Pure & Appl. Chem., Vol. 65, No. 6, pp. 1113–1122, 1993.
Printed in Great Britain.

How Nature builds the pigments of life

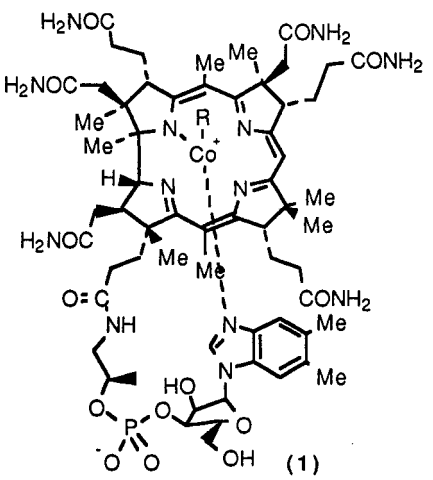
A. R. Battersby[Note a]

University Chemical Laboratory, Cambridge, CB2 1EW, United Kingdom (FAX: 0223 336400)

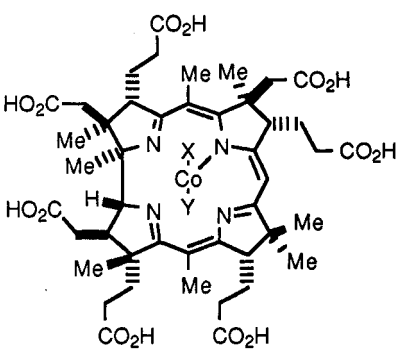
Abstract

Vitamin B_{12} (**1**) is the anti-pernicious anaemia vitamin and the coenzyme form is used by several different enzymes to bring about remarkable rearrangements in living systems. The lecture will focus on the question "How is vitamin B_{12} biosynthesised". A very brief historical survey of what had been achieved earlier will be followed by a full account of the latest exciting advances. One take-home message is that any chemist aiming to solve the biosynthesis of a complex natural product must embrace genetics, molecular biology and enzymology whilst the biologists need the complementary strengths of the most advanced chemistry. A second theme will be to demonstrate the power of experiments, based on multiple labelling of biosynthetically generated intermediates, for studying natural pathways. The lecture will describe research involving molecular biology, enzymology, synthesis, isotopic labelling and the full panoply of modern NMR techniques.

A family of coloured molecules including haem, chlorophyll, the cytochromes and vitamin B_{12} are of such vital importance for living systems that they have become known as the Pigments of Life. The present lecture will focus on just one member, the anti-pernicious anaemia agent, vitamin B_{12} (**1**). It has a structure of considerable complexity and is often referred to as the Everest of biosynthetic problems. The question of how this structure is built in Nature has excited my group since we started on this problem in 1968 and happily much has been discovered about the biosynthetic pathway. I shall first outline very briefly what knowledge was available about those parts of the pathway which had been elucidated before 1990. Time does not allow this to be complete nor to include much detail; just the *essential minimum* information will be given to allow the reader to understand and enjoy the latest exciting developments. Full information and literature references are given in various reviews. (ref. 1).



(1)



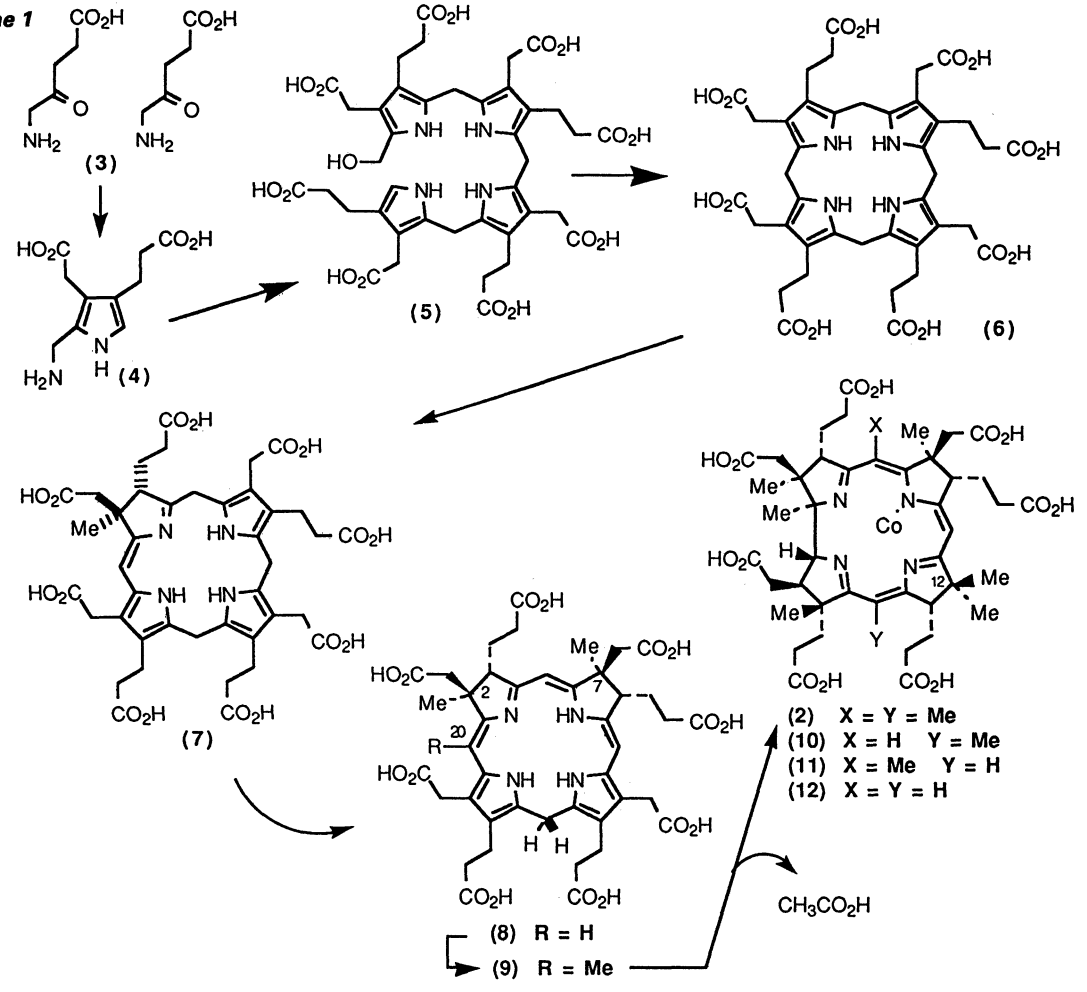
(2)

Note a The author is the spokesman for a team of close colleagues who have generously contributed their different skills and knowledge in making the major advances described in this Lecture. They are, in Paris, Francis Blanche, Beatrice Cameron, Joel Crouzet, Laurent Debussche, Denis Thibaut and in Cambridge, Fumiyuki Kiuchi, Masahito Kodera, Finian Leeper and George Weaver.

A brief history leading to 1990

It was known that the nucleotide loop of vitamin B_{12} (**1**) and the amide functions are added late in the biosynthesis and that cobyrinic acid (**2**) acts in *Propionibacterium shermanii* as a precursor of vitamin B_{12} (**1**). The problem thus becomes "How is the *corrin system* of (**2**) built". The early steps, which are also used for haem and chlorophyll up to the stage of uro'gen III (**6**), are shown in Scheme 1. Aminolaevulinic acid (**3**), ALA, is converted into porphobilinogen (**4**), PBG, four molecules of which by the action of the enzyme hydroxymethylbilane synthase (also called deaminase) afford hydroxymethylbilane (**5**). This is cyclised with rearrangement by the enzyme cosynthetase to generate uro'gen III (**6**). Comparison of structures (**6**) and (**2**) gives a measure of the many steps which are needed to convert the former into the latter. In fact, the step which deflects uro'gen III (**6**) into the B_{12}-pathway is *methylation* at C-2 and the intermediate[Note b] produced is precorrin-1 having the illustrated structure (**7**) or a close double-bond tautomer of it. The same methylase enzyme (SUMPT) responsible for producing (**7**) also adds a second methyl group to C-7 to form precorrin-2 (**8**). Then a second methylase directs the third methyl group to C-20 to generate precorrin-3 (**9**). *S*-Adenosylmethionine (SAM) is the cofactor which provides the *C*-methyl groups of (**7**), (**8**) and (**9**) and it will be seen below that the same holds true for the *C*-methyl groups which are added later.

Scheme 1



Note b Biosynthetic intermediates which precede the first-formed corrin are called *precorrins* and a number is added showing the number of *C*-methyl groups introduced from *S*-adenosylmethionine (SAM) to form that intermediate. If two or more intermediates carry the same number of SAM-derived methyl groups, then letters are added to the number to allow discrimination.

During the initial isolation work by several groups, these labile intermediates (**7**), (**8**) and (**9**) underwent dehydrogenation by air to yield the fully conjugated (aromatic) macrocycles whose structures were determined. But the reduced nature of the true intermediates was later firmly established, *e.g.* by direct isolation (ref. 2) in the case of precorrin-2 (**8**).

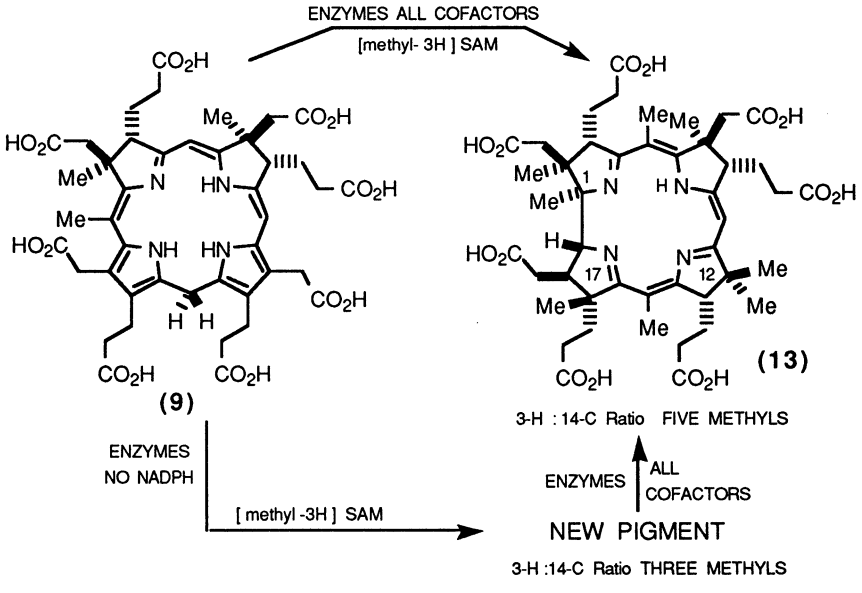
Despite great efforts, no precorrins beyond precorrin-3 (**9**) could be detected prior to 1990. But highly important progress was made in other ways. Thus, labelling studies of the ring-contraction step necessary to form the corrin ring of (**2**) from (**9**) proved that C-20 of (**9**) and the methyl group attached to it are eliminated as *acetic acid* at some later stage on the pathway. Other experiments probed the *order* in which the remaining five *C*-methyl groups are introduced during the conversion of (**9**) into (**2**). It was shown early that C-17 is the fourth methylation site and subsequently it was found that C-12α and C-1 are, respectively, the fifth and sixth sites with methylation at C-5 and C-15 being last. Interestingly, none of the three corrins lacking methyl groups at C-5 (**10**) or C-15 (**11**) or at both these centres (**12**) was enzymically converted (ref. 3) into cobyrinic acid (**2**); it will be seen later that these results fit in perfectly with the recent findings.

The latest developments

In the years preceding 1990, a major effort by Crouzet, Blanche and their colleagues (ref. 4) in France resulted in the detection and sequencing of many of the genes coding for the various enzymes which carry out the biosynthesis of vitamin B_{12} in *Pseudomonas denitrificans*. (Ref. 5). This was followed by the overproduction of these enzymes by the latest molecular biological methods. There is no doubt that this phase of the work laid the foundation on which the recent developments have been built. However, because of the nature of this Symposium, the emphasis here will be on the chemical, structural and spectroscopic research. But it must be remembered for each set of experiments that this enzymic powerhouse is always in place. So the *preparative* production by these overproduced enzymes of *e.g.* precorrin-3 (**9**) in France and the bilane (**5**), uro'gen III (**6**) or precorrin-2 (**8**) in Cambridge is now standard practice for our groups. Similarly the enzymic production of *new intermediates* on a workable scale in Paris allowed determination of their structures; these exciting results will now be outlined.

The work leading to the isolation of the first new biosynthetic intermediate started with the development of a genetically engineered strain of *P. denitrificans* in which eight of the genes involved in

Scheme 2

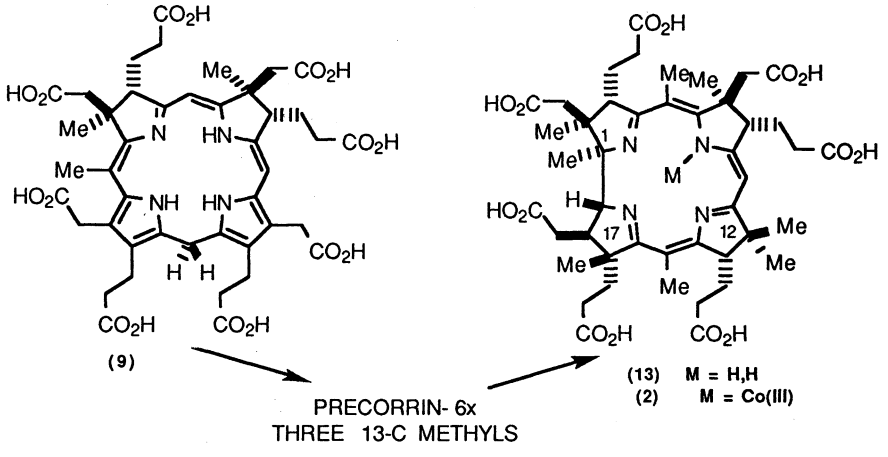


the biosynthesis of vitamin B_{12} had been overexpressed. As a result, this strain produces substantially more of the corresponding eight *normal* enzymes than is the case for the original strain and the full family of enzymes can be released by breaking the cells. The key observation was then made (ref. 6) that whereas this enzyme preparation plus a set of standard cofactors would convert early precursors into the cobalt-free corrin hydrogenobyrinic acid (**13**), omission of reduced nicotinamide adenine dinucleotide phosphate (NADPH) stopped (**13**) being formed but a new yellow pigment accumulated in its place. This new pigment in doubly-labelled form was converted in high yield into hydrogenobyrinic acid (**13**) without significant change in the labelling ratio by incubation with the enzyme system and a full set of cofactors, *i.e.* including NADPH. This demonstrated that a *new intermediate* had been isolated. Notice that the isolated enzyme preparation from *P. shermanii* produces cobyrinic acid (**2**) whilst that from *P. denitrificans* forms the cobalt-free system (**13**); the organic macrocycle is identical in the two cases.

Structure of the new intermediate

A double-labelling experiment summarised in Scheme 2 showed (ref. 6) that during the conversion of precorrin-3 (**9**) into the new intermediate, *three* new *C*-methyl groups are added to the three already present at the outset in (**9**). This showed that the new pigment had been formed by hexamethylation of uro'gen III (**6**) and it was therefore named *precorrin-6x*. A related labelling study then demonstrated that one of the original *C*-methyl groups of precorrin-3 (**9**) had been lost during its conversion into precorrin-6x, indicating that ring-contraction had already occurred at the precorrin-6x stage. The question of which three C-methyl groups of hydrogenobyrinic acid (**13**) correspond to the three new ones in precorrin-6x was answered by a neat experiment. This involved biosynthesis of precorrin-6x from precorrin-3 (**9**) using [*methyl*-^{13}C]SAM followed by further enzymic transformation of

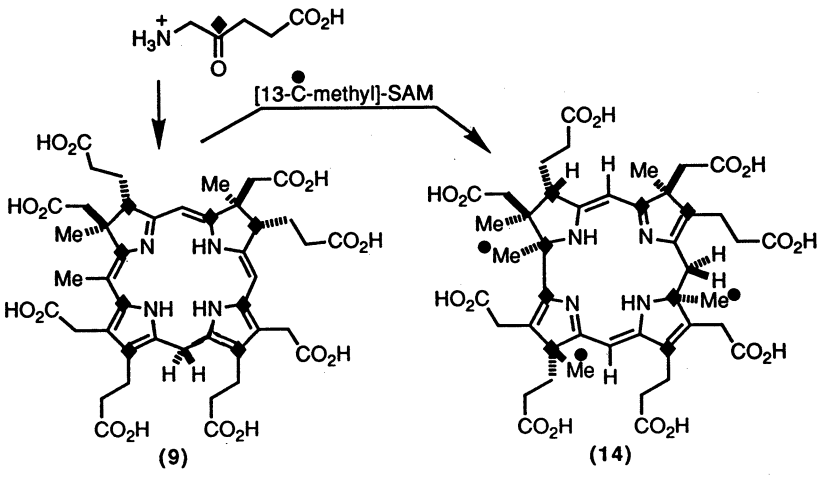
Scheme 3



the ^{13}C-labelled precorrin-6x into (**13**) now using unlabelled SAM. NMR spectroscopy (ref. 6) on a derivative of (**13**) then revealed that the *C*-methyl groups which appear in the corrin macrocycle (**13**) at C-17, C-12α and C-1 are present in precorrin-6x, Scheme 3. Notice that only the methyl groups at C-5 and C-15 of (**13**) are not present in precorrin-6x and it will be recalled that the pulse-labelling studies are in accord by showing that these are the last two methyls to be added.

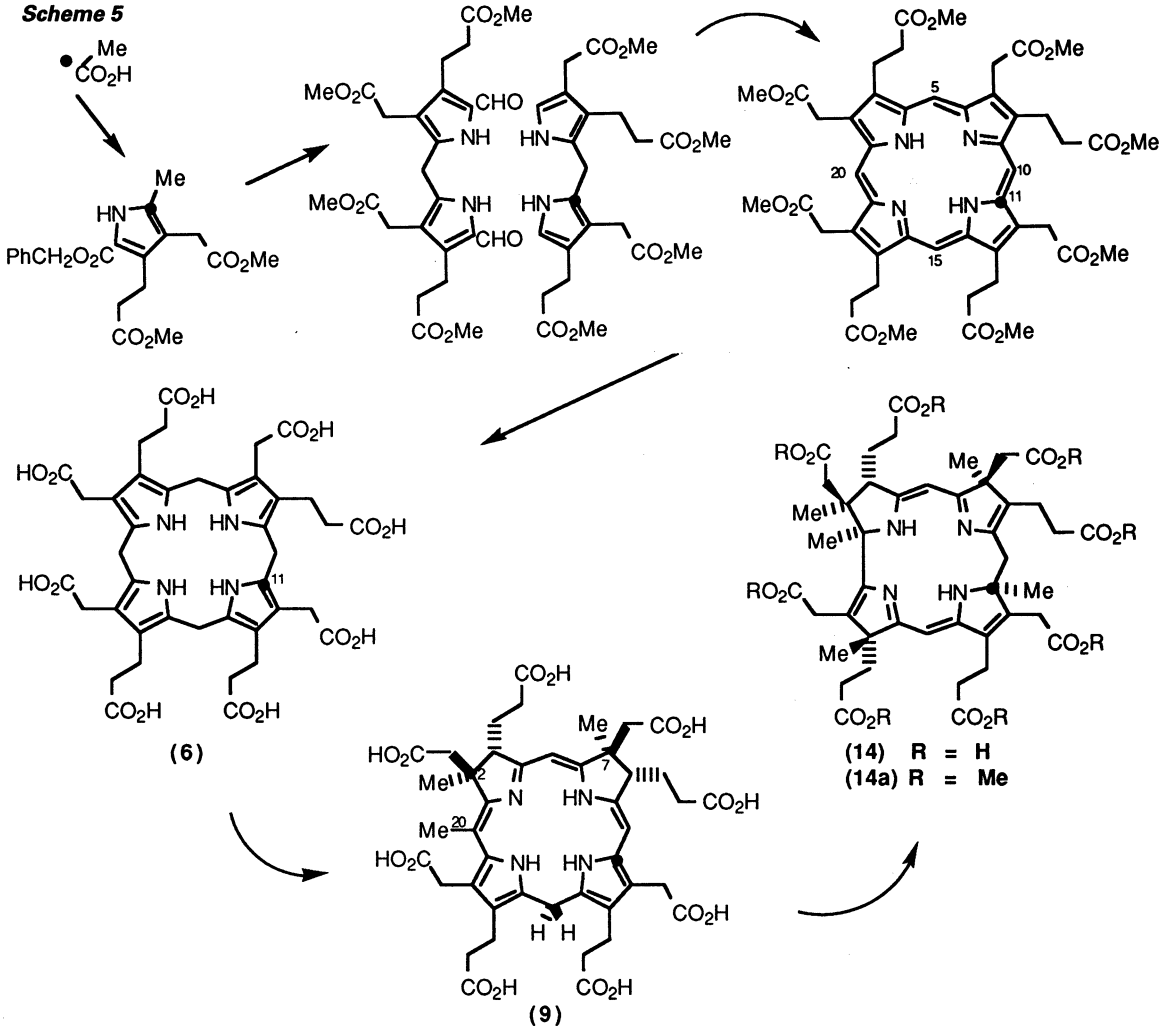
The foregoing results gave much very surprising information about precorrin-6x but more was to come from mass spectrometry which showed that precorrin-6x is an octa-acid, *i.e.* the acetic acid residue at C-12 is intact. In addition, precorrin-6x (seven double bonds) was shown to be at the oxidation level of a dehydrocorrin rather than a corrin (six double bonds), so reduction is needed at some stage of the biosynthetic pathway to generate the corrin (**13**). This fits in with the accumulation of precorrin-6x when the necessary reducing cofactor (NADPH) is excluded.

Scheme 4



All the work on precorrin-6x outlined so far was carried out entirely in Paris (ref. 6) and the great importance of this intermediate for our understanding of B_{12} biosynthesis was obvious. The chemists at Cambridge then joined the biochemists and biologists in Paris for a joint attack on the structural problem. The plan was to biosynthesise precorrin-6x starting from 5-^{13}C-ALA (see 3), also from 4-^{13}C-ALA (as **3**) and finally from 3-^{13}C-ALA (as **3**) with [*methyl*-^{13}C]SAM as the source of methyl groups in the latter two cases. The sum of these three studies was to label every carbon of the

Scheme 5



macrocycle in three interlocking patterns of ^{13}C-labels. In all cases, the precorrin-6x was isolated as its octamethyl ester.

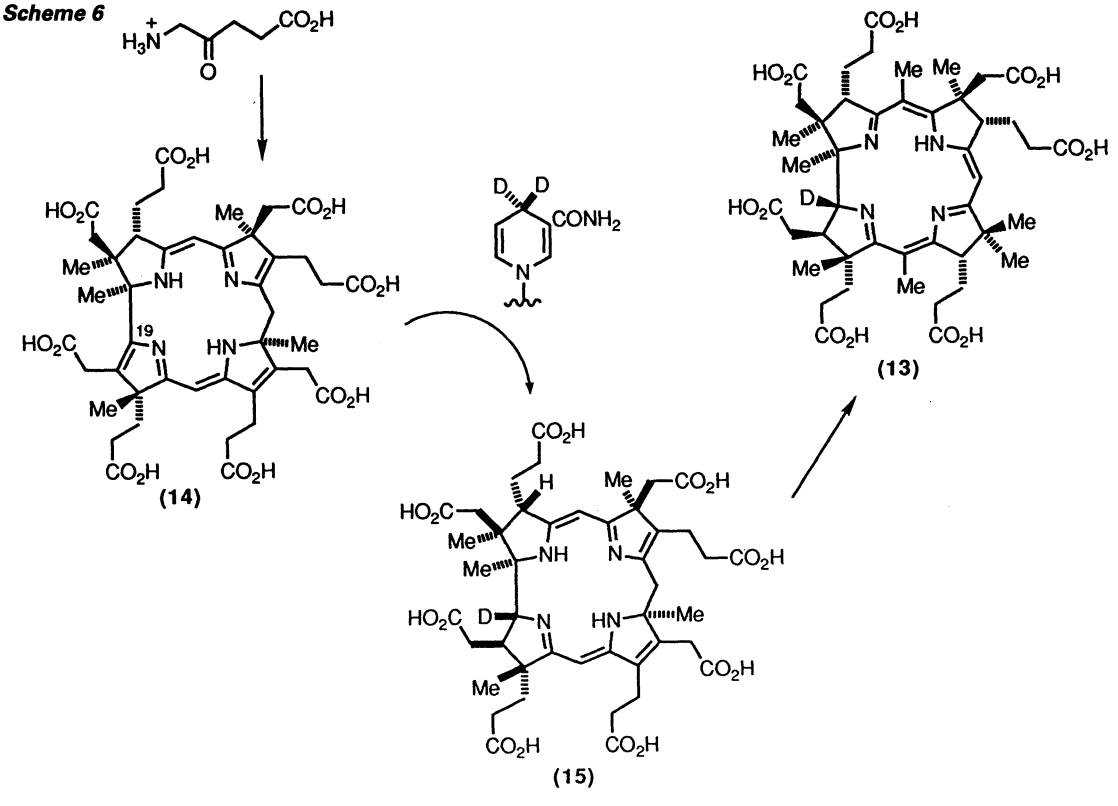
It is beyond the scope of this lecture to give a detailed analysis of the NMR assignments which depended not only on the ^{13}C-signals but also on one-bond and two/three-bond ^{1}H-^{13}C couplings with the full power of ^{1}H-NMR in addition. Suffice it to say that the combined mass of data showed (refs. 6, 7) that precorrin-6x octamethyl ester has the highly unexpected structure (**14a**). Scheme 4 shows the labelling pattern from the experiment with 4-^{13}C-ALA and [*methyl*-^{13}C]SAM. The most surprising features of precorrin-6x are that (*a*) ring-contraction has already occurred (*b*) an additional double-bond is present at C-18/C-19 and the C-12 acetate is still present (*i.e.* not decarboxylated) and (*c*) *C*-methylation has occurred not at C-12α but at C-11. The last feature (*c*) was confirmed (ref. 8) by the enzymic production of precorrin-6x from [*methyl*-^{13}C]SAM and [11-^{13}C]uro'gen III (**6**) which was unambiguously synthesised as in Scheme 5. The ^{13}C-signals from C-11 and from one *C*-methyl group of the isolated ester (**14a**) were split by direct coupling. This methylation at C-11 was totally unexpected; it produces two separated chromophores and so accounts for the pale yellow colour of precorrin-6x (**14**).

It is important to emphasise that though the illustrated double-bond positions are established for precorrin-6x ester (**14a**), the initial octa acid produced enzymically may be a close tautomer. This same *proviso* also holds for the two new biosynthetic intermediates to be described later.

Reduction of precorrin-6x to precorrin-6y

The reduction of the C-18/C-19 double bond of precorrin-6x (**14**) is catalysed by a reductase enzyme which is coded by the *cobK* gene in *P. denitrificans* and the enzyme has been isolated and characterised (ref. 9). It was then possible to develop conditions for isolation of the product, precorrin-6y, which was converted into hydrogenobyrinic acid (**13**) in high yield when it was incubated with the full enzyme system from *P. denitrificans*. With this new intermediate in hand, a joint effort, involving multiple ^{13}C-labelling eactly as for the work on precorrin-6x, proved (ref. 10) that precorrin-6y

Scheme 6

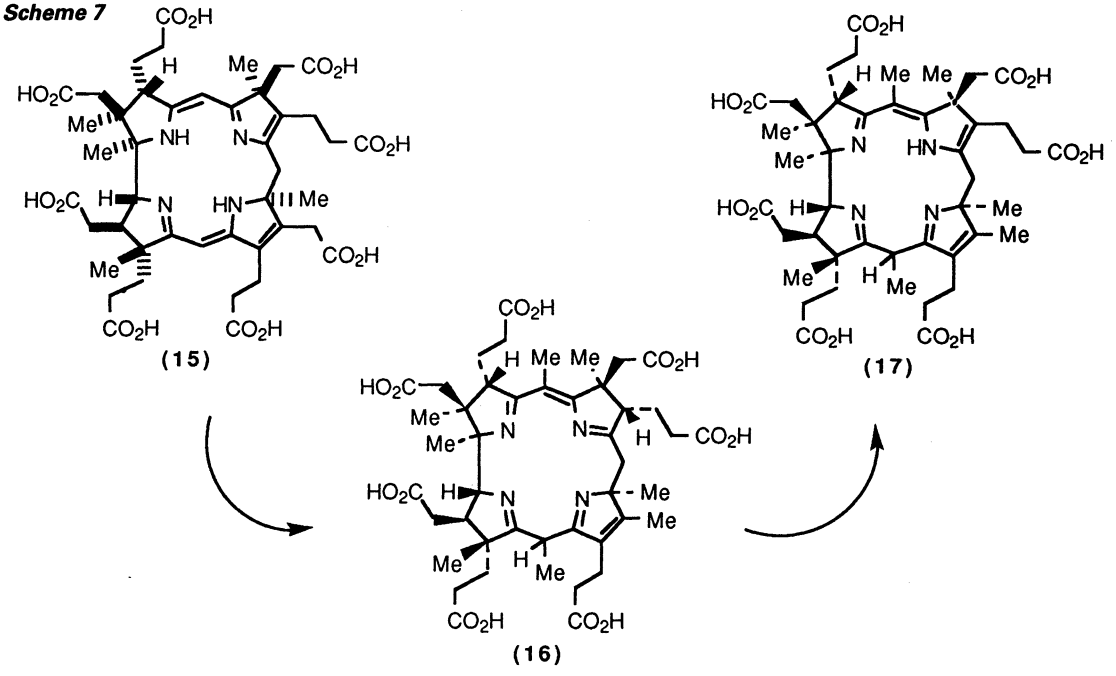


has structure (**15**). Further, experiments in which [4-$^2H_2$]NADPH was used as the reducing cofactor (see Scheme 6) showed (ref. 11) that the hydride equivalent is transferred from the cofactor to C-19 of (**14**). Finally, the preparation of [4R-$^2H_1$]NADPH and [4S-$^2H_1$]NADPH then allowed proof (ref. 12) that the reductase catalyses the transfer of H_R from C-4 of NADPH; H_R is the *forward* hydrogen of NADPH as it is drawn in Scheme 6.

Isolation and structure of precorrin-8x

When one compares the structure of precorrin-6y (**15**) with that of hydrogenobyrinic acid (**13**), it is evident that three different types of reaction are involved: (i) methylation at C-5 and C-15 (ii) decarboxylation of C-12 acetate (iii) rearrangement of the C-11 methyl group to C-12, though not necessarily in the listed order. The search for further enzymes involved in this section of the pathway achieved success by testing for protein fractions having methyl transferase activity towards precorrin-6y. A pure protein, shown to be coded by the *cobL* gene, was isolated (ref. 13) which surprisingly not only

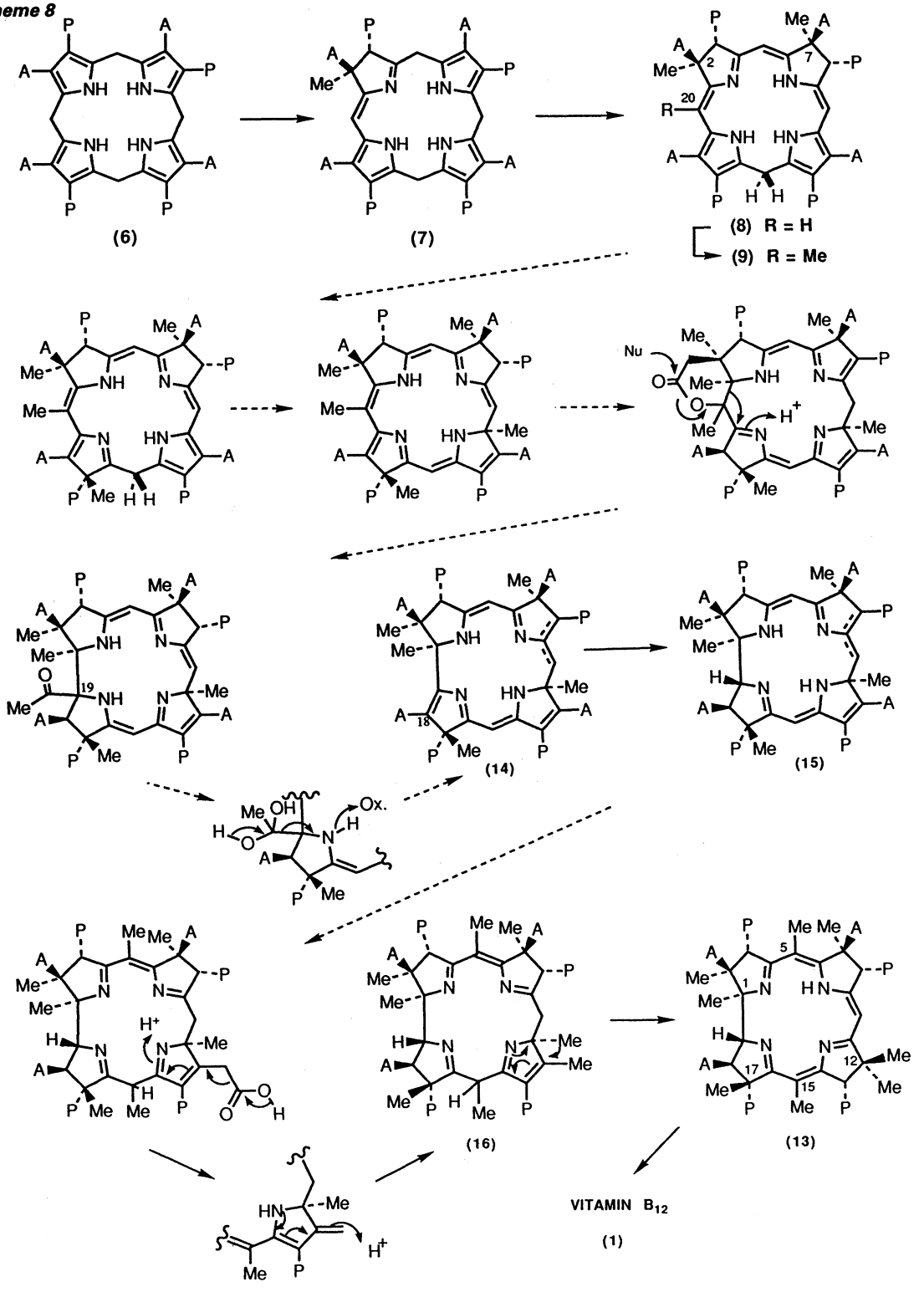
Scheme 7



carried out the C-5 and C-15 methylations but also decarboxylated the acetate at C-12. A small amount of the product of this enzyme was isolated and was named precorrin-8x (see later). This product was identical with material obtained by a more preparative method (ref. 14) based on the strong inhibition by hydrogenobyrinic acid of the final biosynthetic step precorrin-8x → hydrogenobyrinic acid (**13**). Thus, incubation of precorrin-3 (**9**) with the complete *P. denitrificans* enzyme system would normally give (**13**) but addition of (**13**) at the outset caused precorrin-8x to accumulate. Labelling experiments (ref. 14) showed that *five C*-methyl groups had been added during the formation of precorrin-8x from precorrin-3 (**9**), a result which justifies the statements above about the methylation state of this new intermediate. Finally, incubation of doubly-labelled precorrin-8x with the full enzyme system gave a high yield of (**13**) with no significant change in the labelling ratio so establishing the importance of precorrin-8x.

Precorrin-8x was an intensely frustrating substance since it changed in aqueous solution to give a mixture of at least five closely related forms which could be separated. Such behaviour causes great difficulties especially when one is working with *ca.* 300 μg of precious ^{13}C-labelled sample. But when eventually it was realised that these different forms all gradually and spontaneously yielded the same

Scheme 8



final stable form, structural work could start. This followed the earlier approach of extensive ^{13}C-labelling used for precorrin-6x and precorrin-6y and led (ref. 15) to structure (**17**) for the stable form of precorrin-8x, Scheme 7.

Enzymic incorporation experiments then allowed selection of the true biosynthetic intermediate, precorrin-8x itself, from the above set of closely related forms; also, conditions were found

to stabilise this original form. Then by incorporating 2,3-^{13}C-ALA, it was shown that position 12 of the biosynthetic intermediate carries a methyl group not a methylene. This together with much more NMR spectroscopy led to structure (**16**) as the best fit for precorrin-8x (ref. 15). Though some details of this structure (*e.g.* stereochemistry) remain to be established, the main features are clear. All the *C*-methyl groups of (**2**), (**13**) and vitamin B_{12} are attached to the macrocycle of precorrin-8x and only rearrangement of the 11-Me to C-12 is needed to yield (**13**).

The final rearrangement to form corrin

The pure enzyme which catalyses the rearrangement of precorrin-8x (**16**) to give (**13**) has been isolated; (ref. 14) it is encoded by the *cobH* gene. Being relatively small (M, 22,000), this enzyme should be susceptible to both structural and mechanistic study.

SUMMARY

All the information can be combined so that a large part of the pathway to vitamin B_{12} as it has now been revealed can be enjoyed. The steps forward from uro'gen III (**6**) are collected in Scheme 8 to which should be added the first part of Scheme 1 which shows how uro'gen III is built. The various intermediates which are numbered in Schemes 1 and 8 are established as being on the pathway to vitamin B_{12} and their structures are as illustrated, though with the *proviso* emphasised earlier about possible closely related double-bond tautomers in a few cases. The gaps in knowledge are presently at precorrin-4 and precorrin-5 leading on to precorrin-6x (**14**). For this section of the pathway, I thought it would be instructive and sobering to show the hypothetical structures for these tetra- and penta-methylated intermediates as they were viewed last year. What I can say now is that our French colleagues have already gained enough new information to tell us that these ideas must be changed, especially concerning the sequence of events. Of course, the exact nature of the transformations will be revealed by rigorous structural work which will be carried out jointly between Paris and Cambridge, as for all the other new intermediates I have described in this lecture. Attractive suggestions about how the ring-contraction step might occur have been made by Eschenmoser and the principles will hold at whatever stage in the sequence it happens; these involve lactone formation from the ring-A and ring-D acetate residues (ref. 16). For simplicity, Scheme 8 illustrates lactonisation of just the ring-A acetate but the reader must bear in mind that double lactonisation may be involved and should consult ref. 16 for the details. There are experimental results based on ^{18}O-labelling which are in keeping with the lactonisation idea (ref. 17). One way in which the putative C-19 acetyl group of the initial ring-contracted system could be oxidatively cleaved is indicated in Scheme 8 but the oxidation could equally well be initiated at the hydroxyl terminus. Also there are other very reasonable mechanisms by which this cleavage could be achieved. Precorrin-6x (**14**) is then the next biosynthetic intermediate and the remaining steps forward all the way to hydrogenobyrinic acid (**13**) are now known in substantial detail. The final rearrangement of precorrin-8x (**16**) to (**13**) is shown as a suprafacial 1,5-sigmatropic rearrangement of a pyrrolenine. This rearrangement removes the blockage at C-11 so that the macrocycle can gain the stabilisation of the fully conjugated system present in hydrogenobyrinic acid (**13**). The biosynthetic steps going forward from (**13**) to vitamin B_{12} (**1**), including amide formation, cobalt insertion and attachment of the nucleotide loop, are beyond the scope of this Lecture but much is known about these stages (see refs. 5 and 1 and references therein).

The joint Paris-Cambridge teams aim to fill the remaining gaps in knowledge of the pathway to hydrogenobyrinic (**13**) and cobyrinic (**2**) acids, so to reach the "Everest" summit and then move forward to mechanistic and enzymic research on the key steps.

Acknowledgements

I grasp this opportunity to record my great debt to all the outstanding scientists listed at the start of this Lecture. Special praise should be given to the French group who made the latest breakthrough in 1990 which initiated the recent forward surge. It has been a privilege to work with such a fine team where the quality of its members matched the high demands of the B_{12} problem.

REFERENCES

1. For a full review see "B_{12}", D. Dolphin,, Ed.; Wiley: New York 1982, Vol. 1 & 2 and especially A. R. Battersby and E. McDonald, in Vol. 1, p. 107; A. R. Battersby, *Accounts Chem. Res.* **19**, 147, (1986); A. I.Scott, *Accounts Chem. Res.*, **23**, 308, (1990); F. J. Leeper, *Nat. Prod. Rep.*, **6**, 171, (1989) and references therein.
2. A. R. Battersby, K. Frobel, F. Hammerschmidt and C. Jones: *J. Chem. Soc., Chem. Commun.*, 455 (1982).
3. J. Kulka, C. Nussbaumer, and D. Arigoni, *J. Chem. Soc. Chem. Commun.* 1512 (1990); I. Grgurina, S. Handa, G. Weaver, P. A. Cole and A. R. Battersby, *J. Chem. Soc. Chem. Commun.*, 1514 (1990).
4. B.Cameron, K. Briggs, S. Pridmore, G. Brefort and J. Crouzet, *J. Bacteriol.*, **171**, 547 (1989); J. Crouzet, L. Cauchois, F. Blanche, L. Debussche, D. Thibaut, M.-C. Rouyez, S. Rigault, J.-F. Mayaux and B. Cameron, *J. Bacteriol.*, **172**, 5968 (1990); J. Crouzet, B. Cameron, L. Cauchois, S. Rigault, M.-C. Rouyez, F. Blanche, D. Thibaut, and L. Debussche, *J. Bacteriol.*, **172**, 5980 (1990).
5. For scheme showing all genes see J. Crouzet, S. Levy-Schil, B. Cameron, L. Cauchois, S. Rigault, M.-C. Rouyez, F. Blanche, L. Debussche and D. Thibaut, *J. Bacteriol.*, **173**, 6074 (1991).
6. D. Thibaut, L. Debussche and F. Blanche, *Proc. Natl. Acad. Sci. USA* , **87**, 8795 (1990).
7. D. Thibaut, F. Blanche, L. Debussche and F. J. Leeper and A. R. Battersby, *Proc. Natl. Acad. Sci. USA* , **87**, 8800 (1990).
8. F. Blanche, M. Kodera, M. Couder, F. J. Leeper, D. Thibaut, and A. R. Battersby, *J. Chem. Soc. Chem. Commun.*, 138 (1992).
9. F. Blanche, D. Thibaut, A. Famechon, L. Debussche, B. Cameron and J. Crouzet, *J. Bacteriol.* **174**, 1036 (1992).
10. D. Thibaut, F. Kiuchi, L. Debussche, F. J. Leeper, F. Blanche and A. R. Battersby, *J. Chem. Soc., Chem. Commun.*, 139 (1992).
11. G. W. Weaver, F. J. Leeper, A. R. Battersby, F. Blanche, D. Thibaut and L. Debussche, *J. Chem. Soc., Chem. Commun.*, 976 (1991).
12. F. Kiuchi, D. Thibaut, L. Debussche, F. J. Leeper, F. Blanche and A. R. Battersby, *J. Chem. Soc., Chem. Commun.*, 306 (1992).
13. F. Blanche, A. Famechon, D. Thibaut, L. Debussche, B. Cameron and J. Crouzet, *J. Bacteriol.* **174**, 1050 (1992).
14. D. Thibaut, M. Couder, A. Famechon, L. Debussche, B. Cameron, J. Crouzet and F. Blanche, *J. Bacteriol.* **174**, 1043 (1992).
15. D. Thibaut, F. Kiuchi, L. Debussche, F. Blanche, M. Kodera, F. J. Leeper and A. R. Battersby, *J. Chem. Soc., Chem. Commun.*, 982 (1992).
16. A. Eschenmoser, *Angew. Chem. Int. Ed. Engl.*, **27**, 5 (1988).
17. K. Kurumaya, T. Okazaki and M. Kajiwara, *Chem. Pharm. Bull.*, **37**, 1151 (1989).