



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 20, 2023 – 02:42 AM EDT

PDB ID : 2GAI
Title : Structure of Full Length Topoisomerase I from *Thermotoga maritima* in triclinic crystal form
Authors : Hansen, G.
Deposited on : 2006-03-08
Resolution : 1.70 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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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 : 2.35
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) : 2.35

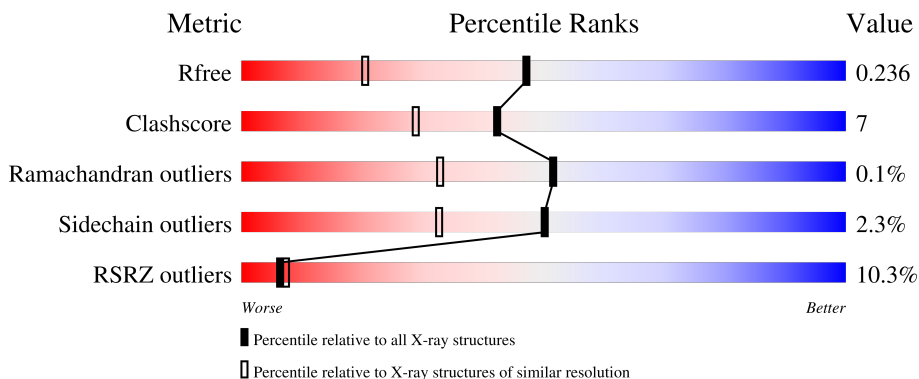
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.70 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	4298 (1.70-1.70)
Clashscore	141614	4695 (1.70-1.70)
Ramachandran outliers	138981	4610 (1.70-1.70)
Sidechain outliers	138945	4610 (1.70-1.70)
RSRZ outliers	127900	4222 (1.70-1.70)

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 of the lower bar indicate the fraction of residues that contain outliers for ≥ 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	 10% 79% 12% • 8%
1	B	633	 9% 78% 12% • 8%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 10600 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
			Total	C	N	O	S			
1	A	581	4770	3059	803	891	17	0	18	0
1	B	581	4776	3065	803	891	17	0	20	0

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	544	Total	O	0	0
			544	544		
2	B	510	Total	O	0	0
			510	510		

PRO
ASP
GLY
ARG
SER
VAL
GLU
GLY
LYS
GLY
ASN
LEU
SER
GLU
LYS
ARG
ARG
LYS
GLY
LYS
LYS
GLY
SER

4 Data and refinement statistics i

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	45.12Å 95.42Å 96.51Å 83.40° 86.15° 84.87°	Depositor
Resolution (Å)	41.56 – 1.70 41.55 – 1.70	Depositor EDS
% Data completeness (in resolution range)	95.9 (41.56-1.70) 95.9 (41.55-1.70)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.53 (at 1.70Å)	Xtrriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.197 , 0.232 0.202 , 0.236	Depositor DCC
R_{free} test set	8387 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	20.3	Xtrriage
Anisotropy	0.055	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.37 , 39.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	0.025 for -h,-l,-k	Xtrriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	10600	wwPDB-VP
Average B, all atoms (Å ²)	12.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.75% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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	0.66	0/4927	0.90	11/6618 (0.2%)
1	B	0.67	2/4941 (0.0%)	0.90	8/6637 (0.1%)
All	All	0.66	2/9868 (0.0%)	0.90	19/13255 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	581	GLY	C-O	9.25	1.38	1.23
1	B	581	GLY	C-N	8.49	1.53	1.34

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	456	ARG	NE-CZ-NH1	7.77	124.19	120.30
1	B	456	ARG	NE-CZ-NH2	-7.10	116.75	120.30
1	A	456	ARG	NE-CZ-NH2	-6.19	117.20	120.30
1	A	108	ARG	NE-CZ-NH1	5.87	123.23	120.30
1	A	528	ASP	CB-CG-OD1	5.73	123.45	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	4770	0	4890	60	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	4776	0	4898	73	0
2	A	544	0	0	8	0
2	B	510	0	0	10	0
All	All	10600	0	9788	133	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 7.

The worst 5 of 133 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:B:344:MET:CE	1:B:349:ALA:HA	1.82	1.09
1:B:344:MET:HE1	1:B:349:ALA:HA	1.43	0.99
1:A:169:GLN:OE1	1:A:471:LEU:HD13	1.74	0.87
1:A:38:ILE:HD11	1:A:99[B]:VAL:HG11	1.57	0.85
1:B:192:HIS:HD2	1:B:213:PHE:H	1.21	0.84

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	595/633 (94%)	584 (98%)	11 (2%)	0	100	100
1	B	597/633 (94%)	587 (98%)	9 (2%)	1 (0%)	47	30
All	All	1192/1266 (94%)	1171 (98%)	20 (2%)	1 (0%)	51	33

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	44	LYS

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	537/562 (96%)	520 (97%)	17 (3%)	39	20
1	B	539/562 (96%)	526 (98%)	13 (2%)	49	31
All	All	1076/1124 (96%)	1046 (97%)	30 (3%)	50	25

5 of 30 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	471	LEU
1	B	397[B]	THR
1	B	12	GLU
1	B	549	ILE
1	B	162	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	334	HIS
1	B	377	GLN
1	B	557	GLN
1	A	557	GLN
1	B	158	ASN

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 monosaccharides 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, 95th 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 ²)	Q<0.9
1	A	581/633 (91%)	0.79	63 (10%) 5 6	2, 11, 26, 41	18 (3%)
1	B	581/633 (91%)	0.77	57 (9%) 7 8	2, 10, 25, 40	20 (3%)
All	All	1162/1266 (91%)	0.78	120 (10%) 6 7	2, 10, 25, 41	38 (3%)

The worst 5 of 120 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	549	ILE	9.1
1	B	228[A]	VAL	8.3
1	B	156	TRP	7.8
1	B	165	ALA	7.6
1	A	572	TYR	7.4

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 monosaccharides 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.