



wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 31, 2016 – 08:27 PM GMT

PDB ID : 1KEZ
Title : Crystal Structure of the Macrocycle-forming Thioesterase Domain of Erythromycin Polyketide Synthase (DEBS TE)
Authors : Tsai, S.-C.; Miercke, L.J.W.; Krucinski, J.; Gokhale, R.; Chen, J.C.-H.; Foster, P.G.; Cane, D.E.; Khosla, C.; Stroud, R.M.
Deposited on : 2001-11-19
Resolution : 2.80 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
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

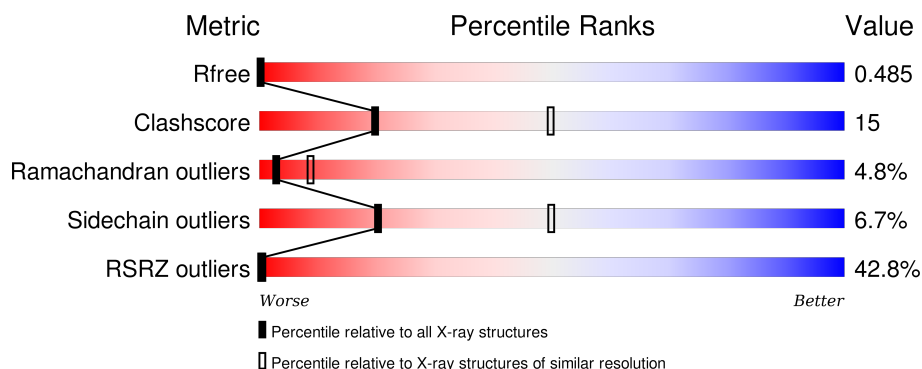
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.80 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	2393 (2.80-2.80)
Clashscore	102246	2827 (2.80-2.80)
Ramachandran outliers	100387	2782 (2.80-2.80)
Sidechain outliers	100360	2784 (2.80-2.80)
RSRZ outliers	91569	2404 (2.80-2.80)

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 ≥ 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	300	<div> <div>27%</div> <div>61% 23% 5% 11%</div> </div>
1	B	300	<div> <div>57%</div> <div>59% 24% 6% 11%</div> </div>
1	C	300	<div> <div>30%</div> <div>61% 27% • 11%</div> </div>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 6426 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 ERYTHRONOLIDE SYNTHASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	267	Total	C	N	O	S	0	0	0
			2011	1263	355	384	9			
1	B	267	Total	C	N	O	S	0	0	0
			2011	1263	355	384	9			
1	C	267	Total	C	N	O	S	0	0	0
			2011	1263	355	384	9			

There are 60 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	CLONING ARTIFACT	UNP Q03133
A	2	ALA	-	CLONING ARTIFACT	UNP Q03133
A	3	SER	-	CLONING ARTIFACT	UNP Q03133
A	284	SER	-	EXPRESSION TAG	UNP Q03133
A	285	SER	-	EXPRESSION TAG	UNP Q03133
A	286	VAL	-	EXPRESSION TAG	UNP Q03133
A	287	ASP	-	EXPRESSION TAG	UNP Q03133
A	288	LYS	-	EXPRESSION TAG	UNP Q03133
A	289	LEU	-	EXPRESSION TAG	UNP Q03133
A	290	ALA	-	EXPRESSION TAG	UNP Q03133
A	291	ALA	-	EXPRESSION TAG	UNP Q03133
A	292	ALA	-	EXPRESSION TAG	UNP Q03133
A	293	LEU	-	EXPRESSION TAG	UNP Q03133
A	294	GLU	-	EXPRESSION TAG	UNP Q03133
A	295	HIS	-	EXPRESSION TAG	UNP Q03133
A	296	HIS	-	EXPRESSION TAG	UNP Q03133
A	297	HIS	-	EXPRESSION TAG	UNP Q03133
A	298	HIS	-	EXPRESSION TAG	UNP Q03133
A	299	HIS	-	EXPRESSION TAG	UNP Q03133
A	300	HIS	-	EXPRESSION TAG	UNP Q03133
B	1	MET	-	CLONING ARTIFACT	UNP Q03133
B	2	ALA	-	CLONING ARTIFACT	UNP Q03133
B	3	SER	-	CLONING ARTIFACT	UNP Q03133

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	284	SER	-	EXPRESSION TAG	UNP Q03133
B	285	SER	-	EXPRESSION TAG	UNP Q03133
B	286	VAL	-	EXPRESSION TAG	UNP Q03133
B	287	ASP	-	EXPRESSION TAG	UNP Q03133
B	288	LYS	-	EXPRESSION TAG	UNP Q03133
B	289	LEU	-	EXPRESSION TAG	UNP Q03133
B	290	ALA	-	EXPRESSION TAG	UNP Q03133
B	291	ALA	-	EXPRESSION TAG	UNP Q03133
B	292	ALA	-	EXPRESSION TAG	UNP Q03133
B	293	LEU	-	EXPRESSION TAG	UNP Q03133
B	294	GLU	-	EXPRESSION TAG	UNP Q03133
B	295	HIS	-	EXPRESSION TAG	UNP Q03133
B	296	HIS	-	EXPRESSION TAG	UNP Q03133
B	297	HIS	-	EXPRESSION TAG	UNP Q03133
B	298	HIS	-	EXPRESSION TAG	UNP Q03133
B	299	HIS	-	EXPRESSION TAG	UNP Q03133
B	300	HIS	-	EXPRESSION TAG	UNP Q03133
C	1	MET	-	CLONING ARTIFACT	UNP Q03133
C	2	ALA	-	CLONING ARTIFACT	UNP Q03133
C	3	SER	-	CLONING ARTIFACT	UNP Q03133
C	284	SER	-	EXPRESSION TAG	UNP Q03133
C	285	SER	-	EXPRESSION TAG	UNP Q03133
C	286	VAL	-	EXPRESSION TAG	UNP Q03133
C	287	ASP	-	EXPRESSION TAG	UNP Q03133
C	288	LYS	-	EXPRESSION TAG	UNP Q03133
C	289	LEU	-	EXPRESSION TAG	UNP Q03133
C	290	ALA	-	EXPRESSION TAG	UNP Q03133
C	291	ALA	-	EXPRESSION TAG	UNP Q03133
C	292	ALA	-	EXPRESSION TAG	UNP Q03133
C	293	LEU	-	EXPRESSION TAG	UNP Q03133
C	294	GLU	-	EXPRESSION TAG	UNP Q03133
C	295	HIS	-	EXPRESSION TAG	UNP Q03133
C	296	HIS	-	EXPRESSION TAG	UNP Q03133
C	297	HIS	-	EXPRESSION TAG	UNP Q03133
C	298	HIS	-	EXPRESSION TAG	UNP Q03133
C	299	HIS	-	EXPRESSION TAG	UNP Q03133
C	300	HIS	-	EXPRESSION TAG	UNP Q03133

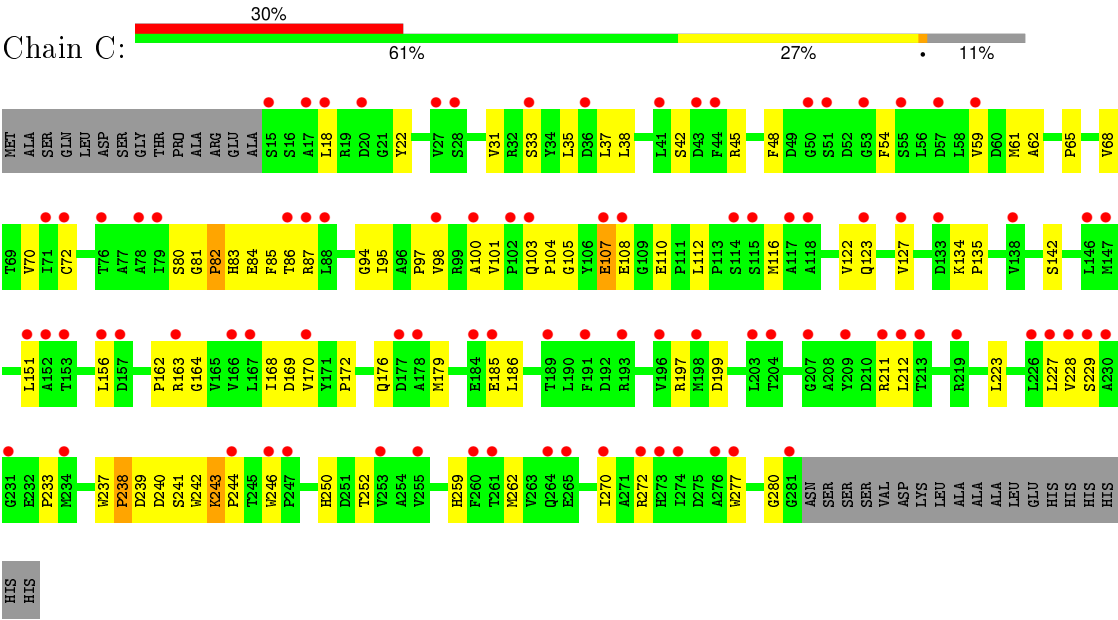
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	83	Total O 83 83	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	217	Total 217	O 217	0	0
2	C	93	Total 93	O 93	0	0



4 Data and refinement statistics

Property	Value	Source
Space group	P 31 2 1	Depositor
Cell constants a, b, c, α , β , γ	130.50Å 130.50Å 208.50Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	30.00 – 2.80 49.68 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.1 (30.00-2.80) 89.6 (49.68-2.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.03	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.43 (at 2.81Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.254 , 0.279 0.474 , 0.485	Depositor DCC
R_{free} test set	2276 reflections (4.95%)	DCC
Wilson B-factor (Å ²)	79.6	Xtriage
Anisotropy	0.107	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.28 , -10.0	EDS
Estimated twinning fraction	0.019 for -h,-k,l	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 50859 reflections	Xtriage
F_o, F_c correlation	0.73	EDS
Total number of atoms	6426	wwPDB-VP
Average B, all atoms (Å ²)	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.44% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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](#)

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	0.35	0/2065	0.58	0/2821
1	B	0.40	0/2065	0.66	0/2821
1	C	0.35	0/2065	0.58	0/2821
All	All	0.37	0/6195	0.61	0/8463

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	2011	0	1925	64	0
1	B	2011	0	1925	66	0
1	C	2011	0	1925	50	0
2	A	83	0	0	12	0
2	B	217	0	0	26	0
2	C	93	0	0	2	0
All	All	6426	0	5775	176	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 15.

The worst 5 of 176 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:69:THR:HG23	1:B:97:PRO:HB2	1.53	0.90
1:C:107:GLU:HG3	1:C:108:GLU:N	1.98	0.78
1:A:71:ILE:HD13	1:A:136:PHE:HB2	1.67	0.76
1:A:169:ASP:HA	1:A:244:PRO:HG3	1.69	0.74
1:A:74:ALA:HB3	1:A:103:GLN:HG2	1.70	0.73

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	265/300 (88%)	219 (83%)	39 (15%)	7 (3%)	7	22
1	B	265/300 (88%)	189 (71%)	54 (20%)	22 (8%)	1	2
1	C	265/300 (88%)	218 (82%)	38 (14%)	9 (3%)	5	16
All	All	795/900 (88%)	626 (79%)	131 (16%)	38 (5%)	3	9

5 of 38 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	193	ARG
1	A	238	PRO
1	B	78	ALA
1	B	131	GLN
1	B	200	ASP

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	204/230 (89%)	189 (93%)	15 (7%)	17	43
1	B	204/230 (89%)	187 (92%)	17 (8%)	14	38
1	C	204/230 (89%)	195 (96%)	9 (4%)	35	69
All	All	612/690 (89%)	571 (93%)	41 (7%)	20	50

5 of 41 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	43	ASP
1	B	134	LYS
1	C	199	ASP
1	B	56	LEU
1	B	103	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	131	GLN
1	B	264	GLN
1	B	103	GLN
1	A	47	HIS
1	B	141	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 [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, 95th 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 ²)	Q<0.9
1	A	267/300 (89%)	1.68	81 (30%) 1 0	52, 64, 88, 123	0
1	B	267/300 (89%)	3.92	172 (64%) 0 0	32, 68, 76, 83	0
1	C	267/300 (89%)	1.77	90 (33%) 0 0	52, 64, 90, 118	0
All	All	801/900 (89%)	2.46	343 (42%) 0 0	32, 65, 86, 123	0

The worst 5 of 343 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	139	ALA	22.3
1	B	71	ILE	21.7
1	B	167	LEU	18.6
1	B	151	LEU	18.3
1	B	72	CYS	18.2

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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.