



# wwPDB X-ray Structure Validation Summary Report

Sep 6, 2023 – 03:53 PM EDT

PDB ID : 4FOO  
Title : Crystal Structure of Shikimate Dehydrogenase (aroE) Q237K Mutant from Helicobacter pylori  
Authors : Cheng, W.C.; Chen, T.J.; Wang, W.C.  
Deposited on : 2012-06-20  
Resolution : 2.55 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 : 4.02b-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

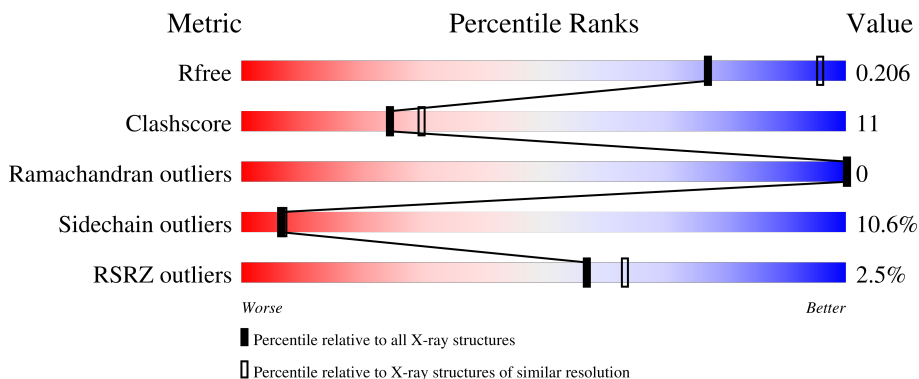
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.55 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	1284 (2.56-2.52)
Clashscore	141614	1332 (2.56-2.52)
Ramachandran outliers	138981	1315 (2.56-2.52)
Sidechain outliers	138945	1315 (2.56-2.52)
RSRZ outliers	127900	1272 (2.56-2.52)

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	271	 70% 22% . .
1	B	271	 72% 21% . .

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 4142 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 Shikimate dehydrogenase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	261	2046	1334	337	365	10	0	0	0
1	B	261	2046	1334	337	365	10	0	0	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	237	LYS	GLN	engineered mutation	UNP P56119
A	264	LEU	-	expression tag	UNP P56119
A	265	GLU	-	expression tag	UNP P56119
A	266	HIS	-	expression tag	UNP P56119
A	267	HIS	-	expression tag	UNP P56119
A	268	HIS	-	expression tag	UNP P56119
A	269	HIS	-	expression tag	UNP P56119
A	270	HIS	-	expression tag	UNP P56119
A	271	HIS	-	expression tag	UNP P56119
B	237	LYS	GLN	engineered mutation	UNP P56119
B	264	LEU	-	expression tag	UNP P56119
B	265	GLU	-	expression tag	UNP P56119
B	266	HIS	-	expression tag	UNP P56119
B	267	HIS	-	expression tag	UNP P56119
B	268	HIS	-	expression tag	UNP P56119
B	269	HIS	-	expression tag	UNP P56119
B	270	HIS	-	expression tag	UNP P56119
B	271	HIS	-	expression tag	UNP P56119

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	24	Total	O	0	0
			24	24		

*Continued on next page...*

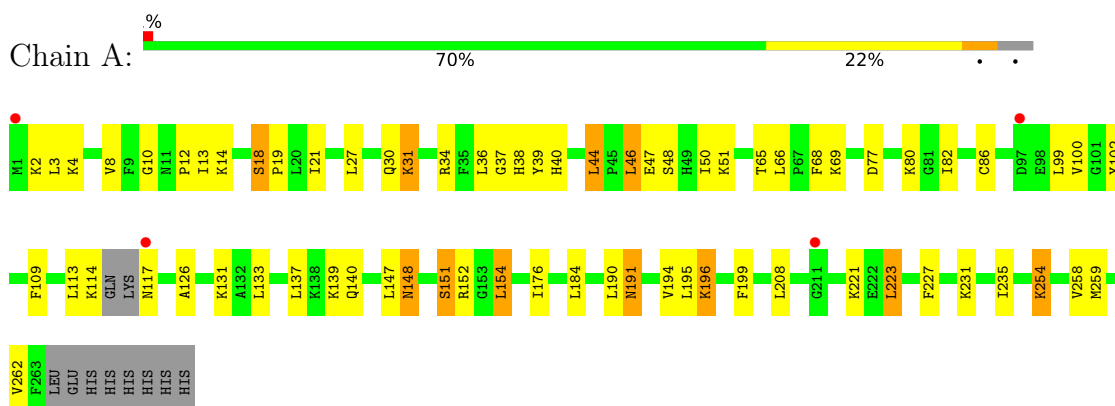
*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
2	B	26	Total	O	0	0
			26	26		

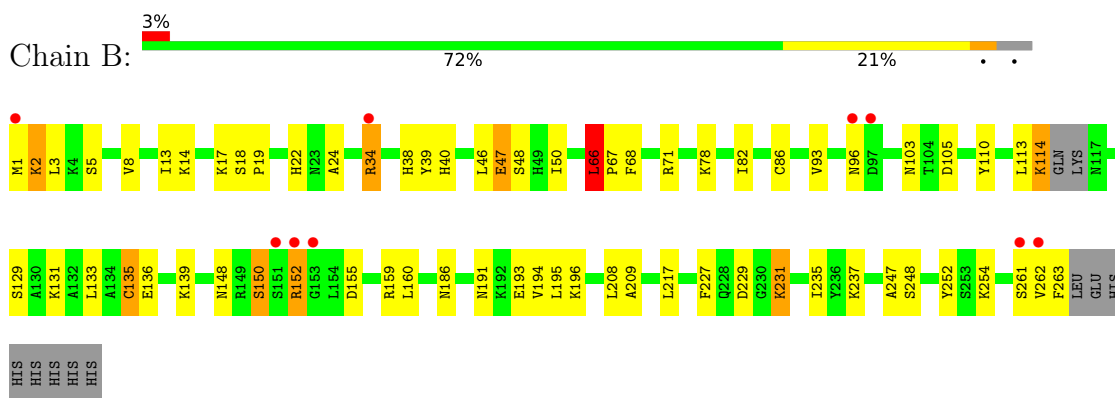
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes 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: Shikimate dehydrogenase



- Molecule 1: Shikimate dehydrogenase



## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	45.81Å 46.09Å 119.08Å 90.00° 99.59° 90.00°	Depositor
Resolution (Å)	30.00 – 2.55 26.95 – 2.55	Depositor EDS
% Data completeness (in resolution range)	99.8 (30.00-2.55) 99.8 (26.95-2.55)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.67 (at 2.54Å)	Xtrriage
Refinement program	REFMAC 5.5.0109	Depositor
R, $R_{free}$	0.192 , 0.231 0.195 , 0.206	Depositor DCC
$R_{free}$ test set	816 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	30.9	Xtrriage
Anisotropy	0.068	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.41 , 61.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.034 for h,-k,-h-l	Xtrriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	4142	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	26.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.67% 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	0.84	0/2092	0.65	1/2815 (0.0%)
1	B	0.89	0/2092	0.69	1/2815 (0.0%)
All	All	0.86	0/4184	0.67	2/5630 (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	66	LEU	CA-CB-CG	8.81	135.56	115.30
1	A	184	LEU	CA-CB-CG	6.96	131.30	115.30

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	2046	0	2086	45	0
1	B	2046	0	2086	46	1
2	A	24	0	0	1	0
2	B	26	0	0	4	0
All	All	4142	0	4172	90	1

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 11.

The worst 5 of 90 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:38:HIS:HD2	1:B:40:HIS:NE2	1.48	1.11
1:B:34:ARG:HH11	1:B:34:ARG:CG	1.78	0.96
1:A:126:ALA:HB2	1:A:148:ASN:OD1	1.66	0.96
1:A:38:HIS:HD2	1:A:40:HIS:NE2	1.67	0.91
1:B:38:HIS:CD2	1:B:40:HIS:NE2	2.40	0.88

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:110:TYR:OH	1:B:155:ASP:OD2[2_655]	1.96	0.24

## 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	257/271 (95%)	255 (99%)	2 (1%)	0	100	100
1	B	257/271 (95%)	251 (98%)	6 (2%)	0	100	100
All	All	514/542 (95%)	506 (98%)	8 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 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	221/231 (96%)	198 (90%)	23 (10%)	7	7
1	B	221/231 (96%)	197 (89%)	24 (11%)	6	6
All	All	442/462 (96%)	395 (89%)	47 (11%)	6	7

5 of 47 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	47	GLU
1	B	114	LYS
1	B	66	LEU
1	B	93	VAL
1	B	136	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 18 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	158	GLN
1	B	228	GLN
1	B	191	ASN
1	A	228	GLN
1	B	143	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](#)

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	261/271 (96%)	-0.23	4 (1%) 73 79	12, 24, 37, 54	0
1	B	261/271 (96%)	-0.12	9 (3%) 45 52	14, 26, 41, 53	0
All	All	522/542 (96%)	-0.18	13 (2%) 57 63	12, 25, 39, 54	0

The worst 5 of 13 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	97	ASP	4.6
1	B	153	GLY	4.3
1	A	1	MET	3.6
1	B	1	MET	3.4
1	A	117	ASN	3.3

### 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.