Tetrahedron Vol. 48, No. 13, pp. 2559-2578, 1992
Printed in Great Britain

0040-4020/92 $3.00+.00
© 1992 Pergamon Press Ltd

Genetically Engineered Synthesis of Complex Natural Products†

A. Ian Scott

Center for Biological Nuclear Magnetic Resonance, Department of Chemistry
Texas A&M University, College Station, Texas 77843-3255, USA

Abstract: *The feasibility of multi-enzyme, one-flask total synthesis of natural products is demonstrated by the preparation of a tetramethyl corphinoid structure and of the alkaloid strictosidine by gene transfer and overexpression of the corresponding gene products in E. coli.*

For the past 40 years 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. Indeed the **first** use of carbon-13 in biosynthesis was Wasserman's elegant study of prodigiosin formation in *Serratia marcescens* in 1970. Now, twenty years on from these pioneering experiments 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 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

† Dedicated, with respect, to Harry Wasserman on the occasion of his 70th birthday, in appreciation of his scholarship in chemistry, music, and art, and as a token of our friendship.

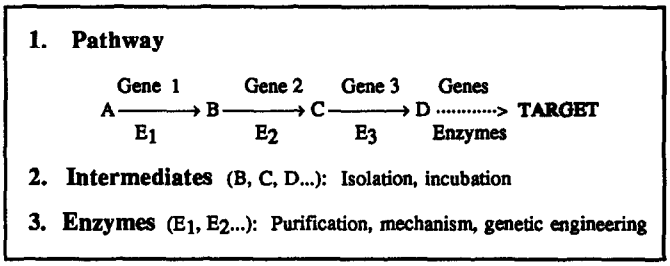


Figure 1 Natural Product Biosynthesis

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

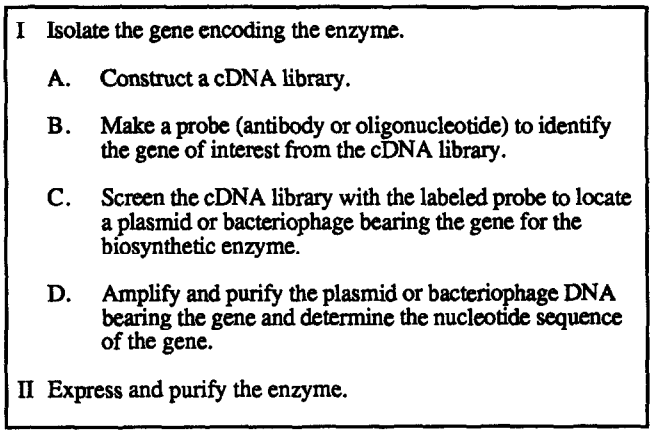


Figure 2 Genetic engineering of biosynthetic enzymes for natural products

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 NMR spectroscopy to assay each biosynthetic step as it occurs. 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

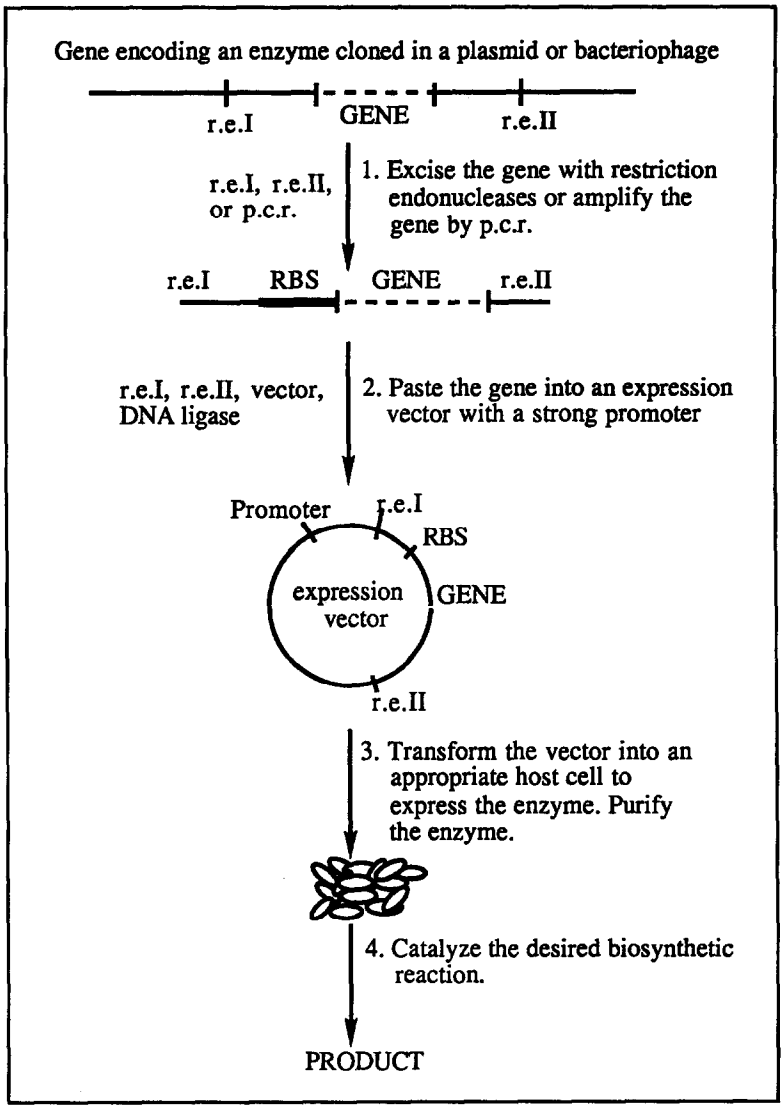


Figure 3 Expression of Biosynthetic Enzymes

substrate specificity thus producing new variations on the biosynthetic theme, as portrayed in Figure 4. This article illustrates the power of such an approach applied to two completely different natural product targets - the corrinoids and the indole alkaloids.

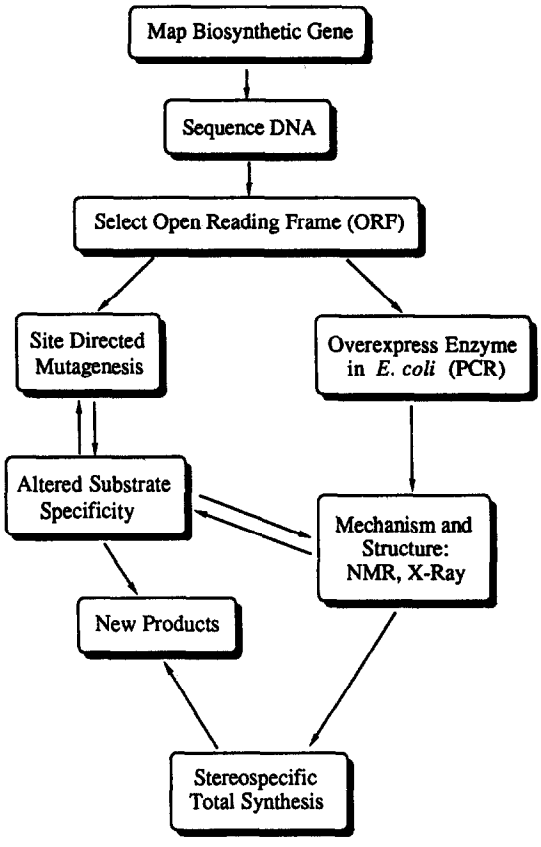
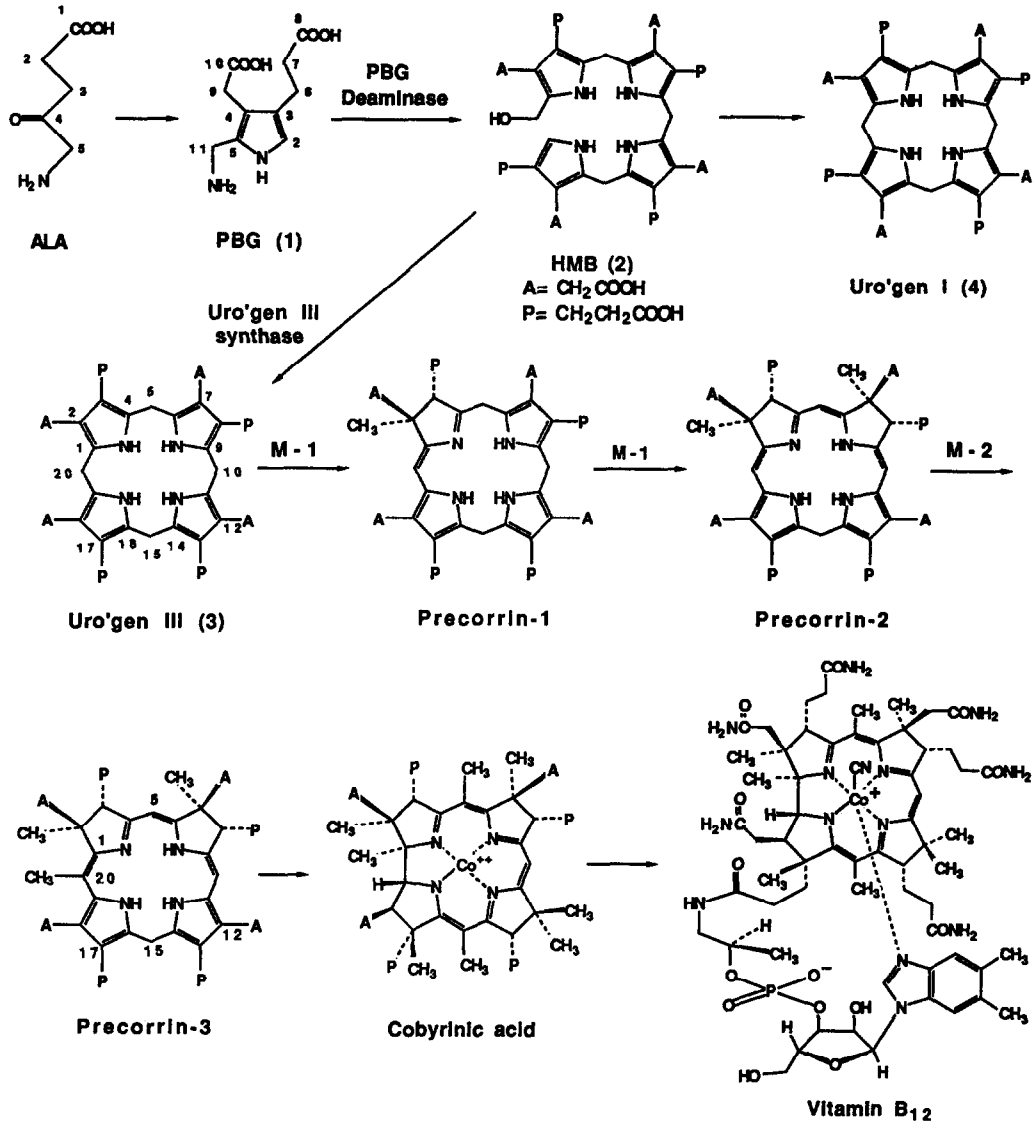


Figure 4 Genetic Engineering of Biosynthetic Enzymes

Porphyrin and Corrin Biosynthesis

Beginning at Yale in 1968 we have been studying Nature's road to porphyrinoids and corrins.[1] The route is summarized in Scheme I which introduces the early intermediates 5-aminolevulinic acid (ALA) and porphobilinogen (PBG, **1**), their enzymatic transformation to the macrocyclic template of uro'gen III (**3**) *via* the hydroxymethyl bilane (**2**), and the partially methylated, reduced isobacteriochlorins, precorrins **1-3**, which represent successive C-methylation of (**3**) by two methyltransferases, M-1 and M-2 with *S*-adenosyl methionine (SAM).

PBG deaminase (EC4.3.1.8) catalyzes the tetramerization of PBG (**1**) to preuro'gen (hydroxymethylbilane, HMB; **2**)[2,3] which is cyclized with rearrangement to the unsymmetrical uro'gen III (**3**) by uro'gen III synthase[3,4,5] (EC4.2.1.75) (Scheme 1). In the absence of the latter enzyme, preuro'gen (**2**) cyclizes to uro'gen I (**4**), which, as discussed below, turns out to be a substrate for the methylases of the vitamin B_{12} pathway. We have used genetic engineering to construct a plasmid pBG 101 containing the *Escherichia coli hemC* gene[6,7] encoding deaminase. *E. coli* (TB1) transformed with this plasmid produces

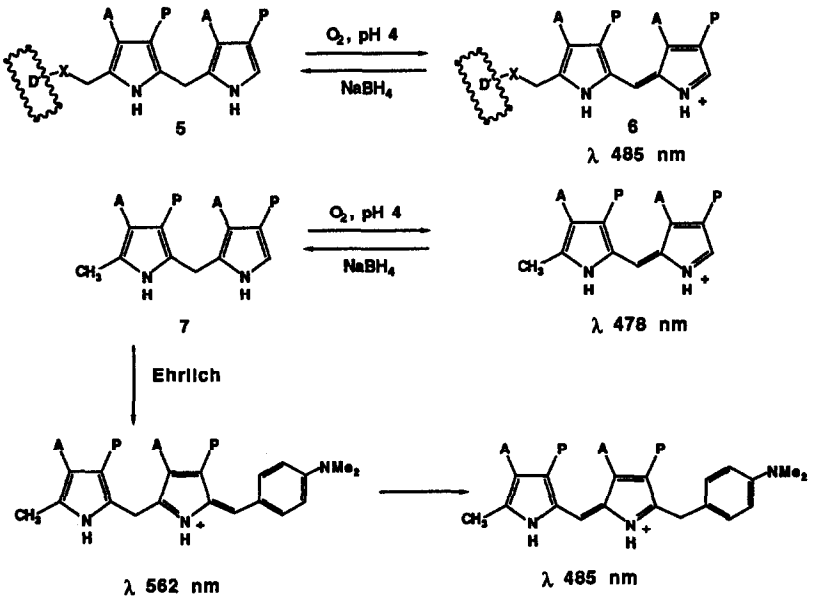


Scheme 1

deaminase at levels greater than 200 times those of the wild strain[7] thereby allowing access to substantial quantities of enzyme for detailed study of the catalytic mechanism.

Previous work with deaminase [8-11] had established that a covalent bond is formed between substrate and enzyme, thus allowing isolation of covalent complexes containing up to 3 PBG units (ES_1-ES_3). Application of ^{3}H-NMR spectroscopy to the mono PBG adduct (ES_1) revealed a rather broad ^{3}H chemical shift indicative of bond formation with a cysteine thiol group at the active site.[2] However, with adequate

supplies of pure enzyme available from the cloning of *hemC* it was shown that a novel cofactor, derived from PBG during the biosynthesis of deaminase, is covalently attached to one of the four cysteine residues of the enzyme in the form of a dipyrromethane which, in turn, becomes the site of attachment of the succeeding four moles of substrate during the catalytic cycle. Thus, at pH < 4, deaminase (**5**) rapidly develops a chromophore (λmax 485 nm) diagnostic of a pyrromethene (**6**), whilst reaction with Ehrlich's reagent generates a chromophore typical of a dipyrromethane (λmax 560 nm) changing to 490 nm after 5-10 min. The latter chromophoric interchange was identical with that of the Ehrlich reaction of the synthetic model pyrromethane (**7**) and can be ascribed to the isomerization shown (Scheme 2) for the model system (**7**). Incubation

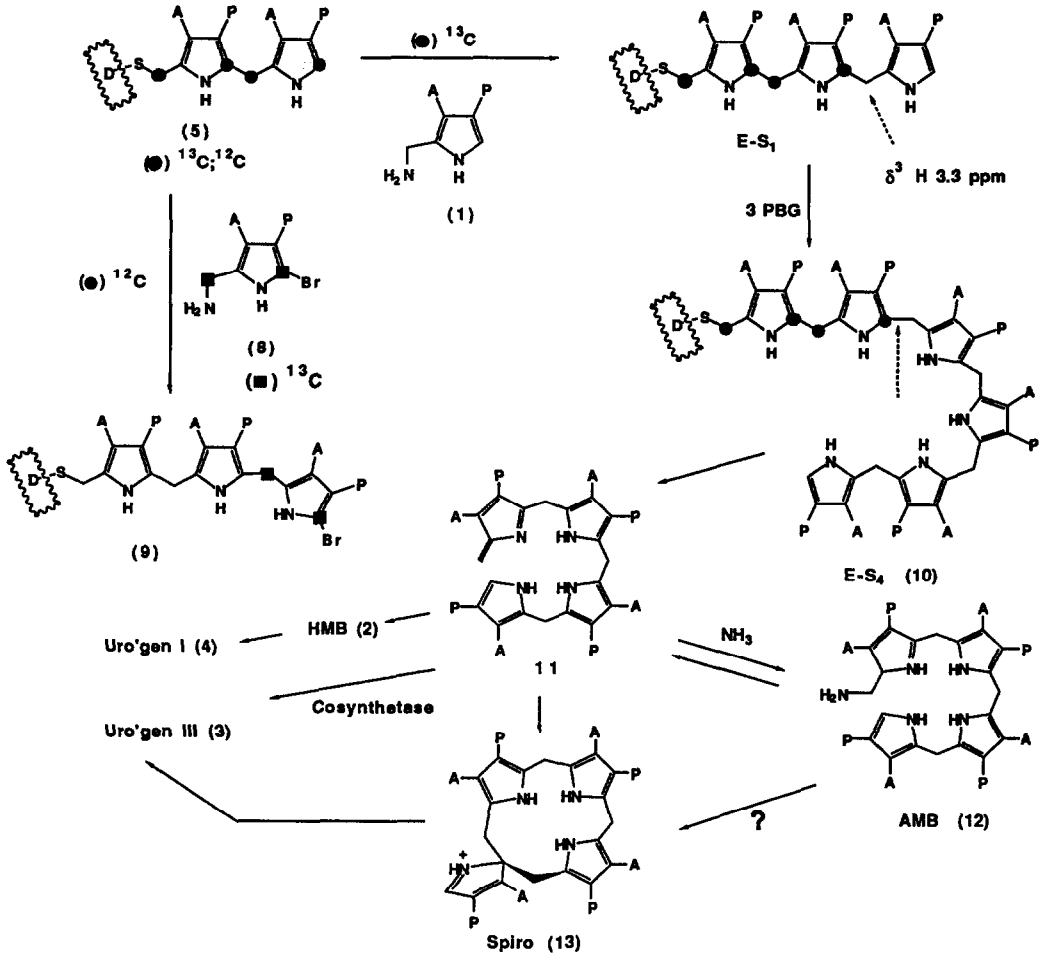


Scheme 2

of *E. coli* strain SASX41B/pBG101 (*hemA*⁻ requiring ALA for growth) with [5-^{13}C]-ALA afforded highly enriched enzyme for NMR studies. At pH8, the enriched carbons of the dipyrromethane (py-CH_2-py) were clearly recognized at 24.0 ppm (py-CH_2-py), 26.7 ppm (py-CH_2-X), 118.3 ppm (α-free pyrrole) and 129.7 ppm (α-substituted pyrrole). The signals are sharpened at pH12 and the CH_2 X-resonance is shifted to δ29.7 in the unfolded enzyme. Comparison with synthetic models reveals that a shift of 26.7 ppm is in the range expected for an α-thiomethyl pyrrole (py-CH_2-SR). Confirmation of the dipyrromethane (rather than oligo pyrromethane) came from the ^{13}C INADEQUATE spectrum taken at pH12 which reveals the expected coupling only between py-CH_2-py (δ 24.7) and the adjacent substituted pyrrole carbon (δ 128.5 ppm). When the enriched deaminase was studied by ^{1}H detected heteronuclear filtered spectroscopy, only 5 protons (in a total of **several thousand!**) attached to ^{13}C-enriched nuclei were observed in accord with structure **5**. A specimen of deaminase was then covalently inhibited with the suicide inhibitor [2,11-$^{13}C_2$]-2-bromo PBG (**8** Scheme 3) to give a spectrum consistent only with structure **9**. The site of covalent attachment of substrate

(and inhibitor) is therefore **the free α-pyrrole carbon at the terminus of the dipyrromethane** in the native enzyme, leading to the structural and mechanistic proposal for deaminase portrayed in Scheme 3.

It was then found that 2 moles of PBG are incorporated autocatalytically into the apoenzyme (isolated *via* overexpression in a *hemB*⁻strain of *E. coli* which **does not make PBG**)[13] and that the first (kinetic) encounter of PBG deaminase with substrate involves attachment of PBG (with loss of NH_3) to the α-free pyrrole position of the dipyrromethane to form the ES_1 complex[14,15] (Scheme 3). The process is repeated to



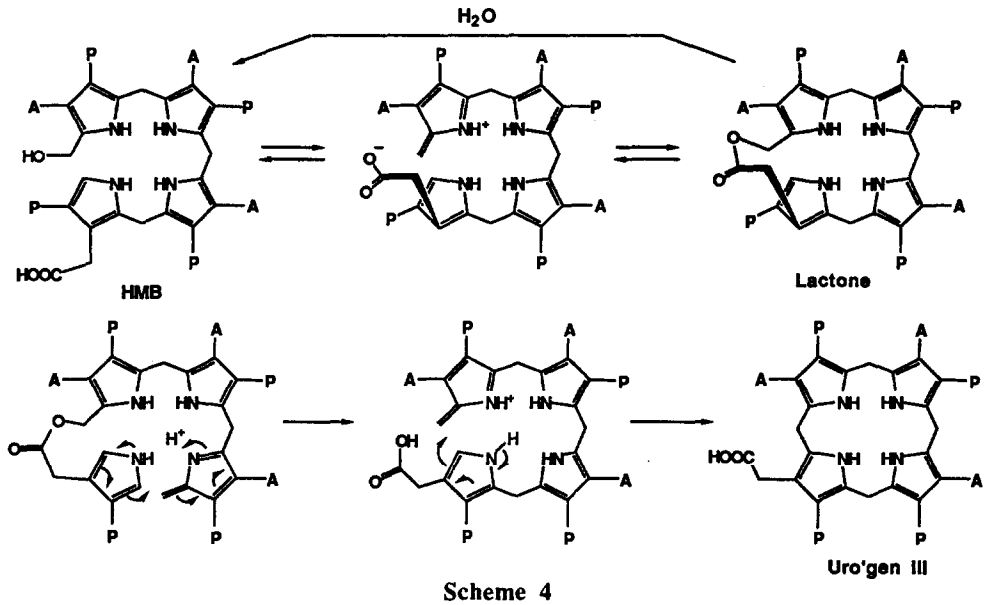
Scheme 3

produce the "tetra PBG" (ES_4) adduct (**10**). At this juncture site-specific cleavage of the **hexapyrrole** chain (at →) releases the azafulvene bilane (**11**) which **either** becomes the substrate of uro'gen III synthase, or in the absence of the latter enzyme, is stereospecifically hydrated[3] to HMB (**2**) at pH12, then cyclized chemically to uro'gen I (**4**) at pH ≤ 8. Independent, complementary work in 1987-8 from two other laboratories[16,17]

reached identical conclusions regarding the catalytic site. The ^{13}C-labeling defines the number of PBG units (two) attached in a head-to-tail motif to the native enzyme at pH8 and reveals the identity of the nucleophilic group (Cys-SH) which anchors the dipyrromethane (and hence the growing oligopyrrolic chain) to the enzyme. Site specific mutagenesis[7] and chemical cleavage[16,17] were employed to determine that Cys-242 is the point of attachment of the cofactor. Thus, replacement[7] of cysteine with serine at residues 99 and 242 gave fully active and inactive specimens of the enzyme respectively.

The use of the α-carbon of a dipyrromethane as the nucleophilic group responsible for oligomerization of 4 moles of PBG (with loss of NH_3 at each successive encounter with an α-free pyrrole) is not only remarkable for the exquisite specificity and control involved, but is, as far as we know, a process unique in the annals of enzymology in that a **substrate** is used not only once, but **twice** in the genesis of the active site cofactor! Even more remarkable is the fact that the apoenzyme is automatically transformed to the active holoenzyme by addition of two PBG units without the intervention of a second enzyme.

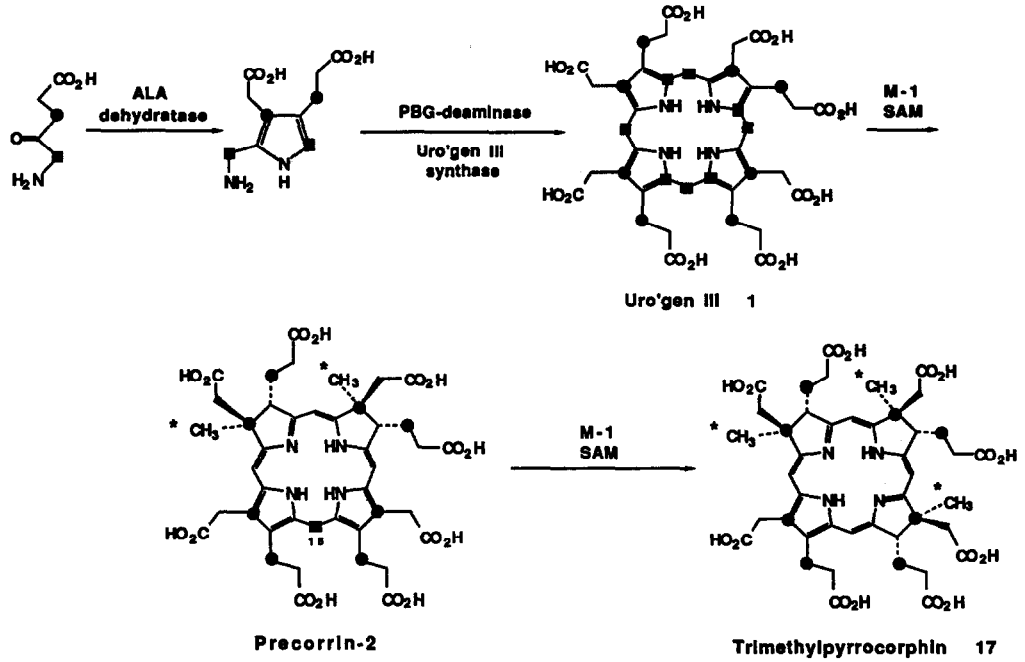
We now turn briefly to the **rearranging** enzyme Uro'gen III synthase. The early ideas of Bogorad[18] involving the aminomethyl bilane (AMB;**12**) as substrate were helpful in finally tracking down the elusive species which, after synthesis by PBG deaminase, becomes the substrate for Uro'gen III synthase. This was shown to be HMB (**2**) and at this stage (1978) the lack of activity of AMB (**12**) as substrate relegated the latter bilane to an interesting artefact produced quantitatively by addition of ammonia to deaminase incubations (Scheme 3). However in 1988 with the acquisition of substantial quantities of pure Uro'gen III synthase obtained by cloning the genes *hemC* and *D* **together**[19a] and overexpression in *E. coli*, now at a level 4,000 times greater than that described elsewhere,[20] the substrate specificity of the synthase (often called cosynthetase) has been reinvestigated. Ever since its conception by Mathewson and Corwin in 1961, the spiro compound (**13**) (Scheme 3) has been a favorite construct with organic chemists, since both its genesis through α-pyrrolic reactivity and its fragmentation - recombination rationalize the intramolecular formation of Uro'gen III from the linear bilane, preuro'gen (HMB;**2**). A careful search for the spiro-compound (**13**) was conducted at subzero temperatures in cryosolvent (-24° C; ethylene glycol/buffer) using various ^{13}C-isotopomers of HMB as substrate.[19] Although the synthase reaction could be slowed down to 20 hours (rather than 20 sec.) no signals corresponding to the quaternary carbon (*;δ ~80 ppm) or to the α-pyrrolic methylene groups (▲; δ 35-40 ppm) could be observed. During these studies, however, it was found that totally synthetic AMB (**12**) (thus free of deaminase which can catalyze interconversion of AMB into HMB **2**) served as a slow but productive substrate for Uro'gen III synthase **at high concentrations** (1 mmole in enzyme and substrate concentrations) and that if care is not taken to remove the ammonia liberated from PBG by the action of deaminase, not only is the enzymatic formation of AMB observed[21] but in presence of Uro'gen III synthase, the product is again Uro'gen III. Regardless of which version of the bilane system (HMB, AMB) is involved, the lack of observation of any intervening species free of enzyme strongly suggests that the intermediate is enzyme-bound and therefore difficult to detect. Indirect evidence for the "spiro" mechanism was adduced when it was found that one of the enantiomers of the corresponding synthetic ring D-**lactam** served as an inhibitor of Uro'gen III synthase.[22a] However a novel alternative to the "spiro intermediate" (**13**) hypothesis, still consistent with the known facts, has been proposed[19] using a self assembly concept involving **lactone** formation as portrayed in Scheme 4. We have suggested that generation of the azafulvene (**12**) (as before)



Scheme 4

is followed not by carbon-carbon bond formation (→ **13**) but by closure of the macrocycle *via* the macrolide (Scheme 4). The regio-specificity of this step is reflected by the failure[5] of synthetic bilanes lacking the acetic acid side chain at position 17 in ring D to undergo enzyme-catalyzed rearrangement to the type-III system, together with the observation that a "switched" bilane carrying a propionate at position 17 (and an acetate at 18) does indeed give a certain amount of (rearranged) uro'gen I (but also Uro'gen III) enzymatically,[5,22b] indicating that lactone formation could also be achieved (but less efficiently) by a propionate group at the 17-position. The subsequent chemistry is quite similar to the fragmentation-recombination postulated for the spiro system except that a "twisted macrolide" becomes the pivotal intermediate species and the sequence proceeds as shown in Scheme 4. Although this novel hypothesis is rather difficult to test experimentally, specific chemical traps for the lactones combined with the use of ^{18}O labeling are being used to confirm or refute the possible role of such macrocyclic lactones as the basis for the mechanism of Uro'gen III synthase.

The first of the methylase enzymes catalyzes the sequential formation of Factors I and II and has been named S-adenosyl methionine Uro'gen III methyl transferase (SUMT). SUMT was first partially purified from *P. shermanii* by G. Müller[23] and recently has been overexpressed in *Pseudomonas denitrificans*.[24] In *E. coli* it was found that the *CysG* gene encodes Uro'gen III methylase (M-1) as part of the synthetic pathway to siroheme, the cofactor for sulfite reductase, and overproduction was achieved by the appropriate genetic engineering.[25] Although, SUMT and M-1 appear to perform the same task, it has beeen found that their substrate specificities differ. Thus, it has been possible to study in detail the reaction catalyzed by M-1 directly using NMR spectroscopy and to provide rigorous proof that the structure of precorrin-2 is that of the dipyrrocorphin tautomer of dihydro-Factor II (dihydrosirohydrochlorin). Uro'gen III (enriched from [5-^{13}C]-ALA at the positions shown in Scheme 5) was incubated with M-1 and [$^{13}CH_3$]-SAM. The resultant



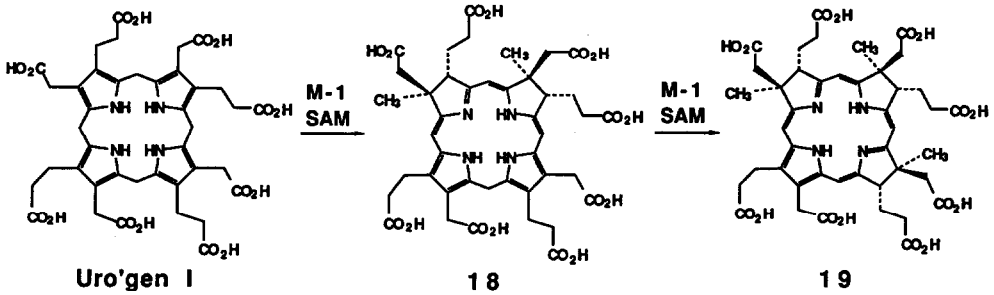
Scheme 5

spectrum of the precorrin-2 revealed an sp^3 enriched carbon at C-15, thereby locating the reduced center (shown as ■). By using a different set of ^{13}C-labels (• from ^{13}C-3 ALA) and [$^{13}CH_3$]-SAM) the sp^2 carbons at C_{12} and C_{18} were located as well as the sp^3 centers coupled to the pendant ^{13}C-methyl groups at (*) C_2 and C_7. This result confirms an earlier NMR analysis[23] of precorrin-2 isolated by careful anaerobic purification of the methyl ester, and shows that no further tautomerism takes place during the latter procedure. The two sets of experiments mutually reinforce the postulate that precorrins-**1**, -**2**, and -**3** all exist as hexahydroporphinoids and recent labeling experiments[26] have provided good evidence that precorrin-**1** is discharged from the methylating enzyme (SUMT) as the species with the structure shown (or a tautomer thereof).

However, prolonged incubation (2 hr.) of Uro'gen III with M-1 provided a surprising result for the UV and NMR changed dramatically from that of precorrin-**2** (a dipyrrocorphin) to the chromophore of a pyrrocorphin, hitherto only known only as a synthetic tautomer of hexahydroporphyrin. At first sight, this event seemed to signal a further tautomerism of a dipyrrocorphin to a pyrrocorphin catalyzed by the enzyme but when $^{13}CH_3$-SAM was added to the incubation, it was found that a **third** methyl group signal appeared in the 19-21 ppm region of the NMR spectrum. When Uro'gen III was provided with the ^{13}C labels (•) (as shown in Scheme 5) 3 pairs of doublets appeared in the sp^3 region (δ 50-55 ppm) of the pyrrocorphin product. The necessary pulse labeling experiments together with appropriate FAB-MS data finally led to the structural proposal (**17**)[27,28] for the novel trimethyl pyrrocorphin produced by "overmethylation" of the normal substrate, uro'gen III, in presence of high concentration of enzyme. Thus M-1 has been recruited to

insert a ring C methyl and synthesizes the long sought "natural" chromophore corresponding to that of the postulated precorrin-4 although in this case the regiospecificity is altered from ring D to ring C. This lack of specificity on the part of M-1 was further exploited to synthesize a range of "unnatural" isobacteriochlorins and pyrrocorphins based on isomers of Uro'gen III. Thus, Uro'gen I produces 3 methylated products corresponding to precorrin-1, precorrin-2 (18), and the type-I pyrrocorphin (19) (Scheme 6). These compounds are reminiscent of a series of tetramethyl type I corphinoids, Factors S_1-S_4 isolated from *P. shermanii*[29,30] which occur as their zinc complexes. Uro'gens II and IV can also serve as substrates for M-1 remarkably producing isobacteriochlorins in both cases.[28b] When uro'gen I was incubated with SUMT,[31] isolation of Factor II of the type I (Sirohydrochlorin I) family revealed a lack of specificity for this methyltransferase also, although in the latter studies no pyrrocorphins were observed. This may reflect a control mechanism in the *P. denitrificans* enzyme (SUMT) which does not "overmethylate" precorrin-2 as is found for the *E. coli* M-1 whose physiological function is to manufacture sirohydrochlorin for sulfite reductase necessary for cysteine synthesis. The fact that *E. coli* does not synthesize B_{12} could reflect an evolutionary process in which the C-methylation machinery has been retained, but is only required to insert the C-2 and C-7 methyl groups.

The sites of C-methylation in both the type I and III series are also reminiscent of the biomimetic C-methylation of the hexahydroporphyrins discovered by Eschenmoser[32] and the regiospecificity is in accord with the principles adumbrated[33] for the stabilizing effect of a vinylogous ketimine system. In principle the



Scheme 6

methylases of the B_{12} pathway, which synthesize both natural and unnatural pyrrocorphins and corphins can be harnessed to prepare several of the missing intermediates of the biosynthetic pathway. It is of note that the instability towards oxygen rationalizes our inability to isolate any new intermediates under aerobic conditions (> 5 ppm O_2).

The possibility was also examined that a complete corrinoid structure based on a type I template could be prepared using uro'gen I, or reduced sirohydrochlorin I, as substrates with the cell-free system capable of converting Uro'gen III to cobyrinic acid. However no type I cobyrinic acid was produced suggesting that Nature, although capable of inserting at least 4 methyl groups into Uro'gen I by C-alkylation is unable to effect the key step of ring contraction using the type-I pattern of acetate and propionate side chains. This

result, although negative, is in accord with a suggestion[33] concerning the requirement for two adjacent acetate side chains in pre-corrinoids, one of which (at C-2) can participate in lactone formation to the C-20 meso position, a postulate supported by experiments[34] with ^{18}O labeled precursors, while the second acetate function (at C-18) is used as an auxiliary to control the necessary activation of C-15 for C-methylation.

Our search for the biosynthetic enzymes leading to precorrin-2 has so far relied on the "heme gene box" (Figure 5) in *E. coli* which controls the synthesis of precorrin-2 as a cofactor for *E. coli* sulfite

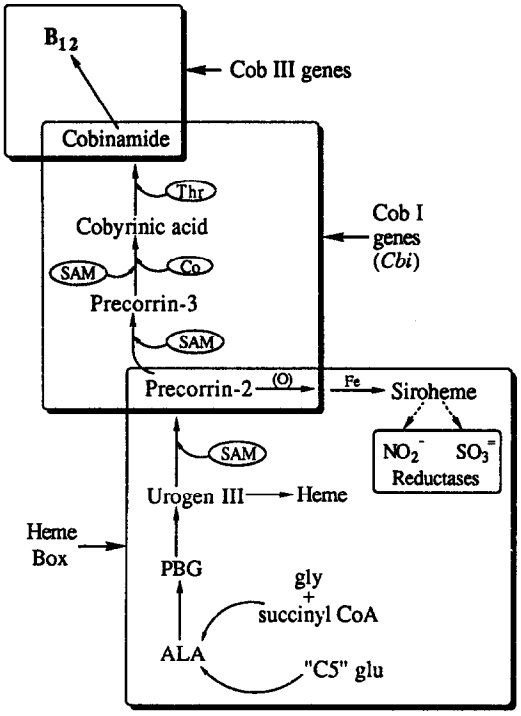


Figure 5

reductase. Fortunately, the heme and corrin pathways intersect in *E. coli* and *Salmonella typhimurium*. Although the former organism does not make B_{12}, the discovery by J. Roth that anaerobic fermentation of *S. typhimurium* produces vitamin B_{12}, has opened the way to enhance our knowledge of vitamin B_{12} synthesis using the vast array of genetic and cloning techniques available with this organism. Three loci at minutes 14, 34 and 42 have been identified by mutation and complementation studies.[35] The main gene cluster at 42 min. contains the machinery (Cob I) necessary for the synthesis of cobinamide from precorrin-2 (Figure 5), a process involving 6 C-methylations at C-1, C-5, C-12, C-15, C-17 and C-20, decarboxylation (at C-12), ring contraction, loss of acetic acid (from C-20 and its attached methyl) amidation and cobalt insertion. Thus the Cob region responsible for cobinamide synthesis probably contains the genes encoding all of these

biosynthetic functions. Through the generosity of Professor Roth we were able to use his sequence data to clone and express the gene products, corresponding to 12 open reading frames (ORF's) necessary to synthesize cobyrinic acid from precorrin-2 (Figure 6). These studies[36] were also aided by the publication of

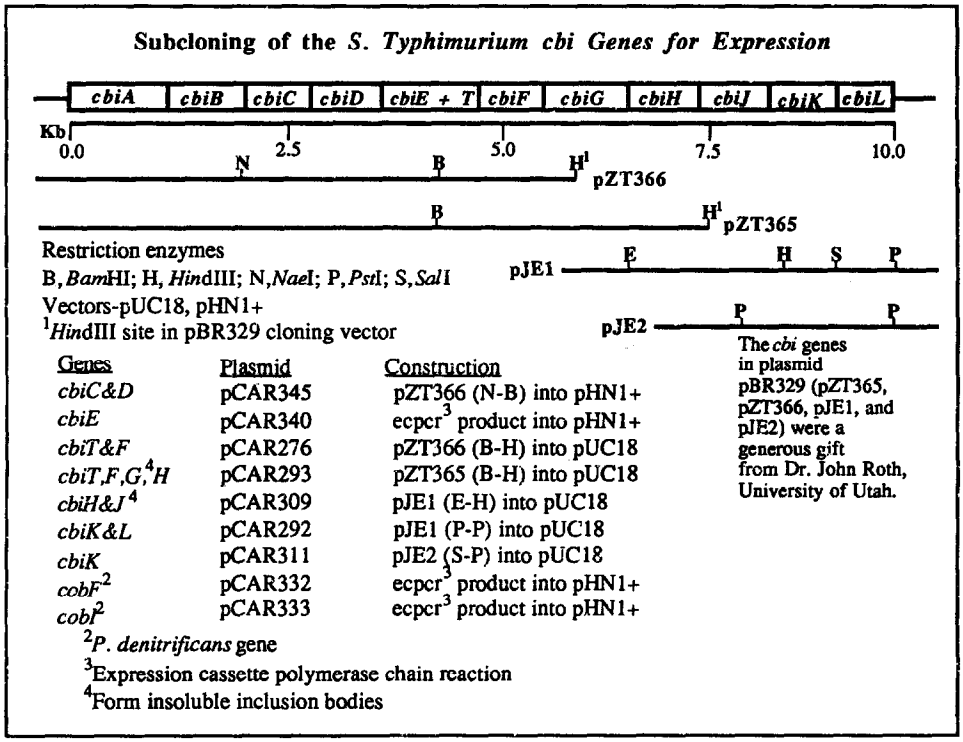


Figure 6

several of the *Pseudomonas denitrificans* DNA sequences encoding corrin synthesis by a French group.[37] In our work ORF 12 (*cbi*L) has been identified as precorrin-3 synthase (M-2).[38] The multi-enzyme synthesis of this complex product from the building block ALA has been accomplished as shown in Figure 7 both in the NMR tube and preparatively (on the multi-milligram scale) by adding the overexpressed enzymes to substrate ALA in the presence of SAM.

Precorrin-3 indeed turned out to be a dipyrrocorphin with the constitution shown. However subtle differences were observed in the ^{13}C-NMR spectrum which reflect the presence of the new methyl group at C-20 and its effect on the conjugated system, resulting in a preponderance of the tautomer shown in Figure 7.

Precorrin 3 was next incubated in turn with each of the putative, overexpressed methyltransferases and SAM. A surprising result was obtained. The only methylation observed was that catalyzed by ORF-7 and turned out to be α-methylation at C-11! The new isolate, identified as the kinetic product by direct NMR observation is a modified corphin bearing a **fourth** methyl group at C-11. Since biochemical conversion to

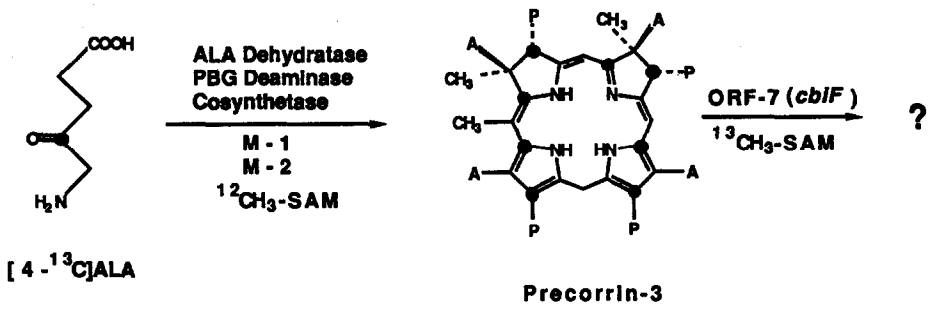
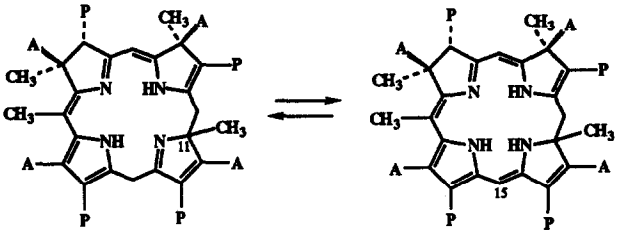


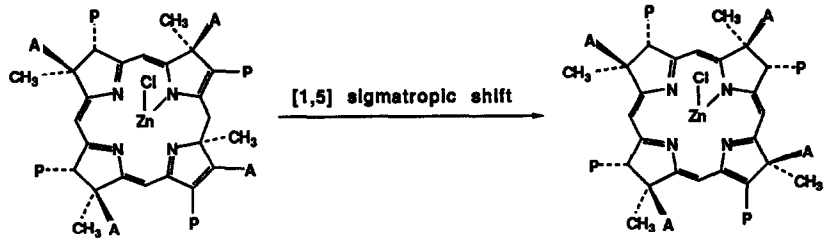
Figure 7 Multienzyme Synthesis of Precorrins

cobyrinic acid has not yet been demonstrated, we have named the new product **4x**.[36] This structural variant and its tautomer **4y** recalls one of the 4 possible structures proposed for Factor S_3.[29,30] Reexamination of the



Compound 4x **4y**

spectroscopic data for this metabolite isolated from *P. shermanii*[29,30] suggests that the correct isomer is in fact the one in which 11 (α-) methylation has taken place on the uro'gen I template and may well be connected to Factor S_1 by [1,5] sigmatropic shift as shown in Figure 8. Also the discovery of precorrin **6x** (Figure 9) by Thibaut *et al*[39] in which C-methylation at C-11, in addition to ring contraction and oxidation at



Factor S_3 Factor S_1

Figure 8

C-18, -19 is evident, requires that the methyl group finally resident at C-12 has migrated from C-11. Thus ORF-7 (*cbi*F; Figure 6) from *S. typhimurium* encodes the methyltransferase enzyme responsible for insertion

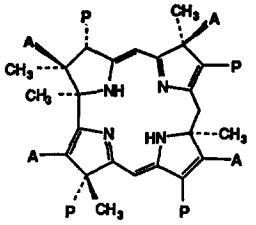


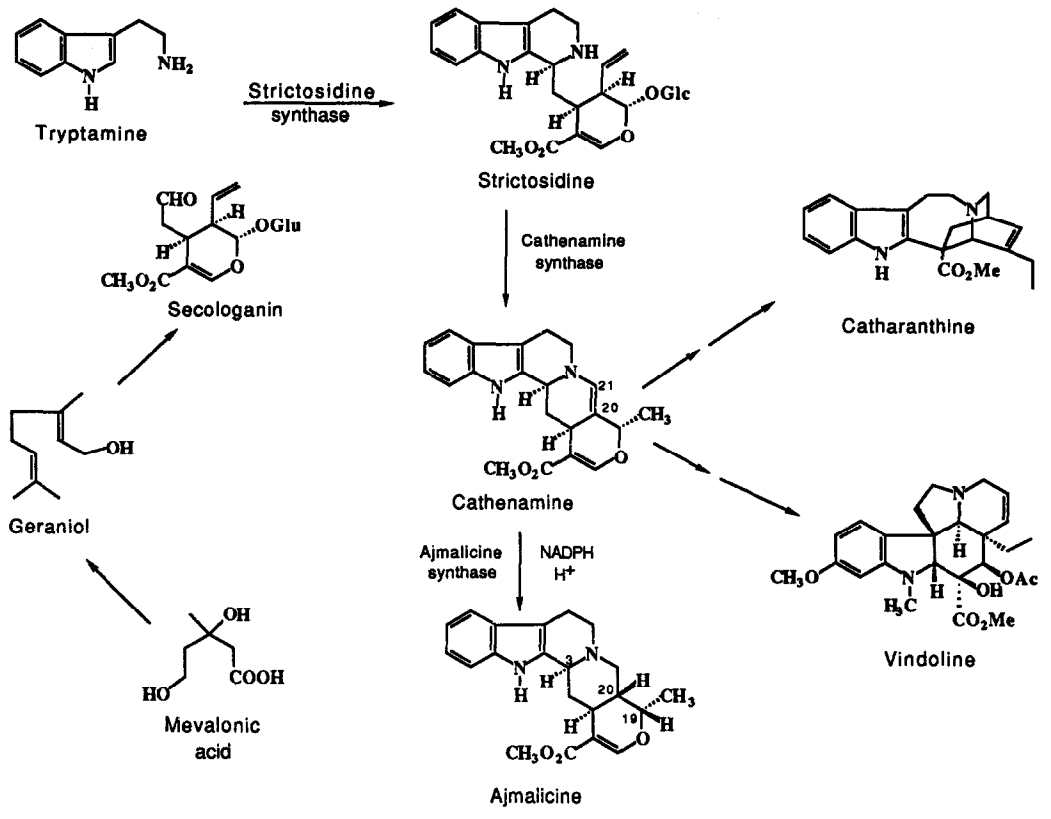
Figure 9 Precorrin 6x

of a methyl group at C-11. We are now at the most exciting stage of the discovery of the rest of the biosynthetic pathway to B_{12} and many of the earlier ideas about the sequence are being modified. Since all of the enzymes have been expressed it now remains to uncover the biosynthetic function of each enzyme. The pathway in *Salmonella* and *P. shermanii* is anaerobic, thus the chosen method of assay is ^{13}C-NMR spectroscopy, using ^{13}C enriched substrate, at <5 ppm of O_2. Although challenging, this logical attack on the problem of corrin biosynthesis is clearly bearing fruit and the prognosis for a successful conclusion is now in the five rather than twenty year time scale.

Transfer of Plant Genes and Expression in Bacteria

The second topic of this review covers very 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.[40] 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[41] which catalyzes the Pictet-Spengler type condensation of the monoterpenoid glucoside secologanin and tryptamine (Scheme 7). The 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[42] 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*.[43] This gene has also been transformed into tobacco plants where it has been shown to express active strictosidine synthase.[44] The overexpression of active *C. roseus* strictosidine synthase in *E. coli* was achieved[45] using the expression cassette polymerase chain reaction (ECPCR) technique[46] and led to the capability of purifying the enzyme in quantities exceeding those previously attainable.



Scheme 7

Expression of strictosidine synthase in *E. coli*

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.[43]

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*.[42]

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.[42] 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 complex 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.

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

The work described in this lecture could not have been possible without the skilled technical achievements of my younger colleagues whose names will be found in the references, and generous financial support from the N.I.H. and Robert A. Welch Foundation. Our long standing and stimulating collaboration with Professor G. Müller (Stuttgart) was aided by a grant from N.A.T.O.

References

1. Scott, A. I. *Acc. Chem. Res.* **1978**, *11*, 29; *idem ibid* **1990**, *23*, 308.
2. Burton, G.; Fagerness, P. E.; Hosazawa, S.; Jordan, P. M.; Scott, A. I. *J. Chem. Soc. Chem. Comm.* **1979**, 202.
3. Battersby, A. R.; Fookes, C. J. R.; Gustafson-Potter, K. E.; Matcham, G. W. J.; McDonald, E. *J. Chem. Soc. Chem. Comm.* **1979**, 1115.
4. Jordan, P. M.; Burton, G.; Nordlov, H.; Schneider, M. M.; Pryde, L. M.; Scott, A. I. *J. Chem. Soc. Chem. Comm.* **1979**, 204.
5. Reviews, Leeper, F. J.; *Nat. Prod. Reports* **1985**, *2*, 19 and 561; **1987**, *4*, 441; **1989**, *6*, 171.
6. Scott, A. I.; *J. Heterocyc. Chem.* **1987**, *14*, S-75.
7. Scott, A. I.; Roessner, C. A.; Stolowich, N. J.; Karuso, P.; Williams, H. J.; Grant., S. K.; Gonzalez, M. D.; Hoshino, T. *Biochemistry* **1988**, *27*, 7984.
8. Anderson, P. M.; Desnick, R. J. *J. Biol. Chem.* **1980**, *255*, 1993.
9. Berry, A.; Jordan, P. M.; Seehra, J. S. *FEBS Lett.* **1981**, *129*, 220.
10. Battersby, A. R.; Fookes, C. J. R.; Hart, G.; Matcham, G. W. J.; Pandey, P. S. J. *J. Chem. Soc. Perkin Trans. I* **1983**, 3041.
11. Jordan, P. M.; Berry, A. *Biochem. J.* **1981**, *195*, 177.
12. Evans, J. N. S.; Burton, G.; Fagerness, P. E.; Mackenzie, N. E.; Scott, A. I. *Biochemistry* **1986**, *25*, 905.
13. Scott, A. I.; Roessner, C. A.; Clemens, K. R.; Stolowich, N. J.; Santander, P. J.; Gonzalez, M. D. *FEBS Lett.* **1988**, *242*, 319.
14. Scott, A. I.; Stolowich, N. J.; Williams, H. J.; Gonzalez, M. D.; Roessner, C. A.; Grant, S. K.; Pichon, C. *J. Am. Chem. Soc.* **1988**, *110*, 5898.
15. Jordan, P. M.; Warren, M. J.; Williams, H. J.; Stolowich, N. J.; Roessner, C. A.; Grant, S. K.; Scott, A. I. *FEBS Lett.* **1988**, *235*, 189.
16. (a) Warren M. J.; Jordan, P. M. *FEBS Lett.* **1987**, *225*, 87.
    (b) Warren M. J.; Jordan, P. M. *Biochemistry* **1989**, *27*, 9020.
17. (a) Hart, G. J.; Miller, A. D.; Leeper, F. J.; Battersby, A. R. *J. Chem. Soc. Chem. Comm.*, **1987**, 1762.
    (b) Miller, A. D.; Hart, G. J.; Packman, L. C.; Battersby, A. R. *Biochem. J.* **1988**, *254*, 915.
    (c) Hart, G. J.; Miller, A. D.; Battersby, A. R.; *Biochem. J.* **1988**, *252*, 909.
    (d) Beifus, U.; Hart, G. J.; Miller, A. D.; Battersby, A. R. *Tetrahedron Letters* **1988**, *29*, 2591.
18. Radmer R.; Bogorad, L. *Biochemistry*, **1972**, *11*, 904.

19. Scott, A. I.; Roessner, C. A.; Karuso, P.; Stolowich, N. J.; Atshaves B. Manuscript in preparation; Scott, A. I. *Pure & Appl. Chem.*, **1990**, *62*, 1269.
20. (a) Alwan, A. F.; Mgbeje, B. I. A.; Jordan, P.M. *Biochem. J.* **1989**, *264*, 397.
(b) Crockett, N.; Alefounder, P. R.; Battersby, A. R.; Abell, C. *Tetrahedron*, **1991**, *47*, 6003.
21. Rosé, S.; Frydman, R. B.; de los Santos, C.; Sburlati, A.; Valasinas, A.; Frydman, B. *Biochemistry* **1988**, *27*, 4871.
22. (a) Stork, W. M.; Baker, M. G.; Raithly, P. R.; Leeper, F. J.; Battersby, A. R. *J. Chem. Soc. Comm.* **1985**, 1294.
(b) Battersby, A. R.; Fookes, J. R.; Matcham, G. W. J.; McDonald, E.; Hollenstein, R. *J. Chem. Soc., Perkin Trans I* **1983**, 3031.
23. Müller, G. in Vitamin B_{12}, Zagalak B.; Friedrich, W., eds, de Gruyter, New York, **1979**, 279-291, who described a methylase from *P. shermanii* capable of using Uro'gen I as substrate. The product was not characterized.
24. Cameron, B.; Briggs, K.; Pridmore, S.; Brefort, G.; Crouzet, J. *J. Bacteriol.* **1989**, *171*, 547-557; Blanche, F.; Debussche, D.; Thibaut, D.; Crouzet, J.; Cameron, B. *ibid*, 4222.
25. Warren, M. J.; Roessner, C. A.; Santander, P. J.; Scott, A. I. *Biochem. J.* **1990**, *265*, 725-729.
26. Brunt, R. D.; Leeper, F. J.; Orgurina, I.; Battersby, A. R. *J. Chem. Soc., Chem. Comm.* **1989**, 428.
27. Warren, M. J.; Stolowich, N. J.; Santander, P.J.; Roessner, C. A.; Sowa, B. A.; Scott, A. I. *FEBS Lett.* **1990**, *261*, 76-80.
28. (a) Scott, A. I.; Warren, M. J.; Roessner, C. A.; Stolowich, N. J.; Santander, P. J. *J. Chem. Soc., Chem. Comm.*, **1990**, 593.
(b) Warren, M. J.; Gonzalez, M. D.; Williams, H. J.; Stolowich, N. J.; Scott, A. I. *J. Am. Chem. Soc.* **1990**, *112*, 5343..
29. Müller, G.; Schmiedl, J.; Schneider, E.; Sedlmeier, R.; Worner, G.; Scott, A. I.; Williams, H. J.; Santander, P. J.; Stolowich, N. J.; Fagerness, P. E.; Mackenzie, N. E. *J. Am. Chem. Soc.* **1986**, *108*, 7875-7877.
30. Müller, G.; Schmiedl, J.; Savidis, L.; Wirth, G.; Scott, A. I.; Santander, P. J.; Williams, H. J. Stolowich, N. J.; Kriemler, H.-P. *J. Am. Chem. Soc.* **1987**, *109*, 6902-6904.
31. Scott, A. I.; Williams, H. J.; Stolowich, N. J.; Karuso, P.; Gonzalez, M. D.; Blanche, F.; Thibaut, D.; Müller, G.; Wörner, G. *J. Chem. Soc., Chem. Comm.* **1989**, 522.
32. Leumann, C.; Hilpert, K.; Schreiber, J.; Eschenmoser, A. *J. Chem. Soc., Chem. Comm.* **1983**, 1404-1407.
33. Eschenmoser, A. *Angew. Chem., Int. Ed. Engl.* **1988**, *27*, 5.
34. Kurumaya, K.; Okasaki, T.; Kajiwara, M. *Chem. Pharm. Bull.* **1989**, *37*, 1151.
35. (a) Jeter, R. M.; Roth, J. R. *J. Bacteriol.* **1987**, *169*, 3189.
(b) Roth, J. R.; Crabau, C.; Doak, T. G. in *"Chemical Aspects of Enzyme Biotechnology: Fundamentals"*, Baldwin, T. O.; Raushel, F. M.; Scott, A. I. Eds., Plenum Publishing Corporation, **1990**, pp 317.
36. Roessner, C. A.; Warren, M. J.; Santander, P. J.; Atshaves, B. P.; Ozaki, S.-I.; Stolowich, N. J.; Iida, K; Scott, A.I., submitted for publication.

37. (a) Thibaut, D.; Corder, M.; Crouzet. J.; Debussche, L.; Cameron, B.; Blanche, F. *J. Bacteriol*, **1990**, *172*, 6245 and references cited therein.
(b) Crouzet, J.; Blanche, F.; Cameron, B.; Thibaut, D.; Debussche, L. in *"Chemical Aspects of Enzyme Biotechnology: Fundamentals"*, Baldwin, T. O.; Raushel, F. M.; Scott, A. I. Eds., Plenum Publishing Corporation, **1990**, pp 299.
38. Warren, M. J.; Roessner, C. A.; Ozaki, S.-I.; Stolowich, N. J.; Santander, P.J.; Scott, A. I. *Biochemistry* **1991**, in press.
39. (a) Thibaut, D.; Debussche, L.; Blanche, F. *Proc. Nat. Acad. Sci. (USA)* **1990**, *87*, 8795.
(a) Thibaut, D.; Blanche, F.; Debussche, L.; Leeper, F. J.; Battersby, A. R. *Proc. Nat. Acad. Sci. (USA)* **1990**, *87*, 8800.
40. Review: Scott A. I. *Acc. Chem. Res.* **1970**, *3*, 151.
41. Mizukami, H.; Nordlov, H.; Lee, S. L.; Scott, A. I. *Biochemistry* **1979**, *18*, 3760.
42. Review: Kutchan, T. M.; Dittrich, H.; Bracher, D.; Zenk, M. H. *Tetrahedron* **1991**, *31*, 5945.
43. McKnight, T. D.; Roessner, C. A.; Devagupta, R.; Scott, A. I.; Nessler, C. L. *Nuc. Acids Res.* **1990**, *18*, 4939.
44. McKnight, T. D.; Berger, D. R.; Burnett, R. J.; Nessler, C. L. *Planta Medica* **1991**, in press.
45. Roessner, C. A.; Devagupta, R.; Hasan, M.; Williams, H. J.; Scott, A. I. *Protein Expression & Purification* **1991**, in press.
46. MacFerrin, K. D.; Terrauera, M.; Schreiber, S. L.; Verdeine, G. *Proc. Nat. Acad. Sci. (USA)* **1990**, *87*, 1937.