Full wwPDB X-ray Structure Validation Report ⓘ

May 14, 2020 – 08:59 am BST

PDB ID	:	5NWA
Title	:	Crystal structure of the complex of Tdp1 with duplex DNA
Authors	:	Richardson, J.M.; Ruksenaite, E.; Morris, E.R.
Deposited on	:	2017-05-05
Resolution	:	3.20 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 ⓘ symbol.

The following versions of software and data (see references ①) were used in the production of this report:

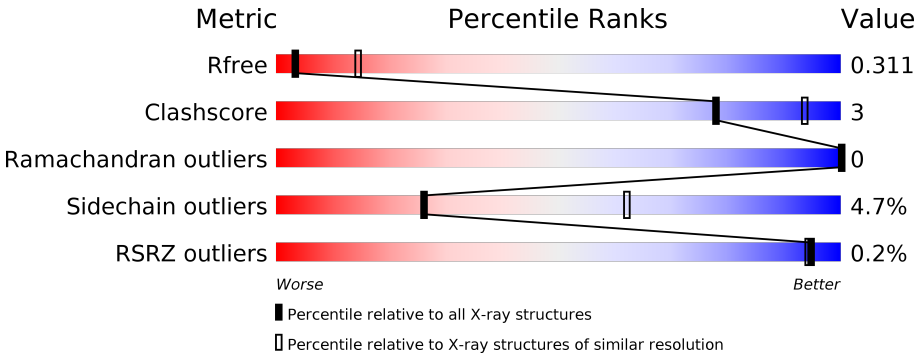
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.20 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	1133 (3.20-3.20)
Clashscore	141614	1253 (3.20-3.20)
Ramachandran outliers	138981	1234 (3.20-3.20)
Sidechain outliers	138945	1233 (3.20-3.20)
RSRZ outliers	127900	1095 (3.20-3.20)

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 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	A	485	80% 10% 9%
1	B	485	78% 11% 11%
2	C	11	82% 18%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 7186 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	441	Total	C	N	O	S	0	0	0
			3531	2292	596	631	12			
1	B	432	Total	C	N	O	S	0	0	0
			3456	2241	585	619	11			

There are 56 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	124	MET	-	initiating methionine	UNP Q9NUW8
A	125	GLY	-	expression tag	UNP Q9NUW8
A	126	SER	-	expression tag	UNP Q9NUW8
A	127	SER	-	expression tag	UNP Q9NUW8
A	128	HIS	-	expression tag	UNP Q9NUW8
A	129	HIS	-	expression tag	UNP Q9NUW8
A	130	HIS	-	expression tag	UNP Q9NUW8
A	131	HIS	-	expression tag	UNP Q9NUW8
A	132	HIS	-	expression tag	UNP Q9NUW8
A	133	HIS	-	expression tag	UNP Q9NUW8
A	134	SER	-	expression tag	UNP Q9NUW8
A	135	SER	-	expression tag	UNP Q9NUW8
A	136	GLY	-	expression tag	UNP Q9NUW8
A	137	LEU	-	expression tag	UNP Q9NUW8
A	138	VAL	-	expression tag	UNP Q9NUW8
A	139	PRO	-	expression tag	UNP Q9NUW8
A	140	ARG	-	expression tag	UNP Q9NUW8
A	141	GLY	-	expression tag	UNP Q9NUW8
A	142	SER	-	expression tag	UNP Q9NUW8
A	143	HIS	-	expression tag	UNP Q9NUW8
A	144	MET	-	expression tag	UNP Q9NUW8
A	145	LEU	-	expression tag	UNP Q9NUW8
A	146	GLU	-	expression tag	UNP Q9NUW8
A	147	ASP	-	expression tag	UNP Q9NUW8
A	148	PRO	-	expression tag	UNP Q9NUW8

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	322	ASN	ASP	conflict	UNP Q9NUW8
A	328	THR	MET	conflict	UNP Q9NUW8
A	548	LEU	PHE	conflict	UNP Q9NUW8
B	124	MET	-	initiating methionine	UNP Q9NUW8
B	125	GLY	-	expression tag	UNP Q9NUW8
B	126	SER	-	expression tag	UNP Q9NUW8
B	127	SER	-	expression tag	UNP Q9NUW8
B	128	HIS	-	expression tag	UNP Q9NUW8
B	129	HIS	-	expression tag	UNP Q9NUW8
B	130	HIS	-	expression tag	UNP Q9NUW8
B	131	HIS	-	expression tag	UNP Q9NUW8
B	132	HIS	-	expression tag	UNP Q9NUW8
B	133	HIS	-	expression tag	UNP Q9NUW8
B	134	SER	-	expression tag	UNP Q9NUW8
B	135	SER	-	expression tag	UNP Q9NUW8
B	136	GLY	-	expression tag	UNP Q9NUW8
B	137	LEU	-	expression tag	UNP Q9NUW8
B	138	VAL	-	expression tag	UNP Q9NUW8
B	139	PRO	-	expression tag	UNP Q9NUW8
B	140	ARG	-	expression tag	UNP Q9NUW8
B	141	GLY	-	expression tag	UNP Q9NUW8
B	142	SER	-	expression tag	UNP Q9NUW8
B	143	HIS	-	expression tag	UNP Q9NUW8
B	144	MET	-	expression tag	UNP Q9NUW8
B	145	LEU	-	expression tag	UNP Q9NUW8
B	146	GLU	-	expression tag	UNP Q9NUW8
B	147	ASP	-	expression tag	UNP Q9NUW8
B	148	PRO	-	expression tag	UNP Q9NUW8
B	322	ASN	ASP	conflict	UNP Q9NUW8
B	328	THR	MET	conflict	UNP Q9NUW8
B	548	LEU	PHE	conflict	UNP Q9NUW8

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

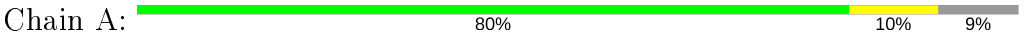
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	C	11	199	93	32	63	11	0	0	0

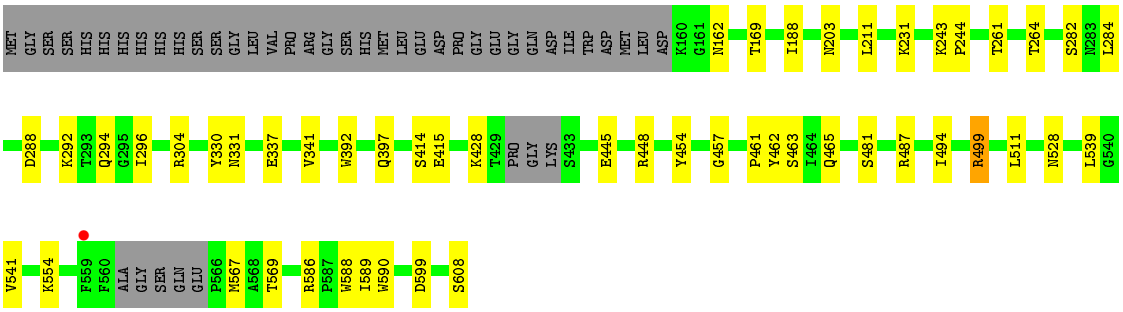
3 Residue-property plots (i)

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: Tyrosyl-DNA phosphodiesterase 1

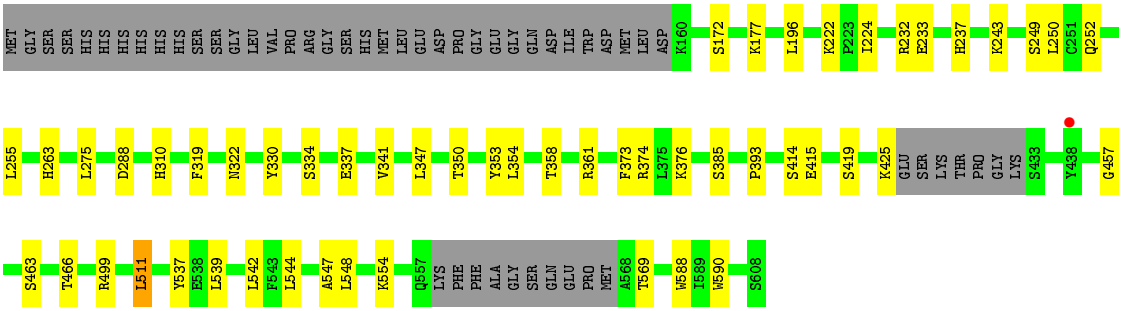
Chain A:





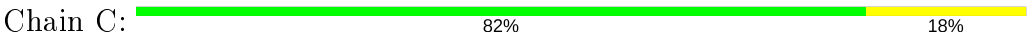
- Molecule 1: Tyrosyl-DNA phosphodiesterase 1

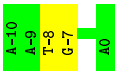
Chain B:



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

Chain C:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	109.19Å 198.19Å 51.33Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	99.10 – 3.20 99.10 – 3.20	Depositor EDS
% Data completeness (in resolution range)	98.7 (99.10-3.20) 98.7 (99.10-3.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.59 (at 3.19Å)	Xtriage
Refinement program	REFMAC 5.8.0107	Depositor
R, R_{free}	0.250 , 0.306 0.259 , 0.311	Depositor DCC
R_{free} test set	952 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å^2)	68.1	Xtriage
Anisotropy	0.290	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.25 , 26.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.90	EDS
Total number of atoms	7186	wwPDB-VP
Average B, all atoms (Å^2)	74.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.95% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ 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

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	0.47	0/3646	0.68	1/4950 (0.0%)
1	B	0.48	0/3568	0.67	0/4848
2	C	0.37	0/221	0.83	0/339
All	All	0.47	0/7435	0.68	1/10137 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	304	ARG	NE-CZ-NH1	5.23	122.92	120.30

There are no chirality outliers.

There are no planarity outliers.

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	3531	0	3480	21	0
1	B	3456	0	3406	19	0
2	C	199	0	108	1	0
All	All	7186	0	6994	41	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 3.

All (41) 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:354:LEU:HD11	1:B:539:LEU:HD11	1.65	0.79
1:A:392:TRP:O	1:A:499:ARG:NH1	2.33	0.62
1:B:554:LYS:HB2	1:B:569:THR:HG22	1.82	0.62
1:B:393:PRO:O	1:B:499:ARG:HD2	2.06	0.55
1:A:461:PRO:HA	1:A:590:TRP:CD1	2.42	0.54
1:A:462:TYR:O	1:A:589:ILE:HG23	2.11	0.51
1:A:463:SER:OG	1:A:465:GLN:OE1	2.29	0.51
1:B:330:TYR:CE1	1:B:537:TYR:CE2	2.98	0.51
1:B:511:LEU:HB2	1:B:542:LEU:HD12	1.92	0.50
1:A:454:TYR:CD2	1:A:599:ASP:HB3	2.48	0.49
1:B:337:GLU:O	1:B:341:VAL:HG23	2.12	0.48
1:B:353:TYR:CD1	1:B:544:LEU:HD12	2.49	0.48
1:A:330:TYR:O	1:A:331:ASN:HB2	2.14	0.48
1:B:319:PHE:CG	1:B:350:THR:HG21	2.49	0.47
1:B:322:ASN:N	1:B:322:ASN:HD22	2.11	0.47
2:C:-8:DT:H2"	2:C:-7:DG:C8	2.50	0.47
1:B:322:ASN:OD1	1:B:374:ARG:NH1	2.46	0.47
1:B:249:SER:C	1:B:250:LEU:HD12	2.35	0.47
1:A:414:SER:OG	1:A:415:GLU:N	2.48	0.46
1:A:337:GLU:O	1:A:341:VAL:HG23	2.16	0.46
1:B:457:GLY:HA3	1:B:588:TRP:CZ2	2.51	0.46
1:A:264:THR:HG23	1:A:539:LEU:HB3	1.97	0.45
1:A:188:ILE:HD11	1:A:296:ILE:HD13	1.99	0.45
1:B:354:LEU:HD11	1:B:539:LEU:CD1	2.43	0.45
1:B:376:LYS:HG2	1:B:419:SER:HA	1.99	0.45
1:A:264:THR:HG23	1:A:539:LEU:CB	2.48	0.44
1:A:169:THR:HG23	1:A:294:GLN:HB2	1.99	0.44
1:A:457:GLY:HA3	1:A:588:TRP:CZ2	2.52	0.44
1:B:414:SER:OG	1:B:415:GLU:N	2.51	0.43
1:A:586:ARG:HB2	1:A:586:ARG:CZ	2.49	0.43
1:A:511:LEU:HD12	1:A:541:VAL:O	2.19	0.42
1:A:397:GLN:NE2	1:A:494:ILE:O	2.52	0.42
1:B:547:ALA:C	1:B:548:LEU:HD23	2.39	0.42
1:A:445:GLU:OE2	1:A:448:ARG:NH2	2.52	0.42
1:B:347:LEU:HD12	1:B:347:LEU:N	2.35	0.42
1:B:554:LYS:CB	1:B:569:THR:HG22	2.50	0.41
1:B:222:LYS:O	1:B:224:ILE:HD12	2.20	0.41
1:A:211:LEU:HD13	1:A:284:LEU:HD13	2.02	0.41
1:A:243:LYS:N	1:A:244:PRO:CD	2.84	0.41
1:A:169:THR:O	1:A:487:ARG:NH2	2.54	0.41
1:A:554:LYS:HB3	1:A:569:THR:HG22	2.03	0.41

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	435/485 (90%)	412 (95%)	23 (5%)	0	100	100
1	B	426/485 (88%)	406 (95%)	20 (5%)	0	100	100
All	All	861/970 (89%)	818 (95%)	43 (5%)	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	384/421 (91%)	371 (97%)	13 (3%)	37	70
1	B	376/421 (89%)	353 (94%)	23 (6%)	18	54
All	All	760/842 (90%)	724 (95%)	36 (5%)	26	62

All (36) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	162	ASN
1	A	203	ASN
1	A	231	LYS
1	A	261	THR
1	A	282	SER
1	A	288	ASP
1	A	292	LYS
1	A	428	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	481	SER
1	A	499	ARG
1	A	528	ASN
1	A	567	MET
1	A	608	SER
1	B	172	SER
1	B	177	LYS
1	B	196	LEU
1	B	232	ARG
1	B	233	GLU
1	B	237	HIS
1	B	243	LYS
1	B	252	GLN
1	B	255	LEU
1	B	263	HIS
1	B	275	LEU
1	B	288	ASP
1	B	310	HIS
1	B	334	SER
1	B	358	THR
1	B	361	ARG
1	B	373	PHE
1	B	385	SER
1	B	425	LYS
1	B	463	SER
1	B	466	THR
1	B	511	LEU
1	B	590	TRP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (8) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	310	HIS
1	A	397	GLN
1	A	446	ASN
1	A	528	ASN
1	B	252	GLN
1	B	363	GLN
1	B	388	ASN
1	B	397	GLN

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

There are no carbohydrates in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

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	<RSRZ>	#RSRZ>2	OWAB(Å^2)	Q<0.9
1	A	441/485 (90%)	-0.24	1 (0%) 95 94	39, 64, 93, 129	0
1	B	432/485 (89%)	-0.05	1 (0%) 95 94	52, 78, 110, 145	0
2	C	11/11 (100%)	-0.28	0 100 100	58, 76, 156, 165	0
All	All	884/981 (90%)	-0.15	2 (0%) 95 94	39, 71, 106, 165	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	559	PHE	2.5
1	B	438	TYR	2.2

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.