0.185 , 0.225	Depositor DCC
$R_{free}$ test set	3946 reflections (5.29%)	DCC
Wilson B-factor (Å <sup>2</sup> )	31.0	Xtriage
Anisotropy	0.077	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 37.7	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	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 <sup>2</sup> )	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.*

<sup>1</sup>Intensities estimated from amplitudes.

## 5 Model quality

### 5.1 Standard geometry

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(°)	Ideal(°)
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 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the

chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	7375	0	12	3	0
2	T	338	0	0	1	0
3	P	264	0	0	1	0
4	A	28	0	0	0	0
5	A	5	0	0	0	0
6	A	520	0	0	1	0
6	P	14	0	0	0	0
6	T	36	0	0	0	0
All	All	8580	0	12	4	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 0.

All (4) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
3:P:105:DG:C2'	3:P:106:DG:C8	2.90	0.54
1:A:668[B]:ARG:NH1	6:A:1351:HOH:O	2.50	0.44
1:A:406:TYR:OH	1:A:638:GLU:OE1	2.39	0.41
1:A:360:SER:OG	2:T:4:3DR:OP2	2.39	0.40

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	903/903 (100%)	885 (98%)	17 (2%)	1 (0%)	59 53

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	258	GLY



### 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	799/797 (100%)	786 (98%)	13 (2%)	75 73

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	303	LEU
1	A	304	LYS
1	A	819	ILE
1	A	295	GLU
1	A	693	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 chains in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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	9,11,12	5.95	1 (11%)	11,14,17	0.93	1 (9%)

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	-	0/4/15/16	0/1/1/1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	T	4	3DR	P-OP1	17.81	1.67	1.46

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	T	4	3DR	O4'-C4'-C3'	2.25	107.50	104.00

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

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	29,29,29	0.91	1 (3%)	42,45,45	1.37	5 (11%)

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	-	0/19/34/34	0/2/2/2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	1001	DCP	C2-N1	3.56	1.42	1.38

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	1001	DCP	C6-C5-C4	4.30	119.25	117.47
4	A	1001	DCP	C2-N3-C4	3.42	120.51	115.57
4	A	1001	DCP	C2'-C1'-N1	-2.97	106.37	114.08
4	A	1001	DCP	O4'-C1'-N1	2.26	111.92	107.68
4	A	1001	DCP	O2B-PB-O3A	2.24	115.75	105.14

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	903/903 (100%)	-0.15	15 (1%) 67 69	17, 30, 56, 148	0
2	T	17/17 (100%)	0.08	0 100 100	23, 38, 79, 102	0
3	P	13/13 (100%)	0.53	1 (7%) 13 12	23, 44, 84, 92	0
All	All	933/933 (100%)	-0.14	16 (1%) 67 69	17, 31, 59, 148	0

The worst 5 of 16 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	257	TYR	10.1
1	A	258	GLY	9.5
1	A	256	MET	8.7
1	A	255	ASN	8.4
1	A	254	GLU	6.3

### 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, 95<sup>th</sup> 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	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
2	3DR	T	4	11/12	0.15	0.67	26,30,40,40	0

### 6.3 Carbohydrates ⓘ

There are no carbohydrates 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. 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, 95<sup>th</sup> 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	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
4	DCP	A	1001	28/28	0.15	0.52	21,25,31,39	0
5	CA	A	1005	1/1	0.12	-1.45	69,69,69,69	0
5	CA	A	1004	1/1	0.05	-3.17	54,54,54,54	0
5	CA	A	1002	1/1	0.11	-3.19	27,27,27,27	0
5	CA	A	1003	1/1	0.08	-4.30	59,59,59,59	0
5	CA	A	1006	1/1	0.06	-11.94	43,43,43,43	0

## 6.5 Other polymers

There are no such residues in this entry.