Pergamon

Tetrahedron, Vol. 50, No. 47, pp. 13315-13333, 1994
Elsevier Science Ltd
Printed in Great Britain
0040-4020/94 $7.00+0.00

0040-4020(94)00810-8

TETRAHEDRON PERSPECTIVE NUMBER 3

THE DISCOVERY OF NATURE'S PATHWAY TO VITAMIN B_{12}. A 25 YEAR ODYSSEY

A. IAN SCOTT
Center for Biological NMR, Department of Chemistry, Texas A&M University,
College Station, Texas 77843, U.S.A.

Abstract—The chronology of the discoveries along the pathway of vitamin B_{12} biosynthesis is reviewed from a personal perspective, including discussion of the most recent, exciting results and a prognosis for the future of natural product biosynthesis.

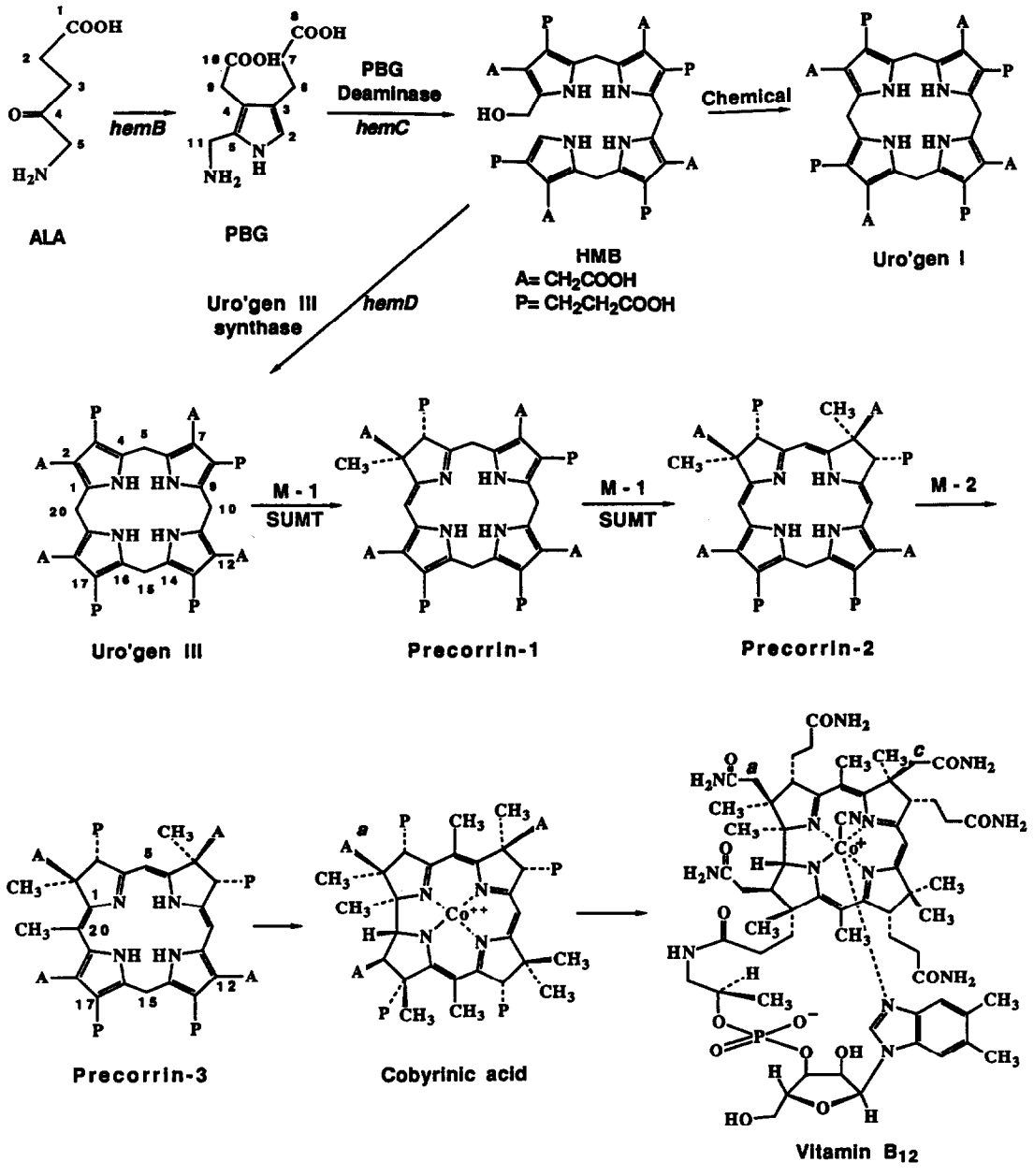
INTRODUCTION

When it was suggested that I might write a *Perspective* describing our research into how nature synthesizes B_{12}, my first reaction was to say that it was perhaps too early, for although one route to the vitamin (unexpectedly involving oxygen) has recently become clear, there is still much work to do in order to solve the mysteries of a second, anaerobic pathway which turns out to be not only different from, but even more fascinating, than the first. But the opportunity to convey both the excitement of the chase just concluded and my optimism for the future, which holds promise of even more surprises in store, was so compelling that I decided to set forth my perspective on natural product biosynthesis, with B_{12} as the central theme.

We begin with a chronological summary of the discoveries along the pathway to corrins.[1,2] The intersection of our researches with those of the other major laboratories working in the area are clearly referenced, not so much from the standpoint of priority of publication—the technical aspects of B_{12} research are sufficiently demanding and complex that time-lags of months between publications from rival laboratories should be regarded as indicators of parallel progess, often involving quite different scientific and philosophical approaches—but rather as testimony to the ultimate convergence of agreement on how B_{12} is made by nature, step by step.

THE EARLY DAYS

From my first encounter with the beautiful structure of vitamin B_{12}, and the monumental total synthesis announced by Woodward[3] and Eschenmoser,[4] the study of vitamin B_{12} biosynthesis became my major interest while at Sussex University. At that time (1966) the only information about the pathway to corrin was Shemin's observation that 5-aminolevulinic acid (ALA) was the source of the tetrapyrrolic template and that methionine supplied six (or was it seven?) of the C-methyl groups, the exact number being difficult to estimate by radiolabeling, the only method available at the time. The involvement of ALA indicated that the first segment of the well-known pathway to the tetrapyrroles of nature was operative, involving first the dehydratase converting 2 mol of ALA to porphobilinogen (PBG) and thence via the combined actions of deaminase and uro'gen III synthase to uro'gen III (Scheme 1), the unsymmetrical tetrapyrrolic macrocycle, already known to be a precursor of heme and chlorophyll.

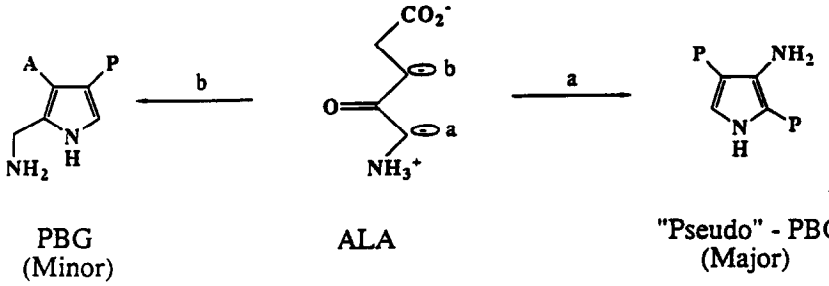


Scheme 1.

Our early experiments were entirely biomimetic, for as organic chemists we were intrigued by the control which nature exerts over the formation of the unique pyrrole PBG, bearing the acetate

(A) and propionate (P) side chains which somehow have to be 'switched' from head to tail sequence at a later stage in the macrocyle uro'gen III.

After much experimentation, we eventually succeeded[5] in simulating the synthesis of PBG from ALA but in very low (10%) yield, the major product being the isomeric pyrrole, 'pseudo'-PBG (Scheme 2), corresponding to the preferred anion (a) of ALA. This small measure of control was achieved by the use of an ion exchange resin which served to bind the substrate in a configuration favorable for the 'correct' dimerization, but the solution of this key problem still remains a synthetic challenge to which we must return one day. Nevertheless, we had taken the first small step towards understanding tetrapyrrole biosynthesis. On arrival at Yale in 1968, the cultivation of B_{12}-producing bacteria began in earnest and we introduced fermentation of *Propionibacterium shermanii* in the Sterling Laboratory—perhaps the first and last biological experiments carried out there. With M. Kajiwara, K. Okada and C. A. Townsend, having learned how to produce 5 mg of B_{12} per liter of culture, we were now ready to try the first feeding experiments with ^{14}C labeled precursors. We also realized that the analytical problems in finding out where the ^{14}C label finished up in B_{12} were enormous, since the degradation chemistry of the vitamin had been left in its infancy when the late Dorothy Hodgkin solved the structure by X-ray crystallography.[1,2] So, as happened so often in the next 25 years, we decided to use the most recent and powerful tools at hand, and luckily chose ^{13}C NMR spectroscopy, although our only equipment at that time was a Bruker FT-90 instrument with 'home-made' FT accessories in Yale Medical School, where our colleague Robert Cushley was able to coax the vital signals above natural abundance to show enrichment when ^{13}C-ALA was fed to whole cells of *P. shermanii*. We had now entered a new world—the combination of NMR and biochemistry to solve not only the nature of the building blocks involved in the architecture of corrin but also their sequence of assembly. In this way we found that B_{12} is indeed derived from PBG, and most importantly, uro'gen III, (Scheme 1) in spite of earlier literature reports to the contrary. Also we could count the number of methyl groups (seven) derived from SAM with confidence, thereby dispelling any previous notion that the C-1 methyl came from a cyclopropane intermediate associated with a preceding ring contraction—yet another attractive concept which had to be abandoned in the face of experimental evidence.



PBG (Minor) ALA "Pseudo" - PBG (Major)

Scheme 2.

These early NMR results were, for us, spectacular—no more tedious degradations were necessary to find the site and extent of radiolabeling—once the ^{13}C NMR spectrum of B_{12} was assigned using the sets of ^{13}C signals from each of the ^{13}C isotopomers of ALA as they found their way (by a still unknown intermediary metabolism) into the vitamin.[6,7]

THE FIRST WAVE OF NEW INTERMEDIATES—FACTORS II, III, AND HMB

The next breakthrough came when Boris Yagen and Eun Lee demonstrated the first cell free system from *P. shermanii* [8] which would incorporate our ^{13}C-enriched building blocks so efficiently that we could label uro'gen III in a unique way and thus determine the fate of each ^{13}C center as it finally appeared in cobyrinic acid—the end product of the cell free system. These methods were rapidly adopted elsewhere so that it was no surprise when my colleague, Gerhard Müller in Stuttgart, by withholding cobalt from the cell free system, isolated the new intermediate, factor II (oxidized precorrin-2; Scheme 1), which remarkably corresponded with the structure of sirohydrochlorin, just solved by Tony Irwin[9] in a joint project with Lewis Siegel (at Duke) and this tied in with the corriphyrins discovered by Bykhovsky[10] in Moscow and sent to Cambridge for structural elucidation. We will return later to the roles played by sirohydrochlorin (factor II; oxidized precorrin-2) and the *cys* genes responsible for cysteine biosynthesis in *Escherichia coli*. By now, in the late 1970s, we realized that in order to solve the complexities of B_{12} biosynthesis we would have to build a new laboratory combining organic chemistry, enzymology, molecular biology and NMR spectroscopy.

This dream came true in the form of an invitation to set up a major interdisciplinary program combining all of these techniques in one department—hence our move to Texas A&M in 1977. There, with Peter Jordan and Gerardo Burton, again using NMR spectroscopy, we first discovered[11] the enzyme-free product HMB (Scheme 1) discharged from the enzyme PBG deaminase which became the substrate for the rearranging enzyme, uro'gen III synthase, and with Tony Irwin and G. Müller[12] we were able to solve the structure of precorrin-3 (in its oxidized form, factor III), which provided the first clue that nature inserts a C-methyl at C-20 in precorrin-2 (Scheme 1) only to be subsequently lost as a two carbon species, during (or after) ring contraction, which was later found to be acetic acid.[12–14]

THE LONG WAIT, AND A CHANGE IN TACTICS

Now came the long wait (1979–1990) where no new intermediates emerged, in spite of processing hundreds of liters of B_{12}-producing lysates incubated with (and without) cobalt. What did emerge, however, was a series of 'factors', S_1–S_4, found by Gerhard Müller whose structures determined by NMR in ^{13}C-enriched form using the various ALA (^{13}C)- isotopomers, turned out to be tetramethyl derivatives of uro'gen I.[15,16] Thus the methyl transferase enzymes of *P. shermanii* are able to use the 'wrong' substrate, the physiologically inactive but omnipresent uro'gen I, the symmetrical isomer of uro'gen III (Scheme 1), to carry out the same type of methylation used in B_{12} synthesis. These biochemical curiosities are authentic natural products, since they are found in fermentations leading to B_{12}.

During this period we also learned by using pulse-labeling techniques developed earlier in Cambridge[20] that the order of insertion of the C-methyl groups was $C_2 > C_7 > C_{20} > C_{17} > C_{12} > C_1 > C_5 > C_{15}$.[17–21] Although obtaining a 'distant' view of what kinds of intermediates must be involved, it still proved impossible to isolate any of these species corresponding to the methylation sequence, (which is conserved in *P. shermanii*, *Clostridium tetanomorphum*, and *Pseudomonas denitrificans*) perhaps due to their instability towards air and/or low concentration. Realizing that no further intermediates were going to be discovered by the 'needle in a haystack' approach, we began in 1987 to assemble the entire repertoire of biosynthetic gene products for B_{12} biosynthesis. Our approach was to differ radically in strategy from work proceeding simultaneously at Genetica (a subsidiary of Rhone–Poulenc) in Paris where a team under A. Rambach had begun to identify the gene clusters and sequences containing whole operons for the B_{12} pathway. Our paths (and swords) were to cross later, but at the outset, having no access to industrial strains of B_{12} producers or their

engineered versions, which could be amplified by sets of gene clusters, we started with the heme box of *E. coli* (Figure1). Here the genetics were well studied, and although *E. coli* does not make B_{12}, it has to carry out the synthesis of precorrin-2 necessary for the source of the iron-containing cofactor, siroheme, as part of the machinery of the giant enzyme sulfite reductase, our earlier source of factor II.

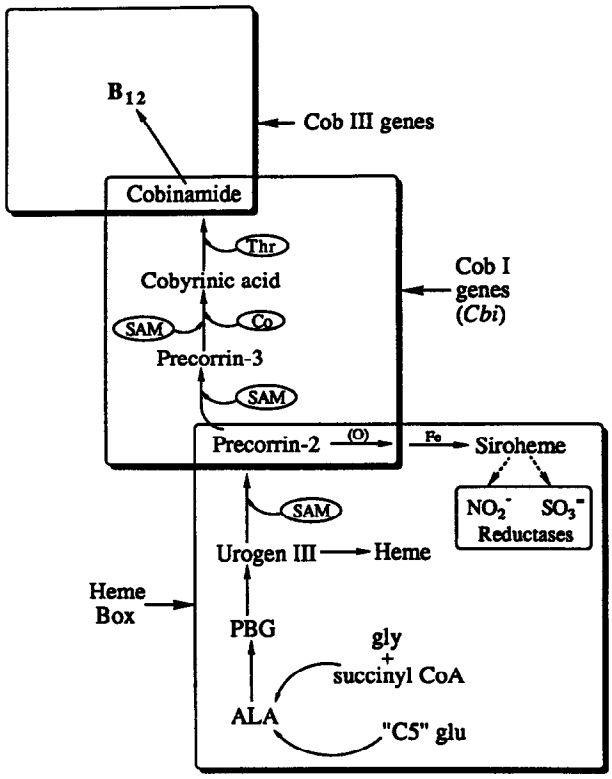


Figure 1. The B_{12} biosynthetic genes showing the heme box of *E. coli* and the *cbi* genes of *Salmonella typhimurium*.

Fortunately the *hem* gene cluster of *E. coli* had already been cloned and through the generosity of molecular biologists worldwide, including Susan Cosloy, A. Sasserman, Peter Jordan and Jeff Cole, we could assemble and overexpress the genes *hemB*, *C* and *D* as well as *cysG*, (M-1; Scheme 1) which, as we suspected, had the necessary methyl transferase activity to convert uro'gen III into precorrin-2. With ample supplies of these biosynthetic enzymes in hand[22] we were next able to study the mechanism of PBG deaminase (*hemC*) and found that the polymerizing enzyme of tetrapyrrole biosynthesis harbored a remarkable dipyrromethane cofactor, made from 2 mol of its own substrate PBG, which was covalently anchored to a cysteine residue (Cys 242) in the 33 kD protein (Figure 2). Again, the ^{13}C NMR technique was, in my view, the unique way to study the assembly of the growing, head to tail, tetrapyrrole chain and we could see each of the new events signaled by a change in the NMR spectrum and later confirmed by electrospray mass

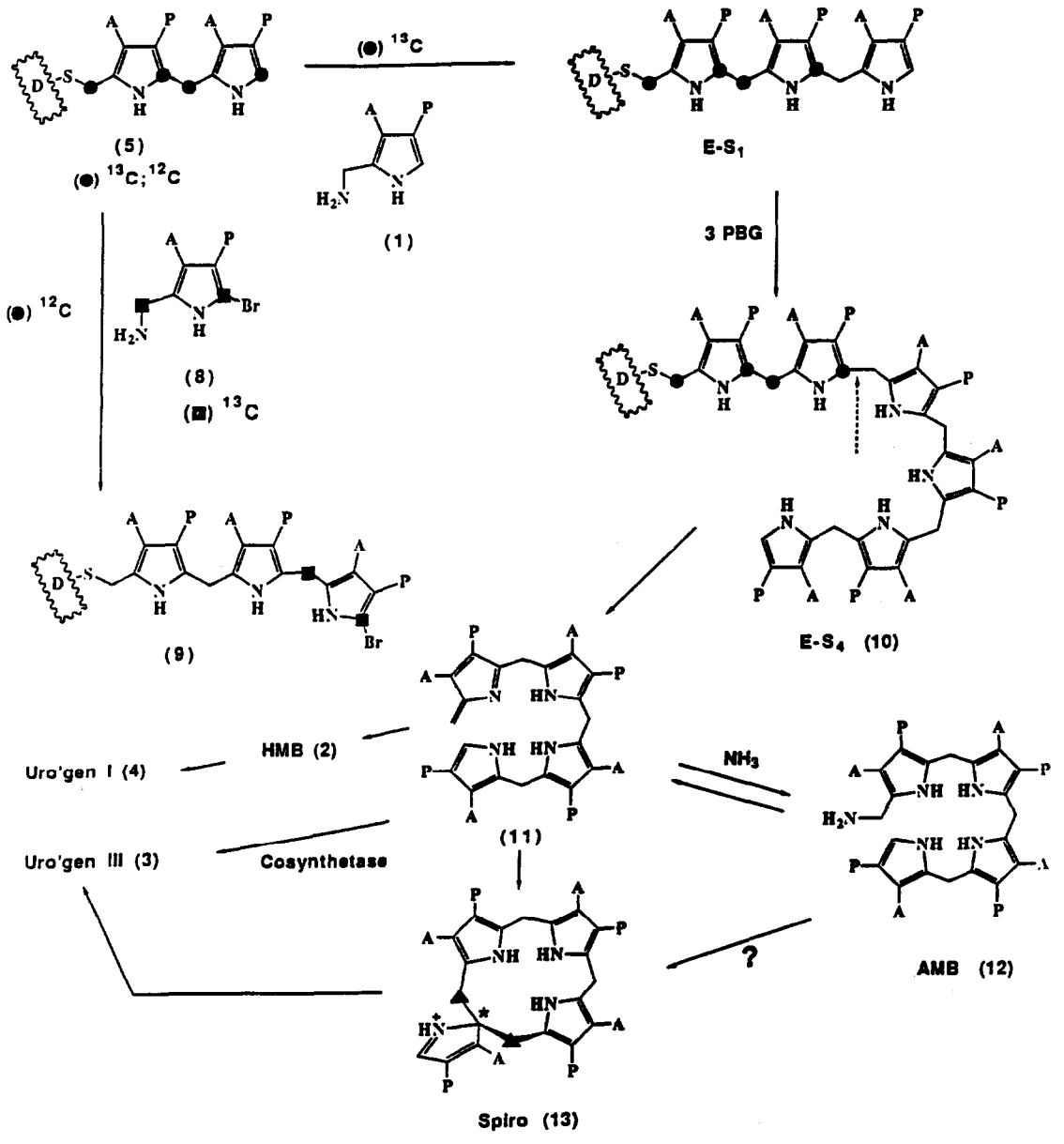
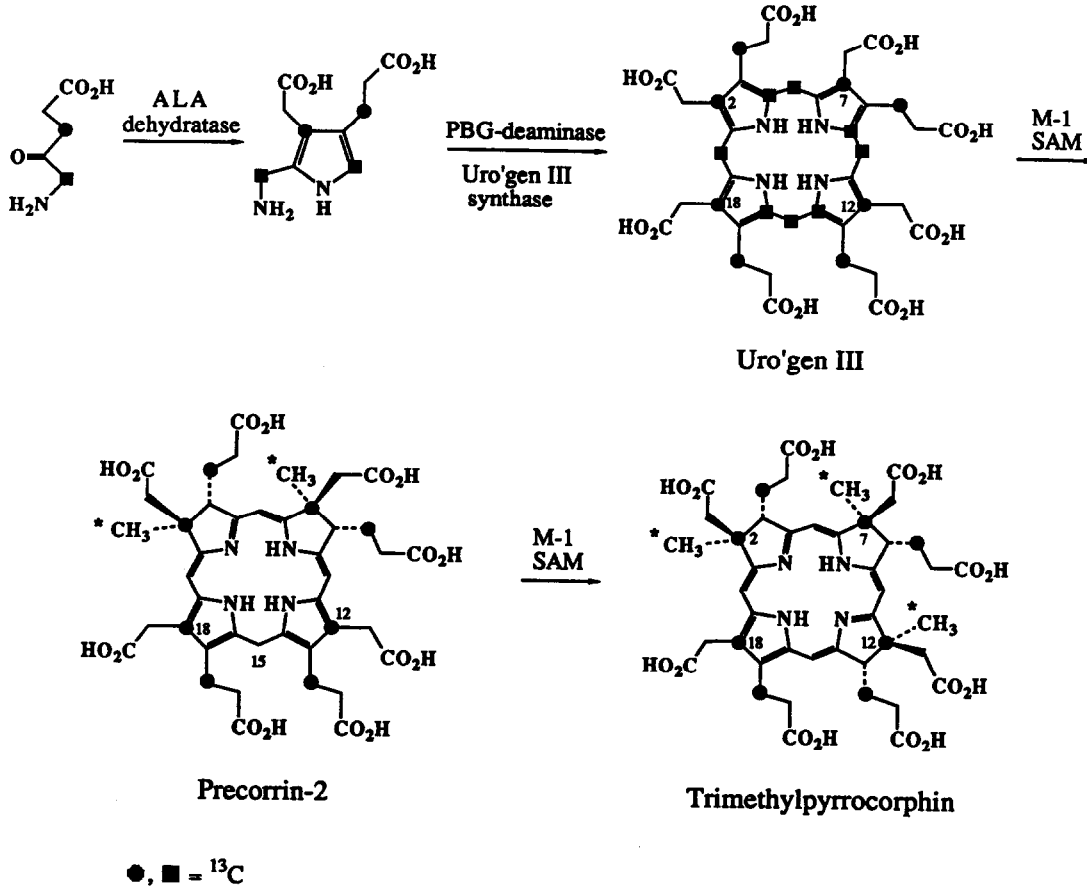


Figure 2. Mechanistic rationale of tetrapyrrole assembly catalyzed by PBG deaminase and uro'gen III synthase. The dipyrromethane cofactor attached to enzyme (**5**) is labeled with 13C(•) and leads to the ES_1–ES_4 complexes. The azafulvene (**11**) is discharged from the ES_4 complex and either becomes the substrate for cosynthetase or can be trapped by NH_3 (or $NaBH_4$). In the absence of cosynthetase, the hydroxymethyl bilane (HMB) accumulates and cyclizes chemically to uro'gen I. Also shown are the putative spiro intermediate (**13**) and the suicide inhibitor bromoPBG (**8**) which forms the inactive complex (**9**).

spectrometry.[23]

As has happened so often in scientific discovery, we were not alone in this search and two groups[24,25] in the U.K. published on the dipyrromethane cofactor within the same five month period. We next found that the apoenzyme reconstitutes the holo-form autocatalytically from the substrate PBG,[23] and when the right number of PBG units (four) has been added to the cofactor, the linear tetrapyrrole is discharged as HMB ready for the next cycle. It was indeed gratifying when our NMR results were confirmed by the 3-D X-ray structure[26] of PBG deaminase.

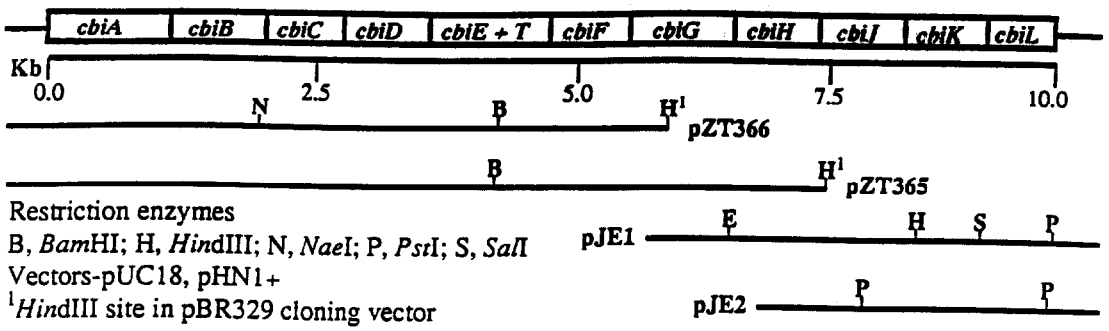
With the capability of making substantial quantities of the pivotal precursor, uro'gen III with overexpressed *hemD* (Scheme 1) now established in our laboratory, we could now move forward into the chiral world of precorrins, one enzyme at a time, beginning with C-methylation by SAM mediated by *cysG* (M-1) which allowed us to study the exact double bond arrangement in precorrin-2.[27] Almost every variant of the arrangement of acetate (A) and propionate (P) side chains was tolerated in the macrocycle when challenged with the methyl transferase enzyme (including the types I and IV urogen). This explained the origins of Müller's factors S_1 and S_3, and we observed that *cysG* could also 'overmethylate' precorrin-2 at the unusual position 12, i.e. out of order from the natural methylation sequence to B_{12}, reactions which, as indicated in Scheme 3, could be studied directly in the NMR tube.[28]



Scheme 3.

THE SEARCH FOR THE BIOSYNTHETIC GENES

At this juncture we undertook yet another major new thrust—the cloning and overexpression of the *Salmonella cbi* genes,[29] so generously provided by John Roth[30] (Utah) who had found that this organism can produce B_{12} anaerobically, and by the end of 1991 had all of the enzymes for corrin synthesis in hand. We had also engineered the necessary genes from a wild strain of *P. denitrificans* into *E. coli* and expressed the biosynthetic gene products, ready to proceed, one enzyme at a time, into the unknown territory beyond precorrin-2 (Figure 3). Both C-20 (*cbiL*) and C-11 (*cbiF*) methyl



Restriction enzymes
B, *Bam*HI; H, *Hin*dIII; N, *Nae*I; P, *Pst*I; S, *Sal*I
Vectors-pUC18, pHN1+
[1]*Hin*dIII site in pBR329 cloning vector

Genes	Plasmid	Construction
cbiC&D	pCAR345	pZT366 (N-B) into pHN1+
cbiE	pCAR340	ecpcr[3] product into pHN1+
cbiT&F	pCAR276	pZT366 (B-H) into pUC18
cbiT,F,G,[4]*H*	pCAR293	pZT365 (B-H) into pUC18
cbiH&J[4]	pCAR309	pJE1 (E-H) into pUC18
cbiK&L	pCAR292	pJE1 (P-P) into pUC18
cbiK	pCAR311	pJE2 (S-P) into pUC18
cobF[2]	pCAR332	ecpcr[3] product into pHN1+
cobI[2]	pCAR333	ecpcr[3] product into pHN1+

[2]*P. denitrificans* gene
[3]Expression cassette polymerase chain reaction
[4]Form insoluble inclusion bodies

Figure 3. Plasmids used to identify and express the open reading frames (ORFs) of the *cbi* genes of *S. typhimurium* encoding B_{12} synthetic enzymes.

transferase activities were found by the ^{13}C NMR assay[29] and using a combination of four enzymes we were able to synthesize factor S_3 as shown in Figure 4.[31] But by now tremendous competition had arisen from the industrial giant, Rhone–Poulenc Santé in Paris, who had absorbed Genetica and had committed a major effort to engineer their production strain of the aerobic organism *P. denitrificans*. Central to these efforts was the amplification of an 8-gene cluster (*cobF–M*), able to synthesize the cobalt free corrinoid, hydrogenobyrinic acid (HBA) from factor III (oxidized precorrin-3). By withholding NADPH or by mutating out several of the *cob* genes, one at a time, it was possible for the French workers to extend the library of intermediates downstream from precorrin-3. They found[32–34] and (in association with the Cambridge group[35–37]) deduced the structures of the ring contracted precorrin-6x, its reduction product, 6y (obtained by restoring

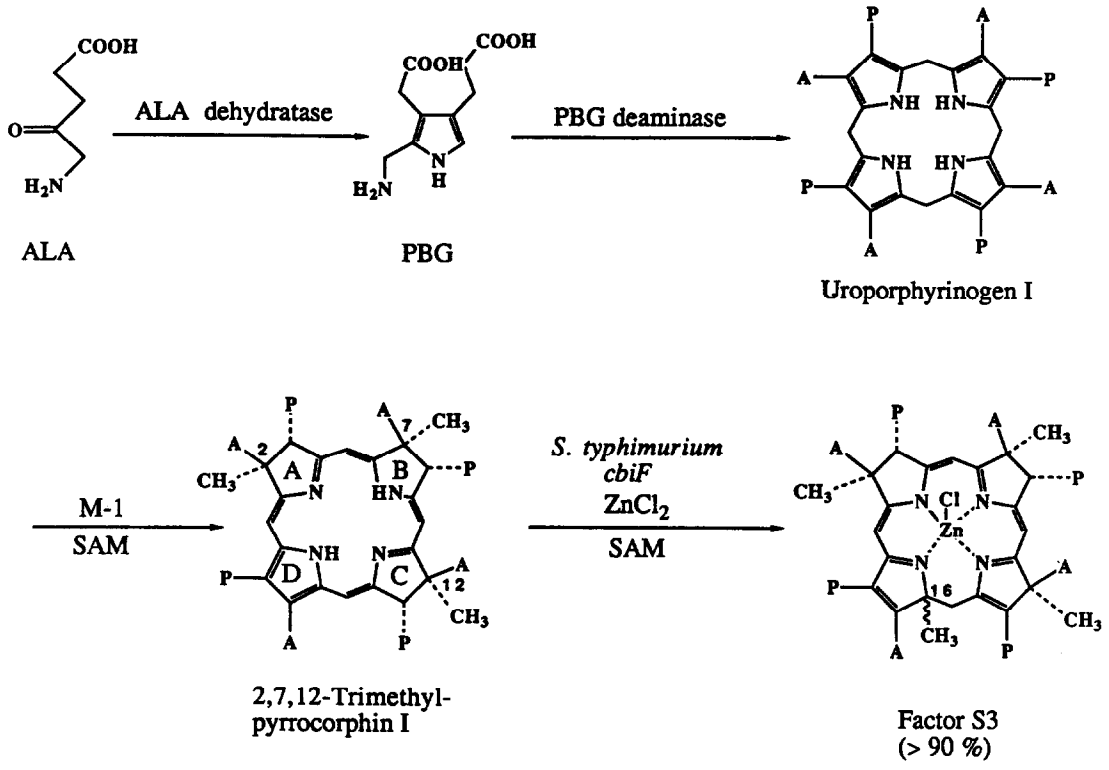


Figure 4. Multienzyme synthesis of factor S_3.

NADPH) and an advanced, fully methylated precorrin-8x, lacking only one step[38] (11→12 methyl shift) to reach corrin (HBA), in a cell free system from the engineered bacterium (Figure 5). Although we were never to have access to their industrial strain, the genes encoding each enzyme in the pathway were available from our 'academic' gene bank and would in principle allow us to proceed, step by step, into the unknown territory beyond precorrin-3.

Still missing from the picture was the remarkable process of ring contraction and at least three intermediates between the dipyrrocorphin structure of precorrin-3, and precorrin-6x, including the unknown tetra- and penta-methylated species, corresponding to precorrins-4 and -5 ,respectively.

ANOTHER LONG WAIT

So, by late 1992 we had the full sets of enzymes for both anaerobic (*Salmonella;* nine enzymes) and aerobic (*P. denitrificans*; eight enzymes) pathways in hand (Figure 6) and felt certain of success, yet in spite of many months, night and day in front of the NMR console, we could see no change in the ^{13}C spectrum of precorrin-3 when challenged with each of the 17 enzymes in turn. We realized that some vital cofactor was missing from our carefully controlled anaerobic incubations carried out in the glove box (< 0.5 ppm O_2). We did find that one of the *Salmonella* genes, *cbiF*, was capable of 'mismethylating' precorrin-3 at the 11-position, i.e. out of the natural order, to yield a tetramethyl pyrrocorphin, compound 4x, whose structure deduced by on-line NMR, our only method of assay,

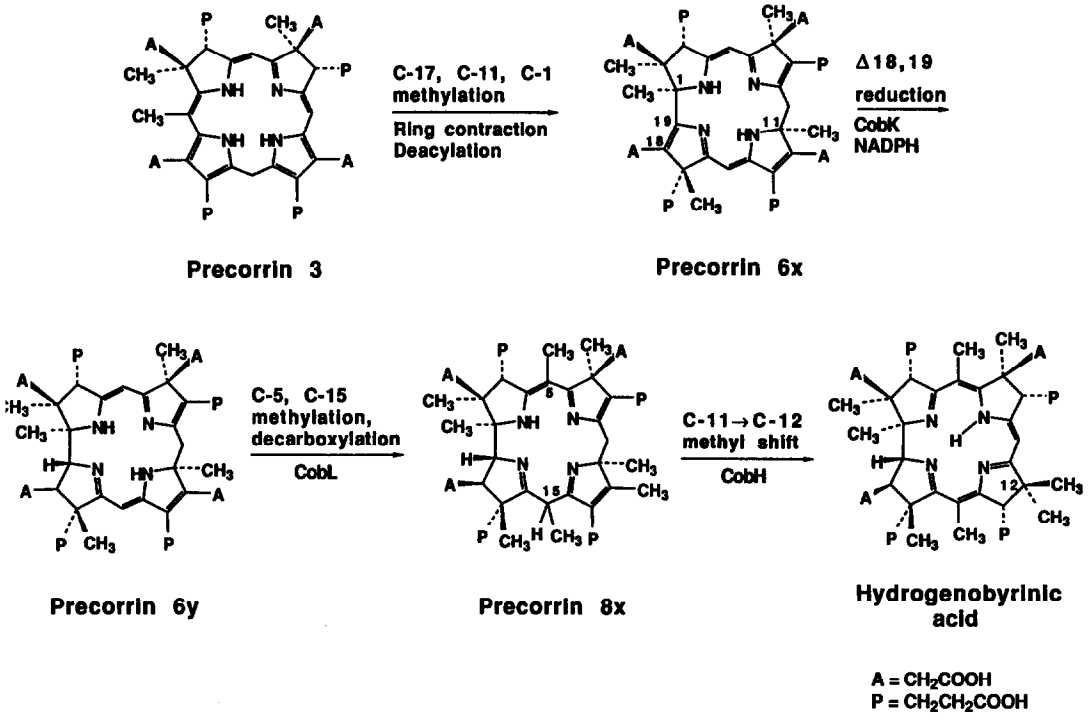


Figure 5. Precorrins-6x, 6y, 8x, recently discovered metal-free precursors of hydrogenobyrinic acid in *P. denitrificans* expression systems.

(Figure 7) was informative about the non-specificity of the methyl transferases but which was not, *per se*, a B_{12} intermediate. We had also learned how to combine as many as five biosynthetic enzymes to prepare substantial amounts of the precursor, precorrin-3 in good yield, endowed with ^{13}C labels ready to explore the ring contraction mechanism[39] (Figure 8). It was a frustrating period until April 1993 when factor IV, the oxidized version of (the still unknown) precorrin-4 from a *cobM* mutant (which could not process the pathway beyond the tetramethylated level) was described by Rhone–Poulenc.[40] From inspection of the structure of factor IV it was at once apparent that the missing cofactor was the very one which the French group had carefully tried to exclude from all of the incubations—oxygen!

Now at last when precorrin-3 was incubated with the prime enzymatic candidate for the ring contraction (*cobG*), this time in the presence of O_2, yet another surprise was in store, for the NMR spectrum changed dramatically, not to a ring contracted product, but to that of a 20-hydroxy lactone, precorrin-3x,[41] formally derived from precorrin-3 by hydroxylation, followed by participation of the ring A acetate carboxyl in γ-lactone formation (Figure 9). All fell quickly into place using $^{18}O_2$ gas to trace these events, as we realized that the function of *cobG* is not to ring contract but rather to install a beautifully designed mechanism for this very purpose, *which only comes into play in the presence of the next methyl transferase.* This turned out to be *cobJ*, a bifunctional enzyme which methylates ring D at C-17 and then (and only then) catalyzes the ring-shrinking process to leave a new methyl ketone function at C-1, corresponding to precorrin-4,[42] which was analyzed in its

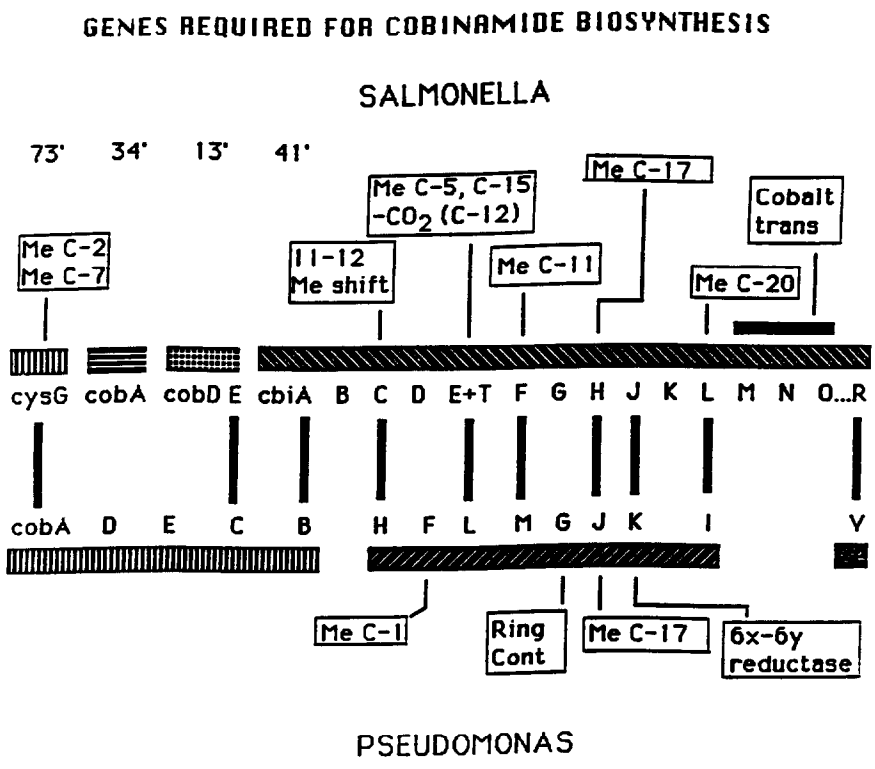


Figure 6. The location and functions of some of the known genes for B_{12} biosynthesis in *S. typhimurium* and *P. denitrificans*. In *S. typhimurium* the genes map at 14' (*cobD*, *cobE*; addition of aminopropanol), 34' (*cobA*; adenosylation), 41' (*cbiA–R*; cobinamide biosynthesis), and 73' (*cysG*; uro'gen III methyltransferase). Homologies between the *S. typhimurium* and *P. denitrificans* gene products are shown. See the text for a discussion of their functions.

biochemically active, reduced form shown in Figure 9. No sooner had our manuscript describing the isolation of precorrins-3x[41a] and 4 been submitted, than the French group also isolated and characterized precorrin-3x and independently decided to name it precorrin-3B.[41b] They had also found factor IV[40] (oxidized precorrin-4) but did not isolate the true intermediate precorrin-4 itself. By the end of 1993 the same group[41c] described factor V which had gained a new methyl at C-11 but unfortunately had lost the acetyl function at C-1 and was not a bio-intermediate. The discovery of the true precorrin-5 is described below. We view the ring contraction process as an acyloin-like rearrangement since we could trace the fate of the original ^{18}O in the C-20 hydroxyl of -3x as it found its way into the carbonyl at C-1 of precorrin-4, by ^{13}C NMR observation of the isotopically shifted resonance of the carbonyl-carbon bearing ^{18}O. By the same isotopic method we showed that a $^{13}C^{18}O_2H$ label in the ring A acetate was retained beyond ring contraction and indeed all the way to HBA,[43,44] in sharp contrast to the fate of this same acetate function, which undergoes exchange with the medium in the anaerobe, *P. shermanii*,[45,46] and realized for the first time that there are two distinct pathways to the complete corrinoid structure of B_{12}, depending on the type of organism, aerobic or anaerobic. The elucidation of the anaerobic pathway has now become our new challenge, since none of the intermediates (or their biosynthetic enzymes) between precorrin-3 and B_{12} in the anaerobes are known, although they are almost certainly cobalt complexes,[47] but that is another story, to be told elsewhere.

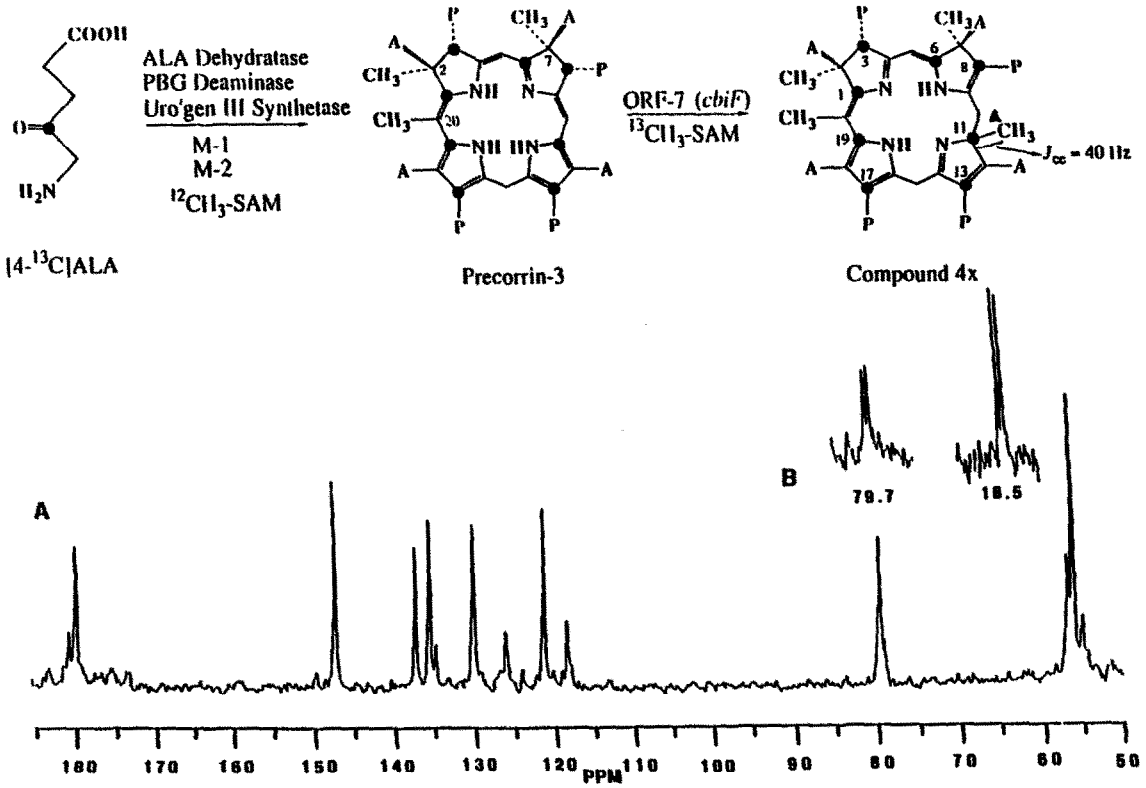


Figure 7. Multienzyme synthesis of compound 4x with the *cbiF* gene product and SAM. The ^{13}C NMR spectrum of 4x (**A**) reveals 3 sp^2 (C-8, C-13 and C-17) but only one sp^2 propionate terminus (C-3) and a signal at δ 79.7 typical of sp^3 carbon (C-11) adjacent to nitrogen. The inset (**B**) shows the coupling (J_{cc} = 40Hz) of the new $^{13}CH_3$ (δ 18.5) to the C-11 signal at δ 79.7.

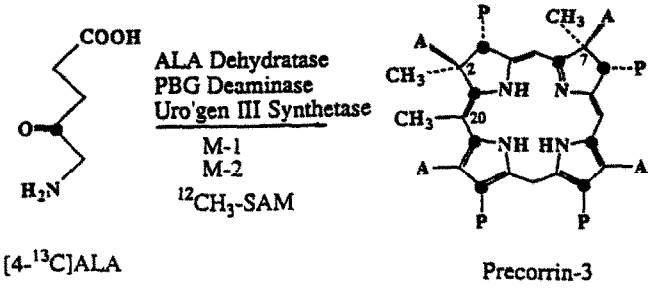


Figure 8.

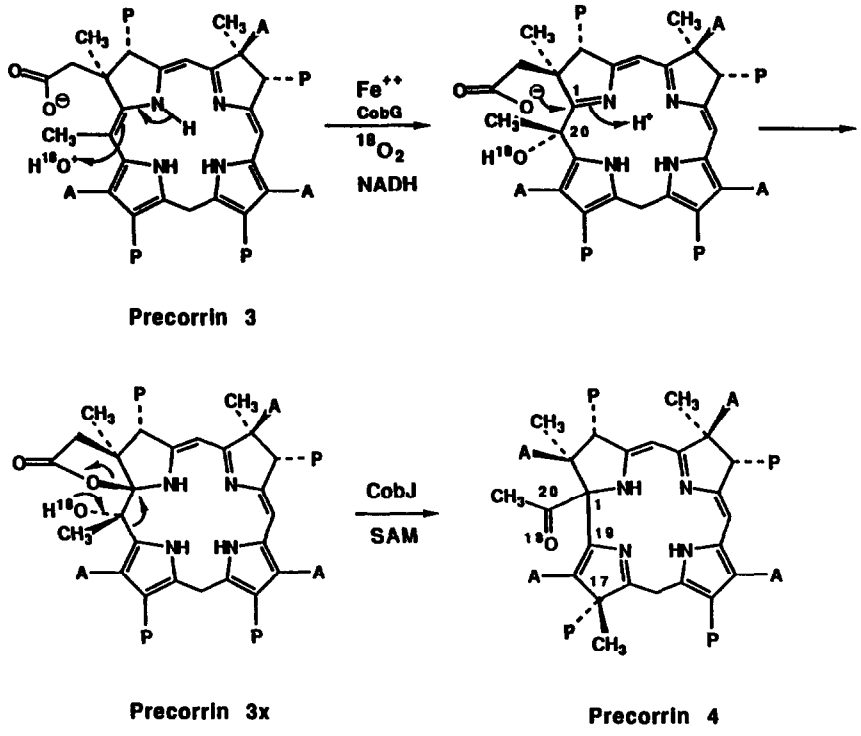


Figure 9. Possible mechanism for the formation of precorrin-3x showing the fate of $^{18}O_2$ in precorrins-3x and 4.

Having reached precorrin-4 and at last beginning to understand nature's method for ring contraction, only one intermediate remained to be discovered and this had to involve yet another methylation, this time at C-11, an event which we had encountered earlier with *cbiF* (Figure 7) but with the wrong substrate! Now everything became clear, for when the gene product, *cobM*, was included in the preparation, we could go directly from precorrin-3 to the long sought precorrin-5,[42] the necessary NMR analysis providing the structure shown in Figure 10. By adding our last unassigned enzyme, *cobF*, we finally reached, by deacylation and methylation at C-1, the known intermediate, precorrin-6x (Figure 11) whose further transformation had been studied genetically[32–34] and chemically[35–37] in Paris and Cambridge, respectively. Now the pathway in all its complexity was revealed—17 discrete steps from ALA mediated by 12 enzymes—a joy to behold!

GENETICALLY ENGINEERED MULTI-ENZYME SYNTHESIS OF CORRIN

As we progressed along the hitherto uncharted road to corrin by the discovery of precorrins-3x, -4 and -5, we became increasingly aware of the potential of multi-enzyme synthesis and could not resist the temptation to try an experiment that would have seemed a pipe dream even five years ago. Why not reconstruct the entire pathway shown in Scheme 4, by incubating ALA with the necessary cofactors SAM, NADPH and of course, oxygen, in a single vessel and look for the final target, HBA? In the spring of 1994 we were ready to assemble the *ex vivo* synthetic machine.

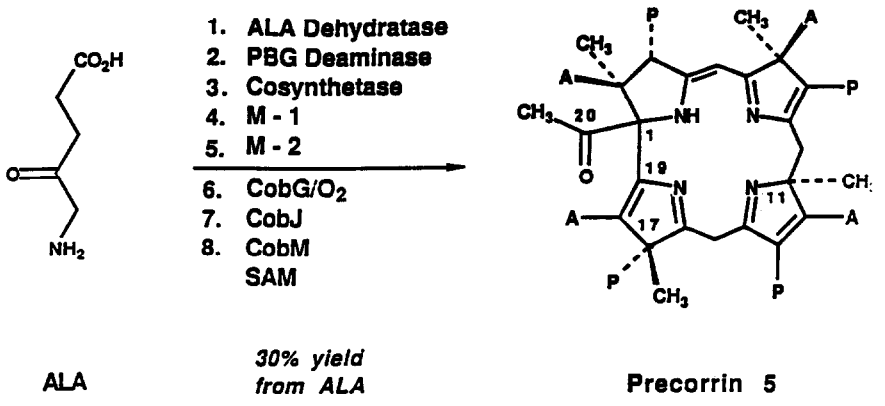


Figure 10. Multienzyme synthesis of precorrin-5.

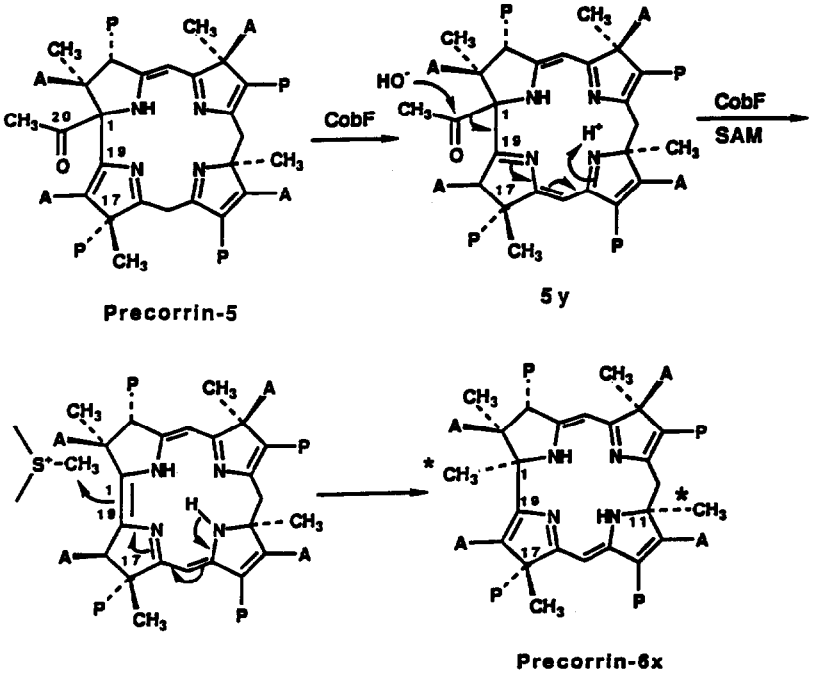
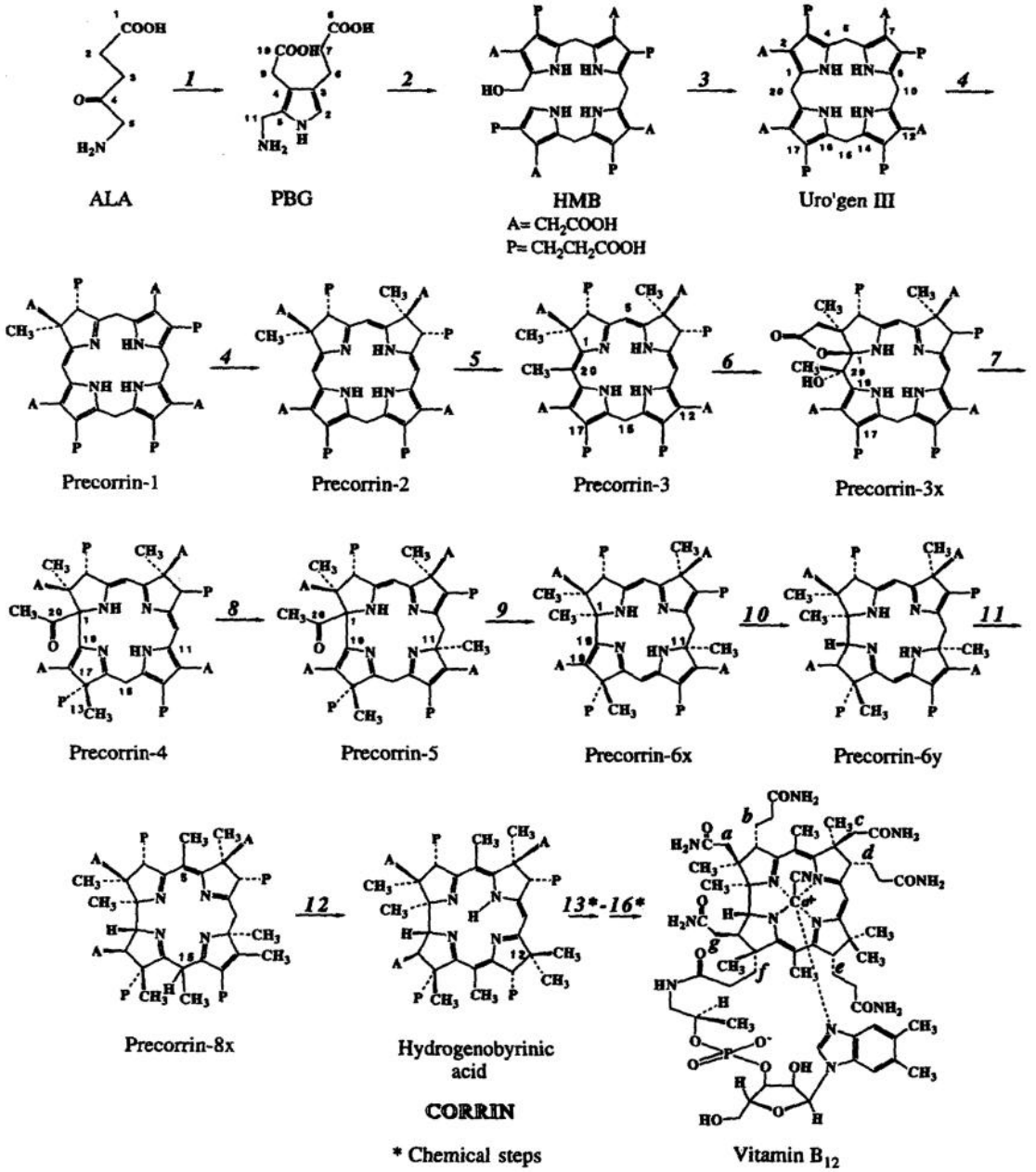


Figure 11. Conversion of precorrin-5 to precorrin-6x catalyzed by *cobF*, showing the tautomeric form 5y.



1. ALA dehydratase (*hemB*); 2. PBG deaminase (*hemC*); 3. Uro'gen III synthase [cosynthase] (*hemD*); 4. Uro'gen III methylase [M-1] (*cysG/cobA*); 5. M-2 (*cobI, cbiL*); 6. Precorrin-3x synthase (*cobG*); 7. Ring contractase/17 methyl transferase [M-3] (*cobJ*); 8. M-4 (*cobM*); 9. M-5 (*cobF*); 10. Reductase (*cobK*); 11. Precorrin-8x synthase (M-6/decarboxylase) (*cobL*); 12. [1,5]-sigmatropic shiftase [hydrogenobyrinic acid synthase] (*cobH*); 13. Insert Co; 14. Esterify; 15. Add nucleotide; 16. Ammonolysis

Scheme 4.

Simply by growing 12, one liter batches of *E. coli* bearing three *hem* and nine *cob* genes, combining the lysates from each with ALA and the above cofactors in buffer (pH7) and, after 12 hours, isolating the corrinoid product, to our delight an overall yield of 20% of HBA from ALA could be reproduced even without trying to optimize the conditions.[48] Complete identity with natural material[19] was established by all possible criteria—^1H, ^{13}C NMR, FAB-MS, UV-Vis, CD and finally by chemical conversion[19] to cobester (Scheme 4; step 13). We had now reached the link with B_{12} itself since the cyanomethyl cobester had already been converted by Eschenmoser[4b] to the complete structure of the vitamin in a remarkable self-assembly process. (Scheme 4, steps 14–16). We believe that, most fortunately for the success of the 12-enzyme, 17-step synthesis of HBA, the last of these enzymes, *cobH*, has such a remarkable affinity for HBA, that it acts like a purification adsorbant and the product can be recovered from protein after simple heat treatment (70°C/10 min), uncontaminated with any other porphyrinoid impurity.

So our saga had come full circle and once again we had returned to synthetic chemistry, but now using the three-dimensional catalysts of nature, and instead of the years required for chemical synthesis, the time for such a genetically engineered synthesis is measured in hours. This is of course an over-simplification, since many person-years had first to be spent in finding and expressing the required biosynthetic genes. It is important at this stage for the general reader to understand the difference between combining several (or many) pure enzymes dedicated solely to reaching a synthetic target in a single flask[49,50] and the construction of genetically engineered strains of bacteria harboring the genes for a biosynthetic pathway, which can either be mutated to accumulate intermediates or used to produce the final target, since the lysates contain the complete metabolic machinery of the cell and can divert the substrates into the major arteries of primary metabolism rather than processing the substrate directly to the target.

WHAT OF THE FUTURE?

Now that the techniques are established, the approach of genetically engineered synthesis should be a general one, for we believe that the most rewarding part of this excursion from synthesis into molecular biology, and back again, is still to come. What if the methods developed for B_{12} synthesis could be applied in a general way to all natural products, regardless of origin? In other words, can the concept be extended to plant, fungal and even human metabolites of potential value as chemotherapeutic agents and exploited to prepare rare natural products with interesting biological activities?

As we enter this most exciting phase of the work, I am certain that the answer must be yes, for already we and others have been able to express the genes for alkaloid synthesis in *E. coli* and at this juncture the literature is beginning to show the first signs of success.[49] Already heterologous gene transfer using cDNA rather than genomic libraries is starting to uncover the activity of biosynthetic enzymes for secondary metabolites. The central problem, still to be solved, is an analytical one. In other words, in a cloned cDNA library, obtained from the action of reverse transcriptase on the isolated mRNA, the biosynthetic factory is usually spread over some 40,000–50,000 plaques, of which only ~20–30 harbor the genes necessary for the synthesis of the target. The breakthrough must come when sensitive assays can be devised to pull the 20 'needles' out of the haystack. When these biosynthetic genes are recombined we will be able to reconstruct the pathway just as was done for B_{12}. We are living in exciting times, since the barrier between chemistry and biology has by now almost disappeared, and from my perspective as a bioorganic chemist, the possibilities presented by combining genetic engineering with organic chemistry are almost limitless. In spite of our recent success with harnessing the 'gene machine' to synthesize B_{12}, I still feel humble in the face of nature's synthetic route, for, just as noted nearly 500 years ago,[51] "Human subtlety will never

devise an invention more beautiful, more simple or more direct than does nature, because in her inventions nothing is lacking and nothing is superfluous."

It has been my good fortune throughout my career to have had many generations of students, postdoctorals and technical assistants whose hard work and perseverance reduced so many of my ideas into practicality. The names of all of those who have given so much over the last 28 years to overseeing and performing the daily (and nightly) experiments involving cloning, enzymology, organic chemistry and NMR spectroscopy, which are now our standard tools for solving the mysteries of biosynthesis, will be found in the two references covering the periods 1966–1993,[1a] and 1993–1994,[48] respectively.

Acknowledgements—We thank N.I.H. and the Robert A. Welch Foundation for generous financial support.

REFERENCES

1. For in-depth reviews covering B_{12} biosynthesis up to April 1993 see:
   (a) Scott, A. I. *Acc. Chem. Res.* **1990**, *23*, 308; *Angew. Chem. Int. Edn. Eng.* **1993**, *32*, 1223.
   (b) Battersby, A. R. *Acc. Chem. Res.* **1986**, *319*, 147; Battersby, A. R.; McDonald, E. in *B_{12}*; Dolphin, D. Ed.; Wiley: New York, 1982, Vol. 1, pp.107–144; Leeper, F. J. *Nat. Prod. Rep.* **1987**, *6*, 171.
   (c) *The Biosynthesis of the Tetrapyrrole Pigments*. Ciba Foundation Symposium 180; Wiley: New York, 1994.
2. Battersby, A. R. *Science* **1994**, *264*, 1551.
3. Woodward, R. B. *Pure Appl. Chem.* **1973**, *33*, 145.
4. (a) Eschenmoser, A.; Wintner, C. E. *Science* **1977**, *196*, 1410; (b) For an account of Eschenmoser's many other contributions to corrin synthesis and chemistry which provide a wealth of mechanistic and structural detail so useful as a guide for our biosynthetic experiments see Eschenmoser, A. *Angew. Chem. Int. Ed. Engl.* **1988**, *27*, 5.
5. Scott, A. I.; Townsend, C. A.; Okada, K.; Kajiwara, M. *Trans. N. Y. Acad. Sci., Series II* **1973**, *35*, 72.
6. Scott, A. I.; Townsend, C. A.; Okada, K.; Kajiwara, M.; Whitman, P. J.; Cushley, R. J. *J. Am. Chem. Soc.* **1972**, *94*, 8267.
7. Scott, A. I.; Townsend, C. A.; Okada, K.; Kajiwara, M.; Cushley, R. J. *J. Am. Chem.. Soc* **1972**, *94*, 8269.
8. Scott, A. I.; Yagen, B.; Lee, E. *J. Am. Chem. Soc.* **1973**, *95*, 5761.
9. Scott, A. I.; Irwin, A. J.; Siegel, L. M.; Shoolery, J. N. *J. Am. Chem. Soc.* **1978**, *100*, 316.
10. Bykhovsky, V. Y. in *Vitamin B_{12}*; Zabalak, B.; Friedrich, W. Eds; de Gruyter: Berlin, 1979.
11. Jordan, P. M.; Burton, G.; Nordlov, H.; Pryde, L. M.; Schneider, M.; Scott, A. I. *J. Chem. Soc., Chem. Commun.* **1979**, 204.
12. Müller, G.; Gneuss, K. D.; Kriemler, H. P.; Scott, A. I.; Irwin, A. J. *J. Am. Chem. Soc.* **1979**, *101*, 3655.
13. Battersby, A. R.; Bushnell, M. J.; Jones, C.; Lewis, N. G.; Pfenninger, A. *Proc. Natl Acad. Sci. USA* **1981**, *78*, 13.
14. Mombelli, L.; Nussbaumer, C.; Weber, H.; Müller, G.; Arigom, D. *Proc. Natl Acad. Sci. USA* **1981**, *78*, 11.
15. Müller, G.; Schmiedl, J.; Schneider, E.; 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.
16. 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.; Kriemler, H.-P. *J. Am. Chem. Soc.* **1986**, *108*, 7875.
17. Scott, A. I.; Mackenzie, N. E.; Santander, P. J.; Fagerness, P. E.; Müller, G.; Schneider, E.; Sedlmeier, R.; Worner, G. *Bioorg. Chem.* **1984**, *12*, 356. (For numbering system see Scheme 1.)
18. Scott, A. I.; Williams, H. J.; Stolowich, N. J.; Karuso, P.; Gonzalez, M. D.; Müller, G.; Hlineny, K.; Savvidis, E.; Schneider, E.; Traub-Eberhard, U.; Wirth, G. *J. Am. Chem. Soc.* **1989**, *111*, 1897.
19. Blanche, F.; Thebout, D.; Frechet, D.; Vuilhorgne, M.; Crouzet, J.; Cameron, B.; Hlineny, K.; Traub-Eberhard, V.; Zboron, M.; Müller, G. *Angew Chem. Int. Edn. Engl.* **1990**, *29*, 884.
20. Uzar, H. C.; Battersby, A. R. *J. Chem. Soc., Chem. Commun.* **1982**, 1204.
21. Uzar, H. C.; Battersby, A. R. *J. Chem. Soc., Chem. Commun.* **1985**, 585.

22. Scott, A. I. *J. Heterocyc. Chem.* **1987**, *14*, S-75.
23. Scott, A. I.; Roessner, C. A.; Stolowich, N. J.; Karuso, P.; Williams, H. J.; Grant, S. K.; Gonzales, M. D.; Hoshino, T. *Biochemistry.* **1988**, *27*, 7984; Scott, A. I., Roessner, C. A.; Clemens, K. R.; Stolowich, N. J.; Santander, P. J.; Gonzalez, M. D. *FEBS Lett.* **1988**, *242*, 319; 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; 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; Aplin, R. T.; Baldwin, J. E.; Pichon, C.; Roessner, C. A.; Scott, A. I.; Schofield, C. J.; Stolowich, N. J.; Warren, M. J. *Bioorg. Med. Chem. Lett.* **1991**, *1*, 503.
24. Warren, M. J.; Jordan, P. M. *FEBS Lett.* **1987**, *225*, 87; Warren, M. J.; Jordan, P. M. *Biochemistry* **1989**, *27*, 9020.
25. Hart, G. J.; Miller, A. D.; Leeper, F. J.; Battersby, A. R. *J. Chem. Soc., Chem. Commun.* **1987**, 1762; Miller, A. D.; Hart, G. J.; Packman, L. C.; Battersby, A. R. *Biochem. J.* **1988**, *254*, 915; Hart, C. J.; Miller, A. D.; Battersby, A. R. *Biochem. J.* **1988**, *252*, 909; Beifus, U.; Hart, G. J.; Miller, A. D.; Battersby, A. R. *Tetrahedron Lett.* **1988**, *29*, 2591.
26. Louie, G. V.; Brownlie, P. D.; Lambert, R.; Cooper, J. B.; Blundell, T. L.; Wood, S. P.; Warren, M. J.; Woodcock, S. C; Jordan, P. M. *Nature* **1993**, *359*, 33.
27. Warren, M. J.; Roessner, C. A.; Santander, P. J.; Scott, A. I. *Biochem. J.* **1990**, *265*, 725.
28. Warren, M. J.; Stolowich, N. J.; Santander, P. J.; Roessner, C. A.; Sowa, B. A.; Scott, A. I. *FEBS Lett.* **1990**, *261*, 76; Scott, A. I.; Warren, M. J.; Roessner, C. A.; Stolowich, N. J.; Santander, P. J. *J. Chem. Soc., Chem. Commun.* **1990**, *8*, 593; Warren, M. J.; Gonzalez, M. D.; Williams, H. J.; Stolowich, N. J.; Scott, A. I. *J. Am. Chem. Soc.* **1990**, *112*, 5343.
29. Roessner, C. A.; Warren, M. J.; Santander, P. J.; Atshaves, B. P.; Ozaki, S.-I.; Stolowich, N. J.; Iida, K.; Scott, A. I. *FEBS Lett.* **1992**, *301*, 73.
30. Jeter, R. M.; Roth, J. R. *J. Bacteriol.* **1987**, *169*, 3189.
31. Roessner, C. A.; Warren, M. J.; Santander, P. J.; Atshaves, B. P.; Ozaki, S.; Stolowich, N. J.; Iida, K.; Scott, A. I. *FEBS Lett.* **1992**, *301*, 73; Ozaki, S.-I.; Roessner, C. A.; Stolowich, N. J.; Atshaves, B. P.; Hertle, R.; Müller, G.; Scott, A. I. *J. Am. Chem. Soc.* **1993**, *115*, 7935.
32. Thibaut, D.; Debussche, L.; Blanche, F. *Proc. Natl Acad. Sci. USA* **1990**, *87*, 8795.
33. Blanche, F.; Thibaut, D.; Famechon, A.; Debussche, L.; Cameron, B.; Crouzet, J. *J. Bacteriol.* **1992**, *174*, 1036.
34. Cameron, B.; Crouzet, J. *J. Bacteriol.* **1992**, *174*, 1050.
35. Thibaut, D.; Blanche, F.; Debussche, L.; Leeper, F. J.; Battersby, A. R. *Proc. Natl Acad. Sci. USA* **1990**, *87*, 8800.
36. Thibaut, D.; Kiuchi, F.; Debussche, L.; Leeper, F. J.; Blanche, F.; Battersby, A. R. *J. Chem. Soc., Chem. Comm.* **1992**, 139.
37. Thibaut, D.; Kiuchi, F.; Debussche, L.; Blanche, F.; Kodera, M.; Leeper, F. J.; Battersby, A. R. *J. Chem. Soc., Chem. Commun.* **1992**, 982.
38. Thibaut, D.; Couder, M.; Famechon, A.; Debussche, L.; Cameron, B.; Crouzet, J.; Blanche, F. *J. Bacteriol.* **1992**, *174*, 1043.
39. Warren, M. J.; Roessner, C. A.; Ozaki, S.; Stolowich, N. J.; Santander, P. J.; Scott, A. I. *Biochemistry* **1992**, *31*, 603.
40. Thibaut, D.; Debussche, L.; Frechet, D.; Herman, F.; Vuilhorgne, M.; Blanche, F. *J. Chem. Soc., Chem. Commun.* **1993**, 513.
41. (a) Scott, A. I.; Roessner, C. A.; Stolowich, N. J.; Spencer, J. B.; Min, C.; Ozaki, S.-I. *FEBS Lett.* **1993**, *331*, 105.
(b) Debussche, L.; Thibaut, M.; Danzer, M.; Debu, F.; Fréchet, D.; Herman, F.; Blanche, F.; Vuilhorgne, M. *J. Chem. Soc., Chem. Commun.* **1993**, 1100.
(c) Debussche, L.; Thibaut, D.; Cameron, B.; Crouzet, J.; Blanche, F. *J. Bacteriol.* **1993**, *175*, 7430.
42. Min, C.; Atshaves, B. P.; Roessner, C. A.; Stolowich, N. J.; Spencer, J. B.; Scott, A. I. *J. Am. Chem. Soc.* **1993**, *115*, 10380.
43. Spencer, J. B.; Stolowich, N. J.; Santander, P. J.; Pichon, C.; Kajiwara, M.; Tokiwa, S.; Takatori, K.; Scott, A. I. *J. Am. Chem. Soc.* **1994**, *116*, 4991.
44. Spencer, J. B., private communication.

45. Kurumaya, K.; Okazaki, T.; Kajiwara, M. *Chem. Pharm. Bull.* **1989**, *37*, 1151; Scott, A. I.; Stolowich, N. J.; Atshaves, B. P.; Karuso, P.; Warren, M. J.; Williams, H. J.; Kajiwara, M.; Kurumaya ,K.; Okazaki, T. *J. Am. Chem. Soc.* **1991**, *113*, 9893.
46. Vishwakarma, R. A.; Balachandran, S.; Alanine, A. I. D.; Stamford, N. P. J.; Kiuchi, F.; Leeper, F. J.; Battersby, A. R. *J. Chem. Soc. Perkin Trans 1* **1993**, 2893.
47. Müller, G.; Zipfel, F.; Hlineny, K.; Savvidis, E.; Hertle, R.; Traub-Eberhard, U.; Scott, A. I.; Williams, H. J.; Stolowich, N. J.; Santander, P. J.; Warren, M. J.; Blanche, F.; Thibaut, D. *J. Am. Chem. Soc.* **1991**, *113*, 9891.
48. Roessner, C. A.; Spencer, J. B.; Ozaki, S.; Min, C.; Atshaves, B. P.; Nayar, P.; Anousis, N.; Stolowich, N. J.; Holderman, M. T.; Scott, A. I. *Protein Expression and Purification* **1994** in press; Scott, A. I. *Synlett* **1994**, in press; Roessner, C. A.; Spencer, J. B.; Stolowich, N. J.; Wang, J.; Nayar, G. P.; Santander, P. J.; Pichon, C.; Min, C.; Holderman, M.; Scott, A. I. *Chemistry and Biology* **1994**, *1*, 119.
49. Scott, A. I. *Pure & Appl. Chem.* **1993**, *65*, 1299; Scott, A. I. *Chem. in Britain* **1993**, *29*, 687.
50. For an excellent review of this field see C.-H. Wong and G. M. Whitesides "Enzymes in Synthetic Organic Chemistry", *Tetrahedron Organic Chemistry Series, Vol. 12;* Elsevier Science: Oxford, 1994.
51. *Notebooks of Leonardo da Vinci 1508–1518*, Paul Richter, Ed.; Dover: New York, 1970.