



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 31, 2016 – 10:31 PM GMT

PDB ID : 1U0V  
Title : An Aldol Switch Discovered in Stilbene Synthases Mediates Cyclization of Specificity of Type III Polyketide Synthases: 18xCHS structure  
Authors : Austin, M.B.; Bowman, M.E.; Ferrer, J.-L.; Schroder, J.; Noel, J.P.  
Deposited on : 2004-07-14  
Resolution : 1.90 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

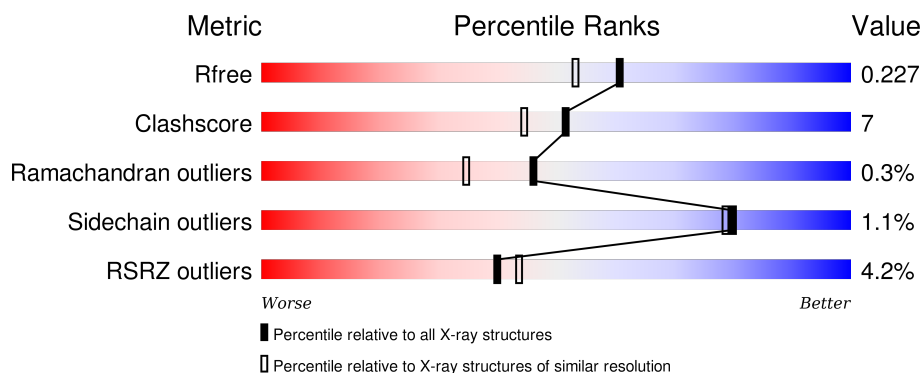
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

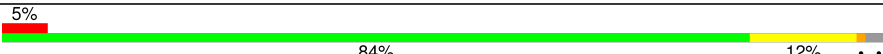
The reported resolution of this entry is 1.90 Å.

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	4755 (1.90-1.90)
Clashscore	102246	5398 (1.90-1.90)
Ramachandran outliers	100387	5338 (1.90-1.90)
Sidechain outliers	100360	5339 (1.90-1.90)
RSRZ outliers	91569	4766 (1.90-1.90)

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	393	 4% 87% 10% ..
1	B	393	 5% 84% 12% ..

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 6597 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 Chalcone synthase 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	388	Total	C	N	O	S	0	0	0
			2970	1889	504	560	17			
1	B	380	Total	C	N	O	S	0	0	0
			2911	1853	493	548	17			

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	GLY	-	CLONING ARTIFACT	UNP P30074
A	-2	SER	-	CLONING ARTIFACT	UNP P30074
A	-1	HIS	-	CLONING ARTIFACT	UNP P30074
A	0	GLY	-	CLONING ARTIFACT	UNP P30074
A	96	ALA	ASP	ENGINEERED	UNP P30074
A	98	LEU	VAL	ENGINEERED	UNP P30074
A	99	ALA	VAL	ENGINEERED	UNP P30074
A	100	MET	VAL	ENGINEERED	UNP P30074
A	131	SER	THR	ENGINEERED	UNP P30074
A	133	THR	SER	ENGINEERED	UNP P30074
A	134	THR	GLY	ENGINEERED	UNP P30074
A	135	PRO	VAL	ENGINEERED	UNP P30074
A	137	LEU	MET	ENGINEERED	UNP P30074
A	157	VAL	TYR	ENGINEERED	UNP P30074
A	158	GLY	MET	ENGINEERED	UNP P30074
A	159	VAL	MET	ENGINEERED	UNP P30074
A	160	PHE	TYR	ENGINEERED	UNP P30074
A	162	HIS	GLN	ENGINEERED	UNP P30074
A	268	LYS	LEU	ENGINEERED	UNP P30074
A	269	GLY	LYS	ENGINEERED	UNP P30074
A	270	ALA	ASP	ENGINEERED	UNP P30074
A	273	ASP	GLY	ENGINEERED	UNP P30074
B	-3	GLY	-	CLONING ARTIFACT	UNP P30074
B	-2	SER	-	CLONING ARTIFACT	UNP P30074
B	-1	HIS	-	CLONING ARTIFACT	UNP P30074

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	GLY	-	CLONING ARTIFACT	UNP P30074
B	96	ALA	ASP	ENGINEERED	UNP P30074
B	98	LEU	VAL	ENGINEERED	UNP P30074
B	99	ALA	VAL	ENGINEERED	UNP P30074
B	100	MET	VAL	ENGINEERED	UNP P30074
B	131	SER	THR	ENGINEERED	UNP P30074
B	133	THR	SER	ENGINEERED	UNP P30074
B	134	THR	GLY	ENGINEERED	UNP P30074
B	135	PRO	VAL	ENGINEERED	UNP P30074
B	137	LEU	MET	ENGINEERED	UNP P30074
B	157	VAL	TYR	ENGINEERED	UNP P30074
B	158	GLY	MET	ENGINEERED	UNP P30074
B	159	VAL	MET	ENGINEERED	UNP P30074
B	160	PHE	TYR	ENGINEERED	UNP P30074
B	162	HIS	GLN	ENGINEERED	UNP P30074
B	268	LYS	LEU	ENGINEERED	UNP P30074
B	269	GLY	LYS	ENGINEERED	UNP P30074
B	270	ALA	ASP	ENGINEERED	UNP P30074
B	273	ASP	GLY	ENGINEERED	UNP P30074

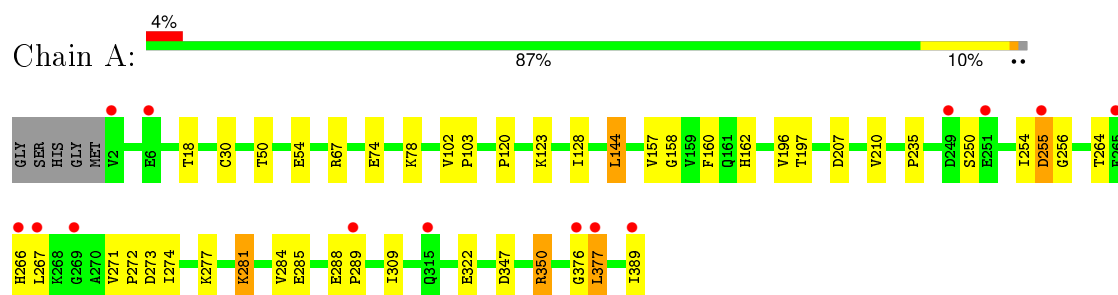
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	360	Total O 360 360	0	0
2	B	356	Total O 356 356	0	0

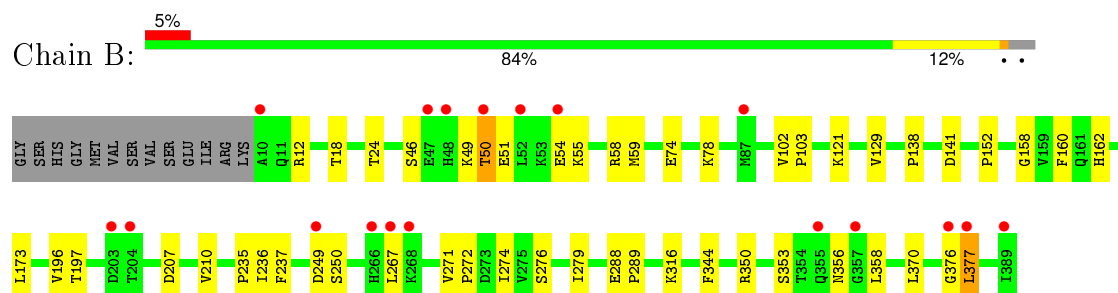
### 3 Residue-property plots [i](#)

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: Chalcone synthase 2



#### • Molecule 1: Chalcone synthase 2



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	71.64Å 59.75Å 82.54Å 90.00° 108.17° 90.00°	Depositor
Resolution (Å)	24.93 – 1.90 24.93 – 1.80	Depositor EDS
% Data completeness (in resolution range)	100.0 (24.93-1.90) 98.7 (24.93-1.80)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.25 (at 1.80Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, $R_{free}$	0.192 , 0.227 0.191 , 0.227	Depositor DCC
$R_{free}$ test set	2639 reflections (5.03%)	DCC
Wilson B-factor (Å <sup>2</sup> )	18.8	Xtriage
Anisotropy	0.331	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 60.7	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	1 of 60789 reflections (0.002%)	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	6597	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	20.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.37% 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, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

## 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	0.34	0/3028	0.61	0/4102
1	B	0.34	0/2969	0.62	0/4022
All	All	0.34	0/5997	0.61	0/8124

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-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 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	2970	0	3010	41	0
1	B	2911	0	2950	41	0
2	A	360	0	0	2	0
2	B	356	0	0	6	0
All	All	6597	0	5960	80	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 7.

The worst 5 of 80 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:250:SER:HB2	1:A:377:LEU:HD12	1.45	0.95

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:207:ASP:HB2	1:A:267:LEU:HD13	1.48	0.94
1:A:274:ILE:HG21	1:A:377:LEU:HD13	1.49	0.93
1:B:274:ILE:HG21	1:B:377:LEU:HD13	1.60	0.83
1:B:210:VAL:HG11	1:B:267:LEU:HD11	1.62	0.80

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	386/393 (98%)	373 (97%)	12 (3%)	1 (0%)	46	35
1	B	378/393 (96%)	363 (96%)	14 (4%)	1 (0%)	46	35
All	All	764/786 (97%)	736 (96%)	26 (3%)	2 (0%)	46	35

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	50	THR
1	A	377	LEU

### 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.

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	320/324 (99%)	316 (99%)	4 (1%)	76	73
1	B	313/324 (97%)	310 (99%)	3 (1%)	82	81
All	All	633/648 (98%)	626 (99%)	7 (1%)	80	79

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

Mol	Chain	Res	Type
1	A	350	ARG
1	B	377	LEU
1	B	249	ASP
1	A	255	ASP
1	B	350	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	11	GLN
1	B	162	HIS
1	B	82	ASN
1	A	143	GLN
1	B	119	GLN

### 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 [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	388/393 (98%)	-0.04	14 (3%) 46 50	8, 16, 29, 40	0
1	B	380/393 (96%)	-0.02	18 (4%) 35 38	8, 17, 34, 44	0
All	All	768/786 (97%)	-0.03	32 (4%) 40 44	8, 17, 31, 44	0

The worst 5 of 32 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	10	ALA	9.0
1	A	269	GLY	8.6
1	B	267	LEU	5.6
1	A	266	HIS	5.3
1	B	50	THR	5.3

### 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 carbohydrates 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.