



# Full wwPDB X-ray Structure Validation Report ⓘ

May 29, 2020 – 08:05 am BST

PDB ID : 5T4I  
Title : A Novel domain in human EXOG converts apoptotic endonuclease to DNA-repair enzyme  
Authors : Szymanski, M.R.; Yin, W.Y.  
Deposited on : 2016-08-29  
Resolution : 2.39 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](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
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

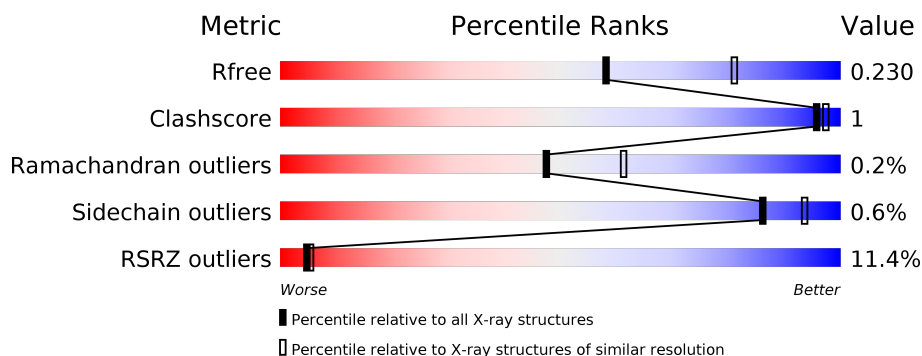
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.39 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	5509 (2.40-2.36)
Clashscore	141614	6082 (2.40-2.36)
Ramachandran outliers	138981	5973 (2.40-2.36)
Sidechain outliers	138945	5975 (2.40-2.36)
RSRZ outliers	127900	5397 (2.40-2.36)

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  $\geq 3$ , 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	317	<div> <div>12%</div> <div>90%</div> <div>6%</div> </div>
1	B	317	<div> <div>8%</div> <div>92%</div> <div>6%</div> </div>
2	C	9	<div> <div>33%</div> <div>89%</div> <div>11%</div> </div>
2	E	9	<div> <div>22%</div> <div>89%</div> <div>11%</div> </div>
3	D	9	<div> <div>11%</div> <div>100%</div> </div>
3	F	9	<div> <div>22%</div> <div>100%</div> </div>

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 10661 atoms, of which 5101 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 Nuclease EXOG, mitochondrial.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	297	Total	C	H	N	O	S	0	0	0
			4721	1516	2340	405	449	11			
1	B	298	Total	C	H	N	O	S	0	0	0
			4743	1522	2353	407	450	11			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	58	MET	-	initiating methionine	UNP Q9Y2C4
A	140	ALA	HIS	engineered mutation	UNP Q9Y2C4
A	369	HIS	-	expression tag	UNP Q9Y2C4
A	370	HIS	-	expression tag	UNP Q9Y2C4
A	371	HIS	-	expression tag	UNP Q9Y2C4
A	372	HIS	-	expression tag	UNP Q9Y2C4
A	373	HIS	-	expression tag	UNP Q9Y2C4
A	374	HIS	-	expression tag	UNP Q9Y2C4
B	58	MET	-	initiating methionine	UNP Q9Y2C4
B	140	ALA	HIS	engineered mutation	UNP Q9Y2C4
B	369	HIS	-	expression tag	UNP Q9Y2C4
B	370	HIS	-	expression tag	UNP Q9Y2C4
B	371	HIS	-	expression tag	UNP Q9Y2C4
B	372	HIS	-	expression tag	UNP Q9Y2C4
B	373	HIS	-	expression tag	UNP Q9Y2C4
B	374	HIS	-	expression tag	UNP Q9Y2C4

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	C	9	Total	C	H	N	O	P	0	0	0
			287	87	102	33	56	9			
2	E	9	Total	C	H	N	O	P	0	0	0
			287	87	102	33	56	9			

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	D	9	Total	C	H	N	O	P	0	0	0
			284	87	102	36	51	8			
3	F	9	Total	C	H	N	O	P	0	0	0
			284	87	102	36	51	8			

- Molecule 4 is MANGANESE (II) ION (three-letter code: MN) (formula: Mn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total	Mn	0	0
			1	1		
4	A	1	Total	Mn	0	0
			1	1		

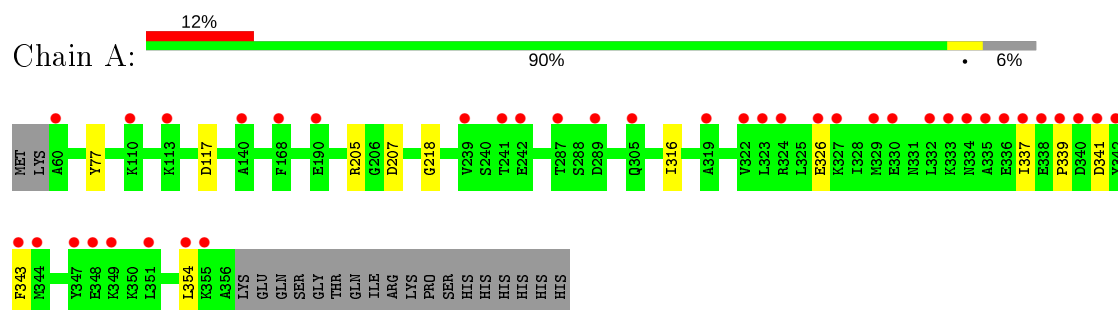
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	24	Total	O	0	0
			24	24		
5	B	29	Total	O	0	0
			29	29		

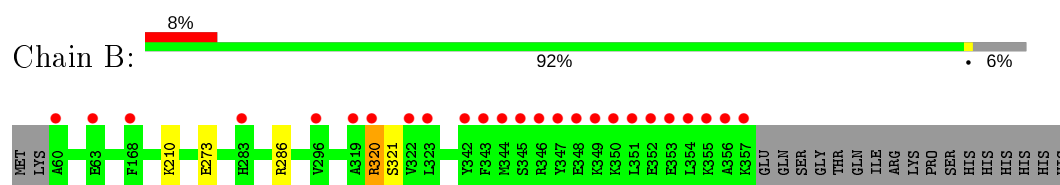
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA 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.

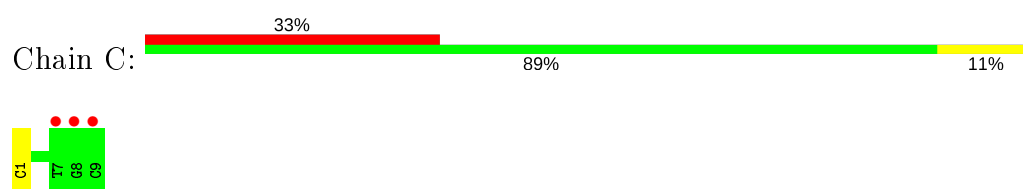
- Molecule 1: Nuclease EXOG, mitochondrial



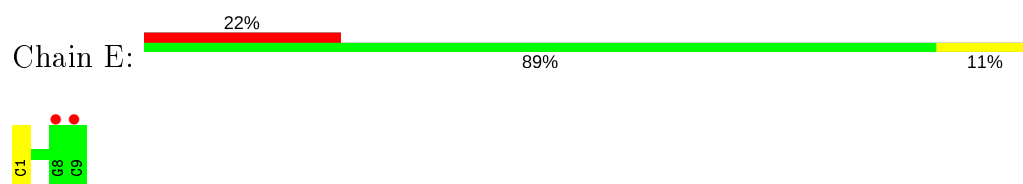
- Molecule 1: Nuclease EXOG, mitochondrial



- Molecule 2: DNA (5'-D(P\*CP\*TP\*GP\*AP\*CP\*GP\*TP\*GP\*C)-3')



- Molecule 2: DNA (5'-D(P\*CP\*TP\*GP\*AP\*CP\*GP\*TP\*GP\*C)-3')



- Molecule 3: DNA (5'-D(\*GP\*CP\*AP\*CP\*GP\*TP\*CP\*AP\*G)-3')





- Molecule 3: DNA (5'-D(\*GP\*CP\*AP\*CP\*GP\*TP\*CP\*AP\*G)-3')



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	73.51Å 79.72Å 138.87Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.03 – 2.39 40.03 – 2.39	Depositor EDS
% Data completeness (in resolution range)	98.4 (40.03-2.39) 97.7 (40.03-2.39)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.68 (at 2.39Å)	Xtriage
Refinement program	PHENIX 1.10.1 _2155	Depositor
R, $R_{free}$	0.191 , 0.229 0.192 , 0.230	Depositor DCC
$R_{free}$ test set	1993 reflections (6.07%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	45.5	Xtriage
Anisotropy	0.411	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.40 , 43.7	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	0.94	EDS
Total number of atoms	10661	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	67.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.02% 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

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: MN

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	0.26	0/2437	0.43	0/3298
1	B	0.25	0/2446	0.41	0/3309
2	C	0.86	1/206 (0.5%)	0.82	0/314
2	E	0.87	1/206 (0.5%)	0.82	0/314
3	D	0.46	0/204	0.77	0/313
3	F	0.46	0/204	0.76	0/313
All	All	0.35	2/5703 (0.0%)	0.50	0/7861

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	E	1	DC	OP3-P	-10.46	1.48	1.61
2	C	1	DC	OP3-P	-10.39	1.48	1.61

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	341	ASP	Peptide



## 5.2 Too-close contacts ⓘ

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	2381	2340	2340	6	0
1	B	2390	2353	2353	3	0
2	C	185	102	102	0	0
2	E	185	102	102	0	0
3	D	182	102	102	0	0
3	F	182	102	102	0	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
5	A	24	0	0	0	0
5	B	29	0	0	0	0
All	All	5560	5101	5101	8	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 1.

All (8) 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:339:PRO:CB	1:A:343:PHE:HB3	2.38	0.54
1:A:205:ARG:NH2	1:A:207:ASP:OD2	2.42	0.51
1:A:339:PRO:HB2	1:A:343:PHE:HB3	1.94	0.50
1:A:77:TYR:OH	1:A:117:ASP:OD2	2.24	0.47
1:B:273:GLU:OE1	1:B:286:ARG:NH2	2.38	0.46
1:A:218:GLY:O	1:B:210:LYS:NZ	2.47	0.46
1:A:316:ILE:HD11	1:A:343:PHE:CZ	2.53	0.44
1:B:320:ARG:HD2	1:B:321:SER:N	2.36	0.41

There are no symmetry-related clashes.

## 5.3 Torsion angles ⓘ

### 5.3.1 Protein backbone ⓘ

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	295/317 (93%)	289 (98%)	5 (2%)	1 (0%)	41	53
1	B	296/317 (93%)	291 (98%)	5 (2%)	0	100	100
All	All	591/634 (93%)	580 (98%)	10 (2%)	1 (0%)	47	61

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	337	ILE

### 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	262/281 (93%)	260 (99%)	2 (1%)	81	91
1	B	263/281 (94%)	262 (100%)	1 (0%)	91	96
All	All	525/562 (93%)	522 (99%)	3 (1%)	86	93

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	326	GLU
1	A	354	LEU
1	B	320	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

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 2 ligands modelled in this entry, 2 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	297/317 (93%)	0.85	39 (13%) 3 3	31, 51, 115, 165	0
1	B	298/317 (94%)	0.70	25 (8%) 11 12	29, 50, 98, 125	0
2	C	9/9 (100%)	1.46	3 (33%) 0 0	43, 60, 130, 149	0
2	E	9/9 (100%)	1.57	2 (22%) 0 0	44, 60, 133, 160	0
3	D	9/9 (100%)	1.06	1 (11%) 5 6	65, 104, 131, 134	0
3	F	9/9 (100%)	0.76	2 (22%) 0 0	70, 88, 150, 153	0
All	All	631/670 (94%)	0.80	72 (11%) 5 5	29, 51, 115, 165	0

All (72) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	340	ASP	10.2
1	B	357	LYS	8.4
1	A	335	ALA	7.6
1	A	341	ASP	7.2
1	A	333	LYS	6.1
1	A	338	GLU	5.8
1	A	332	LEU	5.7
1	B	355	LYS	5.3
1	A	337	ILE	5.1
1	A	344	MET	5.1
1	B	60	ALA	4.9
2	E	8	DG	4.5
2	E	9	DC	4.4
1	A	287	THR	4.4
1	A	326	GLU	4.3
1	A	330	GLU	4.3
1	A	351	LEU	4.3
1	A	241	THR	4.0
2	C	8	DG	4.0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	336	GLU	3.9
1	A	347	TYR	3.8
1	A	323	LEU	3.8
1	A	329	MET	3.7
1	B	351	LEU	3.7
1	A	319	ALA	3.6
1	A	239	VAL	3.5
1	B	320	ARG	3.4
1	B	322	VAL	3.3
3	D	2	DG	3.3
1	B	347	TYR	3.3
1	B	319	ALA	3.2
1	B	346	ARG	3.2
3	F	2	DG	3.2
1	B	343	PHE	3.2
1	B	348	GLU	3.1
1	B	352	GLU	3.1
1	A	343	PHE	3.1
1	B	354	LEU	3.1
2	C	9	DC	3.1
2	C	7	DT	3.0
3	F	3	DC	3.0
1	A	334	ASN	3.0
1	A	168	PHE	2.9
1	A	355	LYS	2.9
1	B	342	TYR	2.8
1	A	327	LYS	2.6
1	A	289	ASP	2.6
1	A	339	PRO	2.6
1	A	110	LYS	2.6
1	B	349	LYS	2.6
1	B	350	LYS	2.5
1	B	168	PHE	2.5
1	A	140	ALA	2.4
1	A	322	VAL	2.4
1	B	296	VAL	2.4
1	B	344	MET	2.4
1	B	323	LEU	2.3
1	A	305	GLN	2.3
1	B	63	GLU	2.2
1	A	242	GLU	2.2
1	A	349	LYS	2.2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	A	113	LYS	2.2
1	B	345	SER	2.2
1	A	348	GLU	2.1
1	A	60	ALA	2.1
1	A	324	ARG	2.1
1	A	354	LEU	2.1
1	A	342	TYR	2.1
1	A	190	GLU	2.1
1	B	356	ALA	2.1
1	B	353	GLU	2.0
1	B	283	HIS	2.0

## 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	MN	A	401	1/1	0.96	0.25	59,59,59,59	0
4	MN	B	401	1/1	0.99	0.14	54,54,54,54	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.