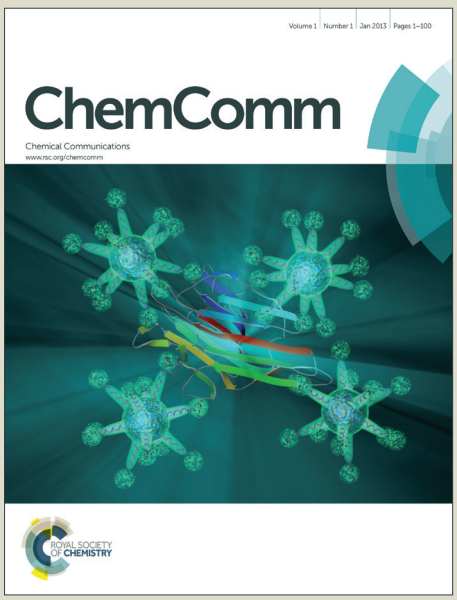
ChemComm

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/chemcomm

Journal Name

RSCPublishing

COMMUNICATION

Directly utilizing an endogenous gene to dissect regulatory elements in the biosynthetic gene cluster of nosiheptide†

Cite this: DOI: 10.1039/xoxxoooooox

Received on xxxx,
Accepted on xxxx

DOI: 10.1039/xoxxoooooox

www.rsc.org/

Min Ma,[a] Yanjiu Xue,[a] Weiying Liu,[a] Hong Zhang,[a] Lingyi Kong,[a] Shuzhen Wang,[a]* and Yijun Chen[a]*

The regulatory elements for nosiheptide biosynthesis were identified by a novel host-vector system with an endogenous gene within the biosynthetic gene cluster as a reporter gene. The present study offers a rapid and reliable method for the identification of regulatory elements in the biosynthesis of various bioactive natural products.

The production of bioactive natural products depends on the biosynthetic machinery and the regulatory mechanism in the producing organisms. [1] Given the importance of promoters on gene regulation, extensive efforts have been made to uncover non-coding DNA regions. To reveal the putative promoters in *Streptomyces* sp., a major source of important bioactive natural products, integrative or self-replicating promoter-probe vectors are typically required in conjunction with the use of an exogenous reporter gene, such as chloramphenicol acetyltransferase gene[2] and *ampC* β-lactamase gene from *E. coli*,[3] brown pigment gene from *S. coelicolor*,[4] *neo* from transposon Tn5,[5] *luxAB* from *V. harveyi*,[6] *xylE* from *P. putida*[7] and *gfp*[8]. Despite the success in promoter identification in the past, the use of most reporter genes requires being compatible with the host, and multiple rounds of screening often result in a large number of false positives and false negatives. Meanwhile, although GFP can be quantitative to simplify the screenings, the effects of solid components and high fluorescent interference in the culture media make it difficult to use in *Streptomyces* sp. Therefore, direct utilization of an endogenous and functional gene as a reporter would be advantageous to avoid the shortcomings from the use of exogenous reporters.

Streptomyces actuosus ATCC 25421 produces nosiheptide, a typical member of thiopeptide, possessing strong activities against gram-positive bacterial pathogens. The biosynthesis of nosiheptide originates from a 50-aa precursor peptide, containing a 37-aa leader peptide and a 13-aa core peptide that undergoes complex post-translational modifications.[9] The biosynthetic gene cluster for nosiheptide (***nos*** cluster) was previously reported (GenBank: FJ438820.1),[9a] from which sixteen different genes are principally involved. Although the functions of a number of genes in the cluster, including *nosABCLMNO*, have been characterized,[9c, 10] the transcriptional regulation of nosiheptide biosynthesis remains elusive.

It was reported that NosA, the *nosA* gene encoding protein, catalyzes an enamide dealkylation to remove the acrylate unit originated from the extended serine residue in the final step of nosiheptide maturation, in which a stable and constantly accumulated bis-dehydroalanine precursor (**1**) is converted to nosiheptide (**2**).[10a] In the present study, we hypothesize that coupling the known function of NosA to construct a host-vector system could be a feasible way to rapidly indentify the promoter(s) in ***nos*** cluster. As shown in Fig. 1, such a system relies on the combination of a previously established *nosA*-inactivated host strain L1101[9b] and a self-replicating vector pWHM7, in which the promoterless *nosA* gene is served as a reporter. This host-vector system allowed to efficiently reveal regulatory elements in the biosynthetic gene cluster of nosiheptide.

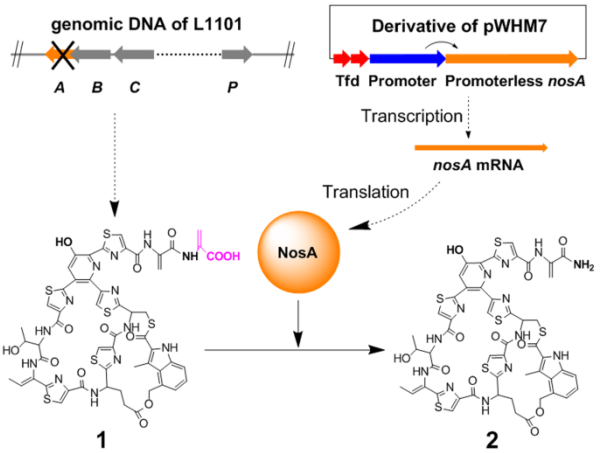


Fig. 1 The host-vector system of L1101-pWHM7 for the identification of promoters in *S. actuosus*. The host strain L1101 is a *nosA*-inactivated mutant of wild-type that only produces a precursor **1** and completely loses the ability to generate mature product **2**. The vector pWHM7 contains a polylinker whose restriction sites are preceded by two fd terminators. The promoterless *nosA* gene was inserted into the polylinker in the right direction to act as a reporter gene. Upon the insertion of a promoter-active DNA fragment in front of the promoterless *nosA* gene into pWHM7, the system is able to convert **1** to **2** by expressed NosA.

To establish such a system for promoter identification, the generation of mutant strain L1101 was implemented by knockout of *nosA* gene as described previously.[9b] The Δ*nosA* mutant L1101 lost the capability of producing **2**, and precursor **1** was generated instead (Fig. 2a). Compound **1** was subsequently purified from the culture broth for structural determination. HR-ESI-MS analysis showed *m/z* 1292.1602$[M+H]^+$ for **1** (Fig. S1), in accordance to the compound previously characterized with a molecular formula of $C_{54}H_{46}N_{13}O_{14}S_6$ from the *nosA*-inactivated strain.[10a]

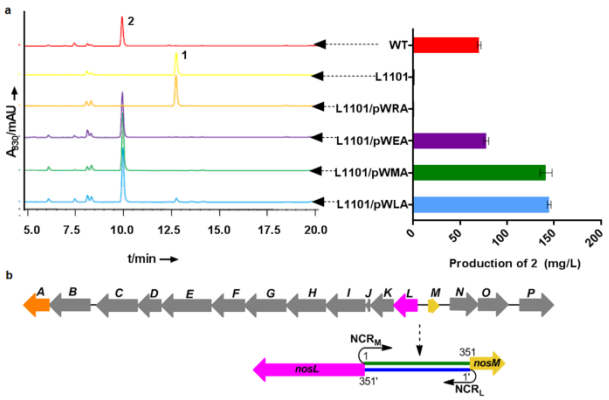


Fig. 2 Analysis of nosiheptide (**2**) production by different mutant strains and the organization of the biosynthetic gene cluster for nosiheptide. (a) HPLC analysis of fermentation extracts from wild-type strain and different mutants of L1101 containing derivatives of pWHM7. Left panel, HPLC chromatograms; Right panel, quantification of HPLC analysis. Data are average values of three independent experiments with standard deviations. WT, *S. actuosus* ATCC25421; L1101, *nosA*-inactivated strain; pWRA, derivative of pWHM7 carrying the promoterless *nosA* gene with original ribosomal binding site; pWEA, derivative of pWHM7 carrying a fusion of promoter P*ermE** and the promoterless *nosA* gene; pWMA and pWLA are derivatives of pWHM7 carrying fusions of $NCR_{L\text{-}M}$ in opposite orientation plus the promoterless *nosA* gene. (b) Nosiheptide biosynthetic gene cluster in *S. actuosus* ATCC 25421 and the non-coding region $NCR_{L\text{-}M}$ (351 bp) located between divergent genes *nosL* and *nosM*. The genes are shown by arrows. Bent arrows indicate the direction of transcription from both strands of $NCR_{L\text{-}M}$. NCR_M and NCR_L represent $NCR_{L\text{-}M}$ in different orientations.

Next, a promoter-probe vector capable of conjugating between *E. coli* and *Streptomyces* was constructed as illustrated in Fig. S2. The vector pWHM7 was generated starting from pWHM3,[11] an *E. coli-Streptomyces* shuttle plasmid. In order to provide additional sites for fragment insertion, pWHM5 was constructed by replacing the *EcoRI-HindIII* fragment on pWHM3 with a 300 bp polylinker (Fig. S3) containing five unique restriction sites (*EcoRI, XbaI, PstI, SphI, HindIII*). Two fd terminators, known to function in *Streptomyces*[5], were added infront of *XbaI* sites to prevent the influence from the promoters on the vector. Then, a fragment containing apramycin resistance gene and the conjugation gene were subcloned from pIJ773 to *EcoRI* site of pWHM5 to yield pWHM7. After extensive sequencing and analysis, the orientation shown in Fig. S1 was confirmed, in which the introduction of apramycin resistance gene was also verified. According to these characteristics, the ability to conjugate between *E. coli* and *Streptomyces* by pWHM7 further increased the potential utility of this plasmid on the genetic manipulation in *Streptomyces*. Moreover, the non-integrative nature of pWHM7 in *Streptomyces* could avoid affecting the structure and function of chromosomal DNA in the host. Given that NosA catalyzes the conversion of **1** to **2**, the production of **2** from **1** could reflect the activity of any promoter(s) inserted upstream of the promoterless *nosA* on pWHM7, clearly indicating the advantages of *nosA*-dependent L1101-pWHM7 system, such as the simple and quantitative assay of compound **2** production and the homologous nature of host strain L1101.

To investigate possible transcriptional read-through effects on the reporter gene from the promoters of pWHM7, the promoterless *nosA* gene retaining its own ribosomal binding site was inserted into pWHM7 at *XbaI-HindIII* sites to create a vector pWRA as a negative control. Meanwhile, to ascertain the utility of the host-vector system, a strong promoter P*ermE** capable of expressing in various *Streptomyces*[5, 12] was cloned to upstream of the promoterless *nosA* gene to result in a vector pWEA as a positive control. These recombinant plasmids were separately introduced into L1101 by conjugation from *E. coli* ET12567/pUZ8002 donor strain, generating the mutant strains, L1101/pWRA and L1101/pWEA, respectively. As anticipated, no production of **2** was observed in L1101/pWRA, whereas L1101/pWEA fully restored the production of **2**(Fig. 2a). In addition, the transcriptional read-through from vector-borne promoters was effectively prevented by the addition of fd terminators. As a result, the expression of promoterless *nosA* gene was solely dependent on the presence of a promoter-active fragment and would not be complicated by background read-through to reliably identify the promoters in *S. actuosus*.

A 351 bp non-coding region ($NCR_{L\text{-}M}$) located between the divergently organized genes *nosL* and *nosM* in ***nos*** cluster was speculated to contain regulatory elements,[9a] we therefore targeted this region for promoter identification (Fig. 2b). To investigate if there are promoters in both orientations in the non-coding DNA, we separately ligated $NCR_{L\text{-}M}$ and its inverse fragment with the promoterless *nosA* gene, and then inserted these fusions into pWHM7 (Fig. 3a), creating plasmids pWMA and pWLA. The resulting plasmids were respectively introduced into L1101, yielding mutant strains L1101/pWMA and L1101/pWLA. The results of HPLC-UV analysis of the fermentation extracts indicated that both mutant strains fully restored the ability of converting **1** to **2** (Fig. 2a), suggesting that $NCR_{L\text{-}M}$ bears strong promoters in both orientations. This divergent promoter region represents a class of bidirectional transcription control element that is widespread in various organisms.[13] Although the regulatory significance of this type of bidirectional promoter has not been well understood, a close correlation between the production of **2** and the DNA fragment inserted upstream of the *nosA* gene was clearly observed in this study. When the fragment of $NCR_{L\text{-}M}$ or inverse $NCR_{L\text{-}M}$ from *S. actuosus* was inserted, the mutant strain produced approximately two-fold more nosiheptide than wild-type and L1101/pWEA strains (Fig. 2a), which could result from the high-copy number of pWHM7 or the sensitivity of promoter-RNA polymerase interactions.

Because mRNA transcriptional start point (TSP) is an essential element in the promoter region, 5' rapid amplification of cDNA ends (5' RACE) was employed to annotate the TSPs of the divergently arranged genes, *nosL* and *nosM*. 5' RACE of the transcripts from *nosL* and *nosM* mRNAs were carried out on $polyC^+$ RNAs from the wild-type strain. The results (Fig. S4) enabled us to map the TSP of *nosM* (55 bp upstream of its start codon GTG) and *nosL* (31 bp upstream of its start codon ATG) as shown in Fig. 3b. Both TSPs are located in $NCR_{L\text{-}M}$, flanking by 265 bp non-coding DNA.

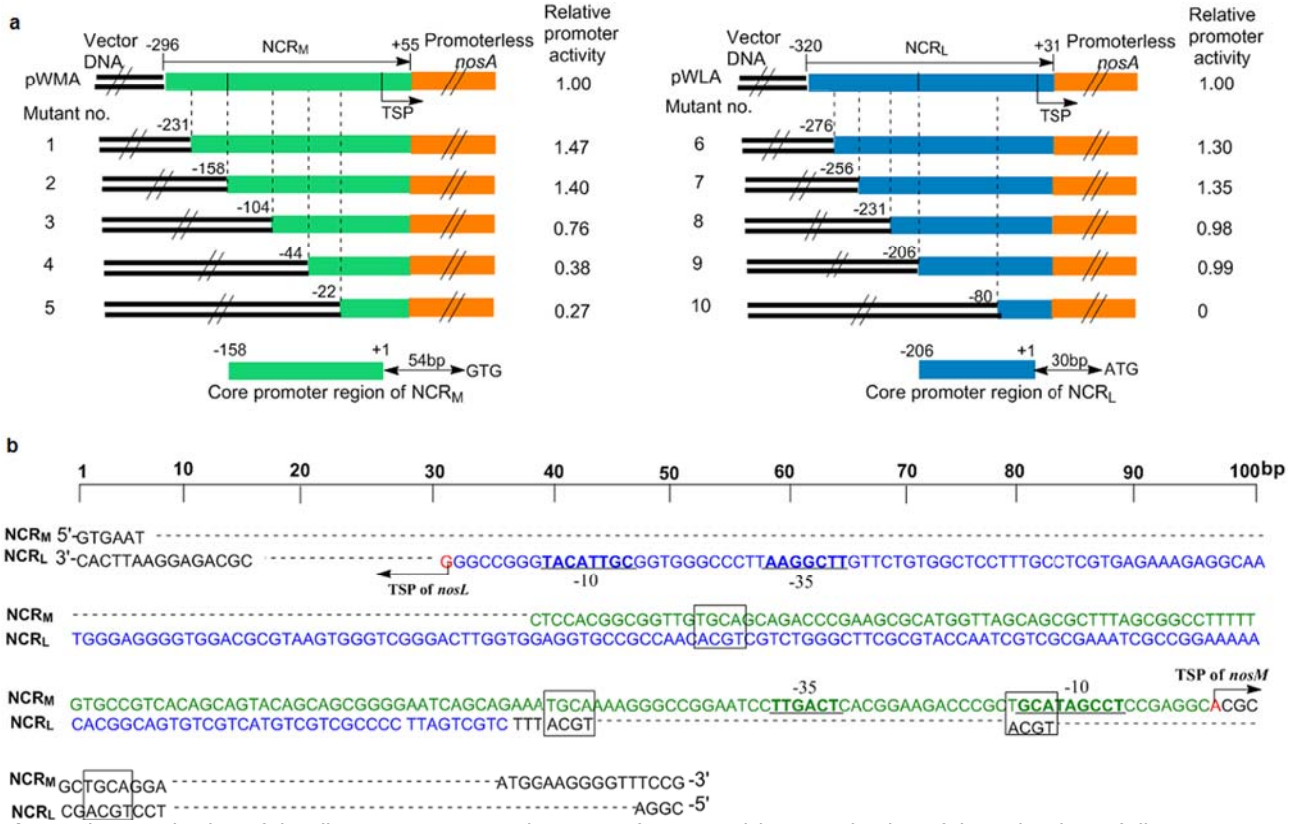


Fig. 3 Characterization of the divergent promoters in NCR_{L-M} fragment. (a) Determination of the 5' borders of divergent promoters in NCR_{L-M} by 5' deletion. Recombinant plasmids pWMX (X=1, 2, 3, 4, 5) and pWLY (Y=6, 7, 8, 9, 10) represent two sets of derivatives of pWHM7 carrying fusions of 5'-truncated NCR_{L-M} in either direction in combination with the promoterless *nosA* gene. After the mutant plasmids were introduced into L1101, the activity of NosA was assayed by analyzing the production of **2** in each mutant strain. (b) Nucleotide sequence of the bidirectional promoter region NCR_{L-M} in ***nos*** cluster. NCR_M and NCR_L indicate different strands of NCR_{L-M} DNA. The TSPs of the promoters are shown by bent arrows. The core promoter regions of the divergent promoters are indicated in blue and green, respectively. The putative -10 and -35 boxes of the divergent promoters are underlined. Boxes show the sequence of TGCA, which is repeated four times with varied spaces from each repeat.

In an effort to further annotate the core promoters in NCR_{L-M}, the host-vector system was used to develop two sets of mutant strains of L1101 bearing recombinant plasmids whose promoterless *nosA* gene was preceded by various 5' deletions ofNCR_{L-M} or its inverse fragments (Fig. 3a). These mutant strains were respectively denoted as L1101/pWMX (X=1, 2, 3, 4, 5) and L1101/pWLY (Y=6, 7, 8, 9, 10). Fig. S5 summarizes the results of HPLC analysis of the fermentation extracts from different mutant strains. As shown in Fig. 3a, mutants 1, 2, 6, 7, 8 and 9 had little or no effects on the expression of the reporter gene, indicating that the deleted regions in these mutants do not have control elements. By contrast, reporter-gene expression was significantly reduced in mutants 3, 4, 5 and 10, confirming that the 5' borders of promoters of *nosM* and *nosL* are located 158 bp and 206 bp upstream of their respective TSPs. A further sequence analysis of this divergent promoter region of NCR_{L-M} revealed same characteristics shared by both strands, in which the sequence of TGCA occurs four times with spaces of 84 bp, 19 bp and 35 bp from each repeat (Fig. 3b). Such repeated sequence was suggested to function as binding sites for regulatory proteins according to previous studies.[14]

Conclusions

We have developed a host-vector system L1101-pWHM7 to characterize the promoters from *S. actuosus* ATCC 25421, the producing strain of nosiheptide. An endogenous *nosA* gene within the biosynthetic gene cluster was directly utilized and inserted on pWHM7 in the right direction to serve as a reporter gene. Based on this coupling system, the promoter activity of inserted DNA fragments in front of the promoterless *nosA* was then examined by quantification of the production of **2** converted from **1** by NosA. Combining the host-vector system with 5' RACE experiments, NCR_{L-M} region in the ***nos*** cluster was unambiguously revealed to be a bidirectional promoter region with two different TSPs on both strands, which could be the binding sites for regulatory proteins in the native host. The current method can facilitate the rapid identification of regulatory elements for the elucidation of regulation mechanism in the biosynthetic machinery and for the genetic manipulation to remarkable increase the production of bioactive natural products.

We thank Dr. Sheng Yang at Institute of Plant Physiology & Ecology, Shanghai Institute of Biological Sciences, CAS for providing the plasmid of pWHM3. This work was supported by the grants from the "111" Project from the Ministry of Education of China and the State Administration of Foreign Export Affairs of China (No: 111-2-07), the National Key Project on Science and Technology of China (No:

2012ZX09103101030), National Science Foundation of China (No: 81172967) and the Priority Academic Program Development of Jiangsu Higher Education Institutions .

Notes and references

[a]*State Key Laboratory of Natural Medicines and Laboratory of Chemical Biology, China Pharmaceutical University, 24 Tongjia Street, Nanjing, Jiangsu Province, 210009, China. Fax: +86 25 83271031; Tel: +86 25 83271045; E-mail: yjchen@cpu.edu.cn, shuzhenwang@cpu.edu.cn*

† Electronic Supplementary Information (ESI) available: Table S1-S2, Fig. S1-S5, experimental details. See DOI: 10.1039/c000000x/

1 G. Liu, K. F. Chater, G. Chandra, G. Niu and H. Tan, *Microbiol. Mol. Biol. Rev.*, 2013, **77**, 112-143.
2 M. J. Bibb and S. N. Cohen, *Mol. Gene Genet.*, 1982, **187**, 265-277.
3 M. Forsman and B. Jaurin, Mol. Gene Genet., 1987, **210**, 23-32.
4 S. Horinouchi and T. Beppu, J. Bacteriol., 1985, **162**, 406-412.
5 J. M. Ward, G. R. Janssen, T. Kieser, M. J. Bibb, M. J. Buttner and M. J. Bibb, Mol. Gene Genet., 1986, **203**, 468-478.
6 A. Schauer, M. Ranes, R. Santamaria, J. Guijarro, E. Lawlor, C. Mendez, K. Chater and R. Losick, *Science*, 1988, **240**, 768-772.
7 C. Ingram, M. Brawner, P. Youngman and J. Westpheling, *J. Bacteriol.*, 1989, **171**, 6617-6624.
8 A. Khetan, W.S. Hu and D. H. Sherman, *Microbiology*, 2000, **146**, 1869-1880.
9 (a) Y. Yu, L. Duan, Q. Zhang, R. Liao, Y. Ding, H. Pan, E. Wendt-Pienkowski, G. Tang, B. Shen and W. Liu, *ACS Chem. Biol.*, 2009, **4**, 855-864; (b) W. Liu, M. Ma, Y. Xue, N. Liu, S. Wang and Y. Chen, *ChemBioChem*, 2013, **14**, 573-576; (c) W. Liu, Y. Xue, M. Ma, S. Wang, N. Liuand Y. Chen, *ChemBioChem*, 2013, **14**, 1544-1547.
10 (a) Y. Yu, H. Guo, Q. Zhang, L. Duan, Y. Ding, R. Liao, C. Lei, B. Shen and W. Liu, *J. Am. Chem. Soc.*, 2010, **132**, 16324-16326; (b) Q. Zhang, Y. Li, D. Chen, Y. Yu, L. Duan, B. Shen and W. Liu, *Nat. Chem. Biol.*, 2011, **7**, 154-160; (c) S. Wang, S. Zhou and W. Liu, Curr. Opin. *Chem. Biol.*, 2013, **17**, 626-634; (d) H. Guo, J. Wang, Y. Li, Y. Yu, Q. Zheng, J. Wu and W. Liu, *Chem. Sci.*, 2014, **5**, 240-246.
11 J. Vara, M. Lewandowska-Skarbek, Y.G. Wang, S. Donadioand C. Hutchinson, *J. Bacteriol.*, 1989, **171**, 5872-5881.
12 M. J. Bibb, G. R. Janssen and J. M. Ward, *Gene*, 1985, 38, 215-226.
13 C. Beck and R. Warren, *Microbiol. Rev.*, 1988, **52**, 318.
14 A. Wietzorrek and M. Bibb, *Mol. Microbiol.*, 1997, **25**, 1181-1184.

