

# Full wwPDB X-ray Structure Validation Report (i)

#### Oct 14, 2023 – 10:10 PM EDT

PDB ID	:	6W7A
Title	:	The crystal structure of the 2009/H1N1/California PA endonuclease mutant
		E119D bound to DNA oligomer TAGCAT (uncleaved, 5mM overnight DNA
		soak)
Authors	:	Cuypers, M.G.; Kumar, G.; Webb, T.; White, S.W.
Deposited on		
Resolution	:	2.09 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 (i) 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 (i)) were used in the production of this report:

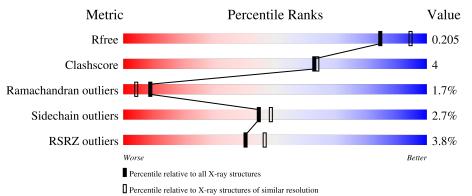
Xtriage (Phenix) EDS buster-report Percentile statistics Refmac CCP4 Ideal geometry (proteins)	: : : : :	1.8.5 (274361), CSD as541be (2020) 1.13 2.36 1.1.7 (2018) 20191225.v01 (using entries in the PDB archive December 25th 2019) 5.8.0158 7.0.044 (Gargrove) Engh & Huber (2001)
Ideal geometry (DNA, RNA)		0
Validation Pipeline (wwPDB-VP)	:	2.36

# 1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:  $X\text{-}RAY \, DIFFRACTION$ 

The reported resolution of this entry is 2.09 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	$egin{array}{c} { m Whole \ archive} \ (\#{ m Entries}) \end{array}$	${f Similar\ resolution}\ (\#{ m Entries,\ resolution\ range}({ m \AA}))$
$R_{free}$	130704	5197(2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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 <=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	А	197	3%	81%		7% • 10%
2	В	8	12% 25%	25%	25%	25%



#### 6W7A

## 2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 1690 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 Protein PA-X,Polymerase acidic protein.

Mol	Chain	Residues		$\mathbf{A}^{\dagger}$	toms			ZeroOcc	AltConf	Trace
1	А	178	Total 1438	C 913	N 244	0 270	S 11	0	0	0

Chain	Residue	Modelled	Actual	Comment	Reference
А	-19	MET	-	expression tag	UNP A0A4P2TE19
А	-18	GLY	_	expression tag	UNP A0A4P2TE19
А	-17	SER	_	expression tag	UNP A0A4P2TE19
А	-16	SER	-	expression tag	UNP A0A4P2TE19
А	-15	HIS	-	expression tag	UNP A0A4P2TE19
А	-14	HIS	-	expression tag	UNP A0A4P2TE19
А	-13	HIS	-	expression tag	UNP A0A4P2TE19
А	-12	HIS	-	expression tag	UNP A0A4P2TE19
А	-11	HIS	-	expression tag	UNP A0A4P2TE19
А	-10	HIS	-	expression tag	UNP A0A4P2TE19
А	-9	SER	-	expression tag	UNP A0A4P2TE19
А	-8	SER	-	expression tag	UNP A0A4P2TE19
А	-7	GLY	-	expression tag	UNP A0A4P2TE19
А	-6	LEU	-	expression tag	UNP A0A4P2TE19
А	-5	VAL	-	expression tag	UNP A0A4P2TE19
А	-4	PRO	-	expression tag	UNP A0A4P2TE19
А	-3	ARG	-	expression tag	UNP A0A4P2TE19
А	-2	GLY	-	expression tag	UNP A0A4P2TE19
А	-1	SER	-	expression tag	UNP A0A4P2TE19
А	0	HIS	-	expression tag	UNP A0A4P2TE19
А	51	GLY	-	linker	UNP A0A4P2TE19
А	52	GLY	-	linker	UNP A0A4P2TE19
А	53	SER	-	linker	UNP A0A4P2TE19
А	119	ASP	GLU	engineered mutation	UNP C6H0Y9

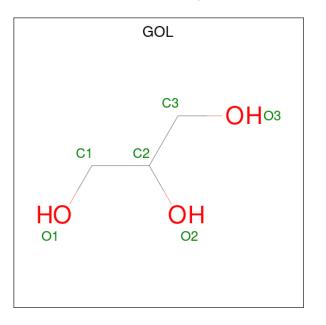
There are 24 discrepancies between the modelled and reference sequences:

• Molecule 2 is a DNA chain called DNA (5'-D(P\*TP\*AP\*GP\*CP\*AP\*T)-3').



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
2	В	6	Total	С	Ν	0	Р	0	0	0
			123	59	22	36	6			5

• Molecule 3 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



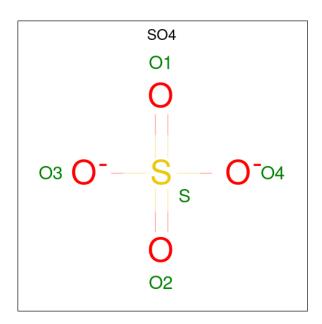
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	А	1	$\begin{array}{ccc} \text{Total} & \text{C} & \text{O} \\ 6 & 3 & 3 \end{array}$	0	0
3	А	1	$\begin{array}{ccc} \text{Total} & \text{C} & \text{O} \\ 6 & 3 & 3 \end{array}$	0	0

• Molecule 4 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	А	1	Total Mn 1 1	0	0

• Molecule 5 is SULFATE ION (three-letter code: SO4) (formula:  $O_4S$ ).





Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	А	1	Total 5	0 4	S 1	0	0

• Molecule 6 is water.

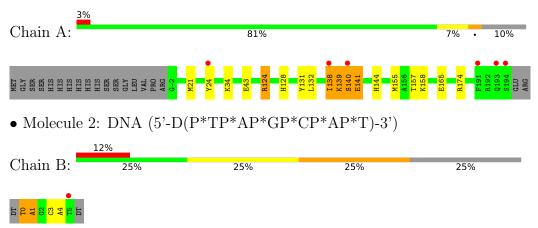
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	А	99	Total O 99 99	0	0
6	В	12	Total O 12 12	0	0



# 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 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 (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.

• Molecule 1: Protein PA-X,Polymerase acidic protein





## 4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 4 2 2	Depositor
Cell constants	89.30Å 89.30Å 134.02Å	Depositor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	$90.00^{\circ}$ $90.00^{\circ}$ $90.00^{\circ}$	Depositor
Resolution (Å)	39.95 - 2.09	Depositor
Resolution (A)	44.65 - 2.09	EDS
% Data completeness	98.8 (39.95-2.09)	Depositor
(in resolution range)	98.8 (44.65-2.09)	EDS
R <sub>merge</sub>	0.08	Depositor
R <sub>sym</sub>	(Not available)	Depositor
$< I/\sigma(I) > 1$	$3.04 (at 2.08 \text{\AA})$	Xtriage
Refinement program	PHENIX 1.17.1_3660	Depositor
D D.	0.174 , $0.205$	Depositor
$R, R_{free}$	0.175 , $0.205$	DCC
$R_{free}$ test set	818 reflections $(5.02\%)$	wwPDB-VP
Wilson B-factor $(Å^2)$	37.4	Xtriage
Anisotropy	0.455	Xtriage
Bulk solvent $k_{sol}(e/Å^3), B_{sol}(Å^2)$	0.36 , $62.5$	EDS
L-test for twinning <sup>2</sup>	$ \langle L  \rangle = 0.48, \langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	1690	wwPDB-VP
Average B, all atoms $(Å^2)$	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 5.11% of the height of the origin peak. No significant pseudotranslation is detected.

<sup>&</sup>lt;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.



<sup>&</sup>lt;sup>1</sup>Intensities estimated from amplitudes.

# 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, MN, GOL

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		Bo	nd lengths	Bond angles		
IVIOI	Chain	RMSZ	# Z  > 5	RMSZ	# Z  > 5	
1	А	0.83	1/1469~(0.1%)	0.90	2/1978~(0.1%)	
2	В	0.66	0/137	1.15	3/209~(1.4%)	
All	All	0.82	1/1606~(0.1%)	0.92	5/2187~(0.2%)	

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	$\operatorname{Ideal}(\operatorname{\AA})$
1	А	43	GLU	CD-OE1	-5.52	1.19	1.25

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Ζ	$\mathbf{Observed}(^{o})$	$\operatorname{Ideal}(^{o})$
2	В	0	DT	C1'-O4'-C4'	-5.52	104.58	110.10
2	В	4	DA	C1'-O4'-C4'	-5.34	104.76	110.10
1	А	174	ARG	NE-CZ-NH2	-5.32	117.64	120.30
2	В	1	DA	O5'-P-OP2	-5.12	101.09	105.70
1	А	124	ARG	NE-CZ-NH2	5.08	122.84	120.30

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	А	1438	0	1373	12	0
2	В	123	0	69	5	0
3	А	12	0	16	0	0
4	А	1	0	0	0	0
5	А	5	0	0	0	0
6	А	99	0	0	1	0
6	В	12	0	0	1	0
All	All	1690	0	1458	13	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 4.

All (13) 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:34:LYS:HD2	2:B:1:DA:C6	2.26	0.70
2:B:3:DC:N3	6:B:101:HOH:O	2.26	0.69
1:A:21:MET:HA	2:B:0:DT:H72	1.90	0.54
1:A:157:THR:HG22	1:A:158:LYS:HG3	1.92	0.51
1:A:128:HIS:HD2	6:A:373:HOH:O	1.93	0.50
1:A:144:HIS:CE1	1:A:157:THR:HB	2.46	0.50
1:A:34:LYS:HD2	2:B:1:DA:N6	2.27	0.49
1:A:138:ILE:O	1:A:140:SER:N	2.46	0.49
1:A:139:LYS:O	1:A:141:GLU:N	2.46	0.48
1:A:131:TYR:C	1:A:131:TYR:CD1	2.88	0.47
1:A:131:TYR:CG	1:A:155:MET:HE2	2.51	0.45
1:A:24:TYR:CB	2:B:0:DT:C4	3.03	0.42
1:A:131:TYR:HB3	1:A:155:MET:HE1	2.02	0.40

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	nalysed Favoured		Outliers	Percentiles	
1	А	176/197~(89%)	167 (95%)	6 (3%)	3~(2%)	9 4	

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	А	139	LYS
1	А	140	SER
1	А	141	GLU

#### 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.

Mo	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	А	150/177~(85%)	146~(97%)	4(3%)	44 48	

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	$\operatorname{Res}$	Type
1	А	124	ARG
1	А	132	LEU
1	А	138	ILE
1	А	165	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	А	92	ASN
1	А	136	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)

Of 4 ligands modelled in this entry, 1 is monoatomic - leaving 3 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		Tink	Bond lengths			Bond angles		
	Type	Unam	nes		Counts	RMSZ	# Z >2	Counts	RMSZ	# Z  > 2
3	GOL	А	201	-	$5,\!5,\!5$	0.13	0	$5,\!5,\!5$	0.33	0
3	GOL	А	202	-	$5,\!5,\!5$	0.13	0	$5,\!5,\!5$	0.35	0
5	SO4	А	204	-	4,4,4	0.31	0	$6,\!6,\!6$	0.14	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	GOL	А	201	-	-	0/4/4/4	-
3	GOL	А	202	-	-	0/4/4/4	-

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,  $95^{th}$  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	$\langle RSRZ \rangle$	#RSRZ>2	$\mathbf{OWAB}(\mathrm{\AA}^2)$	Q < 0.9
1	А	178/197~(90%)	0.22	6 (3%) 45 51	33, 49, 98, 124	0
2	В	6/8~(75%)	0.95	1 (16%) 1 2	60, 69, 134, 138	0
All	All	184/205~(89%)	0.24	7 (3%) 40 46	33, 50, 105, 138	0

All (7) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	А	138	ILE	7.9
1	А	191	PHE	3.3
1	А	194	SER	3.1
1	А	193	GLN	3.0
1	А	24	TYR	2.8
1	А	140	SER	2.7
2	В	5	DT	2.3

### 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 monosaccharides in this entry.

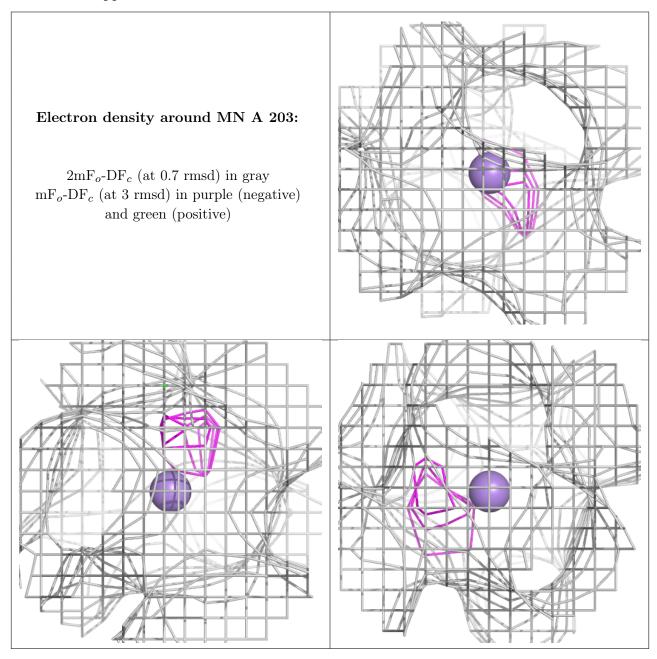
## 6.4 Ligands (i)

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,  $95^{th}$  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	$B-factors(Å^2)$	Q<0.9
3	GOL	А	202	6/6	0.73	0.29	78,82,85,96	0
5	SO4	А	204	5/5	0.81	0.15	103,108,112,114	0
3	GOL	А	201	6/6	0.85	0.25	65,78,81,82	0
4	MN	А	203	1/1	0.99	0.10	40,40,40,40	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.





## 6.5 Other polymers (i)

There are no such residues in this entry.

