

# **Supplementary Information**

## **Identification of (2*S*,3*S*)- $\beta$ -Methyltryptophan as the Real Biosynthetic Intermediate of Antitumor Agent Streptonigrin**

**Dekun Kong<sup>1</sup>, Yi Zou<sup>1</sup>, Zhang Zhang<sup>2</sup>, Fei Xu<sup>1</sup>, Nelson L. Brock<sup>1</sup>, Liping Zhang<sup>2</sup>, Zixin Deng<sup>1</sup> and Shuangjun Lin<sup>1\*</sup>**

<sup>1</sup>State Key Laboratory of Microbial Metabolism, Joint International Laboratory on Metabolic & Developmental Sciences, School of Life Sciences & Biotechnology, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, P. R. China

<sup>2</sup>College of Life Science, Hebei University, Baoding 071002, P. R. China

Correspondence and requests for materials should be addressed to S.-J.L. (email: linsj@sjtu.edu.cn)

## 1. Tables

### 1.1 Table S1. Strains and plasmids used in this study

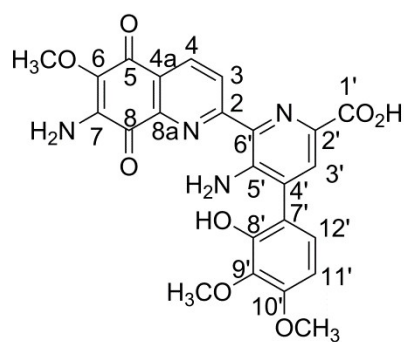
Strains/plasmids	Relevant phenotype	Source/Ref
<b>Strains</b>		
<i>E. coli</i> DH10B	F-mcrA Δ(mrr-hsdRMS-mcrBC)φ80dlacZ ΔM15 ΔlacX74 deoR recA1 endA1 ara Δ139 D (ara, leu)1697 galUgalK λ- rspLnupG	GIBCO BRL
<i>E. coli</i> ET12567/pUZ8002	<i>recF</i> , <i>dam</i> -, <i>dcm</i> -, <i>hsdS</i> , Cml <sup>R</sup> , Kan <sup>R</sup>	1
<i>E. coli</i> BW25113	Δ(araD-araB)567, ΔlacZ4787(::rrnB-4), lacIp-4000(lacIQ), λ-, rpoS369(Am), rph-1, Δ(rhaD-rhaB)568, hsdR514	2
<i>E. coli</i> BL21(DE3)pLysS	F-ompThsdSB(rB-mB-) gal dcm (DE3) pLysS (Chl <sup>R</sup> )	Novagen
<i>S. flocculus</i> CGMCC 4.1223	Wild type producer of Streptonigrin	CGMCC
LS1101( <i>ΔStnQ1</i> )	<i>StnQ1</i> inactivation mutant of CGMCC 4.1223	This study
LS1102	<i>E. coli</i> BL21(DE3)harbors the expression plasmid pLS1102- <i>StnQ1</i>	This study
LS1103	<i>E. coli</i> BL21(DE3)harbors the expression plasmid pLS1103- <i>StnK3</i>	This study
LS1104	<i>E. coli</i> BL21(DE3)harbors the expression plasmid pLS1104- <i>StnR</i>	This study
<b>Plasmids</b>		
pBSK <sup>+</sup>	<i>bla</i> , <i>lacZ</i> , pUC origin, general clone vector	Stragagene
pET28a	Kan <sup>R</sup> , pBR322 origin, T7 promoter, general protein expression vector	Novagen
pIJ790	Chl <sup>R</sup> , including λ-RED ( <i>gam</i> , <i>bet</i> , <i>exo</i> ) for PCR-targeting	3
pIJ773	Apr <sup>R</sup> , source of <i>acc(3)IV</i> and <i>oriT</i> fragment	3
pJTU1289	<i>oriT</i> region of pJTU1278 was deleted for PCR-targeting	4
pLS1100	The cosmid 9A9 was digested with <i>BamHI</i> , and then the 15.2 kb DNA fragment was ligated with pJTU1289 to generate pLS1100, which was used for Δ <i>stnQ1</i> mutant	This study
pLS1101	The <i>StnQ1</i> in pLS1100 was substituted by <i>aac(3)IV+oriT</i> cassette	This study
pLS1102- <i>stnQ1</i>	The <i>stnQ1</i> PCR product ( <i>NdeI-XhoI</i> digested) was ligated into the same sites of pET28a for <i>stnQ1</i> expression	This study
pLS1103- <i>stnK3</i>	The <i>stnK3</i> PCR product ( <i>NdeI-XhoI</i> digested) was ligated into the same sites of pET28a for <i>stnK3</i> expression	This study
pLS1104- <i>stnR</i>	The <i>stnR</i> PCR product ( <i>NdeI-XhoI</i> digested) was ligated into the same sites of pET28a for <i>stnR</i> expression	This study

## 1.2 Table S2. Primers used in this study

Primers code	Sequences of primers (5'-3')
<i>stnQ1TarF</i>	GGGGTCTTCGAGCTGTTGCGGGACGAGAAATCCCTCGAATTCCGGGGATCCGTC GACC
<i>stnQ1TarR</i>	CGTCACCTTCCTCGTTCCGCGGCCTGATCGTGGTGTTCGGTGTAGGCTGGAGCTGCT TC
<i>stnQ1VtarF</i>	CCCGGACACCGCGTTCAACGG
<i>stnQ1VtarR</i>	GGTAGGTGTGCGGGGCGCCG
<i>stnQ128aF</i>	CCTTAC <u>CATATG</u> GTGTGCGCCGCTGAACACG ( <i>NdeI</i> site underlined)
<i>stnQ128aR</i>	CCTTAC <u>CTCGAG</u> TCACGACGCCTCCAG ( <i>XhoI</i> site underlined)
<i>stnK3 28aF</i>	CCTTAC <u>CATATG</u> ATGGGCAGCAGCCATC ( <i>NdeI</i> site underlined)
<i>stnK3 28aR</i>	CCTTAC <u>CTCGAG</u> TCACAGGCGGTCCCCC ( <i>XhoI</i> site underlined)
<i>stnR 28aF</i>	CCTTAC <u>CATATG</u> ACCGCTCCGGTCCGCC ( <i>NdeI</i> site underlined)
<i>stnR 28aR</i>	CCTTAC <u>CTCGAG</u> TCAGCGGAGCCGGCGG ( <i>XhoI</i> site underlined)

**Table S3. NMR data of the compound 4**

Position	<sup>1</sup> H NMR	<sup>13</sup> C NMR
2		159.99
3	9.07 (d, J=8.5 Hz, 1H)	126.25
4	8.40 (d, J=8.5 Hz, 1H)	133.42
4a		126.87
5		175.85
6		135.77
7		141.52
8		180.21
8a		144.06
1'-CO		165.87
2'		134.08
3'	7.68 (s, 1H)	128.11
4'		133.47
5'		146.19
6'		132.02
7'		116.51
8'		148.02
9'		136.63
10'		153.27
11'	6.68 (d, J=8.5 Hz, 1H)	104.10
12'	6.93 (d, J=8.5 Hz, 1H)	124.76
6-OCH <sub>3</sub>	3.83 (s, 3H)	59.67
9'-OCH <sub>3</sub>	3.85 (s, 3H)	55.72
10'-OCH <sub>3</sub>	3.77 (s, 3H)	60.24



## 2. Figures

### 2.1 Figure S1. Alignment of StnQ1/K3/R and MarI/H/G



Figure S1. Identity (I) and similarity (S) of StnQ1/K3/R to MarI/H/G

### 2.2 Figure S2. Expression of StnQ1, StnK3, and StnR

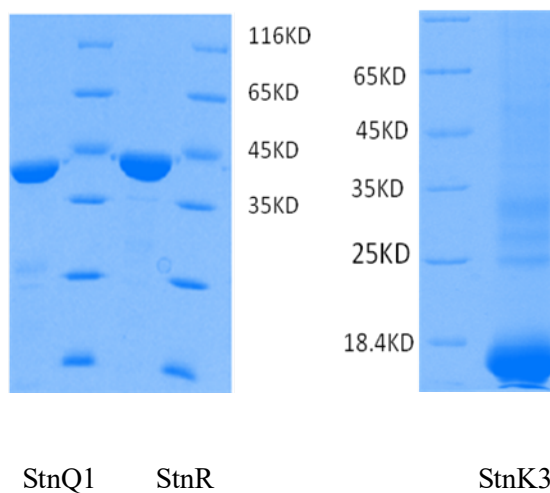
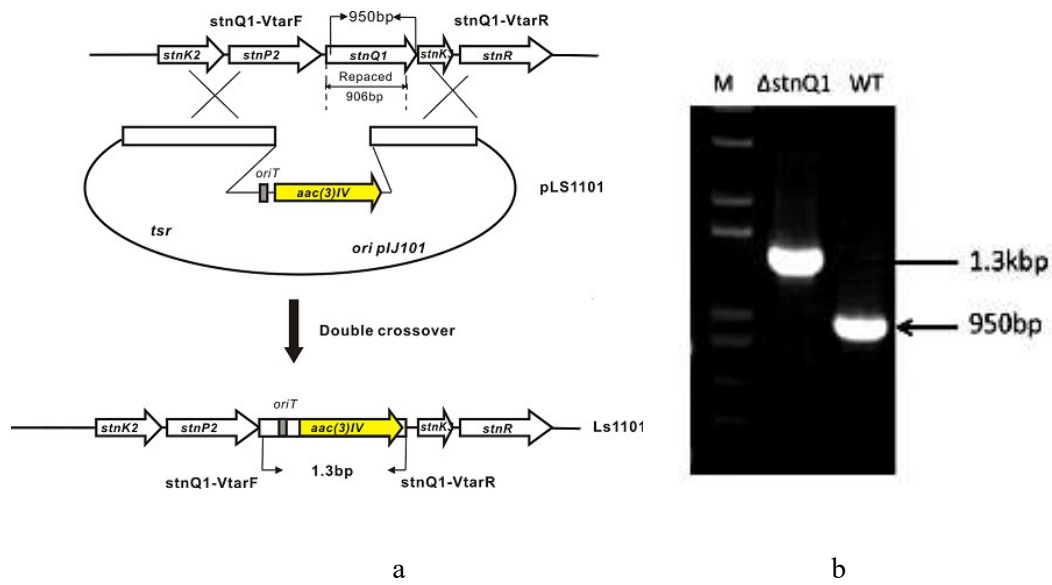


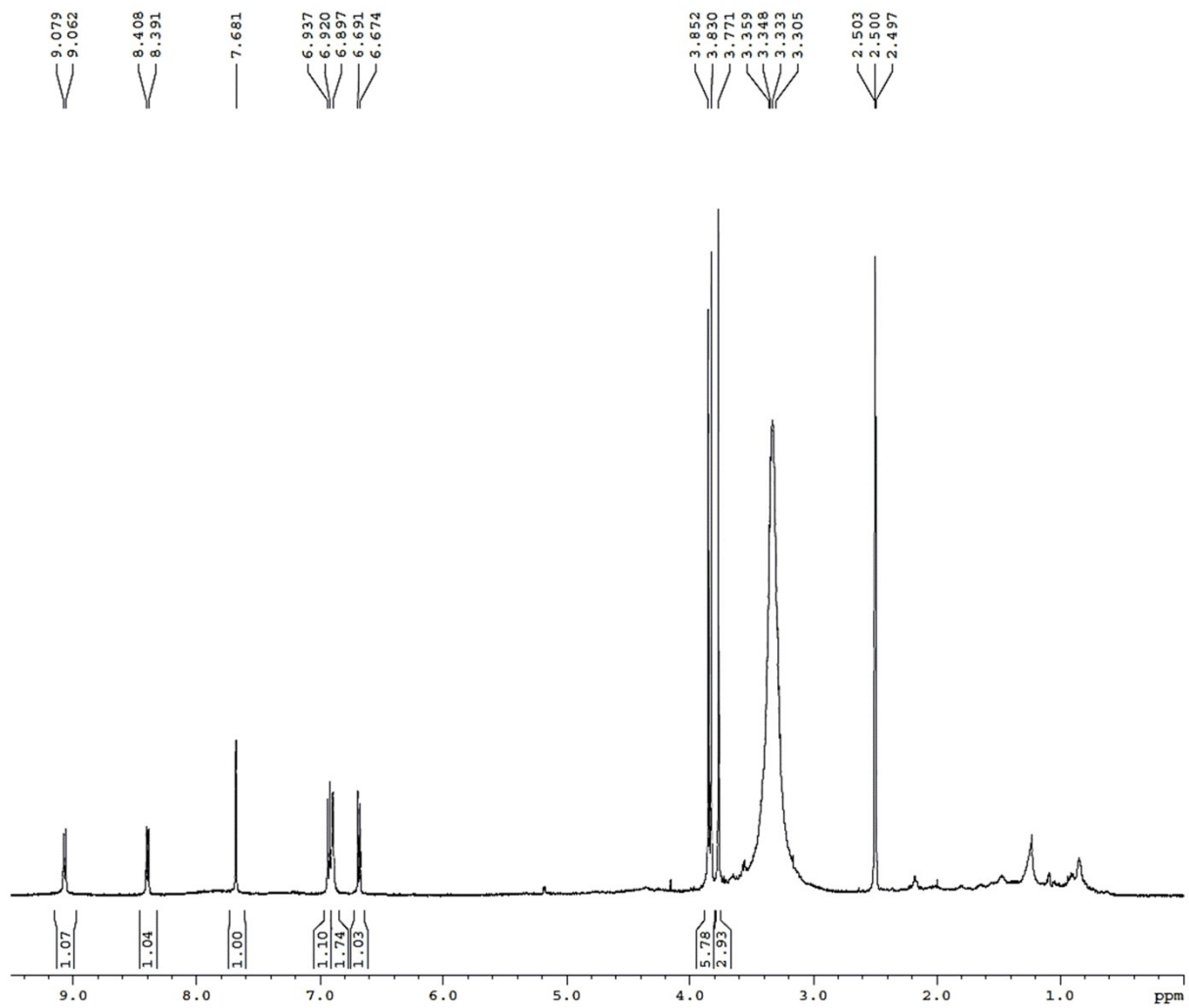
Figure S2. SDS-PAGE of the protein StnQ1, StnR, and StnK3

### 2.3 Figure S3. Construction of the $\Delta$ *stnQ1* mutant

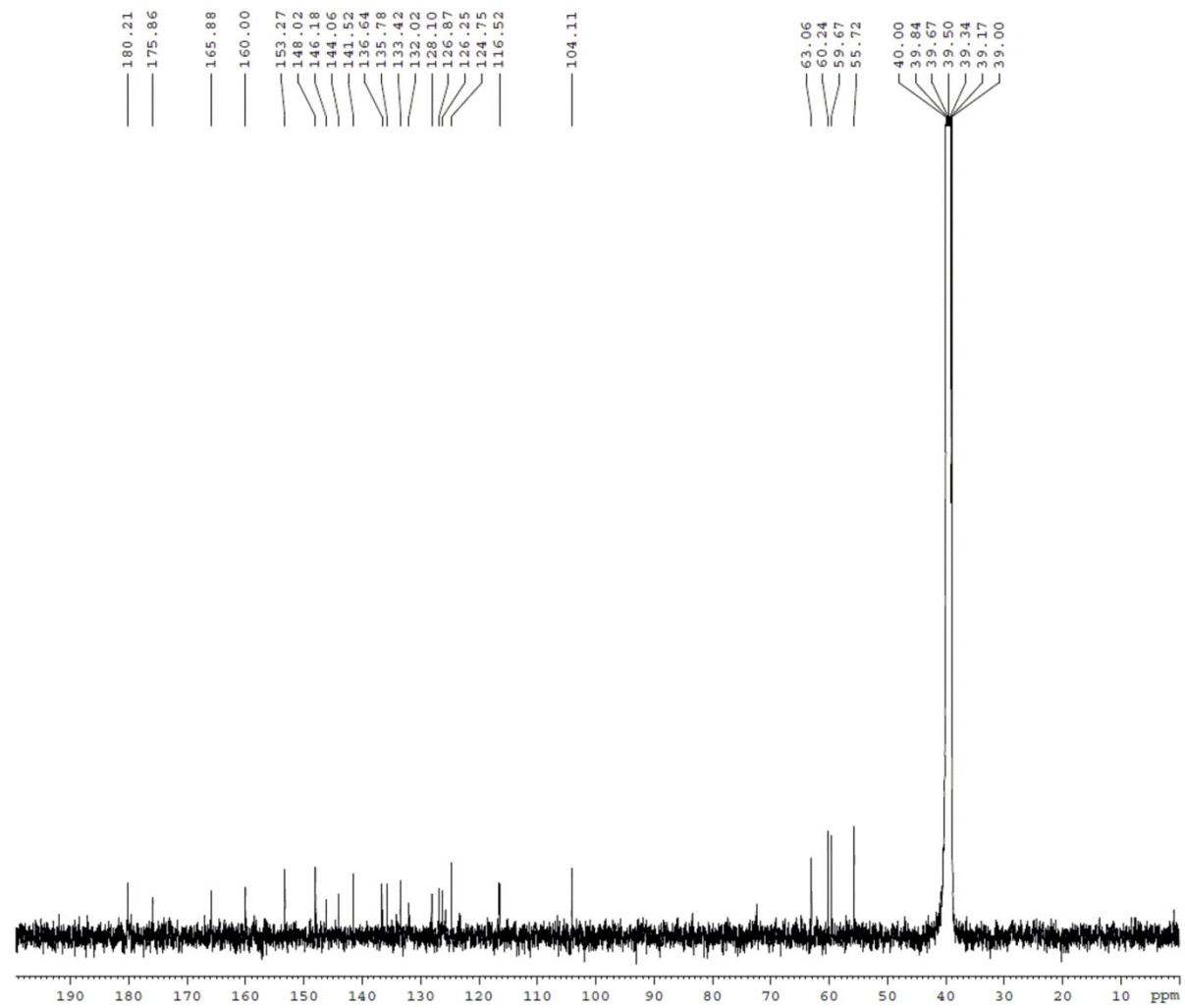


**Figure S3. Construction of the  $\Delta$ *stnQ1* mutant.** **a.** The scheme of *stnQ1* gene replacement with apramycin-resistant/*oriT* cassette (*aac(3)IV-oriT*); **b.** Confirmation of the inactivation of *stnQ1* by insertion of apramycin-resistant/*oriT* cassette (*aac(3)IV-oriT*) by PCR using the primers listed in Table S2

2.4 Figure S4. <sup>1</sup>H NMR spectrum of 4 (500 MHz, DMSO-d<sub>6</sub>)

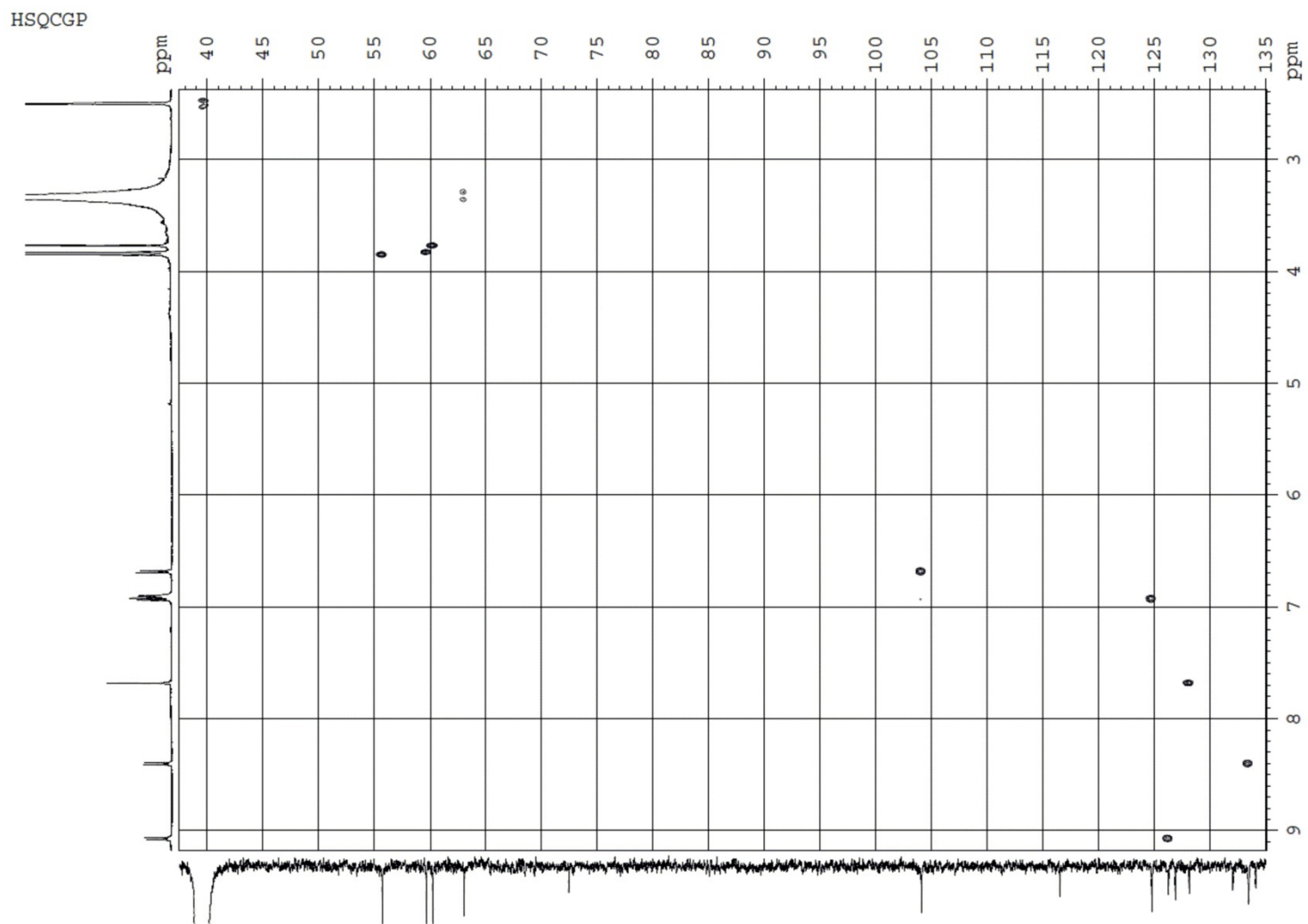


2.5 Figure S5.  $^{13}\text{C}$  NMR spectrum of 4 (125 MHz, DMSO- $d_6$ )





2.6 Figure S6. HMQC NMR spectrum of 4 (DMSO-d<sub>6</sub>)



2.7 Figure S7. HMBC NMR spectrum of 4 (DMSO-d<sub>6</sub>)

