



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 05:40 am BST

PDB ID : 4NOI  
Title : 2.17 Angstrom Crystal Structure of DNA-directed RNA Polymerase Subunit Alpha from *Campylobacter jejuni*.  
Authors : Minasov, G.; Shuvalova, L.; Dubrovskaya, I.; Winsor, J.; Grimshaw, S.; Kwon, K.; Anderson, W.F.; Center for Structural Genomics of Infectious Diseases (CSGID)  
Deposited on : 2013-11-19  
Resolution : 2.17 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

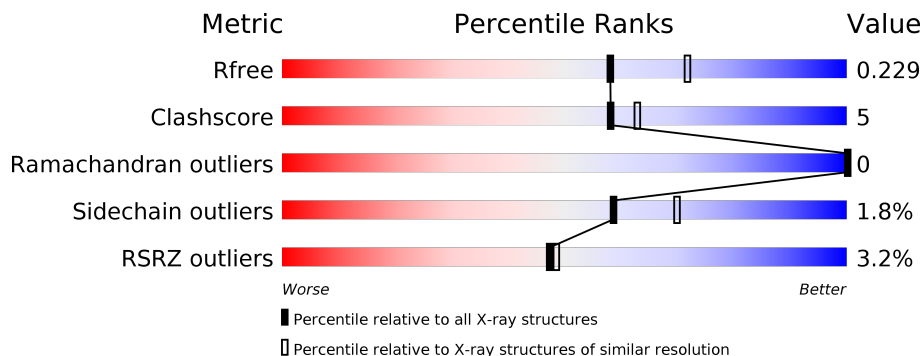
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.17 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	6864 (2.20-2.16)
Clashscore	141614	7689 (2.20-2.16)
Ramachandran outliers	138981	7564 (2.20-2.16)
Sidechain outliers	138945	7564 (2.20-2.16)
RSRZ outliers	127900	6738 (2.20-2.16)

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	361	 2% 56% 7% 37%
1	B	361	 2% 55% 6% 39%

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 3856 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-directed RNA polymerase subunit alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	Se			
1	A	227	1833	1185	286	359	3	0	3	0
1	B	219	1779	1152	277	344	6	0	4	0

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-23	MSE	-	EXPRESSION TAG	UNP Q9PM80
A	-22	HIS	-	EXPRESSION TAG	UNP Q9PM80
A	-21	HIS	-	EXPRESSION TAG	UNP Q9PM80
A	-20	HIS	-	EXPRESSION TAG	UNP Q9PM80
A	-19	HIS	-	EXPRESSION TAG	UNP Q9PM80
A	-18	HIS	-	EXPRESSION TAG	UNP Q9PM80
A	-17	HIS	-	EXPRESSION TAG	UNP Q9PM80
A	-16	SER	-	EXPRESSION TAG	UNP Q9PM80
A	-15	SER	-	EXPRESSION TAG	UNP Q9PM80
A	-14	GLY	-	EXPRESSION TAG	UNP Q9PM80
A	-13	VAL	-	EXPRESSION TAG	UNP Q9PM80
A	-12	ASP	-	EXPRESSION TAG	UNP Q9PM80
A	-11	LEU	-	EXPRESSION TAG	UNP Q9PM80
A	-10	GLY	-	EXPRESSION TAG	UNP Q9PM80
A	-9	THR	-	EXPRESSION TAG	UNP Q9PM80
A	-8	GLU	-	EXPRESSION TAG	UNP Q9PM80
A	-7	ASN	-	EXPRESSION TAG	UNP Q9PM80
A	-6	LEU	-	EXPRESSION TAG	UNP Q9PM80
A	-5	TYR	-	EXPRESSION TAG	UNP Q9PM80
A	-4	PHE	-	EXPRESSION TAG	UNP Q9PM80
A	-3	GLN	-	EXPRESSION TAG	UNP Q9PM80
A	-2	SER	-	EXPRESSION TAG	UNP Q9PM80
A	-1	ASN	-	EXPRESSION TAG	UNP Q9PM80
A	0	ALA	-	EXPRESSION TAG	UNP Q9PM80
B	-23	MSE	-	EXPRESSION TAG	UNP Q9PM80

*Continued on next page...*

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-22	HIS	-	EXPRESSION TAG	UNP Q9PM80
B	-21	HIS	-	EXPRESSION TAG	UNP Q9PM80
B	-20	HIS	-	EXPRESSION TAG	UNP Q9PM80
B	-19	HIS	-	EXPRESSION TAG	UNP Q9PM80
B	-18	HIS	-	EXPRESSION TAG	UNP Q9PM80
B	-17	HIS	-	EXPRESSION TAG	UNP Q9PM80
B	-16	SER	-	EXPRESSION TAG	UNP Q9PM80
B	-15	SER	-	EXPRESSION TAG	UNP Q9PM80
B	-14	GLY	-	EXPRESSION TAG	UNP Q9PM80
B	-13	VAL	-	EXPRESSION TAG	UNP Q9PM80
B	-12	ASP	-	EXPRESSION TAG	UNP Q9PM80
B	-11	LEU	-	EXPRESSION TAG	UNP Q9PM80
B	-10	GLY	-	EXPRESSION TAG	UNP Q9PM80
B	-9	THR	-	EXPRESSION TAG	UNP Q9PM80
B	-8	GLU	-	EXPRESSION TAG	UNP Q9PM80
B	-7	ASN	-	EXPRESSION TAG	UNP Q9PM80
B	-6	LEU	-	EXPRESSION TAG	UNP Q9PM80
B	-5	TYR	-	EXPRESSION TAG	UNP Q9PM80
B	-4	PHE	-	EXPRESSION TAG	UNP Q9PM80
B	-3	GLN	-	EXPRESSION TAG	UNP Q9PM80
B	-2	SER	-	EXPRESSION TAG	UNP Q9PM80
B	-1	ASN	-	EXPRESSION TAG	UNP Q9PM80
B	0	ALA	-	EXPRESSION TAG	UNP Q9PM80

- Molecule 2 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total Na 1 1	0	0

- Molecule 3 is IODIDE ION (three-letter code: IOD) (formula: I).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total I 1 1	0	0
3	A	3	Total I 3 3	0	0

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	5	Total Cl 5 5	0	0
4	A	7	Total Cl 9 9	0	2

- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	101	Total O 106 106	0	5
5	B	116	Total O 119 119	0	3



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	99.17Å 62.74Å 93.62Å 90.00° 112.52° 90.00°	Depositor
Resolution (Å)	29.49 – 2.17 29.49 – 2.17	Depositor EDS
% Data completeness (in resolution range)	98.5 (29.49-2.17) 98.5 (29.49-2.17)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	0.08	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.09 (at 2.18Å)	Xtrriage
Refinement program	REFMAC 5.8.0046	Depositor
R, $R_{free}$	0.174 , 0.225 0.181 , 0.229	Depositor DCC
$R_{free}$ test set	1407 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	42.8	Xtrriage
Anisotropy	0.236	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 43.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	3856	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	49.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.14% 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: NA, IOD, CL

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	0.55	0/1867	0.72	0/2528
1	B	0.63	0/1811	0.76	0/2446
All	All	0.59	0/3678	0.74	0/4974

There are no bond length outliers.

There are no bond angle outliers.

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	1833	0	1822	17	0
1	B	1779	0	1778	19	0
2	A	1	0	0	0	0
3	A	3	0	0	0	0
3	B	1	0	0	0	0
4	A	9	0	0	0	0
4	B	5	0	0	0	0
5	A	106	0	0	0	0
5	B	119	0	0	4	0
All	All	3856	0	3600	35	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 5.

The worst 5 of 35 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:108:SER:HB2	1:A:135:GLU:HG3	1.71	0.71
1:B:74[A]:MSE:SE	1:B:133:ILE:HG12	2.41	0.70
1:A:51:THR:HG21	1:A:207:ILE:HD13	1.78	0.66
1:A:61[A]:ILE:HD13	1:A:141:PHE:CB	2.39	0.53
1:B:151:TYR:HD1	1:B:172:PHE:CD1	2.27	0.52

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	228/361 (63%)	221 (97%)	7 (3%)	0	100	100
1	B	219/361 (61%)	216 (99%)	3 (1%)	0	100	100
All	All	447/722 (62%)	437 (98%)	10 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 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	202/309 (65%)	199 (98%)	3 (2%)	65	76
1	B	196/309 (63%)	190 (97%)	6 (3%)	40	48
All	All	398/618 (64%)	389 (98%)	9 (2%)	59	60

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	71[B]	MSE
1	B	177	ARG
1	B	74[B]	MSE
1	A	185	LYS
1	B	74[A]	MSE

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](#)

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 19 ligands modelled in this entry, 19 are monoatomic - leaving 0 for Mogul analysis.

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 [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	224/361 (62%)	0.05	8 (3%) 42 43	36, 49, 75, 106	0
1	B	216/361 (59%)	-0.05	6 (2%) 53 54	28, 42, 75, 104	0
All	All	440/722 (60%)	0.00	14 (3%) 47 48	28, 46, 77, 106	0

The worst 5 of 14 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	136	ASP	4.8
1	B	136	ASP	4.3
1	B	192	PRO	4.1
1	B	187	LEU	3.7
1	A	19	ILE	3.0

### 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( $\text{\AA}^2$ )	Q<0.9
4	CL	A	411[A]	1/1	0.88	0.10	55,55,55,55	1
4	CL	A	411[B]	1/1	0.88	0.10	63,63,63,63	1
4	CL	A	405	1/1	0.91	0.20	66,66,66,66	0
4	CL	B	403	1/1	0.91	0.07	60,60,60,60	0
4	CL	B	402	1/1	0.92	0.18	51,51,51,51	0
4	CL	B	406	1/1	0.94	0.07	57,57,57,57	0
4	CL	B	405	1/1	0.95	0.06	50,50,50,50	0
4	CL	A	409	1/1	0.96	0.08	62,62,62,62	0
4	CL	A	406	1/1	0.97	0.17	55,55,55,55	0
4	CL	A	407	1/1	0.97	0.06	52,52,52,52	0
3	IOD	A	403	1/1	0.97	0.04	68,68,68,68	1
2	NA	A	401	1/1	0.98	0.09	53,53,53,53	0
4	CL	A	410[B]	1/1	0.98	0.18	37,37,37,37	1
4	CL	A	410[A]	1/1	0.98	0.18	35,35,35,35	1
3	IOD	A	404	1/1	0.98	0.04	62,62,62,62	1
4	CL	B	404	1/1	0.98	0.05	51,51,51,51	0
4	CL	A	408	1/1	0.98	0.06	58,58,58,58	0
3	IOD	B	401	1/1	0.99	0.09	50,50,50,50	1
3	IOD	A	402	1/1	0.99	0.09	51,51,51,51	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.