



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 13, 2017 – 07:03 pm GMT

PDB ID : 1A31  
Title : HUMAN RECONSTITUTED DNA TOPOISOMERASE I IN COVALENT  
COMPLEX WITH A 22 BASE PAIR DNA DUPLEX  
Authors : Redinbo, M.R.; Stewart, L.; Kuhn, P.; Champoux, J.J.; Hol, W.G.J.  
Deposited on : 1998-01-27  
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](mailto: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)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

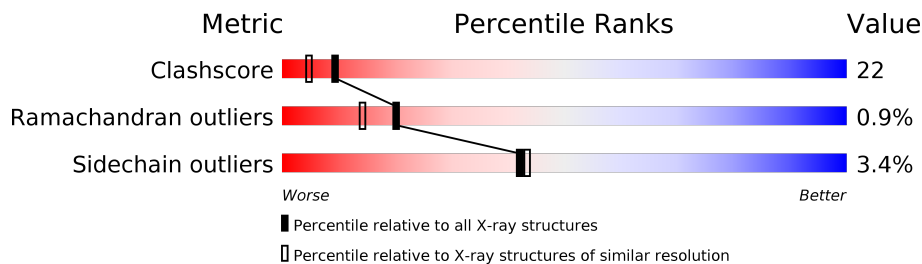
# 1 Overall quality at a glance

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))
Clashscore	112137	4788 (2.10-2.10)
Ramachandran outliers	110173	4740 (2.10-2.10)
Sidechain outliers	110143	4741 (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. 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.

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	C	22	
2	D	22	
3	A	591	

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 4991 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 DNA chain called DNA (5'-D(\*AP\*AP\*AP\*AP\*AP\*GP\*AP\*CP\*5IUP\*5IU\*TP\*GP\*AP\*AP\*AP\*AP\*AP\*5IUP\*5IUP\*5IUP\*5IUP\*T)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	C	22	Total	C	I	N	O	P	0	0	0
			447	213	6	84	124	20			

- Molecule 2 is a DNA chain called DNA (5'-D(\*AP\*AP\*AP\*AP\*AP\*TP\*5IUP\*5IUP\*5IUP\*5IUP\*CP\*AP\*AP\*AP\*GP\*TP\*CP\*TP\*TP\*TP\*T)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	D	22	Total	C	I	N	O	P	0	0	0
			445	214	4	73	133	21			

- Molecule 3 is a protein called PROTEIN (TOPOISOMERASE I).

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	A	458	Total	C	N	O	P	S	0	0	0
			3690	2365	640	664	1	20			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	723	PTR	TYR	MODIFIED RESIDUE	UNP P11387

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	296	Total	O	0	0
			296	296		
4	C	47	Total	O	0	0
			47	47		
4	D	66	Total	O	0	0
			66	66		

### 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 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.

Note EDS was not executed.

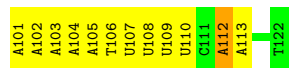
- Molecule 1: DNA (5'-D(\*AP\*AP\*AP\*AP\*AP\*GP\*AP\*CP\*5IUP\*5IU\*TP\*GP\*AP\*AP\*AP\*AP\*AP\*5IUP\*5IUP\*5IUP\*5IUP\*T)-3')

Chain C: 



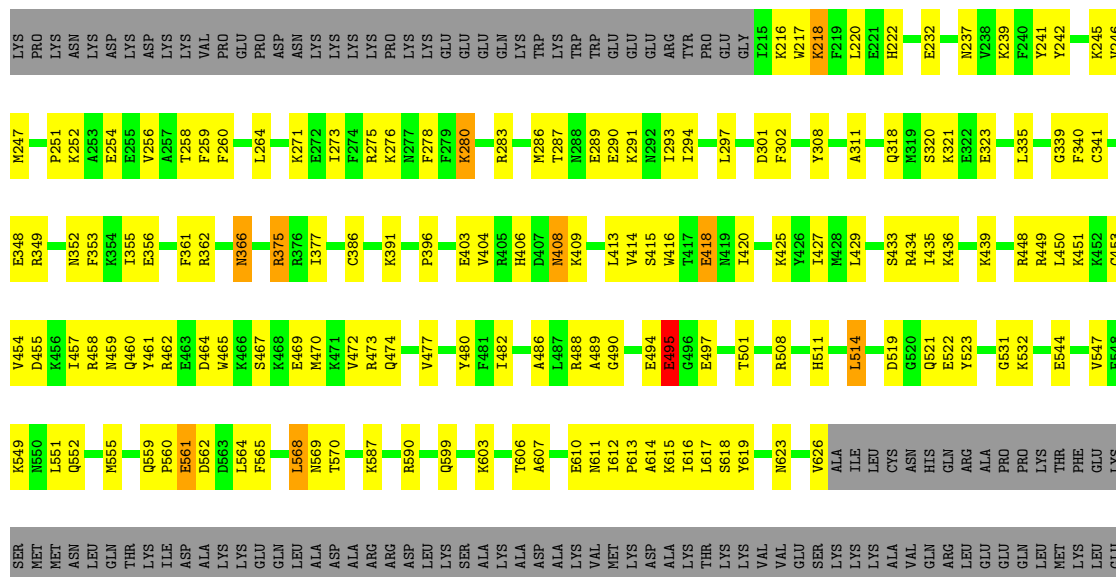
- Molecule 2: DNA (5'-D(\*AP\*AP\*AP\*AP\*AP\*TP\*5IUP\*5IUP\*5IUP\*5IUP\*CP\*AP\*AP\*AP\*GP\*TP\*CP\*TP\*TP\*TP\*TP\*T)-3')

Chain D: 



- Molecule 3: PROTEIN (TOPOISOMERASE I)

Chain A: 



VAL	GLN	ALA	THR	ASP	ARG	GLU	GLU	ASN	LYS	GLN	TLE	ALA	LEU	GLY	THR	SER	K720	L721	N722	Y723	A731	W732	C733	W736	I743	Y744	N745	K746	T747	Q748	R749	E750	W754	W758	A759	D760	E761	D762	Y763	E764	F765
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

## 4 Data and refinement statistics

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	72.00 Å   66.60 Å   71.80 Å 90.00°   98.30°   90.00°	Depositor
Resolution (Å)	20.00 – 2.10	Depositor
% Data completeness (in resolution range)	97.2 (20.00-2.10)	Depositor
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
Refinement program	X-PLOR 3.1	Depositor
R, $R_{free}$	0.247 , 0.310	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	4991	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	49.0	wwPDB-VP

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: 5IU, PTR

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	0.96	0/370	0.91	0/565
2	D	0.92	0/408	0.92	0/625
3	A	0.64	0/3761	0.77	0/5077
All	All	0.70	0/4539	0.80	0/6267

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	2
2	D	0	1
All	All	0	3

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	13	DA	Sidechain
1	C	8	DC	Sidechain
2	D	112	DA	Sidechain

## 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	C	447	0	234	17	0
2	D	445	0	243	23	0
3	A	3690	0	3588	151	0
4	A	296	0	0	30	0
4	C	47	0	0	2	0
4	D	66	0	0	7	0
All	All	4991	0	4065	187	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 22.

The worst 5 of 187 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:C:1:DA:H2''	1:C:2:DA:H5'	1.13	1.12
2:D:101:DA:H2''	2:D:102:DA:H5'	1.15	1.10
1:C:1:DA:H2''	1:C:2:DA:C5'	1.87	1.05
2:D:107:5IU:H5''	4:D:1230:HOH:O	1.59	1.02
1:C:8:DC:H2'	1:C:9:5IU:I5	2.30	1.01

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
3	A	453/591 (77%)	426 (94%)	23 (5%)	4 (1%)	20 14



All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	561	GLU
3	A	721	LEU
3	A	495	GLU
3	A	761	GLU

### 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
3	A	381/534 (71%)	368 (97%)	13 (3%)	42 43

5 of 13 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	418	GLU
3	A	434	ARG
3	A	568	LEU
3	A	408	ASN
3	A	561	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	459	ASN
3	A	748	GLN
3	A	599	GLN
3	A	430	ASN
3	A	576	HIS

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

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

11 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$
3	PTR	A	723	1,3	15,16,17	0.75	0	19,22,24	1.06	1 (5%)
1	5IU	C	10	1,3,2	12,20,22	2.36	3 (25%)	10,28,33	4.90	4 (40%)
1	5IU	C	18	1,2	13,21,22	2.21	3 (23%)	15,30,33	3.88	2 (13%)
1	5IU	C	19	1,2	13,21,22	2.04	2 (15%)	15,30,33	3.83	2 (13%)
1	5IU	C	20	1,2	13,21,22	2.08	2 (15%)	15,30,33	3.92	3 (20%)
1	5IU	C	21	1,2	13,21,22	2.01	2 (15%)	15,30,33	3.90	3 (20%)
1	5IU	C	9	1,2	13,21,22	2.42	3 (23%)	15,30,33	3.77	2 (13%)
2	5IU	D	107	1,2	13,21,22	2.18	2 (15%)	15,30,33	3.95	3 (20%)
2	5IU	D	108	1,2	13,21,22	2.29	3 (23%)	15,30,33	3.81	2 (13%)
2	5IU	D	109	1,2	13,21,22	2.59	3 (23%)	15,30,33	3.76	2 (13%)
2	5IU	D	110	1,2	13,21,22	2.36	3 (23%)	15,30,33	3.82	2 (13%)

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
3	PTR	A	723	1,3	-	0/9/11/13	0/1/1/1
1	5IU	C	10	1,3,2	-	0/3/18/22	0/2/2/2
1	5IU	C	18	1,2	-	0/3/21/22	0/2/2/2
1	5IU	C	19	1,2	-	0/3/21/22	0/2/2/2
1	5IU	C	20	1,2	-	0/3/21/22	0/2/2/2
1	5IU	C	21	1,2	-	0/3/21/22	0/2/2/2
1	5IU	C	9	1,2	-	0/3/21/22	0/2/2/2
2	5IU	D	107	1,2	-	0/3/21/22	0/2/2/2
2	5IU	D	108	1,2	-	0/3/21/22	0/2/2/2
2	5IU	D	109	1,2	-	0/3/21/22	0/2/2/2
2	5IU	D	110	1,2	-	0/3/21/22	0/2/2/2

The worst 5 of 26 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	109	5IU	C5-I5	-8.25	1.92	2.10
1	C	9	5IU	C5-I5	-7.76	1.93	2.10
2	D	110	5IU	C5-I5	-7.51	1.93	2.10
2	D	108	5IU	C5-I5	-7.36	1.94	2.10
2	D	107	5IU	C5-I5	-7.00	1.94	2.10

The worst 5 of 26 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	108	5IU	C5-C4-N3	-6.14	115.05	123.27
1	C	21	5IU	C5-C4-N3	-6.10	115.10	123.27
1	C	18	5IU	C5-C4-N3	-5.95	115.30	123.27
2	D	107	5IU	C5-C4-N3	-5.91	115.36	123.27
1	C	9	5IU	C5-C4-N3	-5.91	115.37	123.27

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

9 monomers are involved in 15 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	723	PTR	1	0
1	C	19	5IU	3	0
1	C	20	5IU	3	0
1	C	21	5IU	2	0
1	C	9	5IU	4	0
2	D	107	5IU	3	0
2	D	108	5IU	1	0
2	D	109	5IU	1	0
2	D	110	5IU	1	0

## 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 [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

### 6.1 Protein, DNA and RNA chains

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

### 6.3 Carbohydrates

EDS was not executed - this section is therefore empty.

### 6.4 Ligands

EDS was not executed - this section is therefore empty.

### 6.5 Other polymers

EDS was not executed - this section is therefore empty.