Simple Reverse Genetics Approach to Elucidating the Biosynthetic Pathway of Complex Thiopeptide Nocathiacin

Maochen Wei,[a] Jing Deng,[a] Shuzhen Wang,[a] Nan Liu,[a] and Yijun Chen[a],[b]*

[a]Dr. M. C. Wei, Dr. J. Deng, K. Feng, Dr. S. Z. Wang, Dr. N. Liu, Prof. Dr. Y. J. Chen
Laboratory of Chemical Biology, China Pharmaceutical University, 24 Tongjia Street, Nanjing, Jiangsu Province, 210009, People's Republic of China
Tel: 86-25-83271045, Fax: 86-25-83271249
Email: yjchen@cpu.edu.cn

[b]Prof. Dr. Y. J. Chen
Department of Chemical Biology, Ernest Mario School of Pharmacy, Rutgers University, Piscataway, New Jersey, USA

Correspondence: Prof. Dr. Yijun Chen

There is an urgent need for the discovery of new antibiotics to fight against multi-drug resistant bacteria.[1-2] Thiopeptides (Figure 1), a large group of highly modified macrocyclic peptides, have attracted intensive attention due to the intriguing molecular architecture and potent activity against various Gram-positive bacteria including methicillin-resistant *Staphylococcus aureus*, vancomycin-resistant *Enterococci* and fully penicillin-resistant *Streptococcus pneumoniae*.[3-5] According to the central heterocyclic feature and the oxidation state, thiopeptides have been classified into five different series, in terms of *a* to *e*.[3] Although almost 100 members of thiopeptides with similar structures have been reported,[6,7] only handful biosynthetic pathways of them have been elucidated by screening genomic library or mining microbial genome databases.[6-15]

Nocathiacin (Noc), a tricyclic thiopeptide in *e* series, exhibited excellent potency against clinically important multi-drug resistant pathogens *in vitro* and *in vivo*.[16] Therefore, it was considered as the most promising molecule, but the poor aqueous solubility prevented its development as a clinical agent. To generate analogues with increased aqueous solubility, semi-synthetic and biotransformation approaches have been utilized to produce new derivatives.[17-20] However, they did not turn out satisfactory results. Consequently, manipulation of the biosynthetic machinery appears to be the only option for generating diverse analogues with improved physiochemical properties, but elucidation of the biosynthetic pathway of such thiopeptide by genomic library screening typically involves tedious and time-consuming steps.

In order to rapidly and efficiently exploit the biosynthetic gene cluster of nocathiacin, we systematically analyzed the reported biosynthetic genes of thiopeptides. Surprisingly, for thiostrepton and siomycin in *a* series, the array and orientation of the enzymes involved in their biosynthesis are identical, suggesting a close relationship between structural peptide and biosynthetic gene cluster. To examine this hypothesis, we next compared the biosynthetic gene clusters of GE2270 and thiomuracin in *d* series, consisting of same macrocyclic core with obviously divergent modifications**.** Similar results were also obtained, in which the enzymes for

the macrocyclic core structure have identical array and orientation. In contrast, cross-series comparison of the biosynthetic genes of thiocillin I in *d* series and nosiheptide (Nos) in *e* series showed completely different array and orientation even though their macrocyclic core structures are nearly identical. Therefore, we postulated that thiopeptides in the same series perhaps have a direct correlation between structural feature and biosynthetic gene cluster.

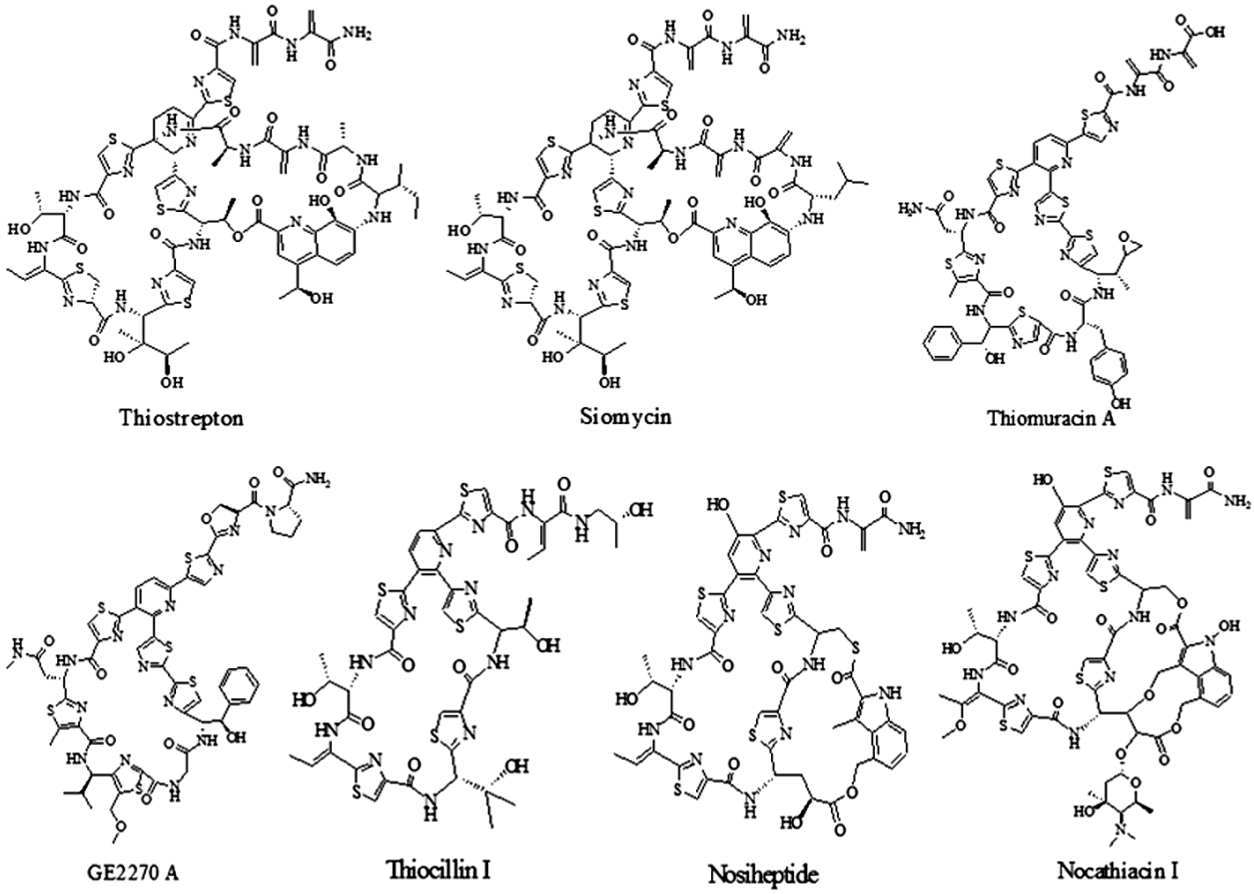

Figure 1. Chemical structures of representative thiopeptides. *a* series: thiostrepton and siomycin; *d* series: thiomuracin A, GE2270A and thiocillin I; *e* series: nosiheptide and nocathiacin I.

In view of the chemical structure, nocathiacin is similar to nosiheptide in *e* series. In addition to the unique tricyclic characteristics, the major difference of nocathiacin from nosiheptide is its glycosylation, which may occur at the last step.[16] It is reasonable to speculate that the arrangement and orientation of the enzymes for the

biosynthesis of nocathiaicin aglycone should be pertinent to nosiheptide, which may offer an opportunity of directly applying reverse genetics to obtain the biosynthetic genes of nocathiacin. Like nosiheptide, [8] nocathiacin possesses an indolic acid at the same position, suggesting that the enzymes for the biosynthesis of indolic acid should be near the precursor peptide. Therefore, amplification of the structural peptide region in combination with the conserved motifs of the enzymes for indolic acid biosynthesis may result in a specific gene from genomic DNA of the producing strain.

To test our hypothesis, a PCR reaction was employed to amplify the biosynthetic genes of nocathiacin from *Amycolatopsis fastidiosa* ATCC 202099. First, primers were designed based on the structural peptide sequence of nocathiacin.[21] Then, comparison of two enzymes flanking the both sides of the precursor peptide for nosiheptide generated specific sequences, in which the conserved motifs containing more than 6 amino acid residues (at least 3 identical amino acid residues at the N-terminus) were randomly selected to pair with the above primers (Table S1 and Figure S1 and S2). As anticipated, two specific DNA fragments were amplified respectively. A gene encoding a precursor peptide, in term of NocA, consisting of a leader peptide (36 amino acids) and a structural peptide (13 amino acids) was obtained (Figure 2 A). The structural peptide of nocathiacin consisted of same sequence as isotope feeding results except for an additional serine located at the C-terminus (Figure 2 B). This serine could be removed during the maturation to form final product in the same fashion as nosiheptide.[8] Although the leader peptide sequence of nocathiacin showed 80% identity to nosiheptide (Figure 2 A), their difference may play a role in mediating different modifications of the macrocyclic core structures. Additionally, the rest of the genes showed high homology to the tailoring enzymes located at both sides of nosiheptide precursor peptide.

To further verify our approach, we continued to amplify the outer sequences from NocA. After genome walking with respective primers by our newly developed method [22] (Table S2 and Figure S3), a complete sequence encoding three tailoring enzymes and a precursor peptide was rapidly obtained and identified (GenBank no. HM236314, submitted on May 16, 2010). As we expected, the arrangement and orientation of

these genes were also same as nosiheptide (Figure 2 C). Bioinformatics analysis of these genes indicated high homology to the tailoring enzymes within the biosynthetic pathway of nosiheptide as well (Table 1). Sequence alignments revealed that NocE, homologous to NosL, belongs to radical SAM-dependent methyltransferase family.[23] This sequence is also homologous to ThiH-like protein,[24] which was proposed to cleave the C*a*-C*ß* bond of L-tryptophan and subsequently lead to the rearrangement of tryptophan for the formation of 3-methylindole-2-carboxylic acid.[8] Meanwhile, NocB, also shares high sequence similarity to a variety of putative radical SAM-dependent methyltransferases,[24] not highly similar to corresponding NosN gene for the biosynthesis of nosiheptide.[8] This type of enzyme was postulated to re-methylate 3-methylindole-2-carboxylic acid at the C-4 position, and the dimethyl carboxylic acid would be hydroxylated to form 4-hydroxymethyl-3-methylindole-2-carboxylic acid, which could be further tethered to the carboxyl group of glutamic acid and the hydroxyl group of serine within the structural peptide *via* two ester bonds.[8] Moreover, NocC, similar to NosO, encodes a protein possibly catalyzing the formation of central 2,3,5,6-tetrasubstituted pyridine domain.[8, 9, 13] Therefore, a proposed biosynthetic pathway for nocathiacin is presented in Figure 3. Although reverse genetics approach has been applied to explore the biosynthesis of ribosomally synthesized peptides for long time,[25,26] thiopeptides with extensive posttranslational modifications have not been a subject for utilizing this type of approach due mainly to the structural complexity and the uncertainty of array and orientation of the biosynthetic genes for these peptides. The present study provides the first example on the rapid and effective elucidation of biosynthetic pathways of these complex molecules, which could generally be applied to various ribosomally synthesized bioactive peptides.

Table 1. Open reading frames in the biosynthetic gene cluster of nocathiacin

Gene	Size[a]	Protein homologue and origin[b]	Identity/similarity (%)	Proposed function
nocA	49	Nosiheptide precursor peptide (FJ438820); *Streptomyces actuosus* ATCC 25421	82/94	Noc precursor peptide
nocB	391	Tlm Orf11 (ABL74954); *Streptoalloteichus hindustanus*	34/67	SAM-dependent oxidase or methyltransferase
nocC	313	NosO (ACR48344); *Streptomyces actuosus* ATCC 25421	40/71	Hypothetical protein
nocE	397	NosL (ACR48341); *Streptomyces actuosus* ATCC 25421	75/90	Radical SAM: Related to biotin and thiamine synthesis

[a] Numbers of amino acids. [b] NCBI accession numbers are given in parentheses.

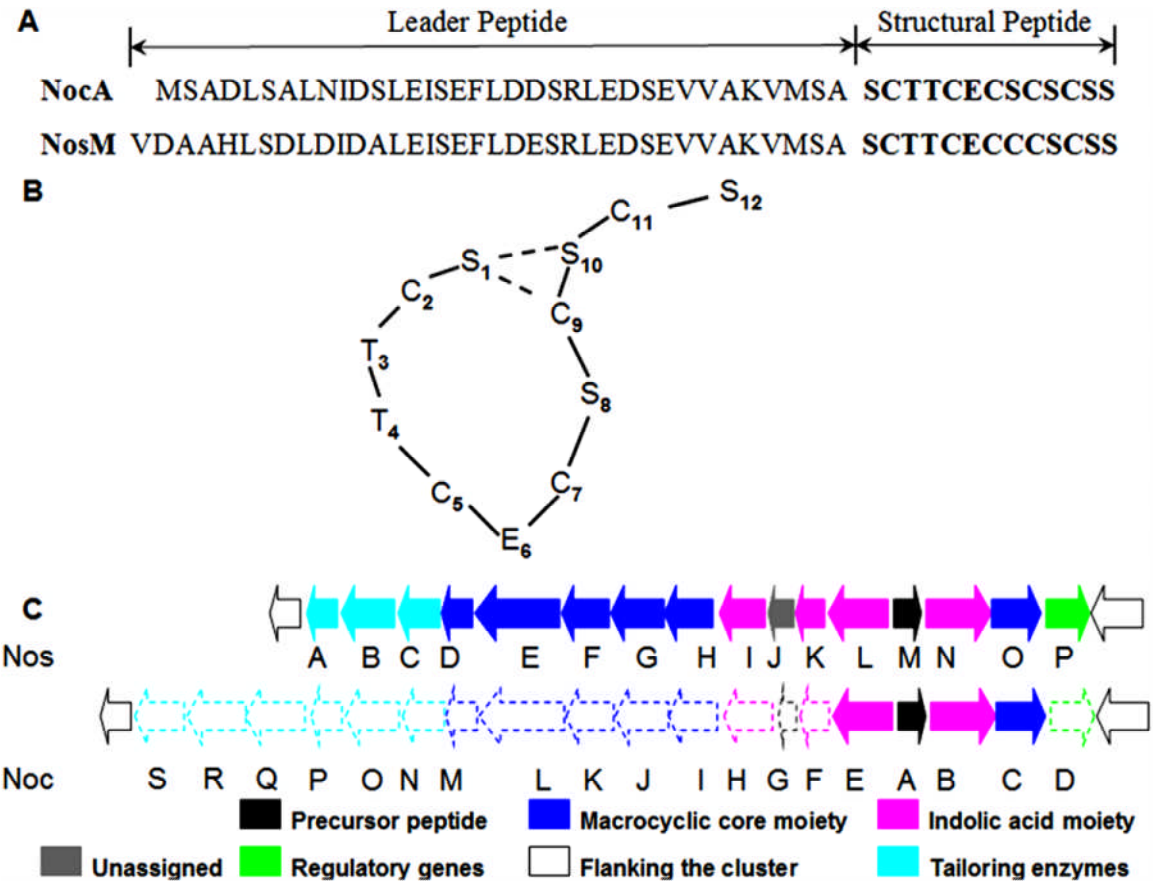

Figure 2. Biosynthetic genes of nocathiacin. (**A**) Precursor peptide of nosiheptide and nocathiacin. (**B**) Core sequence of nocathiacin from isotope feeding. (**C**) Partial biosynthetic gene cluster of nocathiacin.

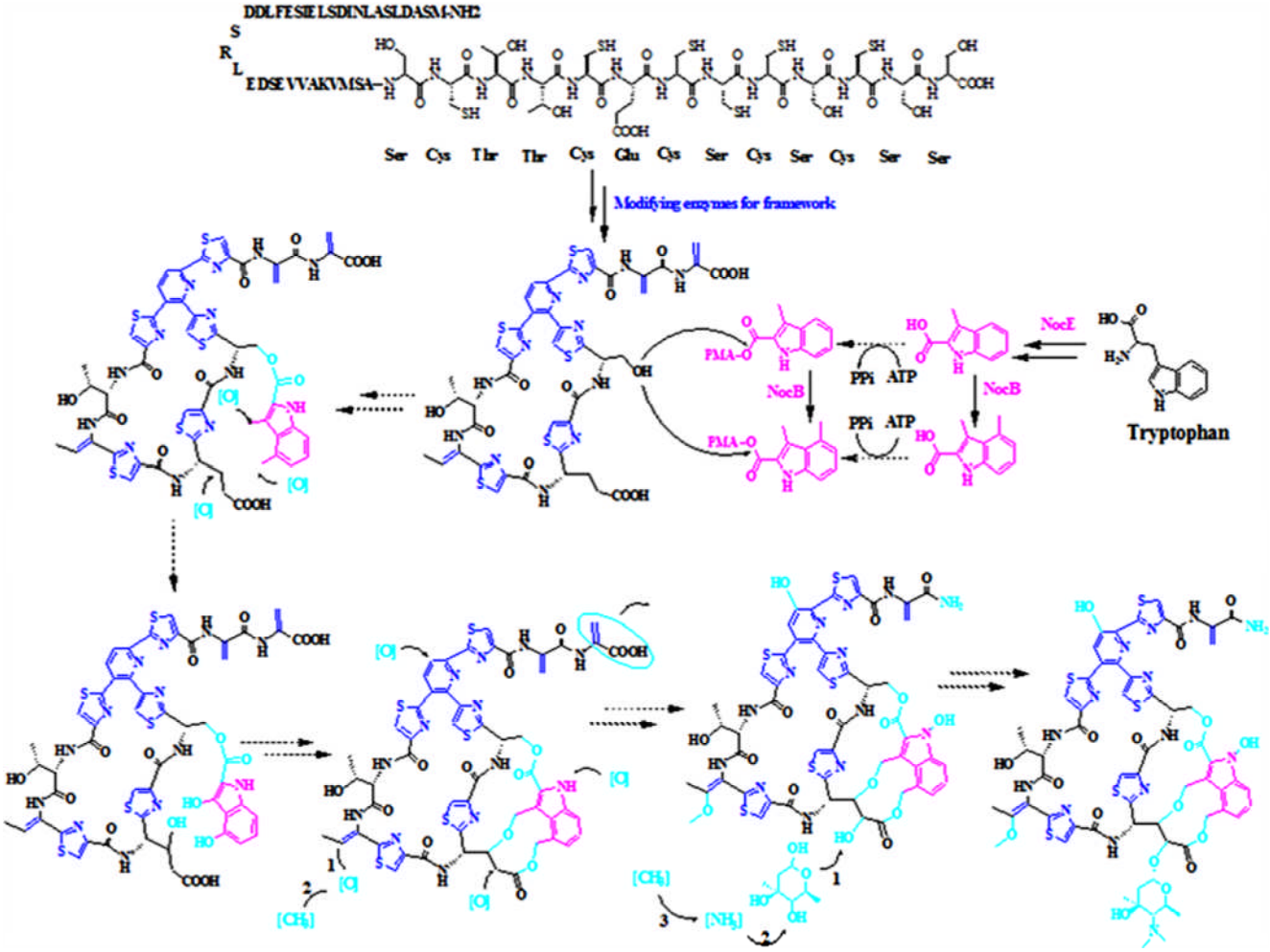

Figure 3. Proposed biosynthetic pathway of nocathiacin. Biosynthesis of indolic acid originated from the rearrangement of tryptophan and subjected to a series of modifications (pink); Modifications of the macyclic core structure (blue); The tailoring steps including hydroxylation, glycosylation, amination, methylation and cleavage of the C-terminal residue to form amide.

While we were preparing this manuscript, the biosynthetic pathway of nocathiacin was reported by Ding *et al* (online available since May 17, 2010) with the method of genomic library screening.[27] However, the detailed sequence information of this biosynthetic gene cluster is currently unavailable to the public except for the precursor peptide. Compared to the published results, our study generated an identical precursor peptide sequence. Moreover, the array and orientation of the related tailoring enzymes involving in the biosynthesis of nocathiacin are also same as the results obtained by genomic library screening, which further confirms the validity of

our reverse genetics approach.

In summary, our study has provided evidence, for the first time, to take advantages of the combination of core structure and related biosynthetic genes to rapidly acquire the information on the biosynthetic pathways of thiopeptides. Moreover, the present study clearly demonstrates the close relationship between structural peptide sequence of thiopeptides within the same series and their biosynthetic genes. Thus, our approach offers a shortcut for speedily elucidating the biosynthetic gene clusters of various thiopeptides, which would facilitate the generation of diverse analogues by subsequent genetic manipulations.

Experimental Section

PCR primers were designed based on the structural peptide sequence of nocathiacin and the conserved motifs of modifying enzymes located at both sides of nosiheptide precursor peptide. When the conserved motifs contained more than 6 identical amino acid residues, conventional primers were designed to avoid the degeneracy of the last base resided in three prime. Otherwise, degenerated primers were designed according to the principle of 'CODEHOP'.[28]

All PCR reactions were performed with 100 ng of genomic DNA of *Amycolatopsis fastidiosa* ATCC 202099,[20] which was prepared as previously described.[22] The reaction mixtures contained 2.5 µl 10 × DNA polymerase buffer, 2.0 mM $MgCl_2$, 0.2 mM of each dNTP, 5% DMSO, 5% glycerol, 2.0 µM of each primer, 1 U Taq polymerase and deionized water with a total volume of 25µl. The amplification, based on the structural sequences combined with conserved motifs, was performed according to 'Conventional PCR' procedure except that the annealing temperature was adjusted to 60 °C. For genome walking, the PCR reaction was conducted according to the method previously described.[22]

The follow-up processes to obtain positive colonies were conducted according to the method in the literature.[29] The positive colonies were sequenced by Invitrogen (Shanghai, China) with ABI Genetic Analyzer 3730 (Applied Biosystems), and DNA sequence analyses were performed by Vector NTI Advance v10.3 (Invitrogen Co.,

USA). Sequence alignments were conducted with BLASTnr from NCBI online service (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

Acknowledgements

This work was supported by the "111 Project" from the Ministry of Education of China and the State Administration of Foreign Expert Affairs of China (No. 111-2-07) and a National Key Project on Science and Technology of China (2009ZX09103-089).

References

[1] C. F. Higgins, *Nature* **2007**, *446*, 749-757.

[2] M. A. Fischbach, C. T. Walsh, *Science* **2009**, *325*, 1089-1093.

[3] M. C. Bagley, J. W. Dale, E. A. Merritt, X. Xiong, *Chem. Rev.* **2005**, *105*, 685-741.

[4] R. A. Hughes, C. J. Moody, *Angew. Chem. Int. Ed. Engl*. **2007**, *46*, 7930-7954.

[5] V. S. Aulakh, M. A. Ciufolini, *J. Org. Chem.* **2009**, *74*, 5750-5753.

[6] H. D. Arndt, S. Schoof, J. Y. Lu, *Angew. Chem. Int*. *Ed. Engl*. **2009**, *48*, 6770-6773.

[7] C. Li, W. L. Kelly, *Nat*. *Prod. Rep*. **2010**, *27*, 153-164.

[8] Y. Yu, L. Duan, Q. Zhang, R. Liao, Y. Ding, H. Pan, E. Wendt-Pienkowski, G. Tang, B. Shen, W. Liu, *ACS Chem. Biol.* **2009**, *4*, 855-864.

[9] W. L. Kelly, L. Pan, C. Li, *J. Am. Chem. Soc.* **2009**, *131*, 4327-4334.

[10] L. C. Wieland Brown, M. G. Acker, J. Clardy, C. T. Walsh, M. A. Fischbach, *Proc. Natl*. *Acad. Sci*. *USA* **2009**, *106*, 2549-2553.

[11] A. S. Geiermann, R. Micura, *Chem. Biol.* **2009**, *16*, 1024-1025.

[12] R. P. Morris, J. A. Leeds, H. U. Naegeli, L. Oberer, K. Memmert, E. Weber, M. J. LaMarche, C. N. Parker, N. Burrer, S. Esterow, A. E. Hein, E. K. Schmitt, P. Krastel, *J. Am. Chem. Soc*. **2009**, *131*, 5946-5995.

[13] R. Liao, L. Duan, C. Lei, H. Pan, Y. Ding, Q. Zhang, D. Chen, B. Shen, Y. Yu, W. Liu, *Chem. Biol*. **2009**, *16*, 141-147.

[14] J. Wang, Y. Yu, K. Tang, W. Liu, X. He, X. Huang, Z. Deng, *Appl. Environ. Microbiol.* **2010**, *76*, 2335-2344.

[15] C. T. Walsh, M. G. Acker, A. A. Bowers, *J. Biol. Chem*. DOI: 10.1074/jbc.R110.135970

[16] W. Li, J. E. Leet, H. A. Ax, D. R. Gustavson, D. M. Brown, L. Turner, K. Brown, J. Clark, H. Yang, J. Fung-Tomc, K. S. Lam, *J. Antibiot. (Tokyo)* **2003**, *56*, 226-231.

[17] K. C. Nicolaou, S. H. Lee, A. A. Estrada, M. Zak, *Angew. Chem. Int*. *Ed. Engl*. **2005**, *44*, 3736-2740.

[18] L. Xu, A. K. Farthing, Y. J. Shi, P. T. Meinke, K. Liu, *J*. *Org*. *Chem.* **2007**, *72*, 7447-7450.

[19] W. Li, S. Huang, X. Liu, J. E. Leet, J. L. Cantone, K. S. Lam, *Bioorg. Med. Chem. Lett*. **2008**, *18*, 4051-4053.

[20] M. Wei, S. Wang, Y. Fang, Y. Chen, *Bioresour. Technol.* **2010**, *101*, 3617-3622.

[21] S. B. Singh, K. Herath, N. X. Yu, A. A. Walker, N. Conors, *Tetrahedron Lett.* **2008**, *49*, 6265-6268.

[22] J. Deng, M. Wei, B. Yu, Y. Chen, *Appl. Microbiol*. *Biotechnol*. **2010**, 87, 757-764.

[23] S. J. Booker, R. M. Cicchillo, T. L. Grove, *Curr. Opin. Chem. Biol.* **2007**, *11*, 543-552.

[24] M. Kriek, F. Martins, M. R. Challand, A. Croft, P. L. Roach, *Angew. Chem. Int. Ed. Engl*. **2007**, *46*, 9223-9226.

[25] T.J. Oman, & van der Donk, W.A. *Nat*. *Chem. Biol.* **2009**, *6*, 9-18.

[26] E.W. Schmidt, *BMC Biol.* **2010**, *8*, 83-86.

[27] Y. Ding, Y. Yu, H. Pan, H. Guo, Y. Li, W. Liu, *Mol. Biosyst.* **2010**, *6*, 1180-1185.

[28] R. Boyce, P. Chilana, T. M. Rose, *Nucleic Acids Res.* **2009**, *37*, W222-W228.

[29] J. Sambrook, Molecular cloning - a laboratory manual, 3rd ed. (Eds.: J. Sambrook, D.W. Russell), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, **2001**.

Supporting Information

Simple Reverse Genetics Approach to Elucidating the Biosynthetic Pathway of Complex Thiopeptide Nocathiacin

Maochen Wei, Jing Deng, Shuzhen Wang, Nan Liu, and Yijun Chen*

Contents

Gene[a]	Primer	Sequence (5'to 3')[b]	Conserved motifs[c]
NocA	Forward	TCGTGCACCACCTGCGAGTG	SCTTCEC
	Reversed	GCACTCGCAGGTGGTGCACGA	SCTTCEC
NocB	Reversed1	GGGATGTCCGTCACCcartcrcaraa	FCDWVQAIP
	Reversed2	GCTGGCCGTGCCGCCGCCCCAG	YWGGGT
NocC	Reversed1	CGCCTGCGCGGCyttnggrtcnc	RDPKAAQA
NocE	Forward	GTACGTCTCGCGGTCGTAGctytcytgraa	FQESYDRETY

Table S1. [a] NocA, precursor peptide of nocathiacin; NocB, tailoring enzyme close to the downstream of NocA; NocC, tailoring enzyme close to the downstream of NocB; NocE, tailoring enzyme close to the upstream of NocA.
[b] IUB (degenerated bases) code was used for degenerated base positions: N = A+C+G+T; Y = C+T; R = A+G.
[c] Conserved amino acid sequences obtained from the alignments of related enzymes in nosiheptide biosynthetic genes with other homologous sequences.

Gene[a]	Primer	Sequence (5'to 3')[b]	Direction
NocC	Down-outer	GCCTGCTGGCAGACCTGGCGGTCGGGTAG	Downstream
	Down-inner	CATCCGTCGCGCATCACCGGCCGCTTGCTG	Downstream
NocE	Up1-outer	GGCTTCGAGCGGGTCTACTTCAACAT	Upstream
	Up1-inner	CGAGATCGACGTGCTCGCCGAGTGGG	Upstream
	Up2-outer	AGGCGATCGGCCTGGCGGGCGTCATCAGC	Upstream
	Up2-inner	GACTCGCAGGCCCGCAACGAGGAGACGAC	Upstream
Random primers			
	Semi-1	GCCAATTCCGGATNGAYKSNGGNTC	Universal walker primer
	Semi-2	GCCTTAAGGCCTANGARMSNCCNAG	Universal walker primer
	Semi-3	CGGTTAAGGCCTANYTCSKNGANGC	Universal walker primer
	Semi-4	GCCAATTCCGGATNSAGYMNCTNCG	Universal walker primer

Table S2. [a] NocC, tailoring enzyme close to the downstream of NocB; NocE, tailoring enzyme close to the upstream of NocA.
[b] IUB (degenerated bases) code was used for degenerated base positions: N = A+C+G+T; S = C+G; Y = C+T; M = A+C; R = A+G; K = T+G

Figure S1. Cop, coproporphyrinogen III oxidase (*Nonomuraea sp.* WU8817); Tlm, Tlm orf11 (*Streptoalloteichus hindustanus*); SAM, SAM-dependent oxidase or methyl transferase (*Streptomyces flavoviridis*). Blue, conserved motifs for primer designs.

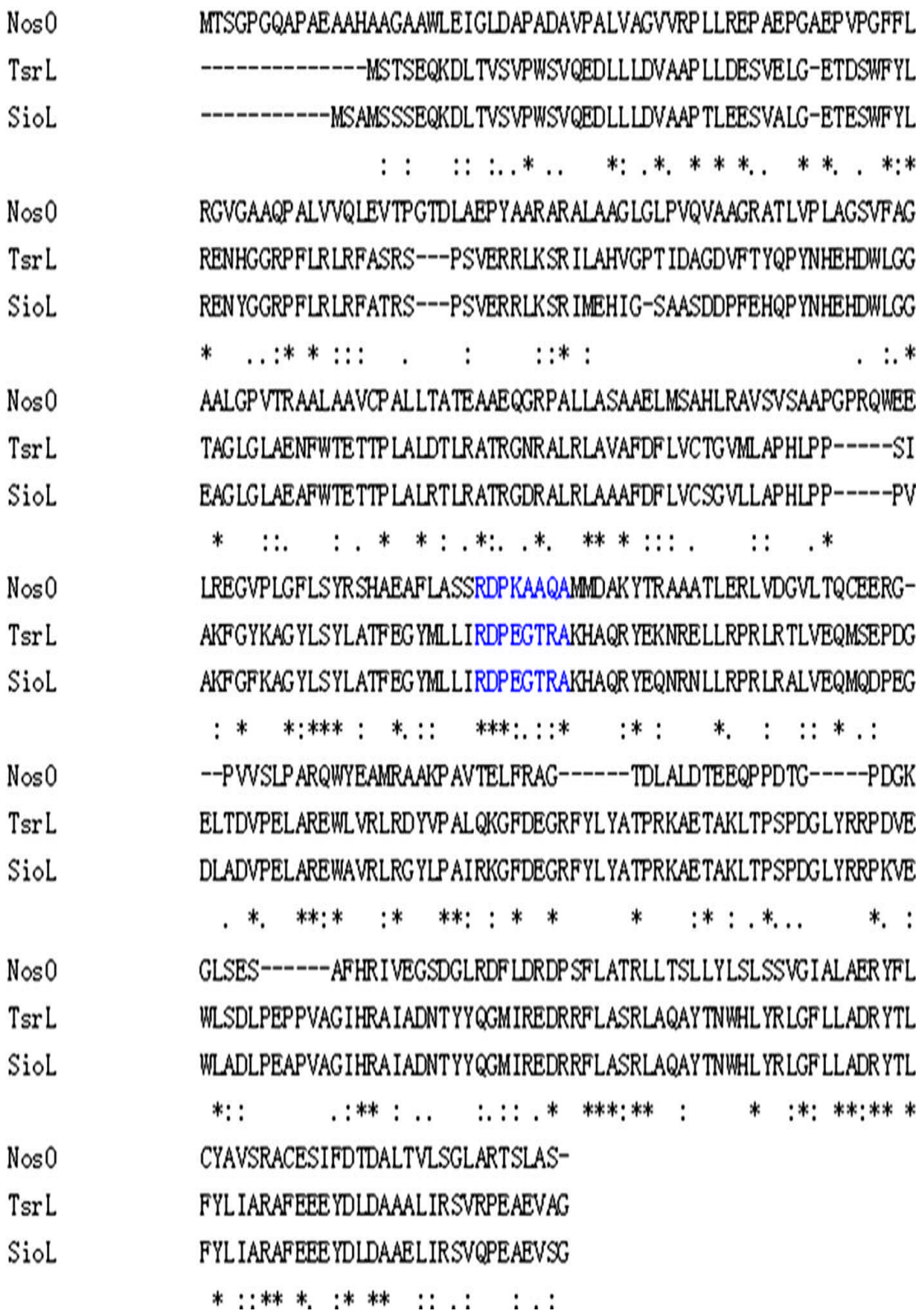

Figure S2. Tsr, thiostrepton (*Streptomyces laurentii*); Sio, Siomycin (*Streptomyces sioyaensis*). Blue, conserved motif for primer design.

Figure S3. Biotin, Biotin and thiamin synthesis associated (*Anaerocellum thermophilum* DSM 6725); ThiH, Thiamine biosynthesis protein ThiH (*Clostridium tetani* E88). Blue, conserved motif for primer design.