



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 10, 2026 – 02:44 AM UTC

PDB ID : 4EJ3 / pdb\_00004ej3  
Title : Crystal structure of a CRISPR associated protein from *Thermus thermophilus* HB8  
Authors : Huang, Q.Q.  
Deposited on : 2012-04-06  
Resolution : 2.52 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 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 : 4-5-2 with Phenix2.0  
Xtriage (Phenix) : 2.0  
EDS : 3.0  
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)  
CCP4 : 9.0.010 (Gargrove)  
Density-Fitness : 1.0.12  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.49

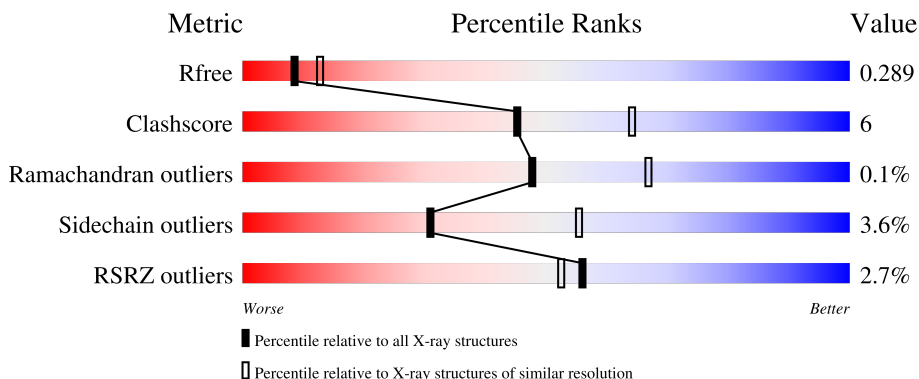
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.52 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))
$R_{free}$	180053	7383 (2.54-2.50)
Clashscore	190562	8079 (2.54-2.50)
Ramachandran outliers	187476	7944 (2.54-2.50)
Sidechain outliers	187428	7946 (2.54-2.50)
RSRZ outliers	180081	7387 (2.54-2.50)

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	514	 4% 77% 13% • 9%
1	B	514	 % 77% 12% • 9%

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 7545 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 CRISPR associated protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	467	3693	2370	659	659	5	0	0	0
1	B	467	3695	2371	662	657	5	0	0	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-11	MET	-	expression tag	UNP Q53VY1
A	-10	ARG	-	expression tag	UNP Q53VY1
A	-9	GLY	-	expression tag	UNP Q53VY1
A	-8	SER	-	expression tag	UNP Q53VY1
A	-7	HIS	-	expression tag	UNP Q53VY1
A	-6	HIS	-	expression tag	UNP Q53VY1
A	-5	HIS	-	expression tag	UNP Q53VY1
A	-4	HIS	-	expression tag	UNP Q53VY1
A	-3	HIS	-	expression tag	UNP Q53VY1
A	-2	HIS	-	expression tag	UNP Q53VY1
A	-1	GLY	-	expression tag	UNP Q53VY1
A	0	SER	-	expression tag	UNP Q53VY1
B	-11	MET	-	expression tag	UNP Q53VY1
B	-10	ARG	-	expression tag	UNP Q53VY1
B	-9	GLY	-	expression tag	UNP Q53VY1
B	-8	SER	-	expression tag	UNP Q53VY1
B	-7	HIS	-	expression tag	UNP Q53VY1
B	-6	HIS	-	expression tag	UNP Q53VY1
B	-5	HIS	-	expression tag	UNP Q53VY1
B	-4	HIS	-	expression tag	UNP Q53VY1
B	-3	HIS	-	expression tag	UNP Q53VY1
B	-2	HIS	-	expression tag	UNP Q53VY1
B	-1	GLY	-	expression tag	UNP Q53VY1
B	0	SER	-	expression tag	UNP Q53VY1

- Molecule 2 is water.

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
2	A	61	Total 61	O 61	0	0
2	B	96	Total 96	O 96	0	0



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	93.78Å 47.83Å 129.18Å 90.00° 97.34° 90.00°	Depositor
Resolution (Å)	49.86 – 2.52 49.86 – 2.52	Depositor EDS
% Data completeness (in resolution range)	100.0 (49.86-2.52) 94.8 (49.86-2.52)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	7.46 (at 2.51Å)	Xtrriage
Refinement program	REFMAC 5.5.0109	Depositor
R, $R_{free}$	0.218 , 0.277 (Not available) , 0.289	Depositor DCC
$R_{free}$ test set	2000 reflections (5.42%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	43.4	Xtrriage
Anisotropy	0.305	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 27.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	7545	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	43.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.78% 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 [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	0.56	0/3784	0.77	0/5141
1	B	0.63	0/3786	0.81	2/5143 (0.0%)
All	All	0.60	0/7570	0.79	2/10284 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	235	THR	CA-C-N	-5.18	116.55	121.57
1	B	235	THR	C-N-CA	-5.18	116.55	121.57

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	3693	0	3706	42	0
1	B	3695	0	3713	48	0
2	A	61	0	0	4	0
2	B	96	0	0	4	0
All	All	7545	0	7419	88	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 6.

The worst 5 of 88 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:272:THR:HG23	1:B:272:THR:HG22	1.42	0.99
1:A:272:THR:HG23	1:B:272:THR:CG2	2.11	0.80
1:B:228:ALA:HB1	1:B:267:GLU:HG3	1.63	0.79
1:A:33:LEU:HD22	1:A:84:GLN:HG3	1.65	0.79
1:A:191:THR:HG21	1:A:340:ARG:HD3	1.75	0.66

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	461/514 (90%)	447 (97%)	14 (3%)	0	100	100
1	B	461/514 (90%)	438 (95%)	22 (5%)	1 (0%)	43	62
All	All	922/1028 (90%)	885 (96%)	36 (4%)	1 (0%)	48	67

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	95	ARG

### 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	376/418 (90%)	361 (96%)	15 (4%)	28	51
1	B	376/418 (90%)	364 (97%)	12 (3%)	34	60
All	All	752/836 (90%)	725 (96%)	27 (4%)	31	55

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

Mol	Chain	Res	Type
1	A	491	ARG
1	B	84	GLN
1	B	355	ARG
1	B	37	GLU
1	B	115	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	273	HIS
1	B	325	GLN
1	B	390	HIS
1	A	487	GLN
1	A	373	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 oligosaccharides 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	467/514 (90%)	0.38	21 (4%) 38 34	28, 48, 69, 86	0
1	B	467/514 (90%)	0.06	4 (0%) 81 79	21, 38, 54, 67	0
All	All	934/1028 (90%)	0.22	25 (2%) 56 52	21, 42, 66, 86	0

The worst 5 of 25 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	227	GLY	3.5
1	A	400	VAL	3.5
1	A	496	LEU	3.5
1	A	96	ASP	3.0
1	B	336	ARG	2.8

### 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 oligosaccharides 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.