PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

The Bakerian Lecture, 1996. Genetically engineered synthesis of natural products

The Royal Society

Phil. Trans. R. Soc. Lond. A 1998 **356**, 1341-1366
doi: 10.1098/rsta.1998.0225

The Royal Society

The Bakerian Lecture, 1996 Genetically engineered synthesis of natural products

By A. I. Scott

Center for Biological NMR, Department of Chemistry, Texas A&M University, College Station, TX 77843-3255, USA

Lecture held 5 June 1996

Contents

Since many natural products are of biological and medicinal importance, methods for studying their biosynthetic pathways, which may eventually result in increased production or novel compounds, are continually being sought. Advances in genetic engineering have made possible the homologous or heterologous expression of many natural product biosynthetic genes from divergent sources resulting in a supply of enzymes not readily available by isolation from the producing organism. Mixing and matching of these enzymes in cell-free reactions provides information about enzyme mechanisms and pathway intermediates, which can be used to design the total cell-free, multi-enzyme synthesis of the target molecule.

Keywords: vitamin B_{12}; multi-enzyme synthesis; biosynthesis; natural products; alkaloids

1. Introduction

For almost 40 years organic chemists have been studying the metabolic pathways leading to the so-called secondary metabolites of nature. More than 100 000 of these small molecules (molecular weight less than 2500) have been characterized, and the biosynthetic origins of the major classes have been firmly established by tracer methodology, first by using $^{14}C-$ and $^{3}H-$ precursors which were fed *in vivo* to plants and microorganisms. The next technical breakthrough involved cell-free

systems and, where appropriate, the isolation and refeeding of intermediates accumulated in mutants of bacterial cells. Slowly, by the end of the 1970s, the three main pathways to secondary metabolites were recognized and the origin of a majority of natural products could be assigned to the mevalonate, polyketide or shikimate routes (and their combinations) by inspection of their structures and, in many instances, by experimental feeding or bioconversion studies now using stable isotopes (^{13}C, ^{2}H, ^{15}N) and nuclear magnetic resonance (NMR) spectroscopy.

During the last decade it has become clear that the complete definition of a pathway to a natural product requires the isolation and characterization of the full set of biosynthetic enzymes, often as many as 15–20 being required for the synthesis of a complex structure. Faced with the enormous task of identifying all of the intermediates and the enzymes which connect them in a multi-step process, researchers in the field began to focus on the genetic analysis of natural product biosynthesis using genomic and cDNA libraries to search for the genes encoding the biosynthetic enzymes of the pathway. In this lecture, I discuss recent advances in natural product biosynthesis in which bio-organic chemists have changed their roles of spectators of nature's strategy for the synthesis of these complex metabolites to those of active participants in their attempts to reconstitute the pathway *in vitro* (and even to alter the structure of the final product) making full use of the combined power of organic chemistry and genetic engineering. The examples I have chosen are representative of a rapidly growing field and are general enough in concept and execution to be applied to most plant, bacterial and fungal systems. First of all, it is important to define 'genetically engineered synthesis' in the context of the ensuing discussion and to separate this endeavour from that of metabolic engineering of microorganisms. Great strides have been made in the latter field, especially in *Streptomyces* spp. (McDaniel *et al.* 1993) where it has been possible to mix genes from different organisms to produce hybrid structures corresponding to the combination of several pathways, and also to amplify selected genes from a metabolic pathway to overproduce a target molecule. Similarly, transgenic plants can be engineered to produce enhanced yields of secondary metabolites or useful proteins, by introducing pathways which are not present in the native species. In many cases, rigid branch-points in the metabolic network can be removed, leading to higher yields of product (Bailey 1991). Our approach, which we have termed genetically engineered synthesis, involves the acquisition of the complete set of gene products requisite for the synthesis of a natural product, followed by their recombination *in vitro*. This procedure eliminates unwanted processing of the key intermediates of the pathway along the main arteries of metabolism necessary for the well-being of the host organism *in vivo*. In other words, by removing all of the metabolic machinery except those enzyme-catalysed reactions dedicated to the synthesis of the desired target, genetically engineered synthesis represents a new departure for the production of desirable, frequently rare natural products of biological importance.

Why has this concept not been exploited previously? One reason is that the biosynthetic enzymes for natural product synthesis are rare, and only recently have organic chemists and molecular biologists cooperated in expressing the enzymes necessary for the synthesis of secondary metabolites by transferring genes from plants to bacteria. Perhaps the principal deterrent to multi-enzyme one-flask synthesis has been the logical question of the selection of favourable enzyme kinetics which may require modulation of pH and solvent, or even genetic alteration of enzymes whose unfavourable

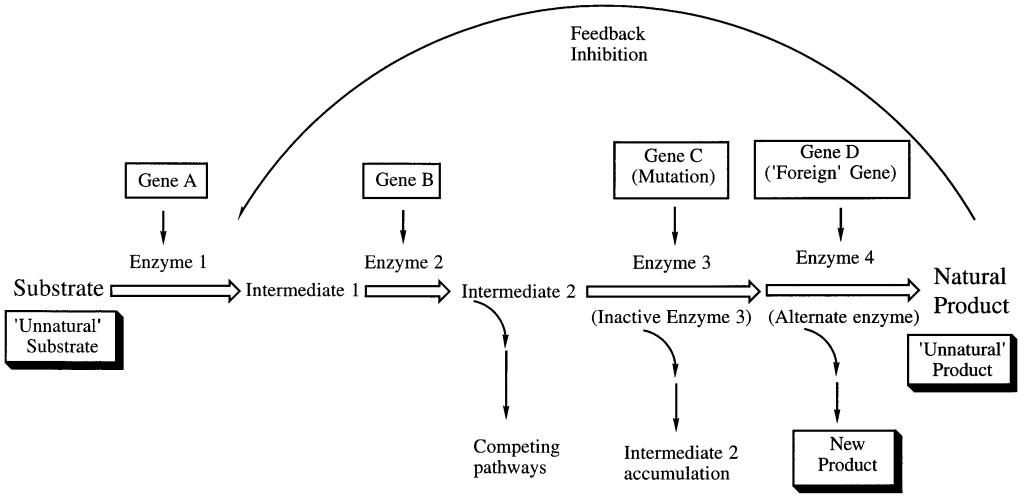


Figure 1. Alternatives in natural product synthesis. The normal biosynthetic pathway from substrate to natural product can be altered for enhanced production or product diversity. Competing pathways can be overcome *in vitro* by using purified enzymes or a large excess of the desired enzymes. Product inhibition may be overcome by replacing regulatory enzymes with isozymes, either naturally occurring or genetically altered, that do not exhibit feedback inhibition. Intermediate accumulation can be induced by mutation *in vivo* or by omission of the enzymes after the desired intermediate in a multi-enzyme *in vitro* system. Unnatural products may be obtained by supplying unnatural substrates. Limitations *in vivo* (cell toxicity, lack of uptake, etc.) may be overcome *in vitro*. New products may be produced by introduction of genes from heterologous organisms *in vivo*, or by addition of the corresponding enzymes *in vitro*, which may reduce problems of expression, toxicity and unfavourable environmental conditions.

K_{m} or k_{cat} values could terminate the synthesis by product (feedback) or substrate inhibition (figure 1).

In fact, it is clear from the examples described in the sequel that the successful acquisition of the biosynthetic enzymes (rather than their kinetic parameters) appears to be the main hurdle, and that the combination of as many as 10–15 enzymes in a single reactor can lead to the target molecule in high overall yield and with perfect stereochemical fidelity. From the practical point of view, many enzyme-catalysed reactions require cofactors which are often too expensive to use stoichiometrically, especially on a large scale. Several efficient methods have been developed for the regeneration of the principal cofactors (nucleoside triphosphates, nicotinamides, etc.) which not only reduce the cost of the process but help drive the reaction towards product either by influencing the equilibrium through coupling to the enzymes required for regeneration or by preventing the accumulation of by-products (Wong & Whitesides 1994). Furthermore, the enantio-selectivity relative to the stoichiometric reaction is frequently increased when the catalytic cycle is coupled to a regeneration system.

2. Synthetic design

Depending on the nature of the organism from which the metabolic route to the natural product is to be reconstituted, the strategy for *in vitro* synthesis must be chosen with care. Since the biosynthetic genes are seldom (if ever) clustered in eukaryotic organisms, a cDNA library must be constructed and probed with oligonucleotides or

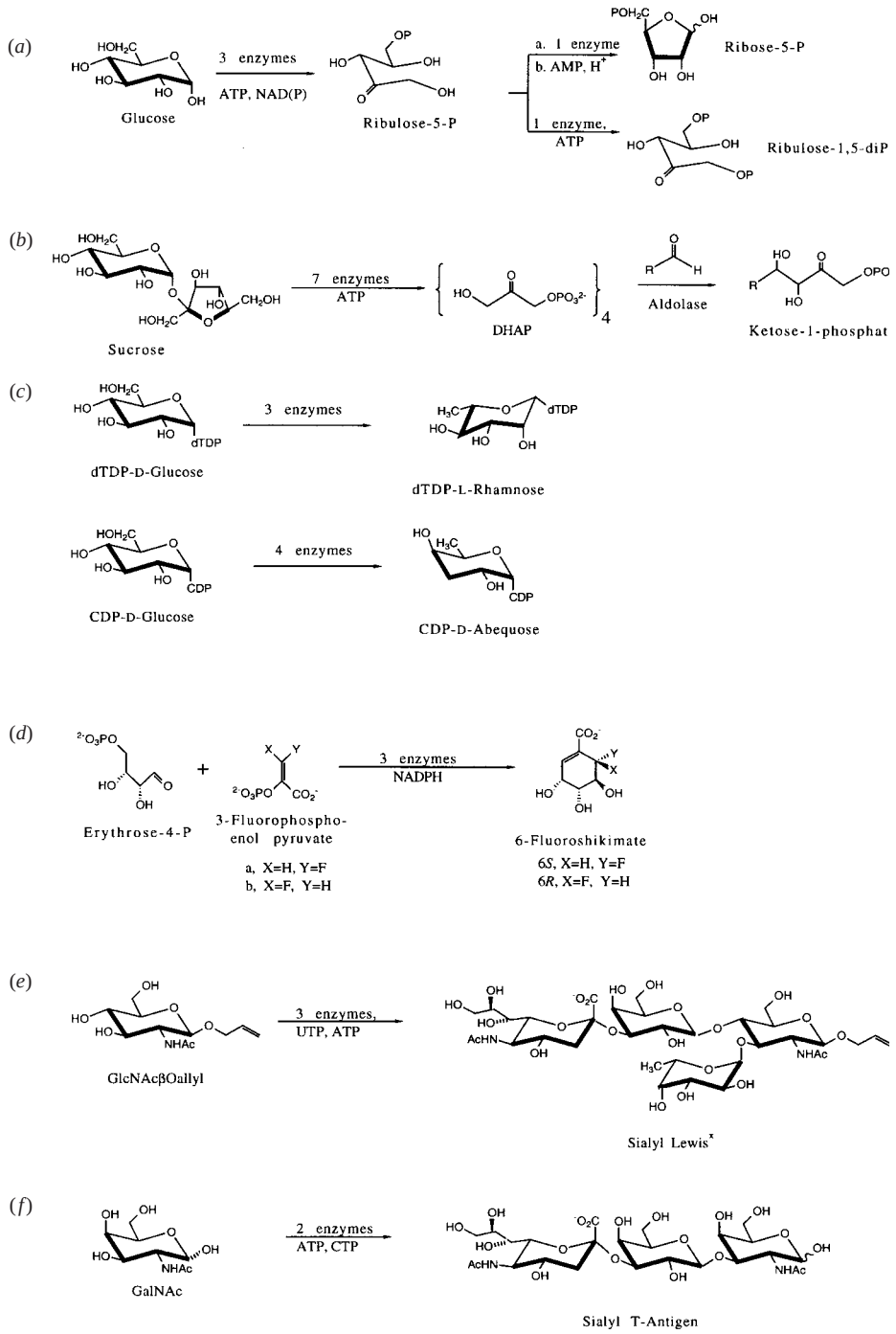


Figure 2. For caption see opposite.

antibodies corresponding to the proteins of interest. In contrast, microorganisms fre-

quently display lengthy segments of DNA which correspond to operons encoding the enzymes for as many as 10–20 steps of a pathway. In such cases, classical selection of mutants and complementation studies with a genomic library will provide a logical, if somewhat tedious, approach which, thanks to modern recombinant DNA techniques, can be accelerated in the later stages of subcloning and expression to produce the complete repertoire of biosynthetic enzymes, which may become available before the exact structures of the intermediates of the pathway have been determined. The disparate strategies for prokaryotes and eukaryotes are summarized in table 1, the choice of expression vector being of paramount importance. The advent of differential display or 'subtraction' libraries is expected to have a profound impact on the currently lengthy procedures for the eukaryotes (table 1, entries 1–6), an area which is ripe for development.

3. Multi-enzyme synthesis exemplified: carbohydrates

The concept of reconstituting a pathway or mixing enzymes from interlinked pathways in a single reactor was reduced to reality by the early work of Wong *et al.* (1980) using commercially available enzymes to synthesize ribose-5-phosphate and ribulose-1,5-bisphosphate from glucose (figure 2*a*). A similar approach was employed to convert sucrose to the versatile substrate, dihydroxy acetone phosphate (DHAP), which was then further processed in the same reaction vessel by the addition of fructose-1,6-diphosphate aldolase in the presence of several aldehyde substrates to produce a variety of unusual branched chain sugars (ketose-1-phosphates, figure 2*b*), a sequence which has been termed 'artificial metabolism' (Fessner & Walter 1992) since the routes evolved are not those of the normal glycolytic pathway. The yields were increased by replacing the rate-limiting enzyme with a recombinant isozyme with more favourable kinetics. Another example of this type is the synthesis of monosaccharides (figure 2*c*) found in the O-antigenic polysaccharide chains (which may be used in vaccine production) of *Salmonella enterica*. Both rhamnose and abequose have been produced using cell-free extracts of *Escherichia coli* (*E. coli*) strains expressing the *S. enterica rbf* genes (Marumo *et al.* 1992; Lindquist *et al.* 1994). In another example, Duggan *et al.* (1995) supplied the unnatural substrate, 3-fluoro-phosphoenolpyruvate, and erythrose-4-phosphate to enzymes of the shikimate biosynthetic pathway to produce equal amounts of the 'unnatural' (*R*)- and (*S*)-6-fluoro analogues of shikimic acid (figure 2*d*), of potential use as antibiotics.

Figure 2. Multi-enzyme synthesis of carbohydrates. An (*) indicates enzymes produced by genetic engineering. (*a*) Synthesis of ribose and ribulose-1,6-diphosphate with hexokinase, glucose-6-phosphate dehydrogenase, 6-phosphogluconate dehydrogenase, phosphoribose isomerase and phosphoribulokinase. (*b*) The multi-enzyme one-flask synthesis of ketose-1-phosphates with invertase, xylose isomerase, hexokinase, glucose-6-phosphate isomerase, fructose-6-phosphate kinase*, fructose-1,6-biphosphate aldolase and triose phosphate isomerase. (*c*) Synthesis of dTDP-L-rhamnose from dTDP-D-glucose with dTDP-D-glucose-4,6-dehydratase*, dTDP-4-keto-rhamnose 3,5-epimerase* and dTDP-6-deoxy-L-*lyxo*-hexulose-4-reductase*. Synthesis of abequose from CDP-D-glucose with CDP-glucose-4,5-dehydratase*, CDP-6-deoxy-D-*xylo*-4-hexulose-3-dehydrase*, CDP-6-deoxy-Δ3,4-glucoseen reductase* and CDP-3,6-dideoxy-D-*xylo*-4 hexulose-4-reductase*. (*d*) Synthesis of (6*R*)- and (6*S*)-fluoroshikimic acids from erythrose-4-phosphate and 3-fluorophosphoenol pyruvate with DHAP synthase*, dehydroquinate synthase, dehydroquinase and shikimate dehydrogenase. (*e*) Synthesis of sialyl-Lewis^x from GlcNAcβOallyl with β-1,4-galactosyltransferase, α-2,3-sialyltransferase* and α-1,3-fucosyltransferase*. (*f*) Synthesis of sialyl T-antigen from N-acetylgalactosamine with β-galactosidase and α-2,3-sialyltransferase*.

Table 1. *Strategies for in vitro genetically engineered synthesis*

prokaryotes	eukaryotes
1. Mutagenesis and selection of mutants 'blocked' in the pathway.	1. Identify pathway by feeding labelled precursors (shikimate, polyketide, mevalonate, amino acid, etc.) *in vivo*.
2. Prepare a genomic library in a plasmid, phage or cosmid vector.	2. Purify the biosynthetic enzymes, often large and multi-functional, e.g. PKS.
3. Complementation of mutants with genomic library and isolation of complementing vectors.	3. Determine amino acid of N-terminal and/or proteolytic fragments.
4. Sequence the insert DNA to identify open reading frames. Other genes of the same pathway are often clustered in polycistronic operons.	4. Synthesize oligonucleotide probes or isolate and label antibodies.
5. Subclone genes into expression vectors providing optimal transciptional and translational signals.	5. Isolate mRNA and construct cDNA library in plasmid or phage vectors. Subtraction libraries for inducible pathways.
6. Express and characterize the encoded enzyme.	6. Screen library with labelled probes. Chemical screening possible if library is in a yeast expression vector.
	7. Isolate vector and sequence insert DNA.
	8. Express gene in heterologous system (*E. coli*, yeast, baculovirus).

Cell surface oligosaccharides are highly diversified in structure and are associated with many cell functions such as receptor binding and as antigenic determinants. Recent developments pioneered by Wong have resulted in the multi-enzyme synthesis of several of these complex oligosaccharides. Representative of Wong's work is the synthesis of the tetrasaccharide, sialyl Lewis^{x} (figure 2*e*), a candidate for the treatment of reperfusion tissue injury, which can now be produced on the kilogram scale using recombinant enzymes and sugar nucleotide regeneration (Wong *et al.* 1995). Along the same lines, Kren & Thiem (1995) have described the single flask preparation of the sialylated trisaccharide epitope of the T-antigen (figure 2*f*), which is expressed on several cell surfaces, including epithelial cancers, and is of potential value for the synthesis of antitumour vaccines. In the above examples, cofactors were regenerated with enzymatic systems at the appropriate steps shown in figure 2.

4. Plant secondary metabolites

Although the primary metabolic sequences leading from the building blocks of acetate (to polyketides and isoprenoids), shikimate (alkaloids) and glucose (to complex carbohydrates) are shared by higher plants and microorganisms, secondary events such as cyclization, hydroxylation and rearrangement of the oligomerized units are, for the most part, quite distinct. For example, whereas squalene serves as the precursor of both plant and microbial steroids, the mechanism of its cyclization to lanosterol and the enzymes which mediate the two processes are not shared by plants and microorganisms. Again, the wide variety of mono-, sesqui-, di- and

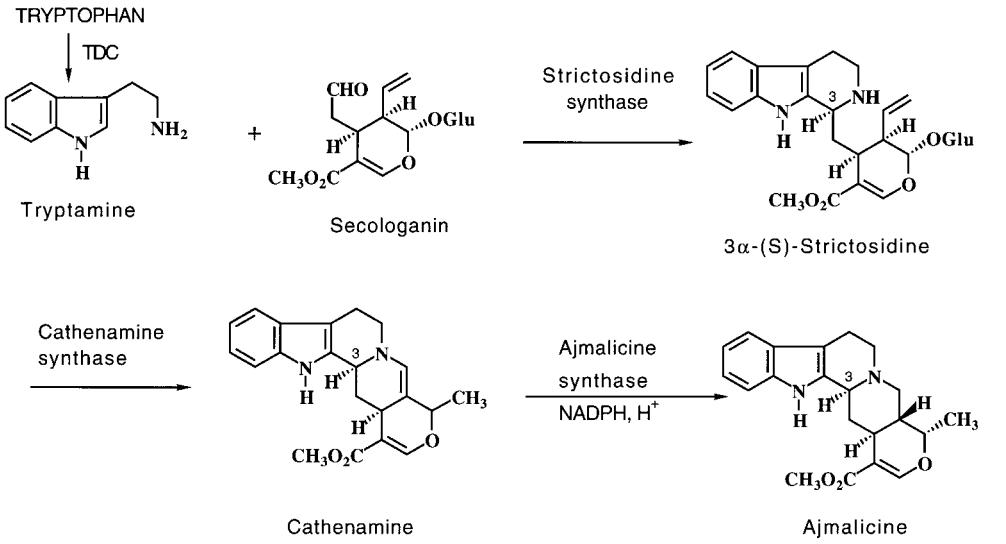


Figure 3. The enzymatic synthesis of ajmalicine from tryptamine and secologanin.

tri-terpene cyclic templates found in plants are, with a few exceptions, not present in bacteria, which have an equally rich but different set of structural motifs in their secondary metabolites. A notable exception is the plant growth hormone, gibberellic acid, which occurs in both plants and microorganisms. Since many plant substances have highly desirable medicinal properties, the possibility of transferring the biosynthetic genes from higher plants to bacteria has now assumed major importance. To prepare for the synthesis of taxol, we have developed overexpression systems for all of the enzymes of the mevalonate pathway to allow the multi-step synthesis of geranyl geraniol biphosphate.

(*a*) *Indole alkaloids*

The pivotal enzyme for the synthesis of this large family of alkaloids (greater than 1500 members), derived from the tryptophan–monoterpenoid pathway, is strictosidine synthase, which catalyses the Pictet–Spengler-like condensation of tryptamine and secologanin to form strictosidine (figure 3). Genes encoding tryptophan decarboxylase and strictosidine synthase have been cloned from cDNA libraries of *Catharanthus roseus* and *Rauwolfia serpentina* (Kutchan 1993; Roessner *et al.* 1992). Overexpression of strictosidine synthase in *E. coli*, yeast and insect cells has allowed acquisition of substantial amounts of this enzyme for synthetic and mechanistic studies (Kutchan *et al.* 1991). When coupled with partly purified cathenamine synthase and ajmalicine synthase in the presence of NADPH, the production of the heteroyohimbine alkaloid, ajmalicine, from tryptamine and secologanin, was demonstrated (Devagupta 1994, figure 3). Optimization of this process must await the expression of the latter two enzymes and the addition of the coupled enzyme system for regeneration of NADPH.

5. Towards a total genetically engineered synthesis of vitamin B_{12}

In the late 1960s, several research groups embarked on a daunting expedition—to discover experimentally how vitamin B_{12} is synthesized by bacteria, a process con-

sidered to have evolved over nearly four billion years (Scott 1993). In the remainder of this lecture, I will describe our own efforts which, along with contributions from several other laboratories, notably those of A. R. Battersby (Cambridge), G. Müller (Stuttgart), Rhône-Poulenc Rorer (Paris) and D. Arigoni (Zürich), have led to the final solution of how nature makes the corrin nucleus. Although the story of B_{12} biosynthesis is about molecular architecture, the actual process of discovering how nature built this incredibly complex structure can be compared to a long voyage of exploration, frequently into unmapped territory. We shall also see how our role as spectators of the natural pathway has changed to that of active participants in harnessing the full power of enzymatic catalysis to reconstitute the entire biosynthetic route *in vitro*.

Thus our odyssey began with synthesis, first using the classical methods of organic chemistry necessary to prepare the biosynthetic 'building blocks' bearing ^{13}C-labels strategically placed to uncover, by NMR spectroscopy, the structures of the novel intermediates encountered in the separate enzyme-catalysed steps. This phase of the exploration proceeded at a steady, but necessarily slow, pace over 20 years and at times there seemed no possibility of expanding the small but precious collection of intermediates beyond precorrin 3 (figure 4), although much was learned during this time about ^{13}C-NMR spectroscopy on the sub-milligram scale. Fortunately, in 1988 the advent of the genetic analysis and manipulation of the corrin operons in two microorganisms accelerated the rate of discovery exponentially. As a result, we were able to pass through a rapidly changing scenery reflecting the increasing structural complexity of the new intermediates on the way to corrin, and we will see how all of the intermediates in the aerobic pathway have been found and their structures determined. For an organic chemist, originally trained in synthesis, the most exhilarating phase was yet to come for, armed with the biosynthetic enzymes, we once again returned to synthetic chemistry to prepare the complete corrin structure in the test tube from a 5-carbon substrate, this time in a matter of hours rather than years. I shall now describe the experiments which led to the multi-enzyme synthesis of corrins, which spanned the last 30 years, as necessary background for understanding the logic of our approach. The extensive literature and bibliography covering the period 1960–1995 has recently been reviewed by the major groups involved (Scott 1993, 1994; Battersby 1994; Blanche *et al.* 1995) and, unless indicated otherwise in the text, the necessary references to previous work will be found in these reviews.

(*a*) *The 'early' stages: from ALA to precorrin 3*

When we began the journey in 1968, the road to corrin was defined only as far as the first signposts set in place by David Shemin, which implicated 5-aminolaevulinic acid (ALA) and S-adenosylmethionine (SAM) as the sources of the tetrapyrrolic ligand and methyl groups, respectively. Our first task was to provide the experimental evidence for the intact incorporation of the smaller building blocks by synthesizing ^{13}C-enriched ALA, porphobilinogen (PBG) and most importantly, the reduced macrocycle, uro'gen III, all of which were found to serve as excellent precursors of B_{12} when fed to whole cells of *Propionibacterium shermanii* by observing the ^{13}C-enrichments above natural abundance in the target, cyanocobalamin, a methodology which also showed unambiguously that seven of the eight methyl groups are derived from SAM. With the establishment of the porphyrinoid–corrin connection in 1972, we entered the arena of cell-free biosynthesis which, by the end of the next decade, had led to the 'library' of intermediates, precorrins 1–3 (figure 4). In spite of prodigious

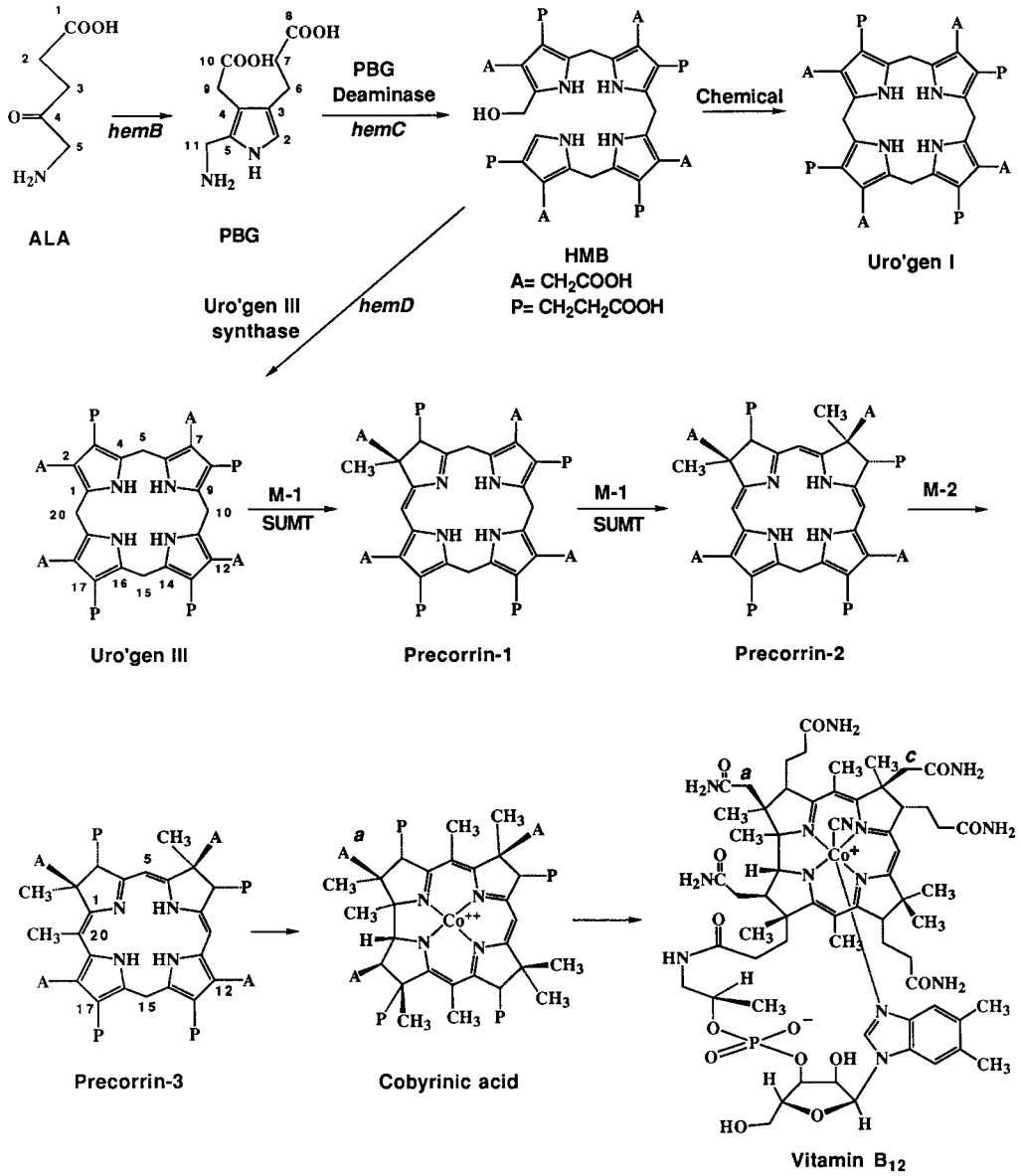


Figure 4. Biosynthesis of vitamin B_{12} in *Propionibacterium shermanii* showing intermediates discovered by 1980.

efforts by all of the participating research groups (Scott 1994; Blanche *et al.* 1995), no further intermediates beyond precorrin 3 came to light between 1979 and 1990. However, during this fallow period we began the process of honing our skills in combining high-field NMR spectroscopy with cryoenzymology to look for intermediates in the reactions catalysed by the enzymes responsible for tetrapyrrole biosynthesis. By adding the powerful techniques of molecular biology to our repertoire, we were ready by 1988 to begin the next stage of our expedition towards the goal of discovering the missing intermediates on the road to corrins, by matching each gene product to its biosynthetic function. The first three enzymes of the pathway, overexpressed in

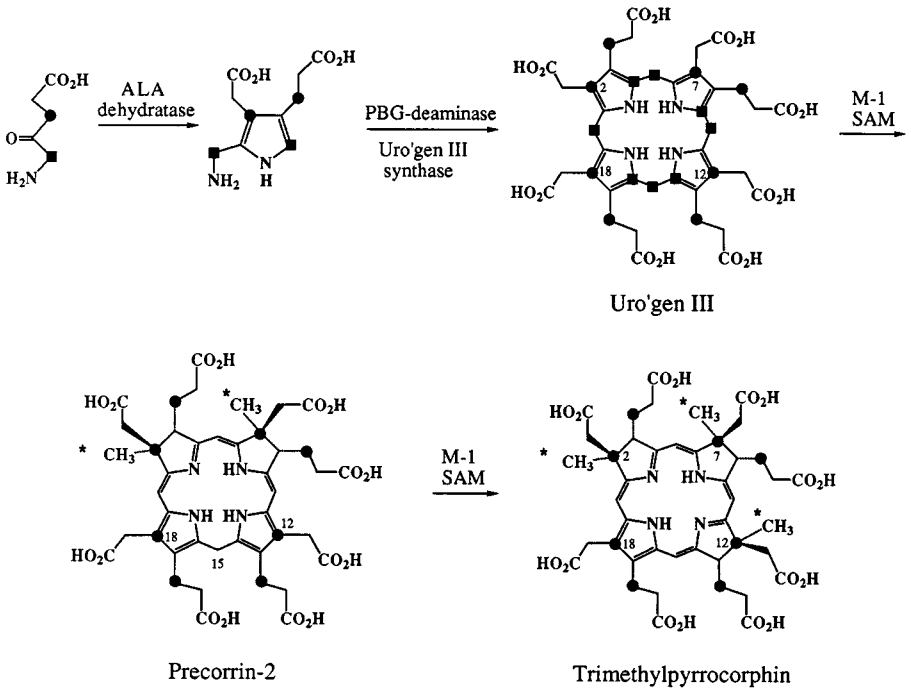


Figure 5. Synthesis of ^{13}C labelled precorrin 2 and its 'overmethylation' at C-12.

E. coli (*hemB*, *C*, *D*) were combined to prepare substantial amounts of uro'gen III from ALA. Next, in order to enter the chiral world of corrins, uro'gen III has to be C-methylated twice by the same enzyme, SAM-uro'gen III methyl transferase (SUMT), which also bears strong homology to *cysG* from *E. coli*.

SUMT was first partly purified from *P. shermanii* by Müller and has been overexpressed in *Pseudomonas denitrificans* as CobA. In *E. coli* it was found that the *CysG* gene encodes uro'gen III methylase (M-1) as part of the synthetic pathway to sirohaeme, the cofactor for sulphite reductase, and overproduction (30 mg l^{-1}) was achieved by the appropriate genetic engineering. Although SUMT and M-1 appear to perform the same task, it has been found that their substrate specificities differ. In fact, the enzyme CysG turned out to be multifunctional (see 'haeme box', figure 6), catalysing the complete synthesis of sirohaeme via the NAD mediated dehydrogenation of precorrin 2, followed by insertion of Fe^{2+}; it has therefore been renamed sirohaeme synthase (Spencer *et al.* 1993). It has been possible to study in detail the reaction catalysed by M-1 directly, by using NMR spectroscopy, and to provide rigorous proof that the structure of precorrin 2 is that of the dipyrrocorphin tautomer shown. Uro'gen III (enriched from [5-^{13}C]-ALA at the positions (■) shown in figure 5) was prepared by mixing dehydratase, deaminase and uro'gen III synthase, and then incubating with M-1 and [$^{13}CH_3$]-SAM. The resultant spectrum of precorrin 2 revealed only one sp^3 enriched carbon assigned to C-15, thereby locating the reduced centre. 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 centres coupled to the pendant ^{13}C-methyl groups at (*) C_2 and C_7, thus confirming an earlier NMR analysis of precorrin 2. The two sets of experiments mutually reinforce the postulate that precorrins 1, 2 and 3 all exist as hexahydroporphyrinoids and labelling experi-

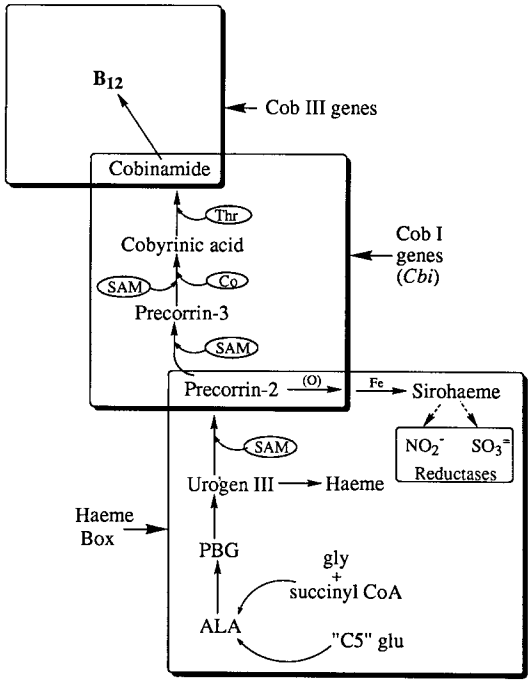
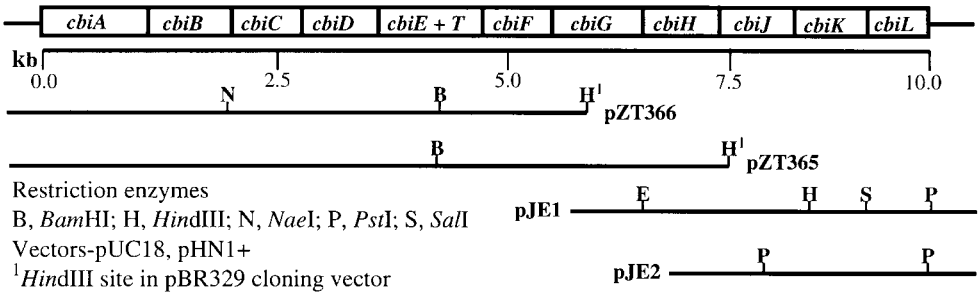


Figure 6. The B_{12} biosynthetic genes showing the 'haeme box' of *E. coli* and the *cbi* genes of *Salmonella typhimurium*.

ments at Cambridge have provided good evidence that precorrin 1 is discharged from the methylating enzyme (SUMT) as the species with the structure shown in figure 4. The above methodology is given in detail since it is typical of, and implicit in, all of the subsequent structure determinations which feature the NMR observation of new ^{13}C-chemical shifts, and the accompanying C–C coupling constants, each time a new $^{13}CH_3^*$ group is introduced or rearrangement occurs. Our search for the biosynthetic enzymes leading to precorrin 2 has so far relied on the 'haeme gene box' (figure 6) in *E. coli* responsible for the synthesis of precorrin 2 as a precursor of sirohaeme, a cofactor for the six-electron *E. coli* sulphite reductase. Fortunately, the haeme and corrin pathways intersect in *E. coli* and *Salmonella typhimurium*. Although the former organism does not make B_{12}, the discovery by Roth *et al.* (1993) that anaerobic fermentation of *S. typhimurium* produces vitamin B_{12} allowed the vast array of genetic and cloning techniques available with this organism to be used in the search for the B_{12} pathway. Three loci at minutes 14, 34 and 42 have been identified by mutation and complementation studies (Roth *et al.* 1993). The main gene cluster at 42 min contains all of the DNA (Cob I) necessary for the synthesis of cobinamide from precorrin 2 (figure 6), a process involving six C-methylations (at C-1, C-5, C-11, C-15, C-17 and C-20), decarboxylation (of the acetate residue at C-12), ring contraction, loss of acetic acid (from C-20 and its attached methyl), amidation and cobalt insertion under control of the *cbi* genes. We were able to use the 10 kb sequence data provided by Roth *et al.* (1993) to clone and overexpress the gene products, corresponding to 12 open reading frames (ORFs) necessary to synthesize cobyrinic acid from precorrin 2 (figure 7).

Ten of the *cbi* genes found in the *S. typhimurium cob* operon were subcloned for



genes	plasmid	construction
cbiC, *D*	pCAR345	pZT366 (N-B) into pHN1+
cbiE	pCAR340	ecpcr[a] product into pHN1+
cbiT, *F*	pCAR276	pZT366 (B-H) into pUC18
cbiT, *F*, *G*[b], *H*	pCAR293	pZT365 (B-H) into pUC18
cbiH, *J*[b]	pCAR309	pJE1 (E-H) into pUC18
cbiK, *L*	pCAR292	pJE1 (P-P) into pUC18
cbiK	pCAR311	pJE2 (S-P) into pUC18
cobF[c]	pCAR332	ecpcr[a] product into pHN1+
cobI[c]	pCAR333	ecpcr[a] product into pHN1+

[a]Expression cassette polymerase chain reaction.
[b]Form insoluble inclusion bodies.
[c]*P. denitrificans* gene.

Figure 7. Plasmids used to identify and express the open reading frames (ORFs) of the *cbi* genes of *S. typhimurium* encoding B_{12} synthetic enzymes.

expression from the four different plasmids (figure 2). The gene products of *cbiE*, *cbiF*, *cbiH* and *cbiL* were shown to be SAM-binding proteins, and based on their homology with other methyltransferases, were considered to be the most likely candidates for methyltransferase activity. Meanwhile, the Rhône-Poulenc group reported that the *P. denitrificans cobL* gene product (Blanche *et al.* 1995) had two functions: methylation (of C-5 and C-15) and decarboxylation of the ring C acetate. SDS-PAGE and NH_2-terminal sequence analysis revealed that two separate gene products in *S. typhimurium* (*cbiE* and *cbiT*) correspond to the *cobL* gene product, with *cbiE* homologous to the methyltransferase region and *cbiT* homologous with the decarboxylase region; i.e. in *Salmonella*, the 5,15-methylase and decarboxylation activities are separated.

From its homology with the *P. denitrificans cobI* gene product (31% identity, 71% conservation), the *cbiL* gene product was predicted to be the *S. typhimurium* precorrin 2 methyltransferase (M-2) and, using this expressed protein (or more efficiently the Cob I enzyme from *P. denitrificans*), the multi-enzyme one-flask synthesis of precorrin 3 from the building block ALA was accomplished as shown in figure 8, both in the NMR tube and preparatively (on the 50 mg scale), by adding the five overexpressed enzymes to the substrate ALA in the presence of SAM (Warren *et al.* 1992). The structure of precorrin 3, a dipyrrocorphin with the constitution shown in figure 3, could then be studied in detail for the first time. This revealed subtle differences in the ^{13}C-NMR spectrum, which reflected the influence of the new

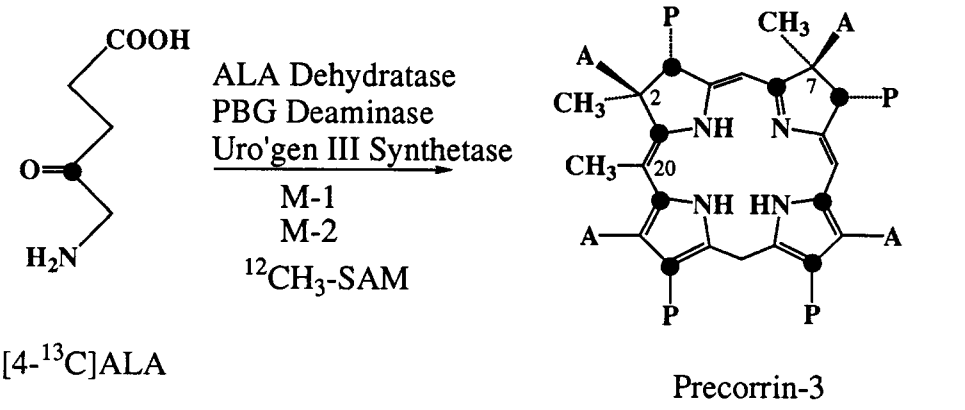


Figure 8. Multi-enzyme synthesis of precorrin 3.

methyl group at C-20 on the conjugated system, resulting in a preponderance of the tautomer shown, whose electronic array is prepared for the next C-methylation step.

As a guide to the anticipated order of insertion of the remaining methyl groups on the periphery of this last intermediate, precorrin 3, we recall earlier pulse-labelling experiments from three research groups (Cambridge, Paris and Texas–Stuttgart), in which the substrates uro'gen III and precorrins 2 and 3 (in their oxidized forms) were incubated with SAM for several hours with a cell-free extract capable of synthesizing cobyrinic acid, followed by a pulse of labelled [$^{13}CH_3$] SAM. By examining the different intensities of the [$^{13}CH_3$]-methyl resonances in the resultant cobester, the *sequence* of methyl group insertion was found to correspond to C-methylation at C-17, followed by C-12 (now known to occur first at C-11, *vide infra*), then at C-1, C-5 and C-15 (see figure 4 for the numbering system). This differentiation between C-5 and C-15 was found in *P. shermanii*, while the reverse order for the last two insertions was deduced using *Clostridium tetanomorphum* extracts. In the third study (*P. denitrificans*), the C-5/C-15 distinction could not be made. However, there is agreement that C-17 is the first site of alkylation on the precorrin 3 template and that this is followed by C-methylation at C-11, then at C-1, imposing certain restrictions on the type of structures expected for the missing intermediate precorrins 4, 5 and 6, corresponding to methyl insertion at C-17, C-11 and C-1, respectively.

In order to begin the task of discovering the remaining precorrin intermediates, precorrin 3 was incubated in turn with each of the putative overexpressed methyltransferases from the genes *cbiE*, *H*, *F*, *L* and SAM. A surprising result was obtained. The only C-methylation observed was that catalysed by ORF-7 (*cbiF*) and this turned out to be methylation at C-11! The new isolate (figure 9) is a modified corphin bearing a *fourth* methyl group at C-11. Since biochemical conversion to cobyrinic acid was not demonstrated, we named the new compound 4x (Roessner *et al.* 1992), whose structural variant recalls one of the four possible structures proposed for factor S_3, an unusual tetramethylated zinc complex based on the type I (symmetrical) porphyrin structure. The spectroscopic data for the latter metabolite isolated from *P. shermanii* could not distinguish between methylation at C-1 or C-11, but it is now clear that the correct isomer is the one in which 11 ($\equiv$ 16) (α-) methylation has taken place on the uro'gen I template. This new structural proposal for factor S_3 was nicely confirmed by a second multi-enzyme synthesis, this time using [4-^{13}C]ALA as substrate in a one-flask reaction mixture containing dehydratase, deaminase ($\rightarrow$ uro'gen I)

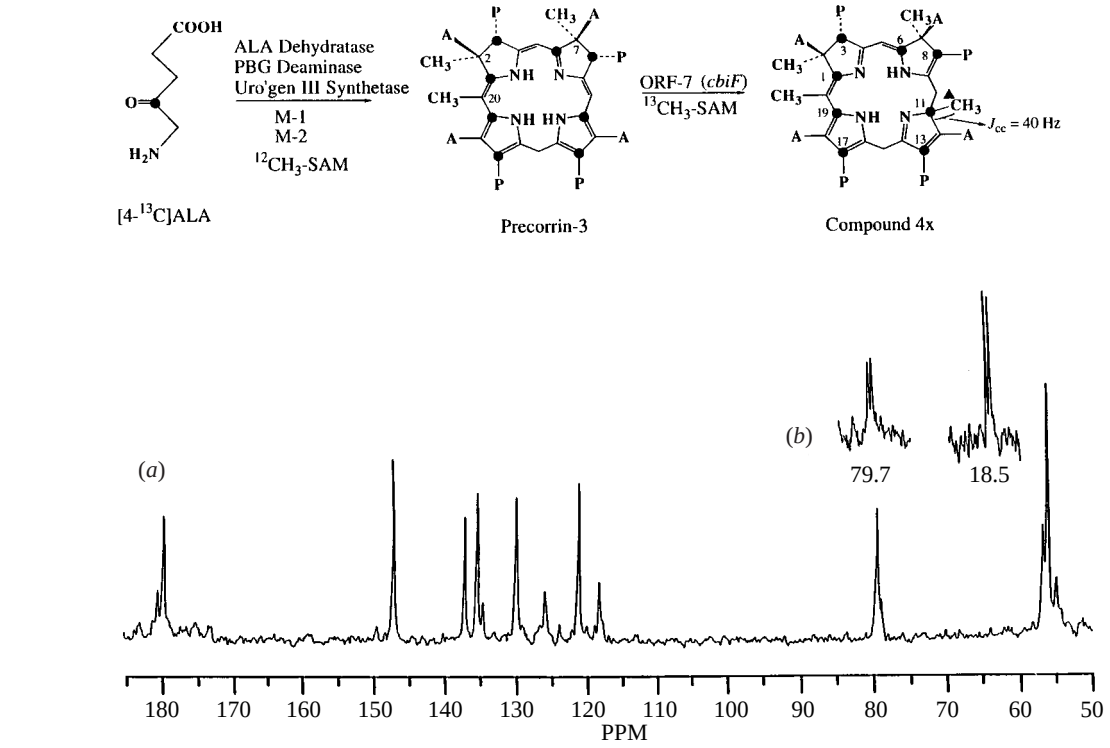


Figure 9. Multi-enzyme synthesis of compound 4x with the *cbiF* gene product and SAM. The ^{13}C-NMR spectrum of 4x (*a*) reveals three sp^2 (C-8, C-13 and C-17) but only one sp^2 propionate terminus (C-3) and a signal at δ79.7 typical of sp^3 carbon (C-11) adjacent to nitrogen. The inset (*b*) shows the coupling (Jcc = 40 Hz) of the new $^{13}CH_3$ (δ18.5) to the C-11 signal at δ79.7.

cysG, *cbiF* and SAM (figure 10). In the absence of the last of these enzymes (*cbiF*), the product 2,7,12-trimethylpyrrocorphin accumulates, but when the fourth enzyme (*cbiF*) was added, a new signal appeared at δ79 heralding the insertion of a fourth (α-) methyl group on the C-16 position (figure 10). The resultant zinc complex (an octamethylester) was identical in every respect with factor S_3 isolated earlier from *P. shermanii*. It only remains to define the absolute stereochemistry of the new chiral centre at C-16. These findings suggest a lack of substrate specificity in C-methylation by the methyl transferases and provide confirmation of the function of *cbiF* as a methyl transferase which attacks the α-position of pyrrole rings of pyrrocorphins, whether type III or type I, as well as opening the door to synthetic chemistry based on C-methylation at electron-rich centres.

The discovery that factor S_3 could be synthesized from uro'gen I, M-1 and *cbiF* has profound implications for the methylation sequence and stereochemistry of corrin biosynthesis, since stereospecific methyl transfer has been effected on an 'unnatural' substrate at an α-pyrrolic position, a process which was later found to be very similar to the way in which a new methyl group is inserted at the 11-position on the pathway to B_{12}.

At this stage, in 1990, the Rhône-Poulenc scientists, in collaboration with A. R. Battersby's group in Cambridge, using engineered *P. denitrificans*, discovered three late intermediates, precorrins 6x, 6y and 8x (Thibaut *et al.* 1990; Blanche *et al.* 1995) (figure 11), whose structures were most informative about the timing of ring

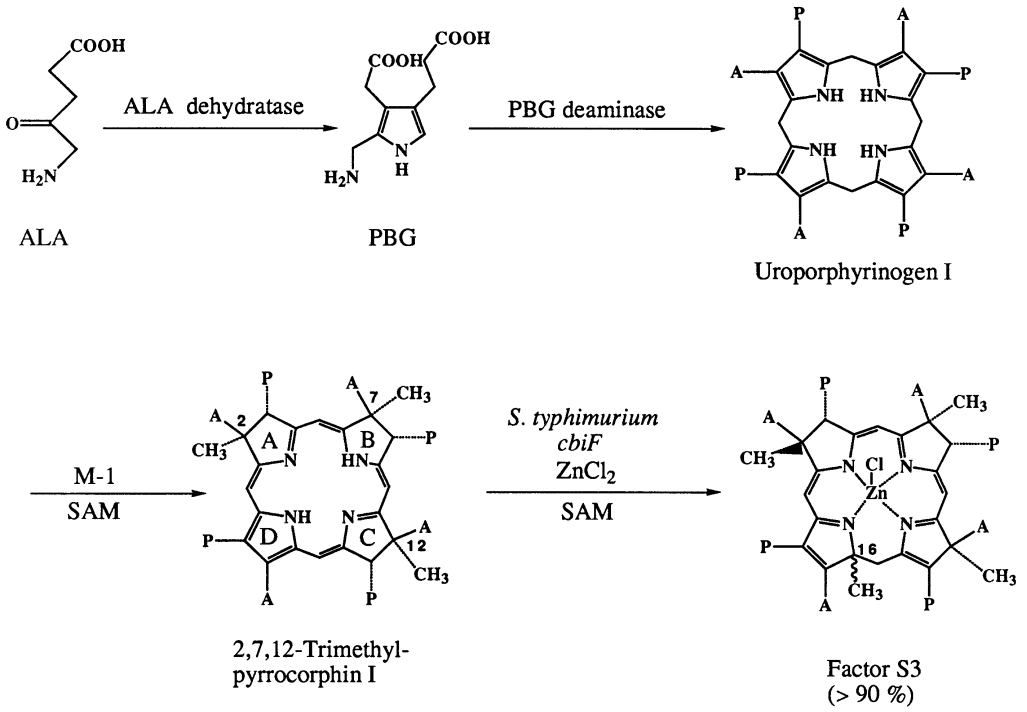


Figure 10. Multi-enzyme synthesis of factor S_3.

contraction and, most interestingly, revealed a final [1,5]-sigmatropic shift which moves the methyl at C-11 in precorrin 8x to its final resting place (C-12) in the corrin structure of hydrogenobyrinic acid (HBA).

Still hidden from view, however, was the nature of the pivotal processes whereby the porphyrinoid ring is contracted at some stage during the conversion of precorrin 3 to 6x. Ever since the discovery of the porphyrinoid–corrin connection, we had always regarded the solution of the ring contraction process as the central problem in B_{12} biosynthesis.

At this juncture, to the sequences of the genes and the necessary plasmids responsible for corrin synthesis from the anaerobic pathway in *Salmonella typhimurium* had been added the corresponding DNA sequences from the aerobic B_{12} producer *P. denitrificans* (Crouzet *et al.* 1990), thus allowing the cloning and overexpression in *E. coli* in the two sets of gene products corresponding to the ORFs shown in figure 12. We were thus in the most unusual situation in having all of the biochemical machinery necessary for the synthesis of corrins in hand without knowing the structures of the intermediates between precorrin 3 and 6x (in the aerobic series) (figure 11) or *any* of the intermediates between precorrin 3 and cobyrinic acid in the anaerobic *S. typhimurium*. Parallel studies on both sets of gene products were therefore initiated. For the anaerobic pathway, our experience with cell-free extracts of *P. shermanii* served as a guide, for we had already shown (with G. Müller) that cobalt is inserted early in this pathway, at the stage of precorrin 2, whereas in the aerobe *P. denitrificans*, cobalt is not inserted until *after* the formation of the complete corrin, HBA (figure 11). So it was already clear that profound differences existed between the aerobic and anaerobic pathways, but due to the complications arising from the necessity

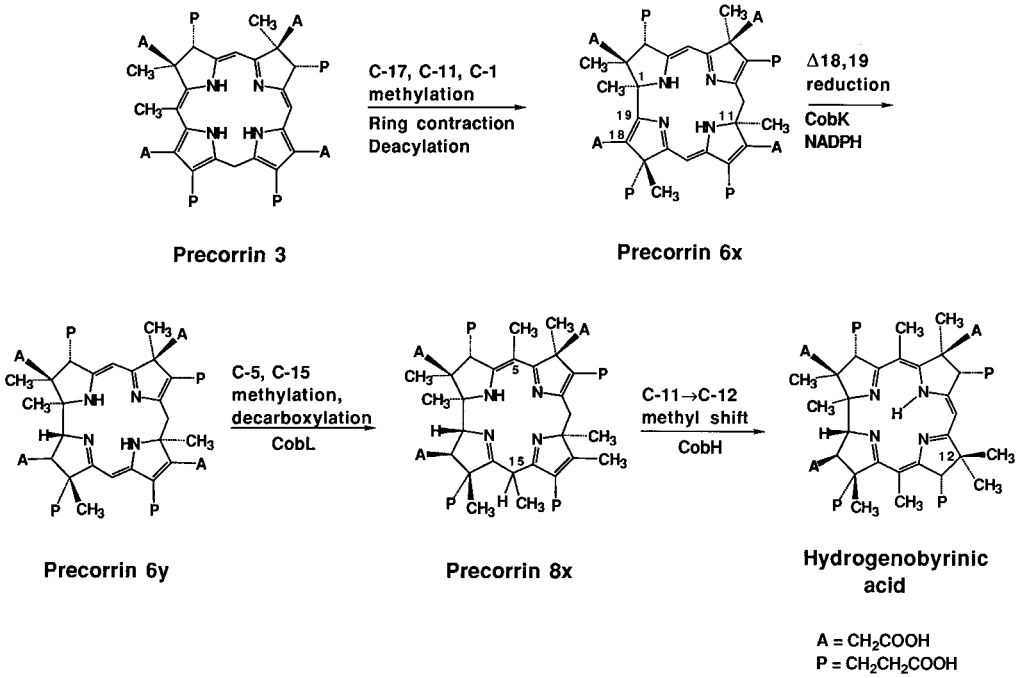


Figure 11. Precorrins-6x, -6y, -8x, recently discovered metal-free precursors of hydrogenobyrinic acid in *P. denitrificans* expression systems.

of learning to work entirely with cobalt complexes (and the attendant problems of Co^{II} NMR spectroscopy), we elected to concentrate first on the metal-free route to corrins, i.e. the aerobic pathway.

(*b*) *The steps beyond precorrin 3 in the aerobic P. denitrificans*

With the five-enzyme synthesis of precorrin 3 developed by the end of 1991 (Warren *et al.* 1992) to produce substantial amounts of ^{13}C-labelled substrate (figure 3), we undertook the systematic screening of the products from the complete set of *P. denitrificans* genes. We had earlier discovered the function of CbiF (i.e. C-11 methylation, *vide supra*), but this enzyme (from *Salmonella*) had been presented with precorrin 3 and had therefore actually handled a close relative, rather than the natural substrate, to make compound 4x (figure 4). In other words, a methyl group should already be at C-17 before C-11 methylation takes place. To make matters worse, in spite of many months of intensive search, no enzyme activity of any of the remaining *cob* genes could be demonstrated. It then became clear that an essential cofactor was missing from our incubations. It turned out that our transatlantic competitors in the B_{12} gene hunt, the research group of Rhône-Poulenc, Rorer had assumed (reasonably) that completely anaerobic conditions were necessary for enzymatic transformation through the processes leading from precorrin 3 to 6x and beyond. When the structure of yet another isolate from a blocked mutant (CobM) of engineered *P. denitrificans*, namely factor IV (oxidized precorrin 4), was discovered by the French workers (Blanche *et al.* 1995), it was immediately apparent that the ring contraction process which led to a methyl ketone pendant from C-1 must be *oxidative* in nature. We therefore revisited our collection of gene products and found to our surprise and delight that indeed the enzyme CobG, which had always been

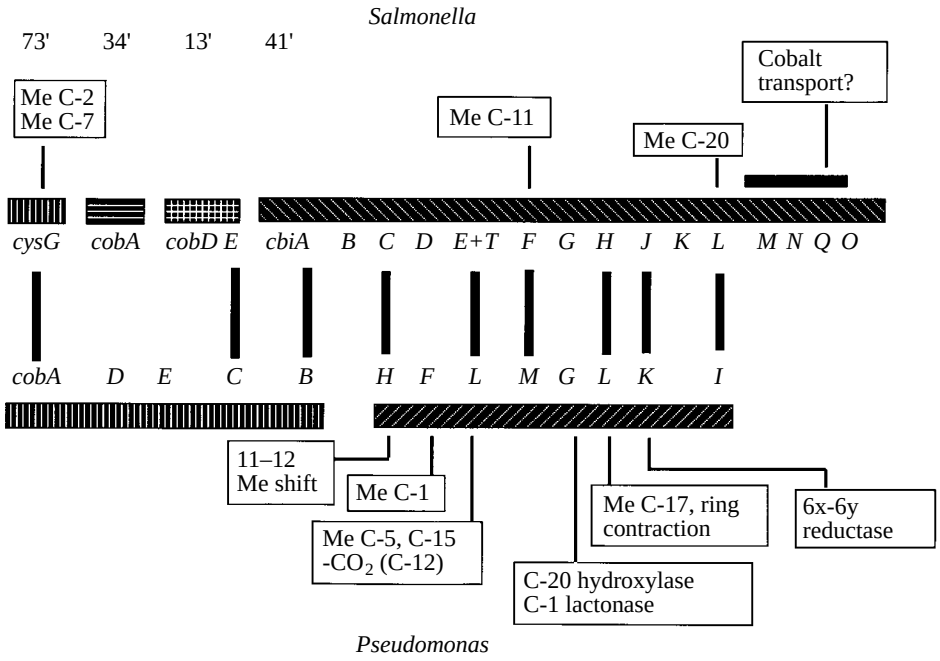


Figure 12. The location and functions of some of the known genes for B_{12} biosynthesis in *Salmonella typhimurium* and *Pseudomonas denitrificans*. In *S. typhimurium*, the genes map at 14′ (*cobD*, *cobE*; addition of aminopropanol), 34′ (*cobA*; adenosylation), 41′ (*cbiA–O*; cobinamide biosynthesis) 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.

tested in the absence of air, served as an excellent catalyst, but only in the presence of O_2 and NADH, not for the ring contraction as anticipated, but for the insertion of one atom of dioxygen into the 20-position of precorrin 3, concomitant with participation of the ring A-acetate in γ-lactone formation terminating at C-1 (figure 13) to yield precorrin 3x (Scott *et al.* 1993). It was then possible to go forward from 3x to the ring-contracted precorrin 4, a process which is formally seen as an acyloin reaction, triggered by methylation at C-17 and catalysed by CobJ, thus discovering nature's wonderful spring-loaded device for oxidative ring-contraction (figure 8). Quite independently, the Rhône-Poulenc group described the formation and structure of precorrin 3x, which they named precorrin 3B. We found that O_2 is required as part of the catalytic machinery and that one atom of $^{18}O_2$ is incorporated at C-20 in the product, precorrin 3x (figure 13), ultimately finding its way into the carbonyl of the methyl ketone in the ring-contracted precorrin 4, as shown by the isotopic shift of ^{18}O on the ^{13}C resonances of C-20 in precorrin 3x and 4. The next step, between precorrin 4 and 6x, could then be shown to involve C-11 methylation, mediated this time by CobM ($\equiv$ CbiF) in the presence of SAM, to reach the last of the missing intermediates, precorrin 5 (Min *et al.* 1993) (figure 14).

The new isolate, precorrin 5, undergoes facile tautomerism and isomerizes on esterification, as seen from the NMR data of its octamethyl ester, revealing a close similarity in methylation pattern with that of factor S_3 (figure 10) in that a fourth methyl group is inserted into the α-pyrrolic position and remains in place until the penultimate step (figure 15). Thus it appears likely that the genes *cobM* and *cbiF* (which

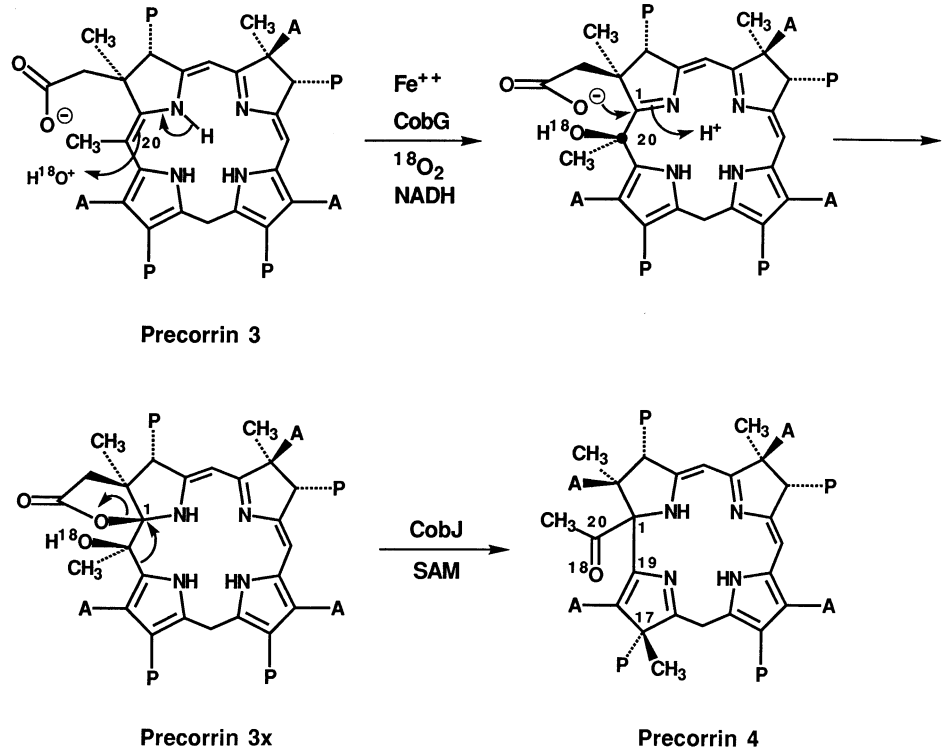


Figure 13. Possible mechanism for the formation of precorrin 3x showing fate of $^{18}O_2$ in precorrins 3x and 4.

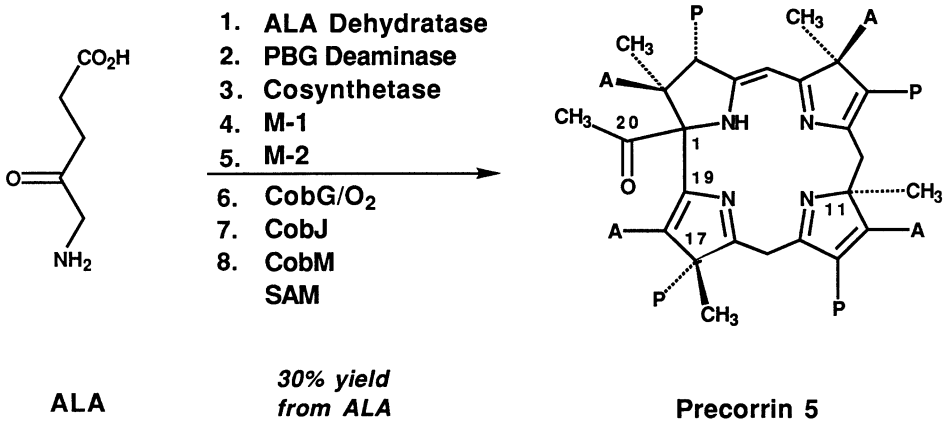


Figure 14. Multi-enzyme synthesis of precorrin 5.

show considerable sequence homology) catalyse the same reaction in their respective pathways, although we recall that *cbiF* 'mis-methylated' the wrong substrate precorrin 3, in forming 4x.

Finally, CobF was found to be yet another bifunctional enzyme, which catalyses both the deacetylation and the methylation at C-1 of precorrin 5 in the presence of SAM. When CobF was included in the incubation containing precorrin 4, CobM and SAM, the signal for C-1 showed that the methyl ketone had been lost from this

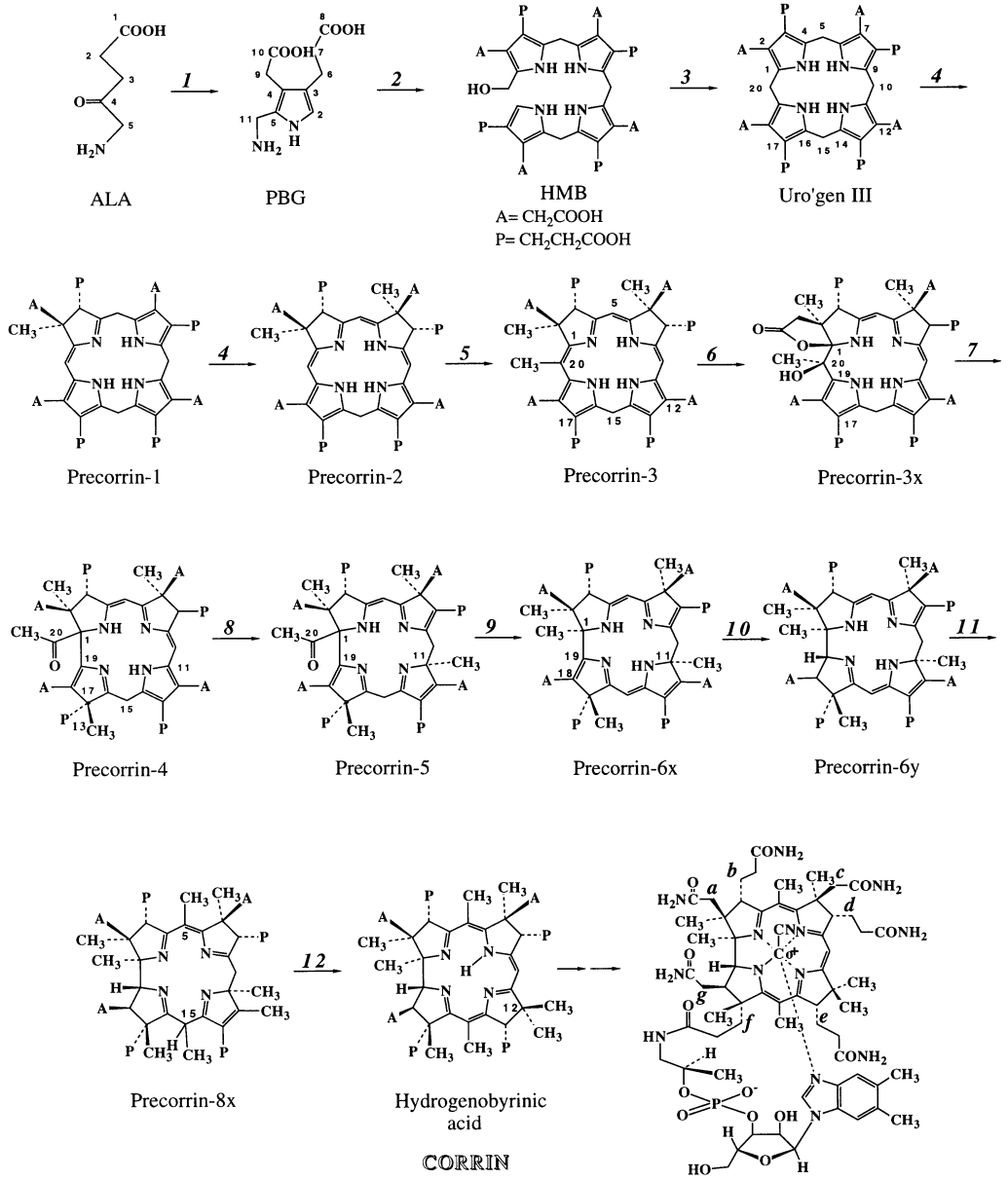


Figure 15. The biosynthetic pathway from ALA to hydrogenobyrinic acid (1995). The enzyme catalysing each reaction and its corresponding gene are: (1) δ-aminolaevulinic acid (ALA) dehydratase (*hemB*); (2) porphobilinogen (PBG) deaminase (*hemC*); (3) uroporphyringen III (uro'gen III) synthase (*hemD*); (4) uro'gen III methyltransferase (*cysG*, *cobA*); (5) precorrin-2-methyltransferase (*cobI*); (6) precorrin-3-oxidase (*cobG*); (7) precorrin-3-hydroxylactone methyltransferase (ring contraction) (*cobJ*); (8) precorrin-4-methyltransferase (*cobM*, *cbiF*); (9) precorrin-5-methyltransferase (loss of acetic acid) (*cobF*); (10) precorrin-6x-reductase (*cobK*); (11) precorrin-6y-decarboxylase and methyltransferase (*cobL*); (12) precorrin-8-methylmutase (*cobH*); (13)* cobalt insertion; (14)* esterification; (15)* add nucleotide loop; and (16)* ammonolysis ((13)*–(16)* are chemical steps).

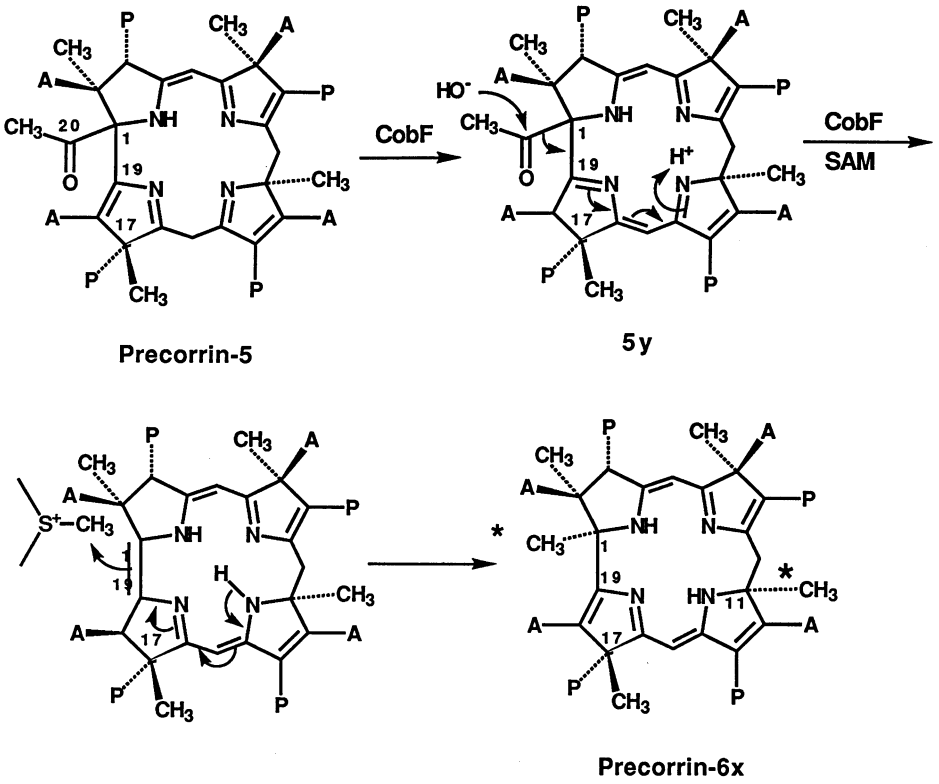


Figure 16. Conversion of precorrin 5 to precorrin 6x catalysed by CobF, showing the tautomeric form 5y.

position and had been replaced by a new C-methyl group. Esterification provided a sample with NMR and fast atom bombardment mass spectrometry data identical with those of precorrin 6x octamethyl ester. This biosynthetic correlation confirms the structure and defines the absolute stereochemistry (except at C-1) of precorrin 5. It is important to stress that, in all of these precorrins, tautomerism plays a major role in their intrinsic chemistry. In other words, the structures elucidated by NMR after isolation may correspond to the most stable tautomer which, as was found for precorrin 8x, *may not be the true biosynthetic intermediate*. Thus, the actual structure of precorrin 5 as released from its biosynthetic enzyme may well be a tautomer (5y), in which the conjugated array is prepared as an electron sink for the deacetylation step which follows (figure 16). It is also conceivable that the enzyme CobF actually prepares the correct tautomer (5y) to allow both the deacetylation and C-1 methyl transferase to operate, as suggested in figure 16. Experiments to test these ideas are in progress.

At this stage we were ready to test the efficiency of multi-enzyme synthesis by incubating ALA, SAM and using the first eight enzymes of the pathway (figure 15), and were gratified to find that the *in vitro* synthesis of precorrin 5 from ALA proceeded in *ca.* 30% overall yield (figure 16) which provided us with considerable optimism for tackling the total synthesis of the complete corrin structure. *En passant* it was possible to trace both the source and the fate of the oxygen functionalities involved at or near the site of ring contraction. Whereas the C-20 hydroxyl in precorrin 3x

is derived from one atom of dioxygen in precorrin 3x, shown by ^{18}O labelling (figure 8), and is eventually lost in the acetic acid excised from C-1 of precorrin 5, the oxygen at C-1 in precorrin 3x stems from the original ring A-acetate carboxyl and is retained in this carboxyl group throughout all of the subsequent rearrangements and processes, culminating in the structure of HBA (Stolowich *et al.* 1996; Li *et al.* 1994). In contrast, it had been found (with M. Kajiwara) that in *P. shermanii* a regiospecific exchange (with water from the medium) of a unique carbonyl function, again the ring A-acetamido group, takes place during formation of B_{12}, some time after precorrin 3, but before cobyrinic acid (see *a* in vitamin B_{12}, figure 4) which has already lost one half of its -C = ^{18}O label at this *a* position. Thus there is a subtle distinction between the role of the acetate side chain in the aerobic pathway compared with the parallel anaerobic pathway in *P. shermanii*, for the latter organism cannot use dioxygen as part of the mechanism and cobalt is inserted at an early stage, into precorrin 2. It is of considerable evolutionary interest that, while aerobes such as *P. denitrificans* carry out the synthesis of the complete corrin, HBA and its a,c-diamide *before* metal insertion, i.e. are driven from porphyrinoids to ring-contracted corrin via functionalization by O_2, the anaerobic pathway employs the two-electron valency changes of cobalt ion to mediate similar, but non-identical, steps most probably using the 'internal' oxygen of the ring A-acetate function as the source of C-20 'hydroxylation'. The pathways, in fact, diverge at precorrin 2 and do not intersect again until the a,c-diamide of cobyrinic acid is reached (Blanche *et al.* 1995; Scott 1994). Once again, thanks to the generosity displayed by the molecular biology community, we recently obtained a genomic DNA library of *P. shermanii*. Together with the complete set of B_{12} genes already cloned from *Salmonella* we are now beginning to solve the mysteries of the anaerobic pathway, whose intermediates, from precorrin 3 onwards, are cobalt complexes and whose structures and mechanisms of formation can be predicted to open up a whole new sequence of events, perhaps even more surprising than the aerobic pathway discussed above. In fact, the first of the new *P. shermanii* B_{12} intermediates, factor IV, has just been discovered (Scott *et al.* 1996) and its structure (figure 17) is indeed a surprise, for not only has ring contraction taken place but the ring A-acetate has formed a unique δ-lactone terminating at C-20 which also bears a proton. This means that, unlike C-20 in precorrins 4 and 5 which is present as a ketone (*vide supra*), the same carbon in factor IV is at the oxidation level of an alcohol and must formally lose acetaldehyde rather than acetic acid as the two-carbon fragment extruded at some later stage. Since ring contraction in the anaerobic *P. shermanii* cannot use O_2, we have suggested (Scott *et al.* 1996) that the process is initiated by attack of the ring A-carboxylate on C-20 using the valency change $Co^{III} \rightarrow Co^{I}$ as the 'electron sink'. So it is clear that the ancient, anaerobic route to B_{12} was able to orchestrate the ring contraction using an internal, rather than an external, source of oxygen functionality, whose installation at C-20 is necessary in both pathways, but whose origin and subsequent fate is quite distinct in aerobic versus anaerobic metabolism.

The structure of factor IV also rationalizes the unique exchange of ^{18}O label at the ring A-acetate carbonyl during anaerobic biosynthesis and the transfer of ^{18}O label to the acetic acid isolated during B_{12} biosynthesis in *P. shermanii* (Arigoni 1994). It is now clear (Scott *et al.* 1996) that the two-carbon fragment is excised as acetaldehyde, which is oxidized to acetic acid during the incubation. That nature should have discovered, and still uses, two independent routes to such a complex structure can only be regarded as fantastic.

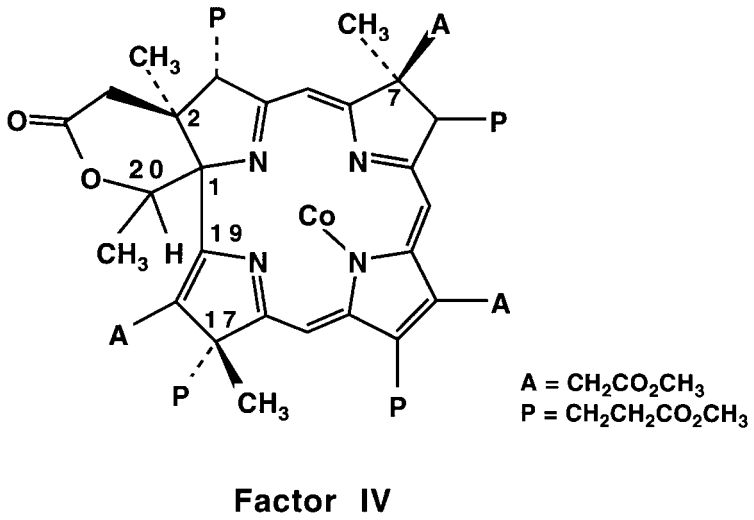


Figure 17. Factor IV, a new intermediate in the *P. shermanii* (anaerobic) pathway, a structure showing evidence of early cobalt insertion, ring contraction and a δ-lactone poised for elimination of acetaldehyde.

(*c*) *Multi-enzyme synthesis of corrins*

The availability of engineered strains of *P. denitrificans* containing the eight-gene cluster necessary for the conversion of precorrin 3 to HBA (figure 15) allowed the isolation of precorrins and determination of their structures, and cell-free lysates of this strain lacking NADPH were used to prepare precorrin 6x from precorrin 3 (Blanche *et al.* 1995). Since we had no access to the production or mutant strains of *P. denitrificans*, the philosophy and strategy of our approach had to be different. As described above, we elected to study the sequences of the gene clusters responsible for B_{12} synthesis in genomic libraries of *Salmonella* (anaerobic) and *P. denitrificans* (aerobic) to compare homologies, especially among the methyl transferases, to overexpress all biosynthetic enzymes in each series and, finally, to incubate the last known intermediate at that time, precorrin 3, strategically labelled with ^{13}C, with each enzyme in turn to find out, by on-line NMR spectroscopy, whether we had transformed the substrate to the next intermediate. The approach was painstaking because we had the necessary genes but no engineered organism to produce large amounts of intermediates. In fact, the single enzyme strategy turned out to have rewards beyond any of our early predictions. We had by now assembled all nine enzymes (by overexpression in *E. coli*) for making precorrin 6x, including the set of four catalysts (HemB, C, D, CobA) whose combination had been used earlier to prepare precorrin 3 from ALA and SAM in 80% yield (figure 10). As the next test of the reconstruction of the pathway beginning with [^{13}C]-ALA and SAM, the nine-enzyme, 12-step synthesis of precorrin 6x was found to proceed in 30% yield and was first used to prepare milligram quantities of the latter, appropriately labelled with ^{13}C, in order to monitor the conversion factor and also to have labelled substrate in hand to follow the last five steps to the complete corrin. Mutant and cell-free studies had already defined these steps, which lead from precorrin 6x, first by reduction (NADPH; CobK) to precorrin 6y, followed by 5,15-methylation and ring C decarboxylation (CobL) to reach precorrin 8x, which is finally isomerized to HBA (figure 15) by the 11 $\rightarrow$ 12 C-methyl shift and prototropic reorganization catalysed by CobH (Blanche *et al.* 1995). With

these gene products also in hand as single enzymes (Roessner *et al.* 1995), we were ready to test the most advanced version of our multi-enzyme synthesis by combining all 12 gene products to synthesize corrin. Although optimistic about the outcome, we were concerned that previous attempts to reconstruct the bioconversion of precorrin 3 to precorrin 6x, using four purified enzymes (CobG, J, M, F) proceeded in low, poorly reproducible yields. This is in sharp contrast to the incubation of precorrin 3 with a cell-free extract of the engineered strain of *P. denitrificans*, which, as described above, afforded precorrin 6x in 20–40% yield using conditions which fortuitously contained enough oxygen to catalyse the CobG reaction. These results highlight the difference between using cell-free extracts which contain many enzymes and cofactors in addition to those required for the biosynthesis of the target, and the multi-enzyme technique, where only the enzymes and cofactors necessary for the synthesis are employed, i.e. it is necessary in the former case to choose a pathway where the intermediates are not diverted, while in the latter, multi-enzyme approach, the proper cofactors must be added. In the case of the low-yielding synthesis of precorrin 6x with four enzymes, the 'missing' cofactor turned out to be oxygen. It is essential to make the above distinction, since multi-enzyme synthesis is a logical procedure using combinations of purified enzymes for the total synthesis of complex natural products, employing a series of well-defined steps, whereas cell-free incubation of engineered organisms (and their mutants) allows preparation of intermediates and final products without necessarily having prior knowledge of the steps involved, and in many instances (other than corrin biosynthesis) the overall conversions using cell-free extracts are extremely low, due to the diversion of the intermediates along the major arteries of metabolism, which can still operate in a cell-free system, since the soluble enzymes necessary for cellular functions are released on lysis. In spite of the possibility that unfavourable substrate/product concentrations could have inhibited several of the enzyme-catalysed steps, to our great satisfaction, the overexpressed enzymes in combination behaved towards substrates and cofactors *in vitro* just as they do in the living cell. It was an exciting moment to see the final corrin target, HBA, being reached before our eyes in 20% overall yield, based on the five-carbon starting material, ALA, when all 12 enzymes were added to a single flask containing ^{13}C-ALA, SAM, NADPH (in buffer, pH 7) under aerobic conditions for 12 h at 30 °C. The resultant ^{13}C-enriched HBA showed the expected set of eight signals in the ^{13}C-NMR spectrum (corresponding to incorporation of eight units of [4-^{13}C]-ALA and, for final characterization, was converted by chemical insertion of cobalt to cobyrinic acid (figure 15), showing complete identity (NMR, CD, MS) with the natural product. The latter serves as a relay to B_{12} since it has already been converted in *ca.* 40% yield to the vitamin (Eschenmoser 1988). Thus the chemo-enzymatic synthesis of vitamin B_{12} was completed in two phases—a 17-step 12-enzyme synthesis of the relay HBA (Roessner *et al.* 1994) and a four-step chemical conversion of the latter to cyanocobalamin, already in place thanks to the work of Podschun & Müller (1985) and Eschenmoser (1988).

On our journey from ALA to the vitamin, we have encountered many unexpected twists and turns of structure and mechanism and can only marvel at the fantastic plan evolved by nature to construct the corrin macrocycle, including the oxidative, spring loading of the ring contraction mechanism, the ring contraction itself, the deacetylation mechanism and finally the methyl migration from C-11 to C-12. Perhaps for the organic chemist, the most satisfying result is that the entire pathway can be reconstituted in a single test-tube, i.e. by removing the cell wall and replacing it

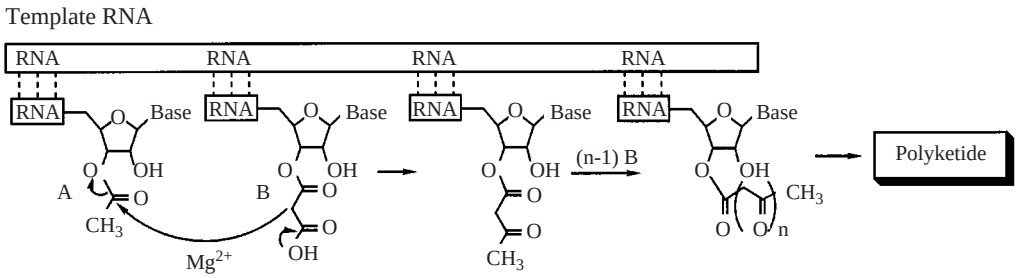


Figure 18. A possible mechanism for polyketide synthesis in the RNA world.

with glass, the soluble, biosynthetic enzymes still behave in perfect concert, allowing the synthesis to continue unaffected by the *in vitro* conditions, which might have militated against the fidelity of the stepwise coupling activities of no less than 12 enzymes.

We regard the successful conclusion of the synthesis of B_{12} not so much as the end of a chapter in biosynthetic exploration, but rather as the beginning of a whole new era in genetically engineered synthesis of complex natural products, regardless of their origin, since it is now possible to prepare and express cDNA or genomic libraries from any species of plant or organism and, as described in the earlier part of the lecture, progress has already been made in the synthesis of rare alkaloids and carbohydrates by this approach.

(*d*) *Epilogue*

In considering the anaerobic pathway to B_{12}, it has recently occurred to the author that since the most primitive organisms, e.g. the Archaeae and methanogens, dating from *ca.* three billion years may have been preceded by a 'breakthrough' organism which used RNAs rather than proteins as catalysts (Benner *et al.* 1989), which would have to produce the necessary membranes to build and maintain a cellular metabolism. I have suggested elsewhere (Scott 1997) that the catalytic machinery of RNA could support a primordial polymerization of acetate, malonate and isoprene units to achieve lipid biosynthesis. The necessary chemical analogy already exists for polyketide synthesis (Scott *et al.* 1975). Moreover, the ribosomal system using poly U as the template has been used to make polyesters from $tRNA^{Phe}$ misacylated with α-hydroxy acids (Fahnestock & Rich 1971). Extension of this concept also rationalizes the formation of porphyrins catalysed by RNA where the first step (catalysed by hemA), using glutamyl-$tRNA^{glu}$ as the substrate, has been conserved in contemporary organisms (Huang *et al.* 1984). The suggested catalytic process for polyketide synthesis on a RNA template in shown in figure 18.

It is a pleasure to pay tribute to the contributions of an enthusiastic team of graduate students, post-doctoral fellows and technical assistants whose names are mentioned in the references and the reviews (Scott 1993, 1994). Over the years we have enjoyed continuing and fruitful collaborations with Professor G. Müller (Stuttgart), Professor M. Kajiwara (Tokyo), Professor J. Roth (Utah) and Professor P. M. Jordan (Southampton).

Last, by no means least, my profound debt to the senior members of the group: Dr C. A. Roessner (genetics/enzymology), Dr N. J. Stolowich (enzymology/NMR) and Dr H. J. Williams (NMR) is gratefully acknowledged.

We thank the National Institute of Health, the Robert A. Welch Foundation and the Texas Advanced Research Program for generous financial support.

References

Arigoni, D. 1994 *CIBA Foundation Symposium 180. The biosynthesis of the tetrapyrrole pigments*, pp. 280–281. Chichester: Wiley.

Bailey, J. E. 1991 Toward a science of metabolic engineering. *Science* **252**, 1668–1675.

Battersby, A. R. 1994 How nature builds the pigments of life: the conquest of vitamin B_{12}. *Science* **264**, 1551–1557.

Benner, S. A., Ellington, A. D. & Tauer, A. 1989 Modern metabolism as a palimpsest of the RNA world. *Proc. Natn. Acad. Sci. USA* **86**, 7054–7058.

Blanche, F., Cameron, B., Crouzet, J., Debussche, L., Thibaut, D., Vuilhorgne, M., Leeper, F. J. & Battersby, A. R. 1995 Vitamin B_{12}: how the problem of its biosynthesis was solved. *Angew. Chem. Int. Ed. Engl.* **34**, 383–411.

Crouzet J, Cameron, B., Cauchois, L., Blanche, F., Rigault, S., Rouyez, M.-C., Thibaut, D. & Debussche, L. 1990 Genetic and sequence analysis of an 8.7-kilobase *Pseudomonas denitrificans* fragment involved in transformation of precorrin 2 to cobyrinic acid. *J. Bact.* **172**, 5980–5990.

Devagupta, R. 1994 Studies on the biosynthesis of indole alkaloids in *Catharanthus roseus*. Ph.D. thesis, Texas A&M University.

Duggan, P. J., Parker, E., Coggins, J. & Abell, C. 1995 Enzymatic synthesis of (6R)- and (6S)-fluoroshikimic acids. *Biorg. Med. Chem. Lett.* **5**, 2347–2352.

Eschenmoser, A. 1988 Vitamin B_{12}: experiments concerning the origin of its molecular structure. *Angew. Chem. Int. Ed. Engl.* **27**, 5–29.

Fahnestock, S. & Rich, A. 1971 Ribosome-catalyzed polyester formation. *Science* **173**, 340–343.

Fessner, W.-D. & Walter, C. 1992 'Artificial metabolisms' for the asymmetric one-pot synthesis of branched chain saccharides. *Angew. Chem. Int. Ed. Engl.* **31**, 614–616.

Huang, D.-D., Wang, W.-Y., Gough, S. P. & Kannangara, C. G. 1984 Glutamyl-t-RNA as a precursor for ALA. *Science* **225**, 1482–1484.

Kren, V. & Thiem, J. 1995 A multi-enzyme system for a one-pot synthesis of sialyl T-antigen. *Angew. Chem. Int. Ed. Engl.* **34**, 893–895.

Kutchan, T. M. 1993 Strictosidine: from alkaloid to enzyme to gene. *Phytochemistry* **32**, 493–506.

Kutchan, T. M., Dittrich, H., Bracher, D. & Zenk, M. H. 1991 Enzymology and molecular biology of alkaloid biosynthesis. *Tetrahedron* **47**, 5945–5954.

Li, Y., Alanine, A. I. P., Vishwakanna, R. A., Balachandran, S., Leeper, F. J. & Battersby, A. R. 1994 Biosynthesis of vitamin B_{12}: mechanistic studies of the transfer of a methyl group from C-11 to C-12 and incorporation of ^{18}O. *J. Chem. Soc. Chem. Commun.* 2507–2508.

Lindquist, L., Schweda, K. H., Reeves, P. R. & Lindberg, A. A. 1994 *In vitro* synthesis of CDP-D-abequose using *Salmonella* enzymes of cloned *rfb* genes. *Eur. J. Biochem.* **225**, 863–872.

McDaniel, R., Ebert-Kosla, S., Hopwood, D. A. & Khosla, C. 1993 Engineered biosynthsis of novel polyketides. *Science* **262**, 1546–1550.

Marumo, K., Lindquist, L., Verma, N., Weintraub, A., Reeves, P. R. & Lindberg, A. 1992 Enzymatic synthesis and isolation of thymidine diphosphate 6 deoxy-d-xylo 4-hexulose and thymidine diphosphate rhamnose. *Eur. J. Biochem.* **204**, 539–545.

Min, C., Atshaves, B. A., Roessner, C. A., Stolowich, N. J., Spencer, J. B. & Scott, A. I. 1993 Isolation, structure, and genetically engineered synthesis of precorrin 5, the pentamethylated intermediate of vitamin B_{12} biosynthesis. *J. Am. Chem. Soc.* **115**, 10 380–10 381.

Podschun, T. & Müller, G. 1985 Hydrogenobyrinic acid and vitamin B_{12}. *Angew. Chem. Int. Edn. Engl.* **24**, 46–48.

Roessner, C. A., Warren, M. J., Santander, P. J., Atshaves, B. P., Ozaki, S.-I., Stolowich, N. J., Iida, K. & Scott, A. I. 1992 Expression of 9 *Salmonella typhimurium* enzymes for cobinamide biosynthesis: identification of the 11-methyl and 20-methyl transferases of corrin biosynthesis. *FEBS Lett.* **301**, 73–78.

Roessner, C. A., Spencer, J. B., Stolowich, N. J. & Santander, P. J. 1994 Genetically engineered synthesis of precorrin 6x and the complete corrinoid, hydrogenobyrinic acid, an advanced precursor of vitamin B_{12}. *Chem. Biol.* **1**, 119–124.

Roessner, C. A., Spencer, J. B., Ozaki, S., Min, C., Atshaves, B. P., Nayer, P., Anoussis, N., Stolowich, N. J., Holderman, M. T. & Scott, A. I. 1995 Overexpression in *Escherichia coli* of 12 vitamin B_{12} biosynthetic enzymes. *Prot. Expr. Purif.* **6**, 155–163.

Roth, J. R., Lawrence, J. G., Rubenfield, M., Kieffer-Higgens, S. & Church, G. M. 1993 Characterization of the cobalamin (vitamin B_{12}) biosynthetic genes of *Salmonella typhimurium*. *J. Bact.* **175**, 3303–3316.

Scott, A. I. 1993 How nature synthesizes vitamin B_{12}. A survey of the last 4 billion years. *Angew. Chem. Int. Edn. Engl.* **32**, 1223–1243.

Scott, A. I. 1994 The discovery of nature's pathway to vitamin B_{12}. A 25 year odyssey. *Tetrahedron* **50**, 13 313–13 333.

Scott, A. I. 1997 How were porphyrins and lipids synthesized in the RNA world? *Tetrahedron Lett.* **38**, 4961–4964.

Scott, A. I., Wiesner, C. J., Yoo, S. & Chung, S.-K. 1975 Acetyl transfer reaction in catechol acetate malonate. A model for the biosynthesis of polyketides and fatty acids. *J. Am. Chem. Soc.* **97**, 6217–6219.

Scott, A. I., Roessner, C. A., Stolowich, N. J., Spencer, J. B., Min, C. & Ozaki, S.-I. 1993 Biosynthesis of vitamin B_{12}: discovery of the enzymes for oxidative ring contraction and insertion of the fourth methyl group. *FEBS Lett.* **331**, 105–108.

Scott, A. I., Stolowich, N. J., Wang, J., Gawatz, O., Fridrich, E. & Müller, G. 1996 Biosynthesis of vitamin B_{12}: factor IV, a new intermediate in the anaerobic pathway. *Proc. Natn. Acad. Sci. USA* **93**, 14316–14319.

Spencer, J. B., Stolowich, N. J., Roessner, C. A. & Scott, A. I. 1993 The *Escherichia coli cysG* gene encodes the multifunctional protein, siroheme synthase. *FEBS Lett.* **335**, 57–60.

Stolowich, N. J., Wang, J., Spencer, J. B., Santander, P. J., Roessner, C. A. & Scott, A. I. 1996 Absolute stereochemistry of precorrin 3x and its relevance to the dichotomy of ring contraction mechanism in vitamin B_{12} biosynthesis. *J. Am. Chem. Soc.* **118**, 1657–1662.

Thibaut, D., Blanche, F., Debussche, L, Leeper, F. J. & Battersby, A. R. 1990 Biosynthesis of vitamin B_{12}: structure of precorrin 6x octamethyl ester. *Proc. Natn. Acad. Sci. USA* **87**, 8800–8804.

Wang, J., Stolowich, N. J., Santander, P. J., Park, J. H. & Scott, A. I. 1996 Biosynthesis of vitamin B_{12}: concerning the identity of the two-carbon fragment eliminated during anaerobic formation of cobyrinic acid. *Proc. Natn. Acad. Sci. USA* **93**, 14320–14322.

Warren, M. J., Roessner, C. A., Ozaki, S.-I., Stolowich, N. J., Santander, P. J. & Scott, A. I. 1992 Enzymatic synthesis and structure of precorrin 3, a trimethyl dipyrrocorphin intermediate in vitamin B_{12} biosynthesis. *Biochemistry* **31**, 603–609.

Wong, C.-H. & Whitesides, G. M. 1994 In *Enzymes in synthetic organic chemistry* (ed. J. E. Baldwin & P. D. Magnus), pp. 19–22. New York: Elsevier.

Wong, C.-H., McCurry, S. D. & Whitesides, G. M. 1980 Practical enzymatic syntheses of ribulose-1,5-bisphosphate and ribose-5-phosphate. *J. Am. Chem. Soc.* **102**, 7938–7939.

Wong, C.-H., Halcomb, R. L., Ichikawa, Y. & Kajimoto, T. 1995 Enzymes in organic synthesis: application to the problems of carbohydrate recognition. II. Cloning and expression of glycosyl transferases and sugar nucleotide synthases. *Angew. Chem. Int. Ed. Engl.* **34**, 521–546.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES