



wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 31, 2016 – 11:29 PM GMT

PDB ID : 1XHX
Title : Phi29 DNA Polymerase, orthorhombic crystal form
Authors : Kamtekar, S.; Berman, A.J.; Wang, J.; Lazaro, J.M.; de Vega, M.; Blanco, L.; Salas, M.; Steitz, T.A.
Deposited on : 2004-09-21
Resolution : 2.35 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

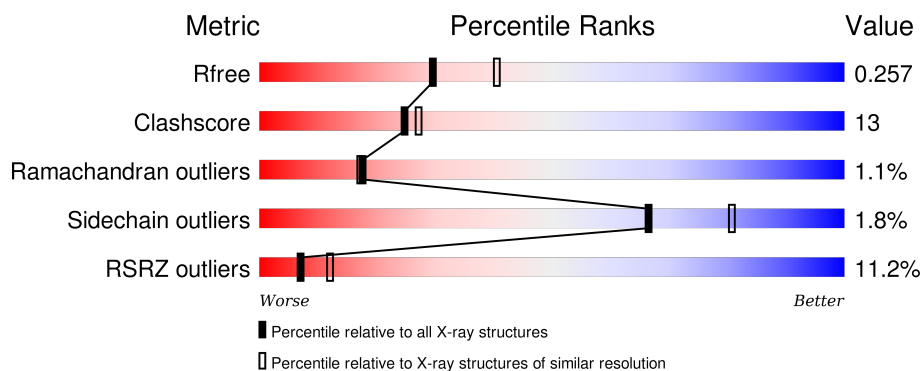
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.35 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	1352 (2.38-2.34)
Clashscore	102246	1456 (2.38-2.34)
Ramachandran outliers	100387	1435 (2.38-2.34)
Sidechain outliers	100360	1436 (2.38-2.34)
RSRZ outliers	91569	1358 (2.38-2.34)

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	575	<div> <div>15%</div> <div>73%</div> <div>25%</div> <div>..</div> </div>
1	B	575	<div> <div>12%</div> <div>72%</div> <div>26%</div> <div>..</div> </div>
1	C	575	<div> <div>11%</div> <div>73%</div> <div>25%</div> <div>..</div> </div>
1	D	575	<div> <div>6%</div> <div>79%</div> <div>19%</div> <div>..</div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	MG	C	5000	-	-	-	X

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 20193 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.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	571	Total	C	N	O	S	0	0	0
			4668	3041	754	852	21			
1	B	571	Total	C	N	O	S	0	0	0
			4668	3041	754	852	21			
1	C	571	Total	C	N	O	S	0	0	0
			4668	3041	754	852	21			
1	D	571	Total	C	N	O	S	0	0	0
			4668	3041	754	852	21			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	12	ALA	ASP	ENGINEERED	UNP P03680
A	66	ALA	ASP	ENGINEERED	UNP P03680
B	12	ALA	ASP	ENGINEERED	UNP P03680
B	66	ALA	ASP	ENGINEERED	UNP P03680
C	12	ALA	ASP	ENGINEERED	UNP P03680
C	66	ALA	ASP	ENGINEERED	UNP P03680
D	12	ALA	ASP	ENGINEERED	UNP P03680
D	66	ALA	ASP	ENGINEERED	UNP P03680

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



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

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	C	2	Total	Mg	0	0
			2	2		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	295	Total	O	0	0
			295	295		
4	B	388	Total	O	0	0
			388	388		
4	C	382	Total	O	0	0
			382	382		
4	D	444	Total	O	0	0
			444	444		

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 ($\text{RSRZ} > 2$). 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.

Chain A:

15% 73% 25%

Chain A is a sequence of 100 amino acids, represented by three-letter codes. The sequence is as follows:

1574 K575 I574

T434 A435 R438 D458 A466 L467 P468 D469 V470 L471 K472 F489 K493 V494 L495 R496 E508 V509 D510 E511 K512 L513 G532 K536 K539 E544 I545 F546 G549 F550 S551 E552 K553 M554 K555 P556 V559 Q560 V561 P562 G563 G564 V565 D570 T571 F572 E573

Chain A is a sequence of 100 amino acids, represented by three-letter codes. The sequence is as follows:

S319 E322 I323 L326 K327 L328 L333 M336 V345 K352 K366 T372 T373 Y390 S395 N396 P397 D398 V399 T400 G401 K402 P403 P404 Y405 L406 E408 H409 A411 L412 G413 F414 R415 L416 G417 E418 A419 E420 T421 P422 K423 P424 V425 Y426 T427 V431

Chain A is a sequence of 100 amino acids, represented by three-letter codes. The sequence is as follows:

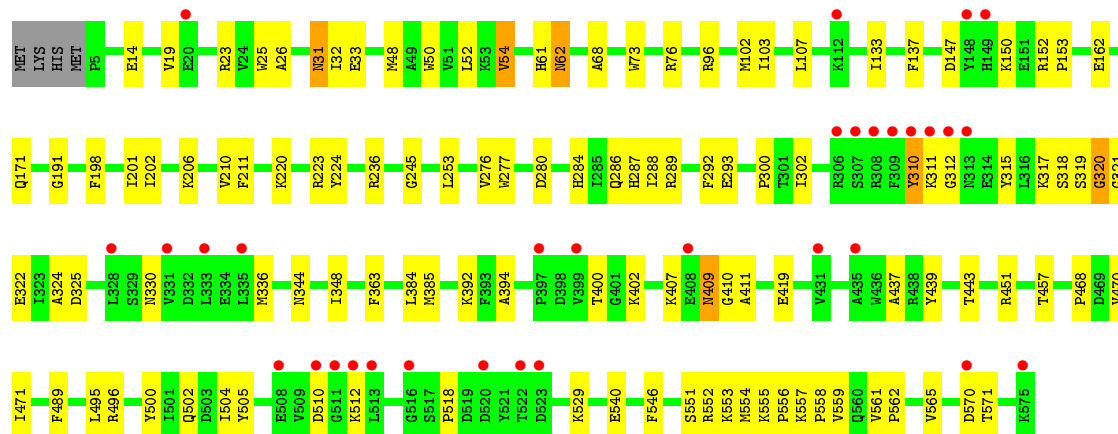
G85 L86 P87 N88 T89 Y90 I94 S95 A96 N97 W100 Y101 M102 I103 D104 I105 G106 L107 G108 G111 K112 R113 V118 I119 P127 D145 I146 D147 Y148 H149 K150 R152 P153 V154 G155 K157 I158 T159 E162 Y163 I166 Q171 E175 L186 D188 L316 L317 K318

Chain A is a sequence of 100 amino acids, represented by three-letter codes. The sequence is as follows:

MET LYS HIS MET P5 M8 C11 A12 F13 E14 T16 V19 E20 R23 Y27 N31 I32 E33 D34 I40 G41 N42 W50 V51 L52 K53 V54 F60 H61 N62 L63 K64 F65 A66 G67 A68 F69 I70 I71 N72 W73 L74 E75 F79 K80 W81 S82 R83 D84

Chain B:

12% 72% 26%



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	96.26Å 149.91Å 199.02Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.96 – 2.35 30.72 – 2.35	Depositor EDS
% Data completeness (in resolution range)	99.7 (19.96-2.35) 99.8 (30.72-2.35)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.16 (at 2.36Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.206 , 0.257 0.207 , 0.257	Depositor DCC
R_{free} test set	11990 reflections (10.01%)	DCC
Wilson B-factor (Å ²)	42.7	Xtriage
Anisotropy	0.251	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 47.3	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 119894 reflections	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	20193	wwPDB-VP
Average B, all atoms (Å ²)	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.72% 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: MG, 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	0.35	0/4788	0.65	3/6459 (0.0%)
1	B	0.36	0/4788	0.62	0/6459
1	C	0.36	0/4788	0.61	0/6459
1	D	0.38	0/4788	0.62	0/6459
All	All	0.36	0/19152	0.63	3/25836 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	5	PRO	CA-N-CD	-13.35	92.81	111.50
1	A	309	PHE	CA-CB-CG	-6.95	97.21	113.90
1	A	309	PHE	CA-C-N	-5.65	104.77	117.20

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	4668	0	4676	134	0
1	B	4668	0	4676	116	0
1	C	4668	0	4676	141	0
1	D	4668	0	4676	97	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	D	10	0	0	0	0
3	C	2	0	0	0	0
4	A	295	0	0	9	0
4	B	388	0	0	4	0
4	C	382	0	0	12	0
4	D	444	0	0	7	0
All	All	20193	0	18704	478	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 13.

The worst 5 of 478 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:A:310:TYR:OH	1:A:322:GLU:HG3	1.37	1.23
1:C:223:ARG:NH2	1:C:424:PRO:HG2	1.70	1.06
1:C:110:LYS:HG2	1:C:115:ILE:HD11	1.36	1.04
1:B:89:THR:HG22	1:B:90:TYR:H	1.22	1.03
1:A:75:GLU:HB3	1:A:406:LEU:HD11	1.45	0.98

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	569/575 (99%)	518 (91%)	44 (8%)	7 (1%)	16	15
1	B	569/575 (99%)	521 (92%)	44 (8%)	4 (1%)	26	29
1	C	569/575 (99%)	525 (92%)	36 (6%)	8 (1%)	14	12
1	D	569/575 (99%)	536 (94%)	28 (5%)	5 (1%)	21	22
All	All	2276/2300 (99%)	2100 (92%)	152 (7%)	24 (1%)	17	17

5 of 24 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	62	ASN
1	A	311	LYS
1	B	311	LYS
1	D	62	ASN
1	A	85	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	502/506 (99%)	492 (98%)	10 (2%)	63	77
1	B	502/506 (99%)	493 (98%)	9 (2%)	66	81
1	C	502/506 (99%)	491 (98%)	11 (2%)	60	75
1	D	502/506 (99%)	496 (99%)	6 (1%)	78	89
All	All	2008/2024 (99%)	1972 (98%)	36 (2%)	66	81

5 of 36 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	223	ARG
1	C	145	ASP
1	D	311	LYS
1	B	552	ARG
1	C	186	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 28 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	396	ASN
1	C	35	HIS
1	D	303	GLN
1	B	545	ASN
1	C	31	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

5.6 Ligand geometry ⓘ

Of 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 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
2	SO4	D	3000	-	4,4,4	3.27	2 (50%)	6,6,6	0.99	0
2	SO4	D	3001	-	4,4,4	3.27	2 (50%)	6,6,6	0.87	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
2	SO4	D	3000	-	-	0/0/0/0	0/0/0/0
2	SO4	D	3001	-	-	0/0/0/0	0/0/0/0

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	3000	SO4	O3-S	-4.71	1.30	1.47

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	3001	SO4	O3-S	-4.29	1.32	1.47
2	D	3000	SO4	O1-S	4.29	1.61	1.47
2	D	3001	SO4	O1-S	4.88	1.63	1.47

There are no bond angle outliers.

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	571/575 (99%)	0.78	88 (15%) 3 5	25, 55, 108, 135	0
1	B	571/575 (99%)	0.65	69 (12%) 6 10	23, 50, 95, 149	0
1	C	571/575 (99%)	0.62	66 (11%) 6 11	22, 47, 106, 147	0
1	D	571/575 (99%)	0.34	32 (5%) 28 42	22, 43, 84, 150	0
All	All	2284/2300 (99%)	0.60	255 (11%) 7 12	22, 48, 103, 150	0

The worst 5 of 255 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	310	TYR	11.6
1	D	306	ARG	8.8
1	A	148	TYR	8.7
1	B	310	TYR	8.5
1	A	309	PHE	8.3

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

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

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
3	MG	C	5000	1/1	0.95	0.22	2.66	47,47,47,47	0
3	MG	C	5001	1/1	0.88	0.18	0.21	53,53,53,53	0
2	SO4	D	3000	5/5	0.97	0.16	-	67,75,77,83	0
2	SO4	D	3001	5/5	0.93	0.19	-	98,99,102,103	0

6.5 Other polymers [i](#)

There are no such residues in this entry.