



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 13, 2017 – 07:02 pm GMT

PDB ID : 1R49  
Title : Human topoisomerase I (Topo70) double mutant K532R/Y723F  
Authors : Interthal, H.; Quigley, P.M.; Hol, W.G.; Champoux, J.J.  
Deposited on : 2003-10-03  
Resolution : 3.13 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
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

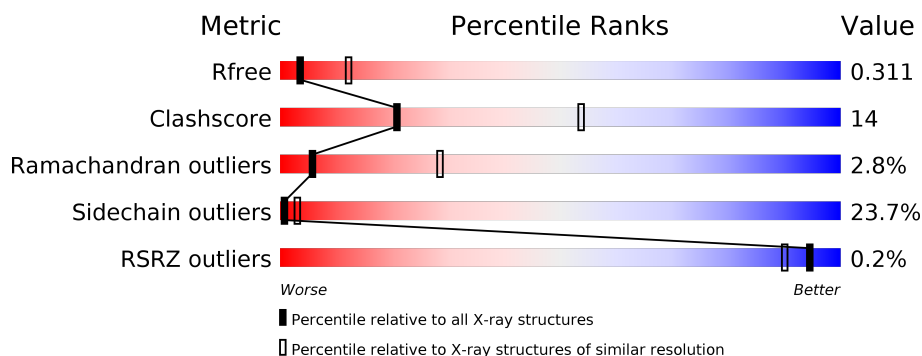
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.13 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))
$R_{free}$	100719	1234 (3.18-3.10)
Clashscore	112137	1345 (3.18-3.10)
Ramachandran outliers	110173	1301 (3.18-3.10)
Sidechain outliers	110143	1301 (3.18-3.10)
RSRZ outliers	101464	1240 (3.18-3.10)

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.

Mol	Chain	Length	Quality of chain
1	B	22	
2	C	22	
3	A	592	

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 5453 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 5'-D(\*AP\*AP\*AP\*AP\*AP\*GP\*AP\*CP\*TP\*TP\*AP\*GP\*AP\*AP\*AP\*AP\*AP\*TP\*TP\*TP\*TP\*T)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	22	Total	C	N	O	P	0	0	0
			451	218	87	125	21			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	22	Total	C	N	O	P	0	0	0
			446	217	70	137	22			

- Molecule 3 is a protein called DNA topoisomerase I.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	A	548	Total	C	N	O	S	40	0	0
			4527	2869	801	831	26			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	532	ARG	LYS	ENGINEERED	UNP P11387
A	723	PHE	TYR	ENGINEERED	UNP P11387

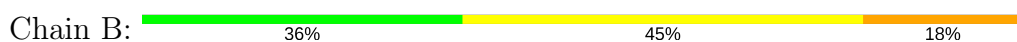
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	25	Total	O	0	0
			25	25		
4	B	1	Total	O	0	0
			1	1		
4	C	3	Total	O	0	0
			3	3		

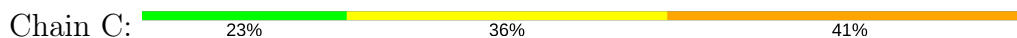
### 3 Residue-property plots

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 ( $\text{RSRZ} > 2$ ). 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.

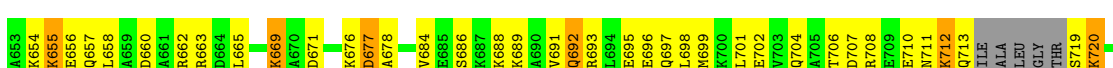
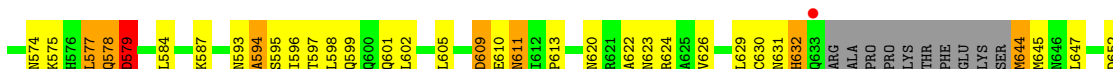
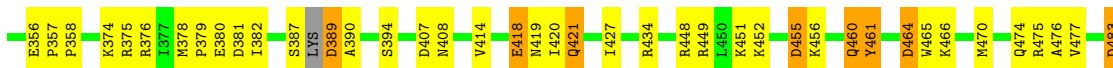
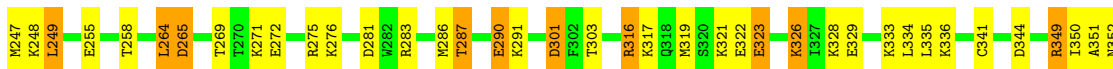
- Molecule 1: 5'-D(\*AP\*AP\*AP\*AP\*AP\*GP\*AP\*CP\*TP\*TP\*AP\*GP\*AP\*AP\*AP\*AP\*AP\*T  
P\*TP\*TP\*TP\*T)-3'



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



- Molecule 3: DNA topoisomerase I



L724	D725	P739	L740	E741	K742	K746	T747	Q748	R749	E750	W754	D757	N758	A759	D760	E761	D762	F765
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	56.94Å 118.64Å 71.65Å 90.00° 98.33° 90.00°	Depositor
Resolution (Å)	47.67 – 3.13 47.60 – 3.13	Depositor EDS
% Data completeness (in resolution range)	100.0 (47.67-3.13) 93.0 (47.60-3.13)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	61.79 (at 3.12Å)	Xtriage
Refinement program	REFMAC 5.1.24	Depositor
R, $R_{free}$	0.277 , 0.339 0.248 , 0.311	Depositor DCC
$R_{free}$ test set	821 reflections (5.59%)	DCC
Wilson B-factor (Å <sup>2</sup> )	65.9	Xtriage
Anisotropy	0.369	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.25 , 55.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	5453	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	66.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.11% of the height of the origin peak. No significant pseudotranslation is detected.*

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

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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	B	1.83	6/508 (1.2%)	1.54	12/782 (1.5%)
2	C	2.47	11/497 (2.2%)	1.93	23/764 (3.0%)
3	A	0.65	8/4621 (0.2%)	0.89	49/6206 (0.8%)
All	All	1.09	25/5626 (0.4%)	1.11	84/7752 (1.1%)

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
3	A	2	6

The worst 5 of 25 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	203	TRP	CA-C	25.38	2.19	1.52
1	B	18	DT	C5-C7	-16.21	1.40	1.50
1	B	21	DT	C5-C7	-16.16	1.40	1.50
2	C	120	DT	C5-C7	-16.16	1.40	1.50
2	C	108	DT	C5-C7	-16.15	1.40	1.50

The worst 5 of 84 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	203	TRP	CB-CA-C	-17.84	74.71	110.40
3	A	203	TRP	N-CA-CB	-15.42	82.84	110.60
3	A	205	TRP	CZ3-CH2-CZ2	-15.30	103.24	121.60
3	A	204	LYS	CB-CA-C	14.32	139.04	110.40
3	A	203	TRP	N-CA-C	-13.69	74.04	111.00

All (2) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
3	A	204	LYS	CA
3	A	208	GLU	CA

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	A	203	TRP	Mainchain
3	A	205	TRP	Peptide
3	A	206	TRP	Mainchain,Peptide
3	A	522	GLU	Peptide
3	A	594	ALA	Peptide

## 5.2 Too-close contacts [i](#)

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	B	451	0	248	16	0
2	C	446	0	252	19	0
3	A	4527	0	4447	108	3
4	A	25	0	0	0	0
4	B	1	0	0	0	0
4	C	3	0	0	0	0
All	All	5453	0	4947	141	3

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 14.

The worst 5 of 141 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)
2:C:121:DT:H2''	2:C:122:DT:H71	1.18	1.14
3:A:521:GLN:HA	3:A:522:GLU:HB2	1.20	1.14
3:A:521:GLN:HA	3:A:522:GLU:CB	1.82	1.08
3:A:531:GLY:O	3:A:532:ARG:O	1.83	0.96
3:A:523:TYR:HB3	3:A:548:PHE:CD1	2.03	0.93

All (3) symmetry-related close contacts are listed below. The label for Atom-2 includes the sym-



metry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:A:204:LYS:NZ	3:A:272:GLU:OE2[1_455]	0.42	1.78
3:A:204:LYS:NZ	3:A:272:GLU:CD[1_455]	1.23	0.97
3:A:204:LYS:CE	3:A:272:GLU:OE2[1_455]	1.54	0.66

## 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	
3	A	540/592 (91%)	494 (92%)	31 (6%)	15 (3%)	6	28

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	204	LYS
3	A	205	TRP
3	A	522	GLU
3	A	532	ARG
3	A	533	ASP

### 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	465/536 (87%)	355 (76%)	110 (24%)	1	3

5 of 110 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	A	506	SER
3	A	567	ARG
3	A	720	LYS
3	A	510	GLU
3	A	533	ASP

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	515	HIS
3	A	521	GLN
3	A	704	GLN
3	A	430	ASN
3	A	657	GLN

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

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

### 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

### 5.6 Ligand geometry [i](#)

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 [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

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	B	22/22 (100%)	-0.58	0	100 100	23, 57, 93, 104	0
2	C	22/22 (100%)	-0.41	0	100 100	20, 57, 91, 106	0
3	A	548/592 (92%)	-0.24	1 (0%)	94 90	23, 64, 106, 135	7 (1%)
All	All	592/636 (93%)	-0.26	1 (0%)	94 90	20, 63, 106, 135	7 (1%)

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	A	633	GLN	2.4

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

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

### 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

### 6.4 Ligands [i](#)

There are no ligands in this entry.

### 6.5 Other polymers [i](#)

There are no such residues in this entry.