



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 13, 2017 – 03:54 am GMT

PDB ID : 4JV2
Title : Ternary complex of gamma-OHPDG adduct modified dna with dna (-1 primer)
polymerase iv and incoming datp
Authors : Banerjee, S.; Shanmugam, G.; Stone, M.P.
Deposited on : 2013-03-25
Resolution : 2.74 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 : 1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix) : 1.9-1692
EDS : trunk28620
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) : recalc28949

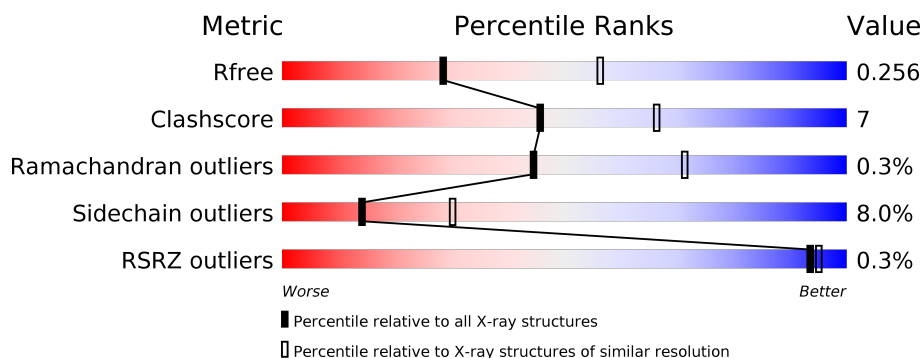
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION




The reported resolution of this entry is 2.74 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}	100719	3342 (2.78-2.70)
Clashscore	112137	3731 (2.78-2.70)
Ramachandran outliers	110173	3670 (2.78-2.70)
Sidechain outliers	110143	3671 (2.78-2.70)
RSRZ outliers	101464	3362 (2.78-2.70)

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	347	 79% 18% . .
2	B	15	 53% 47%
3	C	13	 46% 46% 8%

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 3392 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	341	Total	C	N	O	S	0	0	0
			2744	1760	472	505	7			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-5	HIS	-	EXPRESSION TAG	UNP Q97W02
A	-4	HIS	-	EXPRESSION TAG	UNP Q97W02
A	-3	HIS	-	EXPRESSION TAG	UNP Q97W02
A	-2	HIS	-	EXPRESSION TAG	UNP Q97W02
A	-1	HIS	-	EXPRESSION TAG	UNP Q97W02
A	0	HIS	-	EXPRESSION TAG	UNP Q97W02

- Molecule 2 is a DNA chain called DNA (5'-D(P*TP*(HN1)P*GP*AP*AP*TP*CP*CP*TP*TP*CP*CP*CP*CP*C)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	15	Total	C	N	O	P	0	0	0
			303	146	49	93	15			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	13	Total	C	N	O	P	0	0	0
			273	129	57	75	12			

- Molecule 4 is 2'-DEOXYADENOSINE 5'-TRIPHOSPHATE (three-letter code: DTP) (formula: C₁₀H₁₆N₅O₁₂P₃).



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

- Molecule 6 is water.



- Molecule 1: DNA polymerase IV

I322	I166	HIS
D326	D167	HIS
E327	D168	HIS
	E169	HIS
	V170	HIS
	V171	HIS
	K172	M1
	R173	
	L174	F5
	L175	
	R176	R36
	E177	
	L178	E49
	D179	
	I180	K52
	L194	I59
	K195	
	K196	K65
		K66
		E79
	D211	L92
	M216	Y95
	R230	S103
	D231	I104
	E232	D105
		E106
	R238	
	V241	V115
	R242	E120
	L245	L124
	M251	N130
	N254	K131
		I132
	V264	L133
	E270	K137
	V273	I144
	L276	V149
	R279	I153
	H285	D156
	V296	M157
		A158
	I306	K159
	V312	G162
		I163
		K164
		V165

- Molecule 2: DNA (5'-D(P*TP*(HN1)P*GP*AP*AP*TP*CP*CP*TP*TP*CP*CP*CP*CP*C)-3')

T604	T605	G606	A607	A608		T612	T613	C614		C617	C618
------	------	------	------	------	--	------	------	------	--	------	------

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

G501
G502
G503
G504
G505
A506
A507
G508
G509
C513

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α , β , γ	94.58Å 103.31Å 52.78Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	42.09 – 2.74 42.09 – 2.74	Depositor EDS
% Data completeness (in resolution range)	99.0 (42.09-2.74) 94.4 (42.09-2.74)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.18 (at 2.73Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE: 1.7.2_869)	Depositor
R, R_{free}	0.203 , 0.265 0.196 , 0.256	Depositor DCC
R_{free} test set	1401 reflections (10.01%)	DCC
Wilson B-factor (Å ²)	57.8	Xtriage
Anisotropy	0.244	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 30.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	3392	wwPDB-VP
Average B, all atoms (Å ²)	56.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.93% 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 [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: HN1, CA, DTP

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	0/2783	0.56	0/3736
2	B	0.83	0/306	1.56	6/465 (1.3%)
3	C	0.86	0/308	1.63	6/476 (1.3%)
All	All	0.55	0/3397	0.88	12/4677 (0.3%)

There are no bond length outliers.

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	513	DC	O4'-C1'-N1	12.37	116.66	108.00
3	C	502	DG	O4'-C1'-N9	8.26	113.78	108.00
3	C	504	DG	O4'-C1'-N9	8.15	113.71	108.00
2	B	607	DA	C1'-O4'-C4'	-6.42	103.68	110.10
2	B	617	DC	O4'-C1'-N1	6.21	112.34	108.00

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	2744	0	2889	37	0
2	B	303	0	172	6	0
3	C	273	0	147	3	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	A	30	0	12	4	0
5	A	2	0	0	0	0
6	A	28	0	0	3	0
6	B	7	0	0	0	0
6	C	5	0	0	0	0
All	All	3392	0	3220	46	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 7.

The worst 5 of 46 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)
2:B:605:HN1:N9	2:B:605:HN1:C1'	1.73	1.51
2:B:605:HN1:H	3:C:513:DC:C4	2.03	0.92
2:B:605:HN1:C4	2:B:605:HN1:C1'	2.57	0.83
2:B:605:HN1:N9	2:B:605:HN1:C2'	2.53	0.68
1:A:242:ARG:HD2	1:A:245:ILE:HD11	1.77	0.66

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	339/347 (98%)	320 (94%)	18 (5%)	1 (0%)	44 71

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	167	ASP

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	300/306 (98%)	276 (92%)	24 (8%)	14	30

5 of 24 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	199	ILE
1	A	230	ARG
1	A	331	ARG
1	A	211	ASP
1	A	216	MET

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

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	HN1	B	605	2	19,29,37	5.51	6 (31%)	19,43,55	3.08	6 (31%)

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	HN1	B	605	2	-	0/3/31/43	0/3/4/4

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	605	HN1	O5'-C5'	-21.41	1.14	1.44
2	B	605	HN1	C3-N2	-2.02	1.42	1.47
2	B	605	HN1	C2-N2	2.28	1.39	1.32
2	B	605	HN1	C6-N1	4.43	1.44	1.38
2	B	605	HN1	C4-N3	5.72	1.41	1.33

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	605	HN1	C2'-C1'-N9	-5.18	101.99	114.23
2	B	605	HN1	C4'-O4'-C1'	-2.37	103.63	109.42
2	B	605	HN1	C4-C5-N7	-2.07	107.41	109.41
2	B	605	HN1	C-C3-N2	2.77	121.68	113.49
2	B	605	HN1	O11-C1-N1	6.16	124.30	109.66

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	605	HN1	5	0

5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 3 ligands modelled in this entry, 2 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	DTP	A	401	5	26,32,32	1.12	3 (11%)	26,50,50	2.03	4 (15%)

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	DTP	A	401	5	-	0/18/34/34	0/3/3/3

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	401	DTP	C5'-C4'	-2.28	1.44	1.51
4	A	401	DTP	O3'-C3'	-2.15	1.38	1.43
4	A	401	DTP	C6-N6	2.60	1.44	1.34

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	401	DTP	N3-C2-N1	-7.86	122.01	128.86
4	A	401	DTP	C4-C5-N7	-3.21	106.31	109.41
4	A	401	DTP	O5'-C5'-C4'	2.87	119.17	109.00
4	A	401	DTP	O4'-C1'-N9	4.07	114.65	107.78

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	401	DTP	4	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	341/347 (98%)	-0.05	1 (0%) 93 95	40, 54, 74, 88	0
2	B	14/15 (93%)	-0.56	0 100 100	46, 55, 65, 65	0
3	C	13/13 (100%)	-0.63	0 100 100	42, 47, 68, 72	0
All	All	368/375 (98%)	-0.09	1 (0%) 93 95	40, 54, 74, 88	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	196	LYS	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	HN1	B	605	26/34	0.92	0.17	-	40,55,65,70	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
5	CA	A	402	1/1	0.79	0.19	0.64	50,50,50,50	0
4	DTP	A	401	30/30	0.95	0.14	-1.01	43,49,61,67	0
5	CA	A	403	1/1	0.90	0.30	-	63,63,63,63	0

6.5 Other polymers [i](#)

There are no such residues in this entry.