



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 15, 2017 – 10:34 PM EST

PDB ID : 5VS2
Title : Human DNA polymerase beta pre-catalytic 8-oxoG:dA extension complex with dTTP bound in Watson-Crick conformation
Authors : Reed, A.J.; Suo, Z.
Deposited on : unknown
Resolution : 2.33 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 : 1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20030345
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
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) : rb-20030345

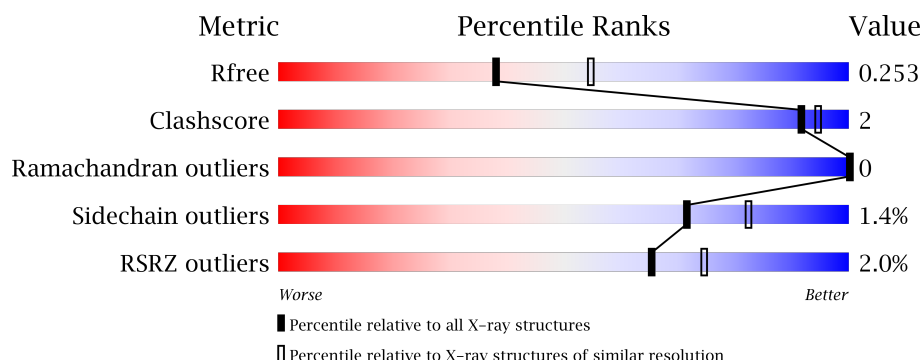
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION





The reported resolution of this entry is 2.33 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}	100719	1570 (2.36-2.32)
Clashscore	112137	1673 (2.36-2.32)
Ramachandran outliers	110173	1654 (2.36-2.32)
Sidechain outliers	110143	1655 (2.36-2.32)
RSRZ outliers	101464	1576 (2.36-2.32)

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	T	16	
2	P	10	
3	D	5	
4	A	341	

2 Entry composition [i](#)

There are 8 unique types of molecules in this entry. The entry contains 3423 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 DNA chain called DNA (5'-D(*CP*CP*GP*AP*CP*AP*(8OG)P*GP*CP*GP*CP*AP*TP*CP*AP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	T	16	Total	C	N	O	P	0	0	0
			326	154	65	92	15			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	P	10	Total	C	N	O	P	0	0	0
			202	97	38	58	9			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	5	Total	C	N	O	P	0	0	0
			106	49	20	32	5			

- Molecule 4 is a protein called DNA polymerase beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	A	328	Total	C	N	O	S	0	1	0
			2608	1647	457	495	9			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	336	HIS	-	expression tag	UNP P06746
A	337	HIS	-	expression tag	UNP P06746
A	338	HIS	-	expression tag	UNP P06746
A	339	HIS	-	expression tag	UNP P06746
A	340	HIS	-	expression tag	UNP P06746
A	341	HIS	-	expression tag	UNP P06746

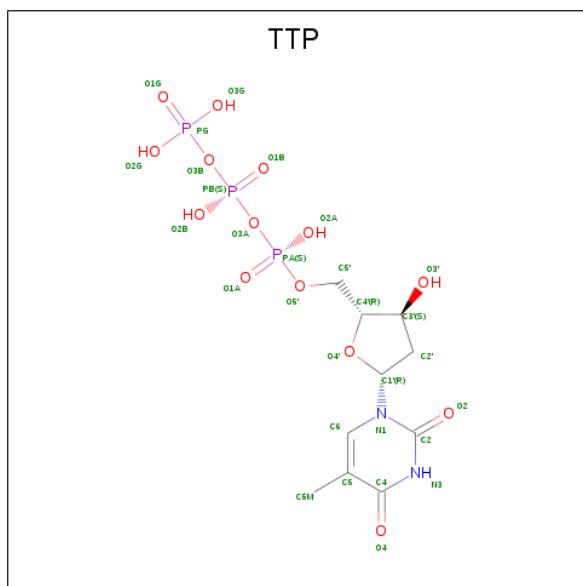
- Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	P	2	Total	Ca	0	0
			2	2		
5	A	1	Total	Ca	0	0
			1	1		

- Molecule 6 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	1	Total	Na	0	0
			1	1		
6	D	1	Total	Na	0	0
			1	1		

- Molecule 7 is THYMIDINE-5'-TRIPHOSPHATE (three-letter code: TTP) (formula: C₁₀H₁₇N₂O₁₄P₃).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
7	A	1	Total	C	N	O	P	0
			29	10	2	14	3	0

- Molecule 8 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	T	15	Total	O	0	0
			15	15		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	P	6	Total 6	O 6	0	0
8	D	1	Total 1	O 1	0	0
8	A	125	Total 125	O 125	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 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: DNA (5'-D(*CP*CP*GP*AP*CP*AP*(8OG)P*GP*CP*GP*CP*AP*TP*CP*AP*G)-3')

Chain T: 



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

Chain P: 



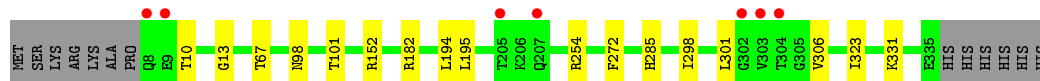
- Molecule 3: DNA (5'-D(P*GP*TP*CP*GP*G)-3')

Chain D: 



- Molecule 4: DNA polymerase beta

Chain A: 



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	50.94Å 80.49Å 55.66Å 90.00° 107.28° 90.00°	Depositor
Resolution (Å)	48.64 – 2.33 53.15 – 2.33	Depositor EDS
% Data completeness (in resolution range)	99.7 (48.64-2.33) 99.7 (53.15-2.33)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	5.07 (at 2.34Å)	Xtriage
Refinement program	REFMAC 5.8.0073	Depositor
R, R_{free}	0.187 , 0.250 0.195 , 0.253	Depositor DCC
R_{free} test set	879 reflections (5.03%)	DCC
Wilson B-factor (Å ²)	24.7	Xtriage
Anisotropy	0.125	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 36.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	3423	wwPDB-VP
Average B, all atoms (Å ²)	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.95% 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, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: NA, CA, TTP, 8OG

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	T	0.45	0/339	0.91	0/518
2	P	0.47	0/226	0.95	0/347
3	D	0.94	1/118 (0.8%)	0.74	0/179
4	A	0.66	0/2657	0.78	4/3576 (0.1%)
All	All	0.64	1/3340 (0.0%)	0.81	4/4620 (0.1%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	1	DG	OP3-P	-9.27	1.50	1.61

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	254	ARG	NE-CZ-NH2	-6.01	117.30	120.30
4	A	182	ARG	NE-CZ-NH2	-5.59	117.50	120.30
4	A	254	ARG	NE-CZ-NH1	5.58	123.09	120.30
4	A	152	ARG	NE-CZ-NH1	5.07	122.83	120.30

There are no chirality outliers.

There are no planarity outliers.

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	T	326	0	179	4	0
2	P	202	0	113	1	0
3	D	106	0	57	0	0
4	A	2608	0	2591	7	0
5	A	1	0	0	0	0
5	P	2	0	0	0	0
6	A	1	0	0	0	0
6	D	1	0	0	0	0
7	A	29	0	13	0	0
8	A	125	0	0	0	0
8	D	1	0	0	0	0
8	P	6	0	0	1	0
8	T	15	0	0	3	0
All	All	3423	0	2953	12	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.

The worst 5 of 12 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:T:16:DG:N3	8:T:102:HOH:O	2.15	0.79
1:T:12:DA:OP2	8:T:101:HOH:O	2.02	0.77
2:P:3:DG:N7	8:P:201:HOH:O	2.28	0.66
1:T:4:DA:OP2	8:T:103:HOH:O	2.17	0.57
4:A:194:LEU:HD23	4:A:195:LEU:N	2.27	0.49

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	
4	A	327/341 (96%)	314 (96%)	13 (4%)	0	100	100

There are no Ramachandran outliers to report.

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	
4	A	281/301 (93%)	277 (99%)	4 (1%)	71	82

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	A	67	THR
4	A	272	PHE
4	A	306	VAL
4	A	331	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
4	A	245	ASN
4	A	285	HIS

5.3.3 RNA ⓘ

There are no RNA molecules 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
1	8OG	T	7	1,2	17,25,26	1.51	3 (17%)	22,37,40	1.92	6 (27%)

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
1	8OG	T	7	1,2	-	0/3/21/22	0/3/3/3

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	T	7	8OG	C8-N7	-3.36	1.30	1.34
1	T	7	8OG	C6-C5	2.93	1.46	1.41
1	T	7	8OG	C5-C4	3.20	1.47	1.40

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	T	7	8OG	C5-C6-N1	-3.32	118.75	123.48
1	T	7	8OG	N3-C2-N1	-2.09	124.41	127.46
1	T	7	8OG	C1'-N9-C4	2.22	131.11	126.99
1	T	7	8OG	C6-N1-C2	2.57	119.76	116.06
1	T	7	8OG	O4'-C1'-N9	3.50	111.42	108.15

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 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$
7	TTP	A	401	5	22,30,30	1.17	1 (4%)	25,47,47	2.00	5 (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
7	TTP	A	401	5	-	0/18/34/34	0/2/2/2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	A	401	TTP	PG-O3B	3.56	1.65	1.60

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
7	A	401	TTP	C5-C4-N3	-4.06	120.76	125.24
7	A	401	TTP	C2'-C3'-C4'	-2.62	97.16	102.73
7	A	401	TTP	O2A-PA-O1A	2.05	122.91	112.28
7	A	401	TTP	C5M-C5-C6	2.99	124.63	118.67
7	A	401	TTP	C4-N3-C2	7.12	121.39	115.16

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	T	15/16 (93%)	-0.05	0	100	100	14, 32, 51, 56	0
2	P	10/10 (100%)	-0.19	0	100	100	14, 34, 43, 44	0
3	D	5/5 (100%)	-0.01	0	100	100	26, 28, 43, 48	0
4	A	328/341 (96%)	-0.12	7 (2%)	64	73	10, 23, 44, 76	0
All	All	358/372 (96%)	-0.12	7 (1%)	65	74	10, 24, 45, 76	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	A	302	GLY	3.6
4	A	303	VAL	3.1
4	A	8	GLN	2.9
4	A	205	THR	2.8
4	A	304	THR	2.7

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
1	8OG	T	7	23/24	0.97	0.12	-	14,15,17,18	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
7	TTP	A	401	29/29	0.96	0.14	1.07	11,12,28,33	0
6	NA	D	101	1/1	0.87	0.15	0.76	32,32,32,32	0
5	CA	A	402	1/1	0.98	0.06	-2.26	20,20,20,20	0
5	CA	P	102	1/1	0.97	0.07	-3.48	37,37,37,37	0
5	CA	P	101	1/1	0.98	0.06	-12.15	13,13,13,13	0
6	NA	A	403	1/1	0.96	0.09	-	14,14,14,14	0

6.5 Other polymers [i](#)

There are no such residues in this entry.