



wwPDB X-ray Structure Validation Summary Report ⓘ

May 16, 2020 – 11:03 am BST

PDB ID : 5VBN
Title : Crystal Structure of human DNA polymerase epsilon B-subunit in complex with C-terminal domain of catalytic subunit
Authors : Baranovskiy, A.G.; Gu, J.; Suwa, Y.; Babayeva, N.D.; Tahirov, T.H.
Deposited on : 2017-03-29
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

<https://www.wwpdb.org/validation/2017/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 : 1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.11
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) : 2.11

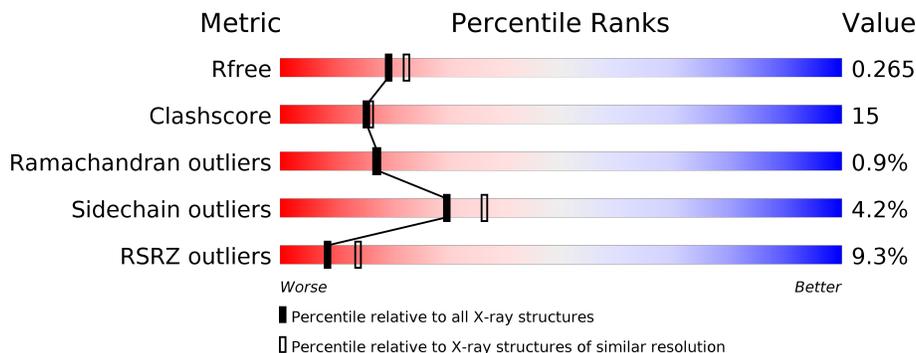
1 Overall quality at a glance

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}	130704	1164 (2.36-2.36)
Clashscore	141614	1232 (2.36-2.36)
Ramachandran outliers	138981	1211 (2.36-2.36)
Sidechain outliers	138945	1212 (2.36-2.36)
RSRZ outliers	127900	1150 (2.36-2.36)

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 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 $\leq 5\%$. 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	527	 4% 58% 20% • 19%
1	E	527	 5% 60% 19% • 19%
2	B	145	 14% 53% 30% • 14%
2	F	145	 23% 43% 41% • 14%

2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 9058 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.

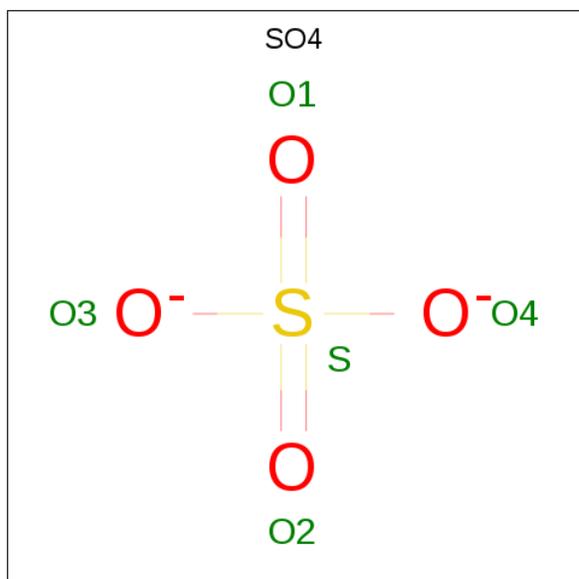
- Molecule 1 is a protein called DNA polymerase epsilon subunit 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	427	Total 3425	C 2219	N 565	O 624	S 17	0	0	0
1	E	427	Total 3425	C 2219	N 565	O 624	S 17	0	0	0

- Molecule 2 is a protein called DNA polymerase epsilon catalytic subunit A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	125	Total 991	C 631	N 161	O 184	S 15	0	0	0
2	F	125	Total 991	C 631	N 161	O 184	S 15	0	0	0

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



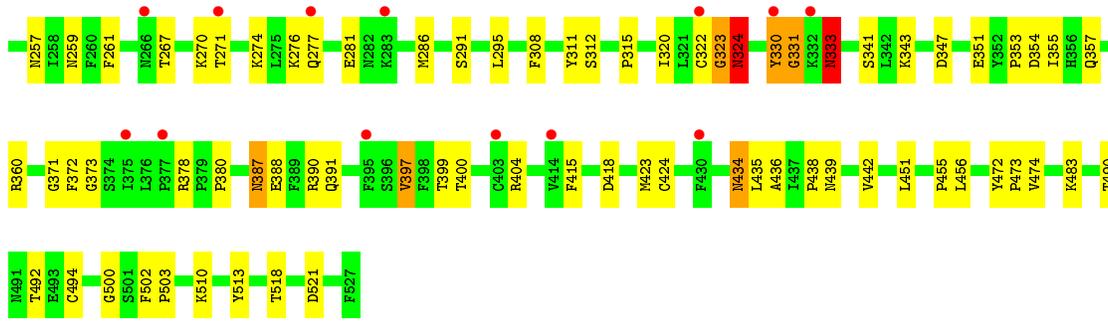
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total O S 5 4 1	0	0
3	E	1	Total O S 5 4 1	0	0

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

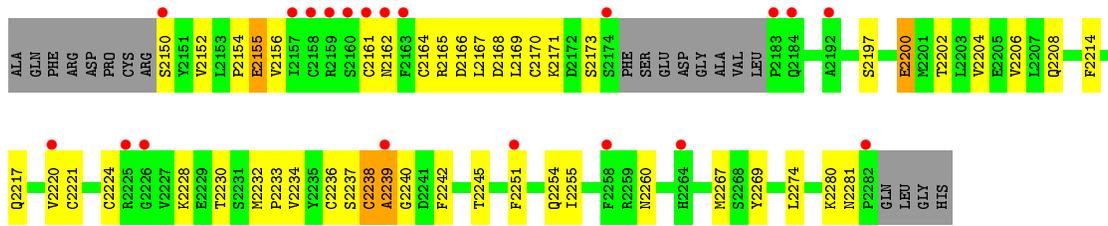
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	2	Total Zn 2 2	0	0
4	F	2	Total Zn 2 2	0	0

- Molecule 5 is water.

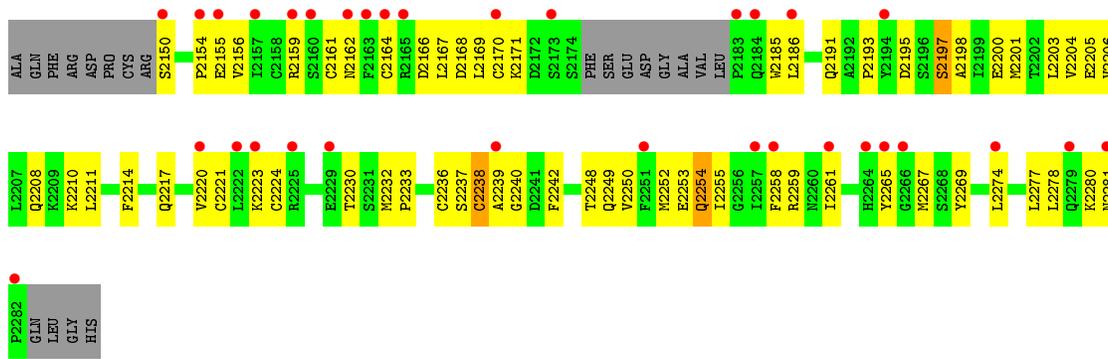
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	108	Total O 108 108	0	0
5	B	6	Total O 6 6	0	0
5	E	94	Total O 94 94	0	0
5	F	4	Total O 4 4	0	0



• Molecule 2: DNA polymerase epsilon catalytic subunit A



• Molecule 2: DNA polymerase epsilon catalytic subunit A



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α , β , γ	92.26Å 201.46Å 78.34Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.75 – 2.35 39.75 – 2.33	Depositor EDS
% Data completeness (in resolution range)	94.7 (39.75-2.35) 93.9 (39.75-2.33)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.57 (at 2.34Å)	Xtrriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.225 , 0.264 0.228 , 0.265	Depositor DCC
R_{free} test set	3117 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å ²)	37.3	Xtrriage
Anisotropy	0.592	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 42.7	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	0.94	EDS
Total number of atoms	9058	wwPDB-VP
Average B, all atoms (Å ²)	46.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 34.31 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 6.9667e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹ Intensities estimated from amplitudes.

² Theoretical 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: ZN, 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	0.48	0/3526	0.70	3/4794 (0.1%)
1	E	0.46	0/3526	0.69	3/4794 (0.1%)
2	B	0.39	0/1010	0.68	1/1365 (0.1%)
2	F	0.40	0/1010	0.68	1/1365 (0.1%)
All	All	0.45	0/9072	0.69	8/12318 (0.1%)

There are no bond length outliers.

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	331	GLY	N-CA-C	6.93	130.44	113.10
1	A	331	GLY	N-CA-C	6.91	130.38	113.10
1	E	324	ASN	N-CA-C	-6.37	93.79	111.00
1	A	324	ASN	N-CA-C	-6.31	93.95	111.00
1	E	323	GLY	N-CA-C	5.65	127.22	113.10

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	3425	0	3361	100	0
1	E	3425	0	3361	86	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	991	0	977	39	0
2	F	991	0	977	52	0
3	A	5	0	0	0	0
3	E	5	0	0	0	0
4	B	2	0	0	0	0
4	F	2	0	0	0	0
5	A	108	0	0	4	0
5	B	6	0	0	0	0
5	E	94	0	0	2	0
5	F	4	0	0	0	0
All	All	9058	0	8676	261	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 15.

The worst 5 of 261 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:513:TYR:HB2	1:A:518:THR:HG22	1.55	0.85
1:E:513:TYR:HB2	1:E:518:THR:HG22	1.57	0.84
2:B:2167:LEU:HD22	2:B:2169:LEU:HD23	1.58	0.84
1:A:488:THR:HG22	1:A:510:LYS:HZ1	1.41	0.83
1:E:343:LYS:HE3	1:E:388:GLU:OE1	1.80	0.81

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	423/527 (80%)	401 (95%)	18 (4%)	4 (1%)	17 17
1	E	423/527 (80%)	402 (95%)	17 (4%)	4 (1%)	17 17

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	121/145 (83%)	111 (92%)	9 (7%)	1 (1%)	19	20
2	F	121/145 (83%)	111 (92%)	9 (7%)	1 (1%)	19	20
All	All	1088/1344 (81%)	1025 (94%)	53 (5%)	10 (1%)	17	17

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	333	ASN
2	B	2238	CYS
1	E	333	ASN
2	F	2238	CYS
1	A	174	SER

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	383/471 (81%)	365 (95%)	18 (5%)	26	31
1	E	383/471 (81%)	368 (96%)	15 (4%)	32	40
2	B	115/131 (88%)	109 (95%)	6 (5%)	23	27
2	F	115/131 (88%)	112 (97%)	3 (3%)	46	56
All	All	996/1204 (83%)	954 (96%)	42 (4%)	30	36

5 of 42 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	2155	GLU
2	B	2260	ASN
1	E	434	ASN
2	B	2165	ARG
2	B	2197	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 26 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	214	HIS
1	E	333	ASN
2	F	2254	GLN
1	E	259	ASN
1	E	297	GLN

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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 6 ligands modelled in this entry, 4 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	$\# Z > 2$
3	SO4	A	601	-	4,4,4	0.30	0	6,6,6	0.20	0
3	SO4	E	601	-	4,4,4	0.21	0	6,6,6	0.11	0

There are no bond length outliers.

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

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 [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	427/527 (81%)	0.30	22 (5%) 27 39	21, 38, 64, 80	0
1	E	427/527 (81%)	0.42	28 (6%) 18 26	22, 39, 67, 80	0
2	B	125/145 (86%)	1.00	20 (16%) 1 3	41, 58, 68, 71	0
2	F	125/145 (86%)	1.31	33 (26%) 0 0	42, 62, 72, 77	0
All	All	1104/1344 (82%)	0.54	103 (9%) 8 14	21, 44, 68, 80	0

The worst 5 of 103 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	332	LYS	6.3
2	F	2258	PHE	6.2
1	E	144	HIS	5.4
1	A	332	LYS	5.4
1	E	175	THR	5.3

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. 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	B-factors(\AA^2)	Q<0.9
4	ZN	F	2301	1/1	0.90	0.05	62,62,62,62	0
4	ZN	B	2302	1/1	0.95	0.03	66,66,66,66	0
4	ZN	F	2302	1/1	0.96	0.04	66,66,66,66	0
3	SO4	A	601	5/5	0.97	0.13	62,63,64,64	0
3	SO4	E	601	5/5	0.98	0.12	60,62,63,64	0
4	ZN	B	2301	1/1	0.99	0.04	56,56,56,56	0

6.5 Other polymers [i](#)

There are no such residues in this entry.