Molecular cloning, sequence analysis and functional characterization of the gene cluster for biosynthesis of K-252a and its analogs†

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

Received 19th March 2009, Accepted 24th June 2009
First published as an Advance Article on the web 4th August 2009
DOI: 10.1039/b905293c

Among the indolocarbazole alkaloids of antitumor antibiotics, K-252a represents a structurally unique indolocarbazole glycoside and exhibits potent neuroprotective and broad anticancer activities. K-252a consists of K-252c and the unusual dihydrostreptose moiety, linked together with oxidative and glycosidic C–N bonds. Herein, we reported a complete sequence of an approximately 45 kb genomic fragment harboring the gene cluster for the biosynthesis of indolocarbazole alkaloids in *Nocardiopsis sp*. K-252 (NRRL15532). The sequence of 35 open reading frames discovered several new, critical genes, hence shedding new light on biosynthesis, resistance and regulation of K-252a and its analogs. To functionally characterize the gene cluster *in vitro* and in enzyme level, a multigene expression cassette containing the K-252c biosynthetic genes was constructed and successfully overexpressed in *Escherichia coli* to yield soluble proteins for cell-free tandem enzymatic assays. Consequently, the heterologous expression with soluble NokA and NokB led to *in vitro* production of chromopyrrolic acid (CPA), thereby providing functional evidence for K-252c biosynthesis. Moreover, a facile production of CPA in culture broth was successfully accomplished by using an *in vivo* biotransformation of L-tryptophan with *E. coli* harboring the gene cassette. Importantly, by sequence analysis and the functional characterization here and in the companion paper, biosynthetic pathways leading to formation of K-252a and its analogs were hence proposed. Together, the results provide critical information and materials useful for combinatorial biosynthesis of K-252a and its analogs for therapeutic applications.

Introduction

The family of indolocarbazole natural products has been a valuable source of lead compounds with potential therapeutic applications in the treatment of cancer and neurodegenerative disorders (Fig. 1).[1–5] *Nocardiopsis sp*. K-252 (*Nonomuraea longicatena*, NRRL15532) produces indolocarbazole alkaloids of antitumor antibiotics, among which K-252a represents a structurally unique indolocarbazole glycoside exhibiting neuroprotective activity and displaying potent cytotoxic activities against numerous cancer cells by inhibiting protein kinases.[6–8] Structurally distinct in the family, K-252a is characterized by the dihydrostreptose moiety and two C–N covalent linkages, one of which is generated by a catalytic action of N-glycosyltransferase (N-Gtf) and the other by an oxidative coupling enzyme.[9] To date, the exact biosynthetic pathways and regulations leading to formation of K-252a and its analogs in *Nocardiopsis sp*. K-252 remains still obscure. And the enzymes for the pathways have not been characterized *in vitro* to reveal actual functional roles and substrate specificity. In particular, there is a serious lack of information on the enzymes responsible for the formation of the modified dihydrostreptose moiety of K-252a. Thus far, many of the studies in deoxysugar biosynthesis are limited to the biogenesis of pyranoses.[10] Such information can be of great value for useful applications of the enzymes in expanding the structural diversity of bioactive natural product glycosides by incorporating various special furanoses like the dihydrostreptose. Moreover, some microorganisms, such as *Nocardiopsis* strains, were found to be capable of producing multiple analogs of indolocarbazole compounds,[3,11,12] raising an interesting question with regards to whether or not the microorganisms may, through proper regulation, utilize a single N-Gtf gene or a single set of K-252c biosynthesis genes for their biosynthesis. Resolution of the above interesting issues must rely on sequence information and functional characterization of the genes involved in the biosynthesis and regulation of K-252a and its analogs, which would be very useful for combinatorial biosynthesis of K-252a analogs for therapeutic applications.

In light of the above important facts and prospects, we constructed a fosmid genomic DNA library of *Nocardiopsis sp*. K-252 and therein cloned and identified the gene cluster for the biosynthesis of the indolocarbazole compounds. We now

Department of Biological Science and Technology, National Chiao Tung University, 75 Po-Ai Street, Hsinchu 300, Taiwan. E-mail: Chiu@mail.nctu.edu.tw; Fax: +886-3-5719605; Tel: +886-3-5131595

† Electronic supplementary information (ESI) available: Experimental procedures for cloning, expression and purification of His-tagged proteins of NokABCD; summary of the discrepancies between *nok* and *ink* genetic loci; degenerate PCR results; PCR primers; NMR data. See DOI: 10.1039/b905293c

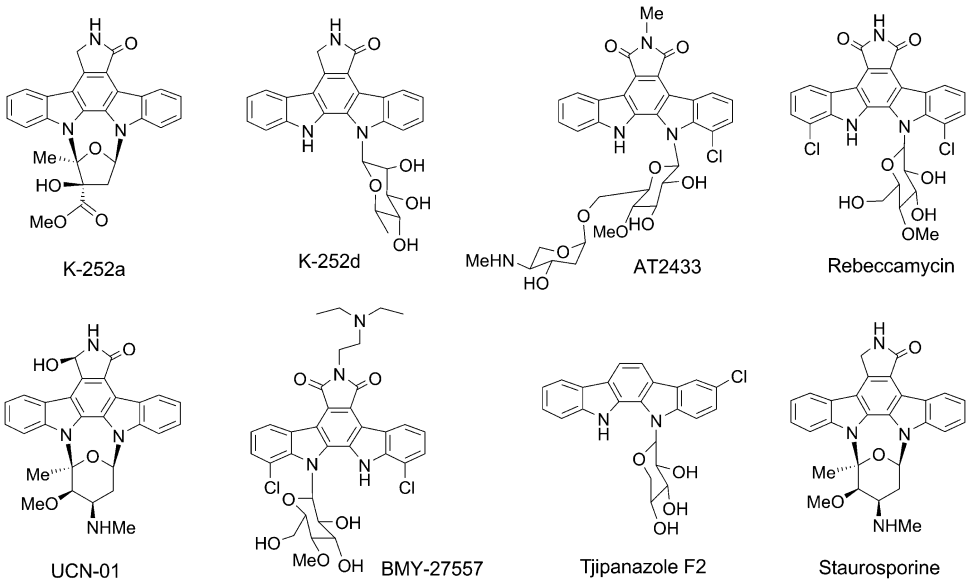


Fig. 1 Representative members (or derivatives) of the indolocarbazole family of natural products.

report the complete 45 kb genomic sequence harboring the gene cluster. Our bioinformatic analysis of the gene cluster has allowed us to functionally deduce the embedded genes and propose biosynthetic pathways leading to formation of the indolocarbazole compounds, especially that of the special dihydrostreptose moiety of K-252a. Most notably, by *in vitro* functional characterization here and in the following paper, we proposed the gene cluster served not only for biosynthesis of K-252a but also for those of its analogs. In addition, construction and overexpression in *E. coli* of the K-252c biosynthetic gene cassette were successfully achieved in this study, whereas the heterologous expression of the gene cassette led to both *in vitro* and *in vivo* functional production of chromopyrrolic acid (CPA) involved in K-252c biosynthesis.

Results and discussion

Construction of the *Nocardiopsis sp.* K-252 fosmid library

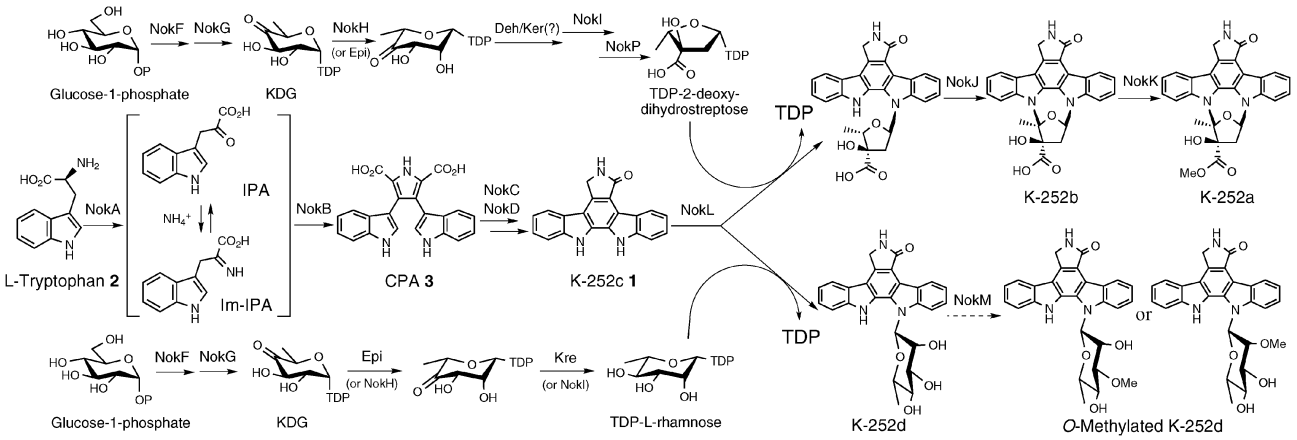
To identify the biosynthetic gene cluster, we constructed a fosmid genomic DNA library of *Nocardiopsis sp.* K-252 by use of a CopyControl fosmid library production kit (Epicentre). Genomic DNA of *Nocardiopsis sp.* K-252 was randomly sheared, size-selected between 30 kb and 70 kb, and then ligated into the *E. coli* fosmid vector pCC1FOS. The resulting ligation products were packaged into λ-phage (Gigapack kit, Stratagene), followed by transfection into *E. coli* EPI300 (Epicentre). As a result, the genomic library was successfully constructed with a total of 5856 fosmid clones, whereas the average sizes of genomic DNA fragments were *ca.* 35 kb per clone.

Cloning of the biosynthetic gene cluster for K-252a and its analogs

As shown in Scheme 1, we proposed the biosynthesis of the modified dihydrostreptose in K-252a utilized the common deoxysugar biosynthetic pathway involving NDP-4-keto-6-deoxy-glucose (KDG) as a key biosynthetic intermediate.[13] In the presence of nucleotide triphosphate, KDG can be synthesized from glucose-1-phosphate by catalytic actions of NDP-glucose synthase and NDP-glucose 4,6-dehydratase.[14,15] Hence, two sets of degenerate primers AG4–AG5 and P1–P2, respectively, were utilized as probes to screen the fosmid library, where the primers were derived from conserved sequences of the homologous genes coding for these two enzymes from various *Streptomyces* species producing glycosylated secondary metabolites.[16,17] Using *Nocardiopsis sp.* K-252 genomic DNA as a template, polymerase chain reaction (PCR) amplifications with the probes yielded DNA fragments of 300 bp and 480 bp (Fig. S1, ESI†), respectively, confirmed to partially code for the enzymes. Designed based on the coding DNA sequences, specific primer pairs were subsequently applied to screen the *Nocardiopsis sp.* K-252 fosmid library by PCR, therefore leading to identification of three overlapping fosmid clones, pJC3B5, pJC40D7 and pJC28B7 (see Materials and methods). Together, the three fosmid clones constituted a genomic DNA fragment of *ca.* 57 kb, from which a 45 kb sequence contig was subsequently obtained by DNA sequencing to cover the entire gene cluster for the biosynthesis of the indolocarbazole compounds, presumably K-252a and its analogs, in *Nocardiopsis sp.* K-252.

Organization of genes in the 45 kb genomic sequence harboring the biosynthetic gene cluster

Sequence analysis of the 45 kb genomic sequence revealed 35 open reading frames (ORFs) as shown in Fig. 2, whereas its detailed information (*e.g.*, annotation and ORF prediction) was deposited in GenBank.[18] In Table 1 are also summarized their deduced functions and related information. As in our 45 kb contig, *nokABCD* genes, encoding enzymatic assembly of K-252c **1** from L-tryptophan **2** (Scheme 1), shared a high degree of similarity to those in *reb*, *sta* and *atm* gene clusters (see also gene cluster comparison as in Fig. 3).[9,19,20] Within the *nok* gene cluster was found *nokJ*, a cytochrome P450 gene,



Scheme 1 The proposed biosynthetic pathway for biosyntheses of indolocarbazole metabolites in *Nocardiopsis sp*. K-252.

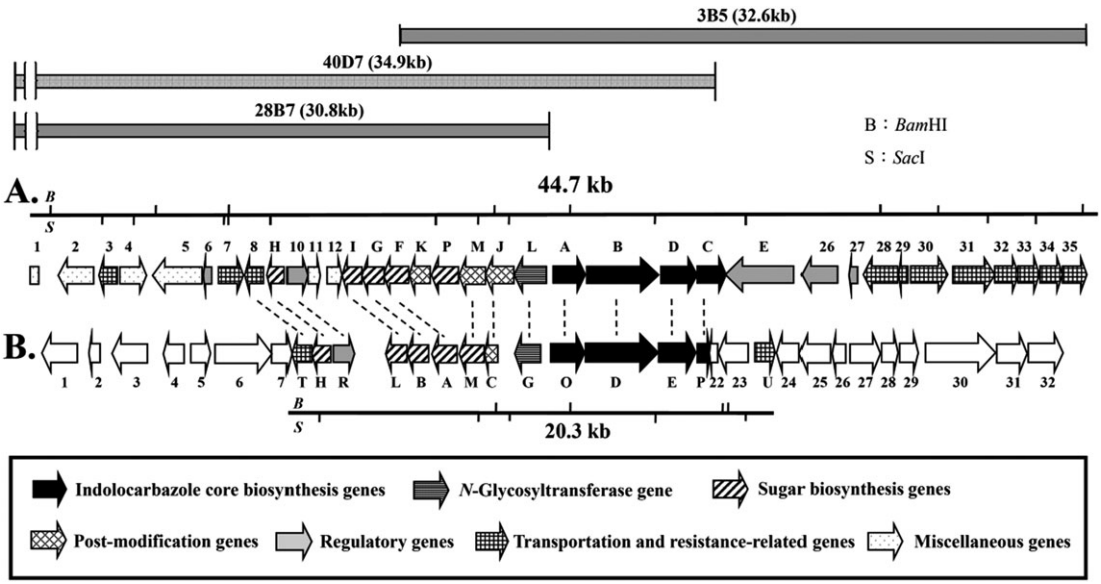


Fig. 2 (A) The genetic map of our 45 kb genomic fragment harboring the *nok* gene cluster for biosyntheses of K-252a and its analogs and (B) comparison with the *ink* genetic locus reported by Kim *et al*.,[35] where only 20.3 kb (as shown in lower restriction map) was actually deposited in GenBank under accession number DQ399653. The dotted lines indicate the matched ORF with the same proposed function. Sketch indicates putative or known functions, and arrows represent the direction of transcription of ORFs.

speculated to be responsible for the oxidative C–N linkage. Like *staN* for staurosporine biosynthesis,[21] the *nokJ* gene resembles in amino acid sequence (45.1% similarity and 30.7% identity) the *eryF* gene responsible for hydroxylation of 6-deoxy-erythronolide B, one of early steps in erythromycin biosynthesis in *Saccharopolyspora erythraea*.[22] Sequence analysis of NokJ revealed the highly conserved heme-binding domain (GXXXCXG), K-helix and oxygen binding region (O_2-BR),[22–24] also preserved in StaN and EryF, as shown in Fig. 4. Very interestingly, NokJ, unlike StaN, shared the same alaine in the O_2–BR of EryF for dioxygen activation. Above findings may suggest NokJ adapt a similar catalytic mechanism as EryF but, however, give a distinct catalytic consequence in form of C–N oxidative coupling.

As shown in Scheme 1, the *nokFGHIP* genes presumably involved in biosynthesis of NDP-2-deoxy-dihydrostreptose (NDP-dStp) were found to be embedded within the cluster. As generally observed in deoxysugar biosynthesis, the 2-deoxygenation required for NDP-dStp biosynthesis may utilize a catalytic action of NDP-sugar 2,3-dehydratase (Deh).[25–27] Interestingly, our experiment showed that the *deh* gene, coding for Deh, was found to be outside of the gene cluster in the genome.[28] The 2-deoxygenation may occur after the 3,5-epimerization presumably catalyzed by NokH.

Most interestingly, in our gene cluster the *nokP* gene was also found to be required to work with *nokI* to complete the biosynthesis of NDP-dStp. The *nokI* gene highly resembled, in amino acid sequence (56% identity and 63% similarity), the dTDP-dihydrostreptose synthase gene (*strL*) in streptomycin biosynthesis.[29] The *nokP* gene coded for a cytochrome p450 hydroxylase/oxidase, highly resembling the homologous enzymes participating in oxidative modification of secondary metabolites, such as leinamycin (*Streptomyces atroolivaceus*), saframycin A (*Streptomyces lavendulae*) and pradimicin

Table 1 Deduced functions for genes in the 45 kb genomic sequence[a] harboring the *nok* gene cluster for biosyntheses of K-252a and its analogs

ORF	Protein	Start/stop	Size (a.a.)	Homolog; accession number	ID/SM (%)[b]	*reb*[c]	*sta*[c]	*atm*[c]
1[d]	NokX	[e]	—	LysR-protein transcriptional regulator; EDX21545	43/58	—	—	—
2[d]		ATG/TGA	504	Amidase; BAC74980	60/68	—	—	—
3[d]		ATG/TGA	282	Integral membrane protein; CAM0193	63/77	—	—	—
4[d]		GTG/TGA	395	Threonine synthase homolog; CAM00339	40/49	—	—	—
5[d]		GTG/TAA	799	ATP-dependent RNA helicase; BAC73942	59/67	—	—	—
6[d]		ATG/TGA	105	Regulatory protein; CAD55454	51/66	—	—	—
7[d]		ATG/TGA	387	Transmembrane efflux protein from the major facilitator superfamily; CAG14958	43/61	*rebT*	—	*atmI*
8	NokN (InkT)[f]	GTG/TGA	277	Putative integral membrane lipid kinase; BAG23064	56/72	—	—	—
9	NokH[g] (InkH)	ATG/TGA	268	Nucleoside-diphosphate-sugar epimerase; ABZ94902	40/57	—	*staE*	—
10	NokO (InkR)	ATG/TGA	300	AraC family transcriptional regulator; BAG17958	64/76	—	—	—
11	NokQ[h]	ATG/TGA	172	Secreted protein; EDX26326	42/53	—	—	—
12	NokR[h]	ATG/TGA	215	Lipoate-protein ligase B; ABD11507	57/68	—	—	—
13	NokI (InkL)	ATG/TGA	282	dTDP-dihydrostreptose synthase; CAA44443	56/63	—	—	—
14	NokG (InkB)	ATG/TGA	309	dTDP-glucose 4,6-dehydratase; BAC55206	71/78	—	*staB*	—
15	NokF (InkA)[h]	GTG/TGA	354	Glucose-1-phosphate thymidyltransferase; BAC55207	72/83	—	*staA*	*atmS7*
16[d]	NokK	ATG/TGA	285	Aklanonic acid methyl transferase; AAF70111	35/47	—	—	—
17[d]	NokP	ATG/TGA	398	Cytochrome P450 hydroxylase; AAN85514	44/58	—	—	—
18	NokM (InkM)	GTG/TGA	377	Sugar O-methyl transferase; CAJ42340	50/63	*rebM*	*staMB*	*atmM*
19	NokJ (InkY)[h]	GTG/TGA	398	Cytochrome P450; BAC55208	52/68	—	—	*staN*
20	NokL (InkG)[h]	TTG/TGA	436	N-glycosyltransferase; BAC55209	57/70	*rebG*	*staG*	*atmG*
21	NokA (InkO)	ATG/TGA	486	L-amino acid oxidase; BAC55210	60/71	*rebO*	*staO*	*atmO*
22	NokB (InkD)	ATG/TGA	1031	Chromopyrrolic acid synthase; BAC55211	53/62	*rebD*	*staD*	*atmD*
23	NokD (InkE)	ATG/TGA	525	Monooxygenase; BAF47693	61/71	*rebC*	*staC*	*atmC*
24[d]	NokC (InkP)[i]	ATG/TGA	409	Cytochrome P-450 RebP; BAC15753	54/64	*rebP*	*staP*	*atmP*
25[d]	NokE	GTG/TAG	962	Transcriptional activator; BAC15755	35/47	*rebR*	*staR*	*atmR*
26[d]	NokS	GTG/TGA	503 or 725	Sensor kinase, two-component system; ABG94428	52/64	—	—	—
27[d]	NokT	GTG/TGA	114	Response regulator receiver protein; ABM12887	56/62	—	—	—
28[d]	NokU	ATG/TAG	510	Amino acid/metabolite permease; CAD55470	45/60	—	—	—
29[d]	NokV	GTG/TAA	102	Putative integral membrane protein; CAJ89547	51/56	—	—	—
30[d]	NokW	ATG/TAA	544	ABC transporter ATP-binding protein; CAB40692	75/86	—	—	—
31[d]		ATG/TGA	601	Putative peptide transport system secreted peptide binding protein; CAM01499	42/59	—	—	—
32[d]		ATG/TGA	348	Putative peptide transport system permease; CAM01500	63/76	—	—	—
33[d]		GTG/TGA	315	Putative peptide transport permease; CAB37469	55/72	—	—	—
34[d]		ATG/TGA	324	ABC transporter ATP-binding protein; CAB37470	69/79	—	—	—
35[d]		ATG/TAA	330	Peptide transport ATP-binding protein; CAB37471	73/80	—	—	—

[a] Totally 35 ORFs were identified with complete sequence available in our study; translated products of some ORFs were named Nok proteins. [b] ID: % identity of amino acid sequences; SM: % similarity of amino acid sequences. The values of identity and similarity were obtained by NCBI BLASTX analysis on protein databases. [c] *reb*, *sta* and *atm* indicate the gene clusters of rebeccamycin, staurosporine and AT2433, respectively. (see refs. 9, 19 and 20). [d] Genes lacking sequence information, not reported or scrambled in *ink* gene cluster. (see Table S1, ESI). [e] Incomplete open reading frame (ORF). [f] ( ) renders the Ink protein reported by Kim *et al.*, see ref. 35. [g] The exact function and possible involvement of NokH in K-252a biosynthesis should be experimentally determined. [h] InkY (171 a.a., InkC in *ink* paper) is 227 a.a. shorter than NokJ; InkG (383 a.a.; 419 a.a. in *ink* paper) is 53 a.a. shorter than NokL; the sequence region harboring *nokQ* and *nokR* was not annotated in *ink*. (see Table S1). [i] *inkA* and *inkP* contain other DNA fragments.

(*Actinomadura hibisca* P157-2).[30–32] NokP may thus satisfy the need for the oxidative maturation of the carboxylate group in the biosynthesis of the modified dihydrostreptose in K-252a. Based on the genetic information, two possible enzymatic pathways, path A and path B, may be proposed to account for the dihydrostreptose biosynthesis, as illustrated in

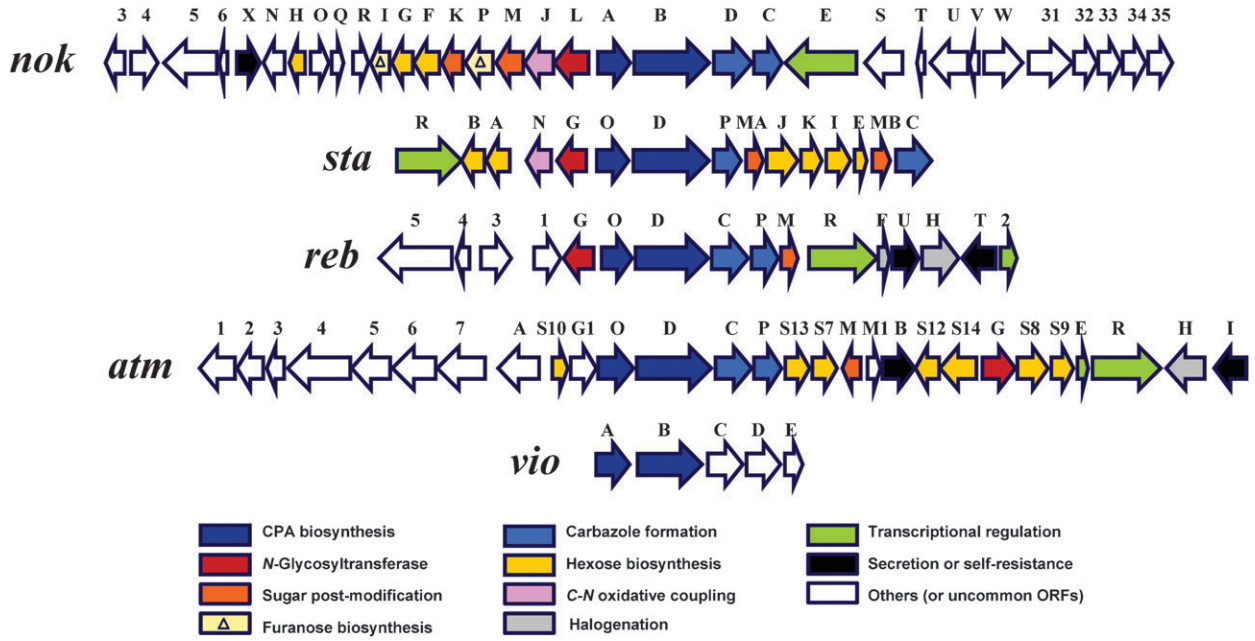


Fig. 3 Comparison of genetic loci responsible for the biosynthesis of K-252a (and its analogs; *nok* genes; accession number FJ031030), staurosporine (*sta* genes; accession number AB088119), rebeccamycin (*reb* genes; accession number AB090952), AT2433 (*atm* genes; accession number DQ297453) and violacein (*vio* genes; accession number AB032799).

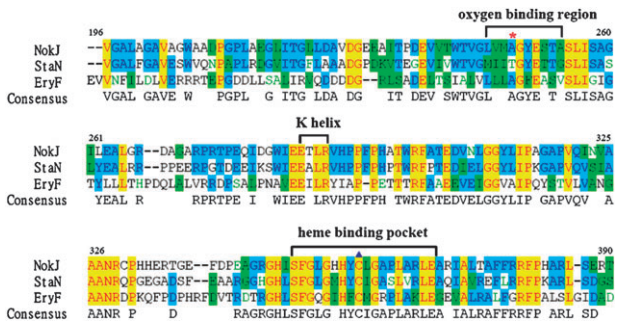
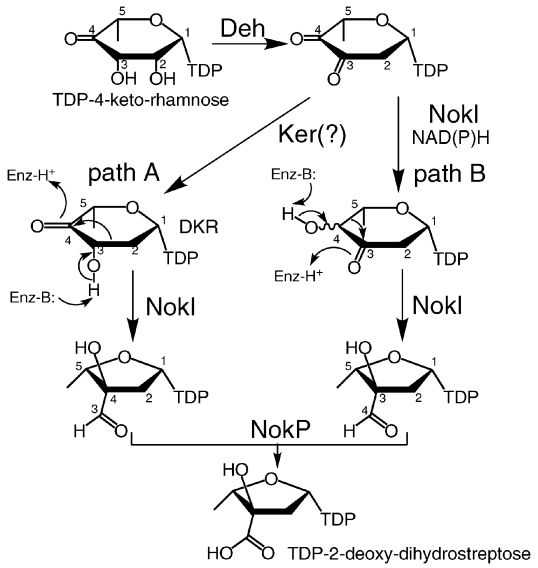


Fig. 4 Sequence alignment of the cytochrome P450 domains of NokJ with StaN and EryF. StaN, cytochrome P450 in staurosporine biosynthesis from *Streptomyces sp.* TP-A0274;[9] EryF, cytochrome P450 in erythromycin biosynthesis from *Saccharopolyspora erythraea*.[22] The amino acids (Ala or Thr) proposed to be involved in the O_2 activation mechanism are marked with an asterisk. The conserved Cys for heme-binding is marked with closed triangle. The indicated region/pocket and helix are the conserved domains in cytochrome P450 enzymes, such as EryF and PicK (picromycin biosynthesis).[22–24]

Scheme 2. In path A, a presumed reductase gene, coding for NDP-sugar 3-keto-reductase (Ker), was proposed to act with *deh* for the formation of NDP-2-deoxy-4-keto-rhamnose (DKR),[25–27,33,34] serving as a precursor for NokI and NokP for final maturation of the activated sugar. Mechanistically, a catalytic base in NokI may abstract the C3–OH proton, followed by 1,2-shift carbon rearrangement between C3 and C4 with concurrent protonation of the C4-keto group to give the aldehyde intermediate. After the ring contraction catalyzed by NokI, the intermediate may then be further oxidized by NokP to give NDP-dStp. In path B, NokI may directly act on the enzymatic product of Deh, and carry out the reduction at C-4 in the presence of NAD(P)H, followed by subsequent ring contraction as indicated. This pathway can be supported by previous study that StrL was proposed to possess both 4-hexulose reductase and ring contraction activities.[36]

Moreover, two ORFs, *nokK* and *nokM*, coding for methyltransferases were both found within the cluster. The *nokK* gene displayed high sequence similarity to *aknG* (46.5% similarity and 33.9% identity)[37] and *chaI* (54.0% similarity and 39.4% identity)[38] proposed to catalyze methyl esterification of the carboxylate group in biosynthesis of aclacinomycin and chartreusin, respectively. As shown in Fig. 5, NokK, ChaI and AknG shared the common motifs involved in binding of S-adenosylmethionine (SAM). However, *nokM* only resembles regular sugar O-methyltransferase genes generally engaged for methylation of sugar hydroxyl groups, *e.g.*, *stfMII* in steffimycin biosynthesis and *spnI*/*spnK* in spinosyn biosynthesis.[34,41] Hence, it should be evident that *nokK*, instead of *nokM*, serves as the best candidate responsible for K-252a methylation. The methyl esterification catalyzed by NokK may be a post-modification step occurring after the N-glycosylation by NokL.

In the gene cluster was found the *nokE* gene coding for ATP-dependent transcriptional activator of LuxR family. NokE shared a high level of sequence similarity (45.6% similarity and 33.8% identity) to RebR in rebeccamycin biosynthesis.[42] The involvement of *nokE* in *nok* gene regulation can be supported by previous experiments that *rebR*-truncated mutant of the rebeccamycin-producer, *Lechevalieria aerocolonigenes*, failed to produce rebeccamycin or related compounds. Adjacent to *nokE* are *nokS* and *nokT*, encoding two-component system sensor kinase and response regulator, respectively. The homolog of highest similarity to



Scheme 2 Proposed enzymatic pathways for biosynthesis of TDP-2-deoxy-dihydrostreptose.

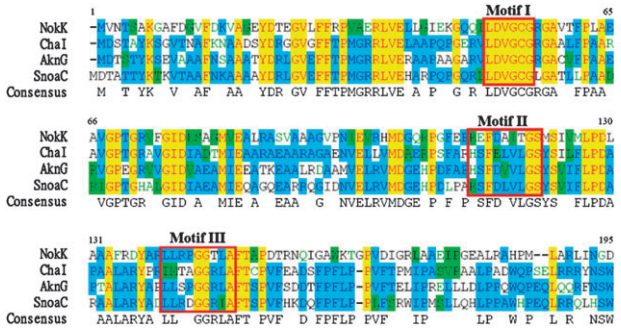


Fig. 5 Alignment of the proposed carboxylate methyltransferase (NokK) of K-252a with similar methyltransferases (MT) encompassing the SAM-dependent MT signature motifs I, II and III.[39,40] ChaI, MT in chartreusin biosynthesis from *Streptomyces chartreusis*; AknG, MT in aclacinomycin biosynthesis from *Streptomyces galilaeus*; SnoaC, MT from *Streptomyces pristinaespiralis*.

nokS was found in *Streptomyces clavuligerus* producing clavulanic acid (57.6% identity)[43] and that to *nokT* in *Mycobacterium vanbaalenii* PYR-1 capable of degrading polycyclic aromatic hydrocarbons (43.8% identity).[44] Downstream of *nokS* and *nokT* were *nokV*, encoding an integral membrane protein, and *nokW*, encoding ABC-transporter ATP-binding protein. NokW displayed very high sequence similarity (>70% identity) to ATP-binding components of various ABC-transporters in many drug-producing actinomycetes, including *Streptomyces coelicolor* A3(2), *Streptomyces griseus* and *Saccharopolyspora erythraea*. Furthermore, on the other side of the *nok* gene cluster was the *nokX* gene, encoding a transmembrane efflux protein from the major facilitator superfamily, similar to the chloramphenicol resistance protein in *Saccharopolyspora erythraea*.

During our work to provide *in vitro* functional evidence for and biochemical characterization of our gene cluster, Kim *et al.* reported an *ink* gene cluster designated solely for K-252a biosynthesis, where a sequence of 20.3 kb was deposited in GenBank.[35] We found, however, many striking discrepancies between the *ink* gene cluster and the *nok* gene cluster in both results and methods (see Fig. 2 and Table 1). Most notably, both of the two key genes (*nokP* and *nokK*) responsible for the formation of the modified dihydrostreptose were absent in the *ink* gene cluster. In addition, in the *ink* cluster the *inkP* gene, corresponding to *nokC*, was found to lack 807 bp at its 3′-end, leading to a truncated gene (423 bp) fused with a small fragment (105 bp) of miscellaneous sequence. The resulting fused protein of InkP (175 a.a.) is shorter than NokC (409 a.a.), whereas NokC in sequence reasonably resembles its homologs, RebP (397 a.a) and StaP (417 a.a.), in rebeccamycin and staurosporine biosyntheses, respectively.[9,19] Importantly, NokC is a cytochrome P450 enzyme that is indispensable for the biosynthesis of K-252a aglycone, K-252c.[9,19,20] Similarly, in the *ink* cluster a 226 bp-segment in 5′-end of *inkA* (corresponding to *nokF*) is composed of other DNA. As compared to the *nok* cluster, several critical genes presumably involved in K-252a regulation and sresistance were absent or in different order in *ink* gene cluster, whereas most of them lacked sequence information. Specifically, downstream of *nokC* there appeared a great discrepancy between *nok* and *ink* in organization and function of genes, as shown in Fig. 2. Within this region of *ca.* 15.6 kb (see *nok orfs* 25∼35) were located the genes coding for a transcriptional activator (NokE), two-component system regulators (NokS and NokT) and transporters (NokV and NokW), presumably critical in the regulation and resistance of K-252a or its analogs. On the other (left) side (*ca.* 8.8 kb) of the gene cluster there was also found a major difference between the two gene clusters (Fig. 2). Notably, both sides (*ca.* 24 kb in total) of the *ink* cluster completely lacked sequence information. For more detailed comparison between *nok* and *ink*, see Table S1 in ESI.†

Comparison with biosynthetic gene clusters of rebeccamycin, staurosporine, AT2433 and violacein

K-252a carries a special furanose moiety cross-bridged with the aglycone (K-252c, **1**) by catalytic actions of the cytochrome p450 enzyme (NokJ) and the N-Gtf (NokL), both of which were also found to be encoded as *staN* and *staG*, respectively, in the *sta* gene cluster for staurosporine biosynthesis.[9] StaG utilizes the same aglycone (K-252c, **1**) as the recipient substrate, but incorporates an amino-pyranose moiety into staurosporine. Similar to the catalytic relationship between NokL and StaG, NokJ and StaN would have evolved, albeit with high sequence homology (51.0% identity and 55.3% identity, respectively), to accept different sugar-attached aglycones for oxidative C–N coupling. Hence, like StaN and StaG, NokJ and NokL act as important catalytic partners, indispensible for the synthesis of the bridged form of indolocarbazole glycosides. In particular, NokJ and NokL may serve as excellent targets and tools for protein engineering in the combinatorial biosynthesis of K-252a analogs. Interestingly, in *nok* and *sta* gene clusters the coding genes for these two enzymes were arranged in the same order and adjacent to those of aglycone biosynthesis (see Fig. 3). In fact, *nokGFJLAB* genes greatly resemble their

homologous genes, *staBANGOD*, in genetic organization (gene order) of the gene clusters.[45] In the senses of genetics and natural product chemistry, it would be reasonable to speculate that the two clusters, *nok* and *sta*, may be closely related in evolutionary relationship, as compared to the only other known gene clusters of indolocarbazole glycosides, *reb* and *atm*, for rebeccamycin and AT2433 biosyntheses, respectively.

Another interesting finding comes from the fact that, among the four genes (*e.g. nokABCD in nok*) required for aglycone biosynthesis, the genes coding for L-amino acid (LAA) oxidase and chromopyrrolic acid (CPA) synthase were found to be intimately coupled in gene order for *nok*, *sta*, *reb* and *atm* clusters (Fig. 3) and shared high sequence homology in amino acid (63.2–68.3% similarity to *nokA* and 58.8–61.8% similarity to *nokB*). Most notably, the *vio* gene cluster bears only two out of the four genes, *i.e. vioA* and *vioB*, indispensible for violacein biosynthesis.[46,47] The finding may be closely associated with the observation, as described in the companion paper,[48] that the enzymatic product of LAA oxidase (NokA) was highly unstable and required CPA synthase (NokB) for immediate, efficient formation of CPA **3**.

The other two genes, *nokD* and *nokC*, highly resemble *staC* (71.0% similarity and 60.8% identity) and *staP* (58.9% similarity and 49.0% identity), respectively, in amino acid sequence, and were proposed to be responsible for the conversion of CPA to K-252c, the common aglycone core of K-252a and staurosporine.[9] Interestingly, the *reb* and *atm* gene clusters also harbor the paired genes (*rebC*/*rebP* and *atmC*/*atmP*, respectively) homologous to *nokD*/*nokC*, whereas both *rebC*/*rebP* and *atmC*/*atmP* were, however, responsible for the turnover of CPA to arcyriaflavin A, the common aglycone core of rebeccamycin and AT2433 (Fig. 1).[19,20] Recent studies showed that the cytochrome p450 enzyme, encoded by *staP*, alone was able to convert CPA to both K-252c and arcyriaflavin A, whereas StaC (FAD monooxygenase) specifically directed the conversion to K-252c only and RebC to arcyriaflavin A.[49] Notably, the *nokD*, *staC*, *rebC* and *atmC* genes shared a similar degree of overall sequence similarity (~70%) and identity (~60%), whereas the *nokC*, *staP*, *rebP* and *atmP* genes also highly resemble each other in amino acid sequence (60~70% similarity and 50~60% identity). Therefore, it would be a particularly intriguing subject of study to resolve the mechanistic difference between the apparently different types (NokD/StaC and RebC/StaC) of FAD monooxygenase. The finding of *nokD*/*nokC* genes from this study may thus provide useful addition to solving the mechanistic puzzle.

Moreover, for all the four gene clusters (*nok*, *sta*, *reb* and *atm*) the transcriptional activator genes (*nokE*, *staR*, *rebR* and *atmR*, respectively) were found to be conserved with good overall sequence homology (41.9–45.6% similarity and 28.9–33.8% identity to *nokE*), indicating the essential role of the genes for precise regulation of the gene clusters. In addition, in the *nok* cluster a AraC family transcriptional regulator gene (*nokO*) was revealed that was similar to the *nanR4* found in the nanchangmycin gene cluster,[50] which however did not show obvious similarity to the genes of other indolocarbazole gene clusters. Nevertheless, the *nokX* gene showed only low sequence homology, in regards to amino acid sequence (33.8% similarity and 18.8% identity), to *rebT* (similar to *atmI*), which has been previously demonstrated to be responsible for rebeccamycin resistance, presumably through the aid of transmembrane electrochemical gradients.[19]

In vitro functional characterization of K-252c biosynthesis

Thus far, the enzymes of K-252a biosynthesis have not been functionally characterized *in vitro* and at the enzyme level. In particular, development of an *in vitro* or *in vivo E. coli* heterologous production system would be valuable, albeit challenging, for the facile and efficient production of K-252a or its biosynthetic intermediates. In light of this prospect and also to functionally characterize our gene cluster for K-252a or its analogs, we proceeded to examine the heterologous expression and functional characterization of the key enzymes involved in K-252c biosynthesis. To achieve the goal, a co-expression plasmid (pCY20) harboring *nokABCD* genes was constructed under the control of a common T7 promoter and a single T7 terminator to investigate the heterologous co-expression of the K-252c biosynthetic genes in *E. coli* (Fig. 6). Hence, we cloned each of *nokABCD* genes by PCR amplifications on the pJC3B5 fosmid as a template. PCR-amplified fragments of *nokA*, *nokB*, *nokC* and *nokD* were then cloned into pET21b expression vectors with *Nde*I and *Nhe*I sites to generate the non-His-tag expression plasmids, pJZ22, pJZ23, pCY10 and pCY5, respectively. Subsequently, the *Xba*I and *EcoR*I digestion fragment (with *nokB*) of pJZ23 was subcloned into pJZ22 (with *nokA*) within *Nhe*I and *EcoR*I sites. Repeated procedures were applied to clone *nokC* and *nokD* sequentially from pCY10 and pCY5, respectively, into the resulting plasmid, thereby generating the *nokABCD* expression cassette, pCY20. In pCY20, the ribosomal binding site of each gene was preserved for expression. This construct will be useful for the additional cloning of more functional genes, *e.g.*, N-glycosyltransferase genes and NDP-sugar biosynthetic genes, into the *Nhe*I and *EcoR*I sites for combinatorial modifications of K-252c, **1**, to generate various glycosylated indolocarbazoles in *E. coli*.

Two approaches could be adapted to investigate the functional expression of the *nokABCD* cassette (pCY20) in the heterologous host *E. coli*, one being the *in vitro* (cell-free) tandem enzymatic reactions of NokABCD and the other being the *in vivo* biotransformation with *E. coli* harboring *nokABCD* genes. Because of the generally high G + C content (> 70%) of actinomycete genes, their heterologous expression (*in vitro*) and subsequent biotransformation (*in vivo*) experiments would be expected to be challenging. To accomplish the functional characterization, we pursued the cell-free tandem enzymatic synthesis of K-252c by *in vitro* functional expression of the *nokABCD* gene cassette in *E. coli* as a first attempt. Hence, the *nokABCD* expression plasmid (pCY20, Fig. 6) was transformed into *E. coli* BL21 (DE3) cells. As a consequence, the co-expression of the four biosynthetic genes resulted, as expected, in insoluble aggregates of NokA, NokB and NokD, and poor expression of NokC, even with lowering the induction temperature to 15 °C and prolonging the IPTG-induction time. To improve the expression, we

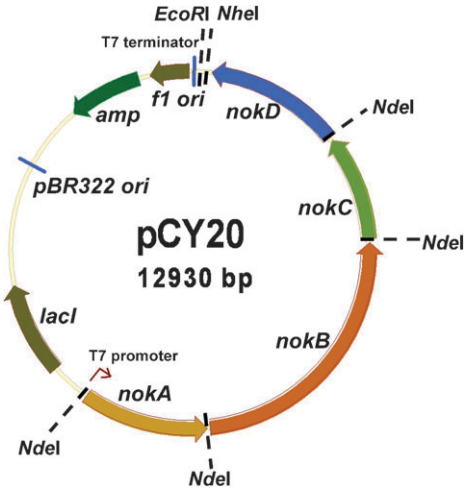


Fig. 6 The expression construct encoding the NokABCD proteins.

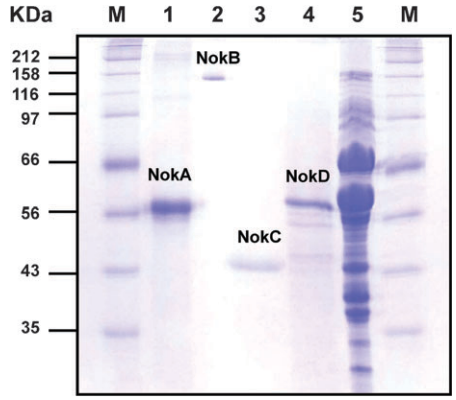


Fig. 7 SDS-PAGE analysis of the nokABCD expression. Purified His-tagged proteins of NokA (1), NokB (2), NokC (3) and NokD (4) from *E. coli* were subjected to SDS-PAGE as reference markers. (M) Molecular weight standards; (5) the cell-free crude extract of the NokABCD proteins.

subsequently carried out the expression in *E. coli* of pCY20 with a combination of chaperones (GroEL, GroES, DnaK, DnaJ, and GrpE) encoded by pG-KJE7.[51] However, the expression of the chaperone proteins troubled the SDS-PAGE analysis of the NokABCD proteins, because the chaperone proteins shared similar molecular weights with NokA and NokD. To accurately assess the optimal co-expression condition for pCY20, we hence cloned, expressed and purified each of the His-tagged proteins of NokA (54.7 kDa), NokB (112.0 kDa), NokC (45.4 kDa) and NokD (57.8 kDa), and therein screened for a common, suitable expression condition under co-expression with the chaperone teams encoded by pG-KJE7 (see ESI†). The best, compromised condition for co-expression of the NokABCD proteins in the presence of the chaperones was finally set to be 250 μM IPTG, 0.2% (w/v) L-arabinose, 25 °C and 10 h for induction, leading to a sufficient quantity of soluble protein for each of the NokABCD proteins as analyzed in Fig. 7. Therefore, the cell-free crude extract of the NokABCD proteins was further utilized for enzymatic assays.

Subsequently, the cell-free crude extract of NokABCD was incubated with the presumed substrate, L-tryptophan (L-Trp, **2**). As a result, RP-HPLC analysis of the reaction mixture revealed the formation of a new, major product (retention time 35.4 min), as shown in Fig. 8 (see profiles C and D). The new product was subsequently prepared at a large scale, where the ethyl acetate extract of the reaction mixture was subjected to purification by semi-preparative RP-HPLC. Extensive NMR analyses by ^{1}H-NMR (Fig. 9), ^{13}C-NMR and 2D-NMR (gCOSY and gHMQC) suggested the product to be chromopyrrolic acid (CPA, **3**), the proposed enzyme product of NokB. Furthermore, high resolution ESI-MS analysis of the product revealed $[M + H]^+$ at *m/z* 386.232 and $[M + Na]^+$ at *m/z* 408.237 (Fig. 9), thereby confirming the product to be CPA ($C_{22}H_{15}N_3O_4$ $[M + H]^+$, calcd. M.W. 386.114; $C_{22}H_{15}N_3O_4$ $[M + Na]^+$, calcd. M.W. 408.097). The observation may suggest that, despite producing sufficient soluble proteins of NokC and NokD, the cell-free extract of the *E. coli* expression system may not sufficiently support the catalytic action of NokC or NokD. This proposition was further supported by a recent observation by Howard-Jones and Walsh that in the staurosporine biosynthesis of StaP, a NokC homolog, the use of CPA as a substrate and exogenous addition of ferredoxin, flavodoxin $NADP^+$-reductase and NAD(P)H was required to display StaP activity.[49] Nevertheless, our experiments with the *nokABCD* expression cassette have clearly demonstrated the proposed functions and *in vitro* (cell-free) enzymatic activities of NokA and NokB involved in the biosynthetic pathway of K-252a and its analogs. And, the cell-free heterologous production of K-252c from *E. coli* could be possible, provided that the genes coding for the electron-transport components to support the NokC activity can be incorporated into the expression cassette.

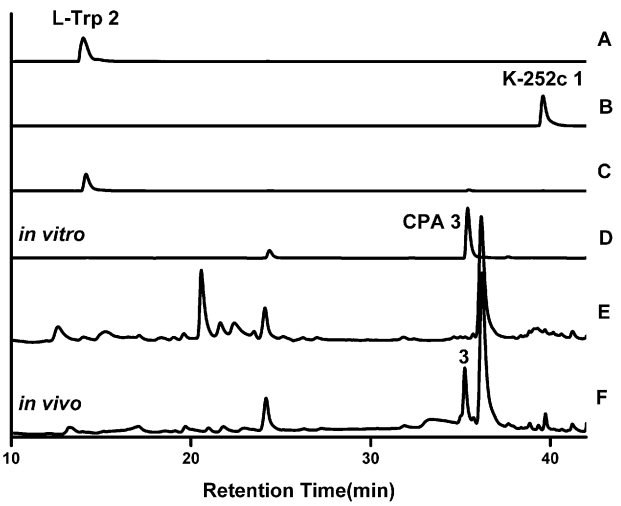


Fig. 8 RP-HPLC analysis of the *in vivo* and *in vitro* assays with NokABCD enzymes and L-tryptophan (L-Trp). The RP-HPLC analysis was monitored at 300 nm. profile A: L-Trp standard; profile B: K-252c standard; profile C: the *in vitro* NokABCD reaction quenched at 0 h; profile D: the *in vitro* NokABCD reaction quenched at 24 h; profile E: the *in vivo* biotransformation control (60 h) with *E. coli*/pET21b; profile F: the *in vivo* biotransformation experiment (60 h) with *E. coli*/pCY20&pG-KJE7. The *in vitro* reaction was conducted with the NokABCD cell-free crude extract, whereas the *in vivo* biotransformation with the *E. coli* cell culture fed with L-Trp.

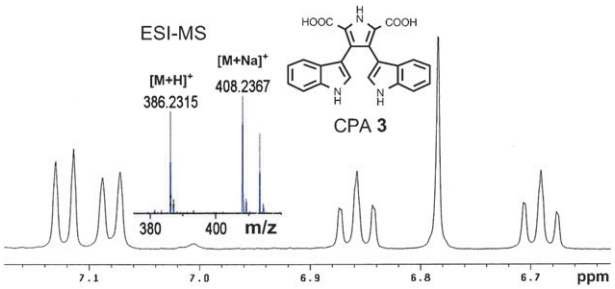


Fig. 9 ^{1}H-NMR and high resolution ESI-MS of the chromopyrrolic acid (CPA) from the NokABCD reaction.

In vivo isolation and production of chromopyrrlic acid from *E. coli*

In light of successful cell-free enzymatic assays of NokABCD, we were prompted to investigate the possibility of utilizing the *E. coli* cells to produce CPA *in vivo*. Subsequently, the *E. coli* cells haboring pCY20 and pG-KJE7 were fed with L-Trp (0.4 g per litre of culture) when induced with 250 μM IPTG and 0.1% (w/v) L-arabinose. After induction and incubation at 25 °C for 60 h, the ethyl acetate extract of the culture broth was subjected to RP-HPLC analysis. Consequently, CPA **3** was detected as confirmed by co-elution with a CPA authentic standard (see profiles E and F in Fig. 8). Control experiments (without the *nokABCD* genes), however, did not give CPA, indicating that CPA was the direct product of *in vivo* NokABCD reactions. In addition, RP-HPLC analysis of cell-free crude extract of the *E. coli* cells from the biotransformation experiment found only a residual amount of CPA (data not shown), suggesting that the majority of CPA had been transported out the cells. The biotransformation experiment successfully demonstrated that the *E. coli* was capable of uptaking L-Trp and transporting CPA out as illustrated in Fig. 10. The *in vivo* biotransformation system may thus provide a useful platform for applications in facile production of CPA or related indolocarbazole compounds.

Multi-functional roles of our gene cluster in *Nocardiopsis sp.* K-252

Our K-252a gene cluster revealed two key genes, *nokP* (encoding hydroxylase/oxidase) and *nokK* (encoding carboxylate methyltransferase) that are indispensible for biosynthesis of K-252a, especially the dihydrostreptose moiety, in *Nocardiopsis sp.* K-252. This observation clearly suggests that the *nok* gene cluster is responsible for K-252a biosynthesis. Nonetheless, our companion paper revealed another important finding that NokL was capable of utilizing TDP-Rha as an alternative substrate to produce K-252d.[48] Moreover, two distinct methyltransferase genes (*nokK* and *nokM*) were disclosed within the *nok* gene cluster in this study. It should be noted that Yasuzawa *et al.* reported production of K-252a, K-252b, K-252c and K-252d in *Nocardiopsis* strains (Scheme 1).[12] Together, these important findings led to a speculation that the *nok* gene cluster may encode not only for biosynthesis of K-252a but also for those of other indolocarbazole metabolites in this strain.

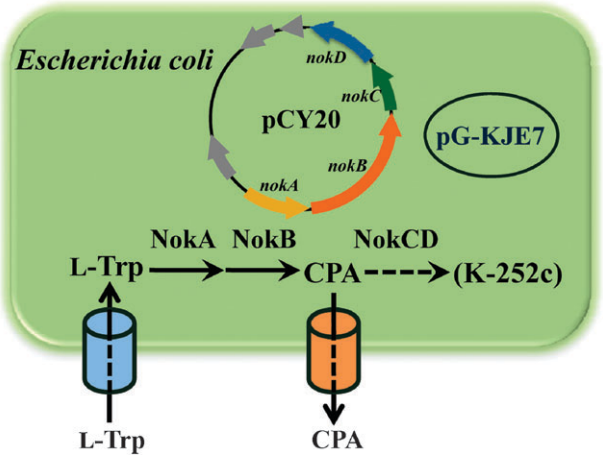


Fig. 10 Graphic representation of the *in vivo* biotransformation of L-tryptophan using *E. coli* harboring pCY20 and pG-KJE7.

Indeed, from our fosmid library clones we also identified genes, located elsewhere in the genome, coding for NDP-hexose 3,5-epimerase (Epi) and putative NDP-4-keto-rhamnose reductase (Kre) (data not shown), which may in principle work with *nokF* and *nokG* of the gene cluster (or their homologous genes in the genome) to accomplish the biosynthesis of TDP-Rha in the strain (Scheme 1). On the other hand, another methyltransferase encoded by *nokM* may possibly be used for the biosynthesis of, yet unidentified, O-methylated K-252d (most likely 2′- or 3′-*O*-methyl-K-252d) in the same strain, as *nokM* was found to share higher sequence similarity with *spnI* (a 2′- or 3′-*O*-methyltransferase gene) than *spnK* and *spnH* in methylation of spinosyn rhamnose.[34,52] Based on our findings here and in the companion paper, we may thus postulate a more complete biosynthetic pathway, as illustrated in Scheme 1, to account for biosyntheses of indolocarbazole metabolites in *Nocardiopsis sp.* K-252. NokK can therefore carry out carboxylate methylation of K-252b to accomplish K-252a biosynthesis, whereas the biosynthesis of K-252d can also be satisfied by the substrate promiscuity of NokL revealed in the following paper.

Materials and methods

Bacterial strains and culture conditions

The *Nocardiopsis sp.* K-252 (NRRL15532, *Nonomuraea longicatena* K-252T) was obtained as a freeze-dried pellet from Agricultural Research Service Culture Collection (USA). The *Nocardiopsis sp.* K-252 was cultivated at 28 °C, 250 rpm in ISP medium 2 (yeast extract 0.4%, malt extract 1%, and glucose 0.4%, pH at 7.3)[53] for approx. 6 to 7 days in baffled flasks for isolation of genomic DNA. The TransforMax EPI300 *Escherichia coli* (Epicentre) were used to construct genomic DNA library. The *E. coli* XL1-Blue and *E. coli* BL21-Codon Plus (DE3)-RP (Stratagene) served as hosts for routine subcloning and protein expression under standard culture conditions as described by Sambrook *et al.*[54]

Plasmids and DNA manipulations

The genomic DNA of *Nocardiopsis sp.* K-252 was extracted by Qiagen Genomic-tip system (Qiagen). The pCC1FOS vector

(Epicentre) was employed for fosmid library construction. The pUC19 (NEB) and pBluescript KS$^+$ (Stratagene) were used to routinely clone restriction fragments for sequencing, whereas the pET21b and pET28a (Novagen) were used to overexpress genes. Restriction endonucleases and T4 DNA ligase were purchased from New England Biolabs (NEB). The *pfu* DNA polymerase (Stratagene) was routinely utilized in polymerase chain reaction (PCR). The mini-preparation of DNA was performed on QIAprep spin miniprep kit (Qiagen). Unless specified otherwise, all chemicals were purchased from Sigma. Sequences of PCR primers are listed in Tables S2 and S3 (ESI†).

Probe design and PCR-screening for identification of the biosynthetic gene cluster for K-252a and its analogs

The degenerate oligonucleotide primers (Operon Biotechnologies) were designed based on conserved regions of homologous genes of NDP-glucose synthase and NDP-glucose 4,6-dehydratase to give expected PCR-amplified fragments of *ca.* 300 bp and 480 bp, respectively. For NDP-glucose synthase, primers were AG4 (forward) and AG5 (reverse); for NDP-glucose 4,6-dehydratase, primers were P1 (forward) and P2 (reverse). The PCR-amplified DNA fragments, as shown in Fig. S1 (ESI†), were verified by sequencing, and then specific primers, dehy and syn, were designed based upon the resulting DNA sequences. Subsequently, the primer pairs (dehy and syn, Table S2†) were utilized as DNA probes to screen the fosmid library for identification of the biosynthetic gene cluster. The PCR (PTC-200 DNA Engine, MJ Research, USA) conditions were as follows: an initial denaturation (5 min at 95 °C); 30 cycles of 1 min at 95 °C, 1 min at 67 °C, and 1 min at 75 °C; and a final extension step of 10 min at 75 °C. Consequently, a PCR-positive fosmid clone, pJC3B5, was found. Furthermore, specific primers designed based upon 3′ and 5′ ends of pJC3B5 insert sequences were utilized for subsequent screening, leading to identification of two other fosmid clones, pJC28B7 and pJC40D7.

DNA sequencing and ORF analysis

The three overlapping fosmid clones, pJC3B5, pJC40D7 and pJC28B7, gave a DNA contig of *ca.* 57 kb in total. The *Bam*HI and *Sac*I restriction fragments of the fosmid clones with approx. 0.4 to 3 kb were subcloned into the cloning vectors for sequencing and assembling of a full-length fosmid contig spanning *ca.* 45 kb as shown in Fig. 3. ORF analysis was performed on Vector NTI program (InforMax), and the computer-aided database searching and sequence analysis were carried out using the BLAST server from the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/BLAST).

Nucleotide sequence accession number

The complete sequence of our 45 kb genomic fragment harboring the entire biosynthetic gene cluster for K-252a and its analogs has been deposited in the GenBank under accession number FJ031030.

Construction of NokA, NokB, NokC, NokD and NokABCD expression plasmids

The K-252c biosynthetic genes, *nokA*, *nokB*, *nokC* and *nokD*, were amplified by PCR on pJC3B5 with the corresponding primers (Table S3) carrying *Nde*I and *Nhe*I restriction sites at the 5′ and 3′ ends, respectively. Each of the PCR-amplified gene fragments preserving the stop codon was individually cloned into the *Nde*I and *Nhe*I sites of pET21b to generate the resulting plasmids, pJZ22, pJZ23, pCY10 and pCY5, harboring *nokA*, *nokB*, *nokC* and *nokD*, respectively. To generate the expression plasmid encoding the biosynthesis of K-252c, the coding genes (*nokA*, *nokB*, *nokC* and *nokD*) were assembled in the same expression construct under the control of the same T7 promoter and a single T7 terminator. To do so, pJZ22 haboring *nokA* served as a starting construct. After digestion of pJZ22 with *Nhe*I and *Eco*RI, the *Xba*I and *Eco*RI digestion fragment containing *nokB* from pJZ23 was ligated with the digested pJZ22 to generate pJZ25. Subsequently, *nokC* and *nokD* gene fragments flanked by *Xba*I and *Eco*RI from pCY10 and pCY5, respectively, were sequentially cloned into pJZ25 by the same method described above to generate the resulting plasmid pCY20.

Heterologous co-expression of NokABCD in *E. coli*

The expression construct, pCY20, was transformed into *E. coli* BL21 (DE3) for protein overexpression. For co-expression with the chaperone teams, pG-KJE7 was co-transformed into the same *E coli* host. Transformants were cultured at 37 °C in Luria-Bertani (LB) medium with 100 μg ml^{-1} ampicillin and 30 μg ml^{-1} kanamycin until an OD_{600} of 0.4–0.5 was reached. Subsequently, the culture was induced with 250 μM isopropyl β-D-1-thiogalacto-pyranoside (IPTG) and 0.2% (w/v) L-arabinose at 25 °C for 10 h. Cells were then harvested by centrifugation at 4 °C (1902 g, 20 min) and subsequently disrupted by sonication at 4 °C in Tris buffer (104 mM Tris-HCl, 10% (v/v) glycerol, pH 7.6). After centrifugation (15 700 g, 20 min) at 4 °C to remove cell debris, the cell-free crude extract was obtained for further experiments.

RP-HPLC analyses of *in vitro* tandem enzymatic reactions of NokABCD

The *in vitro* tandem NokABCD reactions were examined for production of possible biosynthetic intermediates or products. The reaction mixture in a total volume of 52 μl, containing 2 mM substrate (L-tryptophan, L-Trp **2**) and 40 μl NokABCD cell-free crude extract in a reaction buffer (80 mM Tris-HCl, pH 7.8, 7.6% (v/v) glycerol), was incubated at 30 °C for 24 h. The reaction was then terminated by an equal volume of ice-cold MeOH. The reaction mixture was then analyzed by an ODS-C18 RP-HPLC analytical column (4.6 × 250 mm, 5 μm, Zorbax, Agilent). HPLC analyses were performed by Agilent 1100 series equipped with quaternary pump and diode-array detector. The mobile phase for the HPLC elution consisted of methanol (solvent A), 2.5 mM aqueous potassium phosphate at pH 3.5 (solvent B) and acetonitrile (solvent C). The reaction mixture was subjected to RP-HPLC analysis at a flow rate of 1.0 ml min^{-1} with an elution gradient consisting of 8% solvent A, as programmed as follows: 92% solvent B in

5 min, 0–16% solvent C over 10 min, 16–28% solvent C over 15 min, 28–70% solvent C over 15 min, 70–80% solvent C over 2 min, 80% solvent C held for 5 min, and then 80% to 0% solvent C over 3 min. Finally, the column was equilibrated with 92% solvent B for 5 min. The elution was monitored with a full-range of UV-vis wavelengths to trace all possible intermediates or products.

Preparation and characterization of chromopyrrolic acid

To prepare chromopyrrolic acid (CPA **3**) from the NokA/NokB reactions, a reaction mixture (104 μl) containing 4 mM L-Trp **2**, 30 mM ammonium hydroxide, and 80 μl cell-free crude extract of NokABCD in a reaction buffer (80 mM Tris-HCl, 7.6% (v/v) glycerol, pH 7.8) was incubated at 30 °C for 24 h. The reaction was quenched by an equal volume of ice-cold MeOH. After centrifugation with microcon YM-10 (Millipore), the eluted mixture was adjusted to pH 3.0 and extracted with ethyl acetate (EA), subsequently yielding crude CPA after removing EA. The crude CPA was further purified by a semi-preparative C18 RP-HPLC column (5C18-AR-I, 8.0 × 250 mm, 5 μm, Cosmosil, Nacalai tesque, Japan) eluted with the mobile phase consisting of methanol (solvent A) and 2.5 mM ammonium formate (solvent B). The crude product, prepared from 13.5 ml of reaction, was purified by using a HPLC gradient (flow rate at 3.0 ml min^{-1}) programmed as follows: 15% solvent A in 5 min, 15–53% solvent A over 20 min, 53% solvent A held for 5 min, 53–80% solvent A over 3 min, 80% to 15% solvent A over 3 min, and then 15% solvent A for 5 min. The eluted product was neutralized to approximately pH 7.0 by ammonium hydroxide, and then lyophilized to yield a white powder of CPA **3** (2.6 mg, *ca.* 99% purity as estimated by RP-HPLC analysis). The structural elucidation of the product was performed with NMR (Varian INOVA-500) and ESI-MS. ^{1}H-NMR (CD_3OD, 500 MHz), δ_H 6.691 (2H, dd, J = 7.0 Hz), 6.785 (2H, s), 6.858 (2H, dd, J = 7.0 Hz), 7.080 (2H, d, J = 8.0 Hz), and 7.122 (2H, d, J = 8.0 Hz) ppm. ^{13}C-NMR (CD_3OD, 125 MHz), δ_C 110.236, 111.671, 119.328, 121.107, 121.504, 125.252, 125.847, 125.885, 129.374, 137.397, and 165.358 ppm. High resolution ESI-MS calculated for $C_{22}H_{15}N_3O_4$ $[M + H]^+$ 386.114; found 386.232; calculated for $C_{22}H_{15}N_3O_4$ $[M + Na]^+$ 408.097; found 408.237. The gCOSY and gHMQC 2D-NMR results are shown in Fig. S2 (ESI†).

In vivo biotransformation experiments

E. coli BL21 (DE3) co-transformants of pCY20 and pG-KJE7 were cultured and induced similarly as described. For *in vivo* biotransformation, 0.4 g of L-Trp was added into 1 l culture of the *E. coli* upon induction, followed by additional growth at 25 °C for 60 h. The culture was then added with hydrochloric acid to adjust the pH to 3.5. After centrifugation (1902 g, 15 min), the culture broth was collected and extracted with EA. Upon evaporation of EA, the organic extract was subjected to RP-HPLC analysis for production of CPA as described. On the other hand, the cell pellet collected from centrifugation was washed with Tris buffer (20 mM Tris-HCl, pH 7.6) and disrupted by French Press (two passages at 16 000 psi, Spectronic Instruments). After centrifugation (15 700 g, 1 h), the cell debris was removed and the supernatant (soluble fraction) was examined for CPA production as described for the culture broth.

Conclusions

K-252a and its derivatives are promising neuroprotective and anticancer agents. In this study, we reported a complete sequence of a 45 kb *Nocardiopsis sp*. K-252 genomic fragment harboring the gene cluster for the biosynthesis of indolocarbazole metabolites in *Nocardiopsis sp*. K-252. The sequence containing 35 ORFs revealed several new, critical genes, *e.g*., *nokK*, *nokP* and *nokEST* genes, thus shedding new light on the biosynthesis, resistance and regulation of K-252a and its analogs. In addition, the gene cluster was *in vitro* functionally characterized, for the first time, in enzyme level. By constructing a heterologous co-expression system of *E. coli*, soluble proteins of NokABCD enzymes were successfully obtained for *in vitro* functional characterization, subsequently leading to production of chromopyrrolic acid (CPA) supporting K-252c biosynthesis. Furthermore, an *in vivo* biotransformation system with *E. coli* harboring *nokABCD* genes was established and readily afforded CPA in culture broth. Together with our results from the following paper, the findings of this study suggest that the gene cluster could be multifunctional to serve not only for biosynthesis of K-252a but also for those of other indolocarbazole metabolites (including K-252b, K-252c, K-252d and, yet unidentified, methylated K-252d) in *Nocardiopsis sp*. K-252, thus providing new insights into biosynthesis of the indolocarbazole family of antitumor antibiotics.

Acknowledgements

This research was supported by National Science Council (Taiwan) grants (NSC-91-2113-M-009-013 and NSC-92-2113-M-009-025) to Hsien-Tai Chiu. We thank HSP Research Institute (Kyoto Research Park, Japan) for kindly providing pG-KJE7 as a gift.

References

1 G. W. Gribble and S. J. Berthel, in *Studies in Natural Products Chemistry*, Elsevier, Amsterdam, 1993, pp. 365–409.
2 S. Akinaga, K. Sugiyama and T. Akiyama, *Anti-Cancer Drug Des.*, 2000, **15**, 43–52.
3 M. Prudhomme, *Eur. J. Med. Chem.*, 2003, **38**, 123–140.
4 H. Nakano and S. Omura, *J. Antibiot.*, 2009, **62**, 17–26.
5 M. Prudhomme, *Curr. Med. Chem.: Anti-Cancer Agents*, 2004, **4**, 509–521.
6 H. Kase, K. Iwahashi and Y. Matsuda, *J. Antibiot.*, 1986, **39**, 1059–1065.
7 A. H. Ross, C. A. McKinnon, M. C. Daou, K. Ratliff and D. E. Wolf, *J. Neurochem.*, 1995, **65**, 2748–2756.
8 M. M. Berg, D. W. SternbergQll, L. F. Paradall and M. V. Chao, *J. Biol. Chem.*, 1992, **267**, 13–16.
9 H. Onaka, S. Taniguchi, Y. Igarashi and T. Furumai, *J. Antibiot.*, 2002, **55**, 1063–1071.
10 C. Rupprath, T. Schumacher and L. Elling, *Curr. Med. Chem.*, 2005, **12**, 1637–1675.
11 S. Nakanishi, Y. Matsuda, K. Iwahashi and H. Kase, *J. Antibiot.*, 1986, **39**, 1066–1071.
12 T. Yasuzawa, T. Iida, M. Yoshida, N. Hirayama, M. Takahashi, K. Shirahata and H. Sano, *J. Antibiot.*, 1986, **39**, 1072–1078.

13 H. P. Wahl, U. Matern and H. Grisebach, *Biochem. Biophys. Res. Commun.*, 1975, **64**, 1041–1045.
14 J. Oh, S. G. Lee, B. G. Kim, J. K. Sohng, K. Liou and H. C. Lee, *Biotechnol. Bioeng.*, 2003, **84**, 452–458.
15 Y.-L. Chen, Y.-H. Chen, Y.-C. Lin, K.-C. Tsai and H.-T. Chiu, *J. Biol. Chem.*, 2009, **284**, 7352–7363.
16 C. Hyun, S. S. Kim, J. K. Sohng, J. Hahn, J. Kim and J. Suh, *FEMS Microbiol. Lett.*, 2000, **183**, 183–189.
17 H. Decker, S. Gaisser, S. Pelzer, P. Schneider, L. Westrich, W. Wohlleben and A. Bechthold, *FEMS Microbiol. Lett.*, 1996, **141**, 195–201.
18 Our 45 kb genomic sequence harboring the K-252a gene cluster has been deposited in GenBank under accession number: FJ031030.
19 C. Sánchez, I. A. Butovich, A. F. Braña, J. Rohr, C. Méndez and J. A. Salas, *Chem. Biol.*, 2002, **9**, 519–531.
20 Q. Gao, C. Zhang, S. Blanchard and J. S. Thorson, *Chem. Biol.*, 2006, **13**, 733–743.
21 H. Onaka, S. Asamizu, Y. Igarashi, R. Yoshida and T. Furumai, *Biosci., Biotechnol., Biochem.*, 2005, **69**, 1753–1759.
22 S. Nagano, J. R. Cupp-Vickery and T. L. Poulos, *J. Biol. Chem.*, 2005, **280**, 22102–22107.
23 M. C. Betlach, J. T. Kealey, G. W. Ashley and R. McDaniel, *Biochemistry*, 1998, **37**, 14937–14942.
24 N. Parajuli, D. B. Basnet, H. Chan Lee, J. K. Sohng and K. Liou, *Arch. Biochem. Biophys.*, 2004, **425**, 233–241.
25 G. Draeger, S.-H. Park and H. G. Floss, *J. Am. Chem. Soc.*, 1999, **121**, 2611–2612.
26 X. He and H. W. Liu, *Curr. Opin. Chem. Biol.*, 2002, **6**, 590–597.
27 X. M. He and H. W. Liu, *Annu. Rev. Biochem.*, 2002, **71**, 701–754.
28 The *deh* gene was identified by PCR amplification using degenerate oligoprimers designed upon conserved regions of NDP-sugar 2,3-dehydratase homologs from various *Streptomyces* species (unpublished data). In 2,6-dideoxy-sugar biosynthesis, *deh* has often been found to couple with *ker* to achieve the 2-deoxygenation (ref. 25–27 and 33). Locating *ker* elsewhere in the genome of this strain is in progress. (NDP: nucleotide diphosphate).
29 K. Pissowotzki, K. Mansouri and W. Piepersberg, *Mol. Gen. Genet.*, 1991, **231**, 113–123.
30 G. L. Tang, Y. Q. Cheng and B. Shen, *Chem. Biol.*, 2004, **11**, 33–45.
31 L. Li, W. Deng, J. Song, W. Ding, Q. F. Zhao, C. Peng, W. W. Song, G. L. Tang and W. Liu, *J. Bacteriol.*, 2008, **190**, 251–263.
32 B. C. Kim, J. M. Lee, J. S. Ahn and B. S. Kim, *J. Microbiol. Biotechnol.*, 2007, **17**, 830–839.
33 A. Trefzer, J. A. Salas and A. Bechthold, *Nat. Prod. Rep.*, 1999, **16**, 283–299.
34 C. Waldron, P. Matsushima, P. R. Rosteck, Jr, M. C. Broughton, J. Turner, K. Madduri, K. P. Crawford, D. J. Merlo and R. H. Baltz, *Chem. Biol.*, 2001, **8**, 487–499.
35 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.
36 H. Yamase, L. S. Zhao and H. W. Liu, *J. Am. Chem. Soc.*, 2000, **122**, 12397–12398.
37 K. Räty, J. Kantola, A. Hautala, J. Hakala, K. Ylihonko and P. Mäntsälä, *Gene*, 2002, **293**, 115–122.
38 Z. Xu, K. Jakobi, K. Welzel and C. Hertweck, *Chem. Biol.*, 2005, **12**, 579–588.
39 R. M. Kagan and S. Clarke, *Arch. Biochem. Biophys.*, 1994, **310**, 417–427.
40 J. L. Martin and F. M. McMillan, *Curr. Opin. Struct. Biol.*, 2002, **12**, 783–793.
41 S. Gullón, C. Olano, M. S. Abdelfattah, A. F. Braña, J. Rohr, C. Méndez and J. A. Salas, *Appl. Environ. Microbiol.*, 2006, **72**, 4172–4183.
42 H. Onaka, S. Taniguchi, Y. Igarashi and T. Furumai, *Biosci., Biotechnol., Biochem.*, 2003, **67**, 127–138.
43 E. Mellado, L. M. Lorenzana, M. Rodríguez-Sáiz, B. Díez, P. Liras and J. L. Barredo, *Microbiology (Reading, U. K.)*, 2002, **148**, 1427–1438.
44 A. A. Khan, S. J. Kim, D. D. Paine and C. E. Cerniglia, *Int. J. Syst. Evol. Microbiol.*, 2002, **52**, 1997–2002.
45 C. L. Lu, Y. L. Huang, T. C. Wang and H.-T. Chiu, *BMC Bioinf.*, 2006, **7**, 295.
46 P. R. August, T. H. Grossman, C. Minor, M. P. Draper, I. A. MacNeil, J. M. Pemberton, K. M. Call, D. Holt and M. S. Osburne, *J. Mol. Microbiol. Biotechnol.*, 2000, **2**, 513–519.
47 R. V. Antônio and T. B. Creczynski-Pasa, *Genet. Mol. Res.*, 2004, **3**, 85–91.
48 H.-T. Chiu, Y.-C. Lin, M.-N. Lee, Y.-L. Chen, M.-S. Wang and C.-C. Lai, *Mol. BioSyst.*, 2009, **5**, DOI: 10.1039/B912395B, (the companion paper).
49 A. R. Howard-Jones and C. T. Walsh, *J. Am. Chem. Soc.*, 2006, **128**, 12289–12298.
50 Y. Sun, X. Zhou, H. Dong, G. Tu, M. Wang, B. Wang and Z. Deng, *Chem. Biol.*, 2003, **10**, 431–441.
51 K. Nishihara, M. Kanemori, M. Kitagawa, H. Yanagi and T. Yura, *Appl. Environ. Microbiol.*, 1998, **64**, 1694–1699.
52 S. Gaisser, R. Lill, G. Wirtz, F. Grolle, J. Staunton and P. F. Leadlay, *Mol. Microbiol.*, 2001, **41**, 1223–1231.
53 T. Kiser, M. Bibb, M. J. Buttner, K. F. Chater and D. A. Hopwood, in *Practical Streptomyces Genetics*, John Innes Foundation, Norwich, UK, 2000, pp. 117–118.
54 J. Sambrook and D. Russell, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York, 2001.