



# Full wwPDB X-ray Structure Validation Report

Feb 27, 2014 – 08:11 PM GMT

PDB ID : 2EGG  
Title : Crystal Structure of Shikimate 5-Dehydrogenase (AroE) from *Geobacillus kaustophilus*  
Authors : Bagautdinov, B.; Kunishima, N.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)  
Deposited on : 2007-02-28  
Resolution : 2.25 Å(reported)

This is a full wwPDB validation 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>

---

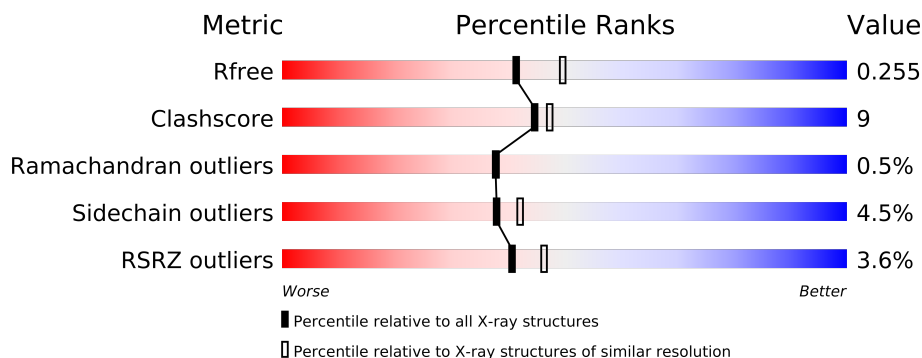
The following versions of software and data (see [references](#)) were used in the production of this report:

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 2.25 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	1108 (2.28-2.24)
Clashscore	79885	1326 (2.28-2.24)
Ramachandran outliers	78287	1291 (2.28-2.24)
Sidechain outliers	78261	1291 (2.28-2.24)
RSRZ outliers	66119	1110 (2.28-2.24)

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	297	
1	B	297	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
2	CL	A	1001	-	X

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 3848 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 Shikimate 5-dehydrogenase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	278	Total	C	N	O	S	0	0	0
			2158	1365	392	394	7			
1	B	166	Total	C	N	O	S	0	0	0
			1286	820	228	233	5			

There are 42 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	EXPRESSION TAG	UNP Q5KWX7
A	2	GLY	-	EXPRESSION TAG	UNP Q5KWX7
A	3	SER	-	EXPRESSION TAG	UNP Q5KWX7
A	4	SER	-	EXPRESSION TAG	UNP Q5KWX7
A	5	HIS	-	EXPRESSION TAG	UNP Q5KWX7
A	6	HIS	-	EXPRESSION TAG	UNP Q5KWX7
A	7	HIS	-	EXPRESSION TAG	UNP Q5KWX7
A	8	HIS	-	EXPRESSION TAG	UNP Q5KWX7
A	9	HIS	-	EXPRESSION TAG	UNP Q5KWX7
A	10	HIS	-	EXPRESSION TAG	UNP Q5KWX7
A	11	SER	-	EXPRESSION TAG	UNP Q5KWX7
A	12	SER	-	EXPRESSION TAG	UNP Q5KWX7
A	13	GLY	-	EXPRESSION TAG	UNP Q5KWX7
A	14	GLU	-	EXPRESSION TAG	UNP Q5KWX7
A	15	ASN	-	EXPRESSION TAG	UNP Q5KWX7
A	16	LEU	-	EXPRESSION TAG	UNP Q5KWX7
A	17	TYR	-	EXPRESSION TAG	UNP Q5KWX7
A	18	PHE	-	EXPRESSION TAG	UNP Q5KWX7
A	19	GLN	-	EXPRESSION TAG	UNP Q5KWX7
A	20	GLY	-	EXPRESSION TAG	UNP Q5KWX7
A	21	HIS	-	EXPRESSION TAG	UNP Q5KWX7
B	1	MET	-	EXPRESSION TAG	UNP Q5KWX7
B	2	GLY	-	EXPRESSION TAG	UNP Q5KWX7
B	3	SER	-	EXPRESSION TAG	UNP Q5KWX7
B	4	SER	-	EXPRESSION TAG	UNP Q5KWX7

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	5	HIS	-	EXPRESSION TAG	UNP Q5KWX7
B	6	HIS	-	EXPRESSION TAG	UNP Q5KWX7
B	7	HIS	-	EXPRESSION TAG	UNP Q5KWX7
B	8	HIS	-	EXPRESSION TAG	UNP Q5KWX7
B	9	HIS	-	EXPRESSION TAG	UNP Q5KWX7
B	10	HIS	-	EXPRESSION TAG	UNP Q5KWX7
B	11	SER	-	EXPRESSION TAG	UNP Q5KWX7
B	12	SER	-	EXPRESSION TAG	UNP Q5KWX7
B	13	GLY	-	EXPRESSION TAG	UNP Q5KWX7
B	14	GLU	-	EXPRESSION TAG	UNP Q5KWX7
B	15	ASN	-	EXPRESSION TAG	UNP Q5KWX7
B	16	LEU	-	EXPRESSION TAG	UNP Q5KWX7
B	17	TYR	-	EXPRESSION TAG	UNP Q5KWX7
B	18	PHE	-	EXPRESSION TAG	UNP Q5KWX7
B	19	GLN	-	EXPRESSION TAG	UNP Q5KWX7
B	20	GLY	-	EXPRESSION TAG	UNP Q5KWX7
B	21	HIS	-	EXPRESSION TAG	UNP Q5KWX7

- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	2	Total Cl 2 2	0	0

- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	291	Total O 291 291	0	0
3	B	111	Total O 111 111	0	0



## 4 Data and refinement statistics

Property	Value	Source
Space group	I 2 3	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	171.39Å 171.39Å 171.39Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	36.54 – 2.25 45.81 – 2.25	Depositor EDS
% Data completeness (in resolution range)	99.1 (36.54-2.25) 99.1 (45.81-2.25)	Depositor EDS
$R_{merge}$	0.08	Depositor
$R_{sym}$	0.06	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.73 (at 2.24Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.218 , 0.255 0.218 , 0.255	Depositor DCC
$R_{free}$ test set	1954 reflections (4.97%)	DCC
Wilson B-factor (Å <sup>2</sup> )	37.0	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 48.3	EDS
Estimated twinning fraction	0.027 for -l,-k,-h	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 39342 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	3848	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.66% 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

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

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	0/2197	0.61	0/2973
1	B	0.31	0/1311	0.56	0/1776
All	All	0.32	0/3508	0.59	0/4749

There are no bond length outliers.

There are no bond angle outliers.

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	2158	0	2186	34	0
1	B	1286	0	1281	29	0
2	A	2	0	0	0	0
3	A	291	0	0	4	0
3	B	111	0	0	1	0
All	All	3848	0	3467	59	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.

All (59) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:B:99:GLU:HA	1:B:102:ARG:HE	1.35	0.91
1:A:112:ASN:HD21	1:A:115:GLY:H	1.23	0.82
1:A:54:ARG:HD3	1:B:54:ARG:HD3	1.66	0.78
1:A:179:GLU:O	1:A:183:ARG:HG3	1.92	0.69
1:A:170:MET:SD	1:A:181:LEU:HD23	2.34	0.68
1:A:48:ARG:NH1	3:A:1017:HOH:O	2.26	0.68
1:B:37:LEU:HD11	1:B:295:LEU:HD13	1.80	0.64
1:A:220:VAL:HG13	3:A:1197:HOH:O	1.97	0.64
1:A:186:ASP:O	1:A:187:GLU:HB2	1.99	0.63
1:A:112:ASN:ND2	1:A:115:GLY:H	1.96	0.62
1:B:134:GLU:HG2	1:B:135:MET:HE3	1.80	0.62
1:A:147:ILE:HD13	1:A:225:LEU:HD22	1.81	0.62
1:B:138:THR:HG22	1:B:140:ASP:H	1.65	0.62
1:A:74:ALA:HA	1:B:64:GLN:NE2	2.15	0.61
1:A:54:ARG:HD3	1:B:54:ARG:CD	2.32	0.59
1:A:197:GLU:O	1:A:201:ARG:HG2	2.01	0.58
1:B:287:MET:O	1:B:291:VAL:HG23	2.04	0.58
1:A:139:LEU:HD12	1:A:163:THR:HG21	1.84	0.57
1:A:112:ASN:HD22	1:A:112:ASN:C	2.06	0.57
1:B:134:GLU:HG2	1:B:135:MET:CE	2.35	0.56
1:A:74:ALA:HA	1:B:64:GLN:HE21	1.69	0.56
1:A:168:ILE:N	1:A:168:ILE:HD12	2.21	0.56
1:A:168:ILE:O	1:A:190:SER:HB3	2.06	0.56
1:B:21:HIS:ND1	1:B:22:MET:N	2.55	0.55
1:B:237:SER:CB	1:B:260:GLN:HB3	2.38	0.52
1:A:167:ARG:HB2	1:A:189:ARG:O	2.10	0.52
1:B:237:SER:HB2	1:B:260:GLN:HB3	1.92	0.51
1:B:239:ILE:O	1:B:262:GLY:HA3	2.11	0.50
1:B:99:GLU:CB	1:B:102:ARG:HH21	2.25	0.50
1:A:231:ARG:O	1:A:234:VAL:HG22	2.11	0.50
1:A:200:THR:O	1:A:229:ARG:HD2	2.12	0.49
1:B:41:MET:HE2	1:B:291:VAL:HG22	1.96	0.48
1:B:39:PRO:HD2	3:B:405:HOH:O	2.15	0.47
1:A:135:MET:O	1:A:137:ILE:HG23	2.14	0.47
1:A:229:ARG:HD3	3:A:1203:HOH:O	2.14	0.47
1:B:99:GLU:HB2	1:B:102:ARG:HH21	1.80	0.46
1:A:112:ASN:ND2	1:A:112:ASN:C	2.68	0.46
1:B:139:LEU:O	1:B:141:GLY:N	2.48	0.46
1:A:44:ASP:OD2	1:A:48:ARG:HD2	2.16	0.45
1:A:216:MET:HG2	1:A:217:HIS:N	2.33	0.45
1:A:209:ILE:HA	1:A:237:SER:O	2.18	0.44
1:A:267:VAL:HG21	1:A:288:LYS:HG2	1.98	0.44

*Continued on next page...*



Continued from previous page...

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:161:LEU:HD23	1:A:168:ILE:HD13	2.00	0.44
1:A:21:HIS:ND1	1:A:21:HIS:N	2.60	0.44
1:A:224:PRO:HD2	1:A:248:TRP:CD1	2.52	0.44
1:B:128:VAL:O	1:B:132:GLU:HG2	2.18	0.43
1:A:195:LEU:O	1:A:199:GLU:HG3	2.18	0.43
1:B:295:LEU:HA	1:B:295:LEU:HD12	1.87	0.43
1:B:135:MET:O	1:B:137:ILE:HG23	2.19	0.43
1:B:260:GLN:HG2	1:B:261:ASN:O	2.19	0.42
1:B:99:GLU:CG	1:B:102:ARG:HH21	2.33	0.42
1:B:91:ILE:HB	1:B:92:PRO:HD3	2.02	0.42
1:B:272:LEU:HD23	1:B:281:TRP:HZ3	1.85	0.42
1:B:261:ASN:HB3	1:B:263:VAL:HG23	2.02	0.41
1:B:38:SER:N	1:B:39:PRO:CD	2.84	0.41
1:A:260:GLN:HG2	1:A:261:ASN:O	2.21	0.41
1:A:176:GLU:HB3	3:A:1282:HOH:O	2.20	0.41
1:A:224:PRO:HD2	1:A:248:TRP:CG	2.56	0.40
1:B:100:HIS:CE1	1:B:104:ILE:HD11	2.57	0.40

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	276/297 (93%)	267 (97%)	9 (3%)	0	100	100
1	B	160/297 (54%)	152 (95%)	6 (4%)	2 (1%)	18	11
All	All	436/594 (73%)	419 (96%)	15 (3%)	2 (0%)	38	38

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	141	GLY
1	B	140	ASP

### 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	222/239 (93%)	211 (95%)	11 (5%)	34	36
1	B	133/239 (56%)	128 (96%)	5 (4%)	44	52
All	All	355/478 (74%)	339 (96%)	16 (4%)	38	42

All (16) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	21	HIS
1	A	54	ARG
1	A	85	PRO
1	A	88	LEU
1	A	112	ASN
1	A	136	ASN
1	A	140	ASP
1	A	145	LEU
1	A	176	GLU
1	A	202	LEU
1	A	225	LEU
1	B	85	PRO
1	B	88	LEU
1	B	136	ASN
1	B	142	LYS
1	B	295	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (8) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	81	ASN
1	A	112	ASN
1	A	113	ASN
1	A	129	GLN
1	A	136	ASN
1	A	289	GLN
1	B	64	GLN
1	B	136	ASN

### 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 ⓘ

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 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	278/297 (93%)	-0.18	5 (1%) 65 72	21, 36, 63, 87	0
1	B	166/297 (55%)	0.05	11 (6%) 18 21	22, 42, 75, 94	0
All	All	444/594 (74%)	-0.09	16 (3%) 41 46	21, 38, 73, 94	0

All (16) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	218	PRO	5.0
1	B	139	LEU	3.4
1	B	125	LEU	3.3
1	B	103	ARG	3.2
1	A	217	HIS	2.9
1	A	188	ARG	2.9
1	B	102	ARG	2.9
1	B	239	ILE	2.8
1	B	237	SER	2.8
1	B	99	GLU	2.7
1	B	101	ALA	2.6
1	B	140	ASP	2.3
1	B	138	THR	2.3
1	A	187	GLU	2.2
1	B	128	VAL	2.1
1	A	92	PRO	2.1

### 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

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	RSR	LLDF	B-factors(Å <sup>2</sup> )	Q<0.9
2	CL	A	1001	1/1	0.22	8.11	65,65,65,65	0
2	CL	A	1002	1/1	0.07	-2.36	35,35,35,35	0

### 6.5 Other polymers

There are no such residues in this entry.