



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 28, 2022 – 12:06 AM JST

PDB ID : 7VUF  
Title : Crystal Structure of the core region of *Thermus thermophilus* MutS2.  
Authors : Fukui, K.; Yano, T.  
Deposited on : 2021-11-02  
Resolution : 3.11 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  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.27  
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) : 2.27

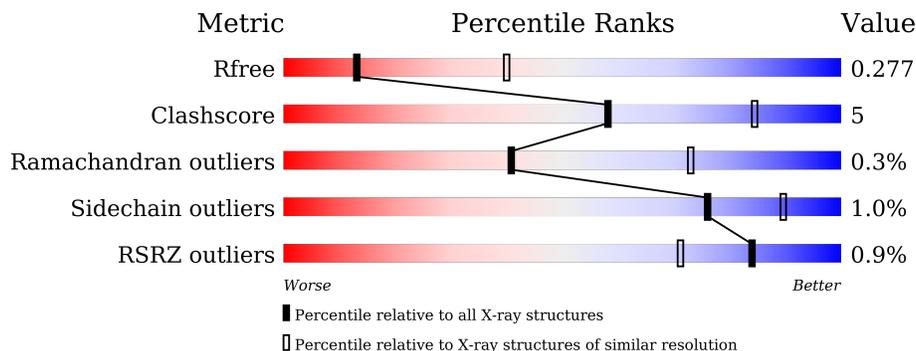
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.11 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	1292 (3.14-3.10)
Clashscore	141614	1389 (3.14-3.10)
Ramachandran outliers	138981	1337 (3.14-3.10)
Sidechain outliers	138945	1337 (3.14-3.10)
RSRZ outliers	127900	1260 (3.14-3.10)

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 of 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	488	 80% 14% • 5%
1	B	488	 84% 11% • •
1	C	488	 83% 13% • •
1	D	488	 81% 13% • 5%

## 2 Entry composition [i](#)

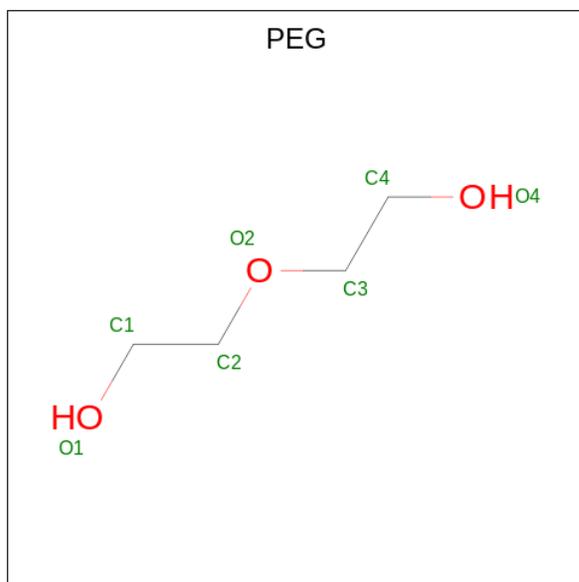
There are 4 unique types of molecules in this entry. The entry contains 14442 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Endonuclease MutS2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	B	471	Total 3584	C 2269	N 642	O 665	S 8	0	0	0
1	A	462	Total 3541	C 2246	N 635	O 651	S 9	0	0	0
1	C	473	Total 3624	C 2296	N 646	O 673	S 9	0	0	0
1	D	463	Total 3536	C 2231	N 643	O 653	S 9	0	0	0

- Molecule 2 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: C<sub>4</sub>H<sub>10</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
2	B	1	Total 7	C 4	O 3	0	0
2	B	1	Total 7	C 4	O 3	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	1	Total C O 7 4 3	0	0
2	B	1	Total C O 7 4 3	0	0
2	A	1	Total C O 7 4 3	0	0
2	A	1	Total C O 7 4 3	0	0
2	C	1	Total C O 7 4 3	0	0

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	C	1	Total Mg 1 1	0	0

- Molecule 4 is water.

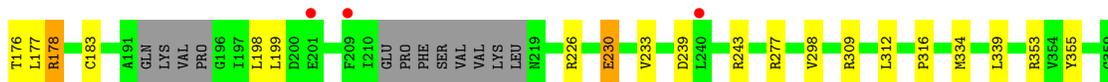
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	22	Total O 22 22	0	0
4	A	35	Total O 35 35	0	0
4	C	31	Total O 31 31	0	0
4	D	19	Total O 19 19	0	0





● Molecule 1: Endonuclease MutS2

Chain D: %



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	93.48Å 114.21Å 292.88Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.89 – 3.11 40.89 – 3.11	Depositor EDS
% Data completeness (in resolution range)	94.0 (40.89-3.11) 94.1 (40.89-3.11)	Depositor EDS
$R_{merge}$	0.27	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.77 (at 3.12Å)	Xtrriage
Refinement program	PHENIX 1.19.2	Depositor
R, $R_{free}$	0.225 , 0.279 0.224 , 0.277	Depositor DCC
$R_{free}$ test set	2587 reflections (4.80%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	57.1	Xtrriage
Anisotropy	0.323	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 39.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	14442	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	52.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.26% 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

Bond lengths and bond angles in the following residue types are not validated in this section: PEG, MG

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	0.24	0/3591	0.52	0/4854
1	B	0.24	0/3633	0.49	0/4915
1	C	0.24	0/3675	0.51	0/4966
1	D	0.24	0/3584	0.52	0/4837
All	All	0.24	0/14483	0.51	0/19572

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	A	0	10
1	B	0	5
1	C	0	14
1	D	0	11
All	All	0	40

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 40 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	315	GLY	Peptide
1	B	316	PRO	Peptide
1	B	318	MET	Peptide
1	B	77	LEU	Peptide
1	B	80	PRO	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	A	3541	0	3625	38	0
1	B	3584	0	3641	33	0
1	C	3624	0	3714	35	0
1	D	3536	0	3592	36	0
2	A	14	0	20	1	0
2	B	28	0	40	0	0
2	C	7	0	10	2	0
3	C	1	0	0	0	0
4	A	35	0	0	1	0
4	B	22	0	0	1	0
4	C	31	0	0	1	0
4	D	19	0	0	1	0
All	All	14442	0	14642	138	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 5.

The worst 5 of 138 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:156:LEU:HA	1:A:158:ARG:H	1.47	0.79
1:B:318:MET:H	1:B:321:LYS:HE2	1.48	0.78
1:A:427:THR:HG22	1:A:429:LEU:H	1.47	0.77
1:A:321:LYS:HE2	1:A:426:THR:HB	1.72	0.72
1:D:152:ARG:HH21	1:D:226:ARG:HH21	1.41	0.68

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	454/488 (93%)	435 (96%)	16 (4%)	3 (1%)	22	56
1	B	461/488 (94%)	450 (98%)	8 (2%)	3 (1%)	22	56
1	C	465/488 (95%)	441 (95%)	24 (5%)	0	100	100
1	D	453/488 (93%)	434 (96%)	19 (4%)	0	100	100
All	All	1833/1952 (94%)	1760 (96%)	67 (4%)	6 (0%)	41	73

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	81	GLU
1	A	176	THR
1	A	210	ILE
1	A	214	SER
1	B	470	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	355/392 (91%)	350 (99%)	5 (1%)	67	85
1	B	357/392 (91%)	355 (99%)	2 (1%)	86	93
1	C	366/392 (93%)	364 (100%)	2 (0%)	88	94
1	D	349/392 (89%)	344 (99%)	5 (1%)	67	85
All	All	1427/1568 (91%)	1413 (99%)	14 (1%)	76	89

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	163	MET
1	C	460	LEU
1	D	447	MET

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	D	160	TYR
1	D	178	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	266	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 monosaccharides in this entry.

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

Of 8 ligands modelled in this entry, 1 is monoatomic - leaving 7 for Mogul analysis.

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	$\# Z  > 2$
2	PEG	A	501	-	6,6,6	0.11	0	5,5,5	0.12	0
2	PEG	B	502	-	6,6,6	0.10	0	5,5,5	0.10	0
2	PEG	B	503	-	6,6,6	0.10	0	5,5,5	0.10	0
2	PEG	B	501	-	6,6,6	0.10	0	5,5,5	0.10	0
2	PEG	A	502	-	6,6,6	0.11	0	5,5,5	0.09	0
2	PEG	B	504	-	6,6,6	0.10	0	5,5,5	0.14	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
2	PEG	C	501	-	6,6,6	0.10	0	5,5,5	0.09	0

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
2	PEG	A	501	-	-	2/4/4/4	-
2	PEG	B	502	-	-	1/4/4/4	-
2	PEG	B	503	-	-	2/4/4/4	-
2	PEG	B	501	-	-	0/4/4/4	-
2	PEG	A	502	-	-	3/4/4/4	-
2	PEG	B	504	-	-	2/4/4/4	-
2	PEG	C	501	-	-	0/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 10 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	502	PEG	O1-C1-C2-O2
2	B	504	PEG	O1-C1-C2-O2
2	B	504	PEG	O2-C3-C4-O4
2	A	502	PEG	C1-C2-O2-C3
2	A	501	PEG	C1-C2-O2-C3

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	502	PEG	1	0
2	C	501	PEG	2	0

## 5.7 Other polymers [i](#)

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 [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	462/488 (94%)	-0.31	3 (0%) 89 79	30, 43, 75, 137	0
1	B	471/488 (96%)	-0.24	3 (0%) 89 79	32, 51, 82, 145	0
1	C	473/488 (96%)	-0.20	7 (1%) 73 56	30, 50, 84, 134	0
1	D	463/488 (94%)	-0.19	4 (0%) 84 71	35, 50, 79, 121	0
All	All	1869/1952 (95%)	-0.24	17 (0%) 84 71	30, 49, 80, 145	0

The worst 5 of 17 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	209	PHE	3.1
1	C	201	GLU	2.9
1	B	52	LEU	2.6
1	A	463	GLY	2.6
1	C	207	THR	2.6

### 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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q < 0.9’ lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
2	PEG	A	501	7/7	0.76	0.26	45,55,77,78	0
2	PEG	B	504	7/7	0.82	0.28	55,60,81,85	0
2	PEG	B	503	7/7	0.82	0.28	62,62,75,83	0
2	PEG	C	501	7/7	0.86	0.25	40,54,68,70	0
2	PEG	A	502	7/7	0.87	0.23	39,59,68,86	0
2	PEG	B	502	7/7	0.92	0.33	46,53,60,61	0
2	PEG	B	501	7/7	0.92	0.35	44,56,64,72	0
3	MG	C	502	1/1	0.98	0.35	40,40,40,40	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.