



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 21, 2020 – 07:34 am BST

PDB ID : 2GAJ  
Title : Structure of Full Length Topoisomerase I from *Thermotoga maritima* in monoclinic crystal form  
Authors : Hansen, G.  
Deposited on : 2006-03-09  
Resolution : 1.95 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

<https://www.wwpdb.org/validation/2017/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)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

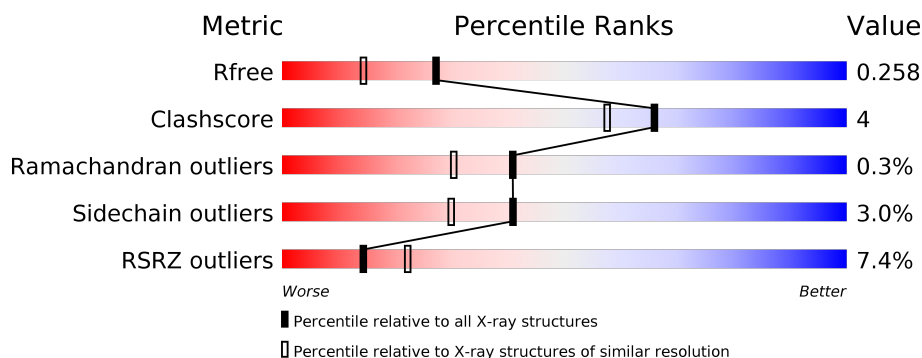
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 1.95 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))
$R_{free}$	130704	2580 (1.96-1.96)
Clashscore	141614	2705 (1.96-1.96)
Ramachandran outliers	138981	2678 (1.96-1.96)
Sidechain outliers	138945	2678 (1.96-1.96)
RSRZ outliers	127900	2539 (1.96-1.96)

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 respectively. 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	A	633	
1	B	633	

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 10305 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 protein called DNA topoisomerase I.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	581	Total	C	N	O	S	0	0	0
			4714	3017	799	881	17			
1	B	581	Total	C	N	O	S	0	0	0
			4714	3017	799	881	17			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	INITIATING METHIONINE	UNP P46799
B	1	MET	-	INITIATING METHIONINE	UNP P46799
A	2	ALA	-	CLONING ARTIFACT	UNP P46799
B	2	ALA	-	CLONING ARTIFACT	UNP P46799

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	427	Total	O	0	0
			427	427		
2	B	450	Total	O	0	0
			450	450		



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	45.05Å 142.71Å 127.64Å 90.00° 90.84° 90.00°	Depositor
Resolution (Å)	15.00 – 1.95 14.99 – 1.95	Depositor EDS
% Data completeness (in resolution range)	99.6 (15.00-1.95) 99.8 (14.99-1.95)	Depositor EDS
$R_{merge}$	0.12	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.10 (at 1.95Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.195 , 0.246 0.209 , 0.258	Depositor DCC
$R_{free}$ test set	5837 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	20.5	Xtriage
Anisotropy	0.062	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.38 , 43.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	0.000 for h,-k,-l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	10305	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	11.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 45.74 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.2655e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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 [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	A	0.67	0/4799	1.07	23/6441 (0.4%)
1	B	0.69	0/4799	1.16	27/6441 (0.4%)
All	All	0.68	0/9598	1.12	50/12882 (0.4%)

There are no bond length outliers.

The worst 5 of 50 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	179	ARG	NE-CZ-NH2	-25.81	107.40	120.30
1	B	179	ARG	NE-CZ-NH1	18.68	129.64	120.30
1	A	456	ARG	NE-CZ-NH2	-11.42	114.59	120.30
1	B	127	ARG	NE-CZ-NH1	10.76	125.68	120.30
1	B	86	ASP	CB-CG-OD1	10.51	127.76	118.30

There are no chirality outliers.

There are no planarity outliers.

### 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	A	4714	0	4821	35	0
1	B	4714	0	4821	42	0
2	A	427	0	0	7	0
2	B	450	0	0	13	0
All	All	10305	0	9642	77	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 4.

The worst 5 of 77 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:A:489:ILE:HD11	1:A:601:LEU:HD21	1.56	0.86
1:A:80:LEU:HD13	1:A:109:ILE:HD11	1.58	0.84
1:A:413:GLU:O	2:A:892:HOH:O	1.96	0.82
1:A:334:HIS:HD2	2:A:729:HOH:O	1.65	0.80
1:B:192:HIS:HD2	1:B:213:PHE:H	1.31	0.77

There are no symmetry-related clashes.

## 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	
1	A	577/633 (91%)	561 (97%)	14 (2%)	2 (0%)	41	30
1	B	577/633 (91%)	562 (97%)	14 (2%)	1 (0%)	47	38
All	All	1154/1266 (91%)	1123 (97%)	28 (2%)	3 (0%)	41	30

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	550	VAL
1	A	503	SER
1	A	165	ALA

### 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	
1	A	519/562 (92%)	505 (97%)	14 (3%)	44	34
1	B	519/562 (92%)	502 (97%)	17 (3%)	38	26
All	All	1038/1124 (92%)	1007 (97%)	31 (3%)	41	30

5 of 31 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	12	GLU
1	B	51	LYS
1	B	549	ILE
1	B	34	MET
1	B	108	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	557	GLN
1	B	26	ASN
1	B	169	GLN
1	A	523	GLN
1	B	158	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 [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	A	581/633 (91%)	0.65	43 (7%)	14 22	2, 10, 24, 34	0
1	B	581/633 (91%)	0.67	43 (7%)	14 22	2, 9, 24, 36	0
All	All	1162/1266 (91%)	0.66	86 (7%)	14 22	2, 9, 24, 36	0

The worst 5 of 86 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	549	ILE	6.8
1	A	165	ALA	6.3
1	A	572	TYR	6.2
1	B	165	ALA	6.2
1	B	156	TRP	5.7

### 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.