



# Full wwPDB X-ray Structure Validation Report

Feb 28, 2014 – 02:11 AM GMT

PDB ID : 4I2I  
Title : Binary complex of mouse TdT with AP5A  
Authors : Gouge, J.; Delarue, M.  
Deposited on : 2012-11-21  
Resolution : 2.50 Å(reported)

This is a full wwPDB 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 <http://wwpdb.org/ValidationPDFNotes.html>

---

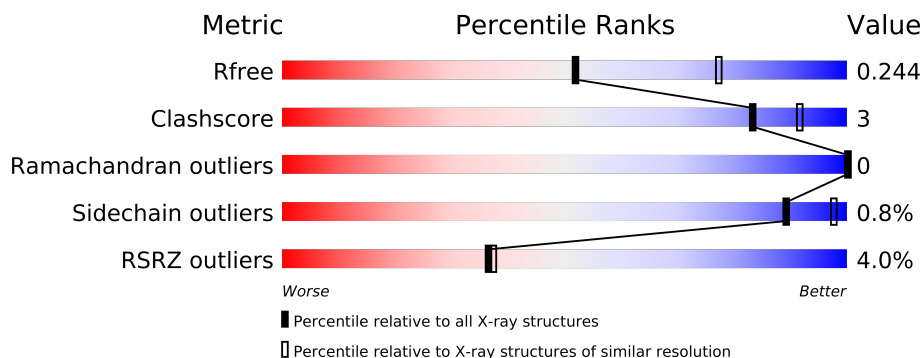
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.15 2013  
Xtriage (Phenix) : dev-1323  
EDS : stable22639  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable22683

# 1 Overall quality at a glance

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))
$R_{free}$	66092	2784 (2.50-2.50)
Clashscore	79885	3562 (2.50-2.50)
Ramachandran outliers	78287	3480 (2.50-2.50)
Sidechain outliers	78261	3482 (2.50-2.50)
RSRZ outliers	66119	2785 (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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	400	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
4	MG	A	609	-	X

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 2849 atoms, of which 0 are hydrogen and 0 are deuterium.

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.

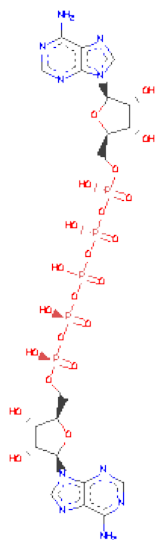
- Molecule 1 is a protein called DNA nucleotidylexotransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	338	Total	C	N	O	S	0	1	0
			2637	1679	459	482	17			

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	111	MET	-	EXPRESSION TAG	UNP P09838
A	112	GLY	-	EXPRESSION TAG	UNP P09838
A	113	SER	-	EXPRESSION TAG	UNP P09838
A	114	SER	-	EXPRESSION TAG	UNP P09838
A	115	HIS	-	EXPRESSION TAG	UNP P09838
A	116	HIS	-	EXPRESSION TAG	UNP P09838
A	117	HIS	-	EXPRESSION TAG	UNP P09838
A	118	HIS	-	EXPRESSION TAG	UNP P09838
A	119	HIS	-	EXPRESSION TAG	UNP P09838
A	120	HIS	-	EXPRESSION TAG	UNP P09838
A	121	SER	-	EXPRESSION TAG	UNP P09838
A	122	SER	-	EXPRESSION TAG	UNP P09838
A	123	GLY	-	EXPRESSION TAG	UNP P09838
A	124	LEU	-	EXPRESSION TAG	UNP P09838
A	125	VAL	-	EXPRESSION TAG	UNP P09838
A	126	PRO	-	EXPRESSION TAG	UNP P09838
A	127	ARG	-	EXPRESSION TAG	UNP P09838
A	128	GLY	-	EXPRESSION TAG	UNP P09838
A	129	SER	-	EXPRESSION TAG	UNP P09838
A	130	HIS	-	EXPRESSION TAG	UNP P09838
A	131	MET	-	EXPRESSION TAG	UNP P09838

- Molecule 2 is BIS(ADENOSINE)-5'-PENTAPHOSPHATE (three-letter code: AP5) (formula: C<sub>20</sub>H<sub>29</sub>N<sub>10</sub>O<sub>22</sub>P<sub>5</sub>).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	P	0	0
			39	10	5	19	5		

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

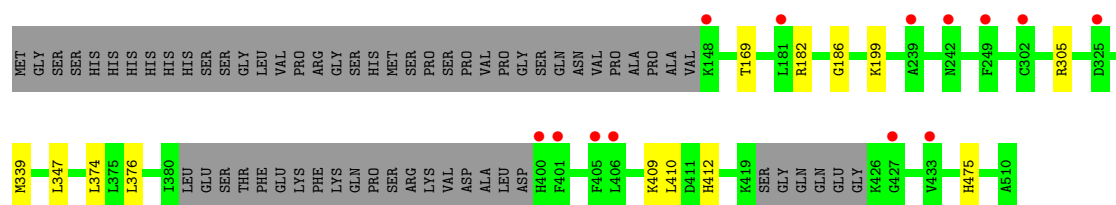
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	7	Total	Zn	0	0
			7	7		

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Mg	0	0
			1	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	165	Total	O	0	0
			165	165		



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	46.75Å 84.10Å 103.89Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	22.10 – 2.50 24.82 – 2.50	Depositor EDS
% Data completeness (in resolution range)	98.8 (22.10-2.50) 97.4 (24.82-2.50)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.78 (at 2.50Å)	Xtriage
Refinement program	BUSTER 2.9.3	Depositor
R, $R_{free}$	0.195 , 0.240 0.199 , 0.244	Depositor DCC
$R_{free}$ test set	722 reflections (5.29%)	DCC
Wilson B-factor (Å <sup>2</sup> )	39.4	Xtriage
Anisotropy	0.318	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 52.2	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.29$	Xtriage
Outliers	0 of 14589 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	2849	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.10% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, MG, AP5

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	0.52	0/2691	0.67	0/3638

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2637	0	0	8	0
2	A	39	0	12	0	0
3	A	7	0	0	0	0
4	A	1	0	0	0	0
5	A	165	0	0	0	0
All	All	2849	0	12	8	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 3.

All (8) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:182:ARG:NH2	1:A:186:GLY:O	2.25	0.70
1:A:169:THR:CG2	1:A:199:LYS:CE	2.81	0.57
1:A:410:LEU:C	1:A:410:LEU:CD1	2.73	0.56
1:A:374:LEU:O	1:A:376:LEU:CD1	2.59	0.51
1:A:347:LEU:CD2	1:A:347:LEU:C	2.80	0.49
1:A:376:LEU:N	1:A:376:LEU:CD1	2.76	0.49
1:A:376:LEU:CD1	1:A:409:LYS:N	2.79	0.45
1:A:305:ARG:CD	1:A:339:MET:SD	3.10	0.40

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	333/400 (83%)	328 (98%)	5 (2%)	0	100 100

There are no Ramachandran outliers to report.

### 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	267/350 (76%)	264 (99%)	3 (1%)	84 96

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

Mol	Chain	Res	Type
1	A	412[A]	HIS
1	A	412[B]	HIS
1	A	475	HIS



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 chains 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 ⓘ

Of 9 ligands modelled in this entry, 8 are monoatomic - leaving 1 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	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
2	AP5	A	601	3,4	41,41,62	2.68	12 (29%)	66,66,98	2.42	15 (22%)

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
2	AP5	A	601	3,4	-	0/34/50/76	0/1/3/6

All (12) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	601	AP5	PD-O3G	9.19	1.76	1.59
2	A	601	AP5	PG-O3G	6.28	1.71	1.59
2	A	601	AP5	C4B-N3B	4.92	1.43	1.35
2	A	601	AP5	PD-O3D	4.91	1.68	1.59
2	A	601	AP5	PE-O3D	4.54	1.68	1.59
2	A	601	AP5	C2B-N3B	4.22	1.40	1.32
2	A	601	AP5	PA-O5F	4.01	1.64	1.50
2	A	601	AP5	C2B-N1B	3.96	1.41	1.33
2	A	601	AP5	C2J-C1J	2.87	1.57	1.53
2	A	601	AP5	PG-O2G	-2.16	1.45	1.55
2	A	601	AP5	PB-O3A	2.10	1.63	1.59
2	A	601	AP5	PE-O2E	-2.08	1.45	1.55

All (15) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	601	AP5	N3B-C2B-N1B	-13.11	117.75	128.71
2	A	601	AP5	O3D-PD-O3G	5.91	113.67	101.66
2	A	601	AP5	O4J-C1J-N9B	5.11	113.20	108.44
2	A	601	AP5	N3B-C4B-N9B	3.75	132.21	125.43
2	A	601	AP5	PG-O3B-PB	-3.52	121.37	131.68
2	A	601	AP5	O4J-C4J-C3J	3.42	112.11	105.17
2	A	601	AP5	C2J-C1J-N9B	3.01	121.00	113.27
2	A	601	AP5	O4J-C1J-C2J	2.90	111.21	106.77
2	A	601	AP5	O3G-PG-O3B	2.78	107.31	101.66
2	A	601	AP5	C2B-N3B-C4B	2.76	121.86	114.01
2	A	601	AP5	C4B-C5B-N7B	-2.39	107.47	109.52
2	A	601	AP5	O3B-PB-O3A	-2.30	96.99	101.66
2	A	601	AP5	C4J-O4J-C1J	-2.19	107.37	109.75
2	A	601	AP5	C8B-N7B-C5B	2.05	109.95	103.58
2	A	601	AP5	N7B-C8B-N9B	-2.01	108.66	114.36

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 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<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	338/400 (84%)	0.07	13 (3%)	38 40	22, 45, 71, 86	0

All (13) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	401	PHE	4.9
1	A	400	HIS	4.0
1	A	239	ALA	3.4
1	A	148	LYS	3.3
1	A	325	ASP	3.1
1	A	405	PHE	2.8
1	A	242	ASN	2.5
1	A	406	LEU	2.5
1	A	302	CYS	2.4
1	A	433	VAL	2.3
1	A	181	LEU	2.3
1	A	249	PHE	2.2
1	A	427	GLY	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 ⓘ

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, 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	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
4	MG	A	609	1/1	0.54	11.95	62,62,62,62	0
2	AP5	A	601	39/57	0.19	0.33	70,86,118,118	0
3	ZN	A	608	1/1	0.03	-2.74	71,71,71,71	0
3	ZN	A	604	1/1	0.05	-2.89	61,61,61,61	0
3	ZN	A	602	1/1	0.09	-3.16	44,44,44,44	0
3	ZN	A	603	1/1	0.07	-3.49	51,51,51,51	0
3	ZN	A	605	1/1	0.03	-3.94	77,77,77,77	0
3	ZN	A	607	1/1	0.03	-5.57	67,67,67,67	0
3	ZN	A	606	1/1	0.05	-8.57	79,79,79,79	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.