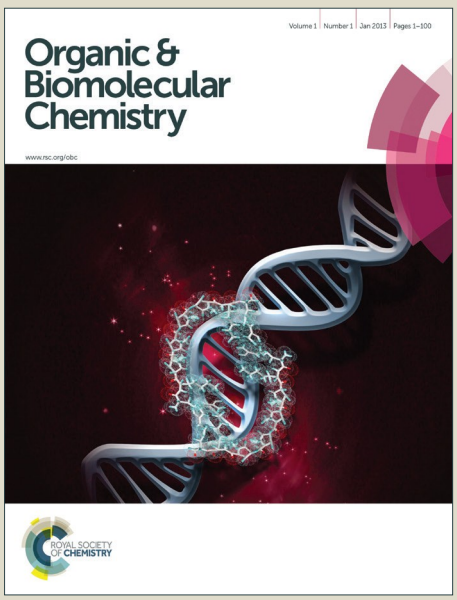
Organic & Biomolecular Chemistry

Accepted Manuscript



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/obc

Journal Name

ARTICLE

Target-Specific Identification and Characterization of the Putative Gene Cluster for Brasilinolide Biosynthesis Revealing Mechanistic Insights and Combinatorial Synthetic Utility of 2-Deoxy-L-Fucose Biosynthetic Enzymes

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Hsien-Tai Chiu*[a], Chien-Pao Weng[a], Yu-Chin Lin[a,b] and Kuan-Hung Chen[b]

Brasilinolides exhibiting potent immunosuppressive and antifungal activities with remarkably low toxicity are structurally characterized by an unusual modified 2-deoxy-L-fucose (2dF) attached to a type I polyketide (PK-I) macrolactone. From the pathogenic producer *Nocardia terpenica* (*Nocardia brasiliensis* IFM-0406), a 210kb genomic fragment was identified by target-specific degenerate primers and subsequently sequenced, revealing a giant *nbr* gene cluster haboring genes (*nbrCDEF*) required for TDP-2dF biosynthesis and those for PK-I biosynthesis, modification and regulation. The results showed genetic and domain arrangements of *nbr* PK-I synthases agreed colinearly with PK-I structures of brasilinolides. Subsequent heterologous expression of *nbrCDEF* in *Escherichia coli* accomplished *in vitro* reconstitution of TDP-2dF biosynthesis. Catalytic functions and mechanisms of NbrCDEF enzymes were further characterized by systematic mix-and-match experiments. The enzymes were revealed to display remarkable substrate and partner promiscuity, leading to establishment of *in vitro* hybrid deoxysugar biosynthetic pathways throughout an *in situ* one-pot (iSOP) method. This study represents the first demonstration of TDP-2dF biosynthesis in enzyme and molecular levels, and provides new hope for expanding structural diversity of brasilinolides by combinatorial biosynthesis.

Introduction

Deoxysugars are indispensable structural determinants in biological activity and potency of many bioactive natural product glycosides. Emerging lines of evidence have promised development of new drugs by creating and modulating structural diversity of the deoxysugars via engineered or combinatorial biosynthesis.[1] Among the deoxysugars, 2-deoxy-L-fucose (2dF) constitutes several known biologically important natural products, such as brasilinolides, elaiophylin, aclacinomycin, marcellomycin and esperamicin.[2-6] Thus far, little is known about biosynthetic origin and mechanism of this unusual deoxysugar (2dF), and its biosynthetic enzymes have never been functionally characterized *in vitro*. Moreover, brasilinolides A and C, isolated from the pathogenic actinomycete *Nocardia terpenica* (*Nocardia brasiliensis* IFM-0406), have received great attention not only due to its presence of 2dF but also its structurally unique 32-membered macrolide (Fig. 1).[7] As a potent immunosuppressant, brasilinolide A has shown remarkably low cytotoxicity and favorable physicochemical properties, and has served as an attractive target for synthetic organic chemistry and biological activity studies.[8] Although the exact molecular mechanisms governing the *in vivo* mode of action of brasilinolide A are still unclear, its structural features distinct from other clinically used immunosuppressants, e.g., cyclosporins, FK506 (Tacrolimus) and rapamycin (Sirolimus), provides a new hope to develop new, specific drugs preventing tissue rejection in human organ transplantation with reduced side effect by triggering an entirely different immune target mechanism.[9] In contrast, brasilinolide B produced by *Nocardia brasiliensis* IFM-0466 did not show immunosuppressive activity but broad antifungal and antibacterial activities.[10] The apparent discrepancy between brasilinolides A and B has been proposed to be attributed to the structural variations in 2dF.[10,11] Also interesting is that brasilinolide C, lacking any modifications on 2dF and macrolide (C23-OH), showed anticancer activity on murine lymphoma, but brasilinolide A did not.[11] Therefore, in drug discovery brasilinolides may act as excellent molecular templates of drugs for glycorandamization and precursor-directed polyketide engineering throughout combinatorial biosynthesis.

In light of the critical roles of 2dF and brasilinolides, the enzymes responsible for biosynthesis of the activated sugar donor, nucleotide diphosphate (NDP)-2dF (Fig. 1), require to be mechanistically characterized in molecular level. In addition, the structure features of brasilinolides also imply participation

[a] *Department of Chemistry, National Cheng Kung University, No.1, University Road, Tainan 701, Taiwan. E-mail: hchiu@mail.ncku.edu.tw*
[b] *Department of Biological Science and Technology, National Chiao Tung University, Hsinchu 300, Taiwan.*

†Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

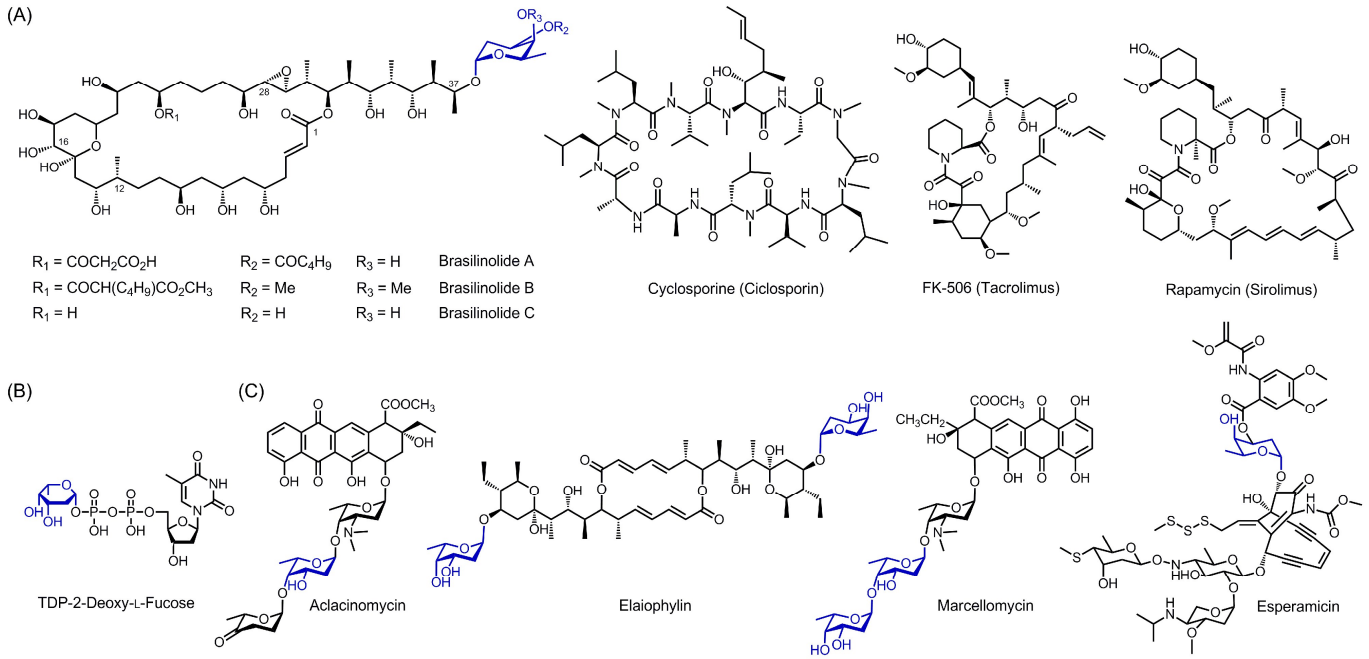


Fig. 1 Chemical structures of brasilinolides and immunosuppressants (A), TDP-2-deoxy-L-fucose (B) and bioactive natural products containing the 2-deoxy-L-fucose moiety (C).

of type I modular polyketide synthases (PKS-I) for scaffold biosynthesis and several important post-modification enzymes promoting (i) C37-*O*-glycosylation (ii) C23-*O*-(alkyl)malonyl acid esterification (iii) C28-29 epoxidation (iv) C16-hydroxylation and (v) *O*-acylation/methylation on the sugar moiety. Thus far, nothing is known about these enzymes, which are very useful for expanding structural diversity of brasilinolides and other macrolides upon conjugation with tunable glycosyl patterns. Herein, we describe molecular cloning, sequence analysis and characterization of a putative gene cluster for brasilinolide biosynthesis, where the 2dF enzymes were obtained in large scale by heterologous expression in *E. coli* and successfully characterized by a functional mix-and-match approach. Furthermore, the 2dF enzymes were utilized for *in vitro* combinatorial synthesis of various NDP-deoxysugars by taking advantage of their striking substrate and partner promiscuity revealed in this study. The results have shed new light on biosynthetic mechanisms leading to formation of NDP-2dF and brasilinolides and the applications of the enzymes in combinatorial biosynthesis.

Results and discussion

Molecular cloning and sequence analysis of the putative gene cluster for brasilinolide biosynthesis.

Based on structural analysis, brasilinolide biosynthesis is featured with PKS-I and the 2-deoxygenated form (2dF) of NDP-sugar. The 2-deoxygenation in deoxysugar biosynthesis is known to be facilitated specifically by catalytic action of NDP-4-keto-6-deoxy-hexose (KDH) 2,3-dehydratase (2,3-Deh). To clone and identify the gene cluster for brasilinolide biosynthesis, a fosmid genomic library of *N. terpenica* was thus constructed (Fig. S1, ESI†). Designed based on the highly conserved regions of various 2,3-Deh homologs and PKS-I ketosynthase (KS) domains, two types of degenerate primer pairs (O_1F/O_3R and KSD1F/KSD1R, respectively) for probing the corresponding genes were used to identify related positive PCR fragments throughout screening of the library (Table S1 and Fig. S2, ESI†). Further screening of the library using the specific primers (PK1F/PK1R and SOHC10F/SOHC10F) derived from the sequences of the PCR fragments led to isolation of the positive fosmid clones containing pYC41D8, pYC41H8 and pYC48C11 (Fig. S3, ESI†). End sequences of the confirmed positive fosmid clones were then utilized to bridge out with other fosmids by PCR screening for extension on both sides, leading to successful construction of a genomic contig spanning *ca.* 210 kb as shown in Fig. 2. Subsequent restriction mapping and sequence analysis revealed a giant *nbr* gene cluster harboring the genes for NDP-2dF biosynthesis (*nbrCDEF*), PKS-I and glycosylation (*nbrAB*), as well as those for polyketide modifications and regulation (Fig. 2; Fig. S3, ESI†). Substantial efforts were made on complete sequencing of the 210kb contiguous segment (101 *orfs*) constituted by DNA inserts of pYC13F9, pYC54H10, pYC60D4, pYC55F7, pYC44F12, pYC41D8, pYC48C11 and pYC14E6 fosmid clones, which was accomplished by three methods: (1) end sequencing of fosmid restriction fragments (2) next-generation sequencing (NGS) of the fosmid clones (3) PCR walking on fosmid clones to close the gaps between assembled sequence contigs obtained from methods 1 and 2. The resulting high-quality sequence information was deposited in GenBank[12], and results of bioinformatics analysis summarized in Table S2 (ESI†). The overall GC content of the 210kb sequenced region was 71.27%, which is higher than the entire GC content

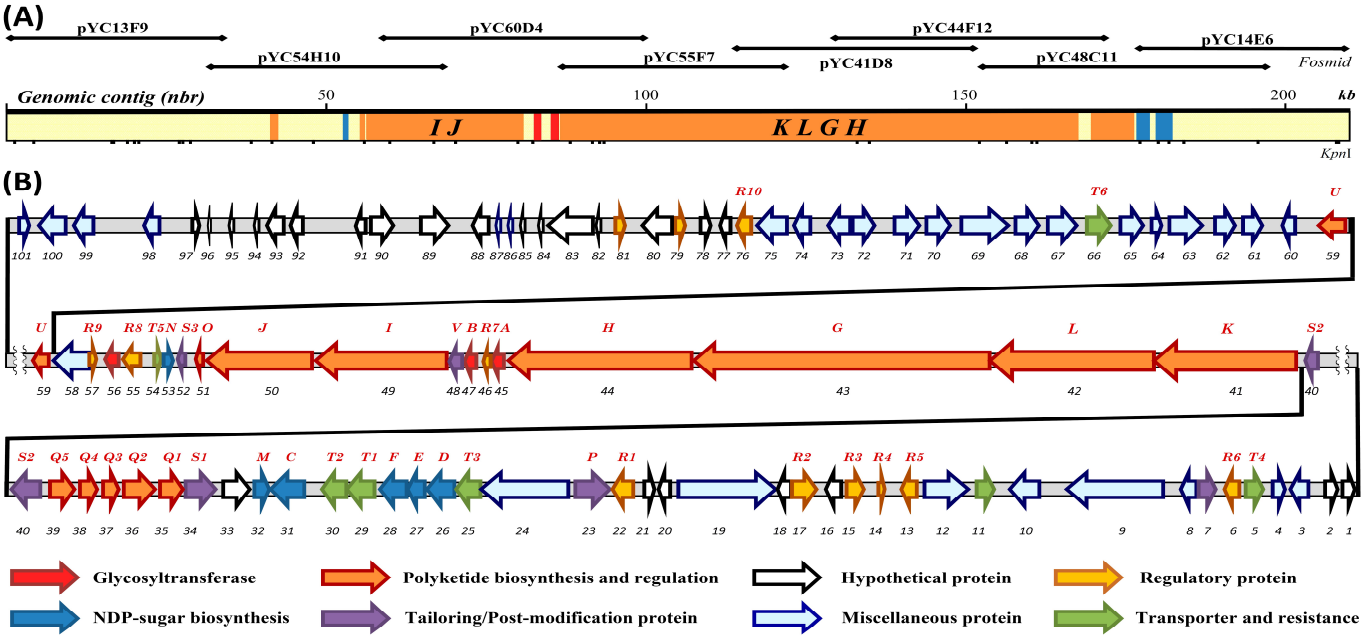


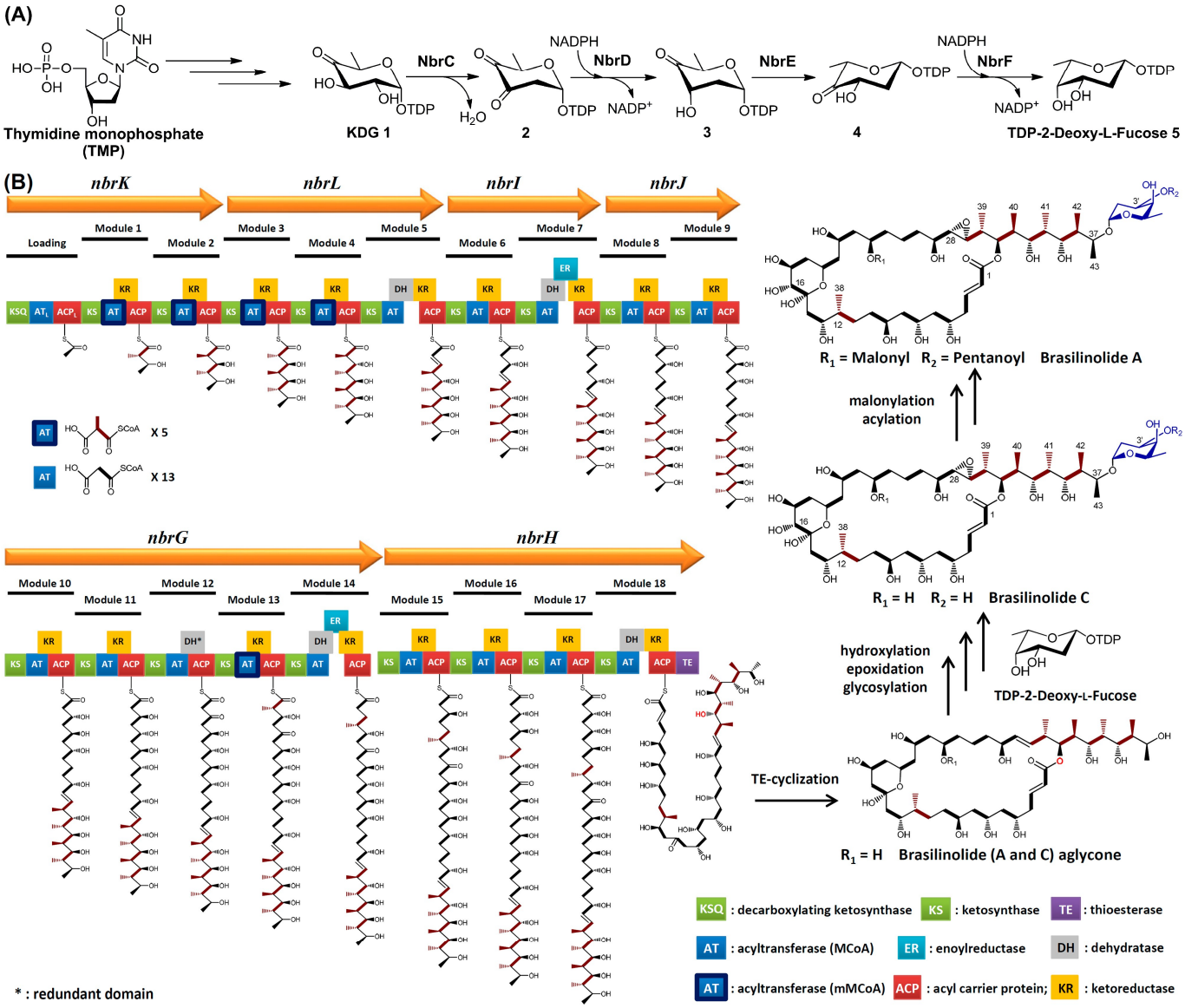
Fig. 2 The 210kb genomic fragment harboring the putative gene cluster for brasilinolide biosynthesis in *N. terpenica*. (A) the genomic contig constituted by the overlapping fosmids (B) genetic arrangement of the *orfs* embedded within the fragment.

(68.0%) of *N. terpenica* and in the lower bound of the characteristic 70~75% for *Streptomyces* DNA.[7]

Genetic and domain arrangements of type I modular polyketide synthase genes in the *nbr* gene cluster

In polyketide biosynthesis, type I refers to linearly arranged and covalently fused catalytic domains organized as modules to form large multifunctional enzymes (PKS), which are well correlated with the linear structural pattern of a given polyketide backbone. Following an initial starter module with loading domains, each extender module harbors ketosynthase (KS), acyltransferase (AT) and acyl carrier protein (ACP) as the three minimum domains required for a full catalytic cycle of ketide chain extension, whereas ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) may be present within the extender module for ketide modification.[13] The cyclization of the polyketide chain may then be accomplished by thioesterase domain (TE) usually attached in the last module. In this study, bioinformatic analysis of the *nbr* gene cluster revealed 6 large PKS-I genes *nbr*GHIJKL (*orfs*43, 44, 49, 50, 41 and 42, respectively) contributing *ca.* 91.7 kb in length and consisting of 19 PKS-I modules and 82 domains in total (Scheme 1). Among transcribed products of the 6 PKS-I genes, NbrK may act as the first-committed PKS-I enzyme in PK-I biosynthesis since it contains a loading module embedded with a characteristic N-terminal KSQ domain. This KSQ domain possesses a highly conserved glutamine (Q) replacing the active-site cysteine (C) within the conserved amino acid region (VDTACSSS) of a common KS domain (Fig. S4, ESI†).[14] Within the NbrK loading module, the KSQ catalyzing decarboxylation may work closely with the AT_L governing the selection of starter unit to yield an ACP_L thioester, thereby acting as a priming mechanism of PK-I biosynthesis similarly operated in spinosyn, pikromycin and oleandomycin biosyntheses.[15] NbrK has two extender modules, modules 1 and 2, following the loading module to initiate the PK-I ketide chain extension cycle. Immediately downstream of *nbrK*, *nbrL* codes for an independent PKS (NbrL) containing 3 more extender modules (3, 4 and 5). NbrI and NbrJ are transcribed by two adjacent *orfs* of the same direction, both harboring 2 extender modules, modules 6 & 7 and modules 8 & 9, respectively. NbrG is the largest PKS of 8,183 amino acids holding 5 extender modules 10-14, and downstream its coding sequence (*orf43*) is the *orf44* encoding NbrH (ORF44, 6725 amino acids). NrbH contains 4 extender modules (15-18) and is recognized as the last PKS committed in PK-I biosynthesis because of its characteristic C-terminal type I TE domain responsible for cyclization of mature polyketide chain. In the TE domain was also found the catalytic triad, i.e. Ser, His and Asp, generally conserved in functional TE domains of PKS and fatty acid synthases (FAS) (Fig. S5, ESI†).[16]

Thus far, the numbers and overall genetic organization of modules and required domains follow the modular logic of type I modular PK biosynthesis and match well with the proposed polyketide chain extensions in brasilinolide biosynthesis as demonstrated in Scheme 1. To further correlate the sequence information with the substrate specificity and stereochemistry of modifications governed by the domains, intensive sequence analysis with conventional bioinformatics tools, e.g. SBSPKS and NCBI programs, was pursued and also verified with aid from literature studies.[16-19] In Table S3 (ESI†) are shown the summarized results from the comparison of sequence-based and chemical structure (brasilinolides)-based predictions on the *nbr* PKS-I domains, especially ATs and KRs. The *nbr* PKS-I enzymes habor 19 AT domains and 17 KR domains. The five AT domains (AT_{1-4} and AT_{13}) presumed to accept methyl-malonyl-CoA (mMCoA) as the substrate, as judged from the



Scheme 1 The proposed biosynthetic pathway for the formation of TDP-2-deoxy-L-fucose (A) and brasilinolides A and C (B).

structure-based analysis of brasilinolides, indeed carry the signature motifs I~III associated with mMCoA substrate specificity (Fig. S6, ESI†).[20] The predictions from both sequence and chemical structure also agree unanimously that the other twelve ATs (AT_{5-12}, AT_{14-15} and AT_{17-18}) may utilize malonyl-CoA (MCoA) as the extender substrate. Interestingly, AT_{16} contains variants (**GDA**VVQ and Y**P**SH) of the motifs I [RQSED]V[DE]VVQ and III (YASH), respectively, typically controlling the substrate specificity for mMCoA. Therefore, the mutations of these motifs may presumably account for the altered specificity for MCoA, in addition to possible combinatorial results of other structural or sequence factors affecting the substrate specificity, including the ~30-residue C-terminal region (motif IV) of AT.[20,21] Furthermore, a similar variation (**GTA**VVQ) on the mMCoA-specific motif I also occurs on the loading AT domain (AT_L), which bears mMCoA-specificity-like motifs but is expected to adopt MCoA as the substrate for brasilinolide biosynthesis. In fact, phylogenetic analysis revealed that the AT_L stood alone in distant evolutionary relationship from all of the five *nbr* ATs (AT_{1-4} and AT_{13}) clustering together and exhibiting substrate specificity for mMCoA (Fig. S6, ESI†). In light of the obscure substrate specificity of *nbr* AT_L from the sequence-based prediction, the exact uptake or final utilization of the actual substrate (MCoA) by *nbr* AT_L may thus likely be associated with the culture condition producing the metabolite and the regulatory role of *nbr* KSQ acting as a gatekeeper for selecting the proper starter (acetate) unit for downstream chain extension.[14,22,23] The observation is also consistent with the inherent nature of relaxed substrate specificity often found in loading AT domains of other type I PKSs such as erythromycin and avermectin PKSs.[20,22-25]

In *nbr* PKSs, nearly all of the 18 extender modules contain KR for keto reduction except module 12, which is in excellent agreement with the presence of the only keto group (at C15) on brasilinolide polyketide backbone. KRs in PKS-I biosynthesis normally determine the stereochemistry of α-substituents and β-hydroxyl groups of extended ketide products and, based on sequence fingerprints, have been classified into six types (A1, A2, B1, B2, C1 and C2) according

to the model of Keatinge-Clay.[26] As shown in Table S3 (ESI†), the A1-type (α*R*-β*S*) fingerprints possessed by KR_2, KR_{11} and KR_{15} indeed account for the observed stereochemistry of C34-35 (α-unepimerized, *R*-*S*), C16-17 (*X*-*S*) and C8-9 (*X*-*S*), respectively, in brasilinolides (*X* indicates no stereocenter). The A2-type (α*S*-β*S*) characters of KR_1 and KR_3 are also consistent with the *S*-*S* α-epimerized stereocenters of brasilinolide C36-37 and C32-33, respectively. Interestingly, *nbr* PKSs contain eight KRs (KR_{4-8}, KR_{10}, KR_{14} and KR_{18}) of B1-type (α*R*-β*R*), four of which can in fingerprints be nicely referred to the established stereochemistry of C30-31 (*R*-*R*, KR_4), C26-27 (*X*-*R*, KR_6), C22-23 (*X*-*R*, KR_8) and C18-19 (*X*-*R*, KR_{10}) in brasilinolides. The other four B1-KRs are each accompanied with DH within the same module, thereby reasonably obscuring the inherent stereochemistry of the corresponding carbons by DH-catalyzed dehydrations. Furthermore, the *S*-*R* configurations of C12-13 and the *X*-*R* configurations of C6-7 and C4-5 also agree nicely with the B2-type characters of KR_{13}, KR_{16} and KR_{17}, respectively. Among all 17 *nbr* KRs, KR_9 presumably responsible for the stereochemistry (*X*-*S*) of brasilinolide C20-21 bears peculiar, mixed sequence characters of A1- and B1-types, featuring the conserved W (A1), LDD motif (B1) and no intramodular DH companion (A1). Phylogenetic analysis of all the 17 KRs showed that, unlike A-type KRs, most B-type KRs (especially B1-KRs) tended to cluster together in evolutionary relationship, where KR_9 seemed to resemble A-type KRs and exhibited the farthest relationship with most B-type KRs (Fig. S7, ESI†). Previous studies have shown that mutations (or mixings) of the sequence fingerprints may lead to alterations in stereochemical outcomes of the α- and β-substituents,[27] and the exact stereochemistry carried out by such KRs should be experimentally determined.[28] Thus, KR_9 may represent an native case of interesting mixed-type KR worthy of further studies.

The aforementioned loss of the Cα-Cβ stereocenters caused by DH-catalyzed dehydrations may take place on C28-29, C2-3, C24-25 and C10-11 in brasilinolides, which can be well supported by the presence of the DHs appropriately suited in *nbr* modules 5, 18, 7 and 14, respectively (Table S3, ESI†). The intramodular companionship of ERs with DHs in modules 7 and 14 can then allow NADPH-dependent reduction of the DH-dehydrated double bonds to afford the apparent single bonds on C24-25 and C10-11, respectively, in brasilinolides. Moreover, all of the four DHs (DH_5, DH_7, DH_{14} and DH_{18}) basically preserve the reported motifs I (HxxxGxxxxP) and II (D[AV][VA][AL][QH]) containing catalytic dyad (H & D) and NADPH-binding motif,[29] as well as two conserved sequence motifs GxxYGPxF and LPFxW responsible for defining the substrate pocket, generally for PKS-I DHs (Fig. S8, ESI†).[16] The four DHs also display good sequence similarity to each other in an averaged level of *ca.* 41% identity and 54% similarity. Interestingly, in *nbr* PKS module 12 was found a redundant DH domain (DH_{12}), which should be inactive or non-functional due to a serious lack of conservation in the motif II and the latter two motifs associated with substrate pocket. This may be further supported by the observations that DH_{12} shares fairly low sequence homology (20.2% identity and 36.3% similarity, averaged) with the other *nbr* DHs, and that in module 12 there is no KR for generating a hydroxyl group required for subsequent dehydration. Similar situations are commonly found in PKSs, such as those involved in rifamycin B and nystatin biosynthesis.[30,31] However, it cannot be ruled out an unprecedented possibility that in brasilinolide biosynthesis the aberrant DH_{12} might participate in the hemiacetal formation between C15-keto and C19-OH on the acyl-NbrG intermediate at module 12. On the other hand, the two ERs (ER_7 and ER_{14}) mentioned above for the full reduction of the α-β carbons were found to highly resemble each other in sequence (59.9% identity and 71.2% similarity) and both contain the consensus sequence LxHx[GA]xGGVG known for NADPH binding (Fig. S9, ESI†).[32] Notably, about 90 residues N-terminal of the consensus motif was identified a conserved Tyr (Y) as a known indicator of a *si*-facial protonation during Cα-β reduction by NADPH.[15] The intrinsic stereospecificity of ER-catalyzed reduction is however obscured due to the absence of α-substituents at C10 and C24.

As demanded for covalent shuttles of extender units and polyketide intermediates in brasilinolides, 19 ACP domains of *ca.* 80 residue long and good similarity (*ca.* 38.8% identity and 52.9% similarity) were each found in each corresponding *nbr* PKS module (including loading module). A conserved [DE]xGxD**S**L motif, similarly revealed in DEBS ACPs, was recognized across all *nbr* ACPs with the conserved serine serving as the phosphopantetheine attachment site for the shuttling (Fig. S10, ESI†).[33,34] Upon completion of PK chain elongations and modifications, a type I TE (TEI) domain located at C-terminus of NbrH may carry out the macrolactone formation of brasilinolides by nucleophilic attack (presumably activated by His) of C31-OH onto C1 of the acyl-TE thioester (linked through Ser), thereby releasing the polyketide macrolactone from NbrH as illustrated in Scheme 1. The Ser of the aforementioned catalytic triad (S-H-D) is embedded within the conserved Gx**S**xG motif generally found in PKS-I TEIs, such as PICS and DEBS TEIs (Fig. S5, ESI†).[33] Interestingly, *nbr* TEI showed about the same overall sequence homology (*ca.* 34.6% identity and 47.9% similarity, averaged) to various PKS-I TEIs (e.g., TEIs in biosyntheses of erythromycin, nystatin, halstoctacosanolides and pimaricin), regardless of the size and length of the corresponding PK macrolactones processed by these TEIs. Perhaps one of the amino acid sequence segments controlling the regiochemistry of the TEI-catalyzed macro-cyclization is associated with the previously reported hypervariable region (lid) found in both *nbr* TEI and other PKS-I TEIs (Fig. S5, ESI†).

Notably, adjacent to *nbrJ* of the *nbr* PKS gene cluster was found *orf51* (*nbrO*) encoding PK-I type II thioesterase (TEII) known to function as an editing enzyme for selecting starter acyl units, removing aberrant extender acyl units, dehydrating PK intermediates and/or releasing final products during the course of PKS-I biosynthesis.[35] In sequence homology, NbrO displayed high sequence identity with various PKS-I TEIIs, e.g. ScnI (51.6%), BorB (48.9%), PikAV (57.3%) and TylO (46.6%), whereas both the GxSxG motif harboring catalytic active-site Ser and the GGHF[FY]L motif with catalytic His were retained in NbrO (Fig. S11, ESI†).[35,36] Prominently, *nbrO* utilizes AGC as the codon for the active-site serine, which is characteristic of PK-I TEIIs.[37] Moreover, previous studies showed that deletion or inactivation of TEII led to evident reduction to various degrees (10-90%) in PK-associated products, including rifamycin B (*A. mediterranei*), borrelidin (*S. parvulus* Tü4055) and tylosin (*S. fradiae*).[36] Hence, in *nbr* gene cluster the intimate clustering of *nbrO* with PKS-I genes with same

orientation may allow efficient priming and elongation of this giant *nbr* PK-I biosynthetic machinery for improved production of products.

In summary, the bioinformatics analyses indicate that the *nbr* PKSs (NbrGHIJKL) are fully equipped with necessary domains and modules well and collinearly arranged for the orderly events (PK chain elongations, modifications and macrocyclization) proposed in brasilinolide biosynthesis. The various sequence characters presented by each domain of *nbr* PKSs also agreed nicely with the presumed substrate specificity and stereochemistry associated with brasilinolide biosynthesis. The *nbr* PKS genes revealed herein thus represent one of the largest PKS gene clusters obtained by target-specific cloning approach.

Organization of genes for TDP-2-deoxy-L-fucose biosynthesis

In microbial deoxysugar biosynthesis, NDP-4-keto-6-deoxy-glucose (KDG, **1**) is known to serve as a key branch-point intermediate from which various NDP-deoxysugars may be derived by catalytic transformations with different sets of KDG-modification enzymes.[1,38] Biosynthetic maturation of an NDP-deoxysugar is usually furnished by a chemical saturation of the KDG 4-keto group such as ketoreduction or transamination catalyzed by 4-ketoreductase or 4-aminotransferase, respectively. For brasilinolide biosynthesis, a set of such modification enzymes are thus required for the formation of NDP-2dF (**5**), the 2dF moiety of which may then be incorporated into the PK macrolactone/aglycone by catalytic action of an *O*-glycosyltransferase along with release of the NDP leaving group (Scheme 1). Indeed, in the *nbr* gene cluster were identified six *orfs* (*orfs*26-28, *orf*31, *orf*32 and *orfs*53) coding for NDP-deoxysugar biosynthetic enzymes (NbrDEFCMN) intimately clustered together with the aforementioned giant PKS-I genes. Further sequence analysis revealed that NbrCDEF showed high sequence homology to the enzymes functioning as KDH 2,3-dehydratase (2,3-Deh), KDH 2,3-reductase (also named KDH 3-ketoreductase, 3-Ker), KDH 3,5-epimerase (3,5-Epi) and KDH 4-ketoreductase (4-Ker), respectively, which catalytically are sufficient to accomplish the modifications of KDG (**1**) to afford NDP-2dF (**5**) as shown in Scheme 1. On the other hand, NbrM and NbrN encoded by *orf32* and *orf53* highly resembles in sequence the enzymes acting as thymidylate kinase (TMK) and glucose-1-phosphate thymidylyltransferase (Gtt), respectively, which are known to work together with acetyl kinase (ACK) and NDP-glucose 4,6-dehydratase (Gdh) for biosynthesis of KDG (**1**), the key precursor for TDP-2dF biosynthesis. Interestingly, the gene (*ngd*) coding for Gdh was identified elsewhere in the *N. terpenica* genome and found to cluster with cell wall biosynthesis genes for primary metabolism (data not shown), suggesting that the *ngd* gene is likely shared by primary and secondary metabolisms in *N. terpenica*. The results also suggested that the biosynthesis of NDP-2dF in *N. terpenica* may utilize the common deoxysugar biosynthetic pathway involving KDG as the key intermediate.

As mentioned, 2dF is a special, rare sugar moiety found in bioactive natural product glycosides. The activated form of 2dF would normally require four genes coding for 2,3-Deh, 3-Ker, 3,5-Epi and 4-Ker. Interestingly, similar genes encoding these four enzymes were also found to group together in the *ela* and *akn* gene clusters for elaiophylin and aclacinomycin biosyntheses, respectively, as shown in Fig. 3, where none of these NDP-2dF genes have ever been functionally and mechanistically characterized. In both *nbr* and *ela* clusters the four genes are arranged in the same direction, whereas in *akn* cluster they are yet split into two groups (*3-ker* and *2,3-deh*/*3,5-epi*/*4-ker*) heading in different directions on chromosome. Although the reason for such rearrangements is still obscure, the observation might be evolutionarily or functionally relevant (see later discussion). It is worth noting that the enzymatic products of 2,3-Deh and 3,5-Epi are generally labile to degradation and may require immediate captures by the downstream enzymes like reductases (e.g., 3-Ker and 4-Ker, respectively) to neutralize (or chemically saturate) the active keto group. Such arrangements of the genes encoding the catalytically-coupled enzymes would probably be an evolutionary consequence for proper transcriptional (or translational) regulation and efficiency. It might also facilitate possible protein-protein interactions between the coupled enzymes for stability and efficient utilization of intermediates throughout channelling between the enzymes.[39,40]

Post-modification genes coding for polyketide macrolide glycosyltransferases

In biosyntheses of brasilinolides, the decoration of the PK macrolide with the 2dF moiety requires catalytic participation of an *O*-glycosyltransferase. Notably, in the *nbr* gene cluster two *orfs*, *nbrA* and *nbrB*, coding for secondary metabolite glycosyltransferases were identified and embedded in between the giant PKS-I genes (Fig. 2, Table S2 and Fig. S3, ESI†). To resolve functional roles of *nbrA* and *nbrB* in the biosynthesis, bioinformatic analysis and evolutionary comparisons of secondary metabolite glycosyltransferases in the NCBI database were conducted in this study, revealing that the deduced products of *nbrA* and *nbrB* shared high sequence similarity to the enzymes responsible for glycosylation of polyketide natural products. Notably, the protein encoded by

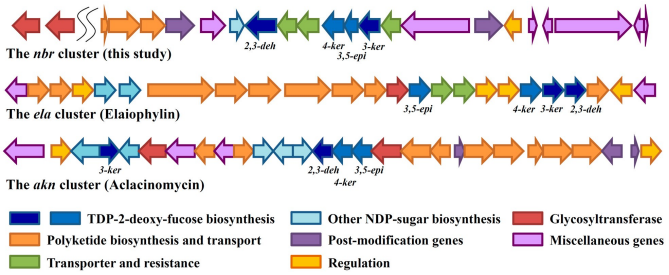


Fig. 3 Comparison of genetic organizations responsible for biosyntheses of the natural products containing 2-deoxy-L-fucose. The biosynthetic gene clusters for comparison are the *nbr* (this study), *ela* (elaiophylin, accession number GP697151) and *akn* (aclacinomycin, accession numbers AF264025 and AF257324) clusters. Transcribed products of the 2-deoxy-L-fucose biosynthetic genes are indicated as *2,3-deh*: KDH 2,3-dehydratase (NbrC in *nbr*); *3-ker*: KDH 2,3-reductase (NbrD in *nbr*); *3,5-epi*: KDH 3,5-epimerase (NbrE in *nbr*); *4-ker*: KDH 4-ketoreductase (NbrF in *nbr*), where KDH is NDP-4-keto-6-deoxyhexose.

nbrA highly resembles in sequence AknK (also named AclK, 49.8% similarity and 34.9% identity), which has been characterized as an L-2-deoxyfucosyltransferase catalyzing the transformation of rhodosaminyl aklavinone to 2-deoxy-L-fucosyl-L-rhodosaminyl aklavinone using TDP-2dF a natural sugar donor.[41] In addition, NbrA shares excellent sequence homology with several polyketide macrolide glycosyltransferases, such as NbmD (53.6/39.4) for norbomycin glycosylation in *S. narbonensis*, OleG1 (54.7/38.8) and OleG2 (56.3/41.7) for oleandomycin in *S. antibioticus*, and DesVII (54.3/40.6) for (neo-)methymycin/picromycin in *S. venezuelae* (%similarity/ %identity). Interestingly, another glycosyltransferase encoded by *nbrB* displays high sequence similarity to the members of MGT family, which include antibiotic resistance macrolide glycosyltransferases for inactivation of macrolide antibiotics. It has been shown that in the endogenous host the resistance to macrolide antibiotics can be achieved by macrolide glycosylation to prevent binding to their cellular targets.[42] Indeed, NbrB exhibits a high level of sequence similarity to OleI (44.3/29.1) and OleD (44.7/29.8) previously shown to carry out the glycosylation and inactivation of oleandomycin and diverse macrolide antibiotics, respectively, in *S. antibioticus*.[42] In light of above observations, NbrA may serve as a preferable candidate to act as a 2-deoxy-L-fucosyltransferase for attachment of the 2dF moiety onto the polyketide aglycone, whereas NbrB may have an opportunity to take part in antibiotic resistance glycosylation. Another interesting finding is that immediately downstream of *nbrH* (the last *nbr* PKS-I gene in biosynthesis) are followed in order *nbrA*, *orf46* (a TetR/AcrR family transcriptional regulator gene) and *nbrB*, where all of them are arranged in the same (head-to-tail) orientation on the chromosome.[43] In light of this genetic organization, it is likely that the *orf46* sitting upstream of *nbrB* might control the expression of *nbrB* in transcriptional level, thereby regulating the glycosylation of brasilinolides (or other antibiotics) as a mode of antibiotic resistance mechanism in *N. terpenica*. On the other hand, downstream of PKS-I genes is situated *orf56* coding for a third glycosyltransferse, which by homology is related to membrane-associated phosphomannose transferase enzymes generally involved in lipoarabinomannan biosynthesis. Similar situation was observed in tautomycetin (*ttn*) biosynthesis, where *ttn* ORF1 (507 a.a., 48.3/31.8), a close homolog of *nbr* ORF56 (509 a.a.), was found adjacent to *ttn* biosynthesis genes and, however, proved to be irrelevant to *ttn* biosynthesis by inactivation experiments.[44] Therefore, *orf56* may as well not be directly associated with brasilinolide biosynthesis.

Monooxygenase genes located within the 210kb genomic contig

Within our 210kb sequenced genomic contig were revealed totally six monooxygenase-encoding *orfs* (*orfs 12*, *23*, *48*, *70*, *74* and *75*), where *orf23* and *orf48* coded for cytochrome P450 enzymes and the others for flavin-dependent enzymes. Physically, *orf23* and *orf48* are closely clustered with PKS-I and NDP-2dF genes, whereas the others located at two sides of the contig with farther distances to PKS-I genes. Sequence analysis of *orf23*- and *orf48*-transcribed products revealed classical motifs of cytochrome P450 monooxygenases.[45] As shown in Fig. 4 (ORF23) and Fig. S12 (ORF48, ESI†), in both ORF23 and ORF48 were found the motifs featured with the highly conserved heme binding pocket (GxxxCxG) holding the invariable cysteine (C) in the β-bulge preceding the L helix, the conserved threonine (T) in the putative I-helix proposed to be involved in O_2 binding, and the E/DxxR motif in the putative K helix also for heme binding.[46,47] In particular, the protein encoded by *orf23* was found to share high sequence homology with several hydroxylases in secondary metabolite biosynthesis, including FcpC (50.7/39.3) in isonuatigenone biosynthesis, FosG (51.2/35.6) in fostriecin biosynthesis, SlnF/SalD (46.0/32.9) in salinomycin biosynthesis from *Streptomyces albus* XM211/DSM41398, and SalC (55.5/40.0) in salinomycin biosynthesis from *S. albus* CCM4719, CYPome (45.8/32.8) in cyclosporine biosynthesis, and PtmO5 (53.5/41.8) in platensimycin/platencin biosynthesis, respectively.[47-50] Most of above enzymes are involved in hydroxylation of type I modular polyketide metabolites similarly as brasilinolides. On the other hand, despite the common P450 motifs in ORF48 and ORF23, ORF48 shared only relatively low sequence homology (37.1/23.7) with ORF23, suggesting these two belong to different subclasses of cytochrome P450 monooxygenases. Notably, our phylogenetic analysis of the hydroxylases and epoxidases involved in biosyntheses of secondary metabolites revealed that, unlike ORF23, ORF48 tended to cluster in evolutionary relationship with several cytochrome P450 and flavin-dependent epoxidases of PK-I biosynthesis in addition to some cytochrome P450 enzymes for hydroxylation (Fig. S13 and Table S4, ESI†). Therefore, it is likely that ORF23 might serve as a preferable candidate responsible for the hydroxylation at C16 of brasilinolides and ORF48 for the C28-29 epoxidation. Interestingly, in *nbr* gene cluster *orf48* and *orf23* seemed to pair up and highly resembled in a.a. sequence *salD* (62.7/49.5) and *salC* (55.5/40.0), respectively, from *S. albus* CCM4719, which also paired together in salinomycin biosynthetic gene cluster involving both epoxidation and hydroxylation.[49,50] This may suggest that ORF48 and ORF23 possibly play similar roles in brasilinolide biosynthesis as SalD

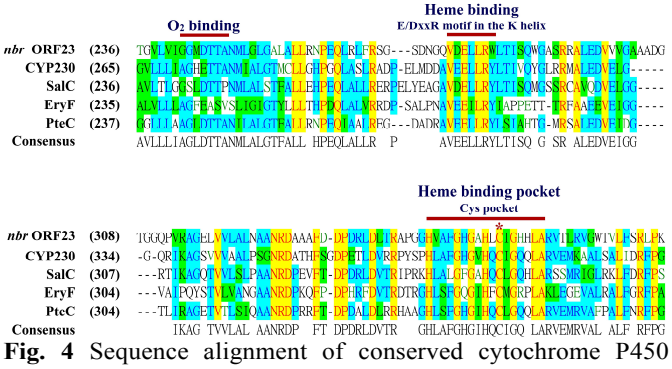


Fig. 4 Sequence alignment of conserved cytochrome P450 domains of ORF23 and its homologs. ORF23 from *Nocardia terpenica*; CYP230 from *Streptomyces tubercidicus*; SalC from *Streptomyces albus* CCM4719; EryF from *Saccharopolyspora erythraea* NRRL 2338; PteC from *Saccharopolyspora erythraea* NRRL 2338. The Cys residue coordinating the heme is indicated by an asterisk (*).

and SalC, respectively, in salinomycin biosynthesis of *S. albus* CCM4719.

It is worth noting that cytochrome P450 monooxygenases serving as epoxidases, instead of hydroxylases, have been found in numerous examples of PK-associated macrolide biosynthesis, e.g., PimD in pimaricin biosynthesis (*S. natalensis*), RosD in rosamicin biosynthesis (*M. rosaria* IFO13697), EpoK in epothilone biosynthesis (*S. cellulosum*), ScnD in natamycin biosynthesis (*S. chattanoogensis*) and CrpE in cryptophycin biosynthesis (*Nostoc sp.* ATCC 53789). These P450 macrolide epoxidases seemed to share only fair overall sequence homology (26-47% similarity and 15-28% identity) but tended to be closer to each other, as well as to FAD-dependent macrolide epoxidases (e.g. MonCI from Monensin A biosynthesis), in phylogenetic relationship (Fig. S13 and Table S4, ESI†). Also worth noting is that the PimD epoxidase shares rather high sequence identity (>54%) with NysL and AmphL P450 monooxygenases, which are in fact hydroxylases in nystatin and amphotericin B biosynthesis, respectively.[51] Hence, it should be an interesting subject of further study to unveil the exact factors determining the identity of a P450 monooxygenase to be either a hydroxylase or epoxidase in macrolide post-modifications. For the rest of the monooxygenase-encoded *orfs*, sequence analysis revealed that the other four *orfs* (*orf12*, *orf70*, *orf74 and orf75*) coded for FAD-dependent monooxygenases of different subclass and might not be directly associated with the biosyntheses of brasilinolides in *N. terpenica* as discussed in details in supplementary information.

Candidate genes coding for tailoring enzymes

In the gene cluster was found *orf34* highly resembling the genes encoding members of polyketide-associated protein A5 (PapA5) family in amino acid sequence. The PapA5 proteins were once speculated to be involved in polyketide transport, but has recently been functionally characterized as a new class of acyltransferases in *Mycobacterium tuberculosis* and *Amycolatopsis mediterranei*.[52,53] In particular, the translational product of *orf34* contains the highly conserved motif, $Hx_3Dx_{14}Y$, in PapA5 and its close homologs (Fig. 5), where by site-directed mutagenesis and protein structure determination His124 and Asp128 in PapA5 from *M. tuberculosis* were found indispensable for catalytic activity by acting as a catalytic base and an active site stabilizer, respectively.[53,54] Interestingly, the PapA5 protein from *M. tuberculosis* displayed high degree of substrate tolerance towards substrate variants of both acyl donors (Acyl-CoA thioesters) and acyl acceptors (alcohol nucleophiles). Most notably, the protein encoded by *orf34* also

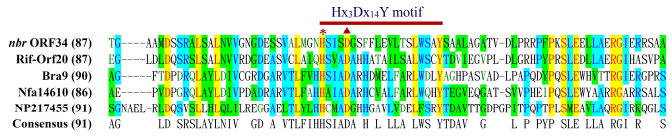


Fig. 5 The conserved motif shown by sequence alignment of the PapA5 family proteins. The aligned proteins are: Rif-Orf20 from *Amycolatopsis mediterranei*; NP217455 from *Mycobacterium tuberculosis* H37Rv; Nfa14610 from *Nocardia farcinica* IFM-10152; Bra9 from *Nocardia terpenica*; ORF34 from *Nocardia terpenica* (this study). Asterisk (*) indicates the histidine proposed as a catalytic base, whereas closed triangle (▲) indicates the aspartate as the active site stabilizer.

shared high sequence similarity and identity to Rif-Orf20 responsible for the acylation of DMDARSV to DMRSV in biosynthesis of rifamycin B, an antibacterial type I polyketide macrolide, where Rif-Orf20 was also found to utilize propionyl-CoA as an alternative substrate in addition to the natural acetyl-CoA substrate.[52] In light of above findings, the *orf34* in our *nbr* gene cluster would possibly take part in the acylation of the C23-OH with malonyl-CoA required in biosynthesis of brasilinolide A. Also possible is the acylation of the 3'-OH on the 2dF moiety of the brasilinolides with pentanoyl-CoA throughout catalytic action of the acyltransferase encoded by *orf34*.

The other candidate possibly serving for the acylation (esterification) of the 3'-OH or the C23-OH in brasilinolide A biosynthesis would be the type B carboxylesterase encoded by *orf40*, which showed high sequence homology with TtmK (63.9/50.9) proposed for the (trans)esterification between the dialkylmaleic anhydride and the PK secondary alcohol in tautomycin (TTM) biosynthesis.[55] ORF40 also displayed good sequence similarity with Cpz21 (45.8/32.9) functionally proved to be an acyltransferase catalyzing the acylation of a linear secondary alcohol with 3-methylglutaryl-CoA in caprazamycin biosynthesis.[56] Similarly, for esterification at the C23-OH, the malonyl-CoA might thus be directly used as a substrate for the *orf40*-encoded carboxylesterase. In the case of the esterification at 3'-OH, the pentanoyl moiety might be incorporated as a CoA-activated form possibly furnished by the *orf39*-encoded protein, which in sequence resembled TtmP (77.9/60.9) of TTM biosynthesis that is highly similar to CaiB (40.8/24.5), a member of the type III acyl-CoA transferase family catalyzing the CoA transfer between γ-butyrobetaine-CoA and carnitine forming carnityl-CoA and γ-butyrobetaine.[57]

Interestingly, immediately downstream of the PKS- and TEII-encoded *orfs* was found *orf52* possibly coding for another acyltransferase. Sequence analysis of *orf52* assigned its transcribed product (ORF52) to be a potential member of α/β-hydrolases. Like many members of the α/β-hydrolase superfamily, ORF52 shared low sequence similarity with the family members, most of which were biochemically uncharacterized. ORF52 (271 a.a.), however, seemed to retain the nucleophile(GxSxG)-acid(D)-base(H) motif (catalytic triad) reported for α/β-hydrolases acting on different substrates in various biological contexts although the acid(D) was replaced by E (and/or the base(H) by D) in ORF52 (Fig. S14, ESI†). Among the characterized α/β-hydrolases, the closest sequence homolog to ORF52 was AidH (271 a.a., 35.8/21.1), which was reported as a novel *N*-acylhomoserine lactonase (AHLase) from an *Ochrobactrum sp.* strain and itself showed little similarity to other AHL-degrading enzymes.[58] On the other hand, of various acyltransferases utilizing different substrates, RhlA (295 a.a., 20.1/11.7) for rhamnolipid biosynthesis in *P. aeruginosa* and PapA5 (422 a.a., 17.1/8.1) as a phthiocerol dimycocerosyl transferase from *M. tuberculosis* showed closer phylogenetic relationship to ORF52 albeit with even lower sequence similarity.[59] Based on these observations, the functional role of *orf52* is hence unclear and remains as a new and interesting subject for further study. However, it may not be absolutely

ruled out the possibility that ORF52 might be capable of serving as an α/β-hydrolase-like acyltransferase acting on the 2dF moiety or the macrolide.

Regulatory and resistance genes in secondary metabolism

In the *nbr* gene cluster were found *orf35*, *orf36* and *orf37*, showing high sequence homology to *ttmN* (82.6/70.7), *ttmR* (65.3/55.7) and *ttmS* (71.3/57.3), respectively, involved in biosynthesis of tautomycin, a potent polyketide phosphatase inhibitor isolated from *Streptomyces spiroverticillatus*.[44,55] Although exact functional roles of their homologs in tautomycetin (TTN) and tautomycin (TTM) biosynthesis are still unclear, TtmS (TtnS) was found to possess a conserved HxGTHxDxPxH motif of the cyclase family. TtmR (TtnR) was homologous to the MmgE/PrpD family protein and assigned to be a putative dehydratase, where PrpD is a known 2-methylcitrate dehydratase involved in propionate catabolism. TtmN (TtnN) was proposed to act as a sensor of cellular energy levels since it highly resembles EhpF (AAN40895) from *Pantoea agglomerans* (35.8% identity) with a conserved AMP-binding domain. Notably, previous inactivation experiments of *ttnR* and *ttnS* (or *ttmS*) alone completely abolished production of TTN (or TTM), and, most importantly, led to complete lack of any polyketide or related metabolites.[44] It was also postulated that *ttnRS* (*ttmRS*) might prohibit the ensuing enzyme-dependent events related to polyketide elongation and/or release, whereas *ttnN* (*ttmN*) could play regulatory functions modulating polyketide biosynthesis. Based on above observations, like *ttnNRS* (*ttmNRS*), *orfs 35~37* clustering together also in *nbr* gene cluster may play similar roles respectively.

Several important *orfs* presumably involved in secondary metabolite regulation and resistance were also identified. In the neighboring region of the NDP-2dF genes were located the *orfs* coding for ABC-2 type transporters (*orf25*, *orf29* and *orf30*) and LuxR family transcriptional regulator (*orf22*), all of which are in the same orientation as NDP-2dF and PKS-I genes. The three transporter *orfs* may be responsible for drug transport and/or resistance. The deduced product of *orf22* shared good sequence homology to the transcriptional regulators generally involved in biosynthesis of type I polyketide macrolides: SalRIII (53.6/36.0) from salinomycin biosynthesis, TetrRIV (46.8/35.6) from tetramycin biosynthesis, PimM (45.4/31.5) from pimaricin biosynthesis, and AmphRIV (47.2/31.7) amphotericin biosynthesis, and ScnRII (46.2/31.5) from natamycin biosynthesis. PimM has been shown to mainly regulate genes involved in initiation and first elongation cycles of polyketide chain extension.[60] By gene knock-in experiments, ScnRII has been demonstrated to promote natamycin production by 3-5 folds.[61] Like PimM and ScnRII as positive regulators, ORF22 harbors an N-terminal PAS-like domain presumably involved in sensing environmental variables and cellular energy, as well as a downstream DNA-binding domain of the LuxR family. The PAS domain of NysRIV in nystatin biosynthesis was suggested to respond to the energy levels, leading NysRIV to directly control nystatin biosynthesis.[62]

Downstream of PKS-I genes were also found several *orfs* coding for transcriptional regulators (*orfs 55*, *57*, *76*, *79* and *81*) and transporters (*orf54* and *orf66*) (Table S2, ESI†). ORF55 (545 a.a.) was found to only contain N-terminal half (DNA-binding domain) of its homolog members (920-935 a.a., *ca.* 30% identity) of LuxR family regulatory proteins often associated with secondary metabolite biosynthesis, whereas ORF57 to be a member of the TetR family proteins (TFRs) involved in regulating antibiotic resistance usually as repressors. ORF54 shares fair sequence similarity with AlbD (35.8/19.7) proposed for transport of albonoursin or other diffusible molecules across membranes in *S. noursei*.[63] ORF66 is homologous to members of the major facilitator superfamily (MFS) transporters, some of which are involved in specific effluxes of secondary metabolites. ORF76 in sequence resembles DNA-binding proteins StrR (41.0/31.6) and KasT (45.5/35.5) serving as pathway-specific activators for regulating streptomycin and kasugamycin biosynthesis gene clusters in *S. griseus* and *S. kasugaensis*, respectively.[64] Based on sequence homology, ORF79 and ORF81 may not be associated with regulation of the *nbr* gene cluster. More regulator-related *orfs* (*orfs 6* and *13-15*) were also found upstream of the 210kb region. As a potential member of the MerR family transcriptional regulators, ORF15 (312 a.a.) is a close homolog of the ORF46 (312 a.a., 61.0/51.7) reported as a candidate regulator appearing in the biosynthetic gene cluster of lipopeptide antibiotic A54145 in *S. fradiae*. By homology, ORF13 is related to CarRI (43.3/31.0), an IclR family regulator associated with carbazole catabolism in *Nocardioides aromaticivorans* IC177, whereas ORF6 only shares limited similarity, mostly on DNA-binding domain, with a TetR family regulator CprB (29.0/16.7) associated with antibiotic production in *S. coelicolor* A3(2).[65] Also, ORF14 is related to Lsr2, which is a regulatory protein involved in multiple cellular processes including cell wall biosynthesis and antibiotic resistance. With such a variety of regulators, a regulation cascade associated with the *nbr* gene cluster is thus of considerable interest for further studies.

Miscellaneous genes embedded in the 210kb genomic contig

On the two sides of the 210kb sequenced region were revealed several *orfs* coding for proteins mainly involved in primary metabolism, as well as those involved in secondary metabolism but not closely associated with brasilinolide biosynthesis, suggesting that the *nbr* gene cluster has reached the boundaries as judged by bioinformatic analysis. Especially, there is an increasing number of *orfs* coding for hypothetical proteins as the sequence moves towards the ends (Fig. 2; Table S2, ESI†). Among all these other *orfs*, *orf7*, *orf9*, *orf24* and *orf59* are of more interest in terms of functional roles and therefore are further discussed in supplementary section (ESI†).

Functional expression and characterization of the TDP-2-deoxy-L-fucose biosynthetic genes for brasilinolide biosynthesis.

Bioinformatic analysis predicted the transcribed products of *nbrCDEF* genes would involve deoxysugar biosynthesis (Table S2, ESI†). Yet, their exact functions, specificity and mechanisms, including stereochemistry of reduction (NbrD and NbrF) and regiochemistry of epimerization (NbrE), remain obscure. To functionally characterize and reconstitute the biosynthesis of NDP-

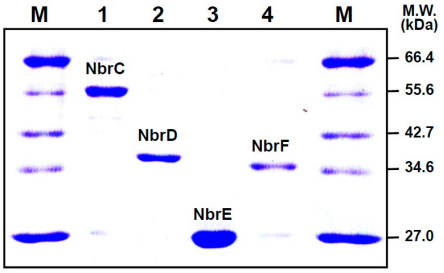


Fig. 6 SDS-PAGE analysis of purified recombinant NbrC, NbrD, NbrE and NbrF. Molecular weights calcd. for C-His$_6$ NbrC, N-His$_{10}$-NbrD, C-His$_6$-NbrE and N-His$_{10}$-NbrF are 53.5 kDa, 39.4 kDa, 23.2 kDa and 35.3 kDa, respectively; M: Protein Marker (NEB).

2dF *in vitro*, a catalytic pathway for conversion of **1** to **5** was constructed as depicted in Scheme 1. Recombinant proteins of NbrCDEF were obtained each by PCR amplification with primers introduced with suitable restriction sites for cloning into pET expression vector system (Novagen). Under control of T7 promoter, the carrying plasmids were heterologously expressed in *E. coli* BL21 (DE3), where the conditions for culture growth and induction were optimized for each enzyme in attempt to obtain sufficient amount of soluble protein. NbrD, NbrE and NbrF, however, suffered from low yield of soluble proteins. The situation was finally resolved with introduction of chaperones, e.g. GroES/EL, to coexpress with target proteins for promoting proper protein folding. Nevertheless, the His-tagged proteins of NbrCDEF were each purified successfully to near homogeneity by Ni-NTA affinity chromatography (Fig. 6).

Measurements of the enzymatic activities displayed by NbrCDEF were accomplished by a split-and-combine manner to dissect and reconstitute the 2dF biosynthetic pathway *in vitro*. The reactions proceeded with first full conversion of KDG **1** to **3** by pre-mixed NbrC/D (molar ratio 1:2), followed by *in situ* co-addition of NbrE/F (7:2), to afford the reduced product **5** in the presence of NADPH (Fig. 7). Such *in situ* one-pot transformations allowed observation or capture of the intermediate state of the biosynthesis, whereas concomitant elimination of TDP from readily degradation of NbrC and NbrE products could be greatly reduced products could be greatly reduced by proper adjustment of assay conditions, e.g. timing and molar ratio in mixing of coupling enzymes. Subsequent SAX-HPLC analysis (Fig. 7) and structural elucidation by ESI-MS ([M-H]$^-$ calcd. 531.1, measured 530.7) suggested formation of TDP-2dF **5**.

Further confirmation of the TDP-2dF biosynthesis by NbrCDEF can be made by characterization of catalytic mechanism and stereochemical courses of the individual enzymes. This was satisfied by conducting mix-and-match pairings of NbrCDEF enzymes with other NDP-sugar biosynthetic enzymes of known function and stereochemistry (e.g., SpnO, SpnN, Epi and DnmV; Figs. 8 & 9),[39,66,67] where the receiving enzyme involved in a coupling (mix) reaction may act a proofreading (match) tool to verify chemical identity of the product from the preceding enzyme, and vice versa, with support from HPLC resolution.

As shown in Fig. 8, in the mix-and-match experiments NbrC (or SpnO) acting as a 2,3-Deh converted KDG **1** to TDP-

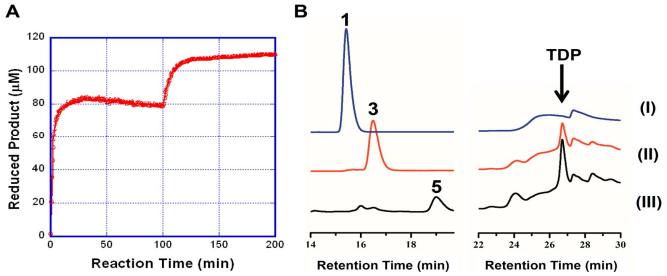


Fig. 7 *In vitro* biosynthesis of TDP-2-deoxy-L-fucose by tandem enzymatic reactions. Panel (A) Spectroscopic detection of NADPH consumption at 340nm in NbrC/D and NbrE/F coupling reactions; Panel (B) SAX-HPLC traces of the tandem reactions: (I) KDG (control) (II) NbrC/D reactions with KDG and NADPH (III) consecutive reactions by *in situ* addition of NbrE/F to the NbrC/D reactions.

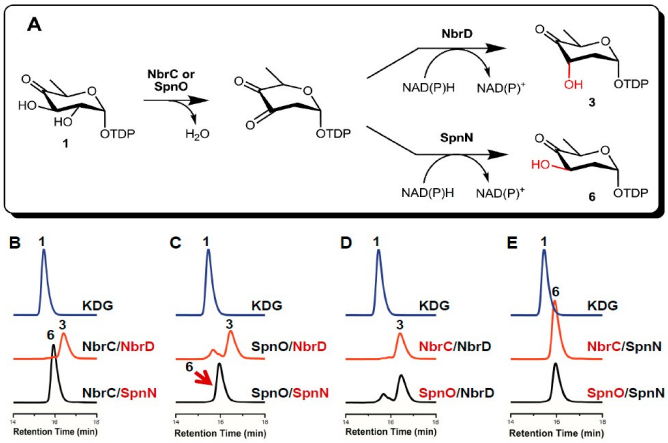


Fig. 8 Systematic comparisons of HPLC analysis profiles of enzymatic coupling reactions of NbrC/SpnO with NbrD/SpnN. (A) Different stereochemical fates operated by NbrD and SpnN acting on the NbrC/SpnO diketo product. (B-E) Comparisons between incubation sets of KDG **1** with NbrC/NbrD, NbrC/SpnN, SpnO/NbrD or SpnO/SpnN.

-3,4-diketo-2,6-dideoxy-D-glucose or its 2,3-enol form. Upon subsequent capture by different 3-Kers (2,3-reductases), the 3-keto group of the diketo intermediate may be attacked by an NADPH hydride on either *si* or *re* face to give a hydroxyl group in *S* or *R* configuration, respectively, resulting in different biosynthetic fates. Previous studies have demonstrated that SpnN catalyzes a *re*-facial reduction yielding the 2-deoxy-KDG **6**.[66] Notably, NbrD adapted a distinct stereochemical course and afforded **3**, giving clearly different retention time on HPLC elution profile. In Figs. 8B-8E are shown four paired-comparisons of HPLC profiles from the coupling reactions catalyzed by a various combination of NbrC/SpnO and NbrD/SpnN, thereby validating the catalytic mechanisms of NbrC and NbrD (Scheme S1, ESI†).

Substrate specificity and partner promiscuity of the TDP-2-deoxy-L-fucose biosynthesis enzymes.

The mix-and-match strategy was also applied on mechanistic characterization of NbrE and NbrF, where Epi (a 3,5-epimerase) and DnmV (a 4-ketoreductase) were employed as match partners.[39,67] Here, KDG **1** was utilized a stable, alternative substrate to functionally characterize NbrE and NbrF, with an

additional advantage of exploring their combinational potential and substrate specificity. Very interestingly, NbrE displayed striking substrate promiscuity towards **1** and gave the product **7**. In the NbrEF coupling reaction, the 4-keto sugar **7** was subsequently captured and stabilized by immediate stereospecific reduction at 4-keto with NbrF, leading to formation of TDP-6-deoxy-L-talose **8** as judged by HPLC (Fig. 9). **8** was further isolated and purified by strong anion-exchange chromatography (MonoQ, Amersham) and gel filtration (Sephadex G10, Amersham), where $NADP^+$, NADPH, TDP and TMP were eliminated. The structural elucidation of **8** was performed with ESI-MS (measured M.W. 547.03) and ^{1}H-NMR, giving characteristic signals well consistent with those previously reported for TDP-6-deoxy-L-talose (calcd. M.W. 547.07) (Figs. S15 & S16, ESI†).

Notably, **8** was also generated upon co-incubation of **1** with Epi and DnmV and also with either combination of Epi/NbrF or NbrE/DnmV by assistance of NADPH (Fig. 9C). The results suggested that in function NbrE match with Epi as a TDP-hexose-3,5-epimerase to generate L-form sugar, where the exact function of Epi has been verified in our previous report in spinosyn and TDP-L-rhamnose biosynthesis.[39] Moreover, NbrF resembles DnmV in both function and stereochemistry and acts as a TDP-hexose 4-ketoreductase by delivering NADPH hydride onto *re* face of the 4-keto group.[39,67] In light of the relaxed substrate specificity displayed by NbrF on **1**, we prompted to examine its degree of substrate tolerance towards various 4-keto sugars. Consequently, NbrF was found to exhibit excellent substrate promiscuity on both D-form (**1** and **6**) and L-form (**4** and **7**) sugars (Scheme 2). Nonetheless, NbrF did not seem to utilize **3** as an acceptable substrate and NADH as the coenzyme. Such intriguing observation indicates that the active site of NbrF may present a peculiar 3-dimentional conformation to selectively accommodate a sugar donor of certain conformation, and is not demanding on the presence of 2-OH.

Bioinformation and catalytic mechanisms of the TDP-2-deoxy-L-fucose biosynthesis enzymes

We hereby demonstrate the first report on the *in vitro* functional expression and mechanistic characterization of the genes (*nbrCDEF*) responsible for the biosynthesis of TDP-2dF, which serves as the key sugar donor for biosynthetic maturation of brasilinolides. In Scheme S1 (ESI†) are shown the catalytic mechanisms of the NbrCDEF enzymes suggested from this study, where the stereospecific courses of reductions by the 3-Ker (NbrD) and the 4-Ker (NbrF) are indicated. Interestingly, these four genes were found to share, in general, good degree of similarity (up to 60-80% similarity and 50-70% identity except AknQ from aclacinomycin biosynthesis) in amino acid sequence to their homologs present in biosynthetic gene clusters of aclacinomycin and elaiophylin, which also hold the deoxyfucose (2dF) as a key structural element in biological activity (Figs. 1 & 3; Table S5, ESI†).[2,4,68,69] Thus, the information gained from this study would be useful for biosynthetic investigations of the related 2dF-containing metabolites. Yet notably, both as a member of 3-Ker family,

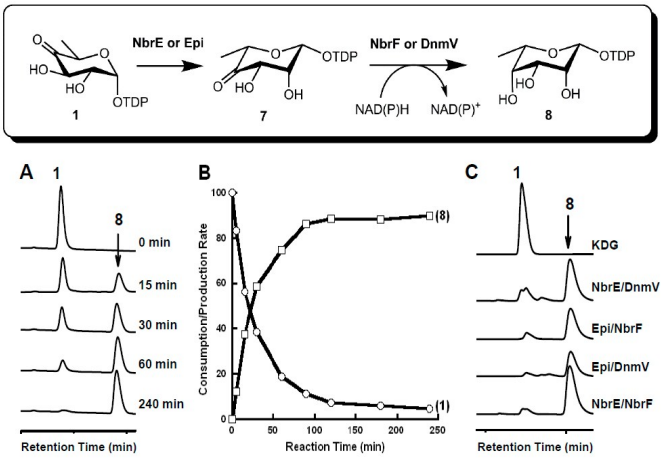
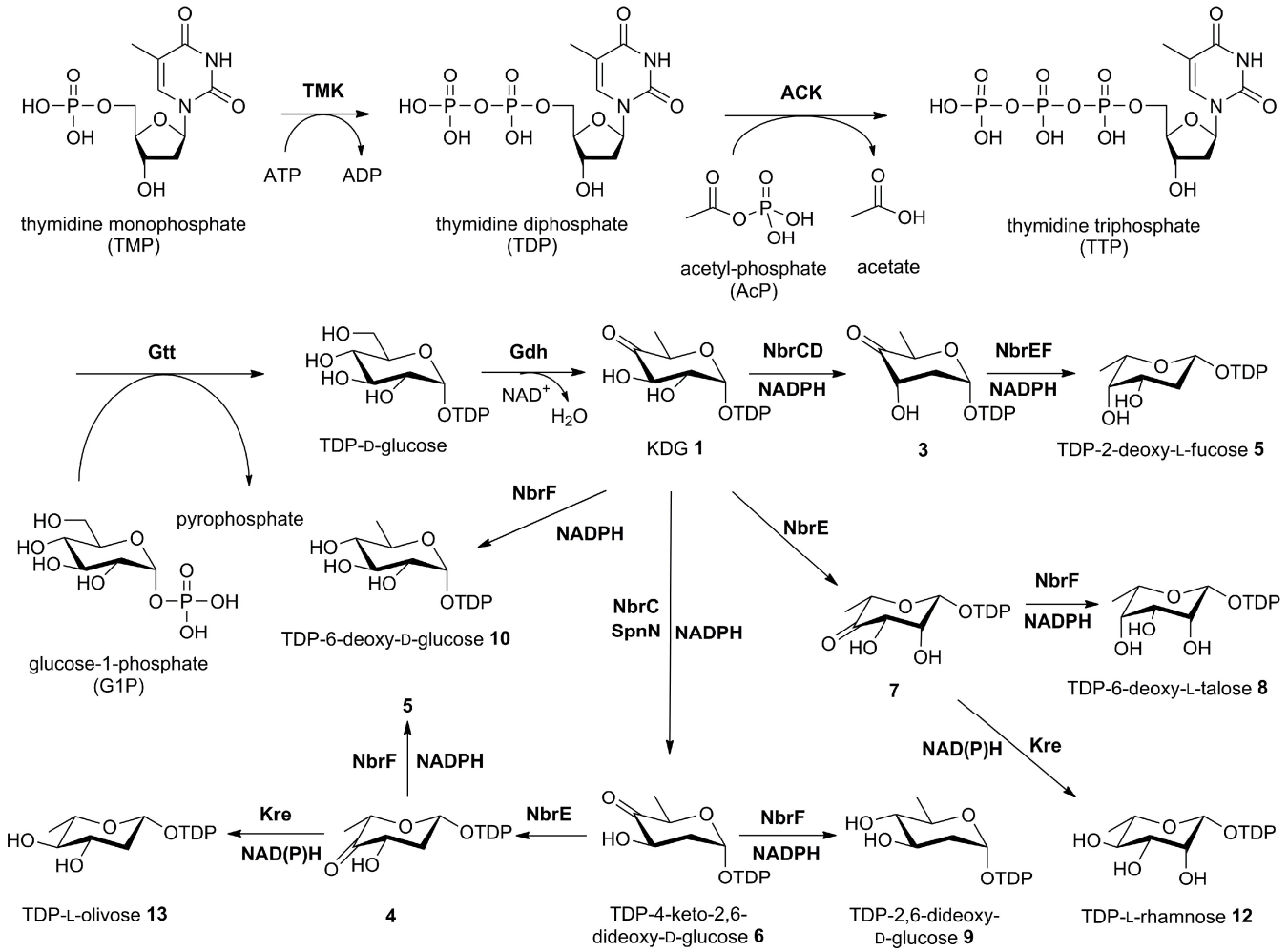


Fig. 9 Time-dependent HPLC analysis of NbrE/Epi and NbrF/DnmV cross-coupling reactions. (A)-(B) HPLC traces and time course of NbrE/F-coupled reactions on KDG (1). (C) HPLC analysis of four cross-coupling reactions between NbrE/Epi and NbrF/DnmV on KDG.

NbrD and AknQ greatly differ from each other in amino acid sequence homology (20.6% similarity and 11.6% identity). Our subsequent phylogenetic analysis of the 3-Ker family revealed two distinct subgroups, Group A and Group B, each represented by NbrD and SpnN, respectively (Fig. S17, ESI†). Very interestingly, our study described here earlier has demonstrated NbrD and SpnN carried out the 3-keto reduction via opposite stereochemical courses. Taken together, these findings may suggest that the mechanistic actions of 3-Ker enzymes in stereochemistry may well correlate with and be coded by the characteristics of their protein sequences and, presumably, protein structures. As shown in Fig. S17 (ESI†), the phylogenetic analysis also revealed that AknQ belongs to the Group B of SpnN. Hence, the observation may suggest that AknQ conduct the 3-keto reduction in opposite direction of stereochemistry as compared to NbrD (Group A). As a consequence, the above findings may allow us to postulate that biosynthesis of aclacinomycin could adapt an alternative path to accomplish the TDP-2dF formation as illustrated in Scheme 3, where the different stereochemical outcome caused by AknQ can actually be compensated (or reversed) by catalytic action of a 3,5-Epi presumably encoded by *aknL*. On the other hand, this may also provide a possible, and yet reasonable, explanation for the observation described for Fig. 3 that, unlike the *3-ker* genes in *nbr* and *ela* gene clusters, the one coding for AknQ in *akn* gene cluster stands alone toward different direction as compared to other NDP-2dF biosynthetic enzymes. In other words, the NDP-2dF biosyntheses in *nbr* and *ela* may thus likely share a better, closer level of evolutionary relationship than that in *akn*.

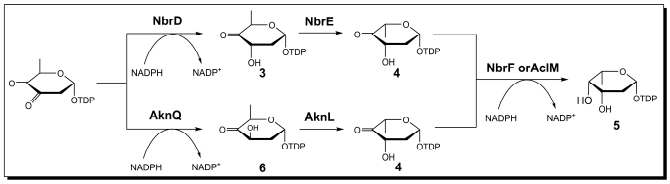
Combinatorial biosynthesis of NDP-deoxysugars by the TDP-2-deoxy-L-fucose biosynthesis enzymes.

The *in vitro* characterization experiments here revealed striking promiscuous nature of the TDP-2dF enzymes (NbrCDEF) towards their substrate and catalytic partners, which constitutes



Scheme 2 The *in situ* one-pot (iSOP) biosyntheses of various TDP-deoxysugars.

the key favourable factors to succeed the combinatorial biosynthesis for diversification of useful compounds.[70] To explore combinatorial synthetic utility of NbrCDEF in deoxysugar biosynthesis, in this study was also established an *in situ* one-pot (iSOP) synthetic platform (Scheme 2) for efficient and reproducible syntheses of various NDP-deoxysugars as useful activated sugar donors for prospective glycodiversification of natural or unnatural compounds.[70] The platform was derived from modifications of the reported procedure and was featured with a straightforward method carried out in one pot without complication by repetitive isolation and purification.[71] In principle, KDG (**1**) serving as the key branch point of NDP-deoxysugar biosynthesis cascade must be made during the process of the iSOP platform starting from commercially available and relatively cheap materials, e.g. thymidine-5'-phosphate (TMP), as illustrated in Scheme 2. With KDG generated *in situ*, various modification enzymes of KDG can be placed in the same one pot to give corresponding modified deoxysugar products. For final maturation of the modified products, the 4-Ker enzymes, NbrF and Kre (from spinosyn biosynthesis), were adapted here to introduce two different stereochemical outcomes of reduction at C4. Similarly, opposite stereochemistry of reduction at C3 can be furnished by 3-Kers (2,3-reductases), SpnN and NbrD. Furthermore, the 2,3-Deh (NbrC or SpnO) can be utilized for deoxygenation at C2. In addition, in this study was further examined the partner promiscuity of the coupling enzymes possibly involved in protein-protein interaction. RmlC (a 3,5-Epi) and RmlD (a 4-Ker), for example, were previously speculated to form a protein complex, which may facilitate substrate channelling to avoid decomposition of the unstable intermediate.[39,40] To examine the issue of partner promiscuity, either member of the NbrC-NbrD couple was substituted with the homologous enzyme, i.e. NbrC with SpnO or NbrD with SpnN, whereas, for the NbrE-NbrF couple, Epi and Kre (or DnmV) were used as the substitutes, respectively (Scheme 2). As a consequence, our experiments showed that the enzymes tested above displayed complete promiscuity towards their coupling-reaction partners and accomplished their own obligatory catalytic roles.



Scheme 3 Catalytic actions of NbrD and AknQ in TDP-2dF formation

In practice, the iSOP reactions operate with two sets of enzymes, the KDG-forming enzymes (TMK, ACK, Gtt and

Gdh) and the KDG-modifying enzymes (e.g., NbrCDEF, Epi, Kre, SpnON and DnmV) (Scheme 2). All enzymes were purified in large sale by Ni-NTA affinity chromatography and can be stored at -20℃ for at least 12 months without loss of activity (Fig. 6; Figs. S18 & S19, ESI†). The enzyme reactions were analyzed by SAX HPLC analysis (Fig. S20, ESI†). A typical iSOP reaction for KDG formation was carried out by direct mixing of (phospho-)nucleotide, phosphate donor, glucose-1-phosphate and magnesium ions with the KDG-forming enzymes in one pot. In this cost-effective iSOP system, phosphatase and ATP were not required to eliminate pyrophosphate ion and to serve as an exogenous phosphate donor, respectively. In such manner, KDG **1** was obtained with nearly quantitative conversion. To accomplish the iSOP version of syntheses of various TDP-deoxysugars, to the same one pot was added a various combination of KDG-modifying enzymes along with required cofactor(s). In some instances, the KDG-modifying enzymes needed to be added as separate sets of enzymes in sequential order with time intervals to afford single deoxysugar products. Nevertheless, all the reactions were carried out still in the same one pot without need to remove any undesired components. For exploring the combinatorial synthetic utility of the 2dF enzymes, in this study such system has successfully applied NbrCDEF onto the syntheses of several special TDP-deoxysugars, such as TDP-6-deoxy-L-talose (**8**), TDP-2,6-dideoxy-D-glucose (**9**), TDP-6-deoxy-D-glucose (**10**), TDP-L-rhamnose (**12**) and TDP-L-olivose (**13**), as shown in Scheme 2. Notably, glycosyltransferases can also be included in the platform for immediate glycosylation (data not shown), which is particularly useful for chemically labile TDP-deoxysugars.

Experimental

Bacterial strains and culture conditions

Nocardia terpenica was kindly provided by Professor Yuzuru Mikami (Chiba University, Japan). *N. terpenica* was cultivated in GSM medium (containing 1% polypeptone, 0.5 % meat extract and 2 % glycerol in 100 ml ddH_2O) at 28℃, 250 rpm for 55 hr to isolate genomic DNA.[7,72] *Escherichia coli* EPI300 were employed for genomic DNA library construction according to the protocol of CopyControl™ Fosmid Library Production Kit (Epicentre). *E. coli* XL1-Blue and *E. coli* BL21-Codon Plus (DE3)-RP (Stratagene) were used as the host strains in Luria-Broth (LB) medium and agar plates with appropriate antibiotics for DNA manipulation and protein expression, respectively.[73]

Materials and general methods

Restriction endonucleases and T4 DNA ligase were purchased from New England Biolabs (NEB). Polymerase chain reaction (PCR) was conducted using *pfu* DNA polymerase from Stratagene. The pCC1FOS vector (Epicentre) was employed for fosmid library construction. The pUC19 (NEB) were used as a cloning vector, whereas the pET21b and pET19b (Novagen) as expression vectors. Oligo primers were products of Operon Biotech Inc. All chemicals and organic solvents were commercially available from Sigma-Aldrich and Merck. Methods for cultivation of *E. coli*, molecular cloning and DNA isolation were performed under the standard procedures as described by Sambrook *et al*.[73]

Construction of the *Nocardia terpenica* fosmid library

Nocardia terpenica, a brasilinolide-producing strain, was used as the source of DNA for the construction of the genomic DNA library. The CopyControl fosmid library production kit (Epicentre) was used for *N. terpenica* fosmid library construction. The randomly sheared genomic DNA with size selected at *ca.* 33 kb were ligated into fosmid vector pCC1FOS, and then packaged into λ-phage (Gigapack kit, Stratagene), followed by transfection into *E. coli* EPI300 (Epicentre). Consequently, the fosmid library was successfully constructed with a total of 5,760 fosmid clones.

Probe design and PCR-screening for identification of the putative gene cluster for brasilinolide biosynthesis

Two pairs of degenerate primers (KSD1F/KSD1R and O_1F/O_3R) designed based on highly conserved regions among homologous domains and genes of PKS-I KS and KDH 2,3-Deh, respectively, were used to amplify the target gene-containing DNA fragments from *N. terpenica* genomic DNA. The resulting PCR-amplified fragments of expected size (*ca.* 330 bp and 350 bp, respectively) were cloned and verified by DNA sequencing to be the target genes. Subsequently, the specific primer pairs (PK1F/PK1R for the KS and SOHC10F/SOHC10R for the 2,3-Deh) designed based on the obtained DNA sequences were applied to screen our constructed fosmid library for the putative gene cluster for brasilinolide biosynthesis. As a consequence, three positive fosmid clones (pYC41D8, pYC41H8 and pYC44F12) containing the PKS gene fragment were identified and found to overlap with each other, thereby giving a genomic contig as analyzed by restriction mapping with restriction enzymes (e.g. *Bam*HI) (Fig. S3, ESI†). Furthermore, two specific primer pairs (41H8T7F/41H8T7R and 41D8HPNF/41D8HPNR) designed based on single end sequences of pYC41H8 and pYC41D8, respectively, were utilized to identify more fosmid clones (pYC50F9 and pYC48C11). Similarly, such strategy of "chromosomal walking" by PCR screening with a series of specific primer pairs designed based on single end sequences of the positive fosmid clones subsequently identified was applied to expand contig length towards each sides of the genomic contig on the chromosome. As a result, pYC12C1, pYC55F7 and pYC60D4 were found by specific primer pairs (41D8HPNF/41D8HPNR, PK50F9F/PK50F9R and PK55F7HPNF/PK55F7HPNR) based on pYC41D8, pYC50F9 and pYC55F7, respectively. Subsequently, the primer pairs (PK60D4HPNF/PK60D4HPNR) designed from pYC60D4 were utilized to find pYC49G9, pYC50A8 and pYC54H10. pYC12B2 and pYC12B10 were further identified by the primer pairs (PK54H10T7F/ PK54H10T7R) designed from pYC54H10, whereas pYC21D1 and pYC13F9 by the primer pairs (PK12B10HPNF/PK12B10HPNR) from pYC12B10.

On the other hand, the specific primer pairs (SOHC10F/SOHC10R) for 2,3-Deh were also employed to identify the positive clones pYC12C1 and pYC48C11, which were also found positive by way of KS primer probes as described. In a similar fashion, the genomic contig was further expanded, leading to identification of pYC14E6 by the primer pairs (PK48C11HPNF/PK48C11HPNR) designed based on pYC48C11. The restriction mapping of all 18 positive fosmid clones gave and concluded the final genomic contig spanning *ca.* 210 kb as shown in Fig. S3 **(**ESI†). The degenerate and specific primers used for PCR screening and their corresponding sequences are listed in Table S1 **(**ESI†).

DNA sequencing and ORF analysis

For thoroughly analysis of the giant genomic contig spanning *ca.* 210 kb, the insert DNAs of the 18 overlapping fosmids were carefully restriction-mapped and, at first, partially sequenced by end sequencing of restriction fragments (Fig. S3, ESI†). For complete sequencing of the 210kb contig, DNA sequencing was first conducted on the pYC48C11 fosmid harboring the TDP-2dF biosynthesis genes. The insert DNA of pYC48C11 was systematically digested by various combinations of appropriate restriction enzymes (e.g., *Bam*HI, *Sph*I, *Sac*I, *Kpn*I and *EcoR*I). The resulting restriction fragments of approx. 0.4 to 3 kb were subcloned into the cloning vectors (e.g. pUC19) for end sequencing. Assembly of the sequences resulted in *ca.* 2-3x sequence coverage over the insert DNA of pYC48C11, thereby revealing and confirming the TDP-2dF biosynthesis genes. As a second stage, an Illumina next-generation sequencing (NGS) method was utilized to obtain the contiguous DNA sequences for total sequencings of pYC14E6, pYC41D8, pYC55F7, pYC60D4, pYC54H10 and pYC13F9 fosmid DNAs. Each of the target fosmid DNA was first analyzed and quantified using the NanoDrop-1000 spectrophotometer (Thermo Scientific) and Qubit Fluorometer (Invitrogen). Subsequently, sequencing library of each fosmid was prepared, indexed and run simultaneously at Welgene Biotech or Genomics BioSci & Tech, Taiwan with paired reads of 250 bp each. Raw Illumina reads were then cleaned, and all the downstream analyses were based on the clean data with high quality. The resulting clean sequence reads were analyzed and assembled using Velvet 1.2.10 or Clc Genomics WorkBench 7.0. Software. As a final stage, a PCR-walking method was applied to close the gap between insert DNAs of pYC55F7 and pYC41D8 using pYC44F12 as a template. The DNA sequence assembly, contiguous alignments, and systematic analysis of the fosmids were performed on Vector NTI program (InforMax). Together with Vector NTI, the NCBI BLAST program (http://www.ncbi.nlm.nih.gov/BLAST) was utilized for identification of open reading frames (ORFs) as well as predictions of their functions by computer-aided database searching. The complete sequence of the entire genomic fragment spanning 210303 bp was thereby obtained and deposited in GenBank under accession number KP161205.

Construction of expression plasmids for His-tagged NbrC, NbrD, NbrE and NbrF proteins

Each of *nbrC*, *nbrD*, *nbrE* and *nbrF* genes was amplified by PCR on pYC48C11, and the corresponding PCR product (1428 bp for *nbrC*, 996 bp for *nbrD*, 612 bp for *nbrE* and 951 bp for *nbrF*) was each cloned into blunt-ended pUC19 at *Sma*I site for DNA manipulation. The corresponding primer pairs carrying appropriate restriction sites (underlined) designed for construction of expression plasmid were listed as follows: for *nbrC*, 5'-CAT ATG AAC ATG CCC GGC CTT C-3' and 5'-CTC GAG CCC CAA CCC GCG CAG-3'; for *nbrD*, 5'-CAT ATG CAC TAC TCC CGA C-3' and 5'-GGA TCC TCA TCG GGC CCA CGC CTC-3'; for *nbrE*, 5'-CAT ATG CAG TCA CGC AAG-3' and 5'-CTC GAG GAC CCG CCG TAG CG-3'; for *nbrF*, 5'-CAT ATG GGT GCT TCG AC-3' and 5'-CTC GAG TTA CCC AGA CTC CG-3'. The *nbrC* and *nbrE* PCR products carrying *Nde*I (5' end) and *Xho*I (3' end) sites were cloned into the corresponding sites of pET21b, giving pQH19 and pQH25 plasmids, respectively, for expression of C-terminal His_6-tagged recombinant proteins. On the other hand, the *nbrD* and *nbrF* PCR products preserving the stop codons were cloned into corresponding restriction sites (*Nde*I at 5' end for both genes; *Bam*HI at 3' end for *nbrD* and *Xho*I at 3' end for *nbrF*) of pET19b, resulting in the expression plasmids pQH28 and pQH29, respectively. All expression plasmids described above were verified by DNA sequencing.

Heterologous expression and purification of His-tagged NbrC, NbrD, NbrE and NbrF proteins

The expression plasmid containing *nbrC*, *nbrD*, *nbrE* or *nbrF* gene was independently transformed into *E. coli* BL21 (DE3). The *E. coli* transformant with the gene coding for C-terminal His_6-tagged NbrC was grown at 37℃ in Luria-Bertani (LB) medium with 100 μg/mL ampicillin until OD_{600} reached 0.5-0.6. After induction with 50 μM isopropyl *β*-D-1-thiogalactopyranoside (IPTG), the culture was incubated at 15℃, 250 rpm for 24 hr. For N-terminal His_{10}-tagged NbrD expression, the co-expression of NbrD with the chaperone encoded by pG-KJE7 were conducted and the transformant was grown at 37℃ in LB medium with 50 μg/ml ampicillin and 35 μg/ml kanamycin until OD_{600} reached 0.5-0.6. Subsequently, the culture was induced by 250 μM IPTG and 0.1 %($^{w}/_{v}$) arabinose at 15℃, 250 rpm for 48 hr. For expression of C-terminal His_6-tagged NbrE and N-terminal His_{10}-tagged NbrF, the cultures were both incubated in LB medium with 50 μg/ml ampicillin and 35 μg/ml kanamycin at 37 ℃ until OD_{600} reached 0.5-0.6. Subsequently, the co-expressions with GroES/EL chaperones were allowed to proceed at 30℃, 250 rpm for additional 12 hr after induction with 50 μM and 250 μM IPTG, respectively.

The cells were then harvested by centrifugation (1902 xg, 15 min) and resuspended with 50 mM KH_2PO_4 buffer containing 15 % glycerol (pH 7.5). After breaking and disrupting cells by two passages through a French press (Spectronic Instruments) at 16,000 psi, cell debris was removed by centrifugation at 30,000 xg for 1 hr. The supernatant was then loaded onto Ni resin to purify the desired protein. For NbrC and NbrD purification, the His-tagged proteins were eluted with a linear

gradient of imidazole (0-500 mM) at a flow rate of 1 ml/min. The gradient was constituted with buffer A (50 mM KH_2PO_4, 10 % (v/v) glycerol, pH 7.5) and buffer B (50 mM KH_2PO_4, 500 mM imidazole, 10 % (v/v) glycerol, pH 7.5). The desired fractions were exchanged with 50 mM KH_2PO_4 buffer containing 15 % (v/v) glycerol (pH 7.5) by Amicon Ultra YM-30. To purify NbrE and NbrF, the similar procedure was used except that the buffer was at pH 7.8. The purified His-tagged recombinant proteins were then subjected to 12.5 % SDS-PAGE analysis and the concentration of each protein was determined by Bradford assay.[74]

Preparation of enzymes for the mix-and-match enzymatic assays

The genes encoding SpnO (NDP-hexose-2,3-dehydratase) and SpnN (NDP-hexose-3-ketoreductase) were obtained by PCR from the genomic DNA of *Saccharapolyspora spinosa* and individually cloned into pET21b to generate expression plasmids pKH8 and pJZ7, respectively. The corresponding PCR primer pairs carrying appropriate restriction sites (underlined) designed for construction of expression plasmid were listed as follows: for *spnO*, 5'-CAT ATG AGC AGT TCT GTC GAA GC-3' and 5'-CTC GAG TCG CCC CAA CGC CCA C-3'; for *spnN*, 5'-CAT ATG CGA AAG CCG GTG CGC AT-3' and 5'-CTC GAG TGT GGA CCC GCA CCG A-3'. The recombinant proteins were co-expressed with GroES/EL chaperones in *E. coli* BL21 (DE3) and the cultures were grown at 37 °C in LB medium with appropriate antibiotics until OD_{600} reached 0.4-0.6. For SpnO and SpnN, the cultures were induced with 250 μM and 50 μM IPTG, respectively, and then grew at the designated temperatures (15℃, 12 hr for SpnO and 30℃, 12 hr for SpnN). The cells were then harvested by centrifugation (1,902 xg, 15 min), broken by two passages through a French press (Spectronic Instrument) at 16,000 psi, and further centrifuged (30,000 xg, 1.5hr, 4℃) to remove cellular debris. The resulting crude extracts were purified using Ni-NTA affinity chromatography in 50 mM KH_2PO_4 (pH 7.5) with 10 % glycerol. Consequently, SpnO was eluted with the elution buffer containing 150-250 mM imidazole, and SpnN with the buffer containing 100-200 mM imidazole. The eluted fractions were concentrated and exchanged against 50 mM potassium phosphate buffer (pH 7.5) containing 20 % glycerol, and then stored at -20℃ and -80℃ for SpnO (*ca.* 6.7 mg/1L culture) and SpnN (*ca.* 11 mg/1L culture), respectively. Furthermore, the gene encoding DnmV was cloned into pET vectors from *Streptomyces peucetius*,[67] overexpressed in *E. coli*, and purified to near homogeneity in large scale for enzymatic assays (Chiu, H.-T., unpublished data). Cloning, expression and protein purification of Epi from *Saccharopolyspora spinosa* was as previously described.[39]

Tandem enzymatic reactions of NbrCDEF

The NbrCDEF enzymatic reactions were performed in two steps and the reduction activities were monitored by UV-VIS absorption spectrometer. As the first step, NbrC/NbrD with molar ratio of 1:2 were incubated with KDG (13.3 nmol) and NADPH (20 nmol) in 50 mM KH_2PO_4 buffer (pH 7.5). The resulting reaction mixtures were incubated for 100 min. As the second step, NbrE/NbrF with molar ratio of 7:2 were added into the reaction mixtures, followed by addition of NADPH (20 nmol). The reduction activity was monitored for each step by measuring decrease in absorption intensity at 340nm at 30℃. Corresponding enzyme activity was calculated using NADPH molar absorption coefficient of 6220 M^{-1} cm^{-1} at 340 nm.

To analyze the formation of reaction intermediates in the tandem enzymatic reactions, NbrC (0.96 nmol) and NbrD (1.9 nmol) were first added into the reaction mixture (50 mM KH_2PO_4, pH 7.8) containing150 nmol NADPH and 100 nmol KDG as a substrate. The reaction was allowed to proceed at 30 ℃ for 180 min before consecutive additions of NbrE (7.92 nmol), NbrF (1.86 nmol) and NADPH (150 nmol) to furnish the TDP-2-deoxy-L-fucose biosynthetic pathway. Aliquots of the reaction mixtures were removed at various time points and subsequently mixed with an equal volume of ice-cold methanol to terminate the reaction. After centrifugation, the enzyme-free solutions were further subjected to HPLC analysis.

Mix-and-match enzymatic coupling reactions

The mix-and-match coupling reactions constituted by a various combination of the 2,3-dehydratase (NbrC or SpnO) and the 3-ketoreductase (NbrD or SpnN) were carried out in the reaction mixture (100 μl, 50 mM KH_2PO_4, pH 7.5) composed of 1.5 mM NADPH, 1 mM KDG and the coupling enzymes with molar ratio of 1:2 (dehydratase: ketoreductase). The reaction mixture was incubated in 30℃ water bath for appropriate time periods and subsequently terminated by an equal volume of ice-cold methanol for HPLC analysis.

In TDP-6-deoxy-L-talose formation, the 3,5-epimerase (NbrE or Epi) and the 4-ketoreductase (NbrF or DnmV) mixed with molar ratio of 10:1 were added to the reaction solution (100 μl) containing KH_2PO_4 buffer (50mM, pH 7.8), 1.5mM NADPH and 1mM KDG. The reaction mixture was incubated at 30℃ and then quenched by an equal volume of ice-cold methanol.

HPLC methods

For HPLC analysis of the enzymatic reactions, the reaction mixtures were treated with an equal volume of ice-cold methanol and flash frozen with liquid nitrogen to terminate enzyme activity. The reaction mixtures were then analyzed by HPLC (Agilent 1100) after centrifugation (16,000 xg at 4℃) for 60 min to remove the precipitated proteins. All HPLC analyses were conducted using a SAX column (ZORBAX SAX, 4.6 x 250 mm, 5 μm, Agilent) eluted at 1 ml/min and monitored at 267 nm. Unless otherwise specified, the general elution program was designed with a multi-stage gradient of potassium phosphate (pH 3.5). For HPLC analyses of compounds **4**, **5**, **8** and **9**, the elution sequence was programmed at 51.8 mM KH_2PO_4 for 3 min, a linear gradient of 51.8-196.2 (or 201.2) mM KH_2PO_4 for 6 min, 196.2 (or 201.2) mM KH_2PO_4 for 11 min, a linear gradient of 196.2 (or 201.2)-500 mM KH_2PO_4 for 1 min, and 500 mM KH_2PO_4 for 8 min. For analysis of compound **10**, the elution program was conducted in the order of 126.5 mM KH_2PO_4 for 3 min, a linear gradient of 126.5-

201.2 mM KH_2PO_4 for 17 min, a linear gradient of 201.2-500 mM KH_2PO_4 for 1 min, and 500 mM KH_2PO_4 for 5 min. For examination of compound **12**, the elution program was performed in the following sequence: 51.8 mM KH_2PO_4 for 3 min, a linear gradient of 51.8-171.3 mM KH_2PO_4 for 6 min, a linear gradient of 171.3-181.3 mM KH_2PO_4 for 12 min, a linear gradient of 181.3-500 mM KH_2PO_4 for 1 min, and 500 mM KH_2PO_4 for 8 min. For analysis of compound **13**, the elution was carried out at 51.8 mM KH_2PO_4 for 5 min, followed by a linear gradient of 51.8-76.7 mM KH_2PO_4 for 20 min.

General method for *In Situ* One-Pot (iSOP) biosyntheses of various NDP-deoxysugars

The iSOP reactions generally include two sets of enzymes, KDG-forming enzymes (TMK, ACK, Gtt and Gdh) and KDG-modifying enzymes (e.g., NbrCDEF, Epi, Kre, SpnON and DnmV). A typical iSOP reaction for KDG formation was carried out by direct mixing of thymidine-5'-phosphate **11** (TMP, 0.4 mmole), acetyl phosphate (AcP, 1.2 mmole), glucose-1-phosphate (G1P, 1.6 mmole) and $MgCl_2$ (0.4 mmole) with thymidine monophosphate kinase (TMK, 0.86 μmole), acetate kinase (ACK, 0.20 μmole), glucose-1-phosphate thymidylyltransferase (Gtt, 0.60 μmole) and TDP-glucose-4,6-dehydratase (Gdh, 0.56 μmole) in one pot (20 mM Tris-HCl, pH7.8, 20 mL) incubated at 37℃ for 12 hr for complete transformation. Occasionally, phosphatase could be added to eliminate pyrophosphate ion upon incubation at 37℃ for 1 hr. In such manner, KDG **1** was obtained in nearly quantitative (~97%) conversion (relative to TMP) as judged by analytical SAX HPLC. To create structural diversity of the TDP-deoxysugars, KDG-modifying enzymes of a various combination were added to a specified amount of the iSOP KDG mixture also in one pot for continued transformations. The continued enzymatic transformations could also be traced by SAX-HPLC analysis as illustrated in Fig. S20 (ESI†). In this study, the iSOP system furnished formations of TDP-4-keto-2,6-dideoxy-D-allose (TDP-kdA **3**), TDP-4-keto-2,6-dideoxy-D-glucose (TDP-2dKDG **6**), TDP-6-deoxy-L-talose (TDP-6dT **8**), TDP-2,6-dideoxy-D-glucose (TDP-26dG **9**), TDP-6-deoxy-D-glucose (TDP-6dG **10**), TDP-L-rhamnose (TDP-Rha **12**) and TDP-L-olivose (TDP-Oli **13**) upon treatment with related KDG-modifying enzymes and cofactors. For transformations, incubation of the aforementioned iSOP KDG mixture with NbrC, SpnN and NADPH gave rise to TDP-2dKDG (97% overall conversion), which was further converted to TDP-Oli (68% overall conversion) by treatment with NbrE, Kre and NAD(P)H. TDP-26dG (50~60% overall conversion) was obtained by addition of NbrC, SpnN and NADPH to the iSOP KDG to give TDP-2dKDG (97% overall conversion), followed by addition of NbrF and additional NADPH. Furthermore, to the iSOP KDG mixture TDP-6dG (58% overall conversion) was generated by addition of NbrF and NADPH, TDP-kdA (62% overall conversion) by introduction of NbrC, NbrD and NADPH, TDP-6dT (79% overall conversion) by addition of premixed NbrE, NbrF and NADPH, and TDP-Rha (97% overall conversion) by co-addition of NbrE, Kre and NAD(P)H.

Conclusions

In summary, by target-specific degenerate priming this work has successfully identified, cloned and sequenced a 210kb genomic fragment of *N. terpenica*, which harbors crucial biosynthetic elements, including PK-I scaffold construction, tailoring and regulatory genes, necessary for establishing brasilinolide biosynthesis. Detailed bioinformatic analysis of the 210kb sequence resolved functional relationship between coding sequences and biosynthesis. By using the elements, the iSOP synthetic platform utilized here serves as a demonstrating version of their promising applications in glycorandamization.[70] The functional expression and mechanistic characterization of the four key enzymes NbrCDEF represent the first example ever demonstrated *in vitro* in biosynthesis of TDP-2dF. The striking features of stereoselectivity and substrate/partner promiscuity revealed by use of various substrate analogs provide new insights into the molecular mechanism and recognition adapted by these enzymes. The information and materials obtained in this study thus lays the key foundation for future generation of unnatural natural product glycosides through combinatorial biosynthesis.

Acknowledgements

This research was supported by Ministry of Science and Technology (Taiwan) grants (MOST 102-2113-M-006-003-MY2 and MOST 104-2113-M-006-004) to Hsien-Tai Chiu. We thank Prof. Yuzuru Mikami (Chiba University, Japan) for kindly providing *N. terpenica*, and also HSP Research Institute (Kyoto Research Park, Japan) for providing pG-KJE7 as a gift. We also thank Dr. Yi-Lin Chen and Meng-Na Lee for technical assistance and discussion.

Notes and references

1. C. J. Thibodeaux, C. E. Melancon and H.-w. Liu, *Nature*, 2007, **446**, 1008-1016.
2. B. A. Haltli, U. S. Patent 7,595,187 B2, 2009.
3. S. F. Haydock, T. Mironenko, H. I. Ghoorahoo and P. F. Leadlay, *J. Biotechnol.*, 2004, **113**, 55-68.
4. K. Räty, J. Kantola, A. Hautala, J. Hakala, K. Ylihonko and P. Mäntsälä, *Gene* 2002, **293**, 115-122.
5. E. L. Schwartz and A. C. Sartorelli, *Cancer Res.*, 1982, **42**, 2651-2655.
6. A. C. Weymouth-Wilson, *Nat. Prod. Rep.*, 1997, **14**, 99-110.
7. Y. Tanaka, H. Komaki, K. Yazawa, Y. Mikami, A. Nemoto, T. Tojyo, K. Kadowaki, H. Shigemori and J. Kobayashi, *J. Antibiot.*, 1997, **50**, 1036-1041.
8. I. Paterson, M. P. Housden, C. J. Cordier, P. M. Burton, F. A. Mühlthau and O. Loiseleur, *Org. Biomol. Chem.*, 2015, **13**, 5716-5733.
9. J. Mann, *Nat. Prod. Rep.*, 2001, **18**, 417-430.
10. Y. Mikami, H. Komaki, T. Imai, K. Yazawa, A. Nemoto, Y. Tanaka and U. Gräefe, *J. Antibiot.*, 2000, **53**, 70-74.
11. K. Komatsu, M. Tsuda, Y. Tanaka, Y. Mikami and J. Kobayashi, *J. Org. Chem.*, 2004, **69**, 1535-1541.
12. Our sequence corresponding to the 210kb genomic fragment harboring the *nbr* gene cluster has been deposited in GenBank under accession number KP161205.
13. J. Staunton and K. J. Weissman, *Nat. Prod. Rep.*, 2001, **18**, 380-416.
14. C. Bisang, P. F. Long, J. Cortés, J. Westcott, J. Crosby, A. L. Matharu, R. J. Cox, T. J. Simpson, J. Staunton and P. F. Leadlay, *Nature*, 1999, **401**, 502-505.

15. C. Hertweck, *Angew. Chem. Int. Ed.*, 2009, **48**, 4688-4716.
16. S.-C. Tsai and B. D. Ames, *Methods Enzymol*., 2009, **459**, 17-47.
17. G. Yadav, R. S. Gokhale and D. Mohanty, *J. Mol. Biol.*, 2003, **328**, 335-363.
18. S. Anand, M. V. Prasad, G. Yadav, N. Kumar, J. Shehara, M. Z. Ansari and D. Mohanty, *Nucleic Acids Res.*, 2010, **38**, W487-W496.
19. D. H. Kwan and F. Schulz, *Molecules*, 2011, **16**, 6092-6115.
20. S. Smith and S.-C. Tsai, *Nat. Prod. Rep.*, 2007, **24**, 1041-1072.
21. H. Petković, A. Sandmann, I. R. Challis, H.-J. Hecht, B. Silakowski, L. Low, N. Beeston, E. Kušer, J. Garcia-Bernardo, P. F. Leadlay, S. G. Kendrew, B. Wilkinson and R. Müller, *Org. Biomol. Chem.*, 2008, **6**, 500-506.
22. S. Heia, S. E. Borgos, H. Sletta, L. Escudero, E. M. Seco, F. Malpartida, T. E. Ellingsen and S. B. Zotchev, *Appl. Environ. Microbiol.*, 2011, **77**, 6982-6990.
23. L. Katz, *Methods Enzymol.*, 2009, **459**, 113-142.
24. S. Yuzawa, C. H. Eng, L. Katz and J. D. Keasling, *Biochemistry*, 2013, **52**, 3791-3793.
25. G. F. Liou, J. Lau, D. E. Cane and C. Khosla, *Biochemistry*, **2003**, ***42***, 200-207.
26. A. T. Keatinge-Clay, *Chem. Biol.*, 2007, **14**, 898-908.
27. A. Baerga-Ortiz, B. Popovic, A. P. Siskos, H. M. O'Hare, D. Spiteller, M. G. Williams, N. Campillo, J. B. Spencer and P. F. Leadlay, *Chem. Biol.*, 2006, **13**, 277-285.
28. Y.-O. You, C. Khosla and D. E. Cane, *J. Am. Chem. Soc.*, 2013, **135**, 7406-7409.
29. C. R. Valenzano, Y.-O. You, A. Garg, A. T. Keatinge-Clay, C. Khosla and D. E. Cane, *J. Am. Chem. Soc.*, 2010, **132**, 14697-14699.
30. L. Tang, Y. J. Yoon, C. Y. Choi and C. R. Hutchinson, *Gene*, 1998, **216**, 255-265.
31. T. Brautaset, O. N. Sekurova, H. Sletta, T. E. Ellingsen, A. R. StrŁm, S. Valla and S. B. Zotchev, *Chem Biol.*, 2000, **7**, 395-403.
32. D. H. Kwan, Y. Sun, F. Schulz, H. Hong, B. Popovic, J. C. Sim-Stark, S. F. Haydock and P. F. Leadlay, *Chem. Biol.*, 2008, **15**, 1231-1240.
33. A. T. Keatinge-Clay, *Nat. Prod. Rep.*, 2012, **29**, 1050-1073.
34. V. Y. Alekseyev, C. W. Liu, D. E. Cane, J. D. Puglisi and C. Khosla, *Protein Sci.*, 2007, **16**, 2093-2107.
35. Y. Y. Wang, X. X. Ran, W. B. Chen, S. P. Liu, X. S. Zhang, Y. Y. Guo, X. H. Jiang, H. Jiang and Y. Q. Li, *FEBS Lett.*, 2014, **588**, 3259-3264.
36. (1) M. Kotowska and K. Pawlik, *Appl. Microbiol. Biotechnol.*, 2014, **98**, 7735-7746 (2) C. Olano, B. Wilkinson, C. Sánchez, S. J. Moss, R. Sheridan, V. Math, A. J. Weston, A. F. Braña, C. J. Martin, M. Oliynyk, C. Méndez, P. F. Leadlay and J. A. Salas, *Chem. Biol.*, 2004, **11**, 87-97.
37. M. Kotowska, K. Pawlik, A. R. Butler, E. Cundliffe, E. Takano and K. Kuczek, *Microbiology*, 2002, **148**, 1777-1783.
38. (1) H.-T. Chiu, Y.-L. Chen, C.-Y. Chen, C. Jin, M.-N. Lee and Y.-C. Lin, *Mol. BioSyst.*, 2009, **5**, 1180-1191. (2) H.-T. Chiu, Y.-C. Lin, M.-N. Lee, Y.-L. Chen, M.-S. Wang and C.-C. Lai, *Mol. BioSyst.*, 2009, **5**, 1192-1203.
39. Y.-L. Chen, Y.-H. Chen, Y.-C. Lin, K.-C. Tsai and H.-T. Chiu, *J. Biol. Chem.*, 2009, **284**, 7352-7363.
40. A. Melo, L. Glaser, *J. Biol. Chem.*, 1968, **243**, 1475-1478.
41. W. Lu, C. Leimkuhler, M. Oberthu¨r, D. Kahne and C. T. Walsh, *Biochemistry*, 2004, **43**, 4548-4558.
42. L. M. Quirós, I. Aguirrezabalaga, C. Olano, C. Méndez and J. A. Salas, *Mol. Microbiol.*, 1998, **28**, 1177-1185.
43. W. Deng, C. Li and J. Xie, *Cell Signal.*, 2013, **25**, 1608-1613.
44. W. Li, Y. Luo, J. Ju, S. R. Rajski, H. Osada and B. Shen, *J. Nat. Prod.*, 2009, **72**, 450-459.
45. V. Jungmann, I. Molnaŕ, P. E. Hammer, D. S. Hill, R. Zirkle, T. G. Buckel, D. Buckel, J. M. Ligon and J. P. Pachlatko, *Appl. Environ. Microbiol.*, 2005, **71**, 6968–6976.
46. A. Trefzer, V. Jungmann, I. Molnár, A. Botejue, D. Buckel, G. Frey, D. S. Hill, M. Jörg, J. M. Ligon, D. Mason, D. Moore, J. P. Pachlatko, T. H. Richardson, P. Spangenberg, M. A. Wall, R. Zirkle and J. T. Stege, *Appl. Environ. Microbiol.*, 2007, **73**, 4317-4325.
47. M. J. Lee, H. B. Kim, Y. J. Yoon, K. Han and E. S. Kim, *Appl. Environ. Microbiol.*, 2013, **79**, 2253-2262.
48. (1) W. Wang, F.-Q. Wang and D.-Z. Wei, *Appl. Environ. Microbiol.*, 2009, **75**, 4202-4205. (2) X.-J. Liu, R.-X. Kong, M.-S. Niu, R.-G. Qiu and L. Tang, *J. Nat. Prod.*, 2013, **76**, 524-529. (3) M. J. Smanskia, Z. Yub, J. Casperb, S. Linb, R. M. Peterson, Y. Chen, E. Wendt-Pienkowski, S. R. Rajski and B. Shen, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 13498-13503.
49. R. Knirschová, R. Nováková, L. Fecková, J. Timko, J. Turna, J. Bistáková and J. Kormanec, *Folia Microbiol.*, 2007, **52**, 359-365.
50. (1) C. Jiang, H. Wang, Q. Kang, J. Liu and L. Bai, *Appl. Environ. Microbiol.*, 2012, **78**, 994-1003. (2) M. E. Yurkovich, P. A. Tyrakis, H. Hong, Y. Sun, M. Samborskyy, K. Kamiya and P. F. Leadlay, *ChemBioChem*, 2012, **13**, 66-71.
51. O. Volokhan, H. Sletta, T. E. Ellingsen and S. B. Zotchev, *Appl. Environ. Microbiol.*, 2006, **72**, 2514-2519.
52. Y. Xiong, X. Wu and T. Mahmud, *ChemBioChem*, 2005, **6**, 834-837.
53. K. C. Onwueme, J. A. Ferreras, J. Buglino, C. D. Lima and L. E. Quadri, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 4608-4613.
54. J. Buglino, K. C. Onwueme, J. A. Ferreras, L. E. Quadri and C. D. Lima, *J. Biol. Chem.*, 2004, **279**, 30634-30642.
55. W. Li, J. Ju, S. R. Rajski, H. Osada and B. Shen, *J. Biol. Chem.*, 2008, **283**, 28607-28617.
56. L. Kaysser, L. Lutsch, S. Siebenberg, E. Wemakor, B. Kammerer, and B. Gust, *J. Biol. Chem*., 2009, **284**, 14987-14996.
57. P. Stenmark, D. Gurmu and P. Nordlund, *Biochemistry*, 2004, **43**, 13996-14003.
58. G. Y. Mei, X. X. Yan, A. Turak, Z. Q. Luo and L. Q. Zhang, *Appl. Environ. Microbiol.*, 2010, **76**, 4933-4942.
59. K. Zhu and C. O. Rock, *J. Bacteriol.*, 2008, **190**, 3147-3154.
60. N. Antón, J. Santos-Aberturas, M. V. Mendes, S. M. Guerra, J. F. Martín and J. F. Aparicio, *Microbiol.*, 2007, **153**, 3174-3183.
61. Y. L. Du, S. F. Chen, L. Y. Cheng, X. L. Shen, Y. Tian and Y. Q. Li, *J. Microbiol.*, 2009, **47**, 506-513.
62. O. N. Sekurova, T. Brautaset, H. Sletta, S. E. F. Borgos, O. M. Jakobsen, T. E. Ellingsen, A. R. Strom, S. Valla and S. B. Zotchev, *J. Bacteriol.*, 2004, **186**, 1345-1354.
63. S. Lautru, M. Gondry, R. Genet and J.-L. Pernodet, *Chem. Biol.*, 2002, **9**, 1355-1364.
64. (1) L. Retzlaff and J. Distler, *Mol. Microbiol.*, 1995, **18**, 151-162 (2) S. Ikeno, D. Aoki, K. Sato, M. Hamada, M. Hori and K. S. Tsuchiya, *J. Antibiot.*, 2002, **55**, 1053-1062.
65. (1) K. Inoue, H. Habe, H. Yamane, H. Nojiri, *Appl. Environ. Microbiol.*, 2006, **72**, 3321-3329 (2) H. Bhukya, R. Bhujbalrao, A. Bitra and R. Anand, *Nucleic Acids Res.*, 2014, **42**, 10122-10133.
66. L. Hong, Z. Zhao, C. E. Melançon 3rd, H. Zhang and H. W. Liu, *J. Am. Chem. Soc.*, 2008, **130**, 4954-4967.
67. S. L. Otten, M. A. Gallo, K. Madduri, X. Liu and C. R. Hutchinson, *J. Bacteriol.*, 1997, **179**, 4446-4450.
68. V. H. DuVernay Jr., J. A. Pachter and S. T. Crooke, *Biochemistry*, 1979, **18**, 4024-4030.
69. K. Räty, T. Kunnari, J. Hakala, P. Mäntsälä and K. Ylihonko, *Mol. Gen. Genet.*, 2000, **264**, 164-172.
70. (1) R. W. Gantt, P. Peltier-Pain and J. S. Thorson, *Nat. Prod. Rep.*, 2011, **28**, 1811-1853. (2) B. R. Griffith, J. M. Langenhan and J. S. Thorson, *Curr. Opin. Biotechnol.*, 2005, **16**, 622-630.
71. (1) J. Oh, S. G. Lee, B. G. Kim, J. K. Sohng, K. Liou and H. C. Lee, *Biotechnol. Bioeng.*, 2003, **84**, 452-458. (2) H. Takahashi, Y. N. Liu and H. W. Liu, *J. Am. Chem. Soc.*, 2006, **128**, 1432-1433. (3) Y. S. Chung, D. H. Kim, W. M. Seo, H. C. Lee, K. Liou, T. J. Oh and J. K. Sohng, *Carbohydr. Res.*, 2007, **342**, 1412-1418. (4) M. K. Kharel, H. Lian and J. Rohr, *Org. Biomol. Chem.*, 2011, **9**, 1799-1808. (5) M. D. Shepherd, T. Liu, C, Méndez, J. A. Salas and J. Rohr, *Appl. Environ. Microbiol.*, 2011, **77**, 435-441. (6) G. Wang, M. K. Kharel, P. Pahari and J. Rohr, *Chembiochem*, 2011, **12**, 2568-2571. (7) G. Wang, P. Pahari, M. K. Kharel, J. Chen, H. Zhu, S. G. Van Lanen and J. Rohr, *Angew. Chem. Int. Ed.*, 2012, **51**, 10638-10642. (8) Q. Wang, P. Ding, A. V. Perepelov, Y. Xu, Y. Wang, Y. A. Knirel, L. Wang and L. Feng, *Mol. Microbiol.*, 2008, **70**, 1358-1367. (9) L. Elling, C. Rupprath, N. Günther, U. Römer, S. Verseck, P. Weingarten, G, Dräger, A. Kirschning and W. Piepersberg, *Chembiochem*, 2005, **6**, 1423-1430. (10) see supplementary reference S22 for more of related literatures.
72. T. Kieser, M. J. Bibb, M. J. Buttner, K. F. Chater and D. A. Hopwood, *Practical Streptomyces Genetics*, John Innes Foundation, Norwich, UK, 2000.
73. J. Sambrook and D. Russell, *Molecular Cloning: A Laboratory Manual*, 3rd Edition, Cold Springs Harbor Laboratory Press, New York, 2001.
74. M. M. Bradford, *Anal. Biochem.*, 1976, **72**, 248-254.