Bioorganic Chemistry **27,** 100–118 (1999)
Article ID bioo.1998.1125, available online at http://www.idealibrary.com on IDEAL®

Vitamin B_{12}: Insights into Biosynthesis's Mount Improbable

Evelyne Raux,* Heidi L. Schubert,† Jennifer M. Roper,* Keith S. Wilson,† and Martin J. Warren*

**Department of Molecular Genetics, Institute of Ophthalmology, University College London, Bath Street, London EC1V 9EL; and †X-Ray Crystallography Group, Department of Chemistry, University of York, Heslington, York YO1 5DD, United Kingdom*

Received August 24, 1998

The biosynthesis of the corrin ring component of cobalamin (vitamin B_{12}) is reviewed with regard to how two separate though broadly similar pathways may have evolved. The more ancient "anaerobic" pathway is characterized by the early chelation of cobalt and the release of acetaldehyde whereas the "aerobic" pathway is characterized by an absolute dependency on molecular oxygen, the late chelation of cobalt and the release of acetic acid. Both pathways require the addition of 8 *S*-adenosyl-L-methionine-derived methyl groups to the periphery of the tetrapyrrole framework. The sequences of these enzymes reveal that they are clearly related, most likely having evolved from an ancestral methylase gene. The three-dimensional structure of one of these methyltransferases is highlighted and discussed in light of a common mechanism for this family of enzymes. Moreover, the aerobic and anaerobic chelatases are described and parallels with the chelatases found in heme and chlorophyll synthesis are drawn.

INTRODUCTION

Of all Nature's tetrapyrrole-derived metallo-cofactors, which include heme, chlorophyll, coenzyme F_{430}, and siroheme, perhaps the most fascinating remains cobalamin (vitamin B_{12}), not least because of the beguiling chemistry which it is able to mediate but also because it is, itself, so structurally complex (*1*). Cobalamin is found largely in one of two common biological forms, methylcobalamin and adenosylcobalamin, which are utilized for methylation and rearrangement reactions, respectively (*2–5*).

It has been argued that cobalamin represents an evolutionarily ancient coenzyme that predates the other tetrapyrrolic-based cofactors and was utilized by the earliest forms of life (*6–9*). This hypothesis is based on the observation that adenosylcobalamin is an RNA cofactor that can also act as a coenzyme for class II ribonucleotide reductases in certain members of the Archaea (*10*), therefore playing an essential part in the evolution of DNA based life from the RNA world. Moreover, certain strict anaerobes are able to synthesize corrinoids but have no role for porphyrinoid hemes (*7*). Likewise the description of a prebiotic synthesis of corrins adds further weight to their incorporation into the metabolism of emerging life forms (*6*).

In fact, living systems can be divided into three groups with respect to cobalamin metabolism: (i) those which neither make nor require cobalamin, (ii) those which

0045-2068/99 $30.00

require cobalamin and have the capacity to make their own, and (iii) those who require it but are unable to make it themselves (*11*). Humans fall into the final category since cobalamin is a necessary component for the enzymes methionine synthase and methylmalonyl CoA mutase, defects in which are associated with homocystinurea, megaloblastic anaemias, and pernicious anemia (*12*,*13*).

Several notable systems have evolved to live in a cobalamin-free world, unable to make the coenzyme and not requiring its properties, of which higher plants and yeasts are perhaps the most prominent. The world appears to rely on bacteria for cobalamin synthesis since cobalamin producers appear to be largely restricted to the prokaryotic fraternity. No genetic evidence for cobalamin synthesis has been obtained from any eukaryotic source.

The biological forms of cobalamin represent one of the most complex "small" molecules made by living systems (Fig. 1). They consist of a corrinoid ring, a contracted porphinoid that is missing the bridging carbon between rings A and D. The corrin ring is physically attached to the lower α-ribazole by an aminopropanol linker. The four nitrogens of the corrin ring and the nitrogen of the dimethylbenzimidazole base of the α-ribazole act as ligands for a centrally chelated cobalt ion. Either a methyl or adenosyl group supply the upper, sixth, ligand (Fig. 1). It is the nature of this cobalt–carbon bond and its intrinsic ability for homolytic cleavage that is at the center of much of the unprecedented biochemistry associated with B_{12}. However, the biosynthesis of the corrin component of cobalamins has also attracted the interest of many scientists over the past 50 years and it is to this metabolic aspect that we now turn.

The corrin ring component of cobalamin is derived from the common tetrapyrrole

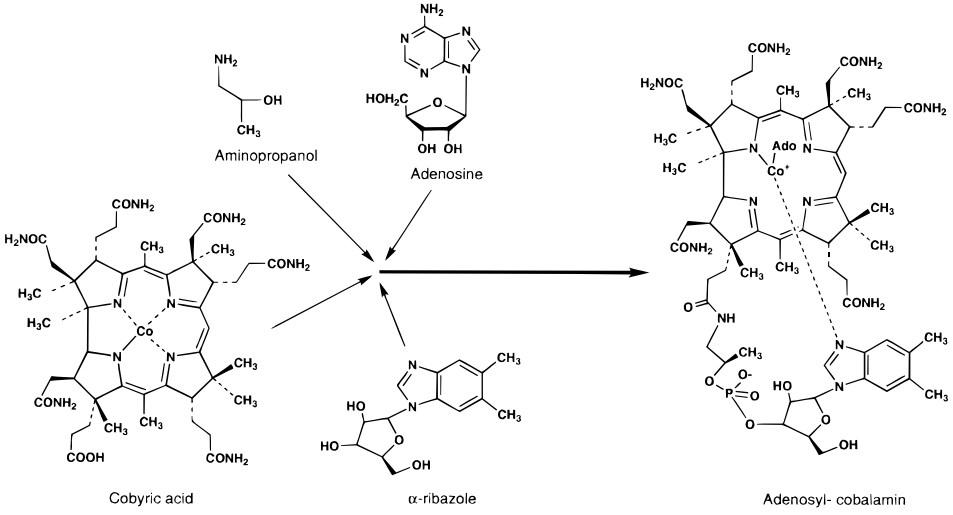

FIG. 1. Adenosylcobalamin and the constituents from which it is made. The corrin component is derived from 5-aminolevulinic acid (see Fig. 3), aminopropanol is synthesized from threonine (*74*), while the dimethylbenzimidazole base of the α-ribazole is derived from either riboflavin or erythrose and methionine.

primogenitor, uroporphyrinogen III, by a large number of enzyme-mediated processes, which include peripheral methylation, ring contraction, decarboxylation, amidation, reduction, cobalt chelation, adenosylation, and aminopropanol attachment (Fig. 2) (*14–17*). The lower base is derived from either riboflavin (*18*) or from smaller units such as erythrose and methionine (*19*). The attachment of the α-ribazole to the adenosylated corrin ring represents the final step in the synthesis of the coenzyme. More than 30 biosynthetic genes are required for the total *de novo* biosynthesis of cobalamin, which represents about 1% of a typical bacterial genome (*20*). For bacteria that make cobalamin this, therefore, represents a substantial genetic investment that must underlie the value of the coenzyme to their metabolism (*21*). Moreover, there appear to be two broadly similar, though quite distinct, routes for the synthesis of cobalamin, representing "aerobic" and "anaerobic" pathways (Fig. 2) (*22*).

The structural complexity of cobalamin and the presence of two large multistep biosynthetic pathways also represents a challenge to neoDarwinists. Using Dawkin's analogy (*23*), cobalamin appears to represent the Mount Improbable of "small" molecule biosynthesis, since its evolution is difficult to describe in terms of variation and selection. As Eschenmoser has highlighted in his masterful account (*6*), a large number of lucky prerequisites must be fulfilled for a multistep biosynthesis of a complex natural product to emerge. In this article we will look at our present understanding of corrin synthesis from an evolutionary perspective, in particular paying attention to the methylation and chelation processes. Overall, our conclusions are that the two pathways for cobalamin synthesis must have appeared early on, prior to any rigid imprinting of substrate recognition by the biosynthetic enzymes. Moreover, some of the biochemical machinery that had evolved for corrin biosynthesis has been subsequently adopted and adapted for the biosynthesis of other modified tetrapyrroles such as hemes and chlorophylls.

COBALAMIN BIOSYNTHESIS

Adenosylcobalamin is biosynthesized from uroporphyrinogen III. This unsymmetrical macrocyclic isomer is, itself, made from eight molecules of 5-aminolevulinic acid by the action of three enzymes, 5-aminolaevulinic acid dehydratase (HemB),[1] porphobilinogen deaminase (HemC)[2] and uroporphyrinogen III synthase (HemD) (Fig. 3) (*24*). In just these three steps the amino ketone is transformed into the molecular framework from which all modified tetrapyrroles are derived (*25*). This involves, initially, the dimerization of two molecules of ALA to give the monopyrrole, porphobilinogen (*26,27*), followed by the polymerization of four molecules of PBG to yield a linear tetrapyrrole, preuroporphyrinogen (hydroxymethylbilane)

[1] In humans this zinc-dependent enzyme is inhibited by lead. Therefore, one of the consequences of lead poisoning is inhibition of the heme biosynthetic pathway and causes a pseudoporphyria. Such a disorder may have been a contributing factor to the fall of the Roman Empire as well as the disastrous Franklin expedition to the arctic in 1865 (*27*).

[2] Molecular defects in this enzyme cause acute intermittent porphyria, which has also been suggested as the underlying cause of the madness of King George III. Although it is debatable whether the king's illness played any significant part in the loss of the American colonies, it was certainly at the heart of the regency crisis of 1789. However, more recent evidence suggests that the King actually suffered with variegate porphyria, which arises from a defect in the enzyme protoporphyrinogen oxidase (*83*).

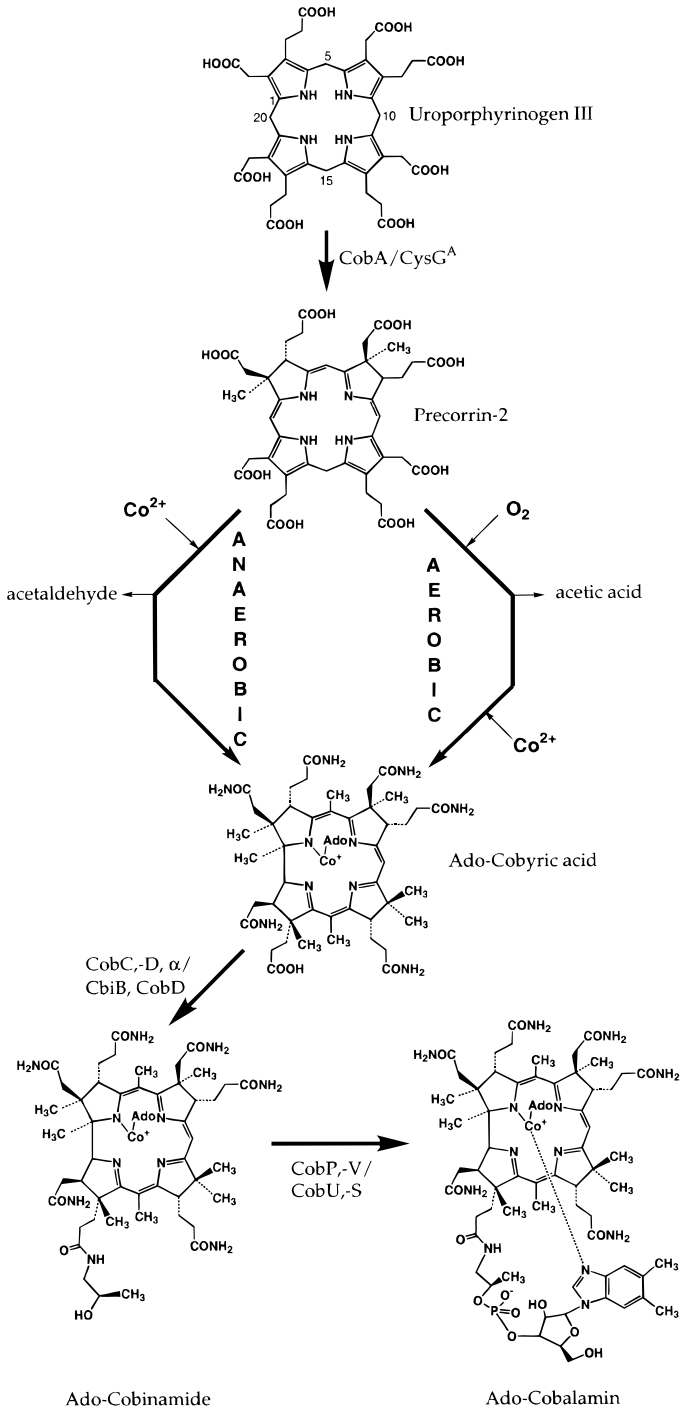

FIG. 2. Schematic representation of the aerobic and anaerobic cobalamin pathways. The two pathways separate at precorrin-2 but rejoin at adenosylcobyric acid. The molecular evolution of the two pathways poses an interesting problem, mainly because they are so similar.

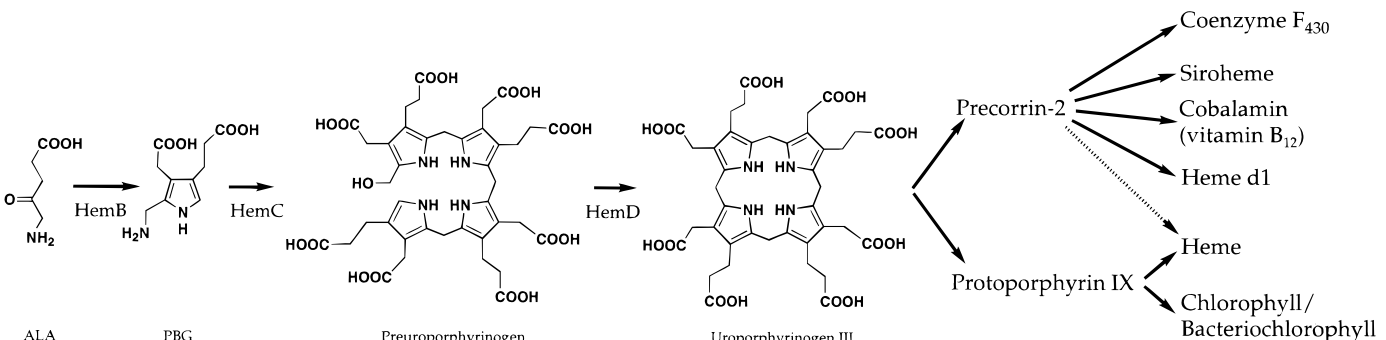

FIG. 3. Biosynthesis of uroporphyrinogen III and its central role in the synthesis of all porphinoids.

(*28*), and, finally, ring inversion of the D-ring of preuroporphyrinogen and cyclization to produce the physiologically important type III isomer of uroporphyrinogen (Fig. 3) (*29*). To direct uroporphyrinogen III toward cobalamin construction and away from the branch of the pathway that leads towards heme and chlorophyll synthesis, the intermediate is methylated by the action of an *S*-adenosyl-L-methionine (AdoMet)-dependent enzyme called uroporphyrinogen III methyltransferase (*30*). This enzyme methylates uroporphyrinogen III at positions 2 and 7, forming a dipyrrocorphin known either as precorrin-2 or dihydrosirohydrochlorin (*30–32*). Precorrin-2 represents another branchpoint in the tetrapyrrole biosynthetic pathway. Siroheme, synthesized by dehydrogenation and ferrochelation of precorrin-2 (*33,34*), coenzyme F_{430} (*35*), and heme d1 (*36*) are all derived from this intermediate (Fig. 3). Moreover, an alternative heme pathway from precorrin-2 has been recently described (*37*).

Precorrin-2 also represents the point at which the aerobic and anaerobic corrin biosynthetic pathways diverge (Fig. 2). To convert precorrin-2 into one of the later intermediates such as cobyric acid both pathways have to perform a further 6 methylations, ring contraction, decarboxylation, cobalt chelation, and amidations before they rejoin again. With the exception of cobalt chelation, both cobalamin pathways undertake these processes in a very similar order using a number of homologous enzymes. However, although there are many similarities between the aerobic and anaerobic routes there are also significant biochemical and genetic differences, which are highlighted in the following sections.

THE AEROBIC PATHWAY

Our detailed understanding of the aerobic pathway comes largely from the pioneering research undertaken by the scientists at Rhone-Poulenc Rorer in France on their commercial production organism, the rather misleadingly named *Pseudomonas denitrificans* (*14,15*). In parallel with major contributions from the laboratories of Battersby (*16*) and Scott (*17*), the complete aerobic biosynthesis of adenosylcobinamide was determined and is outlined in Fig. 4. This represents the fruits of a multidisciplinary approach to solve the problem, with microbial genetics, molecular biology and bioorganic chemistry all employed along the way. The aerobic pathway is initialized by the methylation of precorrin-2[3] at C-20 to give precorrin-3a, a trimethylated pyrrocorphin (*38,39*). This acts as the substrate for a mono-oxygenase, CobG,[4] which generates a γ-lactone with the acetate side chain on ring A of the macrocycle (*14,40*). The new intermediate, precorrin-3b, then undergoes ring contraction during methylation at C-17 in a reaction catalyzed by CobJ (*14*). This results in the formation of precorrin-4 (*41*), which is subsequently methylated at position C-11 by CobM to generate precorrin-5 (*42*). Deacylation of the extruded methylated C-20 position occurs with the action of CobF giving off acetic acid, concomitant with the methylation at C-1 (*42,43*) and yields precorrin-6a (*44*). Reduction of the macrocycle back to the level of a hexahydroporphyrin by CobK produces precorrin-6b (*45*). This acts as the substrate

[3] The numeral after the term precorrin refers to the number of methyl groups that have been added to the tetrapyrrole-derived template during cobalamin biosynthesis (*61*).

[4] Generally, the genes required for corrin biosynthesised in the aerobic pathway are prefixed with the letters *cob*, whereas the genes that encode the equivalent proteins in the anaerobic pathway are prefixed *cbi*.

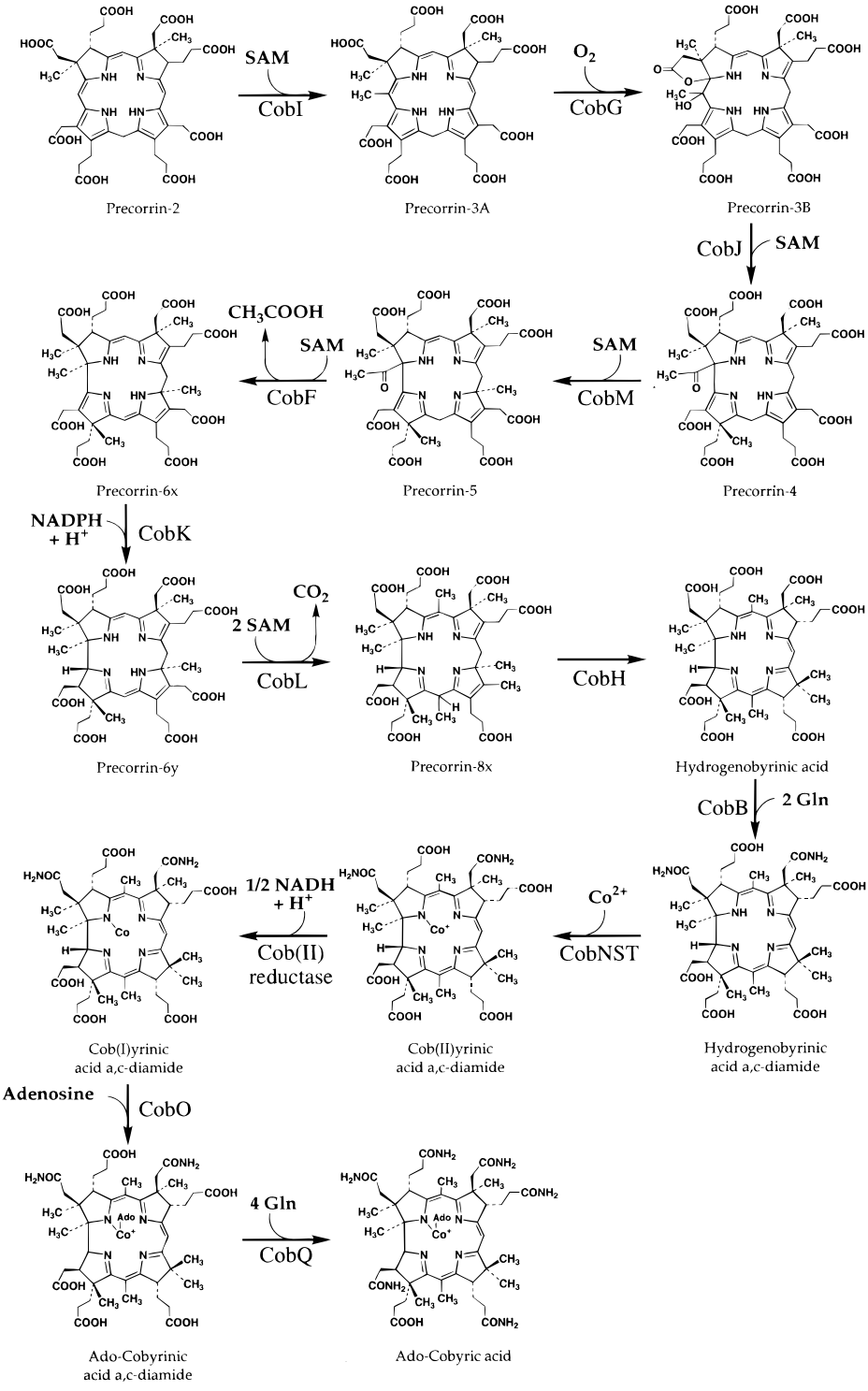

FIG. 4. The aerobic biosynthesis of adenosylcobyric acid from precorrin-2. The aerobic pathway is characterized by the requirement for molecular oxygen, the extrusion of acetic acid, and the relatively late chelation of cobalt.

for the final methyltransferase, CobL, which methylates at the two polar extremes of the macrocycle, C-5 and C-15, and also catalyses the decarboxylation of the acetate side chain on ring C (*46*). In this respect, CobL is a multifunctional enzyme but in some organisms the methylase and decarboxylation steps are undertaken by separate enzymes. A final twist in the complex synthesis of the corrin moiety is afforded by CobH, which rearranges the methyl group, originally added by CobM, from C-11 to C-12 (*47*). This generates a cobalt-free corrinoid that is referred to as hydrogenobyrinic acid. Partial amidation of this intermediate gives hydrogenobyrinic acid a,c diamide (*48*), which acts as the substrate for an ATP-dependent cobalt chelatase, an enzyme complex formed between CobN, S, and T (*49*). Once cobalt is added, yielding cobyrinic acid a,c diamide, reduction of the cobalt ion takes place (*50*) followed by adenosylation (*51*) and the four remaining amidations (*52*). This produces adenosylcobyric acid and represents the point at which the aerobic and anaerobic pathways again join. More complete accounts concerning the unraveling of this fascinating pathway are presented in a number of well-written reviews (*14–17*).

THE ANAEROBIC PATHWAY

The involvement of O_2 as a substrate in the "aerobic" cobalamin pathway immediately suggested that an alternative "anaerobic" pathway must exist since many cobalamin producers were known to be able to synthesize cobalamin in the absence of molecular oxygen. In fact, evidence for an alternative pathway was accumulating even before the elucidation of the aerobic route (*22,53*), although in comparison, proportionately less is still known about anaerobic cobalamin synthesis. Our current understanding of the anaerobic pathway comes from research undertaken on *Propionibacterium shermanii* (*17,22*), *Salmonella typhimurium* (*20,54*), and *Bacillus megaterium* (*55,56*). The anaerobic synthesis is initiated when precorrin-2 is chelated with cobalt to give cobalt-precorrin-2 (Fig. 5) (*53,54,57*). This is subsequently methylated by the C-20 methyltransferase to give cobalt-precorrin-3. The anaerobic cobalamin producers do not contain an enzyme equivalent to the aerobic CobG, the monooxygenase that forms the γ-lactone intermediate precorrin-3b. However, it has been shown that cobalt-precorrin-3 can act as a substrate for the C-17 transmethylase, CbiH, from *S. typhimurium* (*58*). Rather unexpectedly, the product of this *ex vivo* experiment, isolated after an aerobic workup, was a ring-contracted compound called factor IV, which was found to contain a δ-lactone formed from the acetate side chain of ring A. This was very similar to a compound recently isolated from a culture of *P. shermanii* called epi-factor-IV (*59*). It is assumed that the actual intermediate of this part of the pathway is not factor IV but its reduced equivalent, cobalt-precorrin-4 (*58*), and although it can be apparently formed from only cobalt-precorrin-3 and CbiH, the rather low yield of this reaction suggests that, *in vivo* at least, further components are necessary for the synthesis of the δ-lactone. Although none of the intermediates between cobalt-precorrin-4 and cobyrinic acid are known in the anaerobic pathway, it has been shown that the methylated C-20 position is subsequently lost as acetaldehyde (*60*) as opposed to acetic acid in the aerobic pathway. Significantly, the enzyme which catalyses the deacylation in *P. denitrificans*, CobF, has no equivalent in any of the anaerobic-pathway cobalamin producers. In place of the "missing" aerobic components CobG and F, the anaerobic-pathway bacteria contain proteins

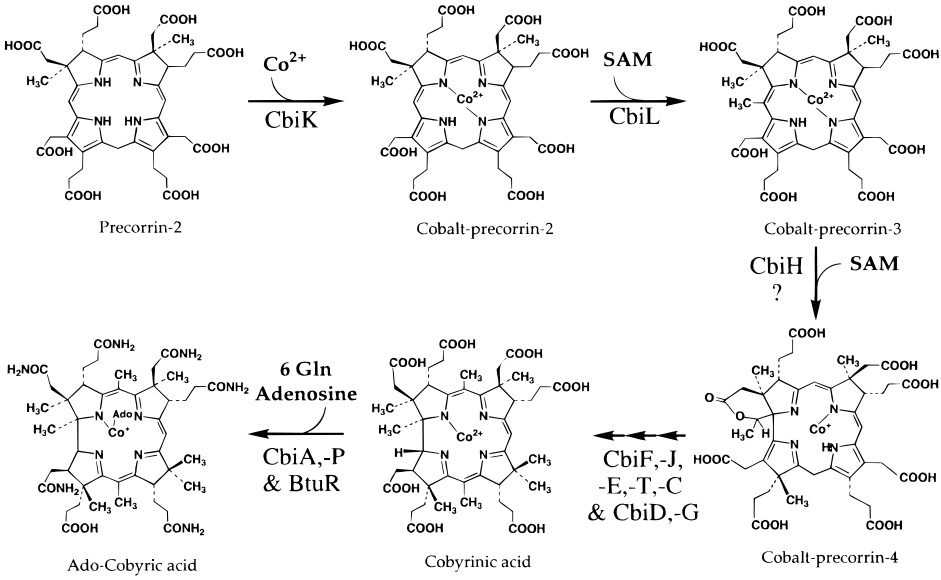

FIG. 5. The anaerobic biosynthesis of adenosylcobyric acid. In comparison to the aerobic route, relatively little is known about this pathway. However, it is characterized by the early chelation of cobalt and the extrusion of acetaldehyde.

termed CbiD and CbiG (*20*) that have been shown to be essential for the operation of the anaerobic pathway (*54–56*). However, the poor solubility of these proteins has made their study technically challenging. It is possible that these proteins may act as the anaerobic equivalents that promote the ring contraction and deacylation processes.

From cobalt-precorrin-4 we can only surmise as to the biochemical events that follow in the anaerobic synthesis, using chemical intuition and analogies with the aerobic pathway. Thus, from cobalt-precorrin-4 further methylations, extrusion of acetaldehyde, decarboxylation, and rearrangement produces cobyrinic acid that after reduction, amidations, and adenosylation yields adenosylcobyric acid. The scheme for anaerobic cobalamin biosynthesis is outlined in Fig. 5.

THE METHYLTRANSFERASES

Both aerobic and anaerobic corrin biosynthesis requires the peripheral addition of eight AdoMet derived methyl groups to the tetrapyrrole framework, a task that is mediated by the action of six methyltransferases. However, the final corrin macrocycle contains only seven of the added methyl groups since the methyl group added to C-20 is extruded during the ring contraction process. As deduced from a series of pulse labeling experiments, the methyl groups are known to be added in a specific order (C-2, C-7, C-20, C-17, C-11, C-1, C-5, C-15) (*61,62*). The methylations prevent reversible C-protonation and lead to the formation of particular tautomeric intermediates. The specific order and timing of the various transmethylations on the macrocycle and the elucidation of the individual steps of the aerobic pathway bears testimony to role of these methylations in forming the correct double bond configurations that are