



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 16, 2020 – 10:54 am BST

PDB ID : 3SQ3  
Title : Crystal Structure Analysis of the Yeast Tyrosyl-DNA Phosphodiesterase H182A Mutant  
Authors : Gajewski, S.; White, S.W.  
Deposited on : 2011-07-04  
Resolution : 2.50 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
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

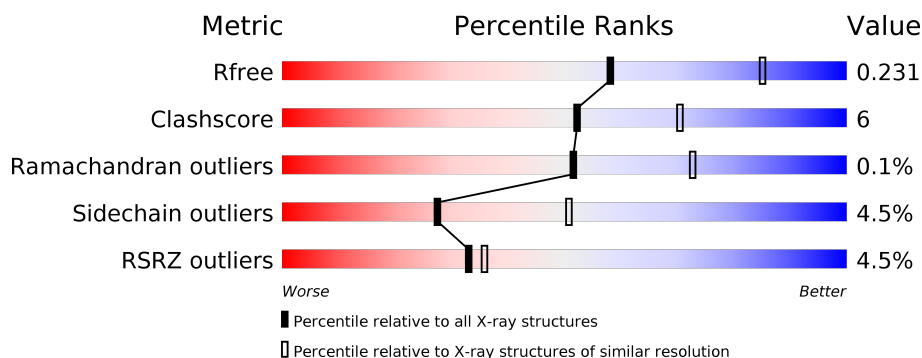
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.50 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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	470	<div> <div>4%</div> <div> <div></div> <div>74%</div> <div>14%</div> <div>•</div> <div>10%</div> </div> </div>
1	B	470	<div> <div>4%</div> <div> <div></div> <div>77%</div> <div>11%</div> <div>•</div> <div>10%</div> </div> </div>
1	C	470	<div> <div>3%</div> <div> <div></div> <div>72%</div> <div>16%</div> <div>•</div> <div>10%</div> </div> </div>
1	D	470	<div> <div>6%</div> <div> <div></div> <div>77%</div> <div>12%</div> <div>•</div> <div>10%</div> </div> </div>

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 13667 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 Tyrosyl-DNA phosphodiesterase 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	424	Total	C	N	O	S	0	0	0
			3434	2229	561	624	20			
1	B	423	Total	C	N	O	S	0	0	0
			3416	2218	557	621	20			
1	C	422	Total	C	N	O	S	0	0	0
			3407	2213	556	618	20			
1	D	424	Total	C	N	O	S	0	0	0
			3410	2216	554	620	20			

There are 40 discrepancies between the modelled and reference sequences:

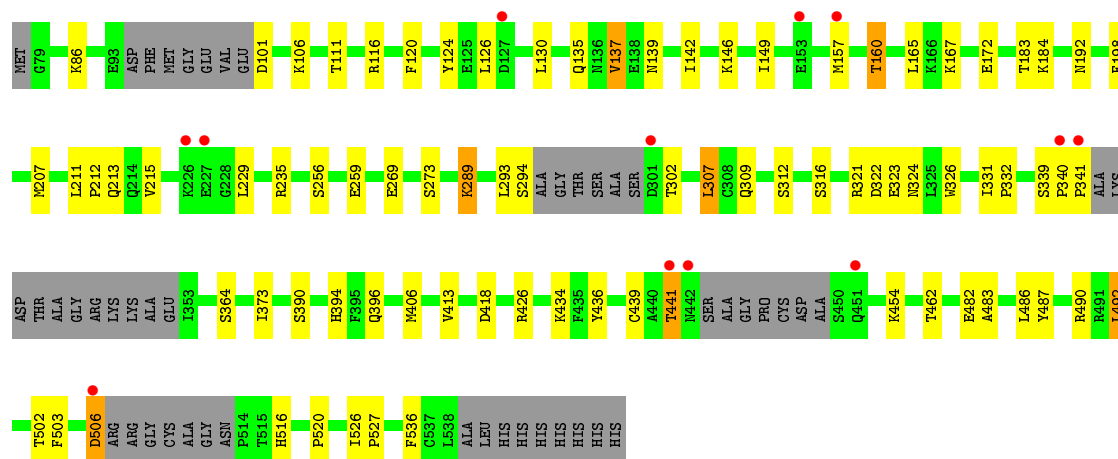
Chain	Residue	Modelled	Actual	Comment	Reference
A	78	MET	-	INITIATING METHIONINE	UNP P38319
A	182	ALA	HIS	ENGINEERED MUTATION	UNP P38319
A	540	LEU	-	EXPRESSION TAG	UNP P38319
A	541	HIS	-	EXPRESSION TAG	UNP P38319
A	542	HIS	-	EXPRESSION TAG	UNP P38319
A	543	HIS	-	EXPRESSION TAG	UNP P38319
A	544	HIS	-	EXPRESSION TAG	UNP P38319
A	545	HIS	-	EXPRESSION TAG	UNP P38319
A	546	HIS	-	EXPRESSION TAG	UNP P38319
A	547	HIS	-	EXPRESSION TAG	UNP P38319
B	78	MET	-	INITIATING METHIONINE	UNP P38319
B	182	ALA	HIS	ENGINEERED MUTATION	UNP P38319
B	540	LEU	-	EXPRESSION TAG	UNP P38319
B	541	HIS	-	EXPRESSION TAG	UNP P38319
B	542	HIS	-	EXPRESSION TAG	UNP P38319
B	543	HIS	-	EXPRESSION TAG	UNP P38319
B	544	HIS	-	EXPRESSION TAG	UNP P38319
B	545	HIS	-	EXPRESSION TAG	UNP P38319
B	546	HIS	-	EXPRESSION TAG	UNP P38319
B	547	HIS	-	EXPRESSION TAG	UNP P38319
C	78	MET	-	INITIATING METHIONINE	UNP P38319

*Continued on next page...*

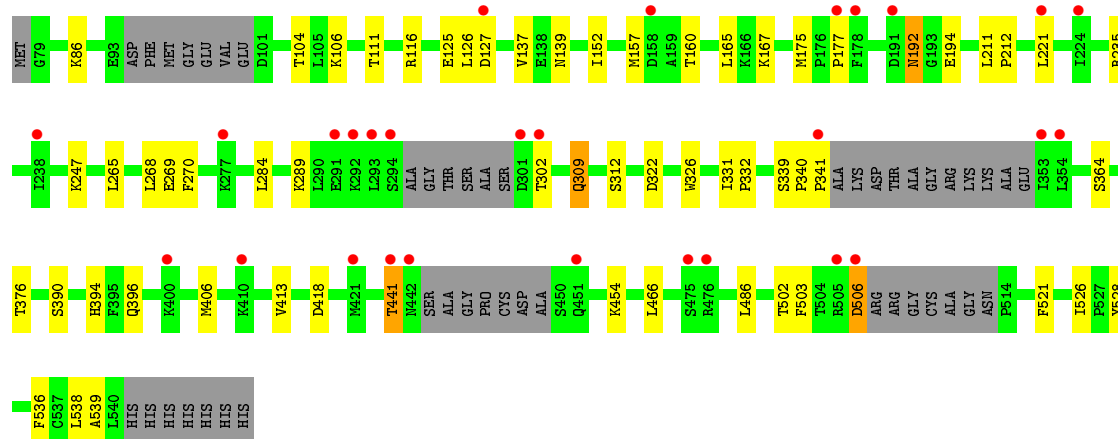
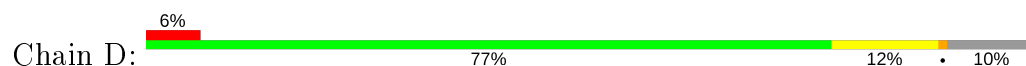
*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	182	ALA	HIS	ENGINEERED MUTATION	UNP P38319
C	540	LEU	-	EXPRESSION TAG	UNP P38319
C	541	HIS	-	EXPRESSION TAG	UNP P38319
C	542	HIS	-	EXPRESSION TAG	UNP P38319
C	543	HIS	-	EXPRESSION TAG	UNP P38319
C	544	HIS	-	EXPRESSION TAG	UNP P38319
C	545	HIS	-	EXPRESSION TAG	UNP P38319
C	546	HIS	-	EXPRESSION TAG	UNP P38319
C	547	HIS	-	EXPRESSION TAG	UNP P38319
D	78	MET	-	INITIATING METHIONINE	UNP P38319
D	182	ALA	HIS	ENGINEERED MUTATION	UNP P38319
D	540	LEU	-	EXPRESSION TAG	UNP P38319
D	541	HIS	-	EXPRESSION TAG	UNP P38319
D	542	HIS	-	EXPRESSION TAG	UNP P38319
D	543	HIS	-	EXPRESSION TAG	UNP P38319
D	544	HIS	-	EXPRESSION TAG	UNP P38319
D	545	HIS	-	EXPRESSION TAG	UNP P38319
D	546	HIS	-	EXPRESSION TAG	UNP P38319
D	547	HIS	-	EXPRESSION TAG	UNP P38319





• Molecule 1: Tyrosyl-DNA phosphodiesterase 1



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	64.15Å 81.79Å 98.58Å 86.92° 85.53° 67.10°	Depositor
Resolution (Å)	50.00 – 2.50 39.39 – 2.50	Depositor EDS
% Data completeness (in resolution range)	95.8 (50.00-2.50) 95.9 (39.39-2.50)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.64 (at 2.51Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, $R_{free}$	0.205 , 0.231 0.205 , 0.231	Depositor DCC
$R_{free}$ test set	3095 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	58.5	Xtriage
Anisotropy	0.043	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 48.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.006 for -h,-h+k,-l	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	13667	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	74.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.01% 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](#)

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	1.09	7/3521 (0.2%)	0.92	1/4765 (0.0%)
1	B	1.04	3/3503 (0.1%)	0.86	1/4742 (0.0%)
1	C	1.02	6/3494 (0.2%)	0.89	2/4730 (0.0%)
1	D	0.95	1/3497 (0.0%)	0.85	1/4737 (0.0%)
All	All	1.03	17/14015 (0.1%)	0.88	5/18974 (0.0%)

The worst 5 of 17 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	308	CYS	CB-SG	-9.22	1.66	1.82
1	C	487	TYR	CD1-CE1	5.95	1.48	1.39
1	C	439	CYS	CB-SG	-5.85	1.72	1.81
1	C	436	TYR	CG-CD2	5.82	1.46	1.39
1	B	239	GLU	CD-OE1	5.79	1.32	1.25

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	C	207	MET	CG-SD-CE	-7.07	88.90	100.20
1	C	101	ASP	CB-CG-OD2	5.68	123.41	118.30
1	A	101	ASP	CB-CG-OD2	5.13	122.92	118.30
1	B	290	LEU	CB-CG-CD1	-5.12	102.30	111.00
1	D	125	GLU	N-CA-C	-5.02	97.46	111.00

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	3434	0	3423	41	0
1	B	3416	0	3397	33	0
1	C	3407	0	3388	45	0
1	D	3410	0	3382	36	0
All	All	13667	0	13590	150	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 6.

The worst 5 of 150 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:B:89:PHE:HD2	1:C:229:LEU:HD21	1.32	0.90
1:B:89:PHE:CD2	1:C:229:LEU:HD21	2.13	0.82
1:B:394:HIS:CE1	1:B:396:GLN:HG2	2.21	0.75
1:A:503:PHE:O	1:A:506:ASP:HB2	1.88	0.73
1:B:322:ASP:O	1:B:406:MET:HE1	1.89	0.72

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	412/470 (88%)	391 (95%)	21 (5%)	0	100	100
1	B	411/470 (87%)	396 (96%)	14 (3%)	1 (0%)	47	68
1	C	410/470 (87%)	393 (96%)	17 (4%)	0	100	100
1	D	412/470 (88%)	395 (96%)	17 (4%)	0	100	100
All	All	1645/1880 (88%)	1575 (96%)	69 (4%)	1 (0%)	51	73

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	211	LEU

### 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	385/422 (91%)	367 (95%)	18 (5%)	26	49
1	B	382/422 (90%)	366 (96%)	16 (4%)	30	54
1	C	381/422 (90%)	362 (95%)	19 (5%)	24	46
1	D	380/422 (90%)	364 (96%)	16 (4%)	30	54
All	All	1528/1688 (90%)	1459 (96%)	69 (4%)	27	51

5 of 69 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	454	LYS
1	C	235	ARG
1	D	364	SER
1	B	492	LEU
1	C	111	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 24 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	288	ASN
1	C	139	ASN
1	D	309	GLN
1	B	394	HIS
1	C	135	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](#)

There are no ligands in this entry.

### 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	424/470 (90%)	0.12	17 (4%) 38 41	44, 68, 101, 108	0
1	B	423/470 (90%)	0.20	20 (4%) 31 33	45, 71, 106, 109	0
1	C	422/470 (89%)	0.15	12 (2%) 53 56	46, 72, 108, 112	0
1	D	424/470 (90%)	0.29	28 (6%) 18 19	49, 78, 116, 120	0
All	All	1693/1880 (90%)	0.19	77 (4%) 33 36	44, 72, 108, 120	0

The worst 5 of 77 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	341	PRO	8.1
1	B	442	ASN	7.4
1	D	442	ASN	7.2
1	C	442	ASN	7.2
1	A	442	ASN	7.1

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.