



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jul 31, 2017 – 09:03 PM EDT

PDB ID : 5I0K  
Title : Insights into Substrate Modification by Dehydratases from Type I Polyketide Synthases  
Authors : Faille, A.; Mourey, L.; Pedelacq, J.D.  
Deposited on : unknown  
Resolution : 3.20 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	:	1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20029824
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	rb-20029824

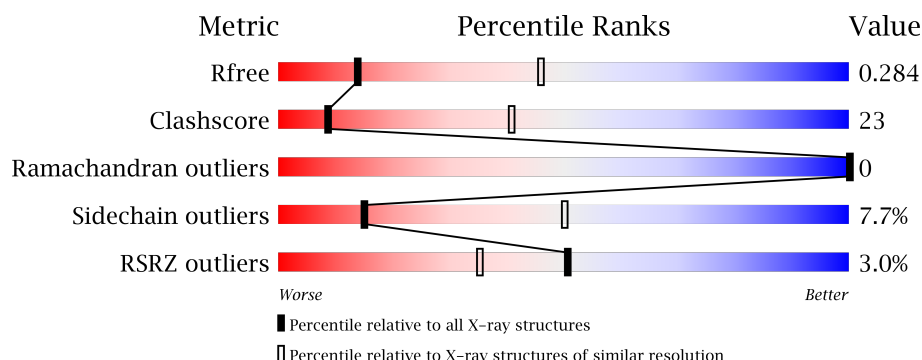
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.20 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))
$R_{free}$	100719	1015 (3.22-3.18)
Clashscore	112137	1009 (3.20-3.20)
Ramachandran outliers	110173	1118 (3.22-3.18)
Sidechain outliers	110143	1117 (3.22-3.18)
RSRZ outliers	101464	1020 (3.22-3.18)

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	322	

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 2026 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.

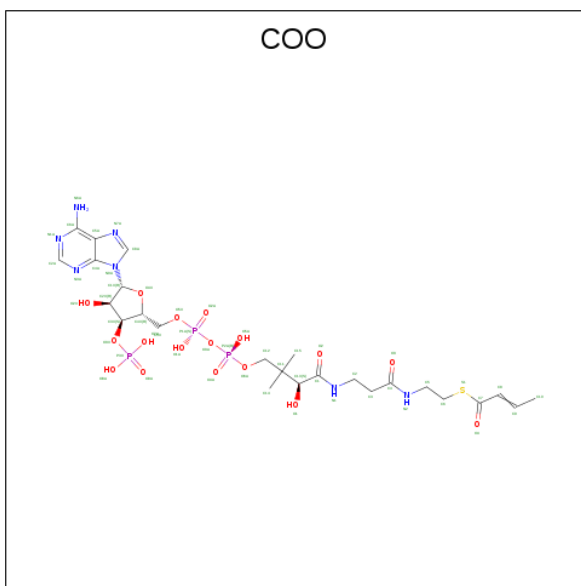
- Molecule 1 is a protein called Phthiocerol synthesis polyketide synthase type I PpsC.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	265	Total	C	N	O	S	0	0	0
			1973	1235	344	387	7			

There are 26 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	900	MET	-	initiating methionine	UNP P96202
A	901	GLY	-	expression tag	UNP P96202
A	902	SER	-	expression tag	UNP P96202
A	903	SER	-	expression tag	UNP P96202
A	904	HIS	-	expression tag	UNP P96202
A	905	HIS	-	expression tag	UNP P96202
A	906	HIS	-	expression tag	UNP P96202
A	907	HIS	-	expression tag	UNP P96202
A	908	HIS	-	expression tag	UNP P96202
A	909	HIS	-	expression tag	UNP P96202
A	910	SER	-	expression tag	UNP P96202
A	911	SER	-	expression tag	UNP P96202
A	912	GLY	-	expression tag	UNP P96202
A	913	LEU	-	expression tag	UNP P96202
A	914	VAL	-	expression tag	UNP P96202
A	915	PRO	-	expression tag	UNP P96202
A	916	ARG	-	expression tag	UNP P96202
A	917	GLY	-	expression tag	UNP P96202
A	918	SER	-	expression tag	UNP P96202
A	919	HIS	-	expression tag	UNP P96202
A	920	MET	-	expression tag	UNP P96202
A	959	PHE	HIS	engineered mutation	UNP P96202
A	1218	THR	-	expression tag	UNP P96202
A	1219	SER	-	expression tag	UNP P96202
A	1220	GLY	-	expression tag	UNP P96202
A	1221	SER	-	expression tag	UNP P96202

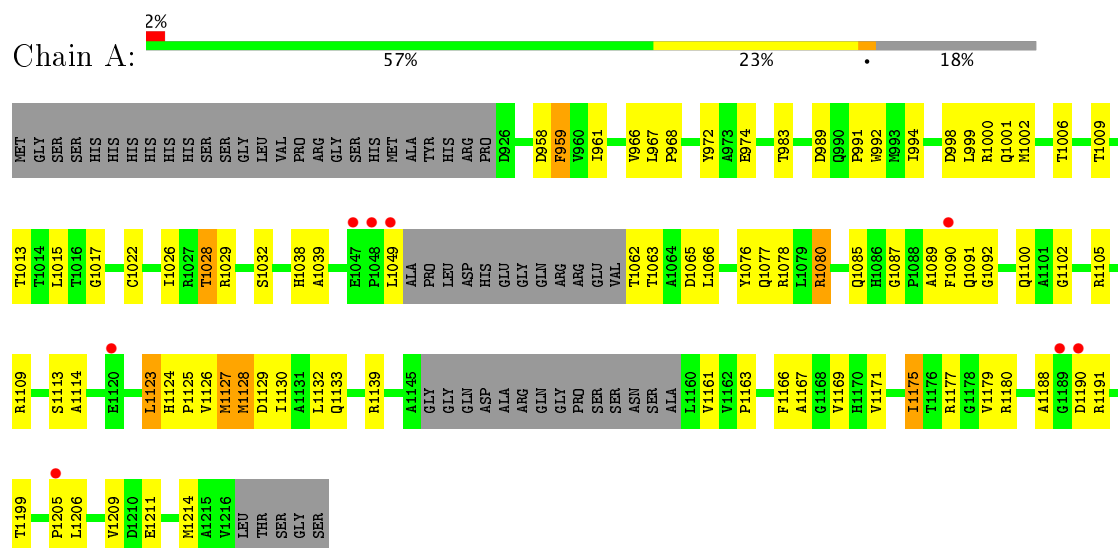
- Molecule 2 is CROTONYL COENZYME A (three-letter code: COO) (formula:  $C_{25}H_{40}N_7O_{17}P_3S$ ).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
2	A	1	Total	C	N	O	P	S	0	0
			53	25	7	17	3	1		



- Molecule 1: Phthiocerol synthesis polyketide synthase type I PpsC



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	83.95Å 83.95Å 166.98Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.39 – 3.20 46.39 – 3.20	Depositor EDS
% Data completeness (in resolution range)	89.1 (46.39-3.20) 71.2 (46.39-3.20)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.60 (at 3.19Å)	Xtriage
Refinement program	PHENIX (1.11.1 _2575: ???)	Depositor
R, $R_{free}$	0.251 , 0.289 0.272 , 0.284	Depositor DCC
$R_{free}$ test set	802 reflections (9.70%)	DCC
Wilson B-factor (Å <sup>2</sup> )	70.7	Xtriage
Anisotropy	0.694	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 58.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.89	EDS
Total number of atoms	2026	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	89.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.67% 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](#)

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

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	0.58	0/2005	0.58	0/2744

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 [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	1973	0	1950	84	0
2	A	53	0	37	18	0
All	All	2026	0	1987	91	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 23.

The worst 5 of 91 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:1076:TYR:CE2	1:A:1091:GLN:HG2	1.80	1.15
1:A:1163:PRO:HD2	2:A:1301:COO:H102	1.43	0.98

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:1076:TYR:CE2	1:A:1091:GLN:CG	2.47	0.96
1:A:1076:TYR:HE2	1:A:1091:GLN:HG2	1.31	0.89
1:A:1076:TYR:OH	1:A:1091:GLN:HA	1.74	0.88

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	259/322 (80%)	228 (88%)	31 (12%)	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	207/254 (82%)	191 (92%)	16 (8%)	15	50

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1065	ASP
1	A	1077	GLN
1	A	1127	MET

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
1	A	1063	THR
1	A	1128	MET

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1133	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 carbohydrates in this entry.

### 5.6 Ligand geometry [i](#)

1 ligand is 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	COO	A	1301	-	46,55,55	0.80	1 (2%)	53,81,81	1.36	4 (7%)

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	COO	A	1301	-	-	3/49/70/70	0/3/3/3

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1301	COO	C5A-C4A	3.11	1.47	1.40

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1301	COO	N3A-C2A-N1A	-5.91	123.71	128.86
2	A	1301	COO	C4A-C5A-N7A	-2.87	106.64	109.41
2	A	1301	COO	C10-C9-C8	-2.66	120.09	125.36
2	A	1301	COO	C3X-C2X-C1X	2.73	106.09	99.95

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	1301	COO	O4-C7-S1-C6
2	A	1301	COO	C8-C7-S1-C6
2	A	1301	COO	C3-C4-N2-C5

There are no ring outliers.

1 monomer is involved in 18 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1301	COO	18	0

## 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	265/322 (82%)	0.14	8 (3%) 51 35	30, 84, 115, 134	0

The worst 5 of 8 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1048	PRO	2.6
1	A	1049	LEU	2.5
1	A	1047	GLU	2.5
1	A	1090	PHE	2.2
1	A	1120	GLU	2.2

### 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(Å <sup>2</sup> )	Q<0.9
2	COO	A	1301	53/53	0.77	0.32	0.63	86,205,242,245	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.