

Rapid characterization and engineering of natural product biosynthetic pathways *via* DNA assembler†Zengyi Shao,<sup>a</sup> Yunzi Luo<sup>a</sup> and Huimin Zhao<sup>\*ab</sup>

Received 18th December 2010, Accepted 31st January 2011

DOI: 10.1039/c0mb00338g

**We report a synthetic biology strategy for rapid genetic manipulation of natural product biosynthetic pathways. Based on DNA assembler, this method synthesizes the entire expression vector containing the target biosynthetic pathway and the genetic elements required for DNA maintenance and replication in various hosts in a single-step manner through yeast homologous recombination, offering unprecedented flexibility and versatility in pathway manipulations.**

Microorganisms and plants have evolved to produce a myriad array of complex molecules known as natural products or secondary metabolites that are of biomedical and biotechnological importance.<sup>1–3</sup> For example, it has been estimated that 77% of antibacterial drugs and 78% of anticancer drugs are natural products or have been derived from natural products.<sup>4</sup> While many effective drugs have been discovered in the past decades, a pressing need for the development of new therapeutic agents remains. This is mainly due to the rapid development of antibiotic resistance to existing drugs, the recent threat of bioterrorism, and the ever-increasing demand for new therapeutic agents for treatment of cancers and infectious diseases, to name a few.

Sequenced genomes and metagenomes represent a tremendously rich source for discovery of novel pathways involved in natural product biosynthesis.<sup>5,6</sup> Over the last decade, the complete genome sequences of more than 1500 organisms have been determined, with more than 7000 organisms in the pipeline (<http://www.genomesonline.org/cgi-bin/GOLD/bin/gold.cgi>). However, only a tiny fraction of these biosynthetic pathways have been characterized, and discovery and sustainable production of natural products are often hampered by our limited ability to manipulate the corresponding biosynthetic pathways.

Existing strategies for functional analysis of biosynthetic pathways can be broadly classified into two groups including the native host gene inactivation method and the heterologous

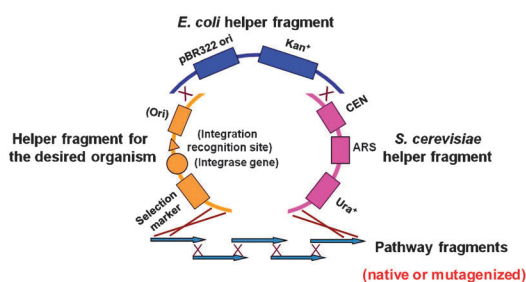
host gene expression method.<sup>6</sup> Compared to the native host-based method, the heterologous host-based method offers several advantages. First, most of native producers are not commonly cultivable in the laboratory and many microorganisms grow very slowly and produce minute amount of target compounds. It has been estimated that only 1% of bacteria and 5% of fungi have been cultivated in the laboratory.<sup>7–9</sup> Second, maximizing expression of a biosynthetic pathway in a well-characterized organism is usually easier and comparison of metabolic profiles is often less difficult. Third, the genetic tools available for a well-characterized organism enable the overproduction of a target natural product through metabolic engineering and the generation of new derivatives of a target compound through combinatorial biosynthesis. The heterologous host-based method has been successfully demonstrated in discovery and characterization of many natural products, such as methylenomycin furans,<sup>10</sup> phosphonic acids,<sup>11,12</sup> and polyketides.<sup>13,14</sup> However, this method also has several limitations. Major drawbacks include but are not limited to, the difficult transfer and genetic manipulation of large biosynthetic pathways such as performing scar-less targeted gene disruption and introducing single/multiple point mutations, problems with promoter recognition and gene expression, limited availability of expression hosts, and low yield.<sup>15</sup> Therefore, it is highly desirable to establish novel efficient modification, transfer, and expression technologies.

Due to its high efficiency and ease to work with, *in vivo* homologous recombination in *Saccharomyces cerevisiae* has been widely used for gene cloning and library creation.<sup>16–18</sup> Larionov and coworkers developed a transformation-associated recombination (TAR) cloning method in *S. cerevisiae* and used it to isolate large chromosomal fragments or entire gene clusters from complex genomes.<sup>19–22</sup> We developed the DNA assembler approach to assemble multiple gene expression cassettes into biochemical pathways either on a plasmid or on a chromosome in *S. cerevisiae* in a single-step fashion.<sup>23</sup> The recent success of constructing the entire *Mycoplasma genitalium* and *Mycoplasma mycoides* genomes further demonstrated the efficacy of using *S. cerevisiae* as a robust factory to assemble DNA molecules.<sup>24,25</sup>

Here we extended the DNA assembler approach, enabling it as a genomics-driven, synthetic biology-based method, for discovery, characterization, and engineering of natural product biosynthetic pathways. The strategy is illustrated in

<sup>a</sup> Institute for Genomic Biology,  
Department of Chemical and Biomolecular Engineering,  
University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA  
<sup>b</sup> Departments of Chemistry, Biochemistry, and Bioengineering,  
600 S. Mathews Avenue, Urbana, IL 61801, USA.  
E-mail: zhao5@illinois.edu; Fax: +1 217333-5052

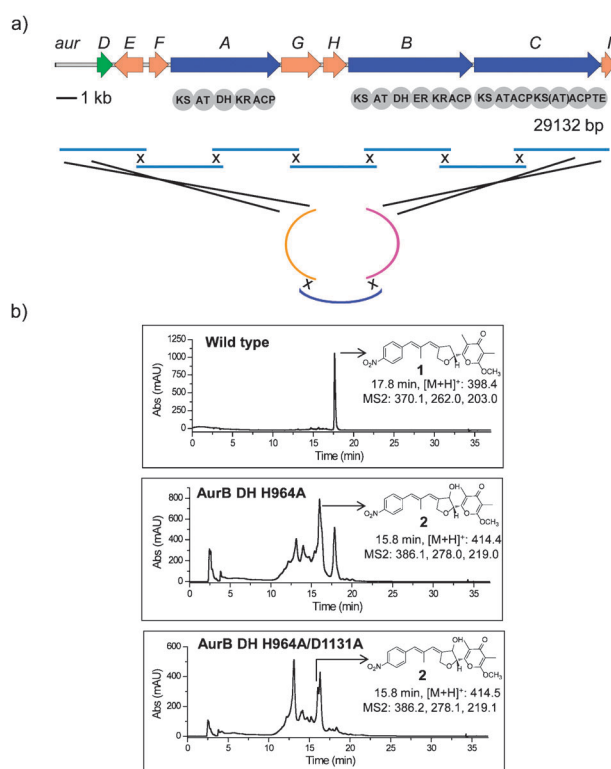
† Electronic supplementary information (ESI) available: Experimental, supplementary figures and tables. See DOI: 10.1039/c0mb00338g



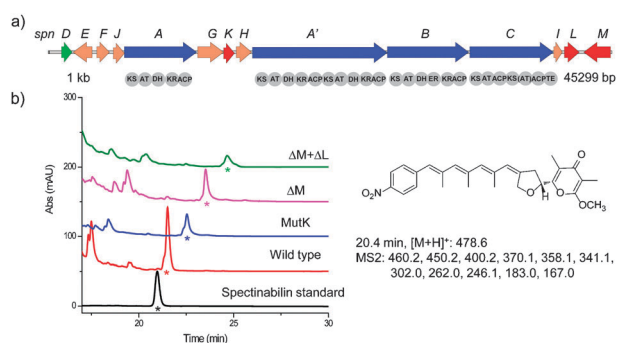
**Scheme 1** The DNA assembler based strategy for efficient manipulation of natural product biosynthetic pathways. Various genetic modifications can be introduced in the pathway fragments to be assembled.

Scheme 1. Briefly, pathway fragments encoding the target biosynthetic pathway and helper fragments carrying the genetic elements needed for DNA maintenance and replication in *S. cerevisiae*, *E. coli*, and the target heterologous expression host are amplified from the genome of the native producer and the corresponding vectors, respectively, and subsequently co-transformed into *S. cerevisiae*. Since polymerase chain reaction (PCR) primers are designed to generate an overlap region between two adjacent fragments, these fragments will be assembled into a single DNA molecule in *S. cerevisiae* through homologous recombination. The isolated plasmids are transformed into *E. coli* for plasmid enrichment and verification, and the correct construct is transformed into the desired host for heterologous expression of the target biosynthetic pathway. Of special note, because the pathway fragments can be readily modified by PCR, various genetic manipulations such as site-directed mutagenesis and scar-less gene deletion and insertion can be easily introduced into the target pathway.

We sought to study two related biosynthetic pathways, the aureothin<sup>26,27</sup> (Fig. 1a) and the spectinabilin<sup>28</sup> (Fig. 2a) biosynthetic pathways. Both pathways contain multiple type I polyketide synthases (PKSs) that are composed of domains including  $\beta$ -ketosynthase (KS), acyl transferase (AT), acyl carrier protein (ACP), ketoreductase (KR), dehydratase (DH), and enoylreductase (ER) domains. Because of the importance of polyketides in medicine and the modular, programmable nature of the biosynthetic machinery, PKSs have been extensively investigated and engineered to produce new derivatives.<sup>29</sup> Aureothin is a rare nitroaryl-substituted polyketide (Fig. 1b, compound 1) from *Streptomyces thioluteus*, which exhibits antitumor, antifungal, and insecticidal activities.<sup>26,27</sup> Seven pathway fragments of 4–5 kb, were PCR-amplified from the genomic DNA, and co-transformed with the three helper fragments into *S. cerevisiae*, leading to their assembly into a single circular DNA molecule of 35.7 kb (29.1 kb from the aureothin biosynthetic pathway and 6.6 kb from the helper fragments). The details of designing the overlaps between adjacent fragments were described in Fig. S1 in the ESI†. Six out of the ten randomly picked constructs were shown to be correctly assembled, as verified by restriction digestion analysis, indicating a high assembly efficiency of 60%. Three of the correct constructs were subsequently conjugated to a heterologous host, *Streptomyces lividans*. Note that a *Streptomyces* phage-derived  $\phi$ C31 integrase gene, a  $\phi$ C31 recognition site, and an origin of conjugal transfer



**Fig. 1** Assembly of the wild type aureothin biosynthetic pathway and the mutant aureothin biosynthetic pathways for creation of new derivatives. (a) Scheme of the aureothin biosynthetic pathway. (b) LC-MS analysis of the *S. lividans* clones carrying the wild type aureothin biosynthetic pathway and the mutant aureothin biosynthetic pathways.



**Fig. 2** Gene knock-out and inactivation studies on the spectinabilin biosynthetic pathways. (a) Scheme of the spectinabilin biosynthetic pathway. (b) LC-MS analysis of the *S. lividans* clones carrying the wild type spectinabilin biosynthetic pathway and the mutant spectinabilin biosynthetic pathways. The asterisks indicate the spectinabilin peaks. MutK, the mutant with a stop codon introduced to SpnK;  $\Delta$ M, the mutant with SpnM deleted;  $\Delta$ M +  $\Delta$ L, the mutant with SpnM and SpnL deleted.

(Fig. S2 in the ESI†) were included in the *Streptomyces* helper fragment so that the target biosynthetic pathway would be transferred and integrated to the chromosome of *S. lividans*. As a result, all three *S. lividans* clones produced aureothin with a titer of  $6.1 \pm 0.2$  mg L<sup>-1</sup> (Fig. 1b and Fig. S3 in the ESI†),

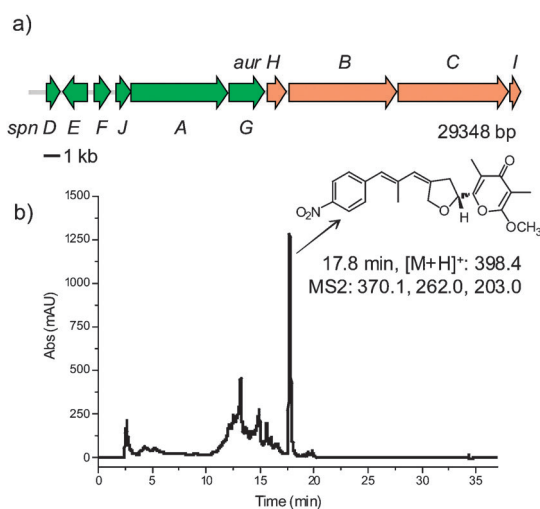
indicating that such a cluster of 29 kb has been successfully assembled and heterologously expressed.

Based on this system, site-directed mutagenesis can now be readily used to generate new derivatives. Unlike the labor-intensive and time-consuming procedures used in the conventional methods,<sup>30–32</sup> DNA assembler only requires adding site-specific mutation(s) into the PCR primers used to generate pathway fragments. As an example, the active sites of the DH domain of AurB were targeted. The motifs HXXXGXXXXP and DXXX(Q/H) were found to be conserved among the DH domains of PKSs and the histidine and the aspartic acid were identified as the catalytic residues.<sup>33–35</sup> Based on this information, AurB DH H964A mutant and AurB DH H964A/D1131A double mutant were generated. We expected them to produce compounds **2** (Fig. 1b) if the corresponding domain was completely inactivated, or a mixture of both compound **2** and aureothin if the activity was only reduced. The mutation led to significant changes in HPLC analysis in which several new peaks appeared and the aureothin peak was drastically decreased (Fig. 1b). The distinguishable peak at 15.8 min was confirmed to be compound **2** by its MS and MS2 patterns (Fig. S4 in the ESI†). Interestingly, it was found that the wild type pathway could also produce a minute amount of compound **2** due to the incomplete modification catalyzed by the AurB DH domain on the upstream intermediate. In order to demonstrate that the derivative was produced mainly due to the corresponding mutation(s), the relative yield, defined as the ratio between the level of compound **2** being produced and the level of the generated aureothin, was determined. The wild type pathway produced compound **2** at a relative yield of 0.023:1, while the AurB DH H964A mutant and the AurB DH H964A/D1131A mutant gave a relative yield of 1.03 : 1 and 26.3 : 1, respectively (Fig. 1b). From the evolutionary point of view, the wild type pathway should be most efficient for the synthesis of the product because almost no accumulation of intermediates was observed, whereas the modified pathways could accumulate some intermediates if the downstream enzymes cannot take those intermediates with modified structures as perfect substrates and only process them at a slower rate. On the other hand, the residual production of aureothin suggests that the corresponding mutation(s) either did not completely inactivate the enzymes, or completely inactivated the enzymes but another unknown enzyme with dehydratase activity played a role in the dehydration step. In addition, the wild type pathway can naturally synthesize a tiny amount of the aureothin analogue. This could take place when a small amount of an intermediate skips a certain enzymatic reaction and at the same time is processed by the downstream enzymes. Overall, our results have demonstrated that single/multiple site-directed mutagenesis can be easily performed without going through the complicated multi-step procedures used by other conventional approaches.<sup>30–32</sup>

Compared to aureothin, spectinabilin is a longer nitro-phenyl containing polyketide, which exhibits antimalarial and antiviral activities.<sup>36,37</sup> The biosynthetic pathway from *Streptomyces spectabilis* consists of 4 PKSs (Fig. 2a)<sup>28</sup> and shares a very high sequence homology with the spectinabilin biosynthetic pathway from *Streptomyces orinoci*<sup>27</sup> (Fig. S5 in the ESI†). However, only the biosynthetic pathway from *S. spectabilis* was successfully

expressed in *S. lividans*,<sup>28</sup> while the biosynthetic pathway from *S. orinoci* failed to produce any spectinabilin in the heterologous host.<sup>27,38</sup> Comparison of the two biosynthetic pathways revealed one major difference: the *S. spectabilis* cluster contains three extra genes, *spnK*, *spnL*, and *spnM* (Fig. S5 in the ESI†). To investigate their role in spectinabilin biosynthesis, we used DNA assembler to inactivate *spnK* by site-directed mutagenesis and completely delete *spnM* and *spnM* + *spnL*. Initially, we attempted to assemble the spectinabilin biosynthetic pathway using the same procedure as we did for the aureothin biosynthetic pathways. However, a low efficiency (<10%) was observed and most of the constructs underwent severe gene deletion(s), possibly due to the high sequence identity among PKS domains. To address this issue, a modified two-step assembly strategy was devised (Fig. S6 in the ESI†), through which we were able to assemble this biosynthetic pathway of 45 kb with an efficiency of 30%. To inactivate *spnK*, the codon TGC encoding Cys199 was changed to a stop codon TGA in the PCR primers. The genes *spnM* and *spnM* + *spnL* were individually removed from the biosynthetic pathway by redesigning the reverse primers for amplifying the last pathway fragments. All three mutants produced spectinabilin (Fig. 2b and Fig. S7 in the ESI†), although their titers were affected by the corresponding mutations (Wild type,  $31.8 \pm 4.5 \mu\text{g L}^{-1}$ ; MutK,  $17.2 \pm 0.5 \mu\text{g L}^{-1}$ ; DelM,  $21.2 \pm 2.1 \mu\text{g L}^{-1}$ ; DelML,  $10.5 \pm 1.6 \mu\text{g L}^{-1}$ ), indicating that *spnK*, *spnL*, and *spnM* are not required for spectinabilin biosynthesis, supporting our previous bioinformatics study showing that they are only involved in up-regulating substrate concentration and product transportation.<sup>28</sup>

To further demonstrate the versatility of our approach in manipulating large gene clusters, we created a hybrid biosynthetic pathway between the aureothin and the spectinabilin pathways. Because the aureothin pathway lacks the counterpart of the *spnA'* gene in the spectinabilin biosynthetic pathway, we removed *spnA'* and created a hybrid pathway comprising



**Fig. 3** Assembly of a hybrid biosynthetic pathway between the spectinabilin biosynthetic pathway and the aureothin biosynthetic pathway. (a) Scheme of the hybrid biosynthetic pathway. (b) LC-MS analysis of the *S. lividans* clone carrying the hybrid biosynthetic pathway.



*spnDEFJAG* and *aurHBCI* (Fig. 3a). The analysis of the N-terminal sequences revealed that SpnA' and AurB share a very high homology at their N-terminals while SpnA' and SpnB are very different at this end (Fig. S8 in the ESI†), indicating SpnA could possibly recognize AurB rather than SpnB. As expected, the resulting construct produced aureothin in the heterologous host with a titer of 0.17 mg L<sup>-1</sup> (Fig. 3b). The generation of aureothin by a hybrid pathway was achieved previously, but a much more complicated experimental protocol has to be used.<sup>27</sup>

## Conclusions

In summary, we have demonstrated that DNA assembler enables sophisticated genetic manipulations such as point mutagenesis and scar-less gene substitution and deletion, which can be used to confirm the gene function, locate key amino acid residues, study the biosynthetic mechanism, express biosynthetic pathways heterologously, and generate new derivatives. Moreover, with little modification, DNA assembler may be used for many other applications such as promoter/regulator replacement, cryptic pathway activation, artificial operon construction, rapid library creation and pathway optimization. All these sophisticated genetic manipulations are difficult to accomplish by other conventional methods. Overall, because the DNA fragments to be assembled are completely mobile and amenable to all sorts of sophisticated genetic manipulations accessible to PCR or can be chemically synthesized *de novo* with optimized codons, this DNA assembler based strategy offers the ultimate versatility and flexibility in characterizing and engineering natural product biosynthetic pathways.

We thank Frances H. Arnold for critical reading of this manuscript. This work was supported by the National Academies Keck Futures Initiative on Synthetic Biology (NAKFI SB13) and the National Institutes of Health (GM077596).

## Notes and references

- P. M. Dewick, *Medical natural products. A biosynthetic approach*, John Wiley and Sons, Chichester, UK, 2nd edn, 2002.
- R. B. Herbert, *The biosynthesis of secondary metabolites*, Chapman and Hall, London, UK, 2nd edn, 1989.
- J. W. H. Li and J. C. Vederas, *Science*, 2009, **325**, 161–165.
- D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2007, **70**, 461–477.
- G. L. Challis, *Microbiology*, 2008, **154**, 1555–1569.
- M. Zerkly and G. L. Challis, *ChemBioChem*, 2009, **10**, 625–633.
- A. T. Bull, M. Goodfellow and J. H. Slater, *Annu. Rev. Microbiol.*, 1992, **46**, 219–252.
- A. L. Demain, *J. Ind. Microbiol. Biotechnol.*, 2006, **33**, 486–495.
- J. R. Leadbetter, *Curr. Opin. Microbiol.*, 2003, **6**, 274–281.
- C. Corre, L. Song, S. O'Rourke, K. F. Chater and G. L. Challis, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 17510–17515.
- R. D. Woodyer, Z. Shao, P. M. Thomas, N. L. Kelleher, J. A. Blodgett, W. W. Metcalf, W. A. van der Donk and H. Zhao, *Chem. Biol.*, 2006, **13**, 1171–1182.
- A. C. Eliot, B. M. Griffin, P. M. Thomas, T. W. Johannes, N. L. Kelleher, H. Zhao and W. W. Metcalf, *Chem. Biol.*, 2008, **15**, 765–770.
- C. Li, K. E. Roeger and W. L. Kelly, *ChemBioChem*, 2009, **10**, 1064–1072.
- S. C. Wenzel, H. B. Bode, I. Kochems and R. Muller, *ChemBioChem*, 2008, **9**, 2711–2721.
- H. Gross, *Appl. Microbiol. Biotechnol.*, 2007, **75**, 267–277.
- P. L. Gunyuzlu, G. F. Hollis and J. H. Toyn, *BioTechniques*, 2001, **31**, 1246, 1248, 1250.
- C. K. Raymond, T. A. Powner and S. L. Sexson, *BioTechniques*, 1999, **26**, 134–141.
- C. Schaerer-Brodbeck and A. Barberis, *BioTechniques*, 2004, **37**, 202–206.
- M. Cocchia, N. Kouprina, S. J. Kim, V. Larionov, D. Schlessinger and R. Nagaraja, *Nucleic Acids Res.*, 2000, **28**, E81.
- V. Noskov, N. Kouprina, S. H. Leem, M. Koriabine, J. C. Barrett and V. Larionov, *Nucleic Acids Res.*, 2002, **30**, E8.
- V. N. Noskov, N. Kouprina, S. H. Leem, I. Ouspenski, J. C. Barrett and V. Larionov, *BMC Genomics*, 2003, **4**, 16.
- Z. Feng, J. H. Kim and S. F. Brady, *J. Am. Chem. Soc.*, 2010, **132**, 11902–11903.
- Z. Shao, H. Zhao and H. Zhao, *Nucleic Acids Res.*, 2009, **37**, e16.
- D. G. Gibson, G. A. Benders, K. C. Axelrod, J. Zaveri, M. A. Algire, M. Moodie, M. G. Montague, J. C. Venter, H. O. Smith and C. A. Hutchison, 3rd, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 20404–20409.
- D. G. Gibson, J. I. Glass, C. Lartigue, V. N. Noskov, R. Y. Chuang, M. A. Algire, G. A. Benders, M. G. Montague, L. Ma, M. M. Moodie, C. Merryman, S. Vashee, R. Krishnakumar, N. Assad-Garcia, C. Andrews-Pfannkoch, E. A. Denisova, L. Young, Z. Q. Qi, T. H. Segall-Shapiro, C. H. Calvey, P. P. Parmar, C. A. Hutchison, 3rd, H. O. Smith and J. C. Venter, *Science*, 2010, **329**, 52–56.
- J. He and C. Hertweck, *Chem. Biol.*, 2003, **10**, 1225–1232.
- N. Traitcheva, H. Jenke-Kodama, J. He, E. Dittmann and C. Hertweck, *ChemBioChem*, 2007, **8**, 1841–1849.
- Y. S. Choi, T. W. Johannes, M. Simurdiak, Z. Shao, H. Lu and H. Zhao, *Mol. Biosyst.*, 2010, **6**, 336–338.
- J. Staunton and K. J. Weissman, *Nat. Prod. Rep.*, 2001, **18**, 380–416.
- T. Ito, N. Roongsawang, N. Shirasaka, W. Lu, P. M. Flatt, N. Kasanah, C. Miranda and T. Mahmud, *ChemBioChem*, 2009, **10**, 2253–2265.
- F. Karray, E. Darbon, H. C. Nguyen, J. Gagnat and J. L. Pernodet, *J. Bacteriol.*, 2010, **192**, 5813–5821.
- J. A. Blodgett, P. M. Thomas, G. Li, J. E. Velasquez, W. A. van der Donk, N. L. Kelleher and W. W. Metcalf, *Nat. Chem. Biol.*, 2007, **3**, 480–485.
- A. Keatinge-Clay, *J. Mol. Biol.*, 2008, **384**, 941–953.
- T. Moriguchi, Y. Kezuka, T. Nonaka, Y. Ebizuka and I. Fujii, *J. Biol. Chem.*, 2010, **285**, 15637–15643.
- K. Pawlik, M. Kotowska, K. F. Chater, K. Kuczek and E. Takano, *Arch. Microbiol.*, 2007, **187**, 87–99.
- K. Kakinuma, C. A. Hanson and K. L. Rinehart Jr, *Tetrahedron*, 1976, **32**, 217–222.
- M. Isaka, A. Jaturapat, J. Kramyu, M. Tanticharoen and Y. Thebtaranonth, *Antimicrob. Agents Chemother.*, 2002, **46**, 1112–1113.
- N. Traitcheva, *PhD thesis*, Leibniz-Institute for Natural Product Research and Infection Biology (HKI), 2006.