



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 04:40 PM GMT

PDB ID : 4FQ8  
Title : Crystal Structure of Shikimate Dehydrogenase (aroE) Y210A Mutant from Helicobacter pylori in Complex with Shikimate  
Authors : Cheng, W.C.; Chen, T.J.; Wang, W.C.  
Deposited on : 2012-06-25  
Resolution : 2.07 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  
<http://wwpdb.org/validation/2016/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 : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865

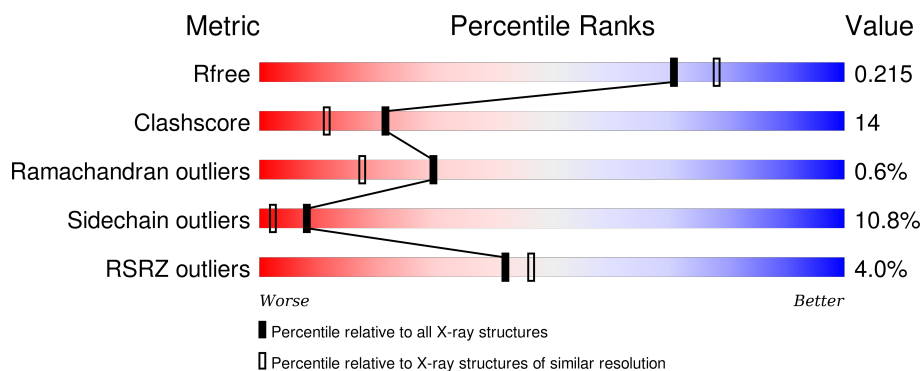
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.07 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))
$R_{free}$	91344	1799 (2.08-2.04)
Clashscore	102246	1910 (2.08-2.04)
Ramachandran outliers	100387	1893 (2.08-2.04)
Sidechain outliers	100360	1893 (2.08-2.04)
RSRZ outliers	91569	1802 (2.08-2.04)

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. 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	
1	B	271	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	SKM	A	500	-	-	-	X

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 4238 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.

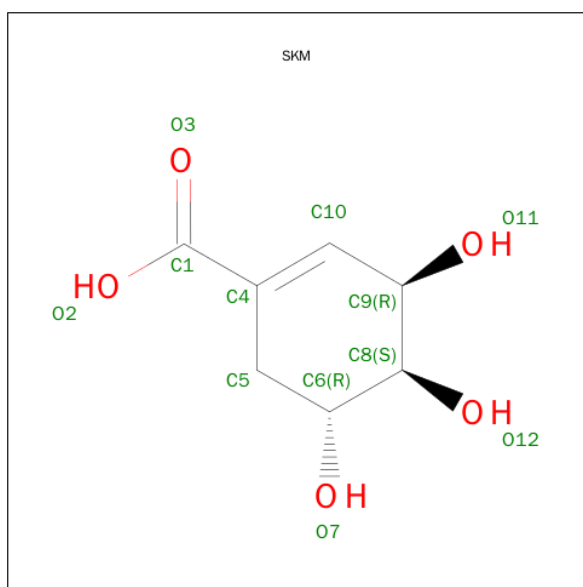
- Molecule 1 is a protein called Shikimate dehydrogenase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	263	Total	C	N	O	S	0	0	0
			2057	1338	341	368	10			
1	B	263	Total	C	N	O	S	0	0	0
			2057	1338	341	368	10			

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	210	ALA	TYR	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	210	ALA	TYR	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 (3R,4S,5R)-3,4,5-TRIHYDROXYCYCLOHEX-1-ENE-1-CARBOXYLIC ACID (three-letter code: SKM) (formula: C<sub>7</sub>H<sub>10</sub>O<sub>5</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			12	7	5		
2	B	1	Total	C	O	0	0
			12	7	5		

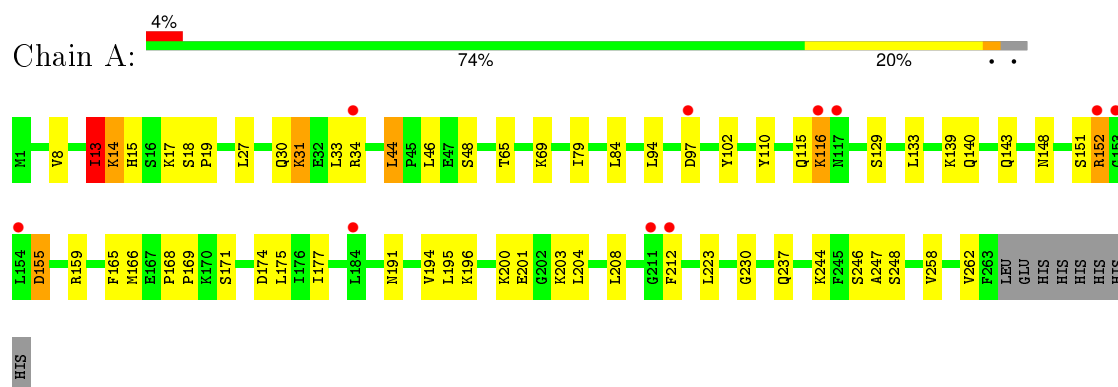
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	49	Total	O	0	0
			49	49		
3	B	51	Total	O	0	0
			51	51		

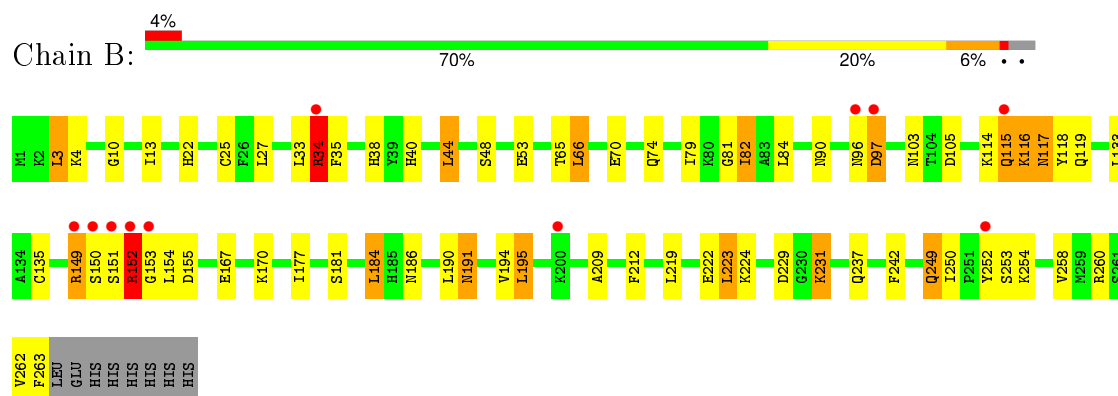
### 3 Residue-property plots

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: Shikimate dehydrogenase



- Molecule 1: Shikimate dehydrogenase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	46.24Å 88.33Å 118.34Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 2.07 28.57 – 2.07	Depositor EDS
% Data completeness (in resolution range)	99.3 (30.00-2.07) 99.3 (28.57-2.07)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	5.93 (at 2.06Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, $R_{free}$	0.209 , 0.251 0.212 , 0.215	Depositor DCC
$R_{free}$ test set	1536 reflections (5.37%)	DCC
Wilson B-factor (Å <sup>2</sup> )	28.4	Xtriage
Anisotropy	0.055	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 47.1	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Outliers	0 of 30150 reflections	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	4238	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 25.52 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 3.1222e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

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

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	0.91	0/2103	0.66	0/2831
1	B	0.93	1/2103 (0.0%)	0.71	0/2831
All	All	0.92	1/4206 (0.0%)	0.68	0/5662

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	B	0	2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	135	CYS	CB-SG	-7.21	1.70	1.82

There are no bond angle outliers.

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	115	GLN	Peptide
1	B	34	ARG	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	2057	0	2099	40	1
1	B	2057	0	2099	78	1
2	A	12	0	9	0	0
2	B	12	0	9	0	0
3	A	49	0	0	3	0
3	B	51	0	0	4	0
All	All	4238	0	4216	118	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 14.

The worst 5 of 118 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:150:SER:C	1:B:152:ARG:HB2	1.49	1.30
1:B:116:LYS:HB3	1:B:117:ASN:CA	1.64	1.28
1:B:116:LYS:CB	1:B:117:ASN:HA	1.64	1.27
1:B:151:SER:N	1:B:152:ARG:HB2	1.60	1.16
1:B:3:LEU:HD21	1:B:38:HIS:CE1	1.80	1.16

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:A:110:TYR:OH	1:B:155:ASP:OD1[4_446]	2.11	0.09

## 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	261/271 (96%)	256 (98%)	4 (2%)	1 (0%)	39	28
1	B	261/271 (96%)	251 (96%)	8 (3%)	2 (1%)	24	12
All	All	522/542 (96%)	507 (97%)	12 (2%)	3 (1%)	30	17

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	152	ARG
1	A	13	ILE
1	B	153	GLY

### 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	222/230 (96%)	202 (91%)	20 (9%)	12	5
1	B	222/230 (96%)	194 (87%)	28 (13%)	5	1
All	All	444/460 (96%)	396 (89%)	48 (11%)	8	2

5 of 48 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	27	LEU
1	B	97	ASP
1	B	249	GLN
1	B	34	ARG
1	B	66	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	228	GLN
1	B	38	HIS
1	B	228	GLN

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	A	237	GLN
1	B	22	HIS

### 5.3.3 RNA ⓘ

There are no RNA molecules 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 ⓘ

2 ligands are modelled in this entry.

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	SKM	A	500	-	10,12,12	2.07	5 (50%)	9,17,17	0.70	0
2	SKM	B	500	-	10,12,12	1.83	3 (30%)	9,17,17	0.59	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	SKM	A	500	-	-	0/0/20/20	0/1/1/1
2	SKM	B	500	-	-	0/0/20/20	0/1/1/1

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	500	SKM	C6-C8	2.07	1.55	1.52
2	B	500	SKM	C5-C4	2.17	1.54	1.50
2	A	500	SKM	C8-C9	2.59	1.58	1.52
2	B	500	SKM	C1-C4	2.65	1.56	1.51
2	A	500	SKM	C10-C4	2.99	1.39	1.34

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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	263/271 (97%)	0.32	10 (3%) 44 49	15, 26, 43, 61	0
1	B	263/271 (97%)	0.36	11 (4%) 40 44	17, 27, 46, 61	0
All	All	526/542 (97%)	0.34	21 (3%) 42 46	15, 27, 44, 61	0

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	152	ARG	6.2
1	B	151	SER	3.9
1	B	34	ARG	3.8
1	B	115	GLN	3.8
1	B	96	ASN	3.7

### 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 carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	SKM	A	500	12/12	0.85	0.21	2.64	32,39,42,42	0
2	SKM	B	500	12/12	0.91	0.15	0.57	22,30,34,35	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.