GOSSYPOL BIOSYNTHESIS IV. THE AMINO ACID SIMILARITY OF TERPENOID CYCLASES

C. Magill, G. Bianchini And C. R. Benedict
Texas A&M University
College Station, TX

Abstract

At least two hundred different families of sesquiterpenoid compounds have been identified in plants. Although related isoprenoid compounds such as giberellins and carotenes are essential to survival, most cotton sesquiterpenes are classified as secondary metabolites. Many have antibiotic activity and are synthesiszed following infection with pathogens, so are considered to be phytoalexins. Representatives of a number of sesquiterpene families are present in cotton, including several with different cyclic structures. Different cyclic sesquiterpenes originate from folding of FPP or its isomers into specific configurations. However, since a single synthase may lead to the formation of several products through variable patterns of condensation, the number of unique cyclases present is undetermined. Further difficulties in identifying and cloning specific synthase genes arise because there are regions of amino acid sequence homology between mono- sesqui- and di- terpene cyclases, and from the fact that some or all of the cyclases are coded by members of small gene families present in the genome.

Introduction

The ability to produce anti-microbial chemicals is a primary mode of defense for plants. Whether the compounds are made and stored as in the glands of cotton, or induced by biotic stresses (phytoalexins) many of the most important defense molecules found in cotton are terpenes. Terpene synthesis begins with mevalonate, the product of condensation of three acetyl-CoA units by HMGR [1]. Phosphorylation and decarboxylation leads to the basic isoprene diphosphate (IPP) unit and its isomer di-methylallyl diphosphate (Fig 1). Head to tail condensation of successive 5C IPP units produces geranyl diphosphate (GPP) from which monoterpenes are derived, farnesyl diphosphate (FPP), the origin of all sesquiterpenoids, and geranyl geranyl diphosphate (GGPP) from which diterpenes are derived. Thirty carbon terpenes and forty carbon terpenes are derived from condensing two FPPs or GGPPs respectively, but in this case, the diphosphates are removed in the condensations.

Cyclic derivatives can made at each step in the pathway beginning with GPP. Cyclic monoterpenes in cotton include a- and b-pinenes, a- and g-tripinenes, limonene, philandrene and cymene. Some of these compounds are insect attractants so are important for pollination. Gossypol is the best known sesquiterpene compound found in cotton. It, along with its biosynthetic precursors are especially important as antifungal, antiherbivore, and anti-insect agents [2]. Related compounds including the cadalenes have been found to have anti-bacterial activity [3], but functional roles, if any, for other volatile cyclic sesquiterpenes remain to be determined. The vital plant hormone GA is a diterpine made following cyclization of GGPP to a kaurine through the action of ent-kaurine synthase. Cyclization of phytoene leads to the synthesis of b-carotene and the xanthophylls and almost certainly to ABA as well. Cyclization of the 30C squalene backbone leads to the various sterols that are present.

Although no varieties of cotton are immune to Verticillium wilt, there is a large difference in the response to infection. Inoculation of the *G. hirsutum* cultivar Rowden by injecting spores of a virulent strain into the stem rapidly leads to defoliation and death, while Sea Brook Sea Island, a *G. barbadense* cultivar is able to recover. The resistant cultivar synthesizes phytoalexins more rapidly and to a higher level than Rowden. Notably, desoxyyhemigossypol (dHG) can accumulate to toxic levels before the fungus has a chance to spread. Since dHG is a cyclic sesquiterpene, it is clear that the terpene pathway is involved. A cloned segment of a cotton HMGR has been used to show that HMGR-mRNA synthesis is induced much more rapidly in SBSI than in Rowden, implicating differential gene expression as the basis for the varietal differences in Verticillium wilt resistance [4]. The same induction system was useful recently in showing that d-cadinine is the first cyclic compound in forming dHG and thus all the related gossypols and heliocides [5]. Interestingly, d-cadinine was also shown to be the cyclic precursor to the antibacterial cadalenes by another group at the same time [6].

Methods

Comparisons of amino acid and base sequences of clones available in the PIR, SwisProt, GenBank and EMBL databases were made based on the basis of BLAST [7] searches using electronic network services or FASTA alignments made with the Wisconsin Package (Version 8) GCG® programs available on a local VAX computer. Internet services were also used to search for motifs (Ogiwara, unpublished) and blocks [8] of conserved sequences shared with other proteins. Optimal alignments were made using MacVector® (Kodak) or by combining PILEUP and PRETTY programs in the GCG package. The occurrence of matching short sequences in other sequences in the databases was detected using the FIND PATTERNS GCG program.

Results and Discussion

The ability to clone a segment of a cotton HMGR was made possible by comparing the base sequences of previously

Reprinted from the *Proceedings of the Beltwide Cotton Conference*
Volume 2:1184-1186 (1996)
National Cotton Council, Memphis TN

cloned plant HMGRs deposited in GenBank. Even though only three plant HMGRs were available, highly conserved regions were readily identified and used to make slightly degenerate PCR primers. Although the same technique proved very effective for cloning segments of 7 other putative defense response genes including chitinase and PAL, only two sesquiterpene cyclase sequences were available at the time; one from tobacco and another from a fungus. Even though the two cyclases make almost identical products, no obvious alignment of conserved amino acid or DNA sequences could be made [9].

Cloning of other plant terpene cyclases has led both to potential solutions and unique problems in using the PCR approach to clone the equivalent genes from cotton. Following the cloning of the elicitor induced 5-epi-aristolochene (a sesquiterpene) synthase from tobacco by Faccini and Chappell in 1993 [9], Croteau's group was able to clone and sequence a monoterpene cyclase, 4S-limonene synthase, from mint [10], and Mau and West [11] cloned a diterpene cyclase (casbene synthase). As was pointed out by the latter authors and reinforced by others since, there is a good deal of sequence homology between these cyclases, each of which represents a different class. The similarity of plant terpene cyclases can be detected and demonstrated in several ways. Using any one of the sequences to search for similarity to sequences in nucleic acid or protein databases with programs such as BLAST or FASTA will return a list of high scores; results of a recent BLAST search show high scores not only for the previously described sequences, three vetispiradiene synthases from *Hyoscyamus,* a medicinal plant [12] and three d-cadinene synthases from cotton submitted by Chen from Heinstein's laboratory, but also identify homologous regions in both maize and *Arabidopsis* kaurine synthases, the diterpene cyclase used to initiate GA synthesis. Significantly, no homology is seen to any of several cloned prenyl transferases, including FPP synthase *(FPS)*, GGPP synthase (*GGPS*), and phytoene synthase (*PSY*) which involve similar reactions for eliminating diphosphates from isoprenoid precursors, or to the cyclase that converts lycopene to b-carotene.

Alignment of related proteins can often help to identify the most highly conserved sequences; such sites are likely to be important to catalysis or enzyme conformation. Based on previously published sequences, Back and Chappell [12] have shown highly conserved spacing for histidine and cysteine residues in each of the plant terpene cyclase exons, as well as an aspartic acid-rich sequence (DDXXD) which is thought to bind a metal cofactor (Mg^{++} or Mn^{++}) critical to substrate (polyprenyl) binding. This sequence is part of two 8-9 amino acid motifs present in eukaryotic FPP synthases. Degenerate primers based on these conserved amino acids have been made and used to clone the FPP synthase from lupins [13]. However, DDXXD by itself is found in nearly 7,000 of the 82,361 proteins currently in the PIR database. It is found in many different types of proteins, but not in the two plant kaurine synthases that have been sequenced, so it cannot be considered an identifying pattern for cyclases. Searches for any other motifs that have been defined from other sets of homologous proteins were negative, as were all searches for conserved blocks. Likewise, a profile made from four aligned terpene cyclases, excluding any signal sequences, did not detect any other proteins with similar profiles. Although the diterpene cyclases that are required to make GA seem to function without the DDXXD motif, they do retain relatively high homology with the other plant cyclases, but over a smaller region. This can be demonstrated by a graphic representation of aligned sequences prepared using the MacVector program, or by actual identification of amino acids conserved across the mon-, sesqui-, and diterpene cyclases. When compared to a similar alignment made without including the ent-kaurine cyclases, the relatively small conserved domain is apparent. When attempting to identify stretches of conserved sequences to use for generating PCR primers to amplify similar sequences, it is often simplest to make pairwise alignments of coding sequences with MacVector® using a Pustell matrix. Regions which contain a predetermined fraction (usually 65% which allows for third position codon degeneracy) of identical base sequences within a window of a set size will be identified. Regions that overlap for each of the pairs can be magnified to identify individual bases and alternatives to include in primer synthesis.

Examination of a GCG "PRETTY" multiple sequence alignment of the actual amino acid sequences for the first two cloned plant sesquiterpene synthases along side the mono- and di terpenene synthases from mint and casbene (Fig. 2), reveals two blocks of five conserved amino acids (sitalics). FINDPATTERNS searches for the LPZYM sequence found several proteins, including gramicidin S, pyoverdine and surfactin synthetases in bacteria. The FKESL sequence was present more often, appearing in several RNA polymerases, an ATP synthase, a heat shock protein, a cytochrome C oxidase and a protease. The diversity of these proteins suggests that chance rather than a conserved motif or binding site accounts for the shared sequence. Interestingly, P(Y,F) ARDR, part of one of the sequences (underlined) used by Back and Chappell to clone the *Hyascyamus* synthase was found only in a polyketide synthase from *Bacillus.* Other relatively conserved sequences where degenerate primers could be made are also apparent. One of these identifies a highly conserved peptide (double underlined) in monoterpene cyclases from pines and sages that has been determined to be part of the active site by virtue of binding to an irreversible inhibitor [14]

If the two most highly conserved plant terpene synthase subsequences are used to make PCR primers it should be anticipated that cyclases in each class may be amplified, and the overall degree of sequence homology could make it very difficult to determine which gene has been cloned from a new species. Even expression tests may not reveal the class of cloned genes, since both mono- and diterpene cyclases are likely to be induced by the same stresses. If conserved sequences that are unique to mono, sesqui or diterpenes can be identified from independently verified clones, it should be possible to narrow the possibilities for amplifying only a desired gene. Obviously, when sequences are to be identified on the basis of homology to previously cloned sequences, it is absolutely critical that the original sequence is correctly identified.

There is another potential problem in attempting to clone specific cotton cyclases. Croteau has shown that while limonene synthase produces predominately limonene from GPP, small amounts of a- and b-pinene and myrcene are also made in vivo and by the recombinant enzyme. Pinene synthases from pine are even less specific and produce several products, some of which are quite different in their ring structures. Chappell, via Cane, has illustrated a basis for the different products; the ability of the carbocation produced by the loss of the diphosphate to form a C-C bond at either end of any double bond in the parent molecule [1]. Prior isomerization of E,E-FPP to nerolidylPP means that even more possibilities exist, and thus it is not even certain at this time how many sesquiterpene cyclases may be present in tetraploid cottons. Examination of the cyclic sesquiterpenes made in cotton suggests that at least two unique synthases must be present to account for at least two substrates. In addition, most of the cloned cyclases have been shown to be members of gene families by virtue of Southern hybridization to genomic DNA digests. If, as in the case of HMGR, different members are differentially expressed, it is likely that it will be possible to identify sequences near the beginning of the message that will serve to identify each of them when active through Northern analysis or ribonuclease protection assays. However, as Back and Chappell have suggested, if slight differences influence the relative amount of different products, it will probably require directed mutagenesis to identify individual genes [12].

References

1. Chappell, J., 1995. The biochemistry and molecular biology of isoprenoid metabolism. Plant Physiology **107**: 1-6.

2. Bell, A.A., Stipanovic, R.D., Mace, M.E. and Kohel, R.J. *Genetic Manipulation of Terpenoid Phytoalexins in Gossypium: Effects on Disease Resistance,* B.E. Ellis, Editor. 1994, Plenum Press: Genetic Engineering of Plant Secondary Metabolites. pp. 231-244.

3. Essenberg, M., B.P. Grover Jr., and E.C. Cover, 1990. Accumulation of antibacterial sesquiterpenoids in bacterially inoculated *Gossypium* leaves and cotyledons. Phytochemistry, **29**: 3107-3113.

4. Joost, O., Bianchini, G., Bell, A.A., Benedict, C.R., Magill, C.W. 1995. Differential induction of 3-hydroxy-3-methylglutaryl CoA reductase in two cotton species following inoculation with Verticillium. MPMI, **8**: 880-885.

5. Benedict, C.R., *et al.*, 1995. The enzymatic formation of d-cadinene from farnesyl diphosphate in extracts of cotton. Phytochemistry, **39**: 327-331.

6. Davis, G.D. and M. Essenberg, 1995. (+)-delta-Cadinene is a product of sesquiterpene cyclase activity in cotton. Phytochemistry, **39**: 553-567.

7. Altshul, S.F., *et al.*, 1990. Basic local alignment search tool. J. Mol. Biol., **215**: 403-410.

8. Henikoff, S. and J.G. Henikoff, 1994. Protein family classification based on searching a database of blocks. Genomics, **19**: 97-107.

9. Facchini, P.J. and J. Chappell, 1993. Gene family for an elicitor-induced sesquiterpene cyclase in tobacco. Proc. Natl. Acad. Sci. USA, **89**: 11088-11092.

10. Colby, S.M., *et al.*, 1993. 4S-Limonene synthase from the oil glands of spearmint (*Mentha spicata*): cDNA isolation, characterization, and bacterial expression of the catalytically active monoterpene cyclase. J. Biol. Chem. **268**: 23016-23024.

11. Mau, C.J.D. and C.A. West, 1994. Cloning of casbene synthase cDNA: evidence for conserved structural features among terpenoid cyclases in plants. Proc. Natl. Acad. Sci. USA, **91**: 8497-8501.

12. Back, K. and J. Chappell, 1995. Cloning and bacterial expression of a sesquiterpene cyclase from *Hyoscyamus muticus* and its molecular comparison to related terpene cyclases. J. Biol. Chem., **270**: 7375-7381.

13. Attucci, S., *et al.*, 1995. Farnesyl pyrophosphate synthase from white lupin: Molecular cloning, expression, and purification of the expressed protein. Archives of Biochemistry and Biophysics, **321**: 493-500.

14. Mcgeady, P. and R. Croteau, 1995. Isolation and characterization of an active-site peptide from a monoterpene cyclase labeled with a mechanism-based inhibitor. Archives of Biochemistry & Biophysics, **327**: 149-155.

Fig. 1. ORIGINS OF CYCLIC TERPENOIDS IN COTTON

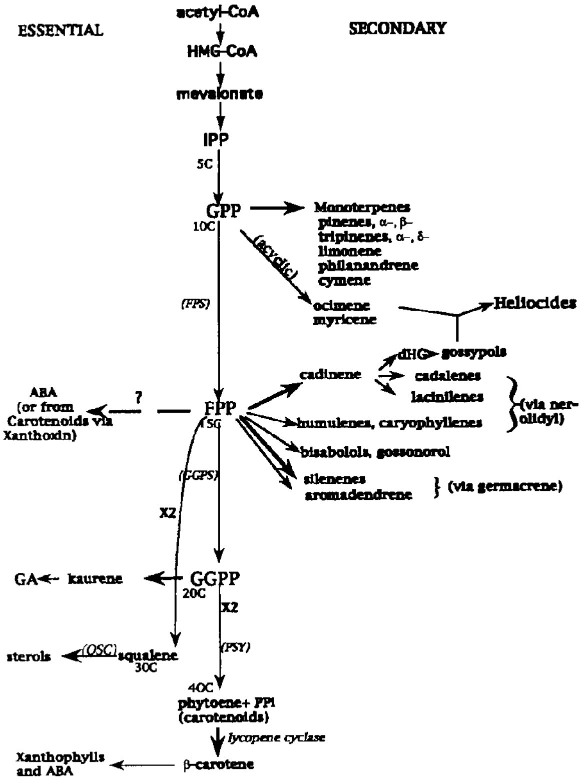

Fig. 2: Identical Amino Acids in Optimally Aligned Clones of Plant Terpene Cyclases

```
Tobacco (Sesqui)     71   MASAAV.ANY  .EEEIVRPVA  DFSPSLWGDQ  F.LSFSIDNQ  VAEKYIYAQE
Hyoscyamus  (Sesqui)      MAPAIVMSNY  EEEEIVRPVA  DFSPSLWGDH  F.HSFSVDNQ  VAEK..YAQE
Casbene (Di)              KPTQACLSST  THQE.VRPLA  YFPPTVWGNR  F.ASLTFN..  PSEFESYDER
Mint (Mono)               RSRLRVYCSS  SQLTTERRSG  NYNPSRWDVN  FIQSLLSDYK  EDKHVIRASE
               Identity   ----------  ------R---  ---P--W---  F--S------  ----------

Tobacco  (Sesqui)  121    IEALKEQTRS  ML.LATGRKL  ADTLNLIDII  ERLGISYHFE  KEIDEILDQI
Hyoscyamus  (Sesqui)      IETLKEQTST  MLSAACGTTL  TEKLNLIDII  ERLGIAYHFE  KQIEDMLDHI
Casbene (Di)              VIVLKKEVKD  ILISSTSDSV  .ETVILIDLL  CRLGVSYHFE  NDIEELLSKI
Mint (Mono)               LVTLVK....  .MELEKETDQ  IRQLELIDDL  QRMGLSDHFQ  NEFKEILSSI
               Identity   ---L------  ----------  -----LID--  -R-G---HF-  ------L--I

Tobacco  (Sesqui)  171    YNQNS.....  ...NCNDLCT  SALQFRLLRQ  HGFNISPEIF  SKFQDENGKF
Hyoscyamus  (Sesqui)      YRADPYFEA.  ..HEYNDLNT  SSVQFRLLRQ  HGYNVSPNIF  SRFQDANGKF
Casbene (Di)              FNSQPDLVD.  ..KKECDLYT  AAIVFRVFRQ  HGFKMSSDVF  SKFKDSDGKF
Mint (Mono)               YLDHHYYKNP  FPKEERDLYS  TSLAFRLLRE  HGFQVAQEVF  DSFKNEEGEF
               Identity   ----------  ------DL--  ----FR--R-  HG-------F  --F----G-F

Tobacco  (Sesqui)  221    KESLASDVLG  LLNLYEASHV  RTHADDILED  ALAFSTIHLE  SAAPH..LKS
Hyoscyamus  (Sesqui)      KESLRSDIRG  LLNLYEASHV  RTHKEDILEE  ALVFSVGHLE  SAAPH..LKS
Casbene (Di)              KESLRGDAKG  MLSLFEASHL  SVHGEDILEE  AFAFTKDYLQ  SSAVE..LFP
Mint (Mono)               KESLSDDTRG  LLQLYEASFL  LTEGETTLES  AREFATKFLE  EKVNEGGVDG
               Identity   KESL--D--G  -L-L-EAS--  -------LE-  A--F----L-  ----------

Tobacco  (Sesqui)  271    PLREQVTHAL  EQCLHKGVPR  VETRFFISSI  YDKEQSKNNV  LLRFAKLDFN
Hyoscyamus  (Sesqui)      PLSKQVTHAL  EQSLHKSIPR  VEIRYFI.SI  YEEEEFKNDL  LLRFAKLDYN
Casbene (Di)              NLKRHITNAL  EQPFHSGVPR  LEARKFIDLY  EADIECRNET  LLEFAKLDYN
Mint (Mono)               DLLTRIAYSL  DIPLHWRIKR  PNAPVWIEWY  RKRPD.MNPV  VLELAILDLN
               Identity   -L-------L  ----H----R  ------I---  -------N--  -L--A-LD-N

Tobacco  (Sesqui)  321    LLQMLHKQEL  AQVSRWWKDL  DFVTTLPYAR  DRVVECYFWA  LGVYFEPQYS
Hyoscyamus  (Sesqui)      LLQMLHKHEL  SEVSRWWKDL  DFVTTLPYAR  DRAVECYFWT  MGVYAEPQYS
Casbene (Di)              RVQLLHQQEL  CQFSKWWKDL  NLASDIPYAR  DRMAEIFFWA  VAMYFEPDYA
Mint (Mono)               IVQAQFQEEL  KESFRWWRNT  GFVEKLPFAR  DRLVECYFWN  TGIIEPRQHA
               Identity   --Q-----EL  -----WW---  ------P-AR  DR--E--FW-  ----------

Tobacco  (Sesqui)  371    QARVMLVKTI  SMISIVDDTF  DAYGTVKELE  AYTDAIQRWD  INEIDRLPDY
Hyoscyamus  (Sesqui)      QARVMLAKTI  AMISIVDDTF  DAYGIVKELE  VYTDAIQRWD  ISQIDRLPEY
Casbene (Di)              NTRMIIAKVV  LLISLIDDTI  DAYATMEETH  ILAEAVARWD  MSCLEKLPDY
Mint (Mono)               SARIMMGKVN  ALITVIDDIY  DVYGTLEELE  QFTDLIRRWD  INSIDQLPDY
               Identity   --R----K--  --I---DD--  D-Y----E--  -------RWD  ------LPEY

Tobacco  (Sesqui)  421    MKISYKAILD  LYKDYEKELS  SAGRSHIVCH  AIERMKEVVR  NYNVESTWFI
Hyoscyamus  (Sesqui)      MKISYKALLD  LYDDYEKELS  KDGRSDVVHY  AKERMKEIVG  NYFIEGKWFI
Casbene (Di)              MKVIYKLLLN  TFSEFEKELT  AEGKSYSVKY  GREAFQELVR  GYYLEAVWRD
Mint (Mono)               MQLCFLALNN  FVDDTSYDVM  KEKGVNVIPY  LRQSWVDLAD  KYMVEARWFY
               Identity   M---------  ----------  ----------  ----------  -Y--E--W--

Tobacco  (Sesqui)  471    EGYMPPVSEY  LSNALATTTY  YYLATTSYLG  MKSATE.QDF  EWLSKNPKIL
Hyoscyamus  (Sesqui)      EGYMPSVSEY  LSNALATSTY  YLLTTTSYLG  MKSATK.EHF  EWLATNPKIL
Casbene (Di)              EGKIPSFDDY  LYNGSMTTGL  PLVSTASFMG  VQEITGLNEF  QWLETNPKLS
Mint (Mono)               GGHKPSLEEY  LENSWQSISG  PCMLTHIFFR  VTDSFTKETV  DSLYKYHDLV
                  Identity  -G--P----Y L-N-------  ----T-----  ----------  --L-------

Tobacco  (Sesqui)  521    EASVIICRVI  DDTATYEVEK  SRGQIATGIE  CCMRDYGIST  KEAMAKFQNM
Hyoscyamus  (Sesqui)      EANATLCRVV  DDIATYEVEK  GRGQIATGIE  CYMRDYGVST  EVAMEKFQEM
Casbene (Di)              YASGAFIRLV  NDLTSHVTEQ  QRGHVASCID  CYMNQHGVSK  DEAVKILQKM
Mint (Mono)               RWSSFVLRLA  DDLGTSVEEV  SRGDVPKSLQ  CYMSDYNASE  AEARKHVKWL
               Identity   -------R--  -D------E-  -RG-------  C-M-----S-  --A-------

Tobacco  (Sesqui)  571    AETAWKDINE  GLL.RPTPVS  TEFLTPILNL  ARIVEVTYIH  NLDGYTHPEK
Hyoscyamus  (Sesqui)      ADIAWKDVNE  EIL.RPTPVS  SEILTRILNL  ARIIDVTYKH  NQDGYTHPEK
Casbene (Di)              ATDCWKEINE  ECM.RQSQVS  VGHLMRIVNL  ARLTDVSYKY  G.DGYTDSQQ
Mint (Mono)               IAEVWKEMNA  ERVSKDSPFG  KDFIGCAVDL  GRMAQLMY.H  NGDGHGTQHP
               Identity   ----WK--N-  ----------  ---------L  -R-----Y--  --DG------
```