1 **Supporting Information**

2

3 **Moving Posttranslational Modifications Forward to Biosynthesize the Glycosylated Thiopeptide**

4 **Nocathiacin I in *Nocardia sp.* ATCC202099**

5

6 Ying Ding, Yi Yu, Haixue Pan, Heng Guo, Yeming Li, and Wen Liu*

7

8 State Key Laboratory of Bioorganic and Natural Products Chemistry, Shanghai Institute of Organic

9 Chemistry, Chinese Academy of Sciences, 345 Lingling Rd., Shanghai 200032, China

10

11 * To whom correspondence should be addressed: Shanghai Institute of Organic Chemistry, Chinese

12 Academy of Sciences, 345 Lingling Rd., Shanghai 200032, China. Wen Liu, Email:

13 wliu@mail.sioc.ac.cn, Tel: 86-21-54925111, Fax: 86-21-64166128

14

15 **Sequence Accession Number:** GU564398

16

17

18

19

20

21

1 **Materials and Procedures**

2

3 **Bacterial Strains, Plasmids, and Reagents**

4 Bacterial strains and plasmids used in this study are summarized in Table S1. Chemicals, biochemicals
5 and media were purchased from Sinopharm Chemical Reagent Co., Ltd. (Shanghai, China) and Oxoid
6 Ltd. (Basingstoke, United Kingdom) unless otherwise stated. Restriction enzymes were purchased
7 from TaKaRa Biotechnology (Dalian) Co., Ltd. (Dalian, China).

8

9 **DNA Isolation and Manipulation**

10 General DNA isolation and manipulation in *Escherichia coli*, *Nocardia* and *streptomyces* were
11 performed according to standard methods.[1,2] PCR amplifications were carried out on an authorized
12 thermal cycler Eppendorf AG 22331 (Hamburg, Germany) using either *LA Taq* DNA polymerase
13 (TaKaRa) or PrimeSTAR HS DNA polymerase (Dalian TaKaRa Biotechnology Co. Ltd., Dalian,
14 China). Primers (summarized in Table S2) synthesis and DNA sequencing were performed at
15 Shanghai Invitrogen Biotech Co., Ltd. (Shanghai, China) and Chinese National Human Genome
16 Center (Shanghai, China). The genomic library of *Nocardia sp.* ATCC202099 was constructed in
17 pOJ446 according to the method described previously.[3] This library was screened by colony
18 hybridization with the 0.3 kb, partial sugar *N*-dimethyltransferase-encoding PCR product, which was
19 amplified by using the degenerate primers (5'-GCT GAC GTC GCC TGC GGS ACS GGN DNN
20 CA-3' and 5'-CGC GAA CGT STC SGG RAA CCA CCA NGG NTC-3'), as a probe for identification
21 of the nocathiacin I (NOC-I) biosynthetic gene cluster. The resulting positive clones were further

1 confirmed by Southern hybridization.

2

3 **Sequence Analysis**

4 ORFs in the *noc* cluster were deduced from the sequenced DNA region with the assistance of
5 FramePlot 4.0beta program (http://nocardia.nih.go.jp/fp4/). The corresponding deduced proteins were
6 compared with other known proteins in the databases by using available BLAST methods
7 (http://www.ncbi.nlm.nih.gov/ blast/). Amino acid sequence alignments were performed with
8 CLUSTALW from BIOLOGYWORKBENCH 3.2 software (http://workbench.sdsc.edu).

9

10 **Inactivation of *nosD*, *nosE*, and *nosI* in the Nosiheptide (NOS)-Producing Strain *S. actuosus***
11 **ATCC25421**

12 To inactivate *nosD*, a 1.9 kb fragment amplified by PCR using the primers D-inf-Ls/D-inf-La and a
13 2.0-kb fragment amplified by using the primers D-inf-Rs/D-inf-Ra were initially cloned into pMD19-T,
14 giving pSL4050 and pSL4051, respectively. After sequencing to confirm the fidelity, the 1.9 kb
15 *Hin*dIII/*Eco*RI and 2.0 kb *Eco*RI/*Xba*I fragments were recovered and co-ligated into the *Hin*dIII/*Xba*I
16 site of pKC1139, yielding the recombinant plasmid pSL4052, in which a 843 bp in-frame coding
17 region (corresponding to AA_{42}-AA_{322} of the deduced product NosD) of *nosD* was deleted. The same
18 strategy was applied to inactivation of *nosE* and *nosI*, and the transient plasmids for each construction
19 were summarized in Table S1. For *nosE* inactivation, the recombinant plasmid pSL4055 contains a
20 mutant *nosE*, in which a 2,451 bp coding region (corresponding to AA_{66}-AA_{882} of the deduced product
21 NosE) was deleted in-frame. For *nosI* inactivation, the recombinant plasmid pSL4058 contains a

1 mutant *nosI*, in which a 1,152 bp coding region (corresponding to AA_{19}-AA_{402} of the deduced product

2 NosL) was deleted in-frame.

3 Introductions of pSL4052, pSL4055 and pSL4058 into *S. actuosus* for inactivation of *nosD*, *nosE* and

4 *nosI*, respectively, were carried out by following the previously described procedure,[5] yielding the

5 corresponding mutant strains SL4050 (△*nosD*), SL4052 (△*nosE*), and SL4054 (△*nosI*).

6

7 ***In trans* Expression of *nocD*, *nocE*, *nocI* or *nocL* from the *noc* Cluster in each corresponding *S.***

8 ***actuosus* Mutant Strain for Heterologous Complementation to NOS Production**

9 To make the *nocD* expression construct, a 1.1 kb *nocD*-containing PCR product that was amplified

10 from the cosmid pSL5001 by PCR using the primers D-hetero-F/D-hetero-R was cloned into pSP72,

11 yielding pSL4059. The 1.1 kb *Hin*dIII/*Xba*I fragment recovered from pSL4059 was co-ligated with a

12 0.45 kb *Eco*RI/*Hin*dIII fragment, harboring the promoter *PermE**, into the *Eco*RI/*Xba*I site of

13 pSET152, yielding the recombinant plasmid pSL4060, in which the 1.1 kb fragment containing *nocD*

14 is under the control of the *PermE** promoter. The same strategy was applied for expressing *nocE*, *nocI*

15 and *nocL*, and the transient plasmids for each construction were summarized in Table S1. For *nocE*

16 expression, the recombinant plasmid pSL4062 carries a 2.7 kb fragment that contains *nocE* under the

17 control of *PermE**. For *nocI* expression, the recombinant plasmid pSL4064 carries a 1.4 kb fragment

18 that contains *nocI* under the control of *PermE**. For *nocL* expression, the recombinant plasmid

19 pSL4066 carries a 1.4 kb fragment that contains *nocL* under the control of *PermE**.

20 Introduction of pSL4060, pSL4062, pSL4064 or pSL4066 into each *S. actuosus* mutant strain, SL4050

21 (△*nosD*), SL4052 (△*nosE*), SL4054 (△*nosI*) or previously made SL4005 (△*nosL*) (5), for heterologous

1 complementation were carried out by following the described procedure,[5] yielding the corresponding
2 recombinant strains SL4051 (SL4050 derivative for expressing *nocD*), SL4053 (SL4052 derivative for
3 expressing *nocE*), SL4055 (SL4054 derivative for expressing *nocI*), and SL4056 (SL4005 derivative
4 for expressing *nocL*).
5
6 **Analysis of NOS Production in *S. actuosus* Strains**
7 Fermentation and isolation of NOS from *S. actuosus* Strains were performed according to the methods
8 described previously.[5] HPLC and HPLC-MS analysis of NOS production were carried out on an
9 Agilent ZORBAX SB-C18 column (4.6×250 mm, part number 880975-902, S/N USCL024998) under
10 the same conditions,[5] with an only exception for ultraviolet detection at 330 nm to give a clearer
11 background.
12
13
14
15
16
17
18
19
20
21

1 **Supplementary Tables**

2

3 **Table S1. Bacterial strains and plasmids used in this study**

4

Strain/Plasmid	**Characteristic(s)**	**Source/ Reference**
E. coli		
DH5α	Host for general cloning	Invitrogen
LE392	Host for genomic library construction	Promega
S17-1	Donor strain for conjugation between *E.coli* and *Streptomyces*	2
Nocardia		
ATCC202099	Wild type strain, NOC-I producing	ATCC
Streptomyces		
S. actuosus ATCC 25421	Wild type strain, NOS producing	ATCC
SL4050	*ΔnosD* mutant, NOS non-producing	This study
SL4051	SL4050 derivative that contains pSL4060 for expressing *nocD in trans*, NOS producing	This study
SL4052	*ΔnosE* mutant, NOS non-producing	This study

SL4053	SL4052 derivative that contains pSL4062 for expressing *nocE in trans*, NOS producing	This study
SL4054	*ΔnosI* mutant, NOS non-producing	This study
SL4055	SL4054 derivative that contains pSL4064 for expressing *nocI in trans*, NOS producing	This study
SL4005	*ΔnosL* mutant, NOS non-producing	5
SL4056	SL4005 derivative that contains pSL4066 for expressing *nocL in trans*, NOS producing	This study
Plasmids		
pMD19-T	*E. coli* subcloning vector	TaKaRa
pSP72	*E. coli* subcloning vector	Promega
pOJ446	*E.coli-Streptomyces* shuttle vector for genomic library construction	NRRL 14791
pKC1139	*E.coli-Streptomyces* shuttle vector for gene inactivation, temperature sensitive replication in *Streptomyces*	1
pSET152	*E.coli-Streptomyces* shuttle vector for heterologous complementation, site-specific integration in *Streptomyces*	1
pSL4001	pOJ446-based, *S. actuosus* ATCC 25421genomic library cosmid	5

pSL5001	pOJ446-based, *Nocardia sp.*ATCC202099 genomic library cosmid	This study
pSL4050	pMD19-T derivative containing a 1.9 kb PCR product from pSL4001	This study
pSL4051	pMD19-T derivative containing a 2.0 kb PCR product from pSL4001	This study
pSL4052	pKC1139 derivative containing a 3.9 kb fragment, construct for *nosD* in-frame inactivation	This study
pSL4053	pMD19-T derivative containing a 1.9 kb PCR product from pSL4001	This study
pSL4054	pMD19-T derivative containing a 1.9 kb PCR product from pSL4001	This study
pSL4055	pKC1139 derivative containing a 3.8 kb fragment, construct for *nosE* in-frame inactivation	This study
pSL4056	pMD19-T derivative containing a 1.9 kb PCR product from pSL4001	This study
pSL4057	pMD19-T derivative containing a 1.9 kb PCR product from pSL4001	This study
pSL4058	pKC1139 derivative containing a 3.8 kb fragment, construct for *nosI* in-frame inactivation	This study

pSL4059	pSP72 derivative containing a 1.1 kb PCR product from pSL5001	This study
pSL4060	pSET152 derivative containing a 1.1 kb fragment, construct for expressing *nocD* in SL4050	This study
pSL4061	pSP72 derivative containing a 2.7 kb PCR product from pSL5001	This study
pSL4062	pSET152 derivative containing a 2.7 kb fragment, construct for expressing *nocE* in SL4052	This study
pSL4063	pSP72 derivative containing a 1.4 kb PCR product from pSL5001	This study
pSL4064	pSET152 derivative containing a 1.4 kb fragment, construct for expressing *nocI* in SL4054	This study
pSL4065	pSP72 derivative containing a 1.4 kb PCR product from pSL5001	This study
pSL4066	pSET152 derivative containing a 1.4 kb fragment, construct for expressing *nocL* in SL4005	This study

1

2

3

4

5

Table S2. Primers used in this study.

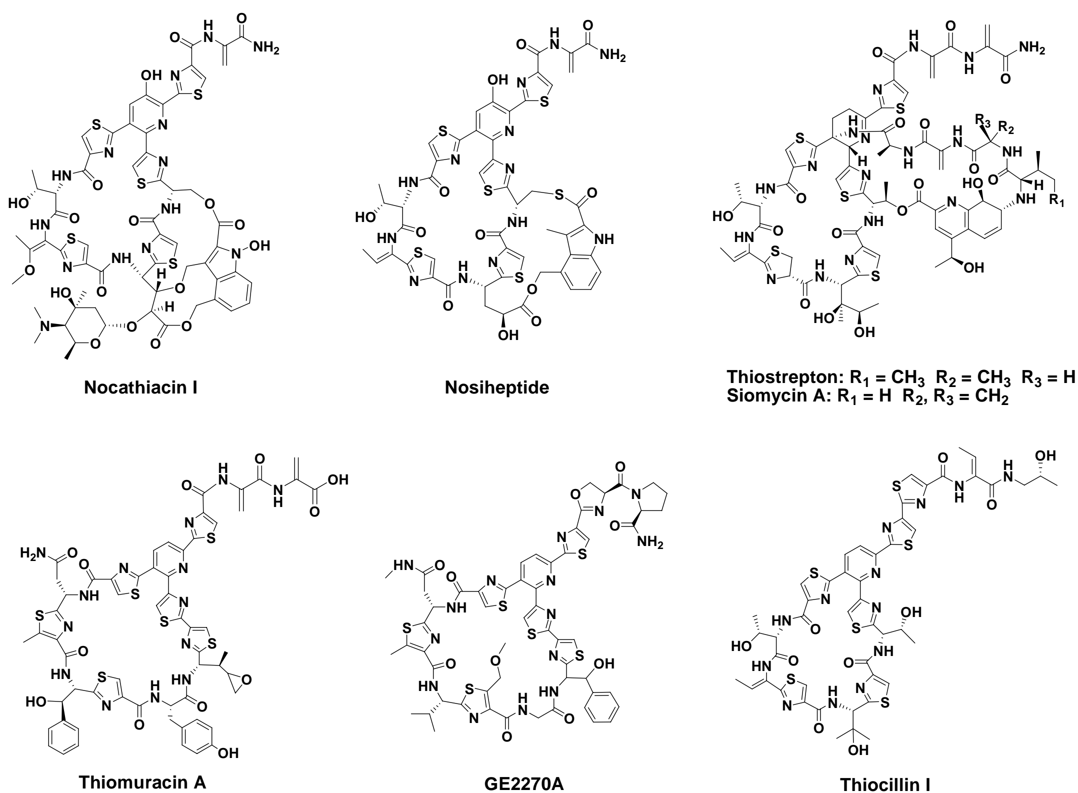
nosD	D-inf-Ls: 5’-GCGC*AAGCTT*CGCGAGAGGCTGGCCGC-3’
	D-inf-La: 5’-GGA*GAATTC*CAGCGCGCCGGAGGCGTGC-3’
	D-inf-Rs: 5’-CCG*GAATTC*GGCCTGCTCGACCG-3’
	D-inf-Ra: 5’-AGC*TCTAGA*GGCAGGTCGAACAGCTCGC-3’
nosE	E-inf-Ls: 5’-CGCA*AAGCTT*CCCGTAGGACATGCCG-3’
	E-inf-La: 5’-CGC*GAATTC*CAGGCGCCCGCGCAGC-3’
	E-inf-Rs: 5’-GTC*GAATTC*CAGGAGTGCCTGCCGC-3’
	E-inf-Ra: 5’-GGG*TCTAGA*AGGCGTTGCCGAGCTGG-3’
nosI	I-inf-Ls: 5’-TGCC*AAGCTT*GGTGTCGGTGTCGGTGAG-3’
	I-inf-La: 5’-CGG*GAATTC*TGCGCCTCCGAGCAGGAG-3’
	I-inf-Rs: 5’-GCG*GAATTC*GGTGAGCAGGTCGTCCCAG-3’
	I-inf-Ra: 5’-GGAG*TCTAGA*TACCACCACGAGGGCGTC-3’
nosL	L-inf-Ls: 5’-C*AAGCTT*ACGCCCCGCAGGAAGAACC-3’
	L-inf-La: 5’-*TCTAGA*ATTGCCTGGGAGTTCTGCG-3’
	L-inf-Rs: 5’-*TCTAGA*CGTCGACCACTTCGTGAACC-3’
	L-inf-Ra: 5’-*GAATTC*GTGGTCGAGGGCGATGG-3’
nocD	D-hete-F: 5’-GT*AAGCTT*CCGGAGCTGAGGGCTACGCCTCCGG-3’
	D-hete-R: 5’-GG*TCTAGA*GGTCATGGCGTGCCGATCAGCTGCG-3’

nocE	E-hete-F: 5’-GT*AAGCTT*GCAGGTGTGCGTGGGCGCCGTC-3’
	E-hete-R: 5’-GG*TCTAGA*GGTCACCGGGCATCGCAACCTCCGGTA-3’
nocI	I-hete-F: 5’-GG*AAGCTT*GCTGTCCTACCTGACCCGCAAGG-3’
	I-hete-R: 5’-GG*TCTAGA*GGTCATGACAGCTTCCCGTTCGGGC-3’
nocL	L-hete-F: 5’-GG*AAGCTT*TGCAGGAACTTCCCCATGACACGGC-3’
	L-hete-R: 5’-GC*TCTAGA*CCTCAGCCGATCGGGATGACGGCC-3’

* Restriction sites for *Hin*dIII (AAGCTT), *Xba*I (TCTAGA) and *Eco*RI (GAATTC) are double-underlined

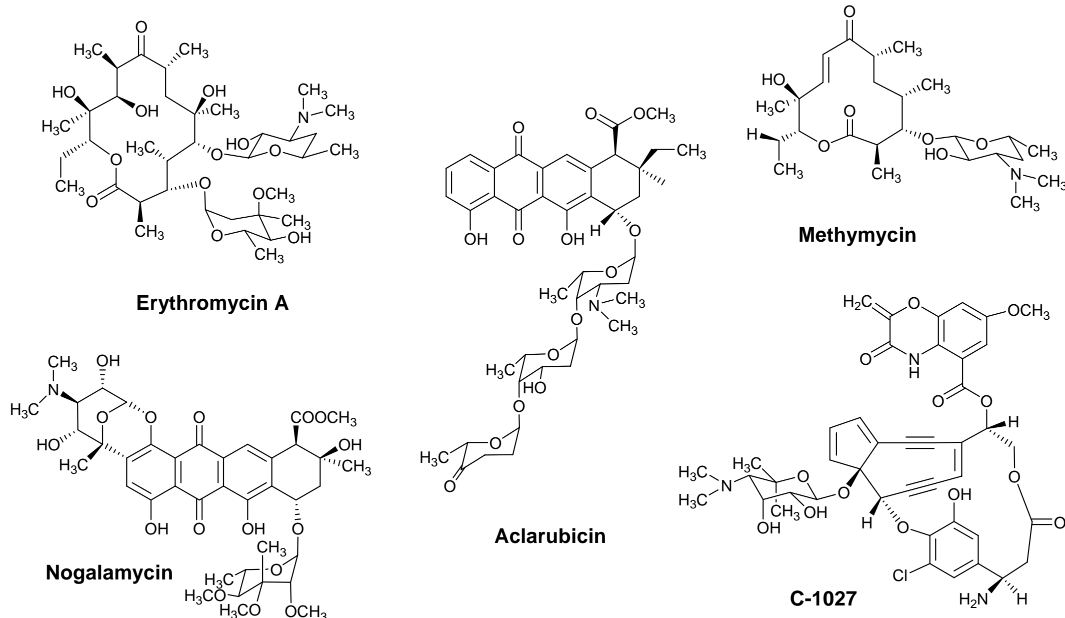
Supplementary Figures.

Fig. S1. Structures of thiopeptides whose biosynthetic gene clusters have been cloned.



1 **Fig. S2.** Structures and sequence comparison. A, glycosylated antibiotics erythromycin, C-1027,
2 aclarubicin, noglamycin, and methymycin which have the *N*-dimethylated deoxyaminosugar moiety. B,
3 amino acid sequence alignment of various *N*-dimethyltransferases, including EryCVI from
4 *Saccharopolyspora erythraea* NRRL 2338 (AAB84073), DesVI from *S. venezuelae* (3BXO_A),
5 SnogA from *S. nogalater* (AAF01819), AknX2 from *S. galilaeus* (AAF73460), SgcA5 from *S.*
6 *globisporus* (AAL06660), and NocS2 from *Nocardia sp.* ATCC202099 in this study. Protein
7 accession numbers are given in parentheses. The conserved amino acid motifs for PCR primer design
8 to clone the NOC-I biosynthetic gene cluster are boxed.[6-10]
9

A



B

```
                                             Forward
             10        20        30        40        50        60
1   MEYGPEHAKFYDLVFRSRGKDFDLEARGLTELILAARPDAVSLLDVACGTGAHLETLATL  SgcA5
1    MYEVDHADVYDLFYLGRGKDYAAEASDIADLVRSRTPEASSLLDVACGTGTHLEHFTKE  DesVI
1    MYGRELADVYEMVYRSRGKSWADEAERVTAEIRSRRPGARSLLDVACGTGAHLEAFRGL  SnogA
1    MYEGGFAELYDRFYRGRGKDYAAEAAQVARLVRDRLPSASSLLDVACGTGTHLRRFADL  EryCVI
1    MYGRELADVYEAIYRSRGKDWGQEAADVSRIITERRPGAGSLLDVACGTGAHLSVFSTL  AknX2
1   MTYGHAHAEHYDLVFRSRGKDWPAESARIAALVRDRAPAAATLLDVGCGTGAHLVTFAKT  NocS2
    -mYg-e-AdvYdlvyrsRGKdwa-Ea--v--lvr-r-P-A-sLLDVaCGTGaHL--f--l  consensus

             70        80        90        100       110       120
61  FGHVEGLEYAPAMLEQAAGRLPGVPLHAGDMRSFDLGRTFDAITCMGNALGEMGSVTELK  SgcA5
60  FGDTAGLELSEDMLTHARKRLPDATLHQGDMRDFRLGRKFSAVVSMFSSVGYLKTTEELG  DesVI
60  FAHTEGLELSDEMRALAERRLPGVPVRPGDMRDFALSGRFDAVVCLFCSIGYLETVADMR  SnogA
60  FDDVTGLELSAAMIEVARPQLGGIPVLQGDMRDFALDREFDAVTCMFSSIGHMRDGAELD  EryCVI
60  FEVAEGLEIAEPMRRLAEQRLPGTTVHAGDMRDFRLPRTYDAVSCMFCAIGYLETLDDMR  AknX2
61  FAAVAGVEPAEAMREIAADRIGAGQVHPGDMRAFDLGRTFDAVTCLGFAIAYMPDTAALD  NocS2
    Fa---GlEl-e-M--lA--rlpg--vh-GDMRdF-L-r-fdAvtcmf--igym-t--el-  consensus

             130       140       150       160       170       180
121 AAVSAMAHHLNPGGVLVAEPWYFPENFLDGHVGGHLHQEEGRVITRMTHSVRQGDKSRLE  SgcA5
120 AAVASFAEHLEPGGVVVVEPWWFPETFADGWVSADVVRRDGRTVARVSHSVREGNATRME  DesVI
120 AAVRTMAAHLVPGGVLVVEPWWFPERFLEGYVAGDLARGEGRTVARVSHSTRQGRRTRME  SnogA
120 QALASFARHLAPGGVVVVEPWWFPEDFLDGYVAGDVVRDGDLTISRVSHSVRAGGATRME  EryCVI
120 AAVRSMAAHLEPGGVLVVEPWWFPENFIEGYVAGDLAREEHRTIARISHTTRKGRATRME  AknX2
121 AAVARMAAHLSPGGVLVVEPWWGPEEFIDGYVGGHLVREDGLVISRVTHSVRRDGATHMR  NocS2
    aAv-smA-HL-PGGVlVvEPWwfPE-FldGyVagdl-reegrti-RvsHsvR-g-atrme  consensus
                      Reverse
```

1

2

3

4

This journal is (c) The Royal Society of Chemistry, 2010

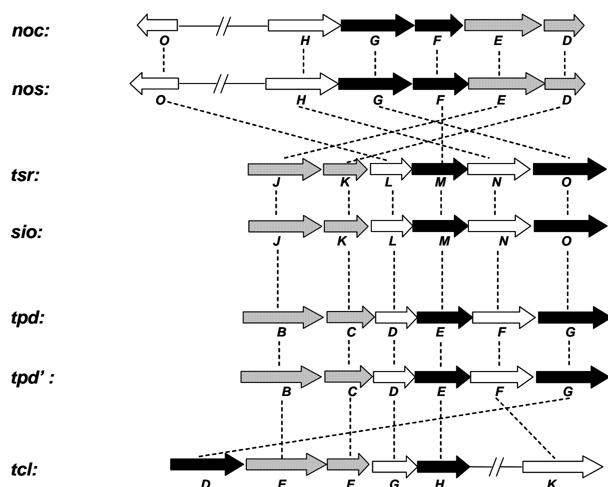
1 **Fig S3.** Features for the thiopeptide framework formation. A, peptide precursors for NOC-I (NocM),
2 NOS (NosM), thiostrepton (TsrH or TsrA), siomycin (SioH), thiocillin ($TclB_{1-4}$ or TclE-H), GE2270A
3 (TpdA), and thiomuracin (TpdA'). B, organization of the thiopeptide framework-forming genes
4 identified from the producers of NOC-I (*noc*), NOS (*nos*), Thiostrepton (*tsr*), Siomycin (*sio*),
5 Thiocillin (*tcl*), GE2270A (*tpd*), and Thiomuracin (*tpd*'). Their deduced functions are labeled in
6 pattern, and homologies in sequence are indicated by dashed lines.

7

A

	Leader Peptide	Structural Peptide
NocM	MSADLSALNIDSLEISEFLDDSRLEDSEVVAKVMSA	**SCTTCECSCSCSS**
NosM	VDAAHLSDLDIDALEISEFLDESRLEDSEVVAKVMSA	**SCTTCECCCSCSS**
$TclB_{1-4}$/TclE-H	MSEIKKALNTLEIEDFDAIEMVDVDAMPENEALEIMGA	**SCTTCVCTCSCCTT**
TsrH/TsrA	MSNAALEIGVEGLTGLDVDTLEISDYMDETLLDGEDLTVTM	**IASASCTTCICTCSCSS**
SioH	MSTAAIVGQEIGVDGLTGLDVDALEISDYMDETLLDGEDLSVTM	**VSSASCTTCICTCSCSS**
TpdA	MSELESKLNLSDLPMDVFEMADSGMEVESLTAGHGMPEVGA	**SCNCVCGFCCSCSPSA**
TpdA'	MDLSDLPMDVFELADDGVAVESLTAGHGMTEVGA	**SCNCFCYICCSCSSA**

B



8 ■ polyazoles ▨ Dehydrated amino acids □ 6-membered nitrogen heterocycle

9

1 **References**

2

3 1. M. Bierman, R. Logan, K. O'Brien, E. T. Seno, R. N. Rao and B. E. Schoner, *Gene,*1992,
4 **116**,43-49.
5 2. T. Kieser, M. J. Bibb, M. J. Buttner, K. F. Chater and D. A. Hopwood, *Practical Streptomyces*
6 *Genetics*, The John Innes Foundation, Norwich, UK, 2000.
7 3. W. Liu and B. Shen, *Antimicrob. Agents Chemother.*, 2000, **44**, 382–392.
8 4. J. Sambrook and D. W. Russell, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor
9 Laboratory Press, New York, 3rd ed, 2001.
10 5. Y. Yu, L. Duan, Q. Zhang, R. Liao, Y. Ding, H. Pan, E. W. Pienkowski, G. Tang, B. Shen and W.
11 Liu. *ACS. Chem. Biol.*, 2009, **4**, 855-864
12 6. H. Chen, H. Yamase, K. Murakami, C. W. Chang, L. Zhao, Z. Zhao and H. W. Liu, *Biochemistry,*
13 2002, **41**, 9165–9183.
14 7. W. Liu, S. D. Christenson, S. Standage and B. Shen. *Science*, 2002, **297**, 1170-1173.
15 8. K. Raty, T. Kunnari , J. Hakala, P. Mantsala and K. Ylihonko. *Mol. Gen. Genet.*, 2000, **264**,
16 164-172.
17 9. R. G. Summers, S. Donadio, M. J. Staver, E. Wendt-Pienkowski, C. R. Hutchinson and L. Katz.
18 *Microbiology*, 1997, **143**, 3251-3262.
19 10. S. Torkkell, K. Ylihonko, J. Hakala, M. Skurnik and P. Mantsala. *Mol. Gen. Genet.*, 1997, **256**,
20 203-209.