A labile point in mutant amphotericin polyketide synthases

Naseem Khan, Bernard Rawlings, Patrick Caffrey

Section: Microbial and Enzyme Technology

A labile point in mutant amphotericin polyketide synthases

Naseem Khan[1], Bernard Rawlings[2], Patrick Caffrey[1†].

[1]School of Biomolecular and Biomedical Science and Centre for Synthesis and Chemical Biology, University College Dublin, Belfield, Dublin 4, Ireland.

[2]Department of Chemistry, University of Leicester, Leicester LE1 7RH, United Kingdom.

Keywords: amphotericin B, polyketide synthase, ketoreductase, polyenyl-pyrone

[†] To whom correspondence should be addressed. E-mail: Patrick.caffrey@ucd.ie, FAX: +353 1 716 1183, Tel.: +353 1 716 1396

Abstract

Streptomyces nodosus produces the antifungal polyene amphotericin B. Numerous modifications of the amphotericin polyketide synthase have yielded new analogues. However, previous inactivation of the ketoreductase in module 10 resulted in biosynthesis of truncated polyketides. Here we show that modules downstream of this domain remain intact. Therefore, loss of ketoreductase-10 activity is sufficient to cause early chain termination. This modification creates a labile point in cycle 11 of the polyketide biosynthetic pathway. Non-extendable intermediates are released to accumulate as polyenyl-pyrones.

Introduction

Amphotericin B **1** (Fig. 1) is used in clinical medicine for treatment of serious systemic fungal infections and visceral leishmaniases. It is obtained from fermentation cultures of *Streptomyces nodosus*. Both chemical modification and engineered biosynthesis have generated analogues with reduced toxicity (Caffrey et al., 2008). The aglycone 19-deoxy-19-oxo-amphoteronolide B **2** (Fig. 1) has potential as a starting material for chemical synthesis of derivatives with altered sugar residues. Small quantities of this compound have been obtained by chemical deglycosylation of amphotericin B (Kennedy et al., 1988). Biosynthesis of 19-deoxy-19-oxo-analogues was expected to result from inactivation of ketoreductase (KR) domain 10 of the amphotericin polyketide synthase (PKS) (Caffrey et al., 2001). However, we recently found that KR10 mutants of *S. nodosus* produce truncated polyketide intermediates instead of full-length macrolactones (Fig. 1) (Murphy et al., 2010). The KR10-1 strain synthesised the pentaenyl-pyrone **3** as a single product whereas the KR10-2 strain gave both the heptaenyl-pyrone **4** and the tetraenyl-pyrone **5** (Fig. 1). These new compounds were produced in good yields of up to 20 mg purified material per litre of culture. Preliminary PCR analysis of strain KR10-1 suggested that modules 5 and 6 had been precisely deleted from the AmphC protein that synthesises most of the polyene unit. This unexpected recombination event was not related to targeted replacement of the KR10 coding sequence, which is located at an approximate distance of 80 kb along the

chromosome. The loss of these two modules accounts for the shorter chain length of pentaenyl-pyrone **3**. Module 5 contains a partly functional enoylreductase domain (Caffrey et al., 2001) that causes formation of both heptaene **4** and tetraene **5** by strain KR10-2.

From the structures of the new pyrone compounds, it can be deduced that no ketoreduction occurs during cycle 10, as expected for a KR10 mutant (Fig. 2). The unprocessed -ketoacyl chains migrate from ACP10 to KS11 and incorporation of methylmalonyl CoA-derived extender units gives polyenyl-2-methyl-3, 5-diketoacyl-ACP11 thioester intermediates (Fig. 2). These products undergo no further enzymatic processing, and are released to accumulate as pyrones. This finding was unexpected because inactivation of KR12 or KR16 did not block subsequent steps in polyketide chain assembly (Power et al., 2008). The co-incidental loss of two AmphC modules from strain KR10-1 raised the possibility that early termination might result from further unplanned deletions downstream of ACP11. This led us to characterise these strains more extensively. Here we show that the remaining PKS is intact. KR10 inactivation alone is sufficient to stall polyketide biosynthesis immediately after chain elongation by module 11.

Materials and methods

Growth of *Streptomyces nodosus* strains and spectrophotometric analysis of polyenyl-pyrones was carried out as described previously (Murphy et al., 2010). Genomic DNA was purified using Qiagen Genomic-tip 500/G ion-exchange columns. Oligonucleotide primers were obtained from MWG Biotech. Genome sequencing was carried out by BaseClear, Leiden, The Netherlands, using an Illumina Genome Analyser GAIIx. Direct sequencing of amplified DNA by the dideoxy method was carried out by Source BioScience, Ireland Sequencing Service, Dublin. PCR products were purified using a Qiagen QIAquick kit. The Artemis programme (downloaded from http://www.sanger.ac.uk/resources/software/artemis/) was used for final assembly of the re-sequenced *amphI* region of strain KR10-1.

Results and discussion

Analysis of AmphI domains in KR10-1 and KR10-2 strains

In KR10-deficient mutants, module 11 incorporates a branched extender unit into the polyketide chain. Since this elongation stage proceeds normally, the lack of ketoreduction in cycle 11 was unexpected. To investigate whether KR11 was intact, the DNA specifying this domain was amplified from strains KR10-1 and KR10-2 and both PCR products were re-sequenced. The PCR primers KR11F and KR11R (Supplementary Table 1) were designed to amplify the regions coding for the *N*- and *C*-terminal interdomain linkers as well as the enzymatic domain. This sequence data showed that no mutations were present in the KR11 domains of either strain. To investigate whether deletions had occurred downstream of module 11, further PCR analysis was carried out on coding sequences for domains within downstream modules (ACP11, KS12, AT12, ACP12, KS13, AT13, KR13, ACP13, ACP14). (The primers used are listed in Supplementary Table 1.) Analysis of amplified DNA by agarose gel electrophoresis indicated that all coding sequences were present in KR10-1 and KR10-2 strains (Supplementary Fig. 1). More thorough analysis of strain KR10-1 was then carried out as described below.

Re-sequencing modules 11-14 in strain KR10-1

The AmphI protein contains modules 9 to 14 of the amphotericin PKS. The structures of the polyenyl-pyrones show that the KR10 PKSs function as expected up to module 10. As next-generation sequencing technology has become affordable, we used this approach to investigate more fully whether the downstream AmphI modules 11 to 14 remain intact. Our previous work has shown that targeted deletion of modules 5 and 6 leads to production of pentaene macrolactone analogues of amphotericin B in high yield (Carmody et al., 2004; Murphy et al., 2010). This shows that deletion of these two modules is not relevant to early chain termination in KR10-1. However, this strain was

selected for re-sequencing so that the spontaneous *amphC* deletion could be investigated as well as the modules downstream of KR10.

Genomic DNA from strain KR10-1 was sent to BaseClear for sequencing. *De novo* assembly provided 3977 contig sequences with an average size of 1818 nucleotides. (In a separate study, we are currently sequencing and annotating the wild-type *S. nodosus* genome.) A second assembly of the strain KR10-1 sequence data was carried out using our unpublished *S. nodosus* genome sequence as a reference. This gave most of the strain KR10-1 *amphI* gene region coding for modules 11 to 14, although several small gaps remained. These regions were amplified by PCR (Supplementary Fig. 2) and directly sequenced by the dideoxy method. The KR10-1 coding sequence for AmphI modules 11 to 14 was re-sequenced completely. The sequence of this region of the protein (amino acid residues 3128 to 9510) was identical to that first published for the amphotericin PKS (Caffrey et al., 2001) with the following exception: $R^{6381}\rightarrow A$ in ACP12. This difference is also present in the *S. nodosus* genome sequence and does not affect amphotericin production; it is thought to represent a neutral substitution or an error in the original sequence. Analysis of the KR10-1 genome indicated that the downstream AmphJ and AmphK PKS proteins were intact, although small gaps in the corresponding gene sequences were not closed. The re-sequencing also revealed the precise location of the deletion in the *amphC* gene (see below).

These results show that early termination occurs even though the programmed pathway specifies -ketoreduction by KR11 and further extension by downstream modules. There are no clear precedents where inactivation of a KR domain in one module prevents ketoreduction by the next. KR16 and KR12 mutants formed 3,5 diketoacyl and 3,5,7-triketoacyl intermediates, respectively, that are extended to form full length macrolactones (Power et al., 2008). The behaviour of the KR10 amphotericin PKSs is therefore surprising.

Proposed mechanisms for early termination in KR10-deficient PKSs

Module 11 normally generates a (2*S*, 3*S*)-2-methyl-3-hydroxyacyl-ACP product. The landmark study of Weissman et al. (1997) first revealed the stereochemical aspects of such a cycle. A (2*S*)-methylmalonyl extender unit is loaded and incorporated, with inversion of stereochemistry, to give a (2*R*)-2-methyl-3-ketoacyl-ACP11. This undergoes epimerisation, via an enol intermediate, to give a (2*S*)-2-methyl-3-ketoacyl intermediate that is reduced to a (2*S*, 3*S*)-2-methyl-3-hydroxyacyl product. The epimerase activity has not been identified for any PKS module but is thought to reside within the KR domain (Valenzano et al., 2009). In a reconstituted module lacking an epimerase, the C-2 proton of a (2*R*)-2-methyl-3-ketoacyl-thioester is protected by the ACP domain and is not exchangeable (Castonguay et al., 2007). Our results suggest that in the unnatural tricarbonyl-ACP11 intermediate, the C-4 protons are labile and this allows tautomerisation of the C-5 ketone. The early enolisation stage of C-2 epimerisation then generates a dienol intermediate that is abnormally stable because it forms an extended system of conjugated double bonds. This is not a substrate for KR11 and is not accepted by KS12. This stalled polyketide intermediate might be off-loaded from the enzyme by the discrete AmphE thioesterase (TE) (Fig. 3, Route 1). These editing TEs mainly recognise decarboxylated extender units attached to ACP domains but can also remove non-extendable polyketide intermediates (Heathcote et al., 2001). The linear acyl chain would form a terminal pyrone ring spontaneously. Alternatively, *cis*-enolisation of the tricarbonyl chain may favour concerted pyrone formation and thioester cleavage, thus bringing about efficient chain release without intervention of a TE (Fig. 3, Route 2). A *cis* geometry has been proposed for enolate intermediates that eventually form (2*S* 3*S*)-2-methyl-3-hydroxyacyl–ACP thioesters such as the product of AmphI module 11 (Starcevik et al., 2007).

With future work, it may still be possible to engineer the biosynthesis of 19-deoxy-19-oxo-amphoteronolides. Mutations in module 11 that abolish the C-2 epimerase activity might allow reduction by KR11 and formation of full-length macrolactones with reversed alkyl group stereochemistry at C-16.

Iterative type I fatty acid and polyketide synthases are known to produce pyrones under abnormal conditions, such as the absence of modification enzymes or NADPH or S-adenosylmethionine co-factors. For example, incomplete lovastatin PKS systems

synthesise a series of aberrant polyketides including polyenyl-pyrones of various chain lengths *in vitro* (Ma et al., 2009). Inactivation of amphotericin PKS domains KR6, KR7, KR8 and KR9 would also result in formation of shorter polyenyl-3,5-diketoacyl intermediates. It seems unlikely that these would form non-extendable dienol intermediates because non-epimerising modules are thought to prevent loss of the C-2 proton from the -ketoacyl-ACP products. However, we plan to investigate this further by determining whether inactivation of KR7, KR8 or KR9 causes stalling of chain growth and accumulation of polyenyl-pyrones.

Mapping of the *amphC* deletion in strain KR10-1

This work also provides information on the evolution of natural product biosynthetic genes. In addition to finding that replacing a KR10 active site residue causes a drastic change in final product structure, we observed contraction of a heptaene synthase to a pentaene synthase. The genome re-sequencing suggested that the internal deletion in the KR10-1 *amphC* gene involves a 47-bp stretch that is directly repeated within the coding sequences for AT5 and AT7 (nucleotides 92165-92211 and 103628-103674, respectively, in sequence accession number AF357202). To confirm this, PCR primers M5F and M7R were designed to match AT5 and AT7 coding sequences flanking the putative deleted region. Direct sequencing of the product confirmed that the recombination had indeed generated an AT5/AT7 hybrid coding sequence, with the fusion point located within the repeated 47 bp sequence (Supplementary Fig 3). In a previous study (Carmody et al., 2004) genetic engineering was used to construct a pentaene macrolactone synthase in which the C-terminal region of KS5 was fused to the interdomain linker preceding AT7. (The coding sequences were joined through KpnI sites representing nucleotides 91698-91703 and nucleotides 103146-103151 in sequence accession AF357202). Extensive screening was required to obtain this gene replacement mutant, indicating that this region does not form a recombination hot-spot. Since the amphotericin PKS genes are stable in wild-type *S. nodosus*, it is unclear why the *amphC* gene should undergo a spontaneous deletion in the KR10-1 strain. Recombination

between directly repeated homologous sequences may have resulted by chance through incorrect repair of a double-strand break. Since **3**, **4** and **5** are of no obvious benefit to the *S. nodosus* producer strains, continued propagation would presumably result in silencing or inactivation of these clusters by mutation. However, polyenylpyrone production remained stable when the KR10 strains were passaged through seven subcultures.

Biotechnological potential of polyenyl-pyrone producer strains

Numerous bioactive polyketides terminate with 2-pyrone groups that presumably protect linear chains from degradation by -oxidation (McGlacken and Fairlamb, 2005). Many of these compounds are biologically active and the polyenylpyrone group includes cytoprotectants, antioxidants, antibacterials and anticancer agents (Clark and Murphy, 2009; Coleman and Waleza, 2006). The starter unit for these polyketides is critical for activity. The KR10-deficient amphotericin PKSs should also be useful for precursor-directed biosynthesis of valuable analogues.

Conclusions

Further work will be required to elucidate fully the precise mechanism for early chain termination. However, the present results prove that inactivation of KR10 introduces a labile point at a later stage in the amphotericin PKS pathway. The production of tetraenyl-, pentaenyl- and heptaenyl-pyrones suggests that the corresponding domain may be critical in polyene PKSs in general. This work has uncovered a restriction on the theoretical number of polyene macrolactone analogues that can be generated by engineered biosynthesis.

Acknowledgments

NK received a PhD studentship from the Irish Higher Education Authority Programme for Research in Third Level Institutions. PC received the financial support of Science Foundation Ireland under grant number 09/RFP/GEN2132.

References

Caffrey, P., Lynch, S., Flood, E., Finnan, S., and Oliynyk, M. (2001) Amphotericin biosynthesis in *Streptomyces nodosus*: deductions from analysis of polyketide synthase and late genes. Chem. Biol. 8: 713 – 723.

Caffrey, P., Aparicio, J. F., Malpartida, F., and Zotchev, S. B. (2008) Biosynthetic engineering of polyene macrolides for generation of improved antifungal and antiparasitic agents. Curr. Topics Med. Chem. 8: 639-653.

Carmody, M., Byrne, B., Murphy, B., Breen, C., Lynch, S., Flood, E., Finnan, S., and Caffrey, P. (2004) Analysis and manipulation of amphotericin biosynthetic genes by means of modified phage KC515 transduction techniques. Gene 343: 107-115.

Castonguay, R., He, W., Chen, A. Y., Khosla, C., Cane D. E. (2007). Stereospecificity of ketoreductase domains of the 6-deoxyerythronolide B synthase. *J. Am. Chem. Soc*. 129: 13758-13769.

Clark, B. R., Murphy, C. (2009) Biosynthesis of pyrrolylpolyenes in *Auxarthron umbrinum Org. Biomol. Chem*. 7: 111-116.

Coleman, R. S., Waleza, M. C. (2006). Total synthesis of gymnoconjugatins A and B.*J. Org. Chem*. 71: 9841-9844.

Heathcote, M. L., Staunton, J., Leadlay, P. F. (2001) Role of type II thioesterases: evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. Chem. Biol. 8: 207-20.

Kennedy, R. M., Abiko, A., Masamune, S. (1988). Degradation of amphotericin B: Cleavage of the glycosidic linkage with aluminum amalgam [Al(Hg)] of dichlorodicyanoquinone (DDQ). Tetrahedron Letters 29: 447-450.

Ma, S. M., Li, J. W.-H., Choi, J. W., Zhou, H., Lee, K. K. M., Moorthie, V. A., Xie, X., Kealey, J. T., Da Silva, N. A., Vederas, J. C., Tang, Y. (2009) Complete reconstitution of a highly reducing iterative polyketide synthase. Science 326: 589-592.

McGlacken, G. P., and Fairlamb, I. J. S. (2005) 2-Pyrone natural products and mimetics: isolation, characterisation and biological activity. Nat. Prod. Rep. 22: 369-385.

Murphy, B., Andersen, K., Borrissow, C., Caffrey, P., Griffith, G., Hearn, J., Ibrahim, O., Khan, N., Lamburn, N., Lee, M., Pugh, K., Rawlings, B. (2010) Isolation and characterisation of amphotericin B analogues and truncated polyketide intermediates produced by genetic engineering of *Streptomyces nodosus*. Org. Biomol. Chem. 8: 3758 - 3770.

Power, P., Dunne, T., Murphy, B., Nic Lochlainn, L., Rai, D., Borissow, C., Rawlings, B., and Caffrey, P. (2008) Engineered synthesis of 7-oxo- and 15-deoxy-15-oxo-amphotericins: insights into structure-activity relationships in polyene antibiotics Chem. Biol. 15: 78-86.

Starcevic, A., Jaspars, M., Cullum, J., Hanrueli, D., Long, P. F. (2007) Predicting the nature and timing of epimerisation on a modular polyketide synthase ChemBioChem 8: 28-31.

Valenzano, C. R., Lawson, R. J., Chen, A. Y., Khosla, C., Cane, D. E. (2009) The biochemical basis for sterochemical control in polyketide biosynthesis. J. Am. Chem. Soc. 131: 18501-18511.

Weissmann, K. J., Timoney, M., Bycroft, M., Grice, P., Hanefeld, U., Staunton, J., Leadlay, P. F. (1997) The molecular basis of Celmer's rules: the stereochemistry of the condensation step in chain extension on the erythromycin polyketide synthase. Biochemistry 36: 13849-13855.

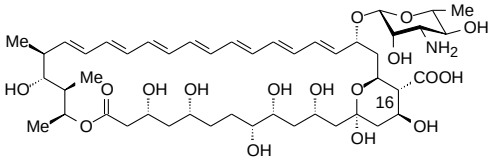
Figure legends

Fig. 1. Structures of amphotericin B **1**, 19-dideoxy-19-oxoamphoteronolide B **2**, pentaenylpyrone **3**, heptaenyl-pyrone **4** and tetraenyl pyrone **5**.

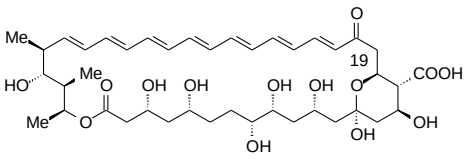
Fig. 2. Formation of pyrones from polyenyl-tricarbonyl-ACP11 intermediates in KR10-deficient amphotericin polyketide synthases. The inactivated domain is shaded.

Fig. 3. Proposed mechanisms for early chain termination in KR10 strains. Biosynthetic cycles 10 and 11 are shown. The -ketone formed by module 10 is unprocessed by the inactivated KR10 domain (shaded). Incorporation of a propionyl extender by module 11 gives a (2*R*)-2-methyl-3-ketoacyl-ACP11. This would be expected to undergo C-2 epimerisation and ketoreduction to give a (2*S*, 3*S*)-2-methyl-3-hydroxyacyl-chain that migrates from ACP11 to KS12. The first stage of C-2 epimerisation gives an enol intermediate that is unusually stable because the unnatural C5 ketone tautomerises to extend fully the system of conjugated double bonds. The resulting dienol intermediate is not reduced by KR11 or accepted by KS12. Two mechanisms for chain release are suggested. Route 1: Release by an editing TE gives a free acid that spontaneously isomerises and forms a pyrone. Route 2: Removal of the C-2 proton gives an intermediate containing a *cis* enol. This adopts a conformation that leads to concerted thioester cleavage and pyrone formation without intervention of a TE.

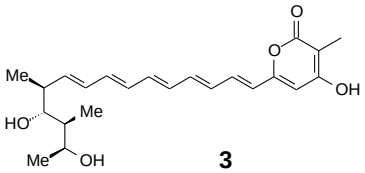
Figure 1



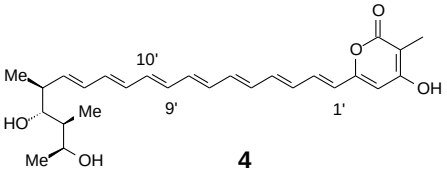
1



2



3



4

5 = 9'-10' - dihydro 4

Figure 2

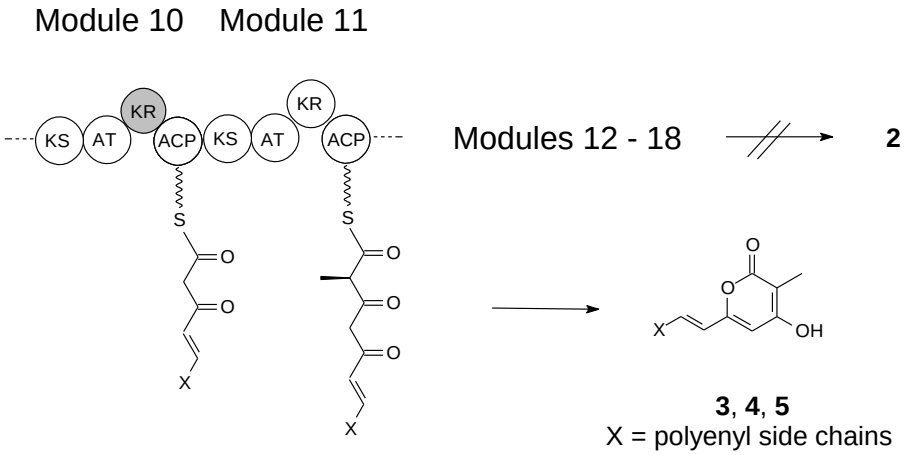
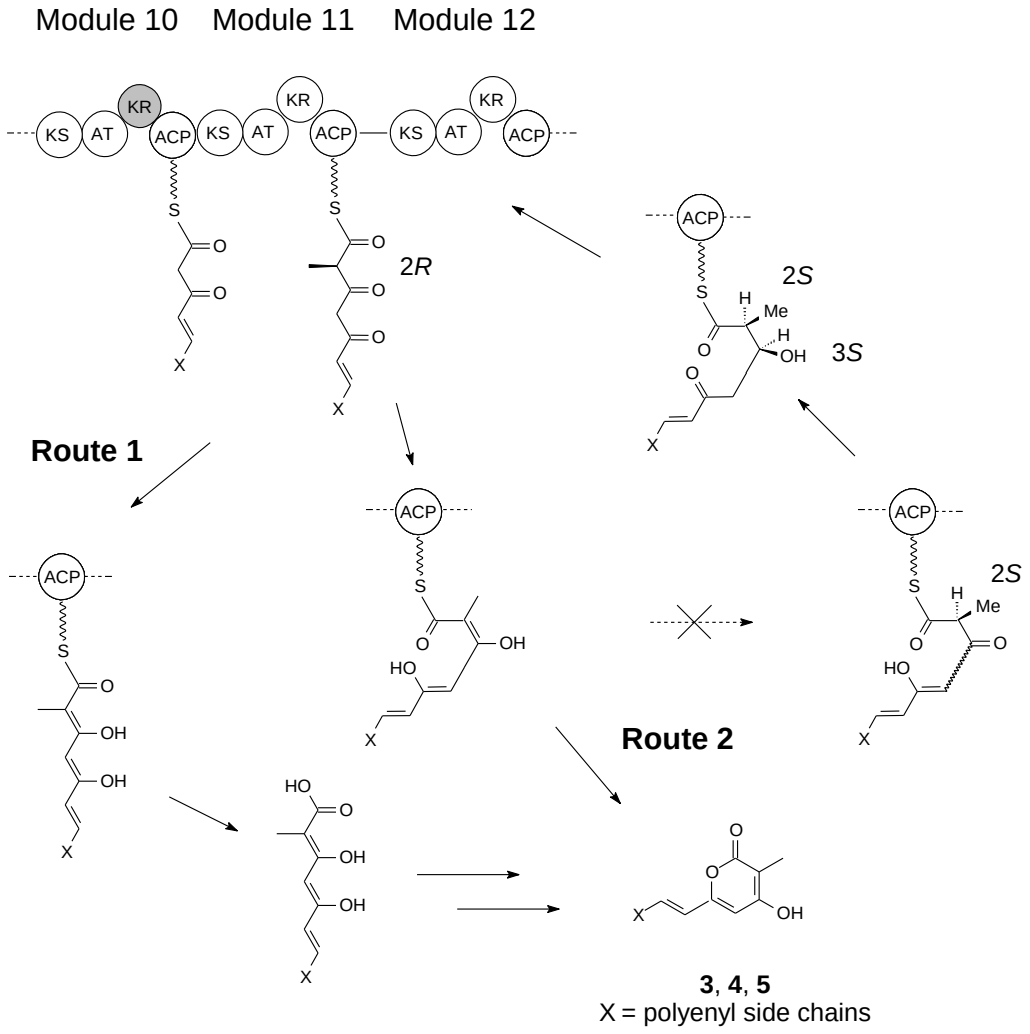

Module 10
Module 11
KS
AT
ACP
KR
KS
AT
ACP
KR
S
O
O
X
S
O
O
O
X
Modules 12 - 18
2
O
O
X
OH
3, 4, 5
X = polyenyl side chains

Figure 3

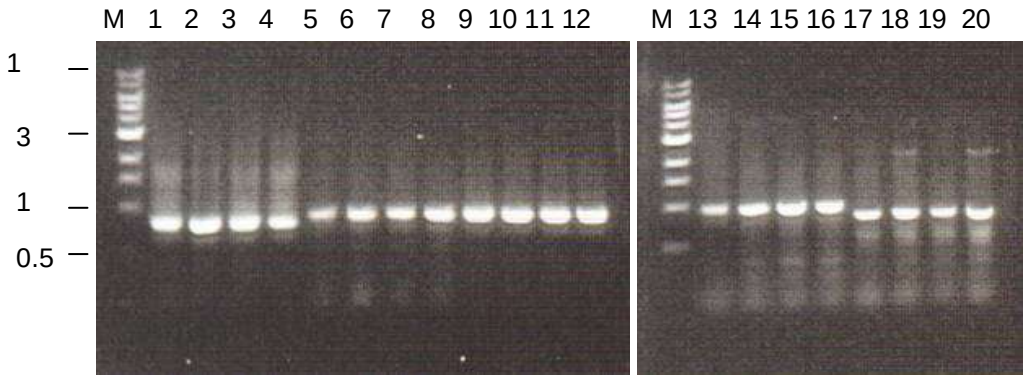

Module 10
Module 11
Module 12
KS
AT
KR
ACP
2R
2S
3S
Route 1
Route 2
OH
HO
Me
H
X
3, 4, 5
X = polyenyl side chains

Supplementary Table 1

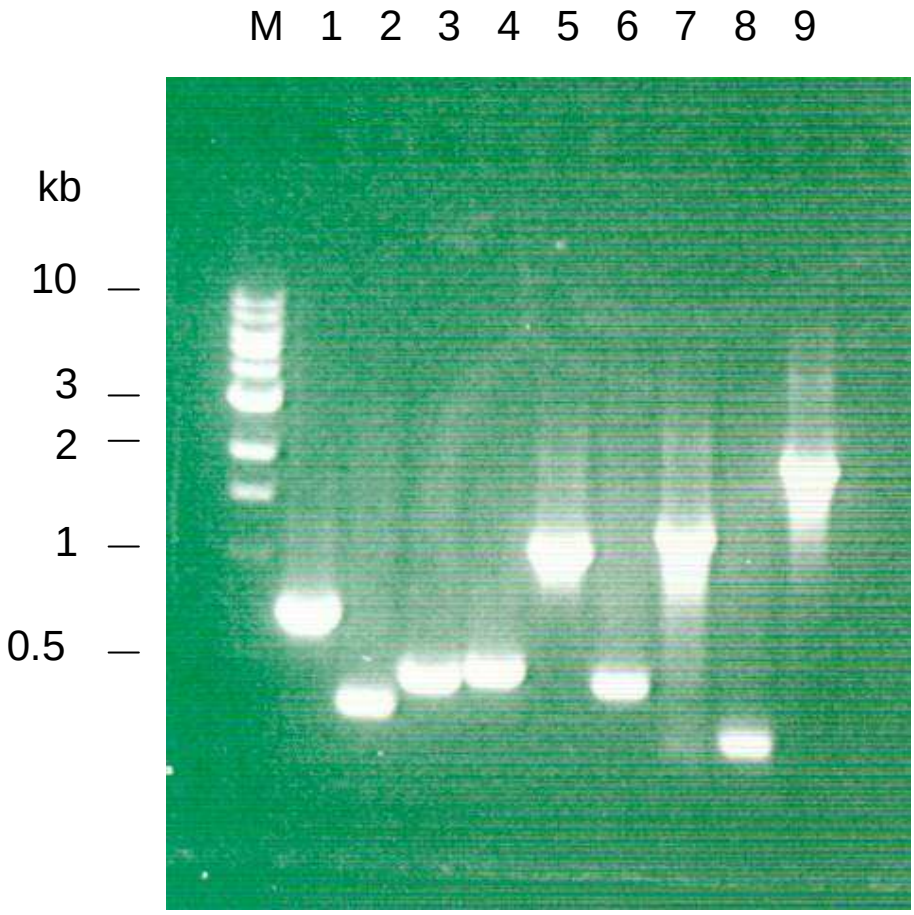
Table 1. Oligonucleotides used in this work. AT12rsF3 and AT12rsR3 were used as primers for complete re-sequencing the PCR product shown in Supplementary Fig. 2 lane 9. AT14rsF2 and AT14rsR2 were used as primers for complete re-sequencing the PCR product shown in Supplementary Fig. 2 lane 7.

KR11F	5′ GATCAAGCTTATGCCTTCCAGCACGAGAAC 3′
KR11R	5′ GATCGAATTCTCCAGCAGCTTGCGGTGCTGTTC 3′
KS12F	5′ GACTTCGACCCCGGCTTCTTC 3′
KS12R	5′ GTGGCCGATGTTGGACTTCAC 3′
AT12F	5′ GCGAACTCTACTAACGCCAC 3′
AT12R	5′ CAGTCGACGTCGAAGCCATG 3′
KS13F	5′ GTGTGTTCGTCGGCACCAACG 3′
KS13R	5′ CTCCACGATGACATGCGCATTG 3'
AT13F	5′ GAACTCTACGAGCGGTACCCG 3′
AT13R	5′ GAAGGCGTACGTCGGCAGTTC 3′
KR13F	5′ GTACGTCCAGTACGGGGACTTC 3′
KR13R	5′ GAAGAGCACGAACGCGTCCAG 3′
AT11rsF	5′ CCGCCGAGGACGAGCAGCTCGC 3′
AT11rsR	5′ ACGGGCAGCGCCACCGACATCAT 3′
AT12rsF	5′ CACCAATGTCCACACCGTCCTG 3′
AT12rsR	5′ CTGTCCGGTGAACAGGAAGGC 3′
AT12rsF2	5′ GCCTTCCTGTTCACCGGACAG 3′
AT12rsR2	5′ TCGCAGGCGGCGATCGTGACA 3′
ACP12rsF	5′ CGTTCGTCCTGTTCTCCTCGACCG 3′
ACP12rsR	5′ GTACGCAGCAGATCGAGGACATG 3′
AT13rsF	5′ GAACTGCCGACGTACGCCTTC 3′
AT13rsR	5′ GTACTGGACGTACTCGGGCTTTC 3′
KR13rsF	5′ TGGACCGTCGTGGTCTCGCTGCTC 3′
KR13rsR	5′ TGCTCCAGGAGCCACCGGCCGTAG 3′
KS14rsF	5′ AGCGGCACCACCGGCGTCTTCGT 3′
KS14rsR	5′ GCTGAACTCGACGAAGCCGAC 3′
AT14rsF	5′ GCGTCCGGGTCCTCCTCGAACT 3′
AT14rsR	5′ ATCGCGTCCAGCACCGCGGCCAG 3′
KR14rsF	5′ TTCGTCCTGTTCTCGTCCGCCTCCT 3′
KR14rsR	5′ TACCAGGTCCACGACGGTGTCCA 3′
M5F	5′ CTGATCCTCTCCGGCAAGTCAC 3′
M7R	5′ GTCCATCAGCGGCGAGTGGAAC 3′
AT12rsF3	5′ AGCAGCGACAGCACACCCTGGAA 3′
AT12rsR3	5′ CCATGGCTTCGACGTCGACTG 3′
AT14rsF2	5′ TCCAGCACCGCGGCCAGCTG 3′
AT14rsR2	5′ CTGGAGTACCCGAACCGCTG 3′

Supplementary data



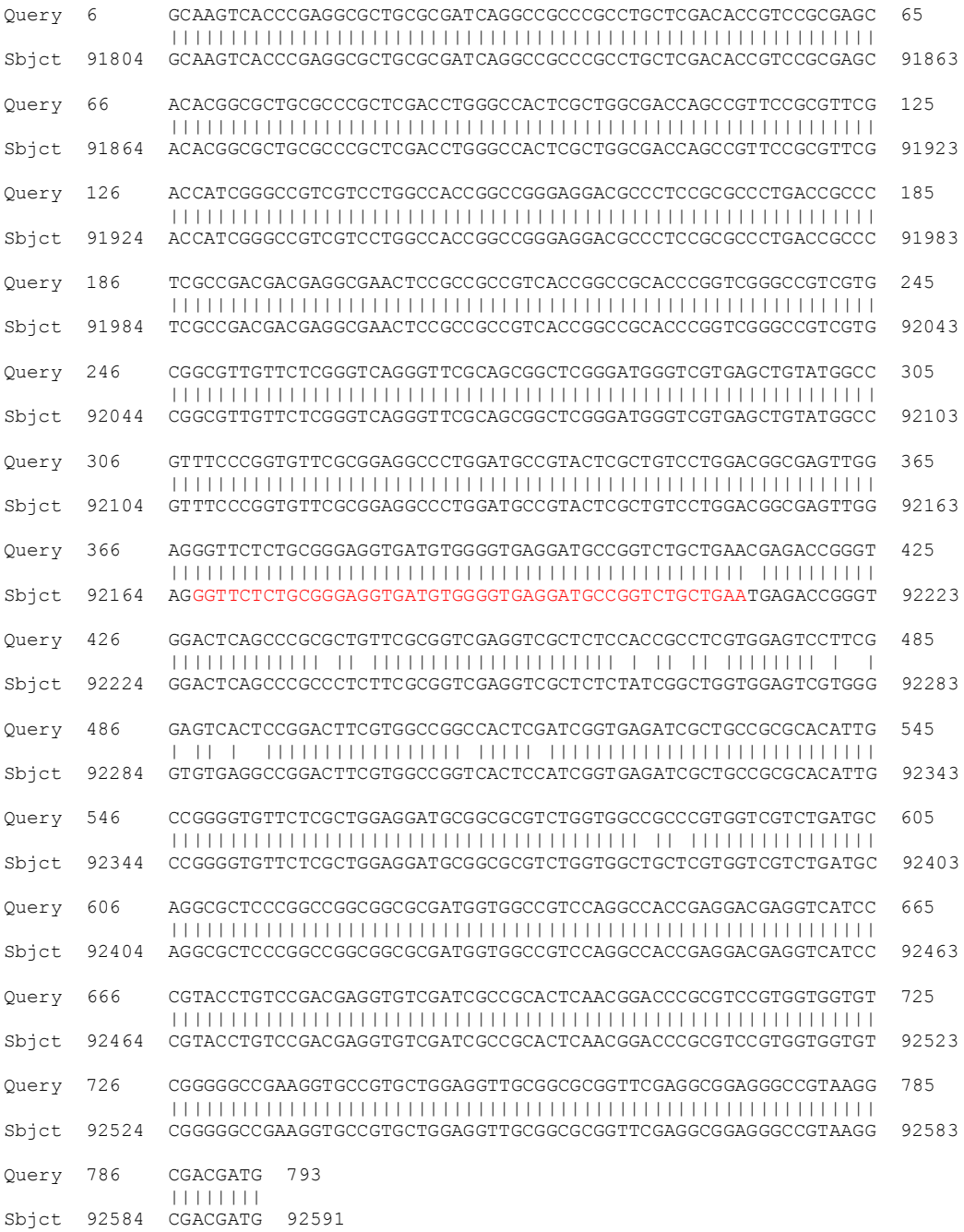
Supplementary Fig. 1. PCR analysis of coding sequences for downstream modules in KR10 mutants. Reactions were carried out with primers designed from sequences coding for domains KS12 (lanes 1 to 4), AT12 (lanes 5 to 8), KS13 (lanes 9 to 12), AT13 (lanes 13 to 16) and KR13 (lanes 17 to 20). The DNA templates were from *S. nodosus* (lanes 1, 5, 9, 13 and 17), strain KR10-1 (lanes 2, 6, 10, 14, 18), strain KR10-2 (lanes 3, 7, 11, 15, 19) and a separate isolate of strain KR10-2 (lanes 4, 8, 12, 16, 20). All strains retained the downstream coding sequences.



Supplementary Fig. 2 Amplification of *amphI* regions unrepresented in strain KR10-1 genomic sequence. PCR products were generated using primers listed in Table S2: AT11rsF + AT11rsR (lane 1); AT12rsF + AT12rsR (lane 2); ACP12rsF + ACP12rsR (lane 3); AT13rsF + AT13rsR (lane 4); KR13rsF + KR13rsR (lane 5); KS14rsF + KS14rsR (lane 6); AT14rsF + AT14rsR (lane 7); KR14rsF + KR14rsR (lane 8); AT12rsF2 + AT12rsF2 (lane 9). The PCR products were purified and directly sequenced by the dideoxy method.

Supplementary Fig. 3.

A.



B.

```
Query  118     CGC-GTTCGACCATCGGGCCGTCGTCCTGGCCACCGGCCG-GGAGGA-CGC-CCTCCGCG  173
               ||| |||||| || || || |||||||| ||||||| | || |||  | |||  | |||||
Sbjct  103378  CGCTGTTCGAGCACCGTGCGGTCGTCCTCGCCACC-GACGCGGA-CACCGCGGC-CCGCG  103434

Query  174     CCCTGACCGCCCTCGCCGACGACGAG-GCGAACTCCGCCGCCGTCA-CCGG-CCGCACCC  230
               ||||| | |||||||| |  | ||||  || || |||| ||||| | | || |||  |||
Sbjct  103435  CCCTGGCGGCCCTCGCGGTGGGCGAGCCCG-ACCCCGCGGCCGT-ATCGGGCCCG-GCCC  103491

Query  231     GGTCGGGCCG-TCGTGCGGCGTTGTTCTCGGGTCAGGGTTCGCAGCGGCTCGGGATGGGT  289
               |  | || || ||| |||||| ||||||||||||||||||||||||||||||| ||||||
Sbjct  103492  GCACCGGTCGCTCG-GCGGCGCTGTTCTCGGGTCAGGGTTCGCAGCGGCTCGGTATGGGT  103550

Query  290     CGTGAGCTGTATGGCCGTTTCCCGGTGTTCGCGGAGGCCCTGGATGCCGTACTCGCTGTC  349
               |||||||||||| ||||| ||||| ||||| || ||||||||||||||||||||||||||
Sbjct  103551  CGTGAGCTGTACGGCCGGTTCCCCGTGTTTGCCGAGGCCCTGGATGCCGTACTCGCTGTC  103610

Query  350     CTGGACGGCGAGTTGGAGGGTTCTCTGCGGGAGGTGATGTGGGGTGAGGATGCCGGTCTG  409
               |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Sbjct  103611  CTGGACGGCGAGTTGGACGGTTCTCTGCGGGAGGTGATGTGGGGTGAGGATGCCGGTCTG  103670

Query  410     CTGAACGAGACCGGGTGGACTCAGCCCGCGCTGTTCGCGGTCGAGGTCGCTCTCCACCGC  469
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  103671  CTGAACGAGACCGGGTGGACTCAGCCCGCGCTGTTCGCGGTCGAGGTCGCTCTCCACCGC  103730

Query  470     CTCGTGGAGTCCTTCGGAGTCACTCCGGACTTCGTGGCCGGCCACTCGATCGGTGAGATC  529
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  103731  CTCGTGGAGTCCTTCGGAGTCACTCCGGACTTCGTGGCCGGCCACTCGATCGGTGAGATC  103790

Query  530     GCTGCCGCGCACATTGCCGGGGTGTTCTCGCTGGAGGATGCGGCGCGTCTGGTGGCCGCC  589
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  103791  GCTGCCGCGCACATTGCCGGGGTGTTCTCGCTGGAGGATGCGGCGCGTCTGGTGGCCGCC  103850

Query  590     CGTGGTCGTCTGATGCAGGCGCTCCCGGCCGGCGGCGCGATGGTGGCCGTCCAGGCCACC  649
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  103851  CGTGGTCGTCTGATGCAGGCGCTCCCGGCCGGCGGCGCGATGGTGGCCGTCCAGGCCACC  103910

Query  650     GAGGACGAGGTCATCCCGTACCTGTCCGACGAGGTGTCGATCGCCGCACTCAACGGACCC  709
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  103911  GAGGACGAGGTCATCCCGTACCTGTCCGACGAGGTGTCGATCGCCGCACTCAACGGACCC  103970

Query  710     GCGTCCGTGGTGGTGTCGGGGGCCGAAGGTGCCGTGCTGGAGGTTGCGGCGCGGTTCGAG  769
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  103971  GCGTCCGTGGTGGTGTCGGGGGCCGAAGGTGCCGTGCTGGAGGTTGCGGCGCGGTTCGAG  104030

Query  770     GCGGAGGGCCGTAAGGCGACGATG  793
               |||||||||||||||||||||||
Sbjct  104031  GCGGAGGGCCGTAAGGCGACGATG  104054
```

Supplementary Fig. 3. BLAST search with the sequence amplified using M5 and M7 primers. The 5′ end of the hybrid sequence perfectly matches the AT5 coding sequence (nucleotide 91804 - 92212) [panel A] whereas the 3′ end matches the AT7 coding sequence (nucleotides 103629 - 104054) [panel B]. The 47 bp direct repeat is in red type (92166-92212 in AT5 coding sequence and 103629-103675 in AT7 coding sequence).