



# wwPDB X-ray Structure Validation Summary Report

Mar 1, 2014 – 12:46 AM GMT

PDB ID : 1X9M  
Title : T7 DNA polymerase in complex with an N-2-acetylaminofluorene-adducted DNA  
Authors : Dutta, S.; Li, Y.; Johnson, D.; Dzantiev, L.; Richardson, C.C.; Romano, L.J.; Ellenberger, T.  
Deposited on : 2004-08-23  
Resolution : 2.10 Å(reported)

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

---

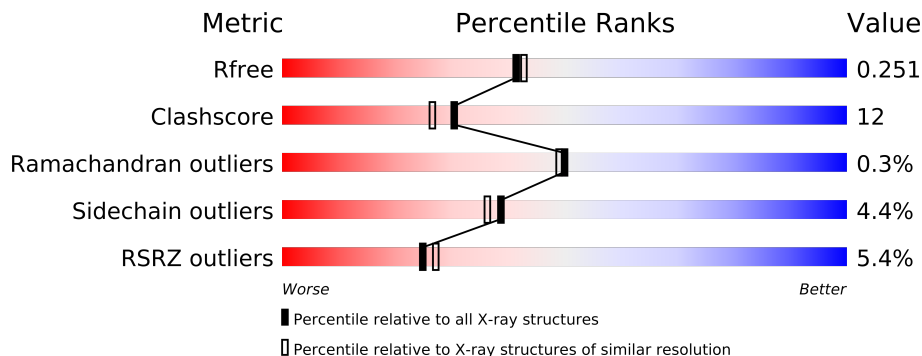
The following versions of software and data (see [references](#)) were used in the production of this report:

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 2.10 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))
$R_{free}$	66092	3012 (2.10-2.10)
Clashscore	79885	3649 (2.10-2.10)
Ramachandran outliers	78287	3610 (2.10-2.10)
Sidechain outliers	78261	3611 (2.10-2.10)
RSRZ outliers	66119	3013 (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  $\geq 3$ , 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	C	22	
2	D	26	
3	A	698	
4	B	108	

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

Mol	Type	Chain	Res	Geometry	Electron density
5	MG	A	3001	-	X

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 6824 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 DNA chain called 5'-D(\*GP\*GP\*AP\*GP\*AP\*GP\*TP\*GP\*AP\*TP\*TP\*GP\*GP\*TP\*AP\*GP\*TP\*GP\*TP\*GP\*AP\*(2DT))-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	10	Total	C	N	O	P	0	0	0
			209	100	38	61	10			

- Molecule 2 is a DNA chain called 5'-D(\*CP\*CP\*CP\*(8FG)P\*AP\*TP\*CP\*AP\*CP\*AP\*CP\*TP\*AP\*CP\*CP\*AP\*AP\*TP\*CP\*AP\*CP\*TP\*CP\*TP\*CP\*C)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	12	Total	C	N	O	P	0	0	0
			258	130	45	71	12			

- Molecule 3 is a protein called DNA polymerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	675	Total	C	N	O	S	0	0	0
			5251	3355	917	955	24			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	?	-	LYS	DELETION	UNP P00581
A	?	-	ARG	DELETION	UNP P00581
A	?	-	PHE	DELETION	UNP P00581
A	?	-	GLY	DELETION	UNP P00581
A	?	-	SER	DELETION	UNP P00581
A	?	-	HIS	DELETION	UNP P00581

- Molecule 4 is a protein called Thioredoxin 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	B	105	Total	C	N	O	S	0	0	0
			790	514	129	144	3			

- Molecule 5 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

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

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	266	Total 266	O 266	0	0
6	B	15	Total 15	O 15	0	0
6	C	11	Total 11	O 11	0	0
6	D	23	Total 23	O 23	0	0

### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

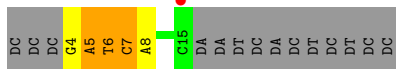
- Molecule 1: 5'-D(\*GP\*GP\*AP\*GP\*AP\*GP\*TP\*GP\*AP\*TP\*TP\*GP\*GP\*TP\*AP\*GP\*TP\*GP\*TP\*GP\*AP\*(2DT))-3'

Chain C: 



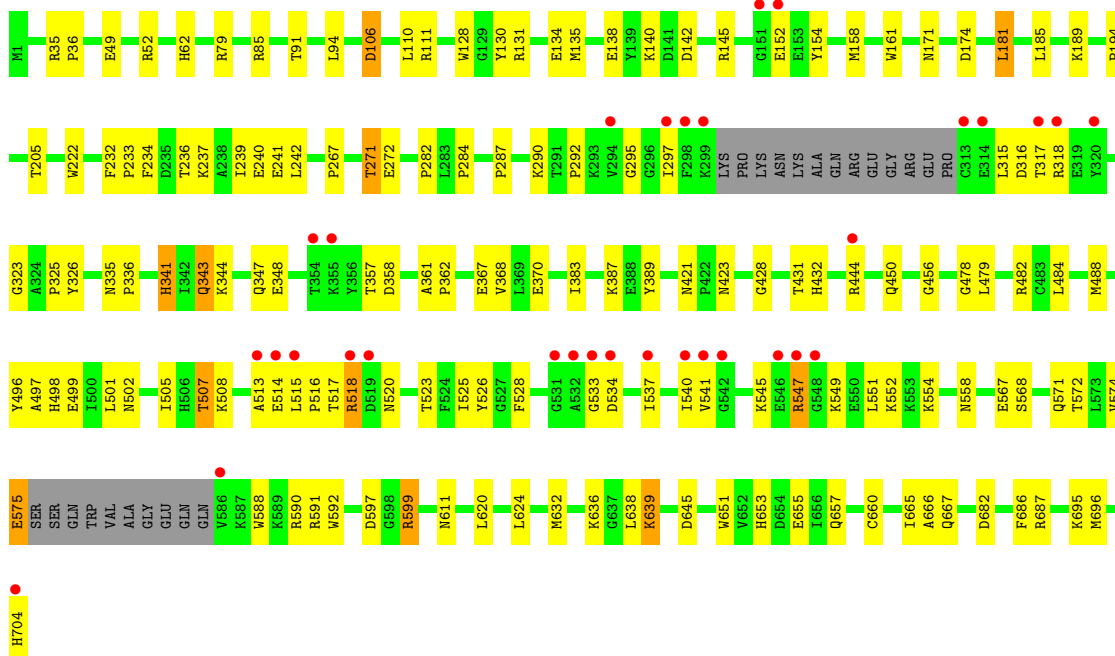
- Molecule 2: 5'-D(\*CP\*CP\*CP\*(8FG)P\*AP\*TP\*CP\*AP\*CP\*AP\*CP\*TP\*AP\*CP\*CP\*AP\*AP\*TP\*CP\*AP\*CP\*TP\*CP\*TP\*CP\*C)-3'

Chain D: 



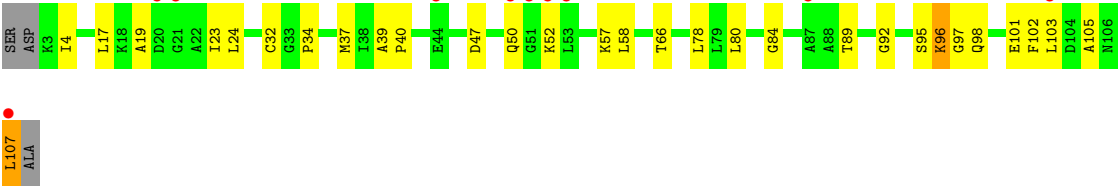
- Molecule 3: DNA polymerase

Chain A: 



- Molecule 4: Thioredoxin 1

Chain B: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	105.11Å 212.31Å 52.07Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.38 – 2.10 47.38 – 2.10	Depositor EDS
% Data completeness (in resolution range)	94.3 (47.38-2.10) 94.4 (47.38-2.10)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.61 (at 2.10Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.213 , 0.236 0.231 , 0.251	Depositor DCC
$R_{free}$ test set	3239 reflections (4.96%)	DCC
Wilson B-factor (Å <sup>2</sup> )	27.1	Xtriage
Anisotropy	0.299	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 39.2	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Outliers	0 of 65243 reflections	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	6824	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	34.0	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.*

<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: 2DT, MG, 8FG

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	C	0.52	0/213	0.87	0/328
2	D	0.53	0/244	0.99	0/372
3	A	0.39	0/5380	0.60	0/7299
4	B	0.33	0/805	0.59	0/1092
All	All	0.40	0/6642	0.63	0/9091

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	C	0	3
2	D	0	3
All	All	0	6

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	18	DG	Sidechain
1	C	20	DG	Sidechain
1	C	21	DA	Sidechain
2	D	5	DA	Sidechain
2	D	6	DT	Sidechain



## 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	C	209	0	115	21	0
2	D	258	0	146	7	0
3	A	5251	0	5084	110	1
4	B	790	0	808	25	0
5	A	1	0	0	0	0
6	A	266	0	0	4	0
6	B	15	0	0	1	0
6	C	11	0	0	0	0
6	D	23	0	0	0	0
All	All	6824	0	6153	151	1

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 12.

The worst 5 of 151 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:C:22:2DT:H6	1:C:22:2DT:H5'	1.30	1.14
1:C:21:DA:H2''	1:C:22:2DT:H5''	1.22	1.12
3:A:271:THR:HG22	3:A:290:LYS:HG2	1.53	0.90
4:B:19:ALA:HB3	4:B:23:ILE:HD11	1.57	0.84
1:C:17:DT:H2''	1:C:18:DG:H5''	1.57	0.83

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
3:A:205:THR:OG1	3:A:687:ARG:NH2[2.655]	2.14	0.06

## 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	
3	A	669/698 (96%)	642 (96%)	25 (4%)	2 (0%)	50	49
4	B	103/108 (95%)	102 (99%)	1 (1%)	0	100	100
All	All	772/806 (96%)	744 (96%)	26 (3%)	2 (0%)	50	49

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	513	ALA
3	A	653	HIS

### 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	
3	A	527/579 (91%)	504 (96%)	23 (4%)	39	36
4	B	81/87 (93%)	77 (95%)	4 (5%)	35	31
All	All	608/666 (91%)	581 (96%)	27 (4%)	39	36

5 of 27 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	507	THR
3	A	547	ARG
4	B	89	THR
3	A	518	ARG
3	A	174	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	347	GLN
3	A	423	ASN
3	A	510	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
3	A	343	GLN
3	A	502	ASN

### 5.3.3 RNA ⓘ

There are no RNA chains in this entry.

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

2 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	2DT	C	22	1,2	18,20,21	0.98	2 (11%)	20,28,31	1.72	5 (25%)
2	8FG	D	4	2	42,44,45	3.78	25 (59%)	57,66,69	12.93	14 (24%)

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	2DT	C	22	1,2	-	0/5/18/19	0/2/2/2
2	8FG	D	4	2	-	0/15/41/42	0/1/6/6

The worst 5 of 27 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	4	8FG	C6-C5	9.75	1.57	1.41
2	D	4	8FG	C8-N29	9.62	1.54	1.34
2	D	4	8FG	C4-N3	7.77	1.48	1.35
2	D	4	8FG	C2-N1	7.27	1.48	1.36
2	D	4	8FG	P-OP1	5.68	1.53	1.46

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	4	8FG	C6-C5-N7	-96.55	129.37	134.22
2	D	4	8FG	C2-N3-C4	5.56	122.90	115.09
2	D	4	8FG	C45-C30-N29	4.87	126.85	117.91
2	D	4	8FG	C4-C5-N7	4.61	113.33	109.58
1	C	22	2DT	C6-N1-C2	-4.42	121.15	122.41

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 1 ligands modelled in this entry, 1 is 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.

## 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	C	10/22 (45%)	0.09	0 100 100	27, 33, 71, 86	0
2	D	12/26 (46%)	0.29	1 (8%) 11 12	19, 31, 63, 81	0
3	A	675/698 (96%)	0.27	32 (4%) 30 33	14, 29, 60, 76	0
4	B	105/108 (97%)	0.60	10 (9%) 8 9	25, 38, 55, 63	0
All	All	802/854 (93%)	0.31	43 (5%) 25 27	14, 31, 60, 86	0

The worst 5 of 43 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	532	ALA	6.9
3	A	294	VAL	5.4
3	A	586	VAL	4.5
4	B	107	LEU	3.9
3	A	298	PHE	3.7

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

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
2	8FG	D	4	39/40	0.27	1.93	44,54,68,68	0
1	2DT	C	22	19/20	0.16	-0.09	28,31,34,35	0

### 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
5	MG	A	3001	1/1	0.27	7.34	42,42,42,42	0

### 6.5 Other polymers ⓘ

There are no such residues in this entry.