ChemPubSoc Europe

DOI: 10.1002/cbic.201300732

Identification of Fluorinases from *Streptomyces* sp MA37, *Norcardia brasiliensis*, and *Actinoplanes* sp N902-109 by Genome Mining

Hai Deng,*[a] Long Ma,[b] Nouchali Bandaranayaka,[b] Zhiwei Qin,[a] Greg Mann,[b] Kwaku Kyeremeh,[d] Yi Yu,[c] Thomas Shepherd,[e] James H. Naismith,[b] and David O'Hagan*[b]

The fluorinase is an enzyme that catalyses the combination of *S*-adenosyl-L-methionine (SAM) and a fluoride ion to generate 5′-fluorodeoxy adenosine (FDA) and L-methionine through a nucleophilic substitution reaction with a fluoride ion as the nucleophile. It is the only native fluorination enzyme that has been characterised. The fluorinase was isolated in 2002 from *Streptomyces cattleya*, and, to date, this has been the only source of the fluorinase enzyme. Herein, we report three new fluorinase isolates that have been identified by genome mining. The novel fluorinases from *Streptomyces* sp. MA37, *Nocardia brasiliensis*, and an *Actinoplanes* sp. have high homology (80–87% identity) to the original *S. cattleya* enzyme. They all possess a characteristic 21-residue loop. The three newly identified genes were overexpressed in *E. coli* and shown to be fluorination enzymes. An X-ray crystallographic study of the *Streptomyces* sp. MA37 enzyme demonstrated that it is almost identical in structure to the original fluorinase. Culturing of the *Streptomyces* sp. MA37 strain demonstrated that it not only also elaborates the fluorometabolites, fluoroacetate and 4-fluorothreonine, similar to *S. cattleya*, but this strain also produces a range of unidentified fluorometabolites. These are the first new fluorinases to be reported since the first isolate, over a decade ago, and their identification extends the range of fluorination genes available for fluorination biotechnology.

The fluorinase (*S*-adenosyl-L-methionine:fluoride adenosyltransferase) isolated from *Streptomyces cattleya* in 2002 is unusual in biochemistry in that it catalyses C–F bond formation from an inorganic fluoride ion.[1] This enzyme mediates a reaction between *S*-adenosyl-L-methionine (SAM) and a fluoride ion to generate 5′-fluorodeoxyadenosine (FDA) and L-methionine, the first step in the biosynthesis of fluoroacetate (FAc) and 4-fluorothreonine (4-FT) from that organism.[2] The biosynthetic pathway to FAc and 4-FT has been evaluated in *S. cattleya*, and is illustrated in Scheme 1.[3] The fluorinase from *S. cattleya* has remained a unique curiosity for a decade. The most closely related genes to this enzyme were the duf-62 family (≈30% homology), and a chlorinase (≈40% homology) from the marine organism *Salinospora tropica*, which carry out related transformations on SAM but with nucleophiles other than fluoride ions.[4] There are over 200 duf-62 proteins in the available genomes, many found in extremophile organisms. They have been shown to catalyse a reaction between SAM and a hydroxide ion, to generate adenosine.[5–7] In *S. tropica* the chlorinase catalyses a similar reaction to the fluorinase except it uses a chloride ion, for C–Cl bond formation, as the first step in salinosporamide production.[8] We now report three new fluorinase enzymes from three additional bacterial strains. These fluorinases have been identified by genome mining, and they represent the first additions in a decade to the original *S. cattleya* fluorinase identification.

First, we identified a new bacterial isolate, *Streptomyces* sp. MA37, which was collected as a soil isolate in Ghana (5°39′32.72″ N, 0°11′55.26″ W). Analysis of the full genome sequence identified *flA*1, a gene with high (87%) homology to that of *flA* from *S. cattleya*.[9] A particularly diagnostic signature is a 21 amino acid loop, which is a unique characteristic of the *S. cattleya* enzyme relative to the chlorinase and duf-62 enzymes. Overexpression of a codon-optimised synthetic *flA1* gene in *E. coli*, with a His_6 tag and a TEV protease cleavage site added,[10] allowed us to be isolate and purify the protein. In vitro assays of the recombinant enzyme, following His_6 tag removal, demonstrated that it was competent to catalyse the conversion of SAM and a fluoride ion into 5′-FDA. Intriguingly, a 4-FT transaldolase (*flFT2*) responsible for the last step in 4-FT

[a] *Dr. H. Deng,[+] Z. Qin*
UK Marine Biodiscovery Centre, Department of Chemistry, University of Aberdeen
Meston Walk, Aberdeen AB24 3UE (UK)
E-mail: h.deng@abdn.ac.uk

[b] *Dr. L. Ma,[+] N. Bandaranayaka,[+] G. Mann, Prof. J. H. Naismith, Prof. D. O'Hagan*
School of Chemistry and Biomedical Sciences Research Centre, University of St Andrews
North Haugh, St Andrews KY169ST (UK)
E-mail: do1@st-andrews.ac.uk

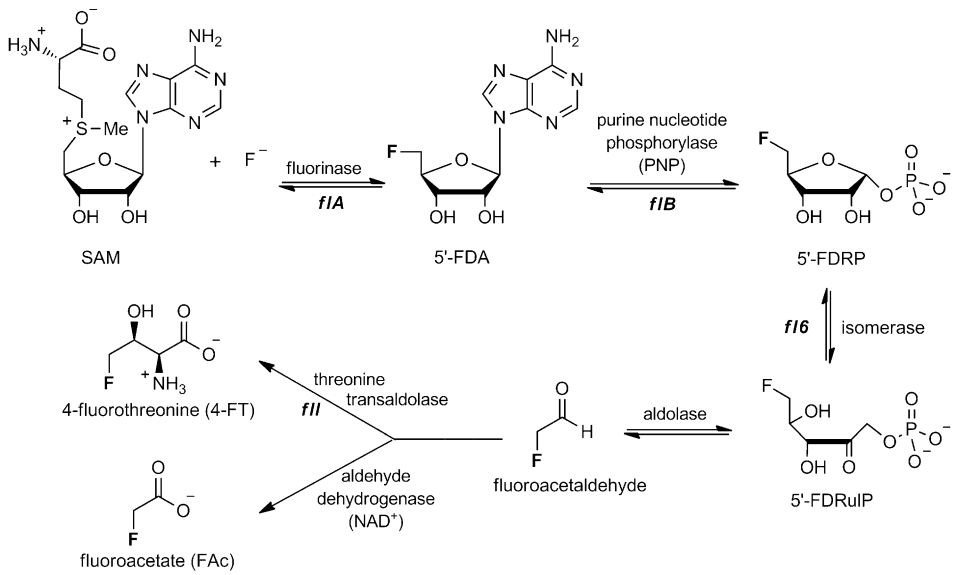
[c] *Dr. Y. Yu*
Key Laboratory of Combinatory Biosynthesis and Drug Discovery (Ministry of Education)
School of Pharmaceutical Sciences, Wuhan University
185 East Lake Road, Wuhan 430071 (P. R. China)

[d] *Dr. K. Kyeremeh*
Department of Chemistry, FGO Torto Building, University of Ghana
P.O. Box LG56, Legon-Accra (Ghana)

[e] *Dr. T. Shepherd*
The James Hutton Institute
Invergowrie, Dundee, DD2 5DA (UK)

[+] *These authors contributed equally to this work.*

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cbic.201300732.



Scheme 1. Biosynthetic pathway to FAc and 4-FT as established in *S. cattleya*.

biosynthesis[15,17] is located adjacent to the *flA1* gene, which suggests that *Streptomyces* sp. MA37 has a capacity for 4-FT biosynthesis. *Streptomyces* sp. MA37 was grown in a shaking flask in a medium supplemented with fluoride (2 mM), and the supernatant was analysed by ^{19}F NMR. The organism produced FAc (−217.44 ppm, t, $^{2}J_{HF}$ = 47.8 Hz) and 4-FT (−232.09 ppm, dt, $^{2}J_{HF}$ = 25 Hz, $^{3}J_{HF}$ = 47 Hz). The identity of 4-FT was further confirmed by GC–MS fragmentation, after *N*-methyl-*N*-(trimethylsilyl)-trifluoroacetamide (MSTFA) derivatisation.[11] Unlike *S. cattleya*, other unidentified fluorometabolites accumulate in the supernatant, however, at lower levels than 4-FT (Figure 2). This ability of *Streptomyces* sp. MA37 to elaborate fluorometabolites, suggests that the *flA1* gene plays a similar role to *flA* in *S. cattleya* in that it encodes the first transformation in fluorometabolite biosynthesis.

The *Streptomyces* sp. MA37 *flA1* gene was amplified by PCR using genomic DNA purified from the mycelium of *Streptomyces* sp. MA37. The overexpressed FlA1 enzyme was crystallised and its structure determined to 1.8 Å resolution (full statistics can be found in

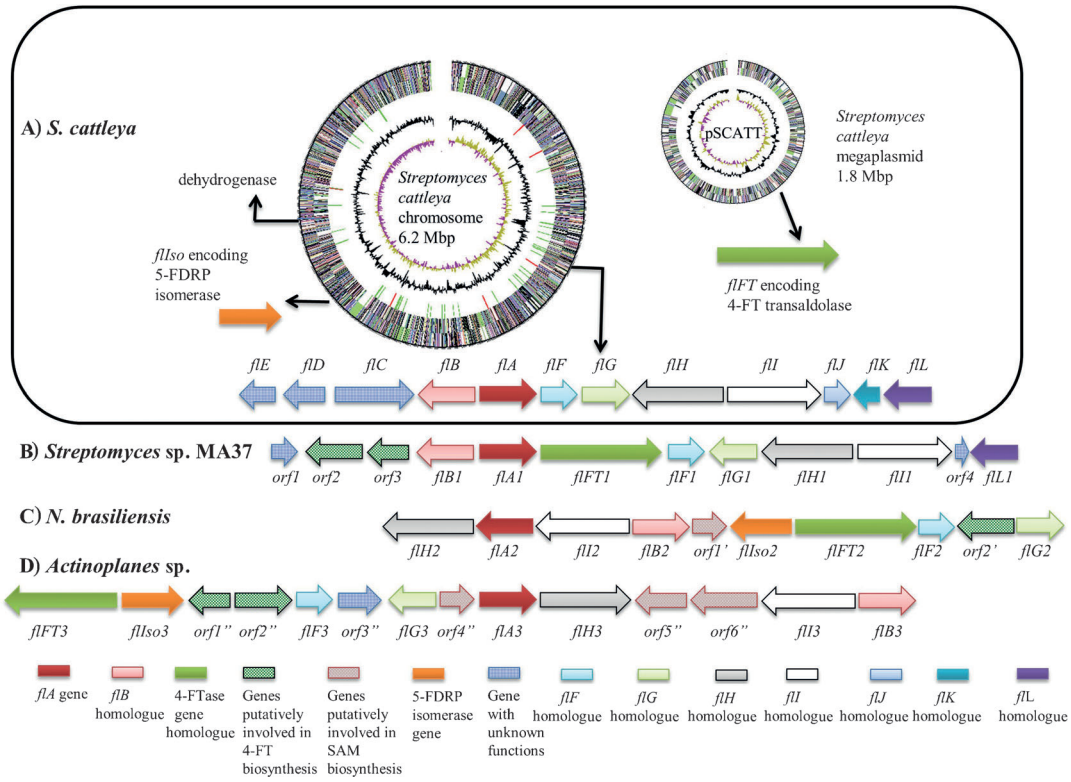


Figure 1. A) Genome map of *S. cattleya* showing the location of the biosynthetic genes, and that the PLP dependent 4-FT transaldolase and FDRP isomerase genes are remote from the *flA* gene cluster. Organisation of genes around the fluorinase (*flA*) from A) *S. cattleya* (Spencer cluster); B) *Streptomyces* sp. MA37; C) *N. brasiliensis*; D) *Actinoplanes* sp. The homologous genes are colour coded for visual comparison: *flA*, fluorinase; *flB*, purine nucleoside phosphorylase; *flF* and *flG*, DNA binding proteins; *flH*, Na^{+}/H^{+} antiporter; *flI*, *S*-adenosylhomocysteine lyase; *flJ* and *flL*, DNA binding proteins; *flK*, fluoroacetyl-CoA lyase; *flFT*, 4-FT transaldolase.

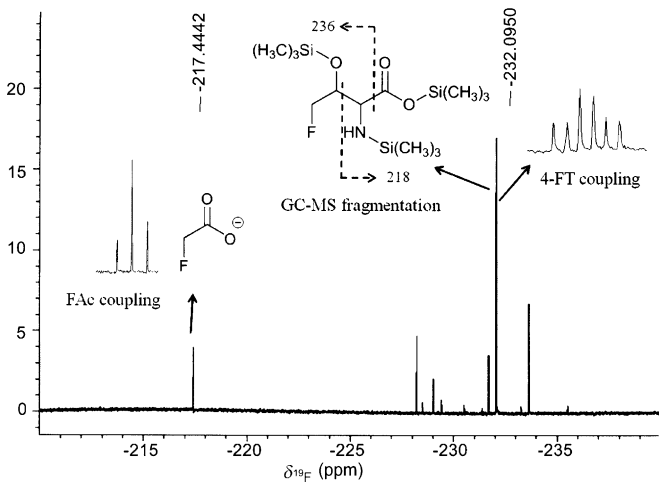


Figure 2. ^{19}F NMR spectroscopic analysis of fluorometabolites in the supernatant of the culture medium from *Streptomyces* sp. MA37 isolated from a Ghanaian soil sample. Inset (middle): MS fragmentation after MSTFA derivatisation.

Table S1 in the Supporting Information) by molecular replacement using the original *S. cattleya* fluorinase coordinates.[12] Superimposition of the *S. cattleya* and *Streptomyces* sp. MA37 structures, as illustrated in Figure 3, shows that they are essentially identical in their monomeric structure. Both proteins form the same hexameric arrangement; essentially a dimer of trimers. The structure of the FlA1 enzyme is found to contain a bound adenosine, which was carried through from the purification. The tight binding of adenosine was first observed in the *S. cattleya* fluorinase enzyme, which can also carry adenosine through the purification and is bound in an identical location.[2]

The residues that comprise the active site of FlA1 are shared between two protein monomers and their constellation is unchanged from our earlier description of the *S. cattleya* fluorinase.[12] The sequence differences between these two enzymes are remote from the active site and do not result in any significant change in the structure. The 21 amino acid loop, shown most clearly in Figure 3 A, emerges as a characteristic of the fluorinase relative to duf-62 and chlorinase. It is partly disordered in the FlA1 structure (Figure 3 A).

In 2012, the genome sequence of the hospital pathogen *Nocardia brasiliensis* HUJEG-1 (ATCC 700358) was placed into the public domain.[13,14] Comparative genome analysis revealed that an open reading frame (ORF; YP_006809254; FlA2) had 81% similarity to the fluorinase from *S. cattleya*. The predicted amino acid sequence indicated conserved active site residues and it also possessed the characteristic 21-residue found in the *S. cattleya* fluorinase.[12] A synthetic codon-optimised *flA2* gene was designed based on the genome sequence. This gene was designed such that it carried a His_6 tag to aid in purification and a TEV protease cleavage site.[10] The protein was successfully overexpressed and purified and was subjected to an in vitro activity assay. The resulting enzyme was also competent in its ability to catalyse the conversion of SAM and a fluoride ion into 5′-FDA. Other key biosynthetic genes are replicated in this genome and in particular a 4-FT transaldolase (*flFT2*), responsible for the last step in 4-FT biosynthesis,[15,17] is located very close to the *flA2* gene indicating a capacity for 4-FT biosynthesis. *N. brasiliensis* HUJEG-1 (ATCC 700358) was grown in shaking-flask cultures with 2 mM fluoride in various media. Although healthy growth was established, in our laboratory, fluorometabolite production could not be detected by ^{19}F NMR; this suggested that the enzyme has only a latent ability for fluorometabolite production.

Earlier this year, a third fluorinase candidate, ORF YP_007949809 (FlA3), was discovered in the deposited genome sequence[16] of the actinomycete, *Actinoplanes* sp. N902-109. From the available sequence data, *flA3* lies in a cluster of 14 ORFs. This gene has 80% identity to the *flA* gene of *S. cattleya* and contains the same 21-residue loop. The organism also has

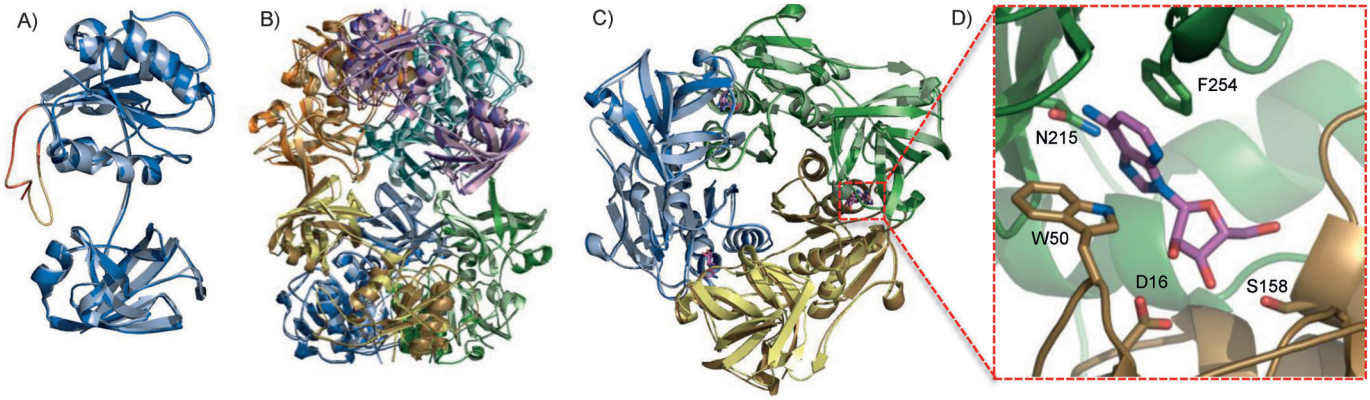


Figure 3. Superimposition of structures of fluorinase enzymes from *S. cattleya* and *Streptomyces* sp. MA37. A) Superimposition of the monomeric fluorinase from *Streptomyces* sp. MA37 (dark blue) and *S. cattleya* (light blue). The 21 amino acid loop overlaps for both structures. Some disorder in the *Streptomyces* sp. MA37 structure does not allow it to be completed. B) Superimposition of the hexameric fluorinases from *Streptomyces* sp. MA37 (darker colours) and *S. cattleya* (lighter colours). This hexameric assembly is essentially a dimer of trimers. C) Superimposition of the trimeric fluorinase structures from *Streptomyces* sp. MA37 (darker colours) and *S. cattleya* (lighter colours). Adenosine molecules are bound at the interface between the monomers. D) Close-up of an adenosine molecule (purple) bound at the active site of the fluorinase from *Streptomyces* sp. MA37. The active site residues, D16, W50, N215, and S158, are all conserved.

a 4-FT transaldolase (FlFT3) located close to the *flA3* gene. A synthetic, *E. coli* codon-optimised gene was also designed, and the product was also shown to catalyse the production of 5′-FDA from SAM in the presence of fluoride and is, therefore, a fluorinase. It remains to be determined if *Actinoplanes* sp. N902-109 is a producer of fluorometabolites. The organism, unlike the genome, is not in the public domain, but given the closely located *flFT3* gene, it is likely that this organism has a capacity for 4-FT biosynthesis.

A study of the reaction kinetics for all four fluorinases (Table 1) indicates that the original *S. cattleya* enzyme is the least efficient of those identified, although they are all rather slow enzymes with turnover numbers (k_{cat}) of less than 0.3 min^{-1}.

Table 1. Comparative kinetic data of the known fluorinase enzymes.

Fluorinase (FlA) source	SAM K_m [μM]	Turnover no. k_{cat} [min^{-1}]	Specificity constant k_{cat}/K_m [$mM^{-1}min^{-1}$]
S. cattleya	29.2 ± 2.41	0.083	2.84
Streptomyces sp. MA37	82.4 ± 18.6	0.262	3.18
N. brasiliensis	27.8 ± 4.23	0.122	4.40
Actinoplanes sp. N902-109	45.8 ± 7.91	0.204	4.44

The identification of these new fluorinases offers four genomes in which to compare and contrast the immediate environment of their respective *flA* genes[17,18] (Figure 1). In *S. cattleya*, *flA* is located immediately adjacent to the second biosynthetic enzyme, a purine nucleotide phosphorylase (PNP; *flB*), which catalyses the conversion of 5′-FDA into 5′-FDRP (Scheme 1). This is also the case for (*flB1*) found in *Streptomyces* sp. MA37, and for the *N. brasiliensis* and *Actinoplanes* sp., the *flB* homologues are situated one and five genes away from their respective *flA* genes. So it is a consistent feature that the fluorinase gene is located close to the PNP-expressing gene. This is also the case for the chlorinase in *S. tropica*.[8] However, the remaining four genes involved in fluorometabolite biosynthesis are scattered across the *S. cattleya* genome (Figure 1), and it is a feature of that genome that the biosynthetic genes are very remote from the (Spencer) cluster that contains the *flA* gene.[17] For all of these recently sequenced organisms, the degree of clustering appears more evolved. This is most striking for the biosynthetic genes (*flFT*) encoding the 4-FT transaldolase, which catalyses the last step in 4-FT biosynthesis. This gene is particularly remote from *flA* in *S. cattleya*, where it is not even on the chromosome but on a mega-plasmid, pSCATT (1.8 Mbp).[18] However, in the three newly sequenced genomes, the *flFT* genes are located very close to their respective *flA* homologues, thus suggesting a more efficient translational coupling of their gene products.

The gene *flIso2* in the *N. brasiliensis* gene cluster encodes a methylthioribose-1-phosphate isomerase[19] and has high sequence homology to the 5-FDRP isomerase (SCATT_20080), the third biosynthetic enzyme in *S. cattleya*. A homologue (*flIso*3) was also identified in the *Actinoplanes* sp. gene cluster. This suggests that *flIso* encodes the enzyme for the biosynthetic conversion of 5′-FDRP into 5′-FDRulP (Scheme 1) during FAc and 4-FT biosynthesis in these organisms. Again the *S. cattleya* homologue (*flIso*) of this gene is not located close to *flA* on the chromosome,[18] and overall *S. cattleya* appears to have the lowest degree of biosynthetic gene clustering of the genomes sequenced thus far.

In conclusion, in silico analysis has revealed three novel fluorometabolite gene clusters in actinomycete species with high homologies to a known fluorinase (80 to 87% identity) from *S. cattleya*. The fluorinase gene products from *Streptomyces* sp. MA37, *N. brasiliensis* and *Actinoplanes* sp. N902-109 were all shown by in vitro activity assays to function as fluorinases, and *Streptomyces* sp. MA37 produced FAc and 4-FT when grown in culture. The newly identified gene clusters have a higher degree of clustering of their biosynthetic genes relative to *S. cattleya*. The identification now of more than one fluorinase, and the associated genes for fluorometabolite production, expands the genetic resources available for developing de novo organic fluorine biotransformations through microbial engineering.

Acknowledgements

D.O'H. acknowledges the Royal Society for a Wolfson Research Merit Award and is the holder of an ERC Advanced grant. K.K. and H.D. acknowledge a Leverhulme–Royal Society Africa Award (AA090088). H.D. is thankful the financial support from the School of Natural and Computing Sciences, University of Aberdeen. Y.Y. is thankful for financial support from the "973" Program (2012CB721006) and the National Natural Science Foundation of China (81102357). G.M. is a BBSRC student.

Keywords: biotransformations · enzyme catalysis · fluorinases · genome mining · *Streptomyces* sp. MA

[1] D. O'Hagan, C. Schaffrath, S. L. Cobb, J. T. G. Hamilton, C. D. Murphy, *Nature* **2002**, *416*, 279 – 280.
[2] X. Zhu, D. A. Robinson, A. R. McEwan, D. O'Hagan, J. H. Naismith, *J. Am. Chem. Soc.* **2007**, *129*, 14597 – 14604.
[3] H. Deng, S. M. Cross, R. P. McGlinchey, J. T. G. Hamilton, D. O'Hagan, *Chem. Biol.* **2008**, *15*, 1268 – 1276.
[4] H. Deng, D. O'Hagan, *Curr. Opin. Chem. Biol.* **2008**, *12*, 582 – 592.
[5] H. Deng, C. H. Botting, J. T. G. Hamilton, R. J. M. Russell, D. O'Hagan, *Angew. Chem.* **2008**, *120*, 5437 – 5441; *Angew. Chem. Int. Ed.* **2008**, *47*, 5357 – 5361.
[6] H. Deng, S. McMahon, A. S. Eustáquio, B. S. Moore, J. H. Naismith, D. O'Hagan, *ChemBioChem* **2009**, *10*, 2455 – 2459.
[7] A. S. Eustáquio, J. Härle, J. P. Noel, B. S. Moore, *ChemBioChem* **2008**, *9*, 2215 – 2219.
[8] A. Eustáquio, F. Pojer, J. P. Noel, B. S. Moore, *Nat. Chem. Biol.* **2008**, *4*, 69 – 74.
[9] The producing microorganism *Streptomyces* sp. MA37 was isolated on ISP2 medium, supplemented with nalidixic acid and nystatin (25 mg L^{-1}), from the rhizosphere soil of a Moracear Bark Cloth tree (*Antiaris toxicaria*, Africa), growing in the University of Ghana Botanical Gardens (Legon, Ghana; 5°39′32.72″ N, 0°11′55.26″ W).
[10] H. Liu, J. H. Naismith, *Protein Expression Purif.* **2009**, *63*, 102 – 111.

[11] K. D. Golden, O. J. Williams, *J. Chromatogr. Sci.* **2001**, *39*, 243 – 250.
[12] C. Dong, F. L. Huang, H. Deng, C. Schaffrath, J. B. Spencer, D. O'Hagan, J. H. Naismith, *Nature* **2004**, *427*, 561 – 565.
[13] L. Vera-Cabrera, R. Ortiz-Lopez, R. Elizondo-Gonzalez, A. A. Perez-Maya, J. Ocampo-Candiani, *J. Bacteriol.* **2012**, *194*, 2761 – 2762.
[14] L. Vera-Cabrera, R. Ortiz-Lopez, R. Elizondo-Gonzalez, J. Ocampo-Candiani, *PLoS One* **2013**, *8*, e65425.
[15] C. D. Murphy, D. O'Hagan, C. Schaffrath, *Angew. Chem.* **2001**, *113*, 4611 – 4613; *Angew. Chem. Int. Ed.* **2001**, *40*, 4479 – 4481.
[16] June 11th 2013; Actinoplanes genome http://www.ncbi.nlm.nih.gov/nuccore/494682416.
[17] F. Huang, S. F. Haydock, D. Spiteller, T. Mironenko, T.-L. Li, D. O'Hagan, P. F. Leadlay, J. B. Spencer, *Chem. Biol.* **2006**, *13*, 475 – 484.
[18] C. Zhao, P. Li, Z. Deng, H.-Y. Ou, R. P. McGlinchey, D. O'Hagan, *Bioorg. Chem.* **2012**, *44*, 1 – 7.
[19] H. J. Imker, A. A. Fedorov, E. V. Federov, S. C. Almo, J. A. Gerlt, *Biochemistry* **2007**, *46*, 4077 – 4089.

Received: November 21, 2013
Published online on January 21, 2014