



# wwPDB X-ray Structure Validation Summary Report

Feb 27, 2014 – 07:48 PM GMT

PDB ID : 2CW5  
Title : Crystal structure of a conserved hypothetical protein from *Thermus thermophilus* HB8  
Authors : Ebihara, A.; Yokoyama, S.; Kuramitsu, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)  
Deposited on : 2005-06-16  
Resolution : 1.94 Å(reported)

This is a wwPDB 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/ValidationPDFNotes.html>

---

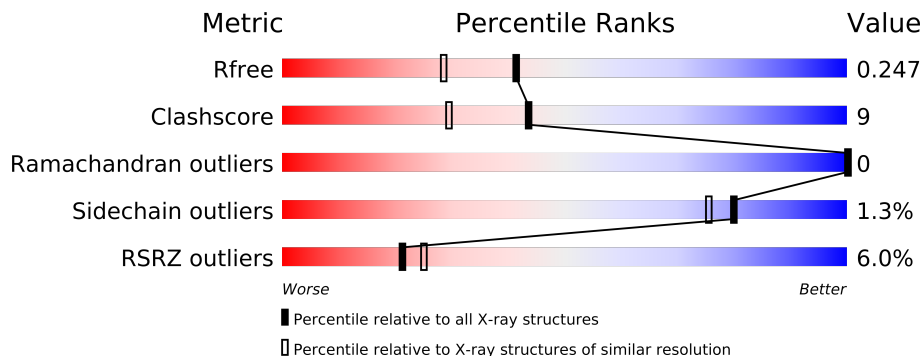
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.15 2013  
Xtriage (Phenix) : dev-1323  
EDS : stable22639  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable22683

# 1 Overall quality at a glance

The reported resolution of this entry is 1.94 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))
$R_{free}$	66092	2024 (1.96-1.92)
Clashscore	79885	2281 (1.96-1.92)
Ramachandran outliers	78287	2255 (1.96-1.92)
Sidechain outliers	78261	2255 (1.96-1.92)
RSRZ outliers	66119	2024 (1.96-1.92)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	255	
1	B	255	
1	C	255	

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 5460 atoms, of which 0 are hydrogen and 0 are deuterium.

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 Bacterial fluorinating enzyme homolog.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	235	Total	C	N	O	Se	0	0	0
			1753	1137	304	310	2			
1	B	224	Total	C	N	O	Se	0	0	0
			1686	1094	296	294	2			
1	C	238	Total	C	N	O	Se	0	0	0
			1794	1167	314	311	2			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	MODIFIED RESIDUE	UNP Q5SLF5
A	250	MSE	MET	MODIFIED RESIDUE	UNP Q5SLF5
B	1	MSE	MET	MODIFIED RESIDUE	UNP Q5SLF5
B	250	MSE	MET	MODIFIED RESIDUE	UNP Q5SLF5
C	1	MSE	MET	MODIFIED RESIDUE	UNP Q5SLF5
C	250	MSE	MET	MODIFIED RESIDUE	UNP Q5SLF5

- Molecule 2 is water.

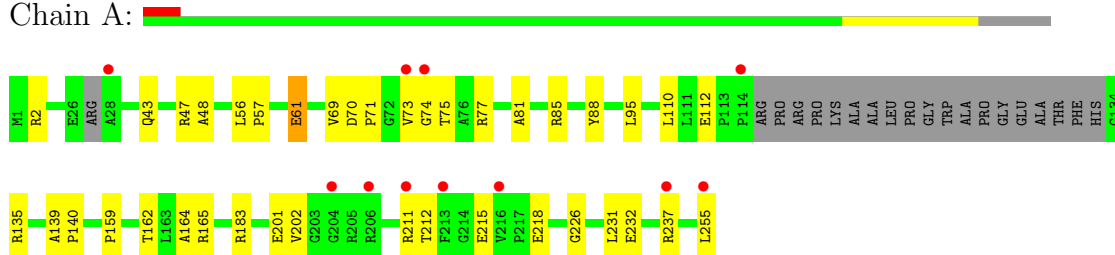
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	77	Total	O	0	0
			77	77		
2	B	92	Total	O	0	0
			92	92		
2	C	58	Total	O	0	0
			58	58		

### 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 errors 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 ( $RSRZ > 2$ ). 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.

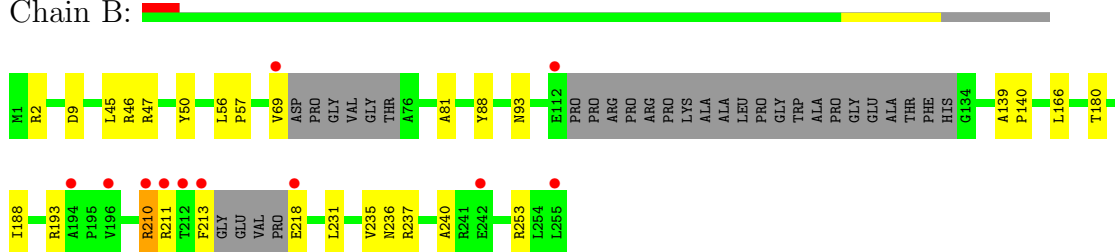
- Molecule 1: Bacterial fluorinating enzyme homolog

Chain A:



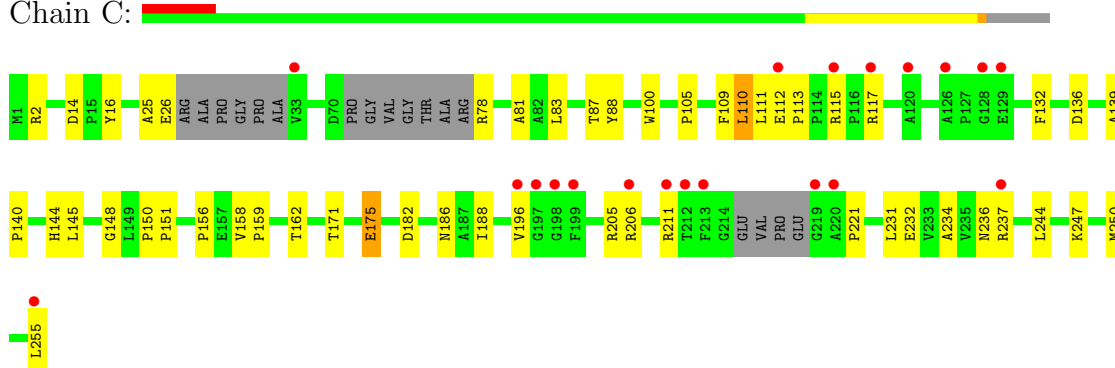
- Molecule 1: Bacterial fluorinating enzyme homolog

Chain B:



- Molecule 1: Bacterial fluorinating enzyme homolog

Chain C:



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	74.79Å 78.12Å 120.58Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	37.40 – 1.94 37.40 – 1.94	Depositor EDS
% Data completeness (in resolution range)	99.5 (37.40-1.94) 99.6 (37.40-1.94)	Depositor EDS
$R_{merge}$	0.06	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.96 (at 1.94Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.212 , 0.246 0.212 , 0.247	Depositor DCC
$R_{free}$ test set	5353 reflections (10.14%)	DCC
Wilson B-factor (Å <sup>2</sup> )	25.8	Xtriage
Anisotropy	0.316	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.38 , 38.9	EDS
Estimated twinning fraction	0.004 for k,h,-l	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 52815 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	5460	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.42% of the height of the origin peak. No significant pseudotranslation is detected.*

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

## 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	A	0.32	0/1797	0.62	1/2458 (0.0%)
1	B	0.32	0/1725	0.59	0/2353
1	C	0.32	0/1841	0.58	0/2515
All	All	0.32	0/5363	0.60	1/7326 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	69	VAL	N-CA-C	-5.63	95.79	111.00

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Close contacts

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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1753	0	1789	35	0
1	B	1686	0	1729	26	0
1	C	1794	0	1829	36	0
2	A	77	0	0	0	0
2	B	92	0	0	1	0
2	C	58	0	0	1	0
All	All	5460	0	5347	92	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 9.

The worst 5 of 92 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:73:VAL:HG22	1:A:135:ARG:HH22	1.38	0.89
1:A:201:GLU:HB2	1:A:255:LEU:HD11	1.67	0.77
1:C:115:ARG:NH1	1:C:115:ARG:HB3	2.02	0.75
1:C:78:ARG:HD2	1:C:110:LEU:HD21	1.69	0.74
1:A:61:GLU:OE2	1:A:85:ARG:HD3	1.87	0.74

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone ⓘ

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	229/255 (90%)	225 (98%)	4 (2%)	0	100	100
1	B	216/255 (85%)	210 (97%)	6 (3%)	0	100	100
1	C	230/255 (90%)	220 (96%)	10 (4%)	0	100	100
All	All	675/765 (88%)	655 (97%)	20 (3%)	0	100	100

There are no Ramachandran outliers to report.

### 5.3.2 Protein sidechains ⓘ

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	175/187 (94%)	173 (99%)	2 (1%)	84	80
1	B	167/187 (89%)	166 (99%)	1 (1%)	92	92

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	C	178/187 (95%)	174 (98%)	4 (2%)	64	54
All	All	520/561 (93%)	513 (99%)	7 (1%)	80	75

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	110	LEU
1	C	175	GLU
1	C	132	PHE
1	A	77	ARG
1	C	171	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	236	ASN
1	C	144	HIS

### 5.3.3 RNA ⓘ

There are no RNA chains 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 ⓘ

There are no ligands in this entry.

## 5.7 Other polymers ⓘ

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 ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

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	235/255 (92%)	0.49	11 (4%) 30 34	16, 25, 50, 58	0
1	B	224/255 (87%)	0.25	11 (4%) 28 33	16, 25, 44, 64	0
1	C	238/255 (93%)	0.55	20 (8%) 11 13	17, 30, 49, 67	0
All	All	697/765 (91%)	0.43	42 (6%) 21 25	16, 27, 48, 67	0

The worst 5 of 42 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	196	VAL	5.5
1	A	73	VAL	5.5
1	A	28	ALA	5.4
1	B	196	VAL	4.7
1	B	213	PHE	4.2

### 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

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

### 6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

### 6.4 Ligands ⓘ

There are no ligands in this entry.

## 6.5 Other polymers ⓘ

There are no such residues in this entry.