



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 21, 2023 – 08:08 PM EDT

PDB ID : 2PY5
Title : Phi29 DNA polymerase complexed with single-stranded DNA
Authors : Berman, A.J.; Kamtekar, S.; Goodman, J.L.; Lazaro, J.M.; de Vega, M.;
Blanco, L.; Salas, M.; Steitz, T.A.
Deposited on : 2007-05-15
Resolution : 1.60 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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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

MolProbity : 4.02b-467
Mogul : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.35
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac : 5.8.0158
CCP4 : 7.0.044 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.35

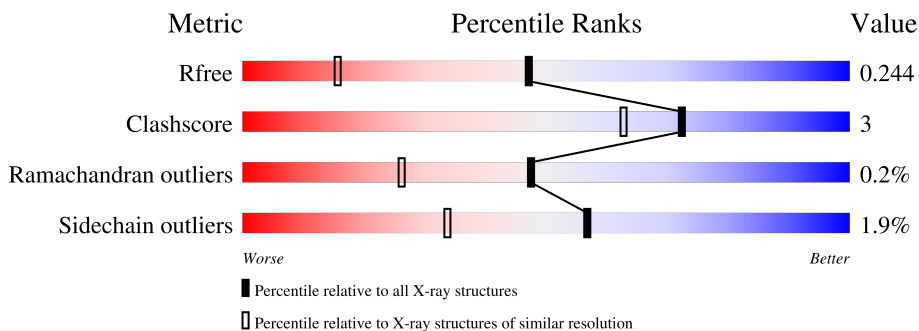
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.60 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}	130704	3398 (1.60-1.60)
Clashscore	141614	3665 (1.60-1.60)
Ramachandran outliers	138981	3564 (1.60-1.60)
Sidechain outliers	138945	3563 (1.60-1.60)

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 of the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria respectively. 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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	D	7	29% (green), 57% (yellow), 14% (grey)
1	E	7	57% (green), 29% (yellow), 14% (grey)
1	J	7	71% (green), 14% (yellow), 14% (grey)
1	L	7	43% (green), 14% (orange), 43% (grey)
1	Y	7	71% (green), 14% (yellow), 14% (grey)
2	A	575	91% (green), 6% (yellow), 3% (orange), 0% (red), 0% (grey) ..
2	B	575	89% (green), 9% (yellow), 2% (orange), 0% (red), 0% (grey) ..

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 11400 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 5'-d(GGACTTT)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	J	6	Total 122	C 59	N 19	O 38	P 6	0	0	0
1	Y	6	Total 105	C 49	N 20	O 31	P 5	0	0	0
1	D	6	Total 121	C 59	N 22	O 35	P 5	0	0	0
1	E	6	Total 105	C 49	N 20	O 31	P 5	0	0	0
1	L	4	Total 79	C 39	N 9	O 27	P 4	0	0	0

- Molecule 2 is a protein called DNA polymerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	A	564	Total 4650	C 3037	N 741	O 850	S 22	28	11	0
2	B	562	Total 4617	C 3014	N 737	O 844	S 22	11	7	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	12	ALA	ASP	engineered mutation	UNP P03680
A	66	ALA	ASP	engineered mutation	UNP P03680
B	12	ALA	ASP	engineered mutation	UNP P03680
B	66	ALA	ASP	engineered mutation	UNP P03680

- Molecule 3 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C₂H₆O₂).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 4 2 2	0	0
3	A	1	Total C O 4 2 2	0	0
3	A	1	Total C O 4 2 2	0	0
3	A	1	Total C O 4 2 2	0	0
3	A	1	Total C O 4 2 2	0	0
3	A	1	Total C O 4 2 2	0	0
3	A	1	Total C O 4 2 2	0	0
3	A	1	Total C O 4 2 2	0	0
3	A	1	Total C O 4 2 2	0	0
3	A	1	Total C O 4 2 2	0	0
3	A	1	Total C O 4 2 2	0	0
3	A	1	Total C O 4 2 2	0	0
3	A	1	Total C O 4 2 2	0	0
3	A	1	Total C O 4 2 2	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 4 2 2	0	0
3	A	1	Total C O 4 2 2	0	0
3	A	1	Total C O 4 2 2	0	0
3	A	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0
3	B	1	Total C O 4 2 2	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	J	31	Total O 31 31	0	0
4	Y	9	Total O 9 9	0	0
4	D	20	Total O 20 20	0	0
4	E	20	Total O 20 20	0	0
4	L	20	Total O 20 20	0	0
4	A	768	Total O 768 768	0	0

Continued on next page...

Continued from previous page...

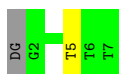
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	625	Total 625	O 625	0	0

3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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. 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. 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: 5'-d(GGACTTT)-3'

Chain J:  71% 14% 14%



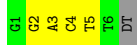
- Molecule 1: 5'-d(GGACTTT)-3'

Chain Y:  71% 14% 14%



- Molecule 1: 5'-d(GGACTTT)-3'

Chain D:  29% 57% 14%



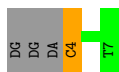
- Molecule 1: 5'-d(GGACTTT)-3'

Chain E:  57% 29% 14%



- Molecule 1: 5'-d(GGACTTT)-3'

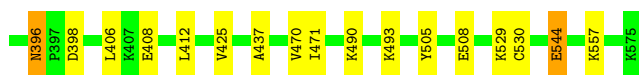
Chain L:  43% 14% 43%



- Molecule 2: DNA polymerase

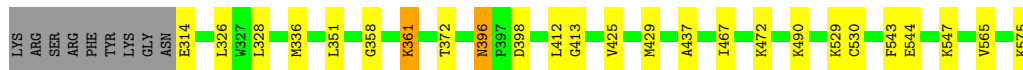
Chain A:  91% 6% ..





- Molecule 2: DNA polymerase

Chain B: 89% 9% ..



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	54.20Å 200.18Å 66.99Å 90.00° 109.40° 90.00°	Depositor
Resolution (Å)	40.59 – 1.60 40.58 – 1.60	Depositor EDS
% Data completeness (in resolution range)	97.8 (40.59-1.60) 97.8 (40.58-1.60)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.23 (at 1.60Å)	Xtrriage
Refinement program	REFMAC 5.2	Depositor
R, R_{free}	0.164 , 0.194 0.224 , 0.244	Depositor DCC
R_{free} test set	8614 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	23.8	Xtrriage
Anisotropy	0.088	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.37 , 53.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	11400	wwPDB-VP
Average B, all atoms (Å ²)	35.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.62% 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: EDO

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	D	0.85	0/135	1.54	2/207 (1.0%)
1	E	0.95	0/117	1.52	2/180 (1.1%)
1	J	0.85	0/135	1.30	1/206 (0.5%)
1	L	1.19	1/86 (1.2%)	1.68	0/130
1	Y	0.69	0/117	1.29	1/180 (0.6%)
2	A	0.66	7/4801 (0.1%)	0.67	4/6477 (0.1%)
2	B	0.45	1/4755 (0.0%)	0.57	0/6418
All	All	0.59	9/10146 (0.1%)	0.71	10/13798 (0.1%)

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	135	LYS	CB-CG	-19.61	0.99	1.52
2	A	544	GLU	CB-CG	-13.40	1.26	1.52
2	A	241	GLU	CB-CG	-9.14	1.34	1.52
2	B	314	GLU	CD-OE2	7.14	1.33	1.25
1	L	4	DC	C3'-O3'	-6.58	1.35	1.44

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	135	LYS	CA-CB-CG	15.16	146.76	113.40
1	D	3	DA	O4'-C1'-N9	-8.52	102.04	108.00
2	A	241	GLU	CA-CB-CG	8.12	131.25	113.40
2	A	544	GLU	CA-CB-CG	7.98	130.96	113.40
1	E	3	DA	O4'-C1'-N9	-7.82	102.53	108.00

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	D	121	0	70	1	0
1	E	105	0	57	0	0
1	J	122	0	70	0	0
1	L	79	0	48	1	0
1	Y	105	0	57	0	0
2	A	4650	0	4683	25	0
2	B	4617	0	4642	39	0
3	A	72	0	108	4	0
3	B	36	0	54	0	0
4	A	768	0	0	4	2
4	B	625	0	0	4	2
4	D	20	0	0	0	0
4	E	20	0	0	0	0
4	J	31	0	0	0	0
4	L	20	0	0	1	0
4	Y	9	0	0	0	0
All	All	11400	0	9789	66	2

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 66 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:328:LEU:CD1	2:B:336:MET:HE1	1.79	1.12
2:B:328:LEU:HD11	2:B:336:MET:HE1	1.06	1.06
2:B:302:ILE:CD1	2:B:336:MET:HE2	1.99	0.93
2:B:302:ILE:HD11	2:B:336:MET:CE	2.05	0.87
2:A:72:ASN:HD22	2:A:76:ARG:HH12	1.31	0.78

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
4:A:9650:HOH:O	4:B:9051:HOH:O[2_646]	1.61	0.59
4:A:9184:HOH:O	4:B:9568:HOH:O[2_646]	1.99	0.21

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
2	A	571/575 (99%)	560 (98%)	10 (2%)	1 (0%)	47 26
2	B	565/575 (98%)	553 (98%)	11 (2%)	1 (0%)	47 26
All	All	1136/1150 (99%)	1113 (98%)	21 (2%)	2 (0%)	47 26

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	425	VAL
2	A	425	VAL

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
2	A	506/506 (100%)	499 (99%)	7 (1%)	67 47
2	B	501/506 (99%)	488 (97%)	13 (3%)	46 21
All	All	1007/1012 (100%)	987 (98%)	20 (2%)	57 31

5 of 20 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	396	ASN
2	B	544	GLU
2	B	575	LYS
2	B	547	LYS
2	A	544	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	303	GLN
2	B	387	ASN
2	B	396	ASN
2	A	171	GLN
2	A	396	ASN

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

5.6 Ligand geometry [i](#)

27 ligands 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
3	EDO	B	9002	-	3,3,3	0.43	0	2,2,2	0.22	0
3	EDO	B	8002	-	3,3,3	0.44	0	2,2,2	0.22	0
3	EDO	A	9001	-	3,3,3	0.49	0	2,2,2	0.15	0
3	EDO	A	9012	-	3,3,3	0.49	0	2,2,2	0.22	0
3	EDO	A	9010	-	3,3,3	0.53	0	2,2,2	0.08	0
3	EDO	A	8012	-	3,3,3	0.37	0	2,2,2	0.32	0
3	EDO	A	8013	-	3,3,3	0.54	0	2,2,2	0.15	0
3	EDO	A	8009	-	3,3,3	0.48	0	2,2,2	0.06	0
3	EDO	A	8011	-	3,3,3	0.49	0	2,2,2	0.16	0
3	EDO	A	9005	-	3,3,3	0.44	0	2,2,2	0.25	0
3	EDO	A	9008	-	3,3,3	0.49	0	2,2,2	0.14	0
3	EDO	A	8001	-	3,3,3	0.59	0	2,2,2	0.44	0
3	EDO	A	8004	-	3,3,3	0.50	0	2,2,2	0.20	0
3	EDO	A	9006	-	3,3,3	0.46	0	2,2,2	0.23	0
3	EDO	B	9004	-	3,3,3	0.50	0	2,2,2	0.26	0
3	EDO	A	9011	-	3,3,3	0.45	0	2,2,2	0.17	0
3	EDO	A	9014	-	3,3,3	0.50	0	2,2,2	0.16	0
3	EDO	B	8005	-	3,3,3	0.43	0	2,2,2	0.29	0
3	EDO	A	8008	-	3,3,3	0.50	0	2,2,2	0.16	0
3	EDO	B	9003	-	3,3,3	0.45	0	2,2,2	0.29	0
3	EDO	B	9009	-	3,3,3	0.51	0	2,2,2	0.19	0
3	EDO	B	8007	-	3,3,3	0.50	0	2,2,2	0.37	0
3	EDO	A	9007	-	3,3,3	0.45	0	2,2,2	0.33	0
3	EDO	B	8006	-	3,3,3	0.34	0	2,2,2	0.47	0
3	EDO	A	8010	-	3,3,3	0.44	0	2,2,2	0.38	0
3	EDO	B	8003	-	3,3,3	0.39	0	2,2,2	0.36	0
3	EDO	A	9013	-	3,3,3	0.53	0	2,2,2	0.15	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
3	EDO	B	9002	-	-	1/1/1/1	-
3	EDO	B	8002	-	-	0/1/1/1	-
3	EDO	A	9001	-	-	0/1/1/1	-
3	EDO	A	9012	-	-	0/1/1/1	-
3	EDO	A	9010	-	-	1/1/1/1	-
3	EDO	A	8012	-	-	0/1/1/1	-
3	EDO	A	8013	-	-	1/1/1/1	-
3	EDO	A	8009	-	-	0/1/1/1	-
3	EDO	A	8011	-	-	0/1/1/1	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	EDO	A	9005	-	-	0/1/1/1	-
3	EDO	A	9008	-	-	0/1/1/1	-
3	EDO	A	8001	-	-	1/1/1/1	-
3	EDO	A	8004	-	-	1/1/1/1	-
3	EDO	A	9006	-	-	1/1/1/1	-
3	EDO	B	9004	-	-	0/1/1/1	-
3	EDO	A	9011	-	-	0/1/1/1	-
3	EDO	A	9014	-	-	0/1/1/1	-
3	EDO	B	8005	-	-	0/1/1/1	-
3	EDO	A	8008	-	-	0/1/1/1	-
3	EDO	B	9003	-	-	0/1/1/1	-
3	EDO	B	9009	-	-	0/1/1/1	-
3	EDO	B	8007	-	-	1/1/1/1	-
3	EDO	A	9007	-	-	1/1/1/1	-
3	EDO	B	8006	-	-	1/1/1/1	-
3	EDO	A	8010	-	-	1/1/1/1	-
3	EDO	B	8003	-	-	1/1/1/1	-
3	EDO	A	9013	-	-	1/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 12 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	8004	EDO	O1-C1-C2-O2
3	A	8010	EDO	O1-C1-C2-O2
3	A	9006	EDO	O1-C1-C2-O2
3	A	9010	EDO	O1-C1-C2-O2
3	A	9013	EDO	O1-C1-C2-O2

There are no ring outliers.

3 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	8012	EDO	1	0
3	A	8001	EDO	1	0
3	A	9006	EDO	2	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

6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.