



# Full wwPDB X-ray Structure Validation Report ⓘ

Nov 15, 2016 – 01:48 PM EST

PDB ID : 5TT4  
Title : Determining the Molecular Basis For Starter Unit Selection During Daunorubicin Biosynthesis  
Authors : Jackson, D.R.; Valentic, T.R.; Patel, A.; Tsai, S.C.; Mohammed, L.; Vasilakis, K.; Wattana-amorn, P.; Long, P.F.; Crump, M.P.; Crosby, J.  
Deposited on : 2016-11-01  
Resolution : 2.50 Å(reported)

This is a Full wwPDB X-ray Structure 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/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	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20028320
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)	:	rb-20028320

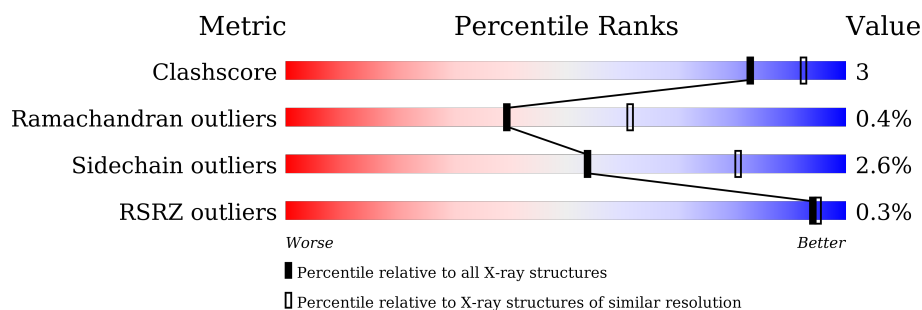
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.50 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))
Clashscore	102246	4242 (2.50-2.50)
Ramachandran outliers	100387	4156 (2.50-2.50)
Sidechain outliers	100360	4158 (2.50-2.50)
RSRZ outliers	91569	3562 (2.50-2.50)

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	373	<div> <div></div> <div>83%</div> <div>8%</div> <div>9%</div> </div>
1	B	373	<div> <div></div> <div>82%</div> <div>9%</div> <div>9%</div> </div>

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 5231 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 Daunorubicin-doxorubicin polyketide synthase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	341	Total	C	N	O	S	0	0	0
			2539	1592	468	467	12			
1	B	339	Total	C	N	O	S	0	0	0
			2519	1581	462	464	12			

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-19	MET	-	expression tag	UNP Q54816
A	-18	GLY	-	expression tag	UNP Q54816
A	-17	SER	-	expression tag	UNP Q54816
A	-16	SER	-	expression tag	UNP Q54816
A	-15	HIS	-	expression tag	UNP Q54816
A	-14	HIS	-	expression tag	UNP Q54816
A	-13	HIS	-	expression tag	UNP Q54816
A	-12	HIS	-	expression tag	UNP Q54816
A	-11	HIS	-	expression tag	UNP Q54816
A	-10	HIS	-	expression tag	UNP Q54816
A	-9	SER	-	expression tag	UNP Q54816
A	-8	SER	-	expression tag	UNP Q54816
A	-7	GLY	-	expression tag	UNP Q54816
A	-6	LEU	-	expression tag	UNP Q54816
A	-5	VAL	-	expression tag	UNP Q54816
A	-4	PRO	-	expression tag	UNP Q54816
A	-3	ARG	-	expression tag	UNP Q54816
A	-2	GLY	-	expression tag	UNP Q54816
A	-1	SER	-	expression tag	UNP Q54816
A	0	HIS	-	expression tag	UNP Q54816
B	-19	MET	-	expression tag	UNP Q54816
B	-18	GLY	-	expression tag	UNP Q54816
B	-17	SER	-	expression tag	UNP Q54816
B	-16	SER	-	expression tag	UNP Q54816
B	-15	HIS	-	expression tag	UNP Q54816

*Continued on next page...*

*Continued from previous page...*

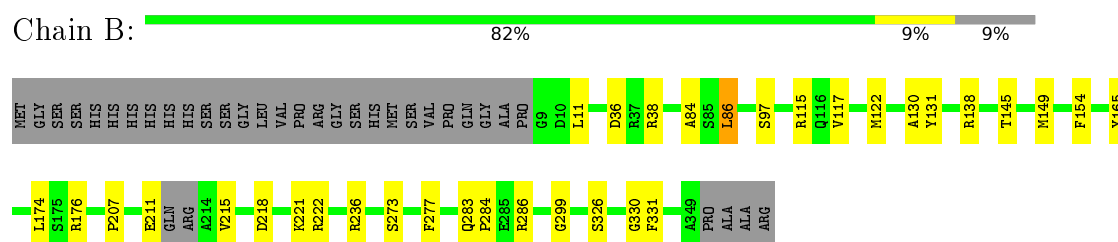
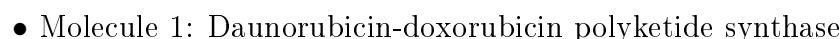
Chain	Residue	Modelled	Actual	Comment	Reference
B	-14	HIS	-	expression tag	UNP Q54816
B	-13	HIS	-	expression tag	UNP Q54816
B	-12	HIS	-	expression tag	UNP Q54816
B	-11	HIS	-	expression tag	UNP Q54816
B	-10	HIS	-	expression tag	UNP Q54816
B	-9	SER	-	expression tag	UNP Q54816
B	-8	SER	-	expression tag	UNP Q54816
B	-7	GLY	-	expression tag	UNP Q54816
B	-6	LEU	-	expression tag	UNP Q54816
B	-5	VAL	-	expression tag	UNP Q54816
B	-4	PRO	-	expression tag	UNP Q54816
B	-3	ARG	-	expression tag	UNP Q54816
B	-2	GLY	-	expression tag	UNP Q54816
B	-1	SER	-	expression tag	UNP Q54816
B	0	HIS	-	expression tag	UNP Q54816

- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	84	Total O 84 84	0	0
2	B	89	Total O 89 89	0	0



- Molecule 1: Daunorubicin-doxorubicin polyketide synthase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	90.03Å 90.03Å 304.69Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	48.01 – 2.50 48.01 – 2.50	Depositor EDS
% Data completeness (in resolution range)	99.5 (48.01-2.50) 99.5 (48.01-2.50)	Depositor EDS
$R_{merge}$	0.14	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	6.66 (at 2.48Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE: 1.8.2_1309)	Depositor
R, $R_{free}$	0.162 , 0.209 0.155 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	DCC
Wilson B-factor (Å <sup>2</sup> )	31.6	Xtriage
Anisotropy	0.351	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 36.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	5231	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.61% 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

### 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	0.45	0/2604	0.55	0/3549
1	B	0.41	0/2583	0.53	0/3520
All	All	0.43	0/5187	0.54	0/7069

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	2539	0	2476	18	0
1	B	2519	0	2454	18	0
2	A	84	0	0	3	0
2	B	89	0	0	1	0
All	All	5231	0	4930	31	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 3.

All (31) 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:113:GLU:HB2	1:B:115:ARG:HB2	1.81	0.63

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:218:ASP:OD2	1:A:222:ARG:NH1	2.39	0.55
1:A:261:ARG:NH1	2:A:405:HOH:O	2.39	0.54
1:B:86:LEU:HD21	1:B:299:GLY:HA2	1.89	0.54
1:B:149:MET:HG3	1:B:165:TYR:HB3	1.89	0.53
1:A:265:PRO:HB2	1:A:267:MET:HG3	1.95	0.48
1:A:330:GLY:N	1:A:331:PHE:HA	2.29	0.48
1:B:218:ASP:OD2	1:B:222:ARG:HD2	2.14	0.48
1:A:316:LEU:HD21	1:A:322:CYS:HB2	1.95	0.48
1:B:330:GLY:N	1:B:331:PHE:HA	2.31	0.46
1:A:266:HIS:NE2	2:A:403:HOH:O	2.31	0.45
1:A:131:TYR:HA	1:B:130:ALA:HB1	1.99	0.45
1:A:92:HIS:HB2	1:B:215:VAL:HG22	1.99	0.45
1:A:99:TYR:CE1	1:B:207:PRO:HD3	2.51	0.45
1:A:12:TYR:OH	1:A:182:ARG:NH2	2.50	0.45
1:A:197:MET:O	2:A:402:HOH:O	2.21	0.44
1:B:11:LEU:HD13	1:B:174:LEU:HD21	1.98	0.44
1:A:37:ARG:HB3	1:A:37:ARG:HE	1.62	0.44
1:B:138:ARG:O	1:B:176:ARG:HD3	2.19	0.43
1:A:130:ALA:HB1	1:B:131:TYR:HA	2.00	0.43
1:B:273:SER:HA	1:B:277:PHE:HB2	1.99	0.43
1:A:86:LEU:HD21	1:A:299:GLY:HA2	2.02	0.42
1:B:236:ARG:HA	1:B:236:ARG:HD3	1.90	0.41
1:A:256:LEU:HA	1:A:256:LEU:HD12	1.91	0.41
1:A:84:ALA:O	1:A:145:THR:HA	2.21	0.41
1:B:283:GLN:HA	1:B:284:PRO:HD3	1.96	0.41
1:B:149:MET:HB3	1:B:154:PHE:CD2	2.56	0.41
1:B:38:ARG:HD3	2:B:427:HOH:O	2.20	0.40
1:B:122:MET:HG3	1:B:326:SER:HB3	2.03	0.40
1:A:273:SER:HA	1:A:277:PHE:HB2	2.02	0.40
1:B:84:ALA:O	1:B:145:THR:HA	2.21	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	339/373 (91%)	331 (98%)	6 (2%)	2 (1%)	30	50
1	B	335/373 (90%)	327 (98%)	7 (2%)	1 (0%)	46	68
All	All	674/746 (90%)	658 (98%)	13 (2%)	3 (0%)	39	61

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	212	GLN
1	A	117	VAL
1	B	117	VAL

### 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	253/278 (91%)	246 (97%)	7 (3%)	51	78
1	B	251/278 (90%)	245 (98%)	6 (2%)	57	82
All	All	504/556 (91%)	491 (97%)	13 (3%)	54	81

All (13) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	86	LEU
1	A	117	VAL
1	A	177	GLN
1	A	182	ARG
1	A	213	ARG
1	A	279	LYS
1	A	283	GLN
1	B	36	ASP
1	B	86	LEU
1	B	97	SER

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	B	211	GLU
1	B	221	LYS
1	B	286	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	212	GLN
1	B	150	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 ⓘ

There are no ligands in this entry.

## 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 [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	341/373 (91%)	-0.54	2 (0%) 90 91	22, 30, 45, 72	0
1	B	339/373 (90%)	-0.51	0 100 100	23, 30, 48, 66	0
All	All	680/746 (91%)	-0.53	2 (0%) 94 95	22, 30, 46, 72	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	212	GLN	3.3
1	A	349	ALA	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](#)

There are no ligands in this entry.

### 6.5 Other polymers [i](#)

There are no such residues in this entry.