Genetically engineered synthesis of natural products

A. Ian Scott

Center for Biological Nuclear Magnetic Resonance, Department of Chemistry
Texas A&M University, College Station, Texas 77843-3255, USA
Fax: (409) 845-5992

Abstract

The feasibility of multi-enzyme, one-flask total synthesis of natural products is demonstrated by gene transfer and overexpression of the corresponding biosynthetic enzymes in *E. coli*.

Since the 1950's organic chemists have been exploring biosynthetic pathways to natural products, at first with radioactive tracers fed to whole plants and growing cells, then more "biochemically" with cell-free extracts and finally with purified enzymes. By the beginning of the 1970's it became clear that techniques of NMR spectroscopy could be used to follow ^{13}C enriched substrates through the maze of multienzyme conversions (frequently 15 or 20 steps) leading to a labeling pattern in the target from which biochemical processing could be deduced - albeit speculatively. Now another powerful technique is beginning to change the way in which many of us interested in Nature's methods of organic synthesis are approaching the problem. The new dimensions afforded by genetics and molecular biology have revolutionized our concepts of what is feasible in natural product biosynthesis. In Figure 1 the pathway

1. Pathway

$$A \xrightarrow[E_1]{\text{Gene 1}} B \xrightarrow[E_2]{\text{Gene 2}} C \xrightarrow[E_3]{\text{Gene 3}} D \cdots\cdots> \text{TARGET}$$

Gene 1, Gene 2, Gene 3, Genes
Enzymes

2. Intermediates (B, C, D...): Isolation, incubation

3. Enzymes (E_1, E_2...): Purification, mechanism, genetic engineering
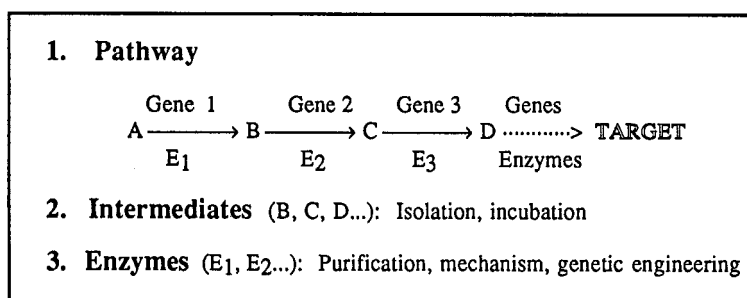
Figure 1   Natural Product Biosynthesis

from substrate A to the target molecule via intermediates B, C, D etc. now includes the gene for each biosynthetic enzyme. The appropriate gene products, i.e. the enzyme for each step, can be purified, sequenced and the cDNA amplified in *E. coli* (or other vector) to produce amounts of enzyme (up to 1 gram per liter of cells) hitherto unimaginable. Thus by employing the techniques shown in Figure 2 using either a cDNA library or more conveniently a series of open reading frames (ORF's) obtained by genetic complementation, mapping and sequencing, the gene products (enzymes) can be expressed "conventionally" or by p.c.r. (Figure 3) and tested for their biosynthetic capabilities, using high field

I   Isolate the gene encoding the enzyme.

   A.   Construct a cDNA library.

   B.   Make a probe (antibody or oligonucleotide) to identify the gene of interest from the cDNA library.

   C.   Screen the cDNA library with the labeled probe to locate a plasmid or bacteriophage bearing the gene for the biosynthetic enzyme.

   D.   Amplify and purify the plasmid or bacteriophage DNA bearing the gene and determine the nucleotide sequence of the gene.

II Express and purify the enzyme.

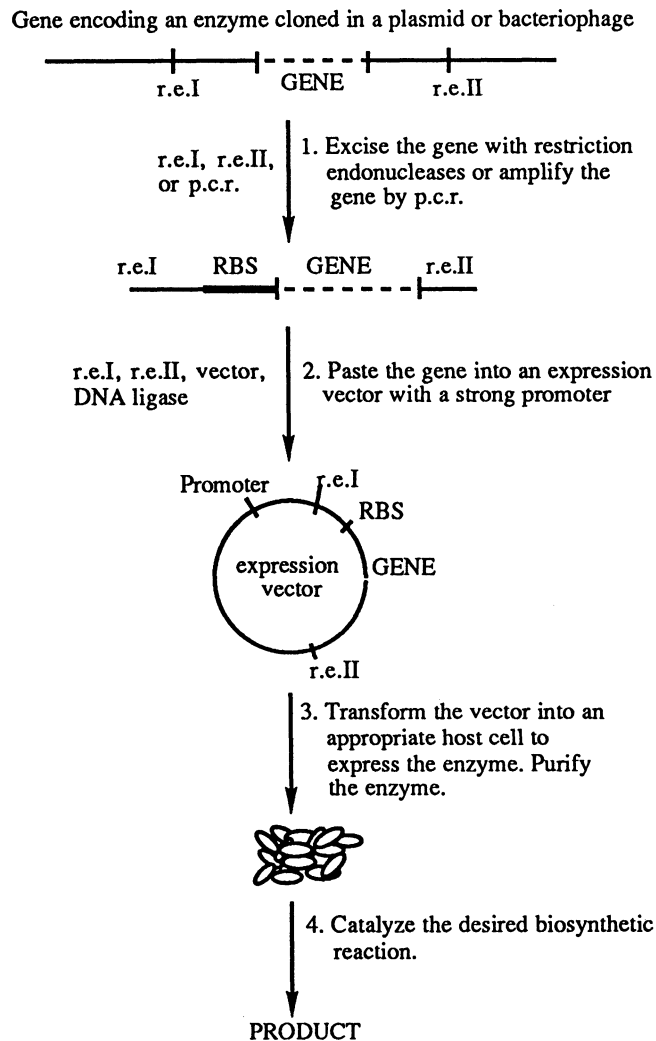Figure 2   Genetic engineering of biosynthetic enzymes for natural products

Gene encoding an enzyme cloned in a plasmid or bacteriophage

r.e.I   GENE   r.e.II

r.e.I, r.e.II,   1. Excise the gene with restriction
or p.c.r.   endonucleases or amplify the
                      gene by p.c.r.

r.e.I   RBS   GENE   r.e.II

r.e.I, r.e.II, vector,   2. Paste the gene into an expression
DNA ligase   vector with a strong promoter

Promoter   r.e.I
                 RBS

expression   GENE
vector

r.e.II

3. Transform the vector into an
   appropriate host cell to
   express the enzyme. Purify
   the enzyme.

4. Catalyze the desired biosynthetic
   reaction.

PRODUCT

Figure 3   Expression of Biosynthetic Enzymes

spectroscopy to assay each biosynthetic step **as it occurs** in the NMR tube. Each enzyme can be studied separately in mechanistic detail, but when **recombined**, the full synthetic machinery can be activated, leading to multi-enzyme **total synthesis** of the target. Site directed mutagenesis can then be used to modify substrate specificity thus producing new variations on the biosynthetic theme, as portrayed in Figure 4. This lecture illustrates the power of such an approach applied to several completely different natural product targets.
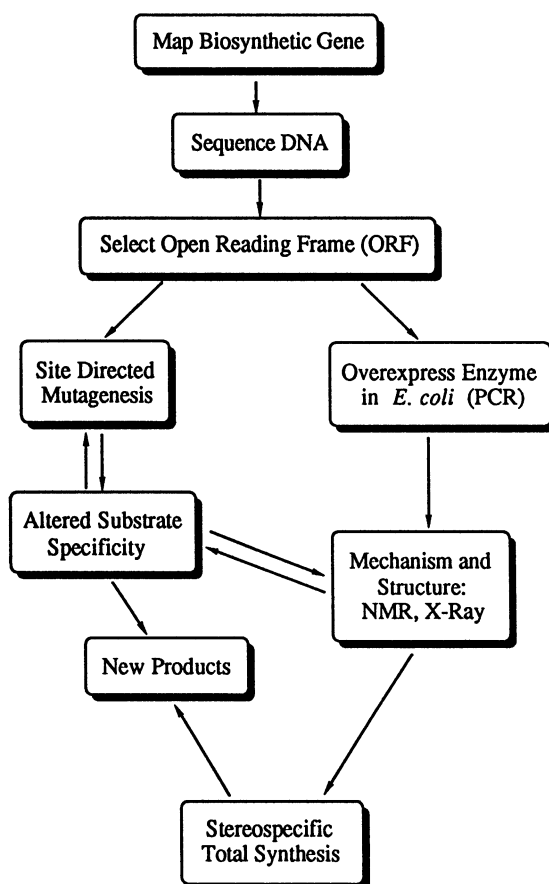
Figure 4   Genetic Engineering of Biosynthetic Enzymes

MECHANISM OF ISOPENICILLIN N SYNTHASE (IPNS)

It is now firmly established that, following the enzymatic assembly of the Arnstein tripeptide, L-aminoadipyl-L-cysteinyl-D-valine [ACV; **1**] in a manner which retains the absolute configuration of the β–methyl groups ($\bullet$ and $\blacktriangle$) of the D-valine residue, cyclization of ACV (**1**) to isopenicillin N (**2**) leads directly to the penicillin nucleus (ref. 1) and thence to cephalosporin C (**6**) via (**3**), (**4**), (**5**) (Fig. 5).

Ever since the discovery by Demain (ref 2) and Abraham (ref. 3) that cell-free preparations from *C. acremonium* catalyze the synthesis of the penicillin nucleus from ACV (**1**), intensive investigations have been undertaken to discover the mechanism(s) of the two bond-forming processes *viz* the removal of 2 hydrogens to fuse the β–lactam ring and a further 2H to generate the second, thiazolidine ring, (**1**) $\rightarrow$ (**8**) $\rightarrow$ (**2**) (Fig. 6). While recent studies with the purified enzyme IPNS (M.W. ~38,000) indicate that the cofactors Fe^{++}, O_2 and ascorbic acid are required in the cell-free system together with dithiothreitol (refs. 1, 4) all attempts to characterize the putative lactam thiol (**8**) directly have so far failed (refs. 5, 6) although kinetic isotopic effect studies strongly suggest that the reaction takes place via initial monocyclic β-lactam formation (ref. 7).
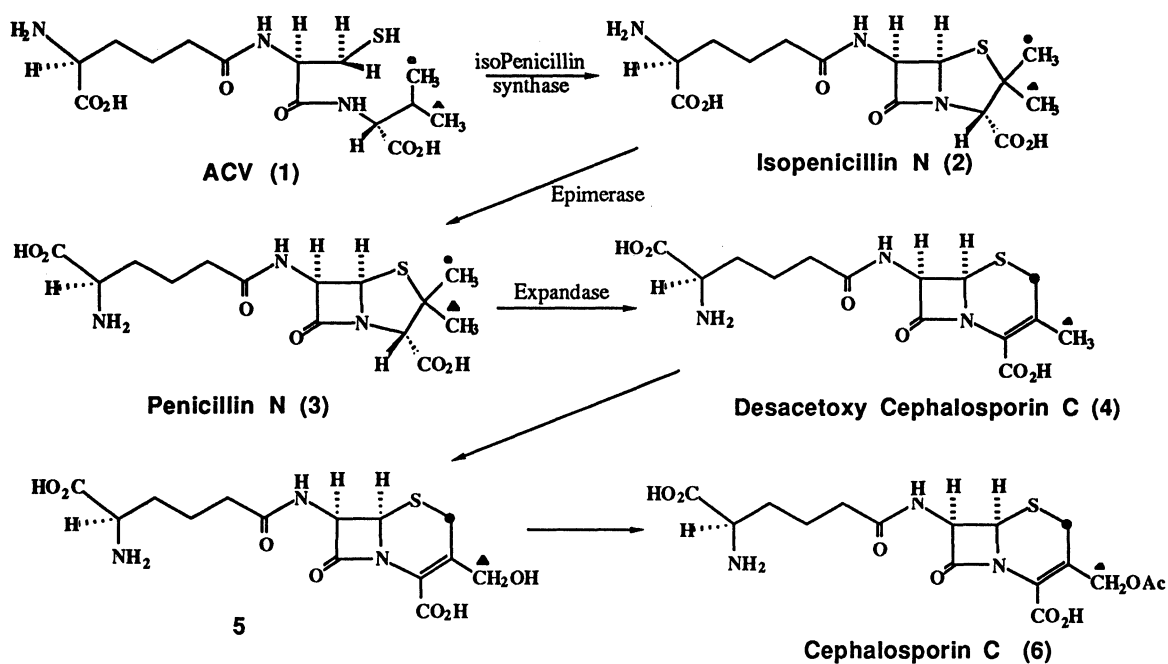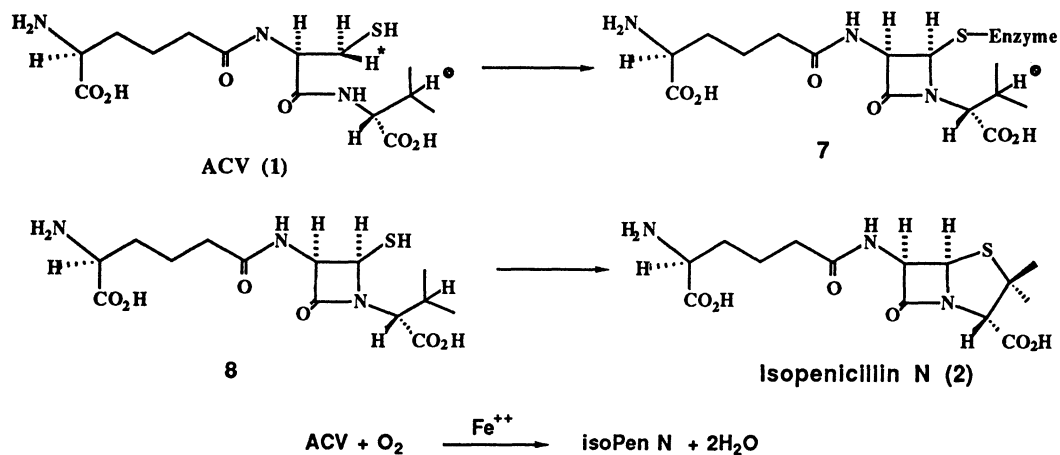
Figure 5

ACV + O₂ $\xrightarrow{\text{Fe}^{++}}$ isoPen N + 2H₂O

$$\text{ACV} + O_2 \xrightarrow{\text{Fe}^{++}} \text{isoPen N} + 2H_2O$$

Figure 6

Synthetic preparation of (**8**) has been achieved in two independent studies (refs. 5, 6) but the attempted conversion (**8**) → (**2**) has not been realized due to the short half-life of (**8**) at pH 6-7, the working range of the cell-free system. Only by indirect kinetic isotope methods can it be deduced that the rate-determining step involves the proton (*) in ACV (**1**) rather than the β–valine proton (•) (refs. 1, 7). In our laboratory the use of **doubly** labeled ACV in which C-3 of Cys and N of Val were enriched with ^{13}C and ^{15}N, respectively, provided an excellent probe (ref. 8) whereby the discharge of (**8**) from the enzyme could have been recognized by ^{13}C-NMR spectroscopy. In the event, only the final molecule of isoPen N (**2**) exhibiting the anticipated ^{13}C-^{15}N coupling (J = 4 Hz) was detectable under the experimental conditions.

Also tested was the ACV hydroxamate (ref. 9) (**9**) which proved to be an inhibitor of IPNS.

(9)
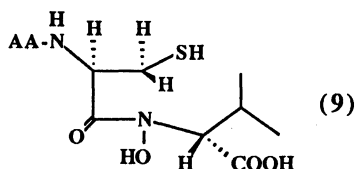
A similar experiment was conducted with the IPNS inhibitor ACG (**10**)[1] synthesized with ^{13}C at the Cys-3 position. This specimen is easily prepared from protected AC which is condensed with protected glycine followed by deprotection. The tripeptide (**10**) can only form one ring, i.e. the β–lactam (**12**) so it was of considerable interest to observe the steady state spectrum of the enzyme-inhibitor complex at 1 mM concentration. Since 50 mg batches of pure IPNS are required for these NMR experiments at high concentrations a new source of the enzyme was required. In collaboration with Dr. Y. Aharonowitz (Tel Aviv) the IPNS gene from *Streptomyces jumonjinensis* cloned into a T7 polymerase system was expressed (ref. 10) and after solubilization of the resultant inclusion bodies could be refolded to give the active enzyme as shown in Figure 7.
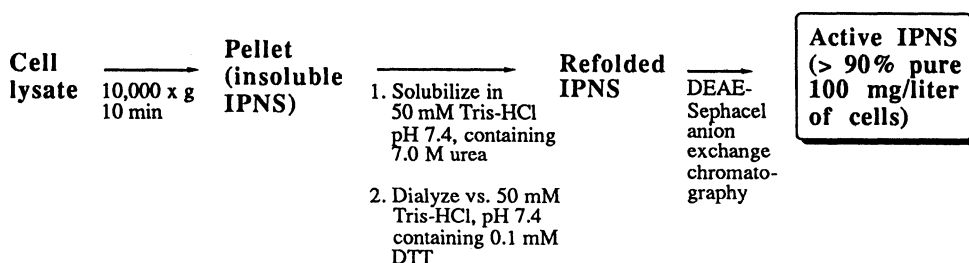

Figure 7   Purification of IPNS

Incubation of ACG (2mM) with IPNS (1mM) at 5°C in the NMR tube revealed (Fig. 8) a persistent new signal at δ66, coincident with the chemical shift of the synthetic β-lactam thiol (**12**) labelled with ^{13}C in the β-carbon of the cysteine moiety. On warming the solution to 25 °C the resonance
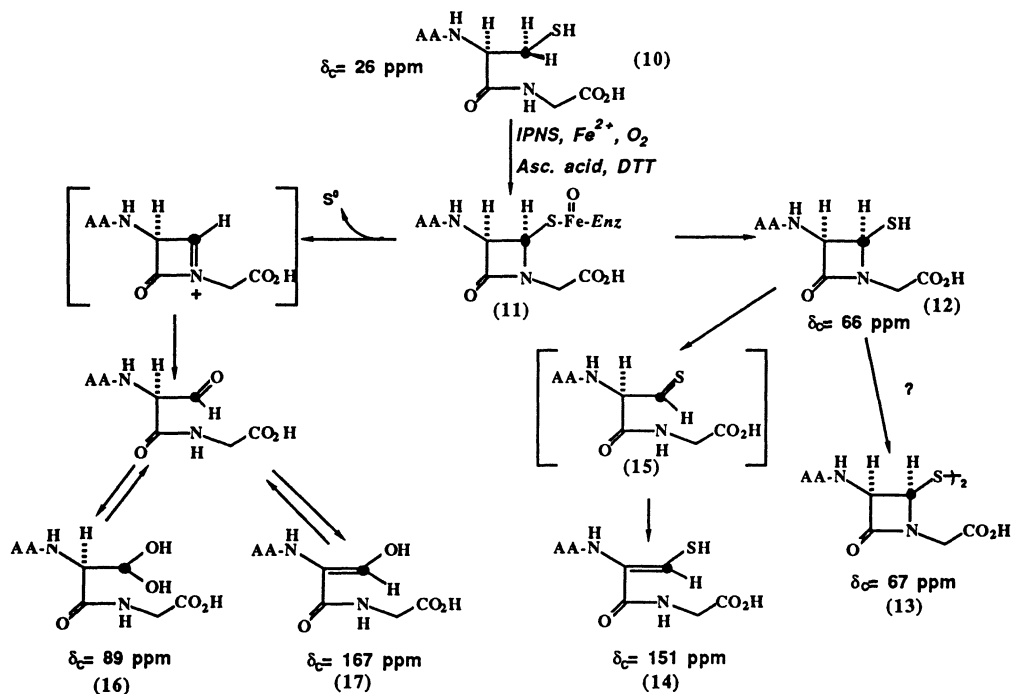

Figure 8

shifted to δ67, ascribed to the disulfide structure (13). At the same time another new signal at δ151 marked the formation of the enethiolate (14), labeled as shown (●) and formed by C-N bond cleavage to the labile thioaldehyde (15) and enolization. This experiment provides the first direct evidence for the monocyclic β-lactam structure (12) generated by IPNS. In addition two other new resonances mark the operation of β elimination of the cysteinyl sulfur to form the aldehyde hydrate (16; δ89) and its **enolate** (17; δ167) which had previously been discovered by Baldwin *et al* (ref. 11) during studies on the mechanism of IPNS employing ACG and by isotopic substitution of ACV with deuterium (ref. 1) to suppress the second ring closure, which also led to the extrusion of sulfur.

By using "antifreeze" solvents such as DMSO-H$_2$O (1:3) it has been possible to extend the catalytic turnover of IPNS with ACV from 20 min. at 25°C to 20 hr. at -15 °C. In this way the observation of any enzyme bound intermediates can be made at 1mM concentration of enzyme-substrate complex. All of the above NMR observations are in accord with current views (ref. 1) on the active site of IPNS in which Fe^{II} is coordinated to 2 histidine residues (ref. 12) and participates in the mechanism by forming a complex with the cysteine thiolate of ACV (Fig. 9). This is followed by addition of O$_2$ and thence via E-I$_1$ and the iron-carbon bonded species (E-I$_2$) both rings are formed with the expulsion of 2H$_2$O. Confirmation of these proposals for the mechanism for IPNS must await the 3-dimensional structure whose solution by X-ray diffraction is now in progress.
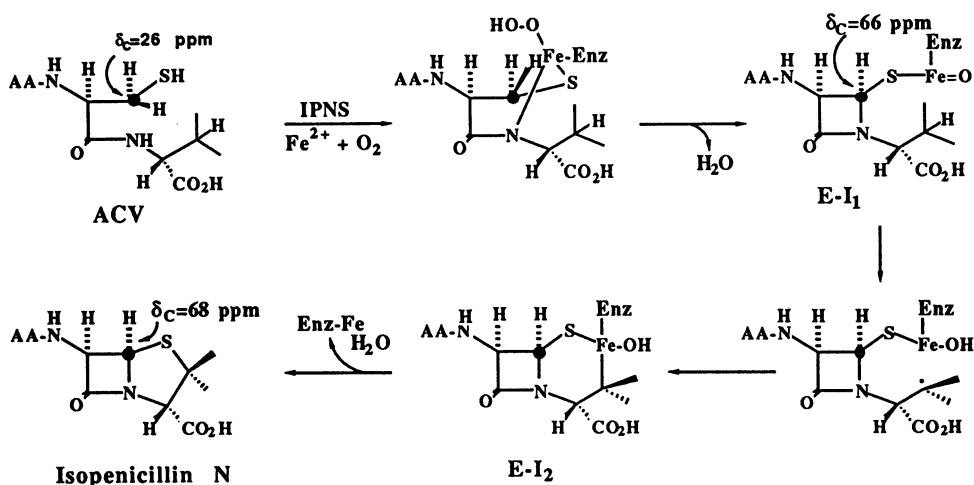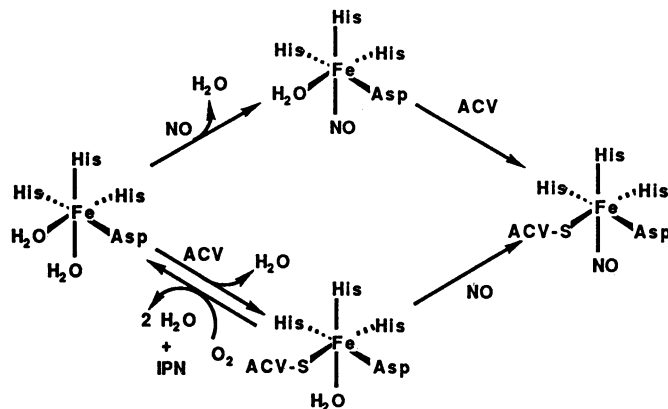

Figure 9a

Figure 9b

TRANSFER OF PLANT GENES AND EXPRESSION IN BACTERIA: SYNTHESIS OF INDOLE ALKALOIDS

The second topic concerns our recent research on the biosynthesis of the indole alkaloids of *Catharanthus roseus*, a subject which has been actively pursued in our laboratories at Vancouver, Sussex, Yale and Texas since 1962 (ref. 13). Again, the techniques of radio and stable isotope tracer feeding, cell-free systems and enzyme purification have been powerfully complemented by genetic engineering by expression of plant enzymes in heterologous systems (bacteria, yeast, insect cells). We chose as our first example the key enzyme strictosidine synthase (ref. 14) which catalyzes the Pictet-Spengler type condensation of the monoterpenoid glucoside secologanin and tryptamine (Fig. 10). The
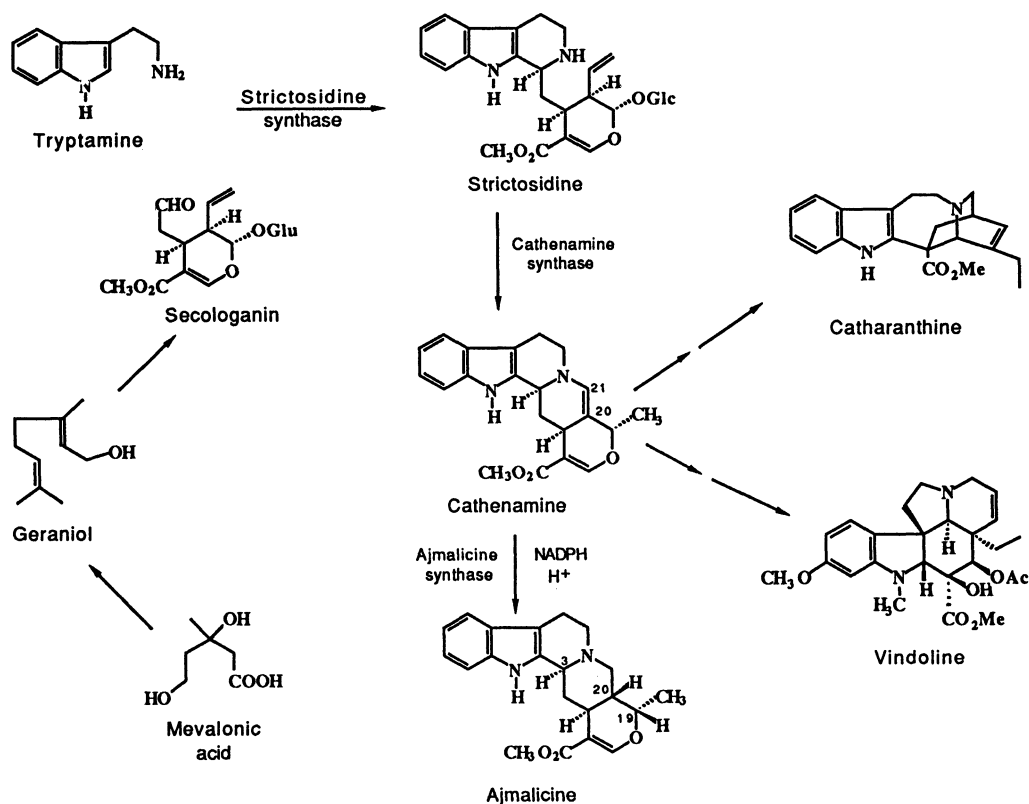

Figure 10

product, 3α-(S)-strictosidine, is the progenitor of more than 1200 indole alkaloids as well as the important quinoline, camptothecin. Pioneering experiments by the Zenk group (ref. 15) have already shown that the cDNA library from *Rauwolfia serpentina* can be probed with oligonucleotide and a cDNA clone isolated and transferred to *E. coli*.

As an aid to our efforts to study the enzymatic mechanism of strictosidine synthase by NMR (which requires 20-40 mg of pure enzyme for each experiment), and to produce large quantities of strictosidine for use as substrate for the next enzyme in the pathway (cathenamine synthase), a glucosidase which forms cathenamine, we have cloned and determined the nucleotide sequence of the homologous gene from *C. roseus* (ref. 16). This gene has also been transformed into tobacco plants where it has been shown to express active strictosidine synthase (ref. 17). The overexpression of active *C. roseus* strictosidine synthase in *E. coli* was achieved (ref. 18) using the expression cassette polymerase chain reaction (ECPCR) technique (ref. 19) and led to the capability of purifying the enzyme in quantities exceeding those previously attainable. Strictosidine synthase has been shown by ultrastructural immunolocalization to be associated with plant vacuoles. To effect transport across the endoplasmic reticulum and into the vacuole, it is normally synthesized with a signal peptide which is, presumably, removed by signal peptidase during transport. Therefore, we designed our ECPCR primers so that the expressed protein would contain a methionine residue immediately followed by the amino acids of the **processed** enzyme which were predicted from the DNA sequence (ref. 16).

When induced with IPTG, *E. coli* strain XA90(pRD1) displayed a new protein band with an M_r of about 34,000 on SDS-PAGE gels that was not seen in uninduced cells. A time course of the expression demonstrated that collecting the cells 2 hours after induction provided good expression of the new protein. After being grown in the presence of IPTG for two hours, the whole cells displayed strictosidine synthase activity as demonstrated by the appearance of strictosidine by cells bearing the pRD1 but not by control cells bearing pHN1+ with no insert. A cell free extract prepared by sonication of the cells followed by centrifugation at 12,000 x g had about four times higher activity than whole cells. This enzymatic activity was in the soluble fraction even though most of the protein associated with the induced band was found in the insoluble fraction, presumably in the form of inclusion bodies. However, only a very small portion of the protein was soluble as there was not enough to be seen by SDS-PAGE and attempts to purify it required up to eight different steps and resulted in less than 1 mg of 50% pure enzyme from eight liters of cells. The insoluble pellets were next washed with buffer and 3 M urea, solublized in 8 M urea, dialyzed and subjected to anion exchange chromatography. The resulting protein was >90% pure as judged by SDS-PAGE and had a specific activity of 31 nkat/mg compared to 104 nkat/mg reported by other workers for enzyme isolated from cell cultures of *C. roseus* (ref. 15). The sequence of the first ten amino acids of the purified enzyme was found to be **Met Ser Pro Ile Leu Lys Lys Ile Phe Ile** corresponding exactly to the sequence predicted from the nucleotide sequence. The M_r of the protein usually appeared to be about 34,000, slightly less than the 36,074 molecular weight predicted from the nucleotide sequence of the gene, suggesting the protein had been partially hydrolyzed by proteolysis. Similar processing of the enzyme has also been reported in both of the previous reports on the expression of the strictosidine synthase in heterologous systems (ref. 15). Therefore, the lower specific activity of our recombinant enzyme (about $1/3$ less than that of the enzyme isolated from *C. roseus* cell cultures) may be due to proteolysis, which would have to occur at the C-terminal since the N-terminal is unaltered. The availability of the recombinant enzyme is highly advantageous, for we now have the capability to purify as much enzyme activity from a **single liter** of a 2 hour induced culture of bacterial cells (about 2.5 g of cells) as was previously possible starting with **1 kg of *C. roseus* cells** which require a complex growth medium and at least a week to grow. The purification procedure is also greatly simplified, requiring 2-3 days and only one chromatography step.

The above technique is applicable to all of the enzymes necessary for the synthesis of the alkaloids of *C. roseus* and indeed is a general solution to targeted natural product synthesis of complex molecules. Desirable but "endangered plant" substances e.g. taxol, camptothecin, can in principle be synthesized once the cDNA library and the repertoire of **sequenced** enzymes is available. We believe that this approach to plant natural products is not only viable, but offers a novel departure from, and a valuable complement to, conventional synthetic chemistry, and may become the method of choice by the turn of this century.

MULTI ENZYME SYNTHESIS OF NATURAL PRODUCTS

During the course of our studies on corrin biosynthesis, a pathway involving over 20 enzymes, we have been fortunate not only in expressing the genes encoding almost all of the desired enzymes but to our initial surprise, it was found that their combination in sets of as many as six enzymes at a time provided a synthetic machine *par excellence* for one flask preparation of several advanced intermedaites endowed both with complex functionality and a wealth of stereochemical detail. From the many experiments performed recently, the syntheses of factor S_3 (ref. 20), an unusual zinc corphinate occurring naturally in *P. shermanii* and of sirohydrochlorin, the iron free prosthetic group of *E. coli* sulfite reductase (ref. 21) have been chosen to illustrate both the potential and success of this approach.

As summarized in Figure 11, the 5-carbon building block, ALA, can be dehydrated to the monopyrrole PBG with the overexpressed enzyme dehydratase encoded by the *hem*B gene. The polymerizing enzyme of tetrapyrrole synthesis is next used to form uro'gen I and thence by way of *triple* C-methylation catalyzed by the *E. coli* enzyme M-1 (*cysG*) the pyrrocorphin is prepared in quantitative yield using S-adenosyl methionine as co-substrate. Finally in the presence of a methyl transferase encoded by the *cbiF* gene of *Salmonella*, the 4th enzymatic step is carried out in the presence of zinc chloride to form the natural product factor S_3, isolated earlier by G. Müller (Stuttgart) from *P. shermanii*. Remarkably, when all four enzymes are added to the substrate ALA in the presence of Zn^{++} and SAM the overall yield of factor S_3 approaches 100%. It now remains to determine the absolute stereochemistry of the ring D methyl insertion, although the enzymatic synthesis reveals the order of C-methylation and leads to a revised structure for S_3.
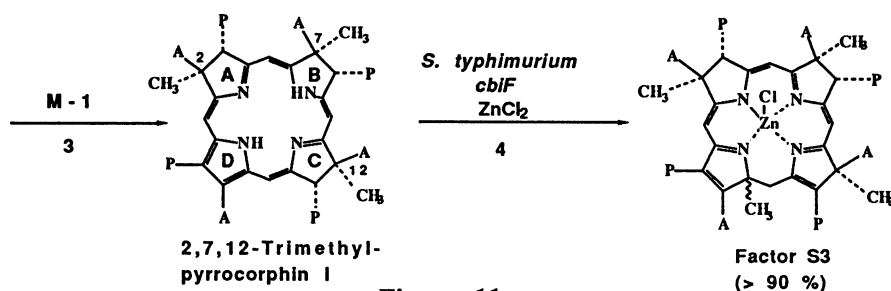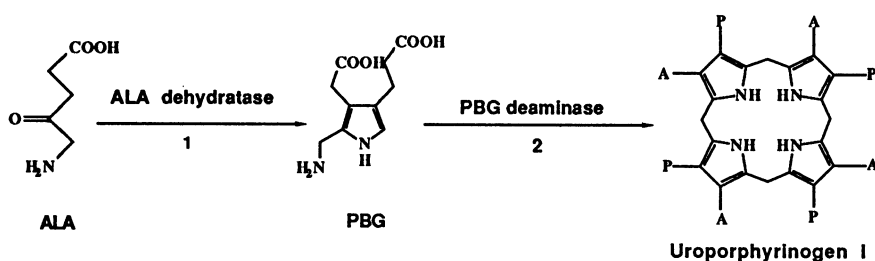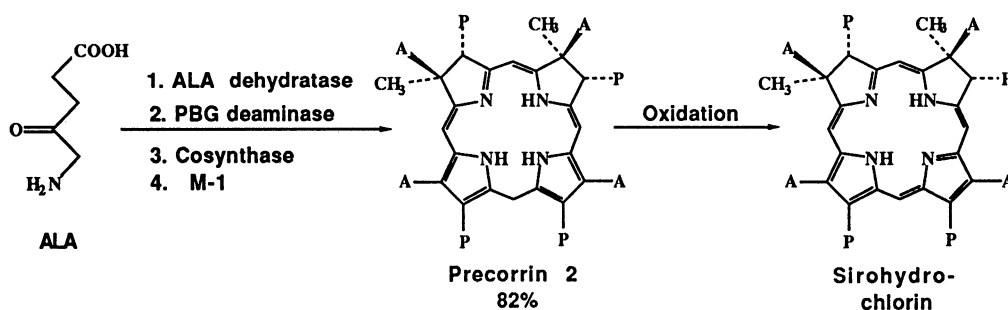
Figure 11

Figure 12

Our second example (Figure 12) again uses ALA as the building block, this time for the rapid assembly of the type III porphyrinoid template by the first 3 enzymes followed by C-methylation with SAM to afford precorrin-2 in 82% yield. Oxidation to sirohydrochlorin completes the synthesis. The 4-enzyme reaction can be conveniently performed in one flask and by omitting the oxidation step, can be coupled to the transformation by 2 more C-methylases from *Salmonella* to form a tetramethyl corphinate in a 6-enzyme, one-flask reaction. These experiments are designed to effect the complete multi-enzyme synthesis of advanced intermediates on the B_{12} pathway, a topic considered in detail elsewhere in this symposium.

In summary, what was once a pipe dream, namely the observation of natural products and their intermediates being formed in a test tube by the sequential action of as many as 5 or 10 enzymes is now a reality, thanks to the powerful combination of organic chemistry, NMR spectroscopy and molecular biology. In other words we have changed our rôle as spectators of natural product biosynthesis to that of imposing control on the genetic machinery to execute a programmed synthesis of the target molecule. These are indeed exciting times for bioorganic chemistry and for those of us who are privileged to be able to stand on the shoulders of the pioneers who gave us superb tools, such as high resolution NMR and cloning techniques, without which none of the experiments described above could even have been contemplated, let alone designed and executed.

Acknowledgements

REFERENCES

1. For recent reviews of research on penicillin biosynthesis see (a) J. E. Baldwin in "Recent Advances in the Chemistry of β–Lactam Antibiotics," Chapter 5 (pp. 62-85), Roy. Soc.Chem. Special Publication No. 52 (A. G. Brown and S. M. Roberts, Eds.), London, 1985: (b) J. E. Baldwin and E. Abraham, *Nat. Prod. Reports*, **5**, 129 (1988).
2. T. Koroni, S. Herchen, J. E. Baldwin, M. Yoshida, N. A. Hunt and A. L. Demain, *Biochem. J.*, **184**, 427 (1979).
3. J. O'Sullivan, R. C. Bleaney, J. A. Huddleston and E. P. Abraham, *Biochem. J.*, **184**, 421 (1979).
4. R. L. White, E. M. John, J. E. Baldwin and E. P. Abraham, *Biochem. J.*, **203**, 791 (1982).
5. E. P. Abraham, R. M. Adlington, J. E. Baldwin, M. J. Crimmin, L. D. Field, G. S. Jayatilake, R. L. White and J. J.Usher, *Tetrahedron*, **40**, 1907 (1984).
6. a) S. K. Chung, R. Shankaranarayan and A. I. Scott, *Tetrahedron Lett.*, 2941 (1983).
   b) A. I. Scott, R. Shankaranarayan and S. K. Chung, *Heterocycles*, **30**, 909 (1990).
7. J. E. Baldwin, R. M. Adlington, J. E. Moroney, L. B. Field and H.-H Ting, *J. Chem. Soc. Commun.*, 984 (1984).
8. R. L. Baxter, C. J. McGregor, G. A. Thomson and A. I. Scott, *J. Chem. Soc. Perkin I*, 369 (1985).
9. R. L. Baxter, G. A. Thomson and A. I. Scott, *J. Chem. Soc., Chem. Commun.*, 32 (1984).
10. O. Landman, D. Shiffman, Y. Av-Gay, Y. Aharonowitz and G. Cohen *FEMS Microbiology Lett.*, **84**, 239-244 (1991).
11. J. E. Baldwin, M. Bradley, R. M. Adlington, W. J. Norris and N. J. Turner *Tetrahedron*, **47**, 457-480 (1991).
12. L.-J. Ming, L. Que, A. Kriauciunas, C. A. Frolik, V. J. Chen, *Inorg. Chem.*, **29**, 1111-1112 (1990).
13. Review: A. I. Scott, *Acc. Chem. Res.* **3**, 151 (1970).
14. H. Mizukami, H. Nordlov, S. L. Lee and A. I. Scott, *Biochemistry*, **18**, 3760 (1979).
15. Review: T. M. Kutchan, H. Dittrich, D. Bracher, M. H. Zenk, *Tetrahedron*, **31**, 5945 (1991).
16. T. D. McKnight, C. A. Roessner, R. Devagupta, A. I. Scott, C. L. Nessler, *Nucl. Acids Res.*, **18**, 4939 (1990).
17. T. D. McKnight, D. R. Berger, R. J. Burnett, C. L. Nessler, *Planta Medica*, in press (1992).
18. C. A. Roessner, R. Devagupta, M. Hasan, H. J. Williams and A. I. Scott, *Protein Expression & Purification*, in press (1992).
19. K. D. MacFerrin, M. Terrauera, S. L. Schreiber, G. Verdine, *Proc. Nat. Acad. Sci. (USA)*, **87**, 1937 (1990).
20. A. I. Scott and S. I. Ozaki, to be published.
21. (a) M. J. Warren, N. J. Stolowich, P. J. Santander, C. A. Roessner, B. A. Sowa, and A. I. Scott, *FEBS Lett.*, **261**, 76-80 (1990); (b) A. I. Scott, *Tetrahedron*, **48**, 2559-2578 (1992).