



wwPDB X-ray Structure Validation Summary Report

Feb 15, 2017 – 02:12 am GMT

PDB ID : 3SQ8
Title : Crystal Structure Analysis of the Yeast Tyrosyl-DNA Phosphodiesterase 1 H432R Mutant (SCAN1 Mutant)
Authors : Gajewski, S.; White, S.W.
Deposited on : 2011-07-05
Resolution : 2.10 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

A user guide is available at

<http://wwpdb.org/validation/2016/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 : 1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix) : 1.9-1692
EDS : trunk28620
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac : 5.8.0135
CCP4 : 6.5.0
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : recalc28949

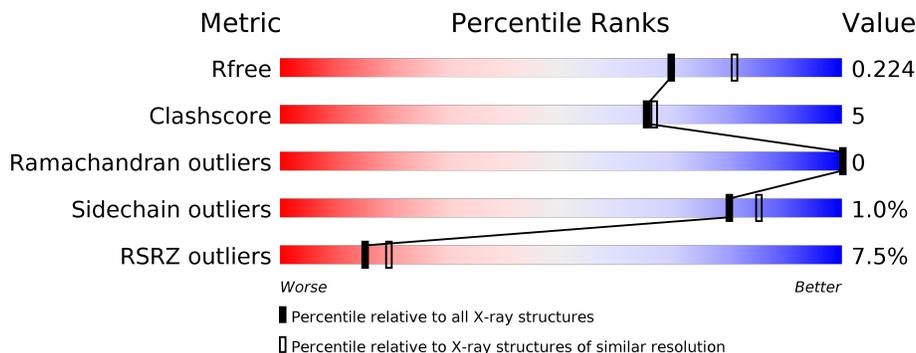
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.10 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}	100719	4243 (2.10-2.10)
Clashscore	112137	4788 (2.10-2.10)
Ramachandran outliers	110173	4740 (2.10-2.10)
Sidechain outliers	110143	4741 (2.10-2.10)
RSRZ outliers	101464	4275 (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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. 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	 5% 81% 9% 10%
1	B	470	 6% 80% 10% 10%
1	C	470	 8% 80% 10% 10%
1	D	470	 8% 80% 10% 10%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 14246 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
			Total	C	N	O	P	S			
1	A	422	3430	2223	564	622	1	20	0	0	0
1	B	423	3450	2235	570	624	1	20	0	1	0
1	C	423	3437	2227	565	624	1	20	0	0	0
1	D	425	3457	2241	569	626	1	20	0	1	0

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	78	MET	-	INITIATING METHIONINE	UNP P38319
A	432	ARG	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	432	ARG	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

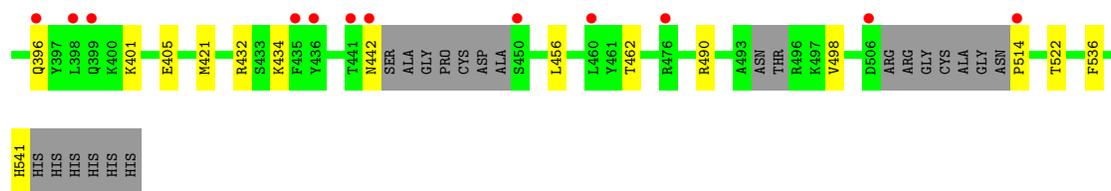
Continued on next page...

Continued from previous page...

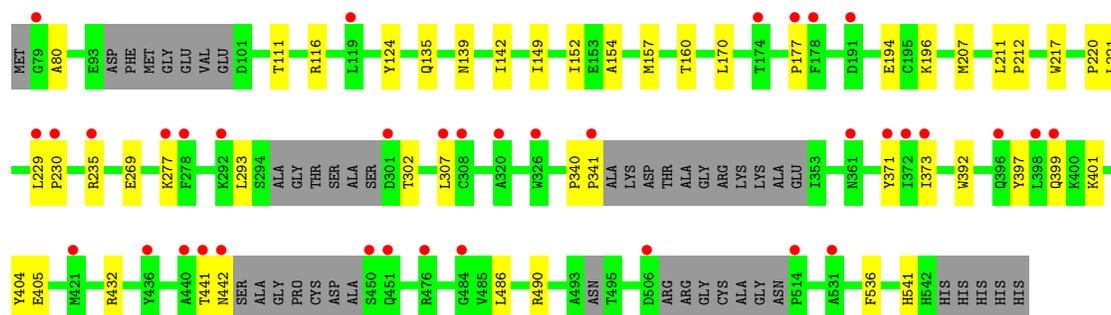
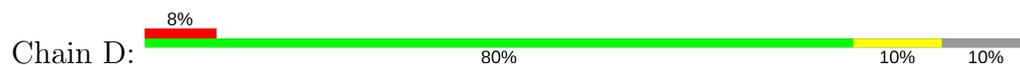
Chain	Residue	Modelled	Actual	Comment	Reference
C	432	ARG	HIS	ENGINEERED MUTATION	UNP P38319
C	540	LEU	-	EXPRESSION TAG	UNP P38319
C	541	HIS	-	EXPRESSION TAG	UNP P38319
C	542	HIS	-	EXPRESSION TAG	UNP P38319
C	543	HIS	-	EXPRESSION TAG	UNP P38319
C	544	HIS	-	EXPRESSION TAG	UNP P38319
C	545	HIS	-	EXPRESSION 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	432	ARG	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 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	168	Total O 168 168	0	0
2	B	146	Total O 146 146	0	0
2	C	95	Total O 95 95	0	0
2	D	63	Total O 63 63	0	0



- Molecule 1: Tyrosyl-DNA phosphodiesterase 1



4 Data and refinement statistics i

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	63.87Å 81.49Å 96.79Å 89.52° 84.90° 67.82°	Depositor
Resolution (Å)	24.79 – 2.10 24.79 – 2.10	Depositor EDS
% Data completeness (in resolution range)	93.4 (24.79-2.10) 93.0 (24.79-2.10)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.38 (at 2.10Å)	Xtrriage
Refinement program	REFMAC 5.6.0117	Depositor
R, R_{free}	0.187 , 0.226 0.183 , 0.224	Depositor DCC
R_{free} test set	4912 reflections (5.27%)	DCC
Wilson B-factor (Å ²)	34.7	Xtrriage
Anisotropy	0.259	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 42.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.003 for -h,-h+k,-l	Xtrriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	14246	wwPDB-VP
Average B, all atoms (Å ²)	49.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.82% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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: NEP

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	0.64	4/3501 (0.1%)	0.69	0/4734
1	B	0.65	2/3525 (0.1%)	0.68	0/4765
1	C	0.58	3/3508 (0.1%)	0.63	1/4743 (0.0%)
1	D	0.55	1/3532 (0.0%)	0.61	0/4776
All	All	0.60	10/14066 (0.1%)	0.66	1/19018 (0.0%)

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	326	TRP	CD2-CE2	5.68	1.48	1.41
1	A	458	TRP	CD2-CE2	5.60	1.48	1.41
1	B	458	TRP	CD2-CE2	5.43	1.47	1.41
1	A	217	TRP	CD2-CE2	5.38	1.47	1.41
1	A	392	TRP	CD2-CE2	5.36	1.47	1.41

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	432	ARG	NE-CZ-NH2	-5.51	117.54	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	A	3430	0	3419	28	0
1	B	3450	0	3436	46	0
1	C	3437	0	3418	29	0
1	D	3457	0	3438	41	0
2	A	168	0	0	2	0
2	B	146	0	0	2	0
2	C	95	0	0	3	0
2	D	63	0	0	1	0
All	All	14246	0	13711	128	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 128 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:207:MET:HE2	1:D:541:HIS:HE1	1.26	0.99
1:B:149:ILE:HD13	1:B:170:LEU:HD11	1.44	0.96
1:D:149:ILE:HD13	1:D:170:LEU:HD11	1.48	0.96
1:C:149:ILE:HD13	1:C:170:LEU:HD11	1.52	0.89
1:C:514:PRO:HD3	1:D:177:PRO:HG3	1.57	0.85

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	409/470 (87%)	398 (97%)	11 (3%)	0	100	100
1	B	409/470 (87%)	401 (98%)	8 (2%)	0	100	100
1	C	408/470 (87%)	398 (98%)	10 (2%)	0	100	100
1	D	411/470 (87%)	403 (98%)	8 (2%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
All	All	1637/1880 (87%)	1600 (98%)	37 (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	383/422 (91%)	378 (99%)	5 (1%)	73	80
1	B	386/422 (92%)	383 (99%)	3 (1%)	85	89
1	C	384/422 (91%)	380 (99%)	4 (1%)	80	85
1	D	386/422 (92%)	382 (99%)	4 (1%)	80	85
All	All	1539/1688 (91%)	1523 (99%)	16 (1%)	80	85

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	399	GLN
1	C	106	LYS
1	D	111	THR
1	B	371	TYR
1	D	196	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 18 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	135	GLN
1	C	139	ASN
1	D	139	ASN
1	B	288	ASN
1	B	541	HIS

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

4 non-standard protein/DNA/RNA residues 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
1	NEP	A	182	1	11,14,15	2.14	5 (45%)	7,20,22	2.49	2 (28%)
1	NEP	B	182	1	11,14,15	2.43	6 (54%)	7,20,22	2.60	4 (57%)
1	NEP	C	182	1	11,14,15	2.36	5 (45%)	7,20,22	2.09	1 (14%)
1	NEP	D	182	1	11,14,15	2.18	5 (45%)	7,20,22	2.26	1 (14%)

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
1	NEP	A	182	1	-	0/4/12/14	0/1/1/1
1	NEP	B	182	1	-	0/4/12/14	0/1/1/1
1	NEP	C	182	1	-	0/4/12/14	0/1/1/1
1	NEP	D	182	1	-	0/4/12/14	0/1/1/1

The worst 5 of 21 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	182	NEP	CD2-NE2	-4.03	1.32	1.39
1	D	182	NEP	CD2-NE2	-3.61	1.32	1.39
1	C	182	NEP	CD2-NE2	-2.99	1.33	1.39
1	A	182	NEP	CD2-NE2	-2.79	1.34	1.39
1	B	182	NEP	P-O1P	2.01	1.58	1.54

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	182	NEP	CG-CD2-NE2	-5.09	102.15	108.96
1	B	182	NEP	CG-CD2-NE2	-4.61	102.80	108.96
1	A	182	NEP	CG-CD2-NE2	-4.57	102.85	108.96
1	C	182	NEP	CG-CD2-NE2	-4.23	103.30	108.96
1	B	182	NEP	O1P-P-O3P	-2.25	105.60	112.30

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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, 95th 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 ²)	Q<0.9
1	A	421/470 (89%)	0.03	24 (5%) 24 30	25, 39, 68, 104	0
1	B	422/470 (89%)	0.12	27 (6%) 20 25	25, 40, 72, 103	0
1	C	422/470 (89%)	0.40	39 (9%) 10 12	32, 51, 81, 132	0
1	D	424/470 (90%)	0.44	37 (8%) 11 14	33, 54, 89, 142	0
All	All	1689/1880 (89%)	0.25	127 (7%) 15 19	25, 46, 80, 142	0

The worst 5 of 127 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	341	PRO	10.6
1	D	341	PRO	10.6
1	D	442	ASN	9.3
1	D	178	PHE	7.2
1	C	442	ASN	6.3

6.2 Non-standard residues in protein, DNA, RNA chains [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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th 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	LLDF	B-factors(Å ²)	Q<0.9
1	NEP	A	182	14/15	0.97	0.10	-	27,31,32,39	4
1	NEP	D	182	14/15	0.96	0.08	-	38,40,41,47	4
1	NEP	B	182	14/15	0.96	0.09	-	27,29,31,37	4
1	NEP	C	182	14/15	0.96	0.09	-	34,35,37,42	4

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.