JOC The Journal of Organic Chemistry

Article

Discovery of the Streptoketides by Direct Cloning and Rapid Heterologous Expression of a Cryptic PKS II Gene Cluster from Streptomyces sp. Tü6314

Zhengyi Qian, Torsten Bruhn, Paul M D'Agostino, Alexander Herrmann, Martin Haslbeck, Noemi Antal, Hans-Peter Fiedler, Ruth Brack-Werner, and Tobias A. M. Gulder

J. Org. Chem., **Just Accepted Manuscript** • DOI: 10.1021/acs.joc.9b02741 • Publication Date (Web): 20 Nov 2019

Just Accepted

is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Discovery of the Streptoketides by Direct Cloning and Rapid Heterologous Expression of a Cryptic PKS II Gene Cluster from *Streptomyces* sp. Tü6314

Zhengyi Qian,[a] Torsten Bruhn,[b] Paul M. D'Agostino,[a,c] Alexander Herrmann,[d] Martin Haslbeck,[e] Noémi Antal,[f] Hans-Peter Fiedler,[f] Ruth Brack-Werner,[d] and Tobias A. M. Gulder[a,c*]

[a]Biosystems Chemistry, Department of Chemistry and Center for Integrated Protein Science Munich (CIPSM), Technical University of Munich, Lichtenbergstraße 4, 85748 Garching bei München, Germany.

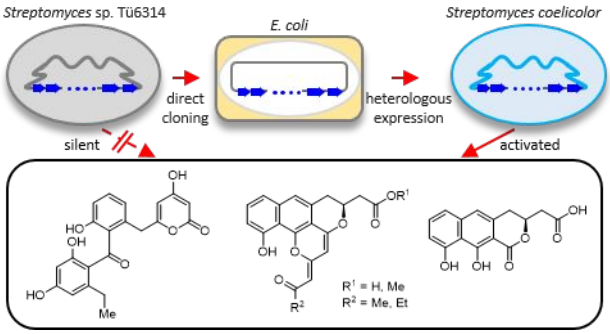
[b]Bundesinstitut für Risikobewertung, Max-Dohrn-Str. 8-10, 10789 Berlin, Germany. [c]Technische Universität Dresden, Chair of Technical Biochemistry, Bergstraße 66, 01602 Dresden, Germany.

[d]Helmholtz Zentrum München, German Research Center for Environmental Health, Institute of Virology, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany.

[e]Department of Chemistry, Technical University of Munich, Lichtenbergstraße 4, 85748 Garching bei München, Germany.

[f]Institute of Microbiology, University of Tübingen, Auf der Morgenstelle 28, D-72076 Tübingen, Germany

[*]Correspondence: tobias.gulder@ch.tum.de, tobias.gulder@tu-dresden.de



ABSTRACT

Genome sequencing and bioinformatic analysis have identified numerous cryptic gene clusters that have the potential to produce novel natural products. Within this work, we identified a cryptic type II PKS gene cluster (*skt*) from *Streptomyces* sp. Tü6314. Facilitated by linear plus linear homologous recombination-mediated recombineering (LLHR), we directly cloned the *skt* gene cluster using the *Streptomyces* site-specific integration vector pSET152. The direct cloning allowed for rapid heterologous expression in *Streptomyces coelicolor*, leading to the identification and structural characterization of six polyketides (three known compounds and the new streptoketides), four of which exhibiting anti-HIV activities. Our study shows that the pSET152 vector can be directly used for LLHR, expanding the Rec/ET direct cloning toolbox and providing the possibility for rapid heterologous expression of gene clusters from *Streptomyces*.

Introduction

Natural products have played a highly significant role in the drug discovery and development process over the last decades.[1] The important drugs penicillin (anti-biotic), avermectins (anti-helmintic and insecticidal) and artemisinin (anti-malaria) are instructive examples of natural products heavily used in current medical applications. Out of all the known microbial natural product producers, actinomycetes produce over half of the antibiotics that exhibit selective biological activities against pathogenic bacteria and fungi, and about 75% of these are produced by *Streptomyces*.[2] However, after the Golden Age of natural product drug discovery in the middle of the 20th century, the late 20th century has seen a sharp decline in such discovery programs from pharmaceutical companies, in part because of the advances in both high throughput screening (HTS) and combinatorial synthesis.[3] The high rediscovery rate of known molecules and the often low quantities of compounds isolated from native producers have further promoted this decrease.[4] Whole-genome sequencing has shown that many microbes have far greater potential to produce secondary metabolites when compared to the chemical diversity already identified.[5-6] Bioinformatic methods and computational tools have been developed to screen this new wealth of genomic data for the identification of biosynthetic gene clusters (BGCs) in a process known as genome mining.[7-9] Genome-based structure prediction then paves the way for the directed discovery of natural products that are likely to yield yet unidentified chemical scaffolds. However, many BGCs are not sufficiently expressed under standard laboratory culture conditions (so-called silent/cryptic pathways) in their natural hosts to allow compound detection, making the discovery of their encoded metabolites a difficult task.[4, 10] Heterologous expression of BGCs is an emerging method in genome mining to overcome this limitation. If successful, metabolites can easily be tracked when a heterologous BGC is expressed in a background-clear host with significantly

simplified target identification, for example by comparative HPLC-MS analysis.[4, 10] *Streptomyces coelicolor* M1152 and *Streptomyces coelicolor* M1154 have specifically been engineered for such secondary metabolites production and are commonly used as heterologous expression hosts for the discovery of novel natural products.[11]

A precondition for the heterologous expression of BGCs is the ability to intercept the respective DNA fragment from the genomic DNA of the natural host. Traditional heterologous expression relied on construction of genomic DNA large-insert clone libraries to screen for colonies harbouring the desired BGCs. Generating and screening such libraries is time-consuming and it is often difficult to find a colony that harbors the entire BGC, especially for gene clusters >40 kb.[4] Recently, new cloning systems have been developed that bypass genomic DNA library construction, thereby largely improving the efficiency of BGCs interception. Important methods include RecET-mediated linear-plus-linear homologous recombination (LLHR), transformation-associated recombination (TAR) and Direct Pathway Cloning (DiPaC).[12-15] LLHR in *E. coli* is suitable for capturing BGCs up to 52 kb in size from *Photorhabdus luminescens*.[12] This method was also used to clone the 106 kb salinomycin gene cluster from *Streptomyces albus* from multiple fragments with downstream whole construct assembly.[16] Extension of the methodology by combination with exonuclease *in vitro* assembly (ExoCET) even facilitated capture of the complete 106 kb gene cluster in a single step.[17] However, the vectors used in LLHR need to be modified by Redαβ recombineering to make them suitable for heterologous expression.[18] Generating a vector system that allows omitting this recombineering step thus has the potential to further streamline LLHR. The pSET152 vector is an *E. coli-Streptomyces* shuttle vector with ϕC31 integration system, enabling its maintenance in *E. coli* and site-specific integration into the *Streptomyces* genome.[18-19] In principle, the pSET152 vector should be applicable for direct cloning

of BGCs from *Streptomyces* genomic DNA for downstream heterologous expression within another *Streptomyces* host.

Type II polyketide synthases (PKSs) produce structurally diverse aromatic metabolites that often possess important biological activities, such as the antibiotic tetracyclines and anticancer drug doxorubicin.[20-22] Type II PKS BGCs have a minimal PKS core consisting of three proteins: ketosynthase alpha (KS_α); ketosynthase beta (KS_β, or chain length factor); and acyl carrier protein (ACP).[23] The minimal PKS is responsible for assembly of the nascent polyketide chain, which is then cyclized to form the aromatic core structure. Complex chemical modifications including oxidation, reduction, methylation and/or glycosylation result in a broad array of structural complexity.[23] Herein, we report the one step capture of a cryptic type II PKS gene cluster from *Streptomyces* sp. Tü6314 by LLHR using the vector pSET152 and its heterologous expression in the host *Streptomyces coelicolor*, leading to the successful recombinant production of six aromatic polyketides, including three new compounds termed streptoketides.

RESULTS AND DISCUSSION

Identification and bioinformatic analysis of the *skt* gene cluster from *Streptomyces* sp. Tü6314

The complete genome of *Streptomyces* sp. Tü6314 was sequenced by PacBio sequencing technology. This resulted in a large (7.76 Mbp) and a small (12.4 kbp) contig, corresponding to a genome size of approx. 7.8 Mbp with a GC content of 71%. Bioinformatic analysis of the genome data using antiSMASH 4.0[24] revealed a cryptic type II PKS gene cluster (*skt*) with ≤ 28% of genes showing similarity to previously characterized pathways in the MIBiG (Minimum Information about a Biosynthetic Gene cluster) database (Figure 1A).[25] The sequence of the *skt* gene cluster was annotated and submitted to GenBank and can be accessed using accession number

MK424349. Detailed BLASTp analysis showed that the *skt* gene cluster has many genes encoding proteins with high homology to oxytetracycline[20] and SF2575[26] biosynthesis (Table 1).

Table 1. Deduced functions of ORFs in the *skt* gene cluster.

ORF	Size[a]	Predicted function	*Streptomyces rimosus* homolog[b]	Accession number	SF2575 homolog[b]	Accession number
L2	77		no hit	-	-	-
L1	300		NmrA/HSCARG family protein (32/45%)	WP_030635220.1	-	-
skt1	81	ACP	OxyC (55/75%)	AAZ78327.1	SsfC (57/77%)	ADE34520.1
skt2	422	KS_{β}/CLF	OxyB (58/69%)	AAZ78326.1	SsfB (55/67%)	ADE34519.1
skt3	423	KS_{α}	OxyA (66/79%)	AAZ78325.1	SsfA (65/79%)	ADE34518.1
skt4	167	regulator	MarR family transcriptional regulator (30/47%)	WP_030371052.1	-	
skt5	151	cyclase	OxyI (54/67%)	AAZ78332.2	SsfY4 (53/65%)	ADE34486.1
skt6	267	regulator	OtcR (42/61%)	AJO26937.1	SsfT1 (41/61%)	ADE34517.1
skt7	518	oxidase	GMC family oxidoreductase (35/50%)	WP_030663369.1	-	
skt8	224	methyl-transferase	SAM-dependent methyltransferase (54/69%)	WP_030682298.1	-	
skt9	267	ketoacyl reductase	OxyJ (66/78%)	AAZ78333.1	SsfU (70/79%)	ADE34491.1
skt10	347	Aromatase / Cyclase	OxyK (44/58%)	AAZ78334.2	SsfY1 (46/61%)	ADE34490.1
skt11*	615	CoA ligase	OxyH?		SsfL2 (45/58%)	ADE34493.1
skt12	246	reductase	SDR family oxidoreductase (35/46%)	WP_030668664.1	SsfU (35/47%)	ADE34491.1
skt13	263	reductase	OxyM (46/59%)	AAZ78336.1	SsfU (28/42%)	ADE34491.1
skt14	417	mono-oxygenase	OxyE (56/68%)	AAZ78329.1	SsfO2 (26/40%)	ADE34483.1
skt15	257	cyclase	OxyN (61/71%)	AAZ78337.1	-	
skt16	328	reductase	aldo/keto reductase (50/65%)	KOT97719.1	SsfF (60/72%)	ADE34525.1
skt17	343	KSIII	3-oxoacyl-ACP synthase (57/68%)	WP_050512192.1	SsfG (35/49%)	ADE34509.1
skt18	333	acyl-transferase	acyltransferase domain-containing protein (59/69%)	WP_030372723.1	SsfV (48/58%)	ADE34485.1
skt19	536	carboxylase	methylmalonyl-CoA carboxyltransferase (81/88%)	GCD42787.1	SsfE (78/84%)	ADE34513.1
R1	556		Membrane protein (66/81%)	KEF22117.1	-	
R2	741		Elongation factor G (75/85%)	GCD47180.1	-	

a. Size in number of amino acids. b. *Streptomyces rimosus* is the producer of oxytetracycline, *Streptomyces* sp. SF2575 is the producer of SF2575. Homolog shows protein sequences identity/similarity%. * OxyH is not correctly annotated in the database.

It has previously been shown that both KS_α and KS_β proteins will group phylogenetically into clades that correlate well with the chain length and initial cyclization pattern of the polyketide precursor produced by a Type II PKS.[27-28] Here, a phylogenetic tree was constructed to compare the *skt* KS_α/KS_β (Skt3/Skt2) with the KS_αs/KS_βs from functionally characterized type II gene clusters.[28] The *skt* KS_α (Skt3) and KS_β (Skt2) grouped well into parallel clades, with a proposed chain length of 21 carbons (Figure S1). Attempts to detect and isolate the respective PKS II products by fermentation of *Streptomyces* sp. Tü6314 were unsuccessful (>10 L fermentation, data not shown). This suggested that the *skt* gene cluster was silent or expressed at low levels in the native host under the used culture conditions. We therefore aimed to activate and express the *skt* BGC in a suitable recombinant host system.

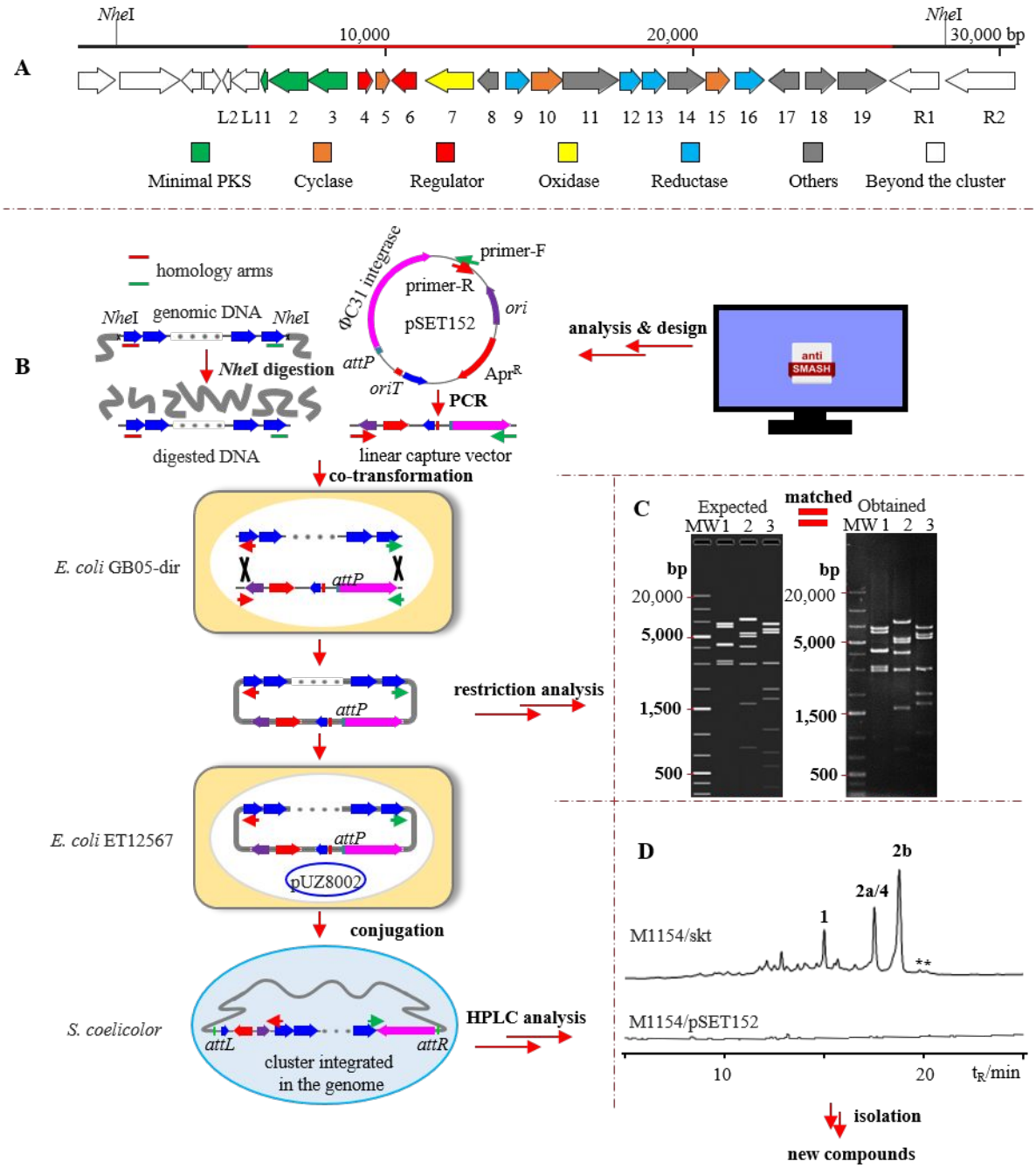


Figure 1. (A) Gene organization of the *skt* gene cluster. The red line indicates the part of the cloned *skt* cluster and was used for heterologous expression within this study. *NheI* was used to excise the cluster from genomic DNA. (B) General workflow for the direct cloning of *skt* by LLHR. pSET152 was used as the expression vector and integrated into the *S. coelicolor* genome for heterologous expression. (C) Left: simulated restriction enzyme analysis of the captured cluster by pSET152 digested with *Blp*I (1), *Pst*I (2) and *Pvu*II (3). Right: authentic restriction enzyme analysis of the captured cluster digested with above restriction enzymes. For details on DNA fragment sizes, please see Figure S4 and Table S3. (D) HPLC-UV analysis of culture extracts from *S. coelicolor*

transformed with emtpy pSET152 and pSET152-*skt*. Putative signals of compounds **3a** (t_R approx. 19.3 min) and 3b (t_R approx. 20.4 min) are labeled with *.

LLHR capture and heterologous expression of the *skt* gene cluster in *S. coelicolor*

For heterologous expression, we selected *S. coelicolor* M1152 and M1154 as hosts. Both M1152 and M1154 are genetically engineered overexpression hosts with a reduced metabolic background.[11] M1152 and M1154 have the ability to produce 20-40 times higher titers of recombinant products when compared to their parental strain M145.[11] They are thus ideally suited for the recombinant production of natural products from other *Streptomyces*. LLHR is a powerful tool to capture BGCs from a broad range of diverse microbes.[12, 16, 18, 29] In this study, we used the pSET152 vector to directly capture the 21.7 kb *skt* BGC via LLHR (Figure 1B). After LLHR, subsequent colony screening PCR resulted in one positive colony out of twelve (Figure S2). Restriction digest analysis showed that the desired *skt* gene cluster was successfully cloned into pSET152 (Figure 1C, Figures S3-S4). The subsequently isolated pSET152-*skt* plasmid was directly introduced into *S. coelicolor* M1152 and M1154 via conjugation for heterologous expression without further modification. The exconjugants with *skt* integrated in the M1154 genome developed strongly red colonies, while the engineered M1152 strain was only slightly red and the strain harboring the empty pSET152 vector developed white to pale grey colonies. Comparative analysis of the metabolite profiles showed that additional compounds were produced in both M1152/*skt* and M1154/*skt* when compared to the empty pSET152 control. The production titers of M1154/*skt* were significantly higher than those of M1152/*skt* (Figure 1D). Consequently, M1154/*skt* was used for downstream compounds isolation.

Isolation and structural elucidation of the natural products

For compound production, strain M1154/*skt* was cultivated in ISP Medium 4. The combined organic extracts of the supernatant and the cell pellet were fractionated by Sephadex LH-20 size exclusion chromatography. Fractions containing compounds absent in the control strain M1154/pSET152 were further purified by semi-preparative HPLC on C18 material. This yielded a total of 6 pure compounds exclusively present in M1154/*skt* (Figure 2). Compound **1** (3.6 mg, HPLC retention time at 14.2 min) possessed a molecular mass of 383.1122 units (Figures S5-S6) which best fits a molecular formula of $C_{21}H_{18}O_7$ in its protonated form (calcd. 383.1125 u), leading to a calculated 13 degrees of unsaturation. Literature search with this and the ^{1}H and ^{13}C NMR data revealed this compound to be UWM5 (**1**), a well-known polyketide shunt product (Table S4, Figures S7-S9). Compound **1** was first isolated from the doxorubicin producing strain with a gene encoding a crucial cyclase disrupted.[30] Compound **2a** (0.7 mg, 16.8 min) had a molecular mass of 353.1017 u (Figures S10-S11) corresponding to a molecular formula of $C_{20}H_{16}O_6$ in its protonated form (calcd. 353.1020 u), likewise possessing 13 degrees of unsaturation. The molecular mass of **3a** (2.8 mg, 19.3 min) was 367.1174 u (Figure S15-S16), resulting in a molecular composition of $C_{21}H_{18}O_6$ in its protonated form, again containing 13 degrees of unsaturation. The observed mass difference of 14 units between **2a** and **3a** therefore suggested **3a** to be a methylated analog (calcd. 367.1176 u) of **2a**. This was also consistent with the differences in the ^{1}H and ^{13}C NMR spectra of **2a** versus **3a**, differing in the additional presence of a methyl ester in **3a** (Tables S5-S8, Figures S12-S14 and S17-S19). Using this combined data, **2a** and **3a** were readily identified as S2502 and S2507, respectively, two molecules previously obtained by heterologous expression of the nogalamycin anthraquinone aglycone genes in *Streptomyces lividans* TK24.[31]

Two additional natural products that seemed to be highly related to **2a/3a** were isolated: compounds **2b/3b**. Compound **2b** (2.6 mg, 17.9 min) possessed an identical molecular formula of $C_{21}H_{18}O_6$ (MS data of the protonated form 367.1174 u, calcd. 367.1176 u) when compared to **3a** (Figures S20-S21). The significantly reduced retention time under identical RP-HPLC isolation conditions (Δt_R = 1.4 min) suggested **2b** to bear a free acid function, with the additional methyl group likely being present in form of a phenolic methyl ester. However, inspection of the ^{1}H and ^{13}C NMR data showed that no methyl ester was present in **2b** (Figure S24). Instead, the compound contained an additional methylene unit. ^{1}H NMR data showed this methylene group to be directly connected to an electron-withdrawing group and a methyl function, based on chemical shifts and signal multiplicity (2H, q with J = 7.4 Hz at 2.35 ppm and 3H, t, with J = 7.4 Hz at 1.02 ppm, also observable in **3b**) (Table S9). 2D NMR (COSY, HMBC) clearly proved the methylene unit to be inserted at the terminal ketone C19 (Figures S25-S26), leading to an ethyl ketone in **2b** versus the methyl ketone in **2a/3a** (Figure 2, box). This suggests the PKS system to be alternatively primed with a propionyl-CoA starter unit (see discussion below), leading to the observed new polyektide named streptoketide A (**2b**). Compound **3b** (6.9 mg, 20.4 min) possessed a molecular mass of 381.1329 u (Figures S22-S23) which is thus 14 u higher than that of **2b**, again suggesting the presence of an additional methyl group (calcd. 381.1333). With the differences in retention times between **2a** and **3a** being identical to those of **2b** and **3b** (Δt_R = 2.5 min), the presence of a methyl ester in **3b** was most likely. This assumption was corroborated by its NMR data (Table S9, Figures S27-S29), revealing **3b** to likewise be a new natural product, which was termed streptoketide B (**3b**).

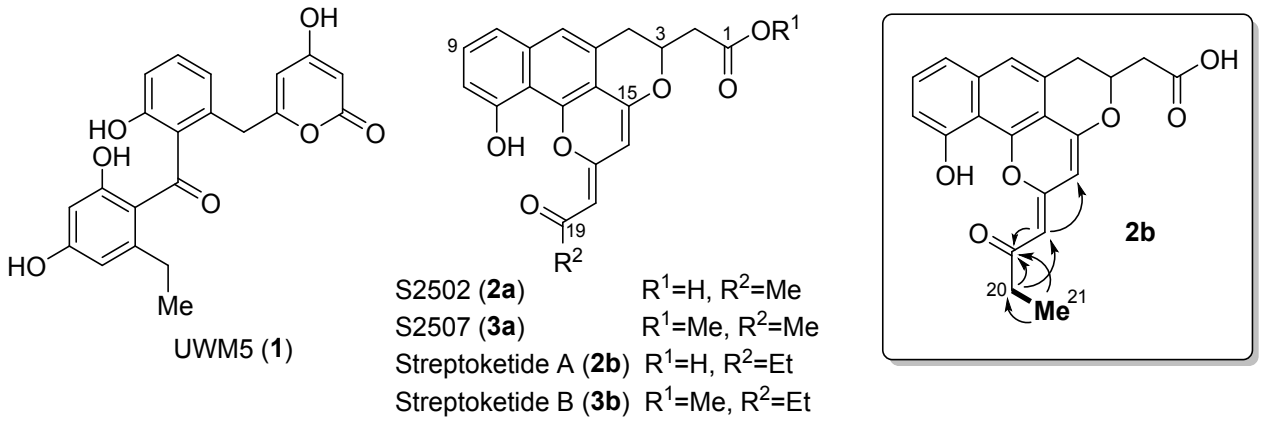


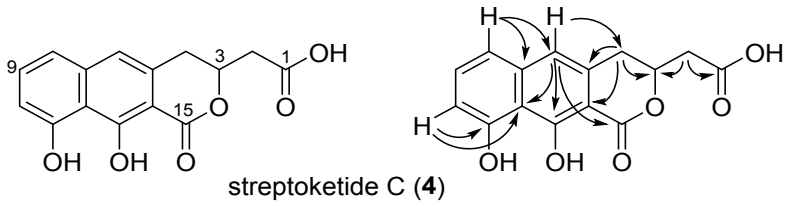
Figure 2. Structures of the known isolated compounds UWM5 (**1**), S2502 (**2a**), S2507 (**3a**) and of the new natural products streptoketide A (**2b**) and B (**3b**). Box: Selection of key COSY (bold bond) and HMBC (arrows) correlations used to elucidate the position of the ethyl substitution, exemplarily shown for **2b**.

Compound **4** (3.6 mg, 16.6. min) had a molecular mass of 287.0562 u in ESI negative mode, corresponding to a composition of $C_{15}H_{12}O_6$ in its deprotonated form (calcd. 287.0561 u) with 10 degrees of unsaturation (Figures S30-S31). The ^{1}H and ^{13}C chemical shifts and ^{1}H signal multiplicities were mostly highly similar to those of C1 to C15 of **2a/b** and **3a/b**, with largest chemical shift differences at C13 (e.g. 149.9 in **2b** versus 162.2 in **4**) and C15 (e.g. 157.4 in **2b** versus 169.7 in **4**). The typical singlet ^{1}H NMR signals of C16 and C18 of **2a/b** and **3a/b** were not present in the spectrum of **4** (Table 2). This suggested atoms C16 to C20/C21 to be absent in **4**. In addition, the retention time of **4** was similar to that of the free acids **2a** and **3a**, pointing at C1 to be a carboxylic acid function. The increase in the chemical shifts at C13 and C15 suggested C15 to be a cyclic ester function. Evaluation of the HMBC NMR data indeed corroborated these assumptions (Figures S32-S33), leading to the overall structure of the new polyketide streptoketide C (**4**) depicted in Figure 3.

Table 2. NMR spectroscopic data of **4** recorded in DMSO-d_6 at 500 MHz (^{1}H) and 125 MHz (^{13}C).

Position	Group	^{1}H (mult., *J* [Hz])	^{13}C	HMBC
1	COOH	13.04 (bs)	171.0	
2	CH_2	2.85 (dd, 16.3, 5.1) 2.77 (dd, 16.3, 7.8)	38.7*	1, 3, 4
3	CH	4.99 (m)	76.2	-
4	CH_2	3.14 (dd, 16.6, 3.3) 3.08 (dd, 16.6, 10.9)	31.8	2, 3, 5, 6, 14
5	C		133.1	
6	CH	7.16 (s)	116.3	4, 7, 8, 12, 13, 14, 15
7	C		138.9	
8	CH	7.25 (d, 7.9)	118.2	6, 7, 9, 10, 11, 13
9	CH	7.49 (t, 7.9)	131.7	7, 8, 10, 11, 12
10	CH	6.84 (d, 7.9)	110.7	8, 11, 12
11	COH	10.06 (bs)	156.5	
12	CH		112.8	
13	COH	12.57 (bs)	162.2	
14	CH		101.3	
15	CO_2R		169.7	

*: extracted from HMBC due to signal overlap with DMSO.



streptoketide C (**4**)

Figure 3. Left: structure of streptoketide C (**4**). Right: selected key HMBC correlations (arrows) used to assemble the structure of **4**.

Compounds **2-4** all bear a stereocenter at C3 with unknown configuration, also for the literature known **2a/3a**. The method of choice to elucidate the configuration at this position was the comparison of experimental ECD spectra to the corresponding calculated spectra for both possible enantiomers. These investigations were initiated with streptoketide C (**4**) with the smallest molecular structure. A comparison of the experimental ECD spectrum of **4** with that calculated for its *S*-configured enantiomer using CAM-B3LYP/def2-TZVP[32-33] revealed a satisfying match between the two curves (Figure 4, left). It has to be kept in mind that the ECD effect of **4** is comparably small and that it is a carboxylic acid. This means that the ECD can be highly influenced

by the solvent and pH. It is thus not surprising that especially the calculated rotational strengths of the first excited states do not fully fit the experiment. Nonetheless, the more pronounced ECD in the smaller wavelength region are very well reproduced and thus allow the robust determination of the absolute configuration of the stereocenter of **4** as *S*.

Above results were further corroborated by also investigating the configuration of S2507 (**3a**) following the identical approach. Comparison of the experimental ECD spectrum with that calculated for the corresponding *S*-enantiomer using CAM-B3LYP/def2-TZVP calculated again revealed *S* configuration at C3 of **3a**. Owing to the close biosynthetic relationship of **2**-**4**, the identical configuration can be conveyed to all other compounds reported on this study.

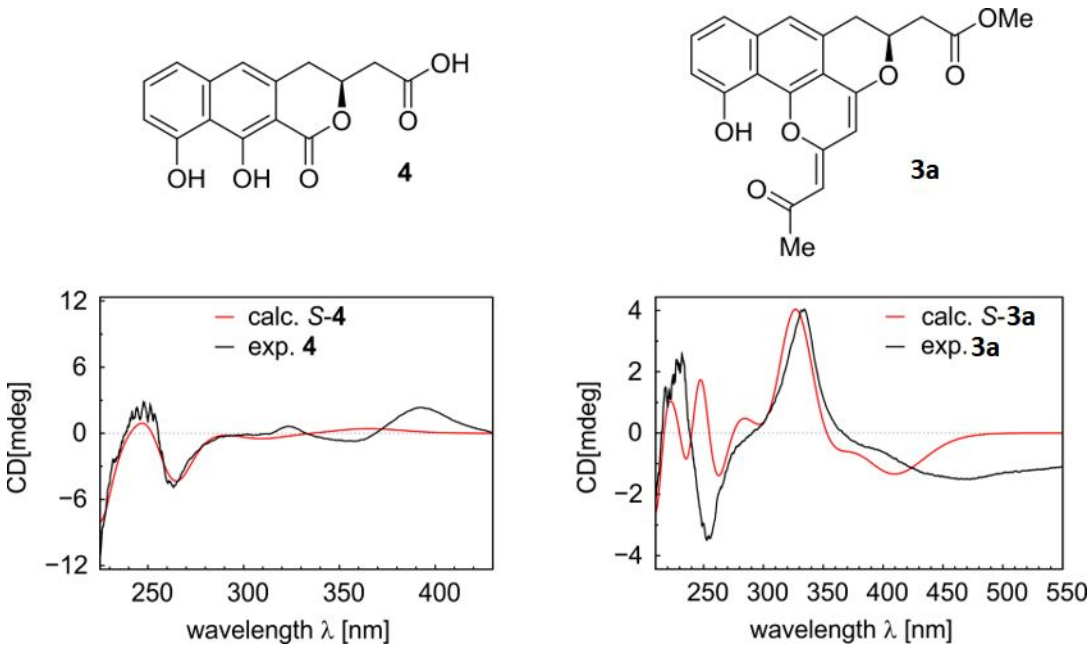


Figure 4. Comparison of experimental and calculated ECD spectra of **4** and **3a** revealing C3 configuration to be *S*.

Proposed *skt* biosynthetic pathway

The *skt* gene cluster encodes many proteins homologous to the oxytetracycline (*oxy*) and SF2575 (*ssf*) gene clusters, including genes for the minimal PKSs: KS_{α} (*skt3*), KS_{β} (*skt2*) and ACP (*skt1*);

the cyclases (*skt5*, *skt10*, *skt15*), the C-9 reductase (*skt9*) and other reductases or oxygenases (Table 1).[20, 26] However, the gene encoding the amidotransferase OxyD in the biosynthetic pathways of tetracyclines is not present. The amidotransferase is responsible for producing the malonamyl starter unit unique of the tetracyclines. In the absence of OxyD, the *oxy* PKS is initiated by an acetate primer.[20] The absence of this gene in *skt* is thus consistent with the structures of all products isolated in this study, which all derive from either acetyl- or propionyl-CoA starter untis.[34] The ability to incorporate both C2 and C3 starter units is well studied in the daunorubicin biosynthetic pathway, in which a PKS III type enzyme DpsC confers starter unit fidelity and the acyltransferase DpsD orchestrates propionate selection.[35-38] Encoded within the *skt* gene cluster, Skt17 indeed shows high similarity to DpsC (58% sequence identity) and Skt18 has high similarity to DpsD (58% sequence identity). Skt18 possesses 48% identity to the acyltransferase SsfV, which is proposed to be involved in initiation of SF2575 biosynthesis with a malonamyl starter unit.[26] Combining the sequence identity and structure, it is thus likely that Skt17 and Skt18 are responsible for priming the PKS system with a propionyl-CoA starter unit to yield the **1/2b/3b** backbone (21-carbon). The occurrence of **2a/3a** can thus be explained by a relaxed substrate selectivity of Skt17/Skt18 enabling alternative initiation with an acetyl-CoA starter unit, resulting in a 20-carbon chain product. This is also consistent with the isolation of UWM5 (**1**, 21-carbon) and SEK43 (20-carbon) from the doxorubicin pathway engineered strains.[34] Skt19 is predicted to be a methylmalonyl-CoA carboxyltransferase, with a possible role of providing propionyl-CoA from methylmalonyl-CoA. Compounds **2a** and **3a** were also previously isolated from an engineered *S. lividans* TK24 by combination of anthracycline and actinorhodin biosynthetic genes.[31, 39] Total synthesis of **2a/3a** indicated that the second pyran ring can be formed easily by a non-enzymatic cyclization/dehydration.[40] Based on this analysis, we propose the following overall

biosynthetic pathway to the streptoketides (Figure 5). The *skt* minimal PKSs (Skt1, Skt2, Skt3) together with Skt17, Skt18, Skt19 initially affords the nascent polyketide chain **5a/b**. Then, Skt9 - which is 66% and 70% identical to the reductases OxyJ and SsfU, respectively - regioselectively reduces the backbone at C-9 to yield **6a/b**. Initial cyclization to **7a/b** is likely directed by Skt10, which shares high similarity to OxyK that catalyzes the first cyclization in oxytetracyclin biosynthesis.[20] **7b** can undergo spontaneous non-enzymatic cyclization to **1**. Controlled cyclization and reduction chemistry catalyzed by cyclases Skt5/Skt15 and reductases Skt12/Skt13/Skt16, respectively, paves the way to **8a/b** and **9a/b**. Hemiacetal formation in **9a/b** delivers intermediate **10a/b**, which upon oxidative *C*,*C*-bond cleavage yields streptoketide C (**4**). Alternatively, dehydration gives **11a/b**, which upon cyclization/dehydration yield **2a/2b**. The latter can be further transformed to **3a/3b**. The BGC encodes a putative methyl transferase, Skt8, which could be involved in this transformation. However, only very limited amounts of **3a**/**3b** can be detected in the raw extract of the recombinant production strain (cf. Figure 1D). In addition, these compounds can be formed from **2a**/**2b** upon dissolving in MeOH during isolation(cf. Figure S34). As the raw extract was dissolved in MeOH and fractionated on Sephadex using MeOH as the eluent, the majority of the isolated **3a**/**3b** is hence likely formed by non-enzymatic methyl ester formation.

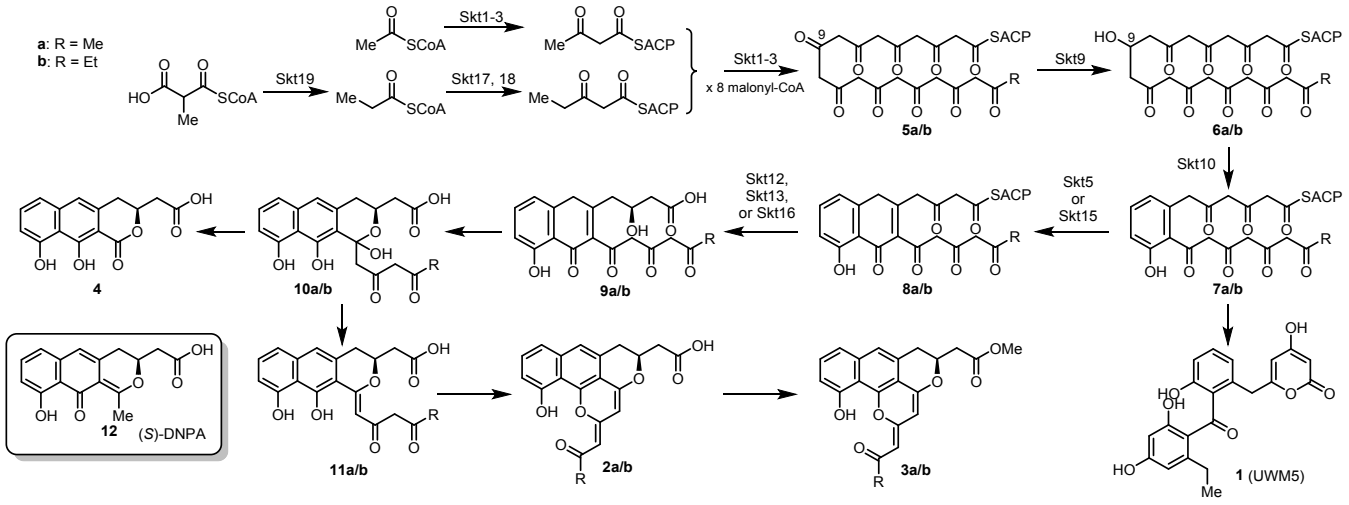


Figure 5. Proposed biosynthetic pathway leading to the isolated polyketides **1-4**.

Anti-viral tests of compounds 1-4

As the previously known compounds also isolated within this study, S2502 (**2a**) and S2507 (**3a**) were reported to have biological activity against adeno-, cytomegalo-, herpes simplex and influenza B viruses at 1 μM test concentration,[40] we were interested in evaluating the inhibitory potential of **1-4** against the human immunodeficiency virus type 1 (HIV-1). Therefore, we performed an EASY-HIT assay[41] using the reporter cell line LC5-RIC and a wild-type HIV-1_{LAI} strain. The LC5-RIC cell line contains a reporter construct encoding the fluorochrome DsRed1 that is expressed upon HIV infection. Antiviral activity of compounds is determined by measuring the intensities of fluorescent signals of HIV-inoculated cultures.[41] We treated HIV inoculated cells with serial dilutions of **1-4** up to a concentration of 50 μM and checked for effects on cell viability conducting a CellTiter Blue® Assay.[42] **2a**, **3a**, **3b** and **4** showed anti-HIV activity, with **4** as the most potent compound, inhibiting HIV infection at a 50% effective concentration (EC_{50}) of 17.3 μM. Neither **1** nor **2b** inhibited HIV infection at concentrations up to 50 μM (Figure 6A). There were no signs of negative cellular impacts during those assays (Figure 6B). Both **2a** and **3a** have anti-viral activities against HIV, but these are weaker than the antiviral effects reported for these compounds against other viruses.[40] Parallel testing of emtricitabine in the EASY-HIT assay yielded an EC_{50} value of 0.7 μM. The difference between the antiviral activities of compounds **2a**, **3a**, **3b**, **4** and this clinical drug is not surprising, since emtricitabine is the result of multiple rounds of optimization.[43]

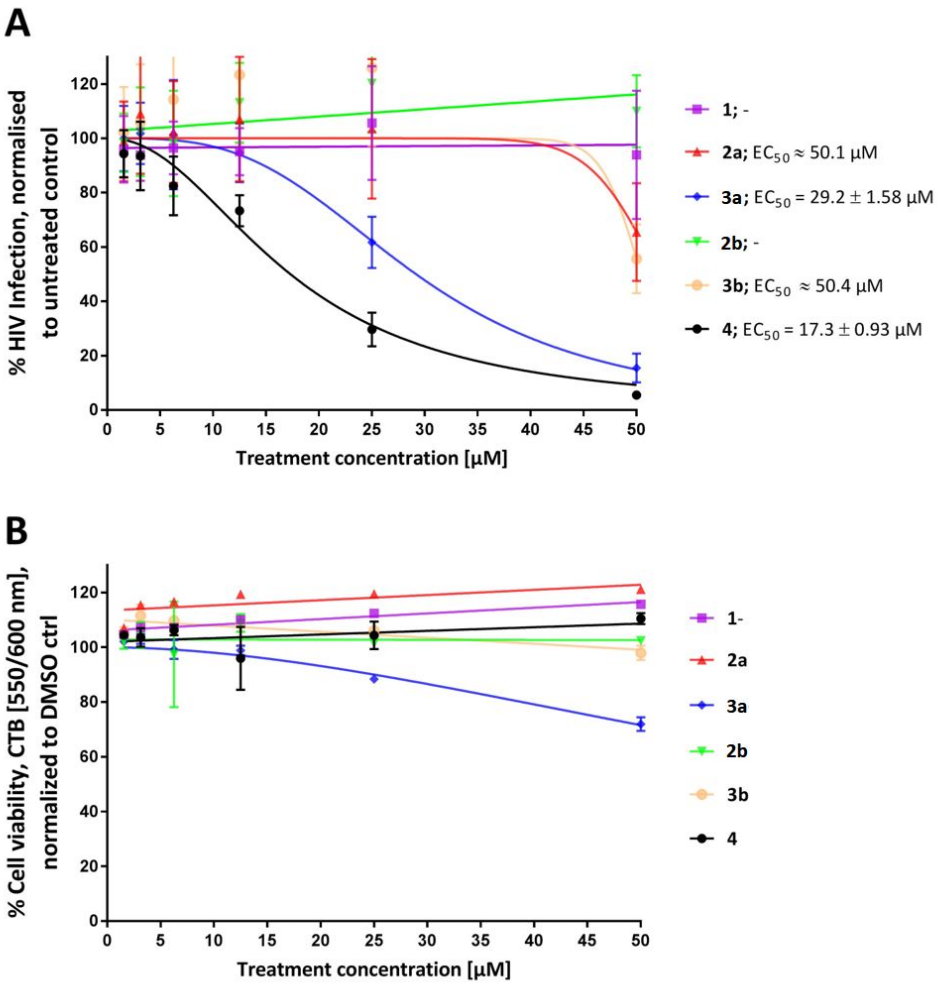


Figure 6. Inhibitory activities of **1-4** against HIV-1 infection. HIV inoculated cells were treated with serial dilutions of **1-4** up to a concentration of 50 µM and checked for effects on (**A**) anti-HIV activity and (**B**) cell viability. Compound effects were evaluated in LC5-RIC cells exposed to HIV-1_{LAI}. Shown are the means and standard deviations (±SD) of three independent experiments with triplicates (*n*=3; m=3).

In conclusion, we have directly captured the type II PKS gene cluster *skt* from *Streptomyces* sp. Tü6314 using the pSET152 vector using the LLHR method. After the rapid heterologous expression in *Streptomyces coelicolor*, we isolated and characterized three known (**1**, **2a**, **3a**) and three new natural products streptoketides A-C (**2b**, **3b**, **4**) and determined the absolute configurations of **2-4** for the first time. Compounds **2a**, **3a**, **3b** and **4** showed anti-HIV activity,

with **4** having the most pronounced effects with an EC_{50} value of 17.3 μM. While this activity is >10 fold weaker than that of the established antiviral drug emtricitabine, there have been no deleterious effects of the compounds observed within our test system, making their structures interesting starting points for further investigations.

Interestingly, during the course of this work, Liu et al. reported a strategy of direct cloning and heterologous expression of natural product biosynthetic gene cluster in *Bacillus subtilis* via Red/ET recombineering.[29] Both their work and ours share the same idea of using a strain specific vector to bypass the requirement of further genetic modifications after LLHR. In this way, heterologous expression of BGCs in a host from the same or closely related species can be simplified.[29] Our study thus contributes to the application of heterologous expression techniques for the efficient production of new natural products encoded by cryptic biosynthetic gene cluster.[4, 10]

EXPERIMENTAL SECTION

Materials and Methods

Strain *Streptomyces* sp. Tü6314

Streptomyces sp. Tü6314 was isolated from a garden soil collected at Egerpatak, Romania. The bacterium was examined for a number of key properties known to be of value in streptomycete systematics. The presence of LL-diaminopimelic acid in the peptidoglycan together with its colonial characteristics and partial sequencing of the 16S rRNA gene allowed its assignment to the genus *Streptomyces*.

Bacterial strains and plasmids

Bacterial strains and plasmids that were used are listed in Table S1. All *E. coli* strains were cultured in Luria-Bertani (LB) liquid or on LB agar media at 37°C. *Streptomyces* strains were maintained on MS agar (20.0 g/L mannitol, 20.0 g/L soya flour, 20.0 g/L agar) at 30°C or cultivated in YMG liquid medium (4.0 g/L yeast extract, 10.0 g/L malt extract, 4.0 g/L glucose, adjusted to pH 7.2 using 1 M KOH) at 28°C with constant shaking at 200 rpm. Liquid ISP Medium 4 (10.0 g/L soluble starch, 1.0 g/L K_2HPO_4, 1.0 g/L $MgSO_4 \cdot 7H_2O$, 1.0 g/L NaCl, 2.0 g/L $(NH_4)_2SO_4$, 2.0 g/L $CaCO_3$, 1.0 mL trace salt solution) was used for large-scale fermentation. The trace salt solution (1.0 g/L $FeSO_4 \cdot 7H_2O$, 1.0 g/L $MnCl_2 \cdot 4H_2O$, 1 g/L $ZnSO_4 \cdot 7H_2O$) was sterile filtered. Apramycin (apr, 30 μg/mL), kanamycin (kan, 50 μg/mL) and chloramphenicol (cm, 15 μg/mL) were added to the media as required. For conjugation between *E. coli* and *Streptomyces*, apr (30 μg/mL) and nalidixic acid (NA, 50 μg/mL) were added.

Bacterial genomic DNA isolation.

The *Streptomyces* genomic DNA isolation was slightly modified from the method described by Wang et al..[18] In brief, *Streptomyces* sp. Tü6314 was cultured in 50 mL YMG medium at 28°C for three days. After centrifugation, the cells were resuspended in 4.5 mL of Solution 1 (10% sucrose [w/v], 50 mM Tris-HCl, pH 8.0, 10 mM EDTA). Then 500 μL of 30 mg/mL lysozyme were added and the mixture was incubated for 1 h at 37°C with occasional inverting. After adding 1.5 mL of 3.3% SDS and 100 μL of Proteinase K (10 mg/mL), the tube was mixed by gentle inversion and incubated at 37°C for at least 1 hour until the solution became clear. The solution was first combined with 2 mL 6 M NaCl and then 1 volume of phenol-chloroform-isoamyl alcohol (25:24:1) was added to create an emulsion. After centrifugation, the aqueous phase was carefully transferred

to a new tube using a wide-bore pipette tip. The DNA was precipitated by adding 0.1 volume of 3 M sodium acetate (pH 5.4), followed by adding 2.5 volumes of absolute ethanol and gently inverting to precipitate DNA. The DNA was dissolved in 4 mL TE buffer and 50 µL RNase A (4 mg/mL) was added. The DNA was precipitated again with ethanol and washed twice with 70% (v/v) ethanol, dried and dissolved in 500 µL 10 mM Tris-HCl, pH 8.0.

Phylogenic tree construction

The phylogenic tree was constructed using the MEGA X program based on ClustalW alignment using the Neighbor-Joining method.[44] The protein sequences were downloaded from the MIBiG database.[25]

Capture of the *skt* gene cluster using LLHR

*Nhe*I restriction sites flanking the proposed *skt* cluster were identified and chosen to release the gene cluster from the genomic DNA (Figure 1A). Approximately 50 µg of the extracted genomic DNA was digested with 50 U of *Nhe*I at 37°C overnight. The digested genomic DNA was purified by ethanol precipitation. The resulting DNA pellet was dissolved in 25 µL ddH_2O.

The linear capturing vector was PCR amplified from the pSET152 backbone using Q5 High-Fidelity DNA polymerase with primer pairs pSET152-cap_cluster21-F/R. These primers were designed to include 50-bp homology arms targeting the flanking region of the *skt* gene cluster, slightly inside the *Nhe*I digested position (Figure 1A-B). The online program Oligo Analyzer (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) was used to find optimal primers with minimum hair pin/dimer formation. The PCR annealing temperature was estimated using the NEB Tm calculator (http://tmcalculator.neb.com/#!/main). PCR conditions were as

follows: 98°C 1 min; 32 cycles of 98°C 10 s, 58°C 20 s, 72°C 5 min; 72°C 10 min. After PCR amplification, *Dpn*I was used to digest the template DNA, followed by agarose gel purification.

LLHR was performed as described in the literature with several modifications.[12, 18] In brief, 300 µL of overnight cultured *E. coli* GB05-dir was inoculated into 15 mL fresh LB medium without antibiotics. The cells were grown at 37°C, 200 rpm for ~1.5 h until OD_{600} reached ~0.3. After adding 300 µL of 1 M L-(+)-arabinose, the cells were grown at 37°C at 200 rpm for 45 min until the OD_{600} reached 0.6-0.8. Then, 1.5 mL cells were transferred into 1.5 mL ice-cold Eppendorf tubes and centrifuged for 30 s at 8,000 *g* at 2°C. The supernatant was discarded and the cell pellet was washed using 1 mL ice-cold water. The washing step was repeated once. Then the cells were resuspended in 30 µL ice-cold water. A total of 5 µg of purified *Nhe*I digested *Streptomyces* sp. Tü6314 genomic DNA and ~500 ng of PCR amplified pSET152 capturing vector were added. Transformation into *E. coli* was achieved by electroporation using ice-cold cuvettes and a BIO-RAD MicroPulser (program E2 for bacteria). Electroporated cells were incubated at 37°C for 120 min while shaking and then spread on LB agar plates supplemented with 30 µg/mL apramycin. Colonies were screened by PCR with the primer pairs pSET152_cap_seq-F/cap_cluster21_verification_L-R. The primers used in this study are listed in Table S2.

Heterologous expression of the *skt* gene cluster

The captured *skt* gene cluster was transformed into the donor strain *E. coli* ET12567/pUZ8002 and then transferred to *S. coelicolor* M1152 or *S. coelicolor* M1154 by intergeneric conjugation as described previously, with minor modifications.[45] A single colony of the donor strain was inoculated into 3 mL LB supplemented with kan/cm/apr and grown at 37°C overnight. A 100 µL

overnight culture was used to inoculate 10 mL fresh LB with kan/cm/apr and grown at 37°C until the OD_{600} reached 0.6-0.8. The grown *E. coli* cells were centrifuged and washed twice with an equal volume of fresh LB. Spores from the host strains were heat activated at 50°C for 10 min in 1 mL YMG medium followed by cooling using tap water and then mixed with the washed donor cells. The mating mixture was then spread on MS agar plates and incubated at 30°C. After 20 h, the MS agar plates were overlaid with 1 mg/plate of NA and apr. Five days later, the Apr^R exconjugants were picked and grown on MS agar containing NA and apr, resulting in single colonies. The empty pSET152 vector was also introduced into the host strains and used as a control. One well-grown single colony was inoculated 100 mL YMG medium and grown at 28°C 200 rpm for 7 days before High Performance Liquid Chromatography (HPLC) analysis. The culture was centrifuged and the supernatant was extracted with ethyl acetate. The crude extracts were dissolved in methanol for HPLC analyses. HPLC was performed on a Jasco HPLC system (UV-1575 Intelligent UV/VIS Detector, DG-2080-53 3-Line Degaser, two PU-1580 Intelligent HPLC Pumps, AS-1550 Intelligent Sampler, HG-1580-32 Dynamic Mixer, Galaxie-Chromatography-Software) with a Eurospher II 100-3 C18 A (150 × 4.6 mm) column purchased from Knauer (Germany). The eluent system consisted of: A = H_2O + 0.05% TFA and B = acetonitrile + 0.05% TFA. The analytical method consisted of a linear gradient: 0-2 min 5% B, 2-30 min 5% to 95% B, 30-35 min 95% B, 35.2-38 min 5% B with a flow rate at 1 mL/min, 25°C.

Compound extraction and isolation.

For product isolation, one well-grown *S. coelicolor* M1154/*skt* single colony was inoculated 100 mL YMG medium and grown at 28°C 200 rpm for 5 days to make a seed culture. The seed culture was used to inoculate 4 × 1 L ISP Medium-4 in four 3-L flasks to a final concentration of 1% (v/v)

and grown at 28°C 200 rpm for 7 days. After fermentation, the cultures were filtered using filter paper and the liquid fraction was adjusted to pH 4 using HCl. The supernatant was extracted three times with an equal volume of ethyl acetate. The cell pellet was extracted four times with 1 L acetone until the solvent became colorless. The organic extracts were combined and dried under reduced pressure to obtain the crude extracts. The crude extracts were dissolved in 50 mL methanol and subjected to a 90 × 3 cm Sephadex LH-20 column using methanol for elution. The eluted fractions were checked by HPLC and fractions containing products not present in the unmodified host strain were collected, followed by purification using a Jasco semi-preparative HPLC system (UV-1575 Intelligent UV/VIS Detector, two PU-2068 Intelligent prep Pumps, a MIKA 1000 Dynamic Mixing Chamber, 1000 µL Portmann Instruments AG Biel-Benken, a LC-NetII / ADC, a Rheodyne injection valve, Galaxie-Chromatography-Software), with a Eurospher II 100-5 C18 A (250 × 16 mm) column with precolumn (30 × 16 mm) purchased from Knauer (Germany). The eluent system consisted of: A = H_2O + 0.05% TFA and B = acetonitrile + 0.05% TFA, with a linear gradient: 0-43 min 25-50% B with a flow rate at 12 mL/min at room temperature. After preparative separation, the fractions containing the desired product were combined and the acetonitrile was removed under reduced pressure. The remaining aqueous phases were frozen in liquid nitrogen and the water was removed by lyophilization (Alpha 2-4 Christ with Chemistry-Hybrid-Pump-RC6 pump). UWM5 (**1**): light-brown amorphous solid (yield 0.9 mg/L, t_R = 14.2 min). ^{1}H and ^{13}C NMR data, see Table S4. HPLC and UV data, see Figure S5. HRMS (ESI-TOF) *m/z*: $[M+H]^+$ calcd. for $C_{21}H_{19}O_7$ 383.1125 (see Figure S6); found 383.1122. The spectroscopic data is in agreement with the literatue.[30] S2502 (**2a**): dark-brown amorphous solid (yield 0.2 mg/L, t_R = 16.8 min). ^{1}H and ^{13}C NMR data, see Table S5 and S6. HPLC and UV data, see Figure S10. HRMS (ESI-TOF) *m/z*: $[M+H]^+$ calcd. for $C_{20}H_{17}O_6$ 353.1020 (see Figure S11); found 353.1017. The

spectroscopic data is in agreement with the literatue.[31, 40] S2507 (**3a**): dark-brown amorphous solid (yield 0.7 mg/L, t_R = 19.3 min). ^{1}H and ^{13}C NMR data, see Table S5 and S6. HPLC and UV data, see Figure S15. HRMS (ESI-TOF) *m/z*: $[M+H]^+$ calcd. for $C_{21}H_{19}O_6$ 367.1176 (see Figure S16); found 353.1174. The spectroscopic data is in agreement with the literatue. [31, 40] Streptoketide A (**2b**): dark-brown amorphous solid (yield 0.7 mg/L, t_R = 17.9 min). ^{1}H and ^{13}C NMR data, see Table S9. HPLC and UV data, see Figure S20. HRMS (ESI-TOF) *m/z*: $[M+H]^+$ calcd. for $C_{21}H_{19}O_6$ 367.1176 (see Figure S21); found 353.1174. Streptoketide B (**3b**): dark-brown white amorphous solid (yield 1.7 mg/L, t_R = 20.4 min). ^{1}H and ^{13}C NMR data, see Table S9. HPLC and UV data, see Figure S22. HRMS (ESI-TOF) *m/z*: $[M+H]^+$ calcd. for $C_{22}H_{21}O_6$ 381.1333 (see Figure S23); found 353.1329. Streptoketide C (**4**): pale beige color white amorphous solid (yield 0.9 mg/L, t_R = 16.6 min). ^{1}H and ^{13}C NMR data, see Table 2. HPLC and UV data, see Figure S30. HRMS (ESI-TOF) *m/z*: $[M-H]^-$ calcd. for $C_{15}H_{11}O_6$ 287.0562 (see Figure S31); found 287.0561.

NMR data collection and CD spectra measurement

For NMR data collection, all purified compounds were dissolved in deuterated DMSO. The NMR spectra were recorded on a Bruker AVHD500 or a Bruker AV500-cryo spectrometer. The chemical shifts δ are listed as parts per million [ppm] and refer to δ(TMS) = 0. The spectra were calibrated using residual undeuterated solvent as internal reference.

CD spectra were measured using a Jasco J-715 spectropolarimeter (Jasco, Gross-Umstadt, Germany). Experiments were performed in quartz cuvettes with 0.1 cm path length at a substance concentration of 0.1 mg/ml and spectra were recorded from 200 to 600 nm in acetonitrile at 22°C. Eight to twelve spectra were accumulated for each substance and subsequently baseline corrected by subtracting the pure solvent spectrum.

Calculation of ECD spectra

The ECD spectra of **4** and **2a** were computed[46] using ORCA[47] for the optimizations and frequency calculations and Gaussian09[48] for the excited states investigations. The conformational analysis was simplified in terms of the rotational freedom of the alkyl side chains. As it is known that these do have a negligible effect on the ECD only the lowest conformation of the side chain was considered for the conformational analysis. More important is the orientation of the side chain relative to the ring (equatorial or axial). The ECD spectra of **4** and **3a** in the equatorial conformation do have mirror image like spectra when compared to the axial one. B3LYP-D3/def2-TZVP[49] calculations including the chain-of-spheres approximation yield very similar energies for both conformers of **4**, while that of **3a** do have a strong difference. According to Gibbs energies **4**-eq and **4**-ax differ only by 0.5 kcal/mol (in favour of eq). In case of **2a**, only the equatorial conformation is populated at RT (energy difference >8 kcal/mol). ECD calculations were performed with CAM-B3LYP/def2-TZVP, tests with wB97XD and B3LYP gave comparable results. To further process the results SpecDis[50] was used. For the Boltzmann weighting Gibbs free energies were utilized. Gaussians were prepared with a value of 0.26 eV and to compensate systematic errors in the calculation of the excited states a UV shift of 25 nm for **4** and of 39 nm for **3a** was applied.

HIV screening assay (EASY-HIT). The EASY-HIT assay is based on the HIV-1 susceptible reporter cell line LC5-RIC that contains a stably integrated reporter construct encoding the fluorochrome DsRed1 that will be expressed upon HIV infection and expression of the early viral

proteins Tat and Rev.[41] Briefly, 10,000 LC5-RIC cells were seeded into each well of black 96-well plates and incubated over night at 37°C, 5% CO_2. Compound stocks (50 mM in DMSO) were screened in a 1:2 series dilution at concentrations from 1.5 to 50 µM at a final DMSO concentration of 0.1% to determine IC_{50} curves. After addition of the compound, LC5-RIC cells were inoculated with HIV-1_{LAI} at a MOI of 0.5 and incubated for 48 hours at 37°C, 5% CO_2. Finally, DsRed1 reporter expression was measured using a fluorescence microplate reader at an excitation filter wavelength of 552 nm and an emission filter wavelength of 596 nm.

Cell viability assay. A CellTiter-Blue® cell viability assay from Promega was performed to check on cell viability of HIV-1_{LAI} inoculated and test compound treated LC5-RIC cultures. This assay measures the ability of metabolically active cells to convert the redox dye resazurin into resorufin, which can be detected by fluorescence spectroscopy. 10,000 LC5-RIC cells were seeded into each well of black 96-well plates and incubated over night at 37°C, 5% CO_2. Compound stocks (50 mM in DMSO) were screened in a 1:2 series dilution at concentrations from 1.5 to 50 µM at a final DMSO concentration of 0.1%. After 48 hours of incubation at 37°C, 5% CO_2 a 1:5 CTB reagent:cell culture medium mix was added to each well and incubated for another hour. Finally, fluorescence signal of resorufin was detected using a fluorescence microplate reader at an excitation filter wavelength of 550 nm and an emission filter wavelength of 600 nm.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS publication website at DOI:XX.

- Plasmids, bacterial strains and primers

- Phylogenetic analysis of ketosynthases
- Restriction analysis of constructs
- HPLC, UV, MS and NMR data of **1-4**
- Heat of formation (B3LYP/def2-TZVP), number of imaginary imaginery frequencies, Cartesian coordinates of **3a** and **4**.

Author Information.

Corresponding Author

*Tobias A. M. Gulder

E-mail: tobias.gulder@ch.tum.de, tobias.gulder@tu-dresden.de

Telephone number: +49-(0)89-289-13833

ORCID

P. M. D'Agostino: 0000-0002-8323-5416

T. Bruhn: 0000-0002-9604-1004

T. A. M. Gulder: 0000-0001-6013-3161

Postal Address

Technische Universität München

Department of Chemistry and Center for Integrated Protein Science Munich (CIPSM)

Biosystems Chemistry

Lichtenbergstraße 4

85748 Garching

Acknowledgments

We thank Prof. Dr. Youming Zhang (Shandong University) for providing the *E. coli* GB05-dir. ZQ thanks China Scholarship Council (CSC) for his PhD scholarship. PMD thanks the European

Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 745435 for funding. This work was supported by the DFG (excellence cluster CIPSM and GU1233/1-1).

References

1. Newman, D. J.; Cragg, G. M., Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **2016,** *79*, 629-661.

2. Demain, A. L., Importance of microbial natural products and the need to revitalize their discovery. *J. Ind. Microbiol. Biot.* **2014,** *41*, 185-201.

3. Shen, B., A New Golden Age of Natural Products Drug Discovery. *Cell* **2015,** *163*, 1297-1300.

4. Rutledge, P. J.; Challis, G. L., Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.* **2015,** *13*, 509-23.

5. Bentley, S. D.; Chater, K. F.; Cerdeno-Tarraga, A. M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D.; Bateman, A.; Brown, S.; Chandra, G.; Chen, C. W.; Collins, M.; Cronin, A.; Fraser, A.; Goble, A.; Hidalgo, J.; Hornsby, T.; Howarth, S.; Huang, C. H.; Kieser, T.; Larke, L.; Murphy, L.; Oliver, K.; O'Neil, S.; Rabbinowitsch, E.; Rajandream, M. A.; Rutherford, K.; Rutter, S.; Seeger, K.; Saunders, D.; Sharp, S.; Squares, R.; Squares, S.; Taylor, K.; Warren, T.; Wietzorrek, A.; Woodward, J.; Barrell, B. G.; Parkhill, J.; Hopwood, D. A., Complete genome sequence of the model actinomycete *Streptomyces* coelicolor A3(2). *Nature* **2002,** *417*, 141-147.

6. Mukherjee, S.; Seshadri, R.; Varghese, N. J.; Eloe-Fadrosh, E. A.; Meier-Kolthoff, J. P.; Goker, M.; Coates, R. C.; Hadjithomas, M.; Pavlopoulos, G. A.; Paez-Espino, D.; Yoshikuni, Y.; Visel, A.; Whitman, W. B.; Garrity, G. M.; Eisen, J. A.; Hugenholtz, P.; Pati, A.; Ivanova, N. N.; Woyke, T.; Klenk, H. P.; Kyrpides, N. C., 1,003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life. *Nat. Biotechnol.* **2017,** *35*, 676-683.

7. Kwiatkowska, M.; Parker, D.; Wiltsche, C., PRISM-Games 2.0: A Tool for Multi-objective Strategy Synthesis for Stochastic Games. *Lect. Notes Comput. Sci.* **2016,** *9636*, 560-566.

8. Blin, K.; Wolf, T.; Chevrette, M. G.; Lu, X.; Schwalen, C. J.; Kautsar, S. A.; Suarez Duran, H. G.; de Los Santos, E. L. C.; Kim, H. U.; Nave, M.; Dickschat, J. S.; Mitchell, D. A.; Shelest, E.;

Breitling, R.; Takano, E.; Lee, S. Y.; Weber, T.; Medema, M. H., antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **2017,** *45*, W36-W41.

9. Ziemert, N.; Alanjary, M.; Weber, T., The evolution of genome mining in microbes - a review. *Nat. Prod. Rep.* **2016,** *33*, 988-1005.

10. Scherlach, K.; Hertweck, C., Triggering cryptic natural product biosynthesis in microorganisms. *Org. Biomol. Chem.* **2009,** *7*, 1753-1760.

11. Gomez-Escribano, J. P.; Bibb, M. J., Engineering *Streptomyces* coelicolor for heterologous expression of secondary metabolite gene clusters. *Microb. Biotechnol.* **2011,** *4*, 207-15.

12. Fu, J.; Bian, X.; Hu, S.; Wang, H.; Huang, F.; Seibert, P. M.; Plaza, A.; Xia, L.; Muller, R.; Stewart, A. F.; Zhang, Y., Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.* **2012,** *30*, 440-6.

13. Yamanaka, K.; Reynolds, K. A.; Kersten, R. D.; Ryan, K. S.; Gonzalez, D. J.; Nizet, V.; Dorrestein, P. C.; Moore, B. S., Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci.* **2014,** *111*, 1957-62.

14. Greunke, C.; Duell, E. R.; D'Agostino, P. M.; Glockle, A.; Lamm, K.; Gulder, T. A. M., Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic potential. *Metab. Eng.* **2018,** *47*, 334-345.

15. D'Agostino, P. M.; Gulder, T. A. M., Direct Pathway Cloning Combined with Sequence- and Ligation-Independent Cloning for Fast Biosynthetic Gene Cluster Refactoring and Heterologous Expression. *ACS Synth. Biol.* **2018,** *7*, 1702-1708.

16. Yin, J.; Hoffmann, M.; Bian, X.; Tu, Q.; Yan, F.; Xia, L.; Ding, X.; Stewart, A. F.; Muller, R.; Fu, J.; Zhang, Y., Direct cloning and heterologous expression of the salinomycin biosynthetic gene cluster from *Streptomyces* albus DSM41398 in *Streptomyces* coelicolor A3(2). *Sci. Rep.* **2015,** *5*, 15081.

17. Wang, H.; Li, Z.; Jia, R.; Yin, J.; Li, A.; Xia, L.; Yin, Y.; Müller, R.; Fu, J.; Stewart, A. F., ExoCET: exonuclease in vitro assembly combined with RecET recombination for highly efficient direct DNA cloning from complex genomes. *Nucleic Acids Res.* **2017,** *46*, e28-e28.

18. Wang, H.; Li, Z.; Jia, R.; Hou, Y.; Yin, J.; Bian, X.; Li, A.; Muller, R.; Stewart, A. F.; Fu, J.; Zhang, Y., RecET direct cloning and Redalphabeta recombineering of biosynthetic gene clusters, large operons or single genes for heterologous expression. *Nat. Protoc.* **2016,** *11*, 1175-90.

19. Bierman, M.; Logan, R.; Obrien, K.; Seno, E. T.; Rao, R. N.; Schoner, B. E., Plasmid Cloning Vectors for the Conjugal Transfer of DNA from Escherichia-Coli to *Streptomyces* Spp. *Gene* **1992,** *116*, 43-49.

20. Pickens, L. B.; Tang, Y., Oxytetracycline biosynthesis. *J. Biol. Chem.* **2010,** *285*, 27509-15.

21. Zhu, T.; Cheng, X.; Liu, Y.; Deng, Z.; You, D., Deciphering and engineering of the final step halogenase for improved chlortetracycline biosynthesis in industrial *Streptomyces* aureofaciens. *Metab. Eng.* **2013,** *19*, 69-78.

22. Aubelsadron, G.; Londosgagliardi, D., Daunorubicin and Doxorubicin, Anthracycline Antibiotics, a Physicochemical and Biological Review. *Biochimie* **1984,** *66*, 333-352.

23. Shen, B., Biosynthesis of aromatic polyketides. *Top. Curr. Chem.* **2000,** *209*, 1-51.

24. Blin, K.; Wolf, T.; Chevrette, M. G.; Lu, X. W.; Schwalen, C. J.; Kautsar, S. A.; Duran, H. G. S.; Santos, E. L. C. D. L.; Kim, H. U.; Nave, M.; Dickschat, J. S.; Mitchell, D. A.; Shelest, E.; Breitling, R.; Takano, E.; Lee, S. Y.; Weber, T.; Medema, M. H., antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* **2017,** *45*, W36-W41.

25. Medema, M. H.; Kottmann, R.; Yilmaz, P.; Cummings, M.; Biggins, J. B.; Blin, K.; de Bruijn, I.; Chooi, Y. H.; Claesen, J.; Coates, R. C.; Cruz-Morales, P.; Duddela, S.; Dusterhus, S.; Edwards, D. J.; Fewer, D. P.; Garg, N.; Geiger, C.; Gomez-Escribano, J. P.; Greule, A.; Hadjithomas, M.; Haines, A. S.; Helfrich, E. J. N.; Hillwig, M. L.; Ishida, K.; Jones, A. C.; Jones, C. S.; Jungmann, K.; Kegler, C.; Kim, H. U.; Kotter, P.; Krug, D.; Masschelein, J.; Melnik, A. V.; Mantovani, S. M.; Monroe, E. A.; Moore, M.; Moss, N.; Nutzmann, H. W.; Pan, G. H.; Pati, A.; Petras, D.; Reen, F. J.; Rosconi, F.; Rui, Z.; Tian, Z. H.; Tobias, N. J.; Tsunematsu, Y.; Wiemann, P.; Wyckoff, E.; Yan, X. H.; Yim, G.; Yu, F. G.; Xie, Y. C.; Aigle, B.; Apel, A. K.; Balibar, C. J.; Balskus, E. P.; Barona-Gomez, F.; Bechthold, A.; Bode, H. B.; Borriss, R.; Brady, S. F.; Brakhage, A. A.; Caffrey, P.; Cheng, Y. Q.; Clardy, J.; Cox, R. J.; De Mot, R.; Donadio, S.; Donia, M. S.; van der Donk, W. A.; Dorrestein, P. C.; Doyle, S.; Driessen, A. J. M.; Ehling-Schulz, M.; Entian, K. D.; Fischbach, M. A.; Gerwick, L.; Gerwick, W. H.; Gross, H.; Gust, B.; Hertweck, C.; Hofte, M.; Jensen, S. E.; Ju, J. H.; Katz, L.; Kaysser, L.; Klassen, J. L.; Keller, N. P.; Kormanec, J.; Kuipers, O. P.; Kuzuyama, T.; Kyrpides, N. C.; Kwon, H. J.; Lautru, S.; Lavigne, R.; Lee, C. Y.; Linquan, B.; Liu, X. Y.; Liu, W.; Luzhetskyy, A.; Mahmud, T.; Mast, Y.; Mendez, C.; Metsa-Ketela, M.; Micklefield, J.; Mitchell, D. A.; Moore, B. S.; Moreira, L. M.; Muller, R.; Neilan, B. A.; Nett, M.;

Nielsen, J.; O'Gara, F.; Oikawa, H.; Osbourn, A.; Osburne, M. S.; Ostash, B.; Payne, S. M.; Pernodet, J. L.; Petricek, M.; Piel, J.; Ploux, O.; Raaijmakers, J. M.; Salas, J. A.; Schmitt, E. K.; Scott, B.; Seipke, R. F.; Shen, B.; Sherman, D. H.; Sivonen, K.; Smanski, M. J.; Sosio, M.; Stegmann, E.; Sussmuth, R. D.; Tahlan, K.; Thomas, C. M.; Tang, Y.; Truman, A. W.; Viaud, M.; Walton, J. D.; Walsh, C. T.; Weber, T.; van Wezel, G. P.; Wilkinson, B.; Willey, J. M.; Wohlleben, W.; Wright, G. D.; Ziemert, N.; Zhang, C. S.; Zotchev, S. B.; Breitling, R.; Takano, E.; Glockner, F. O., Minimum Information about a Biosynthetic Gene cluster. *Nat. Chem. Biol.* **2015,** *11*, 625-631.

26. Pickens, L. B.; Kim, W.; Wang, P.; Zhou, H.; Watanabe, K.; Gomi, S.; Tang, Y., Biochemical analysis of the biosynthetic pathway of an anticancer tetracycline SF2575. *J. Am. Chem. Soc.* **2009,** *131*, 17677-89.

27. Metsa-Ketela, M.; Halo, L.; Munukka, E.; Hakala, J.; Mantsala, P.; Ylihonko, K., Molecular evolution of aromatic polyketides and comparative sequence analysis of polyketide ketosynthase and 16S ribosomal DNA genes from various *Streptomyces* species. *Appl. Environ. Microb.* **2002,** *68*, 4472-4479.

28. Feng, Z.; Kallifidas, D.; Brady, S. F., Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites. *Proc. Natl. Acad. Sci.* **2011,** *108*, 12629-12634.

29. Liu, Q.; Shen, Q.; Bian, X.; Chen, H.; Fu, J.; Wang, H.; Lei, P.; Guo, Z.; Chen, W.; Li, D.; Zhang, Y., Simple and rapid direct cloning and heterologous expression of natural product biosynthetic gene cluster in Bacillus subtilis via Red/ET recombineering. *Sci. Rep.* **2016,** *6*, 34623.

30. Lomovskaya, N.; Doi-Katayama, Y.; Filippini, S.; Nastro, C.; Fonstein, L.; Gallo, M.; Colombo, A. L.; Hutchinson, C. R., The *Streptomyces* peucetius dpsY and dnrX genes govern early and late steps of daunorubicin and doxorubicin biosynthesis. *J. Bacteriol.* **1998,** *180*, 2379-2386.

31. Kunnari, T.; Kantola, J.; Ylihonko, K.; Klika, K. D.; Mantsala, P.; Hakala, J., Hybrid compounds derived from the combination of anthracycline and actinorhodin biosynthetic pathways. *J. Chem. Soc. Perk. Trans. 2* **1999**, 1649-1652.

32. Yanai, T.; Tew, D. P.; Handy, N. C., A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phy. Lett.* **2004,** *393*, 51-57.

33. Weigend, F.; Ahlrichs, R., Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005,** *7*, 3297-3305.

34. Moore, B. S.; Hertweck, C., Biosynthesis and attachment of novel bacterial polyketide synthase starter units. *Nat. Prod. Rep.* **2002,** *19*, 70-99.

35. Jackson, D. R.; Shakya, G.; Patel, A. B.; Mohammed, L. Y.; Vasilakis, K.; Wattana-Amorn, P.; Valentic, T. R.; Milligan, J. C.; Crump, M. P.; Crosby, J.; Tsai, S. C., Structural and Functional Studies of the Daunorubicin Priming Ketosynthase DpsC. *ACS Chem. Biol.* **2018,** *13*, 141-151.

36. Rajgarhia, V. B.; Priestley, N. D.; Strohl, W. R., The product of dpsC confers starter unit fidelity upon the daunorubicin polyketide synthase of *Streptomyces* sp. strain C5. *Metab. Eng.* **2001,** *3*, 49-63.

37. Bao, W.; Sheldon, P. J.; Hutchinson, C. R., Purification and properties of the *Streptomyces* peucetius DpsC beta-ketoacyl:acyl carrier protein synthase III that specifies the propionate-starter unit for type II polyketide biosynthesis. *Biochemistry* **1999,** *38*, 9752-7.

38. Bao, W.; Sheldon, P. J.; Wendt-Pienkowski, E.; Hutchinson, C. R., The *Streptomyces* peucetius dpsC gene determines the choice of starter unit in biosynthesis of the daunorubicin polyketide. *J. Bacteriol.* **1999,** *181*, 4690-5.

39. Kantola, J.; Kunnari, T.; Hautala, A.; Hakala, J.; Ylihonko, K.; Mantsala, P., Elucidation of anthracyclinone biosynthesis by stepwise cloning of genes for anthracyclines from three different *Streptomyces* spp. *Microbiology* **2000,** *146 (Pt 1)*, 155-63.

40. Krohn, K.; Vukics, K., First chemical synthesis of the antiviral agents S2502 and S2507. *Synthesis-Stuttgart* **2007**, 2894-2900.

41. Kremb, S.; Helfer, M.; Heller, W.; Hoffmann, D.; Wolff, H.; Kleinschmidt, A.; Cepok, S.; Hemmer, B.; Durner, J.; Brack-Werner, R., EASY-HIT: HIV full-replication technology for broad discovery of multiple classes of HIV inhibitors. *Antimicrob. Agents Chemother.* **2010,** *54*, 5257-5268.

42. Riss, T. L.; Moravec, R. A.; Niles, A. L.; Duellman, S.; Benink, H. A.; Worzella, T. J.; Minor, L., Cell viability assays. In *Assay Guidance Manual [Internet]*, Eli Lilly & Company and the National Center for Advancing Translational Sciences: **2016**.

43. Liotta, D. C.; Painter, G. R., Discovery and development of the anti-human immunodeficiency virus drug, emtricitabine (Emtriva, FTC). *Accounts Chem. Res.* **2016,** *49*, 2091-2098.

44. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K., MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018,** *35*, 1547-1549.

45. Wang, X. K.; Jin, J. L., Crucial factor for increasing the conjugation frequency in *Streptomyces* netropsis SD-07 and other strains. *FEMS Microbiol. Lett.* **2014,** *357*, 99-103.

46. Pescitelli, G.; Bruhn, T., Good computational practice in the assignment of absolute configurations by TDDFT calculations of ECD spectra. *Chirality* **2016,** *28*, 466-474.

47. Neese, F., Software update: the ORCA program system, version 4.0. *Wiley Interdiscip. Rev.-Comput. Mol. Sci.* **2018,** *8*, e1327.

48. Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G., Gaussian09 Revision D. 01, Gaussian Inc. Wallingford CT. *URL:* http://www.*gaussian. com* **2009**.

49. Grimme, S.; Ehrlich, S.; Goerigk, L., Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011,** *32*, 1456-1465.

50. Bruhn, T.; Schaumlöffel, A.; Hemberger, Y.; Pescitelli, G., SpecDis version 1.71; Berlin, Germany. 2017. *URL:* https://specdis-software.jimdo.com **2017**.