W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Feb 13, 2017 – 05:37 pm GMT

PDB ID	:	1EJ9
Title	:	CRYSTAL STRUCTURE OF HUMAN TOPOISOMERASE I DNA COMPLEX
Authors	:	Redinbo, M.R.; Champoux, J.J.; Hol, W.G.
Deposited on	:	2000-03-01
Resolution	:	2.60 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 (i) symbol.

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

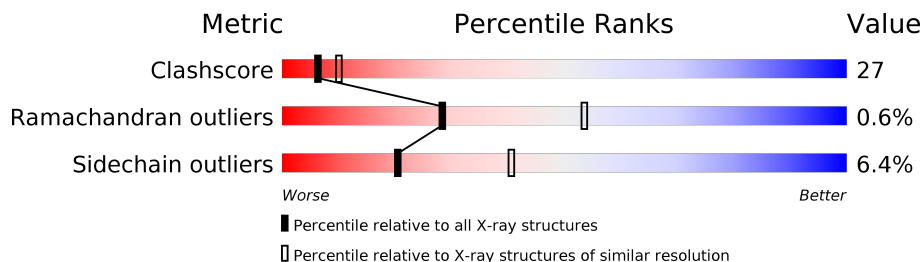|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.60 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	Percentile Ranks	Value
Clashscore		27
Ramachandran outliers		0.6%
Sidechain outliers		6.4%

Worse   *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	112137	2895 (2.60-2.60)
Ramachandran outliers	110173	2848 (2.60-2.60)
Sidechain outliers	110143	2848 (2.60-2.60)

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 <=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	23	52% / 35% / 9% / ·
2	D	23	43% / 48% / · / ·
3	A	563	41% / 42% / · / 14%

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 5215 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(*C*AP*AP*AP*AP*AP*GP*AP*CP*TP*C P*AP*GP*AP*AP*AP*AP*AP*TP*TP*TP*TP*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	22	Total	C	N	O	P	0	0	0
			451	218	88	124	21			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	D	22	Total	C	N	O	P	0	0	0
			445	218	70	136	21			

- Molecule 3 is a protein called DNA TOPOISOMERASE I.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	483	Total	C	N	O	S	0	0	0
			3966	2547	689	708	22			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	634	ILE	ARG	engineered	UNP P11387
A	723	PHE	TYR	engineered	UNP P11387

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	270	Total	O	0	0
			270	270		
4	C	33	Total	O	0	0
			33	33		
4	D	50	Total	O	0	0
			50	50		

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.
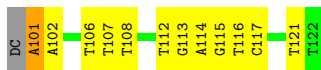
Note EDS was not executed.

- Molecule 1: DNA (5'-D(*C*AP*AP*AP*AP*AP*GP*AP*CP*TP*CP*AP*GP*AP*AP*AP*AP*AP*TP*TP*TP*TP*T)-3')
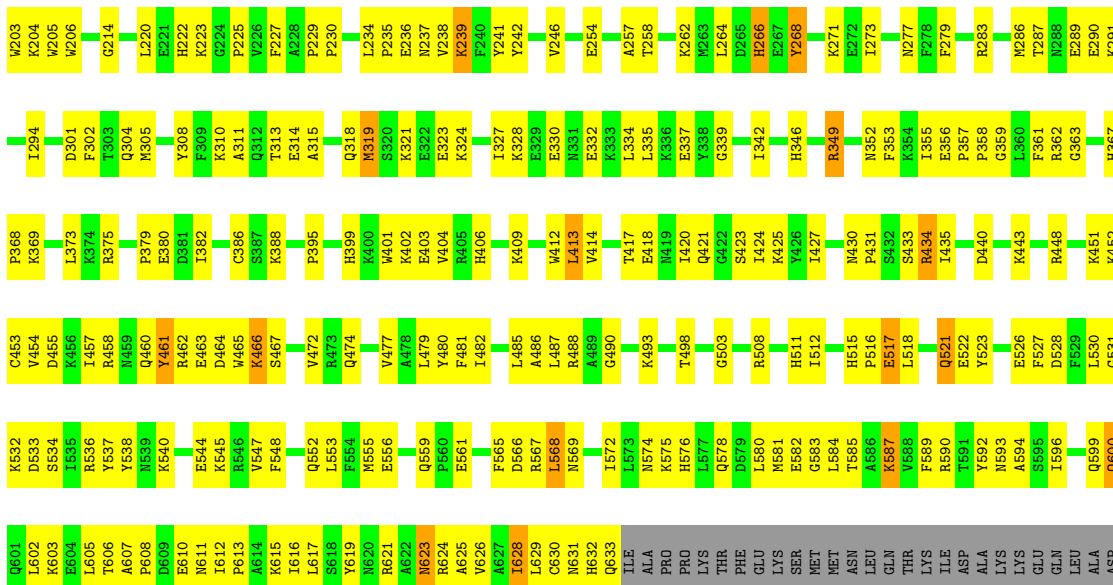
Chain C:

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

Chain D:

- Molecule 3: DNA TOPOISOMERASE I

Chain A:

ALA ARG ARG ASP LEU LYS SER ALA LYS ALA ASP ALA LYS VAL MET LYS ASP ALA LYS THR LYS LYS VAL VAL GLU SER LYS LYS LYS ALA VAL GLN ARG LEU GLU GLU VAL GLN ALA THR ASP ARG GLU GLU GLU ASN LYS GLN I714 A715 L716 S719 F723

L724 D725 P726 V730 A731 W732 C733 K734 K735 W736 G737 V738 P739 I740 E741 K742 I743 Y744 N745 K746 T747 Q748 R749 F752 A753 W754 A755 I756 D757 M758 E761 D762 Y763 E764 F765

4 Data and refinement statistics (i)

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, α, β, γ	56.98Å 124.92Å 72.29Å 90.00° 93.84° 90.00°	Depositor
Resolution (Å)	15.00 – 2.60	Depositor
% Data completeness (in resolution range)	98.7 (15.00-2.60)	Depositor
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	X-PLOR 3.1	Depositor
R, R_{free}	0.214 , 0.281	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	5215	wwPDB-VP
Average B, all atoms (Å^2)	54.0	wwPDB-VP

5 Model quality (i)

5.1 Standard geometry (i)

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	0.99	0/508	1.03	0/782
2	D	1.02	0/496	1.04	3/764 (0.4%)
3	A	0.74	0/4064	0.83	0/5476
All	All	0.80	0/5068	0.88	3/7022 (0.0%)

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

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	112	DT	C1'-O4'-C4'	-5.98	104.12	110.10
2	D	112	DT	C3'-C2'-C1'	-5.60	95.78	102.50
2	D	101	DA	O4'-C1'-N9	-5.03	104.48	108.00

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	17	DA	Sidechain
1	C	8	DC	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	451	0	250	22	0
2	D	445	0	256	23	0
3	A	3966	0	3923	210	0
4	A	270	0	0	23	0
4	C	33	0	0	2	0
4	D	50	0	0	13	0
All	All	5215	0	4429	246	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 27.

The worst 5 of 246 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)
3:A:517:GLU:HG3	3:A:522:GLU:HG3	1.30	1.10
1:C:21:DT:H2'	1:C:22:DT:H72	1.38	1.02
3:A:599:GLN:HE22	3:A:765:PHE:H	1.01	0.97
3:A:367:HIS:HD2	3:A:369:LYS:H	1.12	0.97
2:D:101:DA:H1'	4:D:1064:HOH:O	1.74	0.88

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	479/563 (85%)	448 (94%)	28 (6%)	3 (1%)	28	53

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	319	MET
3	A	264	LEU
3	A	467	SER

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	419/507 (83%)	392 (94%)	27 (6%)	20	40

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

Mol	Chain	Res	Type
3	A	460	GLN
3	A	498	THR
3	A	741	GLU
3	A	461	TYR
3	A	268	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
3	A	460	GLN
3	A	515	HIS
3	A	631	ASN
3	A	430	ASN
3	A	623	ASN

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 ⓘ

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.