



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 26, 2020 – 05:34 am BST

PDB ID : 3VW4  
Title : Crystal structure of the DNA-binding domain of ColE2-P9 Rep in complex with the replication origin  
Authors : Itou, H.; Yagura, M.; Itoh, T.; Shirakihara, Y.  
Deposited on : 2012-07-31  
Resolution : 2.70 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](mailto: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

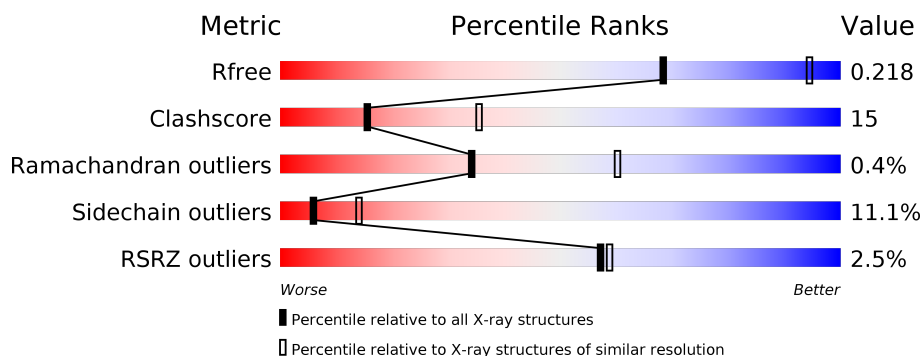
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.70 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	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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  $\geq 3$ , 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	128	<div> <div>2%</div> <div> <div></div> <div>70%</div> <div>20%</div> <div>• • 5%</div> </div> </div>
1	B	128	<div> <div>2%</div> <div> <div></div> <div>72%</div> <div>18%</div> <div>• 6%</div> </div> </div>
2	C	23	<div> <div>4%</div> <div> <div></div> <div>17%</div> <div>35%</div> <div>35%</div> <div>13%</div> </div> </div>
2	E	23	<div> <div>4%</div> <div> <div></div> <div>39%</div> <div>35%</div> <div>17%</div> <div>9%</div> </div> </div>
3	D	23	<div> <div>4%</div> <div> <div></div> <div>17%</div> <div>61%</div> <div>9%</div> <div>13%</div> </div> </div>
3	F	23	<div> <div></div> <div> <div></div> <div>17%</div> <div>48%</div> <div>22%</div> <div>13%</div> </div> </div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 3689 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 Rep.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	122	Total	C	N	O	S	0	0	0
			994	635	196	160	3			
1	B	120	Total	C	N	O	S	0	0	0
			978	626	192	158	2			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	167	MET	-	EXPRESSION TAG	UNP Q06B24
A	168	GLY	-	EXPRESSION TAG	UNP Q06B24
A	169	HIS	-	EXPRESSION TAG	UNP Q06B24
A	170	HIS	-	EXPRESSION TAG	UNP Q06B24
A	171	HIS	-	EXPRESSION TAG	UNP Q06B24
A	172	HIS	-	EXPRESSION TAG	UNP Q06B24
A	173	HIS	-	EXPRESSION TAG	UNP Q06B24
A	174	HIS	-	EXPRESSION TAG	UNP Q06B24
B	167	MET	-	EXPRESSION TAG	UNP Q06B24
B	168	GLY	-	EXPRESSION TAG	UNP Q06B24
B	169	HIS	-	EXPRESSION TAG	UNP Q06B24
B	170	HIS	-	EXPRESSION TAG	UNP Q06B24
B	171	HIS	-	EXPRESSION TAG	UNP Q06B24
B	172	HIS	-	EXPRESSION TAG	UNP Q06B24
B	173	HIS	-	EXPRESSION TAG	UNP Q06B24
B	174	HIS	-	EXPRESSION TAG	UNP Q06B24

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	20	Total	C	N	O	P	0	0	0
			412	196	80	116	20			

*Continued on next page...*

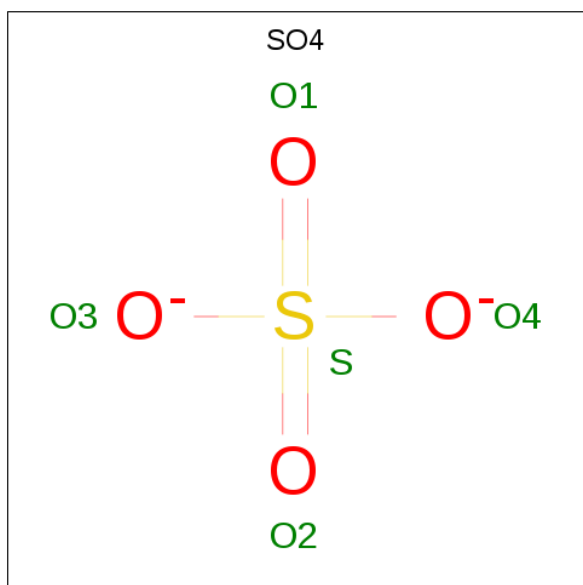
Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	21	Total	C	N	O	P	0	0	0
			433	206	85	121	21			

- Molecule 3 is a DNA chain called DNA (5'-D(P\*GP\*AP\*TP\*AP\*AP\*GP\*GP\*CP\*TP\*TP\*AP\*TP\*CP\*TP\*GP\*GP\*TP\*CP\*TP\*CP\*AP\*TP\*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	20	Total	C	N	O	P	0	0	0
			410	196	71	123	20			
3	F	20	Total	C	N	O	P	0	0	0
			410	196	71	123	20			

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



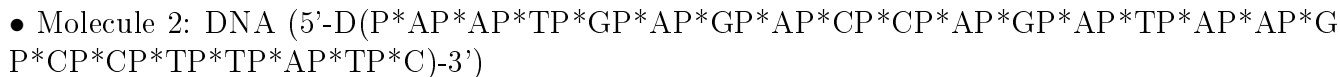
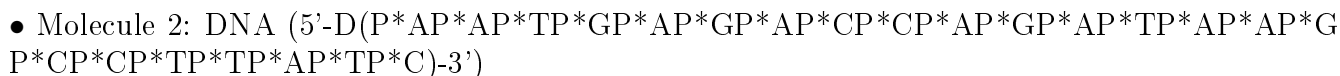
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	O	S	0	0
			5	4	1		
4	B	1	Total	O	S	0	0
			5	4	1		
4	B	1	Total	O	S	0	0
			5	4	1		
4	B	1	Total	O	S	0	0
			5	4	1		
4	B	1	Total	O	S	0	0
			5	4	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	10	Total 10	O 10	0	0
5	B	14	Total 14	O 14	0	0
5	D	2	Total 2	O 2	0	0
5	F	1	Total 1	O 1	0	0



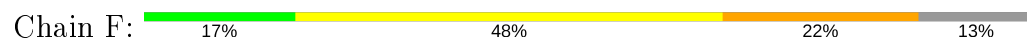
- Molecule 1: Rep



- Molecule 3: DNA (5'-D(P\*GP\*AP\*TP\*AP\*AP\*GP\*GP\*CP\*TP\*TP\*AP\*TP\*CP\*TP\*GP\*GP\*TP\*CP\*TP\*CP\*AP\*TP\*T)-3')



- Molecule 3: DNA (5'-D(P\*GP\*AP\*TP\*AP\*AP\*GP\*GP\*CP\*TP\*TP\*AP\*TP\*CP\*TP\*GP\*GP\*TP\*CP\*TP\*CP\*AP\*TP\*T)-3')



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	47.66 Å 61.82 Å 273.51 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.00 – 2.70 36.69 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.2 (40.00-2.70) 99.3 (36.69-2.70)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.70 (at 2.68 Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, $R_{free}$	0.220 , 0.262 0.222 , 0.218	Depositor DCC
$R_{free}$ test set	1154 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	29.6	Xtriage
Anisotropy	0.337	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 37.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	3689	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.06% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>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: 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	1.07	2/1025 (0.2%)	1.00	2/1382 (0.1%)
1	B	1.17	4/1008 (0.4%)	1.07	6/1359 (0.4%)
2	C	0.51	0/463	1.44	14/712 (2.0%)
2	E	0.49	0/487	1.27	8/749 (1.1%)
3	D	0.66	0/458	1.48	9/705 (1.3%)
3	F	0.74	1/458 (0.2%)	1.56	11/705 (1.6%)
All	All	0.91	7/3899 (0.2%)	1.26	50/5612 (0.9%)

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	275	TRP	CD2-CE2	6.00	1.48	1.41
1	B	186	TRP	CD2-CE2	5.92	1.48	1.41
1	B	195	TRP	CD2-CE2	5.55	1.48	1.41
1	A	243	GLU	CG-CD	5.40	1.60	1.51
1	B	201	TRP	CD2-CE2	5.24	1.47	1.41

The worst 5 of 50 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	F	13	DG	O5'-P-OP1	12.25	125.40	110.70
2	E	21	DT	O5'-P-OP1	10.29	123.05	110.70
2	E	22	DT	P-O3'-C3'	9.50	131.10	119.70
3	D	15	DT	O5'-P-OP2	-9.39	97.25	105.70
3	D	17	DA	O5'-P-OP2	8.82	121.29	110.70

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	994	0	1005	28	0
1	B	978	0	993	15	0
2	C	412	0	225	27	0
2	E	433	0	236	14	0
3	D	410	0	228	21	0
3	F	410	0	228	15	0
4	A	5	0	0	0	0
4	B	20	0	0	0	0
5	A	10	0	0	0	0
5	B	14	0	0	1	0
5	D	2	0	0	0	0
5	F	1	0	0	0	0
All	All	3689	0	2915	93	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 93 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:189:ARG:HH22	2:C:22:DT:H2'	1.17	1.06
1:A:249:GLN:HE22	3:D:18:DG:H21	1.08	1.01
1:A:189:ARG:NH2	2:C:22:DT:H2'	1.75	0.99
3:F:30:DC:H2'	3:F:31:DT:C6	2.00	0.95
3:D:27:DG:H2''	3:D:28:DG:H5''	1.50	0.90

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	120/128 (94%)	113 (94%)	7 (6%)	0	100	100
1	B	118/128 (92%)	108 (92%)	9 (8%)	1 (1%)	19	43
All	All	238/256 (93%)	221 (93%)	16 (7%)	1 (0%)	34	60

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	172	HIS

### 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	100/105 (95%)	89 (89%)	11 (11%)	6	14
1	B	98/105 (93%)	87 (89%)	11 (11%)	6	13
All	All	198/210 (94%)	176 (89%)	22 (11%)	6	14

5 of 22 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	272	LEU
1	B	182	LYS
1	B	282	ARG
1	A	282	ARG
1	B	179	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	174	HIS
1	B	249	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	193	GLN
1	A	172	HIS
1	A	249	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](#)

5 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	$\# Z  > 2$
4	SO4	A	301	-	4,4,4	0.71	0	6,6,6	1.33	0
4	SO4	B	301	-	4,4,4	1.26	1 (25%)	6,6,6	0.96	0
4	SO4	B	304	-	4,4,4	1.25	0	6,6,6	1.40	1 (16%)
4	SO4	B	303	-	4,4,4	1.44	1 (25%)	6,6,6	0.85	0
4	SO4	B	302	-	4,4,4	1.24	0	6,6,6	1.17	0

All (2) bond length outliers are listed below:

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
-----	-------	-----	------	-------	---	-------------	----------

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	301	SO4	O1-S	2.06	1.57	1.46
4	B	303	SO4	O2-S	2.03	1.57	1.46

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	B	304	SO4	O4-S-O3	2.33	119.03	109.06

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	122/128 (95%)	-0.14	2 (1%) 72 74	4, 14, 80, 139	0
1	B	120/128 (93%)	-0.27	3 (2%) 57 59	3, 13, 86, 98	0
2	C	20/23 (86%)	0.16	1 (5%) 28 27	28, 49, 75, 93	0
2	E	21/23 (91%)	0.13	1 (4%) 30 28	26, 49, 77, 95	0
3	D	20/23 (86%)	-0.41	1 (5%) 28 27	7, 24, 66, 74	0
3	F	20/23 (86%)	-0.44	0 100 100	5, 25, 67, 70	0
All	All	323/348 (92%)	-0.19	8 (2%) 57 59	3, 18, 81, 139	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	291	CYS	4.9
1	B	171	HIS	4.2
2	E	23	DA	2.8
2	C	22	DT	2.7
1	B	279	GLY	2.6

### 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, 95<sup>th</sup> 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(Å <sup>2</sup> )	Q<0.9
4	SO4	B	304	5/5	0.94	0.14	34,35,42,47	0
4	SO4	B	303	5/5	0.94	0.19	43,47,55,55	0
4	SO4	A	301	5/5	0.99	0.12	9,10,11,11	0
4	SO4	B	301	5/5	0.99	0.16	14,17,18,18	0
4	SO4	B	302	5/5	0.99	0.12	6,6,6,6	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.