Supplementary Information

Molecular Cloning, Sequence Analysis and Functional Characterization of the Gene Cluster for Biosyntheses of K-252a and Its Analogs

Hsien-Tai Chiu,* Yi-Lin Chen, Chien-Yu Chen, Chyn Jin, Meng-Na Lee, and Yu-Chin Lin

Department of Biological Science and Technology,
National Chiao Tung University, 75 Po-Ai Street,
Hsinchu 300, Taiwan.

CONTENTS

*To whom correspondence should be addressed at the Department of Biological Science and Technology, National Chiao Tung University, 75 Po-Ai Street, Hsinchu 300, Taiwan.

Tel: +886-3-5131595; Fax: +886-3-5719605; E-mail: Chiu@mail.nctu.edu.tw.

Experimental Procedures (Supplementary)

Construction of expression plasmids for His-tagged NokA, NokB, NokC and NokD proteins

The *nokA*, *nokB*, *nokC* and *nokD* genes were amplified from the pJC3B5 fosmid clone as a template by polymerase chain reaction (PCR) amplifications with corresponding primers (Table S3) which introduced unique restriction sites at the 5' and 3' ends as follows. For *nokA* amplification, the primer pair, NKONF1 and NKONhR1, with the restriction sites *Nde*I and *Nhe*I, respectively, was used. NKDNF1 and NKDXR1 introduced with *Nde*I and *Xho*I sites, respectively, were used to amplify *nokB*. Similarly, *nokC* and *nokD* were amplified with corresponding 5' and 3'primers introduced with *Nde*I and *EcoR*I sites, where NKPNdF1 and NKPEcR1, respectively, were used for *nokC* amplification, NKCNdF1 and NKCEcR1, respectively, for *nokD* amplification. The reaction mixture containing 100 ng of DNA template, 5 μM of each primer and 0.3 mM of each dNTP was subsequently added with 1.25 units of *pfu* (Stratagen) to trigger the PCR reaction performed in a thermocycler (PTC-200 DNA Engine, MJ Research, USA). The PCR-amplified fragment of *nokA* was then cloned into an *Nde*I- and *Nhe*I-digested pET28a expression vector, that of *nokB* into an *Nde*I- and *Xho*I-digested pET21b expression vector, and those of *nokC* and *nokD* into an *Nde*I- and *EcoR*I-digested pET21b expression vector with corresponding sites, generating the N-terminal His-tag expression plasmid pJZ38 and the C-terminal His-tag expression plasmids pJZ18, pCY11 and pCY6, respectively.

Heterologous expression of His-tagged NokA, NokB, NokC and NokD proteins

Each DNA construct containing the coding gene with poly-His codons was individually transformed into *E. coli* BL21 (DE3) (Stratagen) for protein overexpression. To improve NokA solubility, a co-expression of N-terminal His_6-tagged NokA with the chaperone team encoded by pG-KJE7 was conducted. For protein production, the *E. coli* transformant was cultured at 37℃ in Luria-Bertani medium with 100 μg/ml ampicillin and 30 μg/ml kanamycin until OD_{600} reached 0.4-0.5. Subsequently, the co-expression was induced with 250μM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 0.2% (w/v) L-arabinose at 25℃ for 10 h. For expression of C-terminal His_6-tagged NokB, the *E. coli* culture was incubated at 37℃ with 100 μg/ml ampicillin, and was subsequently induced with 250 μM IPTG for overexpression at 15℃ for 24 h. For C-terminal His_6-tagged NokC expression, the *E. coli* culture was grown at 37℃ with 100 μg/ml ampicillin until OD_{600} of 0.3-0.4 reached, and was then induced with 50 μM IPTG at 30℃ for 4 h. To express C-terminal His_6-tagged NokD, the *E. coli* transformant was grown at 37℃ with 100

μg/ml ampicillin to reach OD_{600} of 0.4, at which point IPTG was added to give a final concentration of 50 μM. The induced culture was allowed to grow at 15℃ for 10 h.

Determination of the optimal co-expression condition for NokABCD

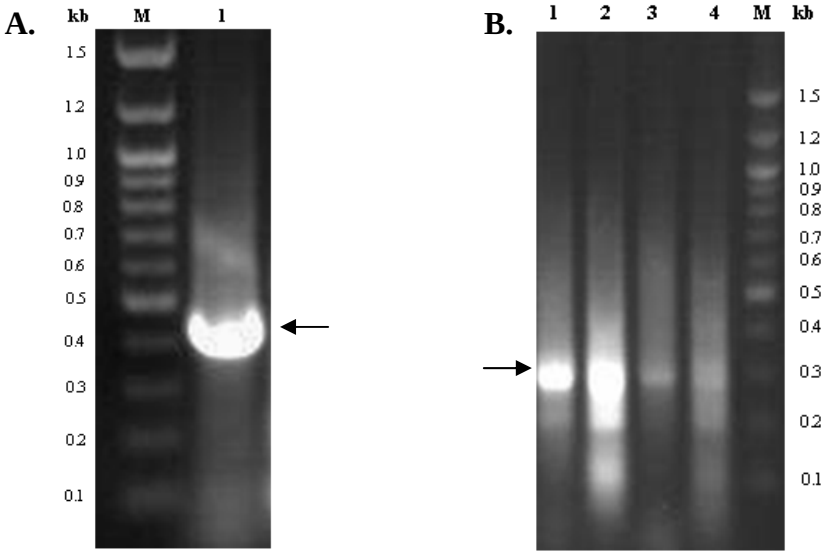
To probe the optimal co-expression condition for NokABCD and chaperone proteins, each of NokA, NokB, NokC and NokD was first examined for its own expression condition to obtain purified His-tagged protein as a molecular standard. Based on each of the four conditions, a selected co-expression condition was then applied to each individual to re-examine the solubility and production yield of each protein when expressed in the presence of the chaperones. After several cycles of the selection process, the final co-expression condition with the chaperones, as described in **Materials and methods**, was determined to be the induction with 250 μM IPTG and 0.2% (w/v) L-arabinose at 25℃ for 10 h for the co-expression construct coding for NokABCD.

Purification of each of His-tagged NokA, NokB, NokC and NokD proteins

For His-tagged NokA purification, the cells were collected by centrifugation (1,902g, 45 min) at 4℃ and were resuspended in buffer A (20 mM Tris-HCl, 200mM NaCl, 10% (v/v) glycerol, pH 7.8). Subsequently, the cells were broken and disrupted by French Press (two passages at 16,000 psi, Spectronic Instruments). After centrifugation at 30,000g at 4℃ for 1 h, the supernatant was subjected to purification by a Ni-NTA column (Qiagen) pre-equilibrated with buffer A at a flow rate of 1 ml/min and monitored at 254 nm. The His-tagged NokA was eluted with a linear gradient of 0-500 mM imidazole, constituted with buffer A and buffer B (20 mM Tris-HCl, 200 mM NaCl, 500 mM imidazole, 10% (v/v) glycerol, pH 7.8). For His-tagged NokB purification, the procedure similarly as described for NokA was used, where buffer A was replaced by buffer C (20 mM Tris-HCl, 10% (v/v) glycerol, pH 8.5) and buffer B by buffer D (20 mM Tris-HCl, 10% (v/v) glycerol, 500mM imidazole, pH 8.5). For His-tagged NokC purification, buffer E (25 mM Tris-HCl, 500 mM NaCl, 15% (v/v) glycerol, pH 7.6) and buffer F (25 mM Tris-HCl, 500 mM NaCl, 15% (v/v) glycerol, 500 mM imidazole, pH 7.6) were used to replace buffer A and buffer B, respectively. For His-tagged NokD, buffer G (20 mM Tris-HCl, 500 mM NaCl with 10% (v/v) glycerol at pH 7.8) and buffer H (buffer G plus 500 mM imidazole) were applied to replace buffer A and buffer B, respectively. The purified His-tagged recombinant proteins were subjected to 12.5% SDS-PAGE analysis.

(Supplementary Figures, Schemes and Tables)

Figure S1 Degenerate PCR Amplification on *Nocardiopsis sp.* K-252 gDNA



(A) The oligonucleotide primers of P1 and P2 pairs were used to amplify the 480bp nucleotide fragment. The 100bp DNA ladder was used a marker as indicated (M) in figure. (B) The oligonucleotide primers of AG4 and AG5 pairs were utilized to amplify the 300bp nucleotide fragment using the genomic DNA as a template. The agarose gel was indicated as follows: Lane 1: 10% DMSO, Lane 2: without DMSO, Lane 3: zNTP with 10% DMSO, Lane 4: zNTP without DMSO. (zNTP is a mixture of dATP, dCTP, dTTP, dGTP and 7-deazaGTP).

Table S1 Summary[a] of the Discrepancies between *nok* and *ink*[b] Genetic Loci

ORF1~ORF7[c]: The *ink* locus lacks the sequence region (8810 bp) of *nok orfs1~7*. Organization of *orf*s in this region was completely different between *nok* and *ink*, and many of them were different in orientation, function or size.
ORF8: Predicted size of InkT (294 a.a.)[b] was different from that of *nok* ORF8 (277 a.a.).
ORF11 and ORF12: The sequences corresponding to *orf11* and *orf12* were not annotated in *ink*.
ORF14: Predicted size of InkB (304 a.a. in NCBI)[b] was different from that of *nok* ORF14 (NokG, 309 a.a.). The size of InkB was reported to be 309 a.a. in *ink* paper.[b]
ORF15: In *ink* locus, the 226 bp in the 5'-end sequence of *inkA* was a 5'-end fragment of *nok orf17* (*nokP*), where *inkA* coded for glucose-1-phosphate thymidyltransferase and *nokP* coded for cytochrome P450 hydroxylase. Thus, *inkA* is a gene composed of two partial fragments, one from *orf15* (*nokF*) and the other from *orf17* (*nokP*). And, the joint site of the two fragments was found to be a *Sau*3AI in DNA sequence.[d] Predicted size of InkA (369 a.a.)[b] was different from that of *nok* ORF15 (NokF, 354 a.a.).
ORF16 and ORF17: A 2049 bp (*ca.* 2 kb) segment of DNA sequence harboring *nok orf16* (*nokK*) and *orf17* (*nokP*) was not found in *ink* gene cluster sequence and genetic locus (see Fig. 3 of the paper). Both *nokK* and *nokP* were proposed to play important roles in K-252a biosynthesis.
ORF19: Predicted size of InkY (171 a.a., named InkC in Kim's paper)[b] is shorter than that of *nok* ORF19 (NokJ, 398 a.a.). NokJ was proposed to be a key enzyme to catalyze the oxidative *C-N* bond formation, and its homolog (StaN, 394 a.a.) in staurosporine biosynthesis has been functionally characterized *in vivo*.[1]
ORF20: The predicted product of *inkG* is 53 amino acids shorter at its *N*-terminal as compared to *nok orf20* coding for NokL (436 a.a.). The 53-amino acid region in InkG (383 a.a., written 419 a.a. in *ink* paper)[b] contains a conserved His-14, an equivalent of which has been proposed to be a catalytic base in some *O*-Gtfs.
ORF24: The *inkP* gene lacks 807 bp at its 3'-end in sequence, leading to a truncated product (423 bp) fused with a small fragment (105 bp) of miscellaneous sequence. The resulting fused protein of InkP (175 a.a.) was shorter than *nok* ORF24 (NokC, 409 a.a.). Homologs of NokC in rebeccamycin (RebP, 397 a.a) and staurosporine (StaP, 417 a.a.) biosyntheses have been functionally proven to be cytochrome p450 enzymes *in vitro*. NokC highly resembled both RebP and StaP in sequence. A *Sau*3AI was found at the truncation/fusion site (near the 3'-end) of *inkP*, and the remaining *ink* gene cluster sequence (2815 bp out of 20300 bp) downstream of the truncation site (*Sau*3AI in *inkP*), was totally different from *nok*.[d]
ORF25~ORF35: The *ink* locus lacks the DNA sequence (15610 bp) of the region harboring *nok orfs25~35*. Moreover, the arrangement of the *inkU* and *orfs22~32* in *ink* locus, corresponding to the region, was completely different from that of *nok orfs25~35*. Notably, *orf25* (*nokE*), *orf 26* (*nokS*) and *orf27* (*nokT*) in *nok* genetic locus coded for the proteins critical in transcription activation and two-component system regulation of K-252a biosynthesis.

a. see also Fig. 3 (of our paper) for the comparison between *nok* locus and *ink* locus.

b. the protein sizes (a.a.) of many *orfs* in *ink* paper[2] were in disagreement with those submitted to GenBank (DQ399653). In this table, the sizes reported by Kim *et al.* in GenBank were used; this however does not change the general features of the discrepancies if otherwise.

c. the *orf* number designated in *nok* genetic locus, as shown in Table 1 (of our paper).

d. Kim's genomic library was constructed by partial *Sau*3AI digestion of chromosomal DNA (see reference S2).

Table S2 Sequences of Degenerate and Specific Primers for Fosmid Library Screening

Primer Name	Primer sequence (5'→3' direction)	Target Gene Function
AG4 (forward)	RYG TCS GTG ATC TCS AGC TCG CCS CG	NDP-glucose synthase[3]
AG5 (reverse)	GAC TTC RTS ATG TAT CTS GGC GAC AA	NDP-glucose synthase[3]
P1 (forward)	CSG GSG SSG CSG GST TCA TSG G	NDP-glucose 4,6-dehydratase[4]
P2 (reverse)	GGG WRC TGG YRS GGS CCG TAG TTG	NDP-glucose 4,6-dehydratase[4]
syn (forward)	ACT TCG TCA TGT ATC TCG GCG ACA AC	NDP-glucose synthase
syn (reverse)	GTG AAC AGG TAG ACG CCG ACC AGC CG	NDP-glucose synthase
dehy (forward)	GTA GTT GTT GCC GCA CCG GGT GAT C	NDP-glucose 4,6-dehydratase
dehy (reverse)	GAG GTG ACC GTG CTC GAC AAG CTC AC	NDP-glucose 4,6-dehydratase

Table S3 Sequences of Primers used for Gene Cloning by PCR

Primer Name	Primer sequence (5'→3' direction)	Restriction site	Gene
NKONF1 (forward)	CATATGTTCAGTCGTCTACC	*Nde*I	*nokA*
NKONhR1 (reverse)	GCTAGCTCATCGGCCCAGATC	*Nhe*I	*nokA*
NKDNF1 (forward)	CATATGAGCGTCTTCAGCCTGC	*Nde*I	*nokB*
NKDXR1 (reverse)	CTCGAGTCTCACGCTTCCCC	*Xho*I	*nokB*
NKDNhR1 (reverse)	GCTAGCTCATCTCACGCTTC	*Nhe*I	*nokB*
NKPNdF1 (forward)	CATATGGAGCCGTTCTCCCC	*Nde*I	*nokC*
NKPEcR1 (reverse)	GAATTCCCTCGCGGGTGG	*Eco*RI	*nokC*
NKPNhR1 (reverse)	GCTAGCTCACCTCGCGGG	*Nhe*I	*nokC*
NKCNdF1 (forward)	CATATGACTCGCAGCGAAG	*Nde*I	*nokD*
NKCEcR1 (reverse)	GAATTCCCCCGCCCCTCGCAC	*Eco*RI	*nokD*
NKCNhR1 (reverse)	GCTAGCTCACCCCGCCCCTC	*Nhe*I	*nokD*

restriction sites are underlined in sequence.

Figure S2 The 2D-NMR Spectra of Chromopyrrolic Acid (CPA, 3)

Figure S2a The gCOSY Spectrum of CPA (3)

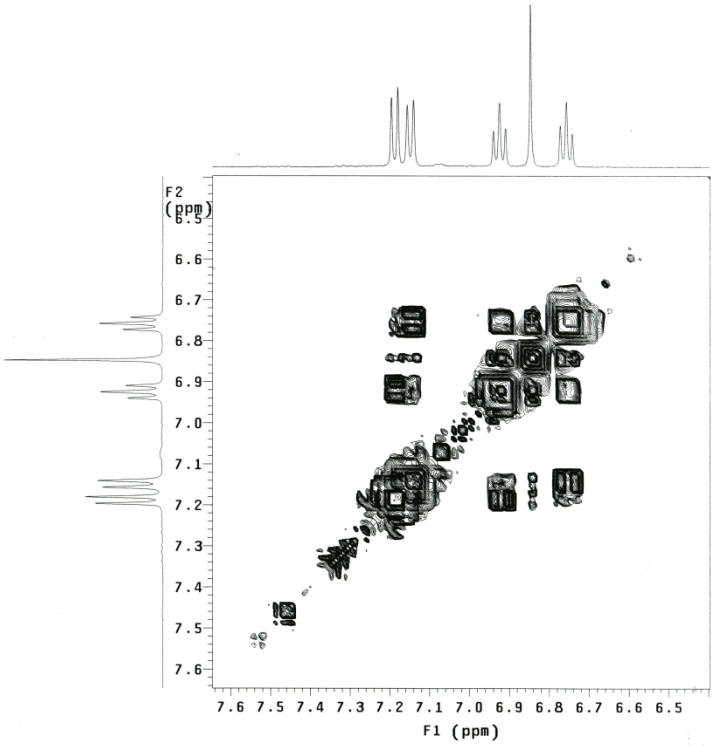
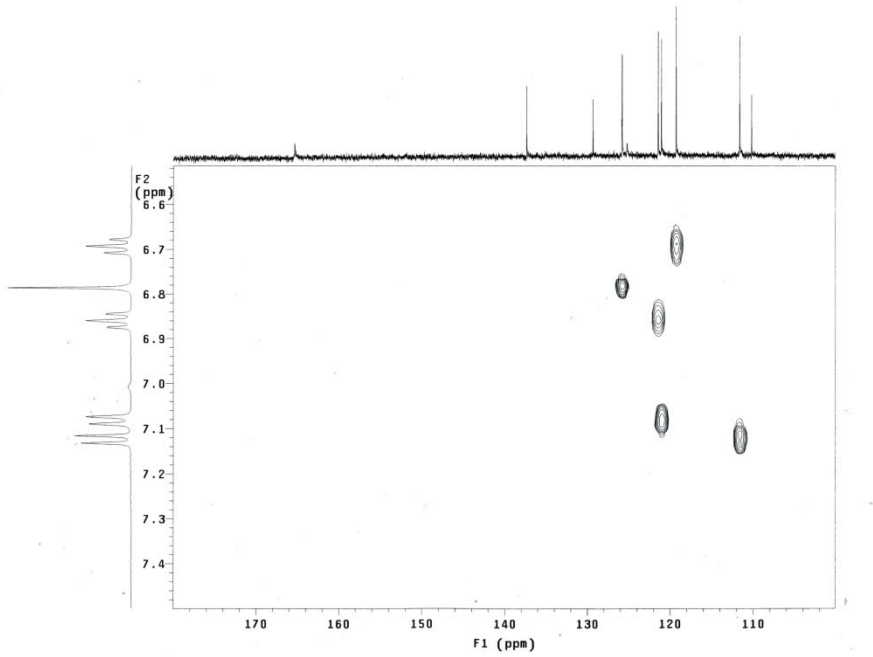


Figure S2b The gHMQC Spectrum of CPA (3)



References (Supplementary Information)

S1. H. Onaka, S. Asamizu, Y. Igarashi, R. Yoshida and T. Furumai, *Biosci. Biotechnol. Biochem.*, 2005, **69**, 1753-1759.

S2. S. Y. Kim, J. S. Park, C. S. Chae, C. G. Hyun, B. W. Choi, J. Shin and K. B. Oh, *Appl. Microbiol. Biotechnol.*, 2007 ,**75**, 1119-1126.

S3. C. Hyun, S. S. Kim, J.K. Sohng, J. Hahn, J. Kim and J. Suh., *FEMS Microbiol. Lett.,* 2000, **183**, 183-189.

S4. H. Decker, S. Gaisser, S. Pelzer, P. Schneider, L. Westrich, W. Wohlleben and A. Bechthold, *FEMS Microbiol. Lett.*, 1996, **141**, 195-201.