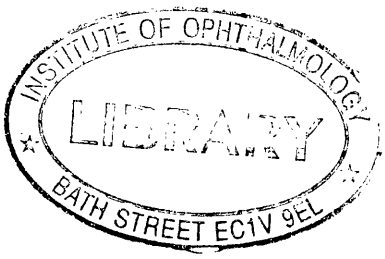
# Biosynthesis of cobalamin (vitamin $B_{12}$) in *Salmonella typhimurium* and *Bacillus megaterium de Bary*; Characterisation of the anaerobic pathway.

By Evelyne Christine Raux

A thesis submitted to the University of London
for the degree of doctorate (PhD.) in Biochemistry.




University College London
Department of Molecular Genetics, Institute of Ophthalmology, London.

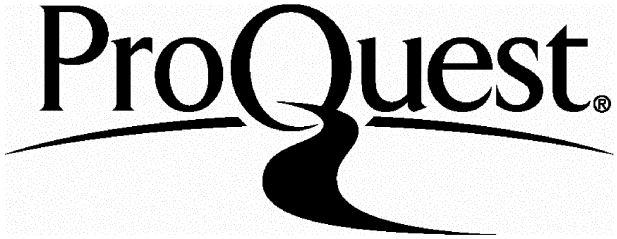
Jan 1999

ProQuest Number: U121800

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.




ProQuest U121800

Published by ProQuest LLC(2016). Copyright of the Dissertation is held by the Author.

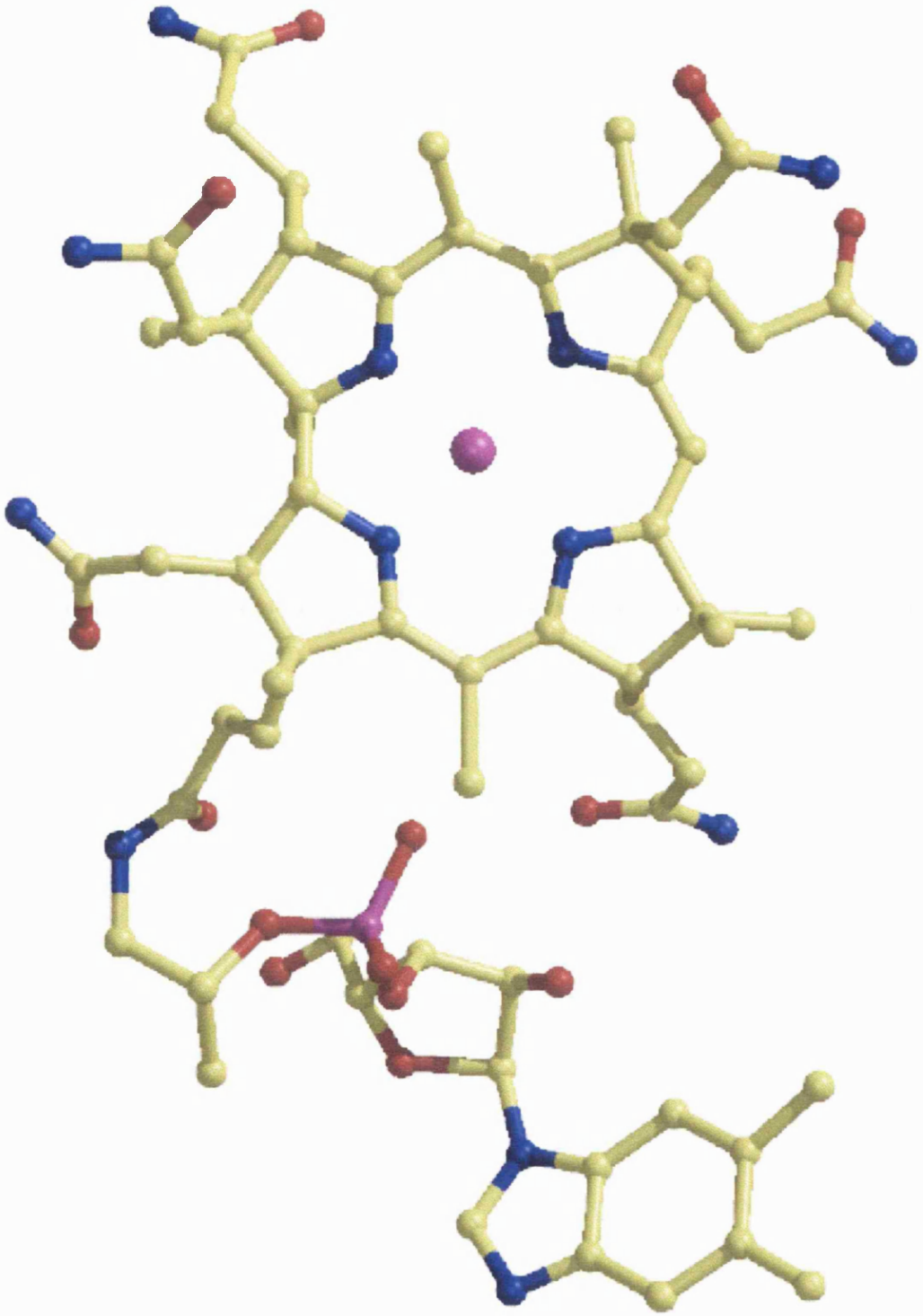
All rights reserved.
This work is protected against unauthorized copying under Title 17, United States Code.
Microform Edition © ProQuest LLC.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

**Abstract**

The transformation of uroporphyrinogen III into cobalamin (vitamin $B_{12}$) requires about 25 enzymes and can be performed by either aerobic or anaerobic pathways. The aerobic route is dependent upon molecular oxygen, and cobalt is inserted after the ring contraction process. The anaerobic route occurs in the absence of oxygen and cobalt is inserted into precorrin-2, several steps prior to the ring contraction. A study of the biosynthesis in both *S. typhimurium* and *B. megaterium* reveals that two genes, *cbiD* and *cbiG*, are essential components of the pathway and constitute genetic hallmarks of the anaerobic pathway. The genes responsible for the cobalt chelation, the *S. typhimurium* CbiK and the *B. megaterium* CbiX, were identified within *cob* operons and were characterised. Moreover, the activity of the multifunctional iron chelatase/dehydrogenase enzymes (*E. coli* CysG and *S. cerevisiae* Met8p) involved in sirohaem biosynthesis have been investigated for their ability to act as a cobalt chelatase in corrin biosynthesis. Cobalamin can be produced from the *S. typhimurium cob* operon with any of these chelatases whereas precorrin-2 dehydrogenase activity is required with the *B. megaterium cob* operon. The X-ray structure of CbiK has been solved at 2.4Å and is highly similar to the structure of the *B. subtilis* protoporphyrin IX ferrochelatase suggesting a common mechanism. Unlike the *P. denitrificans* cobalt chelatase complex, which requires three proteins and ATP (similar to the protoporphyrin IX magnesium chelatase), CbiK belongs to the unique protein-ATP independent chelatase family. Conserved amino acids have been characterised as key residues within the CbiK active site. Genomic comparisons of $B_{12}$-producing organisms highlight divergences between the methyltransferases, which separate into aerobic and anaerobic pathway subgroups. Further insights into the methyltransferases have been gained from the X-ray structure of the *B. megaterium* CbiF (solved at 2.4Å resolution). Finally, the molecular structure of cobyric acid produced from the *S. typhimurium cob* operon in *E. coli* has been deduced from a number of spectrometric studies, an approach which could be used in the future to characterise other intermediates along the anaerobic cobalamin pathway. From the results obtained in this thesis, it becomes apparent that the terms "aerobic" and "anaerobic" pathways are misleading and should be replaced by "late-cobalt insertion" and "early-cobalt insertion" pathways respectively.

## Nature's most beautiful cofactor, cobalamin.



The methyl group linked to the cobalt has been removed in this diagram and the loop has been detached from the cobalt (purple), as it occurs when the methyl-cobalamin is bound in the methionine synthase. Atoms in yellow indicate a carbon; in blue, nitrogen; in red, oxygen and the purple atom in the lower part of the figure is a phosphore [Drennan *et al.* (1994) Science, 266: 1669-1674].

**Acknowledgements**

I started research on vitamin $B_{12}$ biosynthesis in October 1988, and for over ten years, this molecule has never ceased to fascinate me. Many people have been involved in this work and I would like to thank them here.

**In France:**

- I am very grateful to Dr Alain Rambach for having introduced me to the $B_{12}$ world, trusting me when I worked in Gif-sur-Yvette and to have allowed me to use his $B_{12}$ collection strains afterwards.

- I would also like to thank everyone who was working with me in Gif-sur-Yvette between 1988 and 1995. A very special thanks to Dr Claude Thermes, for having taught me the basis of molecular biology and for all those endless discussions about $B_{12}$. Thanks to Anne Lanois and Yves D'aubenton for their continuous help and support during those years.

**In England:**

- I give my most sincere thanks to my supervisor, Dr Martin Warren. This thesis would never have been realised without his help, guidance and discussions. Thanks also for correcting my English with such patience...

- A very warm thanks to Dr Heidi Schubert for her invaluable and efficient collaboration.

- Many friendly thanks for the help and collaboration of Dr Sarah Woodcock and Dr Jenny Roper.
- To everyone in the Molecular Genetics department, and especially to Natalie Senior, Sarah Awan, Richard Newbold, Sussie Dalvin and Seraphina Idowu.

- A big thanks to Dr Helen Wilcock for her everyday support over the three years.
- My warmest thanks to Isabelle, Daniel and Nicole Evain for their advice and support, and to my parents for always having respected my choices, whatever they were.

- And finally, the last word is for you, Phil...

**Remerciements**

J'ai commencé à étudier la biosynthèse de la vitamin $B_{12}$ en octobre 1988, et depuis plus de dix ans, cette molécule n'a pas jamais cessé de me fasciner. Beaucoup de personnes ont participé à cet ambitieux projet, j'aimerai aujourd'hui les remercier.

**En France:**

• Je suis trés reconnaissante à Monsieur Alain Rambach de m'avoir fait découvrir le monde de la $B_{12}$, puis de m'avoir fait confiance lorsque je travaillais à Gif-sur-Yvette, ainsi que de m'avoir permis d'utiliser sa collection de souches productrices de $B_{12}$, après ma venue à Londres.

• Je remercie aussi tous ceux qui travaillaient à Gif-sur-Yvette entre 1988 et 1995. Un merci particulier à Claude Thermes de m'avoir appris les bases de la biologie moléculaire ainsi que pour toutes ces interminables discussions à propos de $B_{12}$. Merci également à Anne Lanois et Yves D'aubenton pour leur complicité de chaque jour durant ces années.

**En Angleterre:**

• J'adresse mes plus sincères remerciements au Dr Martin Warren. Sans son aide, ses conseils et les multiples discussions que nous avons eues, cette thèse n'aurait jamais vu le jour. Merci aussi d'avoir corrigé mon anglais avec tant de patience...

• Je remercie trés chaleureusement le Dr Heidi Schubert pour son inestimable collaboration.

• Un merci amical aux Dr Sarah Woodcock et Jenny Roper pour l'aide qu'elles m'ont fournie chaque jour et pour leur collaboration, ainsi qu'à tous ceux du département de génétique moléculaire, et plus particulièrement à Natalie Senior, Sarah Awan, Richard Newbold, Sussie Dalvin and Seraphina Idowu.

• Un merci spécial au Dr Helen Wilcock pour son soutien de chaque jour pendant plus de 3 ans.

• Mes plus chaleureux remerciements vont à Isabelle, Daniel et Nicole Evain pour leur écoute et leurs conseils, ainsi qu'à mes parents pour avoir toujours respecté mes choix, quels qu'ils soient.

• Le mot de la fin est pour toi, Phil...

**Contents.**

## List of Tables

## List of Figures

## Abbreviations

| | |
|---|---|
| A | Adenine |
| ALA | 5-aminolaevulinic acid |
| ATP | Adenosine triphosphate |
| AP | Aminopropanol |
| C | Cytosine |
| CoA | Coenzyme A |
| Da | Dalton |
| DEAE | Diethylaminoethyl |
| DNA | Deoxyribonucleic acid |
| EDTA | Ethylenediaminetetra-acetic acid |
| EPR | Electron paramagnetic resonance |
| f.p.l.c | Fast protein liquid chromatography |
| G | Guanine |
| IPTG | Isopropyl-β-D-thiogalactopyranoside |
| kb | Kilobase pair |
| LB | Luria-Bertani |
| MCS | Multi cloning site |
| $NAD^+$/NADH | Nicotinamide adenine dinucleotide |
| $NADP^+$/NADPH | Nicotinamide adenine dinucleotide phosphate |
| NMR | Nuclear magnetic resonance |
| OD | Optical density |
| PAGE | Polyacrylamide gel electrophoresis |
| PBG | Porphobilinogen |
| PCR | Polymerase chain reaction |
| PLP | Pyridoxal phosphate |
| SAM | S-adenosyl L methionine |
| SDS | Sodium dodecyl sulfate |
| T | Thymine |
| TEMED | N, N, N', N' -tetramethylenediamine |
| Tris | Tris (hydroxymethyl)aminomethane |
| Uro'gen III | Uroporphyrinogen III |
| UV | Ultra violet |

**Amino acids and their abbreviations**

| Amino acid | Three letter code | Single letter code | Side chain |
|---|---|---|---|
| Alanine | Ala | A | non-polar |
| Arginine | Arg | R | basic |
| Asparagine | Asn | N | polar |
| Aspartate | Asp | D | acid |
| Cysteine | Cys | C | polar |
| Glutamate | Glu | E | acid |
| Glutamine | Gln | Q | polar |
| Glycine | Gly | G | non-polar |
| Histidine | His | H | basic |
| Isoleucine | Ile | I | non-polar |
| Leucine | Leu | L | non-polar |
| Lysine | Lys | K | basic |
| Methionine | Met | M | non-polar |
| Phenylalanine | Phe | F | non-polar |
| Proline | Pro | P | non-polar |
| Serine | Ser | S | polar |
| Threonine | Thr | T | polar |
| Tryptophan | Trp | W | polar |
| Tyrosine | Tyr | Y | polar |
| Valine | Val | V | non-polar |
| Selenocysteine | Sec | U | polar |

## Genetic code

| A | R | N | D | C | Q | E | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| Ala | Arg | Asn | Asp | Cys | Gln | Glu | Gly | His | Ile |
| GCA | CGA | AAC | GAC | UGC | CAA | GAA | GGA | CAC | AUA |
| C | C | U | U | U | G | G | C | U | C |
| G | G | | | | | | G | | U |
| U | U | | | | | | U | | |
| | AGA | | | | | | | | |
| | G | | | | | | | | |

| L | K | M | F | P | S | T | W | Y | V |
|---|---|---|---|---|---|---|---|---|---|
| Leu | Lys | Met | Phe | Pro | Ser | Thr | Trp | Tyr | Val |
| CUA | AAA | AUG | UUC | CCA | UCA | ACA | UGG | UAC | GUA |
| C | G | | U | C | C | C | | U | C |
| G | | | | G | G | G | | | G |
| U | | | | U | U | U | | | U |
| UUA | | | | | AGC | | | | |
| G | | | | | U | | | | |

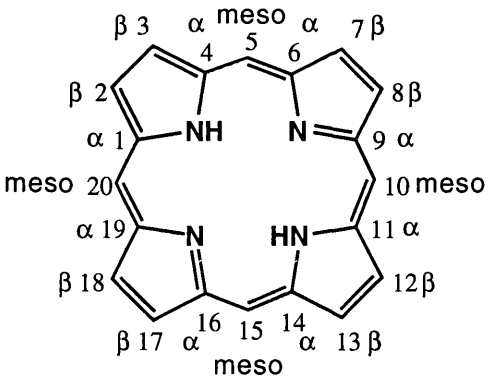
| Termination Signals |
|---|
| UAA (Ochre) |
| UAG (Amber) |
| UGA (Opal) |

## Cobalamin biosynthetic protein names.

| Function | *cob* | *P. denitrificans* | *S. typhimurium* | *B. megaterium* | *E. coli* |
|---|---|---|---|---|---|
| C2/C7 methylations | I | CobA | CysG | CysG[A] | CysG |
| amidation (a & c) | I | CobB | CbiA | CbiA | - |
| AP-P synthase | | CobC | CobD | - | - |
| AP attachment | I | CobD | CbiB | - | - |
| ? | | CobE | - | - | - |
| C1 methylation | I | CobF | - | - | - |
| mono oxygenase | I | CobG | - | - | - |
| methyl migration | I | CobH | CbiC | CbiC | - |
| C20 methylation | I | CobI | CbiL | CbiL | - |
| C17 methylation | I | CobJ | CbiH | $CbiH_{60}$ | - |
| macrocycle reduction | I | CobK | CbiJ | CbiJ | - |
| C5/C15 methylation and decarboxylation | I | CobL | CbiE and CbiT | CbiET | - |
| C11 methylation | I | CobM | CbiF | CbiF | - |
| cobalt chelation | I | CobN | - | - | - |
| adenosylation | I | CobO | CobA | BtuR | BtuR |
| GTP-cobinamide synthase | III | CobP | CobU | - | CobU |
| amidation (b, d, e, g) | I | CobQ | CbiP | - | - |
| cobalt chelation | I | CobS | - | - | - |
| cobalt chelation | I | CobT | - | - | - |
| α-ribazole-P synthase | II | CobU | CobT | - | CobT |
| cobalamin synthase | III | CobV | CobS | - | CobS |
| ? | | CobW | - | - | - |
| ? | I | - | CbiD | CbiD | - |
| ? | I | - | CbiG | CbiG | - |
| cobalt chelatase | I | - | CbiK | - | - |
| cobalt transport | | - | CbiM | - | - |
| cobalt transport | | - | CbiN | - | - |
| cobalt transport | | - | CbiO | - | - |
| cobalt transport | | - | CbiQ | - | - |
| α-ribazole-P Pase | II | - | CobC | - | - |
| ? | | - | - | CbiW | - |
| cobalt chelatase | I | - | - | CbiX | - |
| ? | | - | - | CbiY | - |

P means phosphate; Pase, phosphatase and AP, aminopropanol.

**Tetrapyrrole numbering system and nomenclature.**




| | |
|---|---|
| Bacteriochlorin | two reduced double bonds oppositely. for example: between the C-17 and C-18 and between C-7 and C-8. Tetrahydroporphyrin (9 C=C). |
| Chlorin | one double bond is reduced creating two carbon atoms fully saturated. for example: between the C-17 and C-18. Dihydroporphyrin (10 C=C). |
| Corphin | only saturated carbon at β positions (8 C=C). |
| Corrin | carbon 20 is missing, all carbons at position β are reduced. One meso carbon is also reduced (5 C=C). |
| Isobacteriochlorin | two reduced double bonds on adjacent pyrroles. for example: between the C-2 and C-3 and between C-7 and C-8. Tetrahydroporphyrin (9 C=C). |
| Phlorin | one saturated carbon bond at meso position. Dihydroporphyrin. (10 C=C). |
| Porphyrin | only unsaturated carbon bonds (11 C=C). |

## Meeting/Conferences

Tetrapyrrole meeting, January 96, Southampton:

Poster: *S. typhimurium* cobalamin (vitamin $B_{12}$) biosynthetic genes, functional studies in *S. typhimurium* and *E. coli*.

4th European symposium on vitamin $B_{12}$ and $B_{12}$-proteins, September 96, Innsbruck (Austria):

Poster: Identification and characterisation of a *Bacillus megaterium cob* operon: an obligate aerobe with an anaerobic cobalamin biosynthetic pathway.

Tetrapyrrole meeting, July 97, London, QMW:

Talk: Cobalamin (vitamin $B_{12}$) biosynthesis, the second step of the "anaerobic" pathway.

Tetrapyrrole meeting, March 98, Cambridge:

Poster: Vitamin $B_{12}$ biosynthesis in *S. typhimurium* and *B. megaterium*: different routes to climb Mt Everest without oxygen.

Gordon research conferences, chemistry & biology of tetrapyrroles, July 98, Rhode Island (USA):

Poster: Vitamin $B_{12}$ biosynthesis in *S. typhimurium* and *B. megaterium*: different routes to climb Mt Everest without oxygen.

Tetrapyrrole meeting, January 99, London.

## List of publications


Raux, E., Lanois, A., Levillayer, F., Warren, M. J., Brody, E., Rambach, A. and Thermes, C. (1996) *S. typhimurium* cobalamin (vitamin $B_{12}$) biosynthetic genes; functional studies in *S. typhimurium* and *E. coli*: *J. Bacteriol.* **178**: 753-767.

Raux, E., Thermes, C., Heathcote, P., Rambach, A., and Warren, M. J. (1997) The role of the *Salmonella typhimurium cbiK* in cobalamin (vitamin $B_{12}$) and siroheme biosynthesis. *J. Bacteriol.* **179**: 3202-3212.

Beck, R., Raux, E., Thermes, C., Rambach, A. and Warren, M. J. (1997) CbiX: A novel metal-binding protein involved in sirohaem biosynthesis in *Bacillus megaterium*. *Biochem. Soc. Trans.* **25**: 77S.

Woodcock, S. C., Raux, E., Levillayer, F., Thermes, C., Rambach, A., and Warren, M. J. (1998) Effect of mutations in the transmethylase and dehydrogenase/chelatase domains of sirohaem synthase (CysG) on sirohaem and cobalamin biosynthesis. *Biochem. J.* **330**: 121-129.

Raux, E., Schubert, H. L., Woodcock, S. C., Wilson, K. S., and Warren, M. J. (1998) Cobalamin (vitamin $B_{12}$) biosynthesis: Cloning, expression and crystallisation of the *Bacillus megaterium* S-adenosyl-L-methionine-dependent-cobalt-precorrin-4 transmethylase CbiF. *Eur. J. Biochem.* **254**: 341-346.

Schubert, H. L., Wilson, F. S., Raux, E., Woodcock, S. C. and Warren, M. J. (1998) The X-ray structure of a cobalamin biosynthetic enzyme, cobalt-precorrin-4 methyltransferase. *Nature Structure Biology*. **5**: 585-592.

Raux, E., Lanois, A., Warren, M. J., Rambach, A., and Thermes, C. (1998) Cobalamin (vitamin $B_{12}$) biosynthesis: Identification and characterisation of a *Bacillus megaterium cob* operon. *Biochem. J.* **335**: 159-166.

Raux, E., Lanois, A., Rambach, A., Warren, M. J., and Thermes, C. (1998) Cobalamin (vitamin $B_{12}$) biosynthesis: Functional characterisation of the *Bacillus megaterium cbi* genes required to convert uroporphyrinogen III into cobyrinic acid *a,c*-diamide. *Biochem. J.* **335**: 167-173.

Raux, E., Mc Veigh, T., Peters, S. E., Leusteck, T. and Warren, M. J. (1999) The role of *Saccharomyces cerevisiae* Met1p and Met8p in sirohaem and cobalamin biosynthesis. *Biochem. J.* **338**: In press.

Raux, E., Schubert, H. L., Roper, J. M., Wilson, S. W. and Warren, M. J. (1999) Vitamin $B_{12}$; insights into biosynthesis's Mount improbable. **(Review)** *Bioorganic. Chem.* **27**: 100-118.

Schubert, H. L., Raux, E., Wilson, S. W. and Warren, M. J. (1999) Common chelatase design in the branch tetrapyrrole pathways of heme and anaerobic cobalamin synthesis. In prep.

INSTITUTE OF OPHTHALMOLOGY
LIBRARY
BATH STREET EC1V 9EL

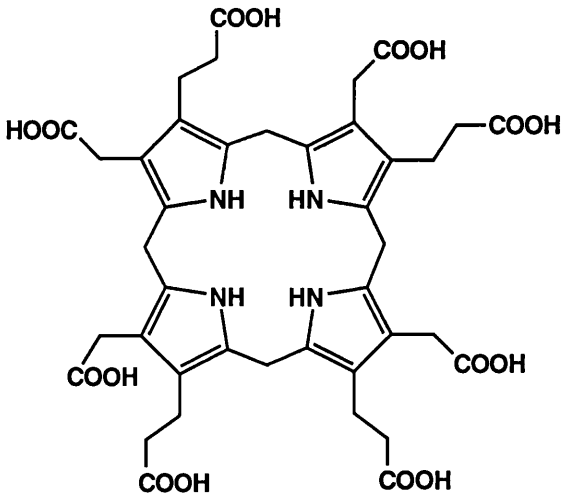
# CHAPTER 1,

# INTRODUCTION TO TETRAPYRROLES AND COBALAMIN IN PARTICULAR

## 1-1- General introduction to Nature's tetrapyrroles.

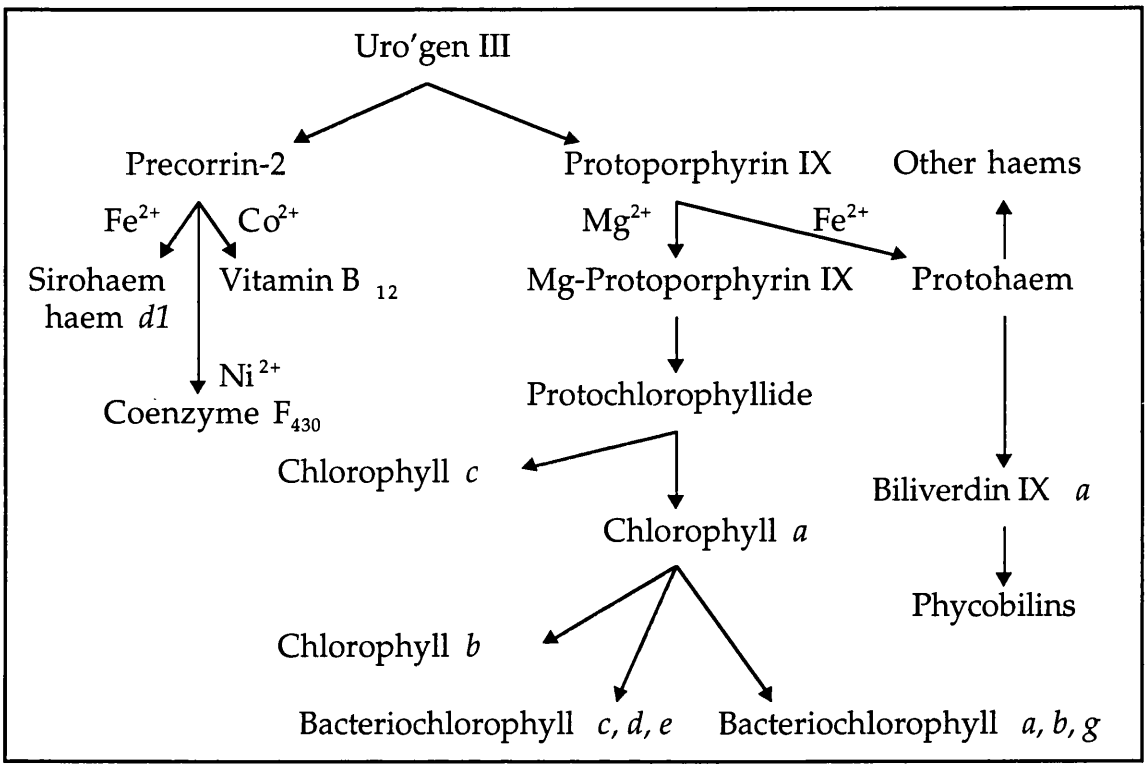
Biological modified tetrapyrrole molecules act as prosthetic groups to many fundamental processes in living organisms. They are commonly referred to as the "pigments of life", a denomination which illustrates their essential biological role as well as their chromophoric properties.

Figure 1-1- Uroporphyrinogen III.




All modified tetrapyrroles are derived from uroporphyrinogen III, the ubiquitous primogenitor which consist of four pyrrole rings joined by carbon bridges to form a larger macrocyclic system (Figure 1-1). Enzymatic modifications of uro'gen III yield the six main types of modified tetrapyrrole found in Nature: haem, chlorophyll and bacteriochlorophyll, bilin, isobacteriochlorin, coenzyme $F_{430}$ and vitamin $B_{12}$ (Figure 1-2).

Figure 1-2- Outline of pathways for modified tetrapyrrole biosynthesis.

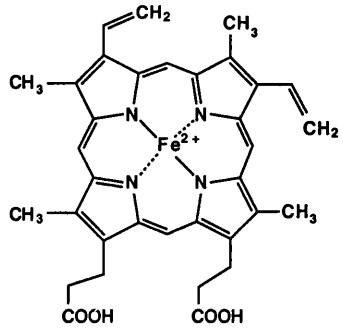



This is a modified version of the figure from Friedmann *et al.* 1991.

## 1-1-1- Haem.

Haem is involved in oxygen and electron transport. It is a porphyrin that contains a central ferrous ion. Three main types of haem are commonly observed which differ only in their side chains. The *a*-type haem is the prosthetic group of cytochrome oxidase which is the terminal reaction of the respiratory chain (cytochromes *b* to *c1* to *c* to *a+a3*). The *b*-type haem (Figure 1-3) is the cofactor of haemoglobin, cytochromes *b*, $P_{450}$, catalase and peroxidase. The *c*-type haem is a component of cytochrome *c* (mitochondria) and cytochrome *f* (plastids) [Beale and Weinstein, 1991]. Finally, other types of haem can be found in bacteria, haem *d1* (presumably derived from precorrin-2) and *o*, which are associated with terminal cytochrome oxidases [Chang *et al.* 1993]. Haem biosynthesis has been determined in *Desulfovibrio vulgaris* by Ishida *et al.* 1998. Protohaem, in this bacterium, is synthesised via a precorrin-2 intermediate, which is subsequently decarboxylated (C-12 & C-18) and deacetylated (C-2 & C-7) to give coproporphyrinogen III, the common intermediate to any haem biosynthetic pathway.

Figure 1-3- Haem *b* (iron protoporphyrin IX).




**Haem**

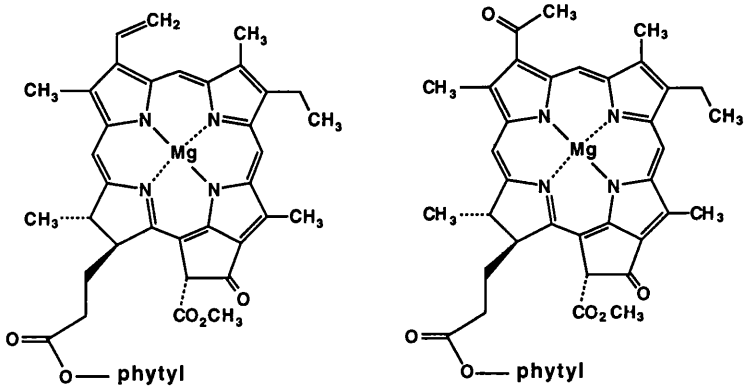
## 1-1-2- Chlorophyll and bacteriochlorophyll.

Chlorophyll and bacteriochlorophyll catalyse the conversion of solar energy to chemical energy by the process of photosynthesis. These pigments not only contain a centrally chelated magnesium ion but they also possess a fifth ring.

Chlorophylls are mainly divided into three classes, *a*, *b* and *c*. Higher plants and green algae contain chlorophyll *a* (Figure 1-4) and a small proportion of chlorophyll *b*. The major structural difference between the two forms is a formyl group (Chl *b*) instead of a methyl group (Chl *a*) on ring B (carbon 7) of the macrocycle. Chlorophyll *c* is found in eukaryotic algae that do not contain chlorophyll *b* (except one species). None of its pyrrole rings are reduced [Beale and Weinstein, 1991].

There are seven different bacteriochlorophylls, *a*, *b*, *c*, *d*, *e*, *f* and *g*. Bacteriochlorophyll *a* (Figure 1-4) and *b* are purple pigments produced by photosynthetic bacteria (for example, *Rhodobacter sphaeroides* (Bchl *a*) and *Rhodopseudomonas viridis* (BChl *b*)). Bacteriochlorophyll *g* was isolated from *Heliobacterium chlorum*. Bacteriochlorophyll *c*, *d* and *e* are pigments of green and brown sulfur bacteria (for example *Prosthecochloris aestuarii* (BChl *c*), *Chlorobium vibrioforme* (BChl *d*) and *Chlorobium pheovibrioides* (BChl *e*) [Smith, 1991].

Figure 1-4- Chlorophyll *a* and bacteriochlorophyll *a*.




Chlorophyll a Bacteriochlorophyll a

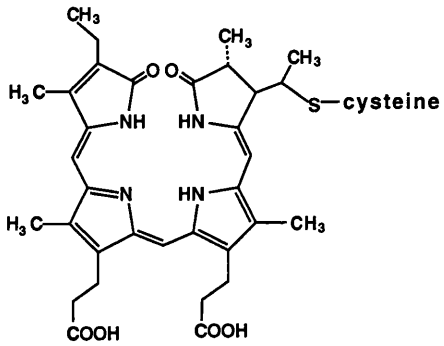
## 1-1-3- Phycobilins and phytochrome.

Phycobilins and phytochrome consist of open-chain modified tetrapyrroles. They function as accessory pigments and act as the main photosynthetic light-harvesting pigments when attached with appropriate apoproteins by a covalent thioester bond linkage (phycobiliproteins).

The two main classes of phycobilins are either a blue pigment, phycocyanin (Figure 1-5) and allophycocyanin, or a red pigment, phycoerythrin. They are found in cyanobacteria and red algae. These accessory pigments allow the organisms to capture more of the available light.

The phytochrome chromophore is found in higher plants and some green algae [Beale and Weinstein, 1991].

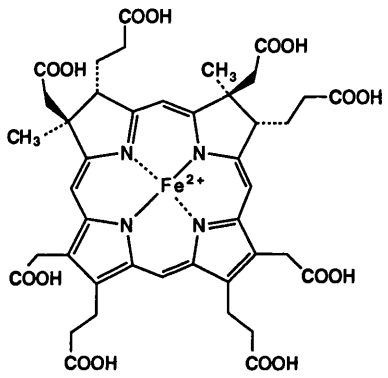
Figure 1-5- Phycocyanin.



## 1-1-4- Isobacteriochlorins.

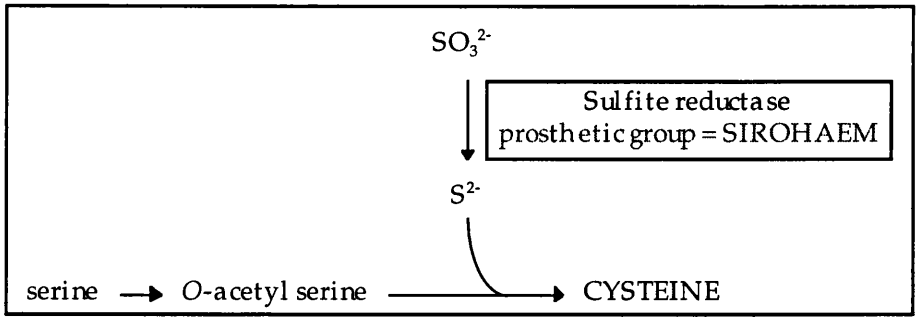
The most common isobacteriochlorin is sirohaem, the prosthetic group of nitrite and sulfite reductases, which convert nitrite to ammonia and sulfite to sulfide respectively. The central metal of sirohaem is ferrous iron (Figure 1-6).

Figure 1-6- Sirohaem.




As shown in Figure 1-7, sulfite reductase produces sulfide which then reacts with *O*-acetylserine to form cysteine. The absence of sirohaem from the sulfite reductase leads to the failure to reduce sulfite and therefore to a cysteine auxotrophy [Kredich, 1987].

Figure 1-7- Role of sirohaem in cysteine biosynthesis.



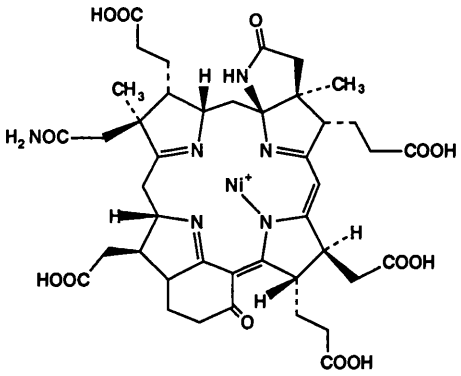

Another example of isobacteriochlorin is the purple pigment $Cobalt^{III}$-isobacteriochlorin found in *Desulphovibrio gigas* and *D. desulphuricans* which are

two sulfite reducing microorganisms thought to be primitive on the evolutionary scale [Battersby and Sheng, 1982].

## 1-1-5- Coenzyme $F_{430}$:

Coenzyme $F_{430}$, the prosthetic group of methyl coenzyme M reductase, is involved in the catalysis of the final step of methane formation in methanogenic bacteria such as *Methanobacterium thermoautotrophicum*. This yellow pigment is a nickel-containing modified tetrapyrrole (Figure 1-8). The crystal structure of this enzyme containing two molecules of coenzyme $F_{430}$ has been recently resolved at 1.45 Å and the coenzyme $F_{430}$ role in the reaction has been described [Ermler *et al.* 1997].

Figure 1-8- Coenzyme $F_{430}$:

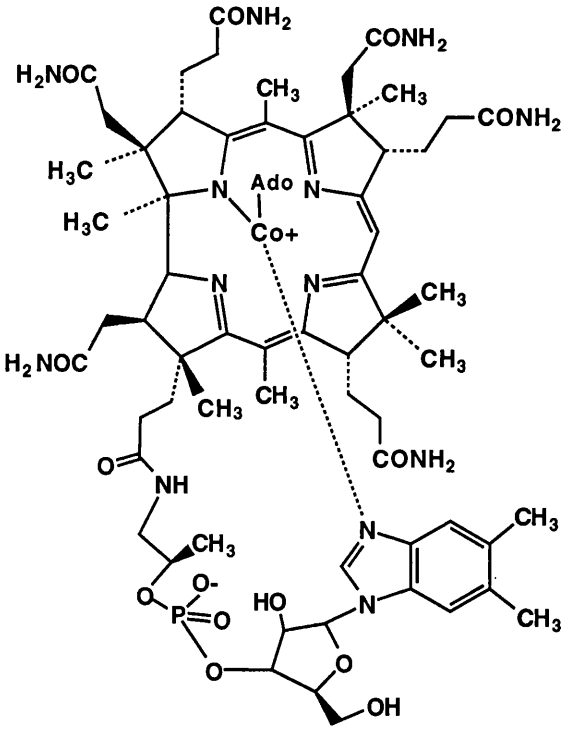



## 1-1-6- Vitamin $B_{12}$:

Cobalamin, or vitamin $B_{12}$, is certainly the most complicated cofactor of all biopigments. It possesses a contracted macrocycle (rings A and D are directly bonded together) and contains a central cobalt ion, modified side chains and various axial ligands (Figure 1-9). The lower ligand is linked from the carbon 17 to the central cobalt by an "arm" constituted of aminopropanol and a dimethylbenzimidazole derivative. From an evolutionary standpoint, vitamin $B_{12}$ can be considered the matriarch of the family through its role in deoxyribonucleotide synthesis despite the fact that this function has been maintained in only few bacteria like *Lactobacillus leichmanii*. Vitamin $B_{12}$ is only

biosynthesised by some micro-organisms, but most higher animals are capable of converting the vitamin into the two required forms, methylcobalamin and adenosylcobalamin. Vitamin $B_{12}$ is involved in methylations and various intramolecular rearrangement reactions.

Figure 1-9- Vitamin $B_{12}$.



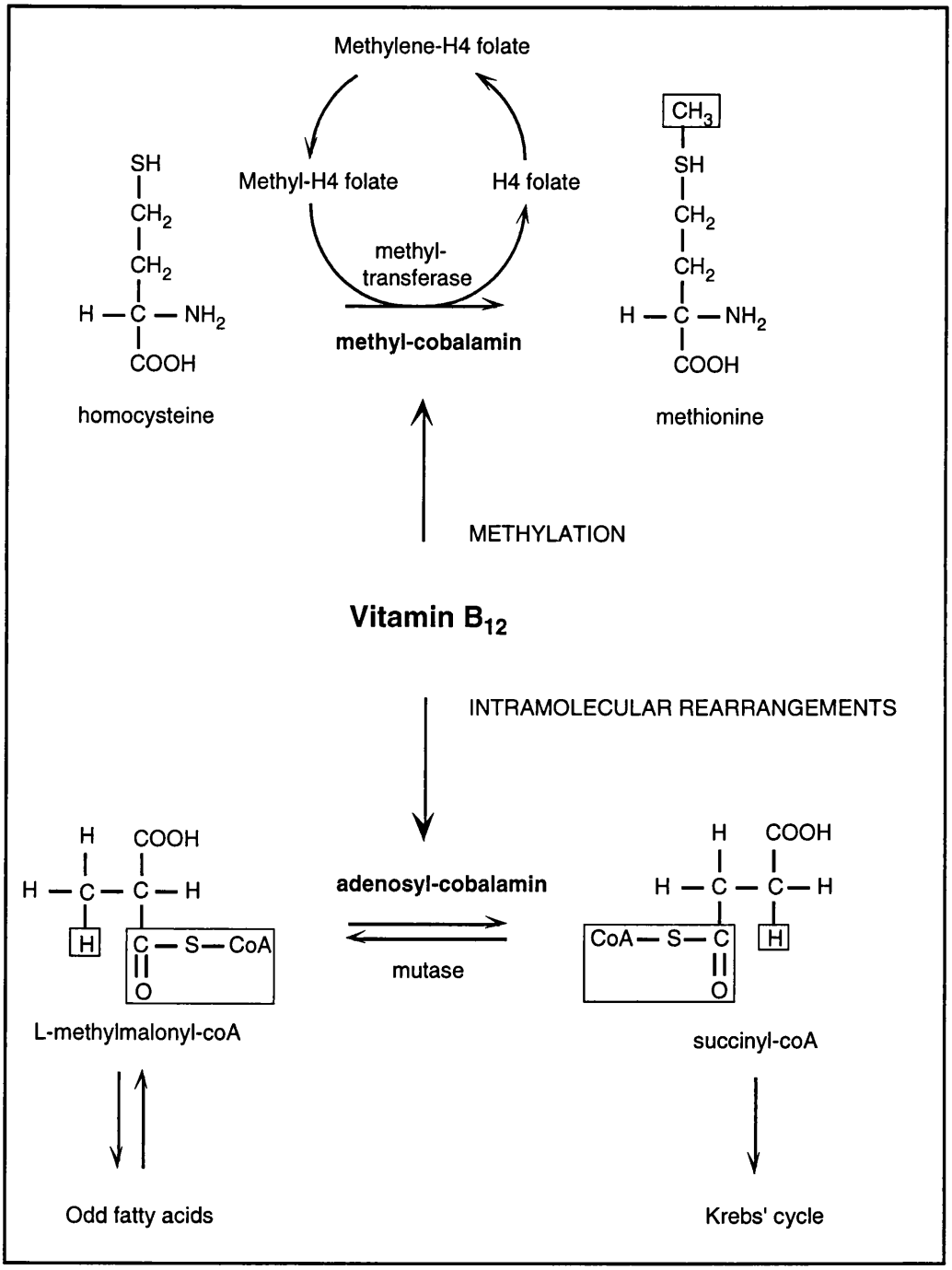

Ado-Cobalamin

Vitamin $B_{12}$ is in fact the cyano form of cobalamin.

## 1-2- Cobalamin dependent enzymes.

Only two enzymes are known to depend on cobalamin in human and animals, methionine synthase and methylmalonyl-CoA mutase (Figure 1-10). However, more than ten different $B_{12}$-dependent enzymes are now known in micro-organisms.

Figure 1-10- The two mammalian reactions requiring vitamin $B_{12}$.



## 1-2-1- Reactions catalysed by cobalamin.

The two active forms of cobalamin, adenosyl- and methyl-cobalamin, catalyse three different types of reactions.

Adenosyl-cobalamin participates in 1, 2 intramolecular rearrangements in which a hydrogen atom is exchanged with a group on an adjacent carbon, via radical intermediates created by adenosyl-cobalamin. The reaction is described in Figure 1-10, which is illustrated with methylmalonyl CoA mutase. All enzymes involved in this kind of rearrangement (and their exchange group) are listed in Table 1-1. Adenosyl-cobalamin also participates in reduction of ribonucleotides to deoxyribonucleotides (ribonucleotide reductase). For this reason, it has been argued that adenosyl-cobalamin is an evolutionary ancient coenzyme.

Methyl-cobalamin participates in methyl group transfer. Methionine synthase catalyses the transfer of the methyl group from methyltetrahydrofolate to cob(I)alamin, which is transformed in the methylcobalamin form. The methyl group is then transferred to homocysteine, generating tetrahydrofolate, cobalamin and methionine (Figure 1-10) [Fenton and Rosenberg, 1989]. Another enzyme which may also require methyl-cobalamin as a cofactor is the N5-methyltetrahydromethanopterin: coenzyme M methyltransferase (Mtr). This membrane associated enzyme complex catalyses the methylation of coenzyme M in the process of methanogenesis [Harms and Thauer, 1996].

Table 1-1- Reactions catalysed by adenosyl-cobalamin.

| Enzyme | Exchanged group | Reaction |
|---|---|---|
| **Carbon skeletal rearrangements:** | | |
| methylmalonyl CoA mutase | -CO-S-CoA | L-methylmalonyl CoA -> succinyl CoA |
| glutamate mutase | -CH-$COO^-$ \| $NH_3^+$ | L-glutamate -> threo-3- methyl aspartate |
| α-methyleneglutarate mutase | -C=$CH_2$ \| $COO^-$ | α-methyleneglutarate -> β-methylita-conate |
| isobutyryl CoA mutase | -CO-S-CoA | isobutyryl CoA -> n-butyryl CoA |
| **Dehydration with rearrangements:** | | |
| propanediol dehydratase | $-OH^-$ | 1,2-propanediol -> propionaldehyde |
| glycerol dehydratase | $-OH^-$ | glycerol -> β-hydroxypropionaldehyde (and acrolein) |
| ethanolamine ammonia lyase | $-NH_3^+$ | ethanolamine -> acetaldehyde and $NH_3$ |
| **Ribonucleotide reductase:** | | |
| ribonucleotide triphosphate reductase | | nucleotide -> deoxynucleotide |
| **Amino-group migrations:** | | |
| β-lysine aminomutase | $-NH_3^+$ | β-lysine -> 3,5-diaminohexanoate |
| D-ornithine aminomutase | $-NH_3^+$ | ornithine -> 2,4-diaminopentanoate |
| D-lysine aminomutase | $-NH_3^+$ | α-lysine -> 2,5-diaminohexanoate |
| **Others:** | | |
| queuosine synthase | | epoxyqueuosine -> queuosine |

## 1-2-2- Vitamin $B_{12}$ in human life.

Human and animals are unable to synthesise cobalamin *de novo* and therefore acquire it from their diet. Humans need about 2 µg of vitamin $B_{12}$ everyday. Natural deficiency in this vitamin is rare as the total amount of vitamin $B_{12}$ in the body is about 3 to 5 mg (the majority is stored in the liver). In the event of cobalamin deficiency, the hepatic store of the vitamin is able to last for about 5 years. Deficiency of vitamin $B_{12}$ leads to a megaloblastic anemia called pernicious anemia, abnormal intestinal function and specific neurological disorders, including neurologic degeneration and dementia.

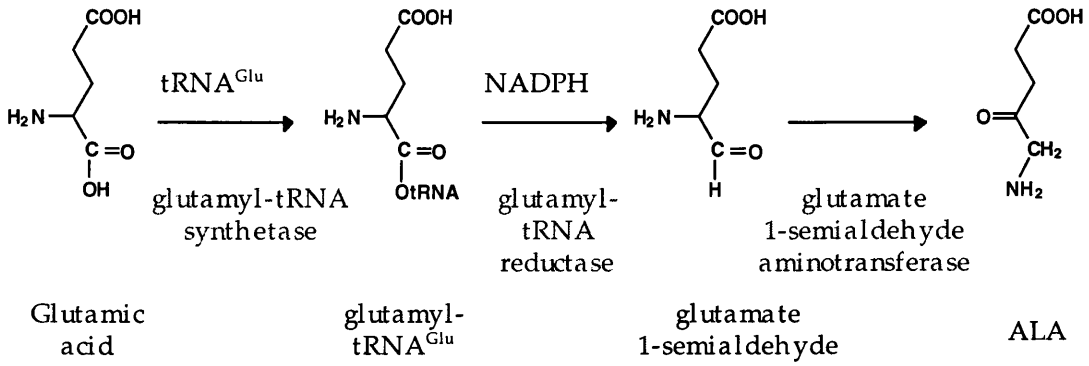
In the absence of adenosyl-cobalamin, abnormalities in metabolism occur due to a blockage in propionate catabolism in which the methylmalonyl CoA mutase reaction is implicated.

The methyl transfer catalysed by methionine synthase also involves the transformation of methyltetrahydrofolate into tetrahydrofolate. The absence of cobalamin involves a functional tetrahydrofolate deficiency. This depletion interferes with other reactions requiring tetrahydrofolate, such as the synthesis of nucleic acid precursors. [Fenton and Rosenberg, 1989].

## 1-3- Uroporphyrinogen III biosynthesis.

The first step in the biosynthesis of all tetrapyrroles, the synthesis of 5-aminolaevulinic acid (ALA), is undertaken by one of two distinct routes. Higher plants, algae, cyanobacteria and several other photosynthetic and non-photosynthetic bacteria produce ALA by the C-5 pathway (Figure 1-11) whereas most eukaryotes (except higher plant and algae) and photosynthetic bacteria (α subclass of the purple bacteria) use the C-4 pathway, also called the Shemin pathway (Figure 1-12).

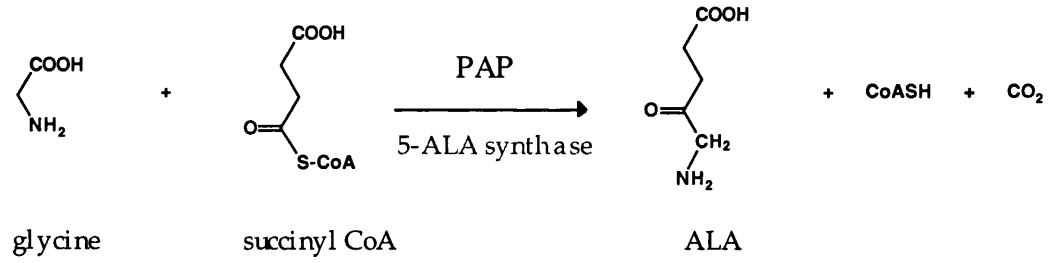
Figure 1-11- Biosynthesis of 5-aminolaevulinic acid via the C-5 pathway.




ALA, 5-aminolaevulinic acid.

The C-5 pathway leads to the synthesis of 5-aminolaevulinic acid from the intact carbon skeleton of glutamic acid by a multi-enzyme pathway. Glutamic acid is activated at the α-carboxyl by ligation to $tRNA^{Glu}$ with an aminoacyl-tRNA synthetase, the glutamyl-$tRNA^{Glu}$ synthetase. An NADPH-dependent reductase (HemA) converts glutamyl-$tRNA^{Glu}$ to glutamate 1-semialdehyde, which is finally transformed by the glutamate 1-semialdehyde aminotransferase (HemL) into 5-aminolaevulinic acid [Kannangara *et al.* 1994].

Figure 1-12- Biosynthesis of 5-aminolaevulinic acid via the Shemin pathway.




ALA, 5-aminolaevulinic acid and PAP, pyridoxal 5' -phosphate.

The C-4 or Shemin pathway leads to the synthesis of 5-aminolaevulinic acid from the decarboxylative condensation of succinyl-CoA and glycine, which is mediated by pyridoxal phosphate in a single step catalysed by the enzyme 5-aminolaevulinic acid synthase (also named HemA).

The C-5 pathway is thought to be the more primitive route since it is found in archaebacteria. Interestingly, *E. coli, S. typhimurium* and *Bacillus megaterium* use this more ancient route whereas *P. denitrificans* and *Saccharomyces cerevisiae* use the more evolutionary recent step.

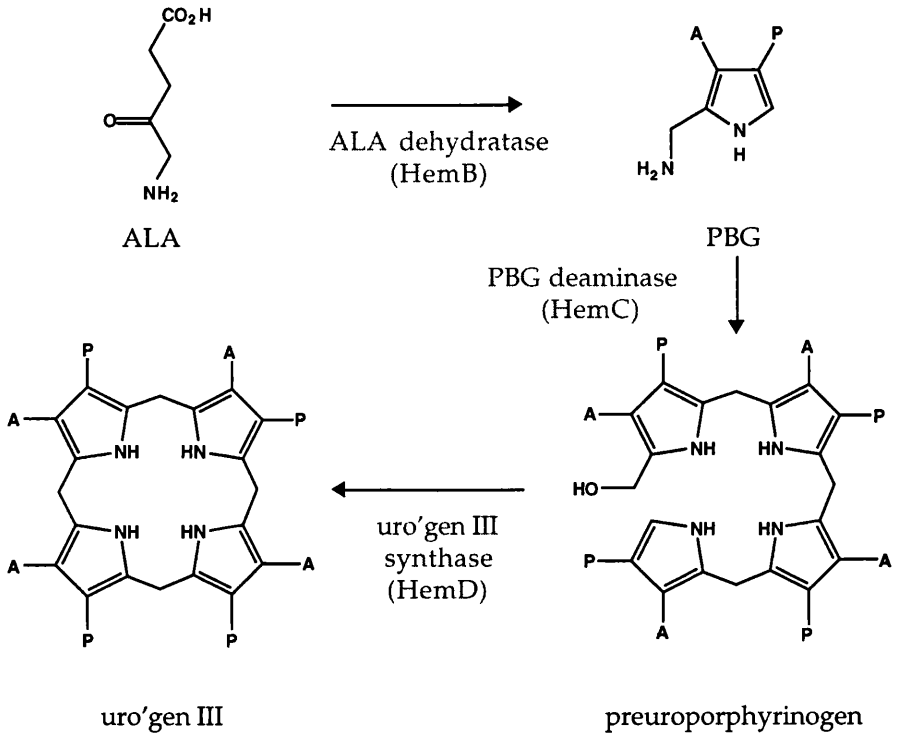
From 5-aminolaevulinic acid, the synthesis of uroporphyrinogen III is similar in all living organisms and requires just three enzyme mediated steps, ALA dehydratase, PBG deaminase and uroporphyrinogen III synthase (Figure 1-13).

Two molecules of aminolaevulinic acid are condensed to yield porphobilinogen, thereby constructing the mono-pyrrole building block. This step is catalysed by 5-aminolaevulinic acid dehydratase (HemB). The enzyme has been purified from a variety of sources and all are thought to use a similar catalytic mechanism. However, the enzymes demonstrate a differential metal dependency such that the *E. coli* HemB requires zinc but is stimulated by magnesium whereas the mammalian and *S. cerevisiae* enzyme are solely zinc dependent whilst, finally, the pea enzyme is solely magnesium dependent [N. Senior PhD thesis, 1996].

The next enzyme in the pathway catalyses the head to tail polymerisation of four molecules of porphobilinogen, leading to the generation of an open tetrapyrrole molecule named either hydroxymethylbilane or preuroporphyrinogen. This step is catalysed by PBG deaminase, an enzyme which contains a dipyrromethane molecule as a cofactor. The dipyrromethane acts as the site for pyrrole attachment and elongation.

Uroporphyrinogen III is subsequently synthesised from preuroporphyrinogen after inversion of ring D and closure of the macrocycle by uro'gen III synthase [Jordan, 1994].

Figure 1-13- From aminolaevulinic acid to uroporphyrinogen III.

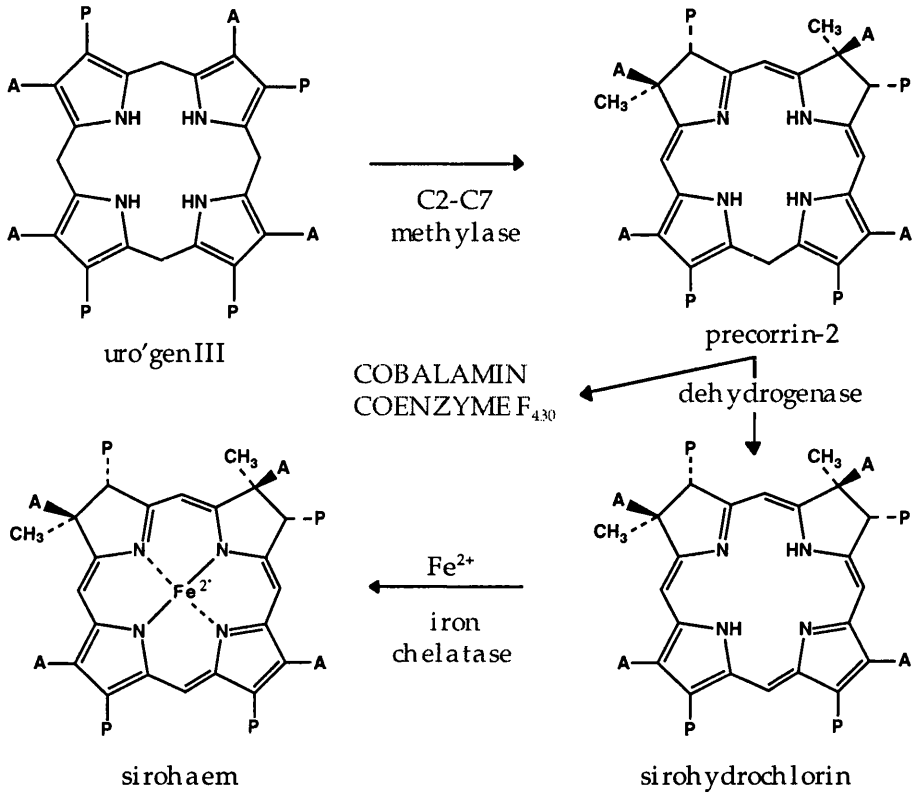



ALA, 5-aminolaevulinic acid; PBG, porphobilinogen; uro'gen III, uroporphyrinogen III.

## 1-4- Sirohaem biosynthesis.

Uroporphyrinogen III is initially methylated at carbon C-2 and C-7 in a reaction that requires two molecules of S-adenosyl methionine (SAM) to give dihydrosirohydrochlorin (precorrin-2). This intermediate is the branch point at which the pathways for sirohaem, coenzyme $F_{430}$ and cobalamin diverge. For the completion of sirohaem synthesis, precorrin-2 is dehydrogenated in an $NAD^+$ dependent manner to give sirohydrochlorin which is subsequently chelated with ferrous iron (Figure 1-14) [Warren and Scott, 1990]. As described in the following section, the enzymes which undertake these three steps vary between organisms.

Figure 1-14- From uro'gen III to sirohaem.




1-4-1- In *Escherichia coli* and *Salmonella typhimurium*

In *E. coli* and *S. typhimurium*, sirohaem is synthesised from uroporphyrinogen III by a unique multifunctional enzyme, the sirohaem synthase. This enzyme is also named CysG because of its involvement in cysteine biosynthesis (Figure 1-7). CysG is a protein of 457 amino acids and can be described as a uro'gen III methylase, a precorrin-2 $NAD^+/NADP^+$ -dependent dehydrogenase and sirohydrochlorin ferrochelatase. The carboxyl domain of CysG is similar to the *P. denitrificans* CobA which only catalyses the double methylation at C-2 and C-7 of uro'gen III. Gene dissection of CysG has shown that a truncated protein containing the C-terminal part of CysG (amino acids 202 to 457) is still able to perform the double methylation. These results strongly suggest that CysG evolved as a consequence of a gene fusion between a uro'gen III methylase and a separate oxidase/chelatase enzyme [Warren *et al*, 1994].

In *S. typhimurium*, it was suggested that CysG was also required for the first reaction specific to cobalamin biosynthesis, namely the cobalt chelation [Fazzio and Roth, 1996]. The exact role of CysG in cobalamin biosynthesis is discussed in Chapter 3.

### 1-4-2- In *Saccharomyces cerevisiae.*

In *S. cerevisiae*, sirohaem is used exclusively in sulfite reduction as this organism does not possess a nitrite reductase. Two distinct *S. cerevisiae* mutant strains were found to incur a methionine auxotrophy, which could be overcome by complementation with the *S. typhimurium cysG*. The genes corresponding to the mutations, MET1 and MET8, have been isolated and sequenced. Met1p is a fusion protein between an unknown N-terminal domain and uro'gen III methylase (similar to the C-terminal domain of CysG). Met8p is not similar to CysG, except for a potential $NAD^+$ binding site. Therefore, because of this common motif and the fact that this enzyme has at least one identical function as CysG, it was suggested that Met8p could be the oxidase/chelatase to complete the synthesis of sirohaem [Hansen *et al.*1997].

## 1-5- Vitamin $B_{12}$ biosynthesis pathway in *Pseudomonas denitrificans.*

About 30 reactions are required to transform uro'gen III into cobalamin. In the obligate aerobe *Pseudomonas denitrificans*, 22 *cob* genes have been identified and localised in four distinct loci within the genome, referred to as complementation groups A, B, C and D. Eight open reading frames located within the groups B, C and D have not been classified as *cob* genes as mutations in these ORF did not lead to a $B_{12}$ deficiency (Figure 1-15).

The process by which uro'gen III is transformed into adenosyl-cobalamin has been almost completely elucidated. At least 20 enzymes are involved in this transformation: CobA, -I, -G, -J, -M, -F, -K, -L, -H, -B, -N, -S, -T, Cob(II) reductase, CobO, -Q, -C, -D, protein α, CobP, -U, and -V (Figure 1-16). No

functions have been ascribed to the *cob* genes *cobE* and *cobW* [Thibaut, 1996; Blanche *et al.* 1995 (Review)].

The first of these enzymes, uro'gen III methyltransferase (CobA), converts uro'gen III into precorrin-2 by two S-adenosyl-L-methionine (SAM) dependent transmethylations at positions 2 and 7 [Blanche *et al.* 1989]. A further methylation at C-20 catalysed by CobI produces precorrin-3A [Thibaut *et al.* 1990].

CobG, a monooxygenase, catalyses the oxidative reaction leading to precorrin-3B. Indeed, CobG adds a hydroxyl group (derived from molecular oxygen) at C-20 which induces the lactonisation of ring A to C-1. [Debussche *et al.* 1993; Spencer *et al.* 1993]. Ring contraction of precorrin-3B is promoted during the subsequent transmethylation at position C-17, a reaction catalysed by CobJ and which yields precorrin-4. This molecule still contains an acetyl group linked to C-1, which must be removed before the methylation at that site [Debussche *et al.* 1993]. A further SAM-dependent transmethylation at C-11 is catalysed by CobM and leads to the production of precorrin-5 [Debussche *et al.* 1993]. After deacylation of the C-1 acetyl group, CobF undertakes the sixth methylation at C-1, reactions that produce precorrin-6x. It is not clear whether the deacylation is also catalysed by CobF [Debussche *et al.* 1993; Min *et al.* 1993].

The oxidation state of the macromolecule, which is changed during the oxidative ring contraction process, is re-adjusted to the level of a hexahydroporphyrin (precorrin-6y) by the action of CobK in a process that requires NADPH [Blanche *et al.* 1992a].

The final methylations of C-5 and C-15 are afforded by CobL which also decarboxylates the acetate side chain at C-12 thereby producing precorrin-8x [Blanche *et al.* 1992b]. A rearrangement of the methyl group previously added to C-11 by CobM to position C-12 is catalysed by CobH which produces precorrin-8y, hydrogenobyrinic acid [Thibaut *et al.* 1992; Blanche *et al*, 1990]. The peripheral side chains a and c are amidated by CobB which gives hydrogenobyrinic acid a, c-diamide [Desbussche *et al.* 1990].

Cobalt chelation by an enzyme complex formed with CobN, -S and -T occurs [Desbussche *et al.* 1992] and is followed by a cobalt reduction catalysed by a cob(II)yrinic acid a, c-diamide reductase [Blanche *et al.* 1992c]. It is only then that the adenosylation catalysed by CobO occurs [Desbussche *et al.* 1991]. CobQ subsequently performs four amidations of the peripheral side chains (b, d, e and g) producing adenosyl-cobyric acid [Blanche *et al.* 1991]. Cobinamide synthesis is completed by the action of CobC and -D, together with another protein termed α, through the attachment of aminopropanol to the cobyric acid intermediate at C-17 [Blanche *et al.* 1995].

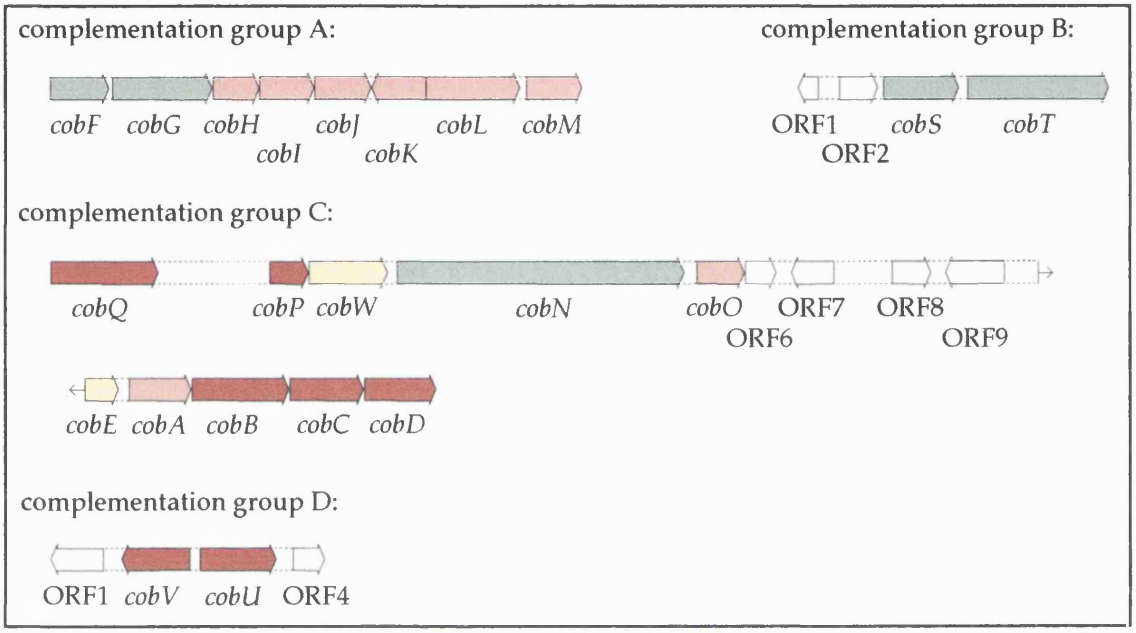
From cobinamide to cobalamin, three further reactions are required. CobP was found to be a bifunctional enzyme catalysing the phosphorylation of cobinamide and also the next step, the transfer of the GMP moiety from GTP to give adenosyl-GDP-cobinamide [Blanche *et al.* 1991].

Before the ultimate transformation of the macromolecule into cobalamin can be achieved, the lower axial ligand, α-ribazole, a molecule of 5,6-dimethylbenzimidazole linked to a ribose by an α-glycosidic bond, has to be synthesised. The first step of the synthesis of α-ribazole is catalysed by CobU and produces α-ribazole-phosphate from 5,6-dimethylbenzimidazole. A mutation in the corresponding gene leads to a cobalamin minus phenotype. A phosphatase (uncharacterised) hydrolyses α-ribazole-phosphate into α-ribazole.

Finally, CobV, or cobalamin synthase, completes the synthesis of cobalamin by exchange of GMP with α-ribazole [Blanche *et al.* 1995; Cameron *et al.* 1991].

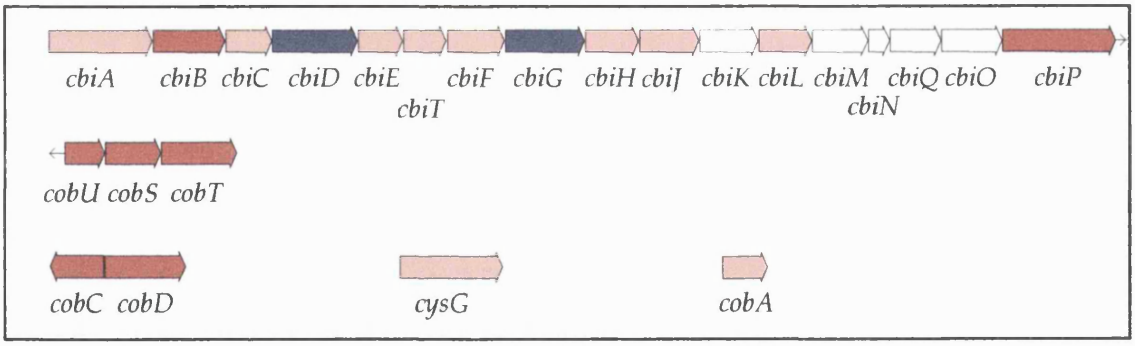
Figure 1-15- Genetic organisation of vitamin $B_{12}$ biosynthetic loci.

## *P. denitrificans*



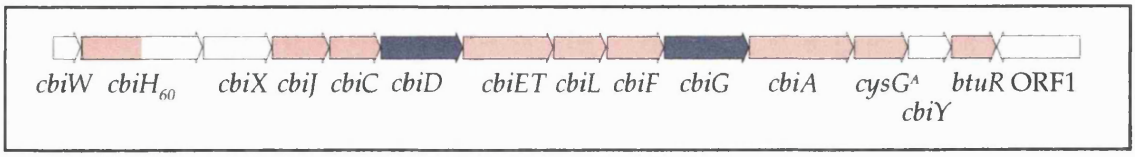

In *P. denitrificans*, the cobalamin biosynthetic genes are mainly located on four separate loci which are described as complementation groups A to D.

## *S. typhimurium*




In *S. typhimurium*, one major cobalamin operon has been described (cob*I-III-II*). Three other loci also contain *cob* genes.

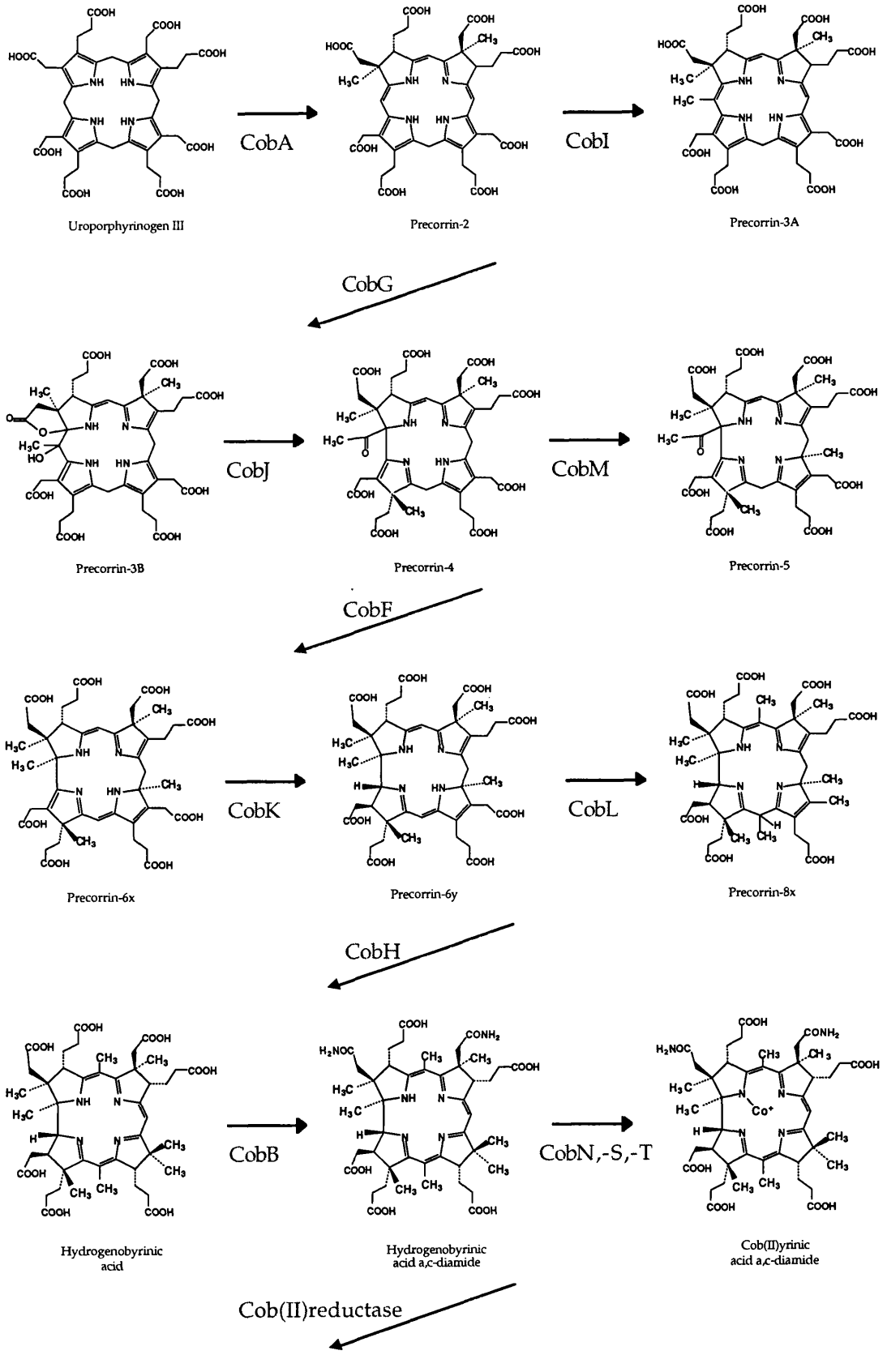
## *B. megaterium*

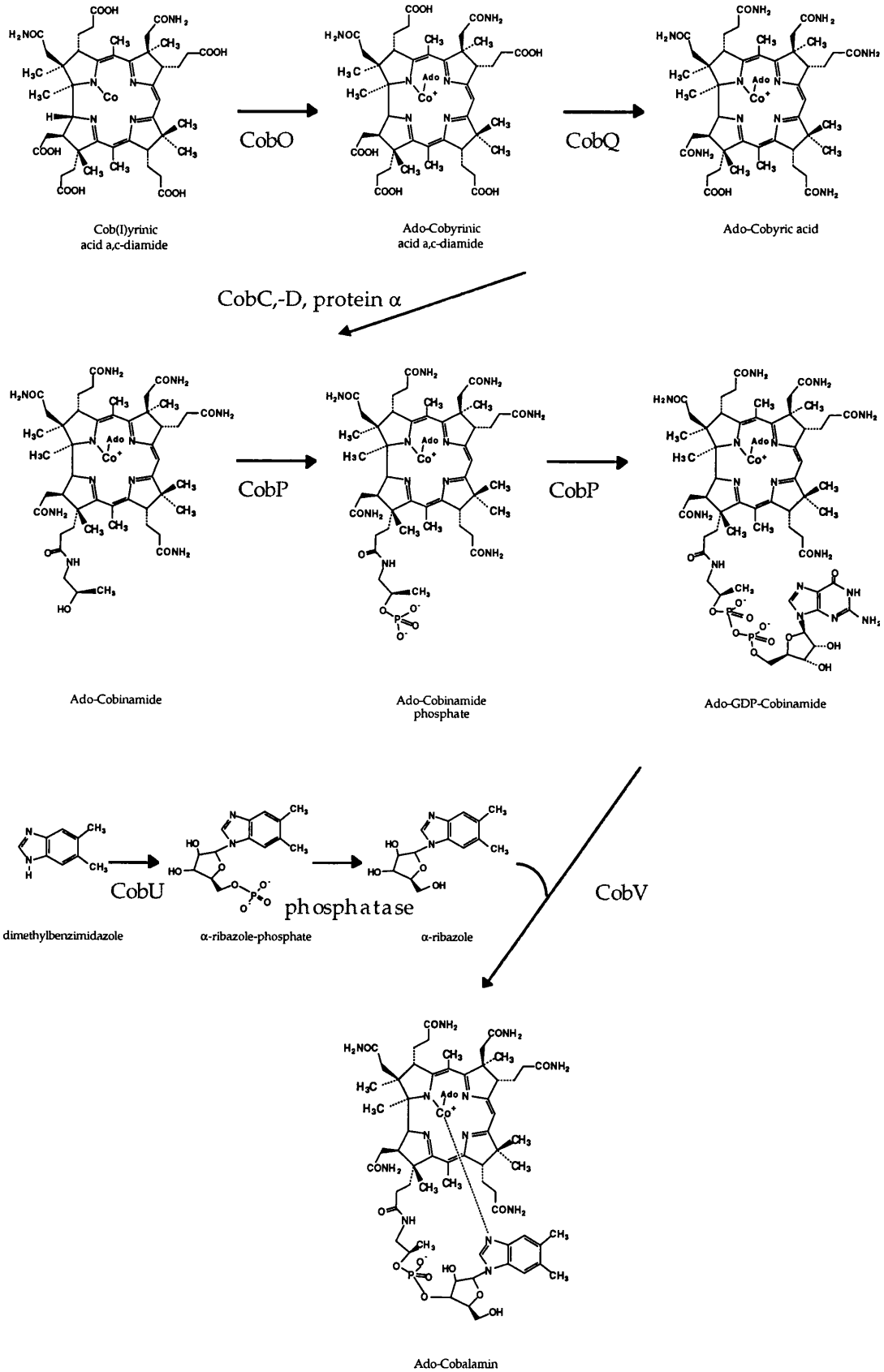



In *B. megaterium*, only one major cobalamin operon has been described. It contains *cobI* genes. Other *cobI* genes and the *cobII* and *cobIII* are located somewhere else on the genome.

The boxes filled in pink indicate that these genes are involved in cobyrinic acid a, c diamide synthesis and are similar in the three species. The boxes filled in red are involved in the transformation of cobyrinic acid a, c diamide to cobalamin. The boxes filled in purple are genes found in *S. typhimurium* and *B. megaterium* only. The boxed filled in green are genes specific to *P. denitrificans* and finally the boxes filled in yellow are genes involved in cobalamin biosynthesis in *P. denitrificans* but these two genes are not characterised.

Figure 1-16- The biosynthesis of cobalamin in *P. denitrificans.*






Cob(I)yrinic acid a,c-diamide

Ado-Cobyrinic acid a,c-diamide

Ado-Cobyric acid

Ado-Cobinamide

Ado-Cobinamide phosphate

Ado-GDP-Cobinamide

dimethylbenzimidazole

α-ribazole-phosphate

α-ribazole

Ado-Cobalamin

## 1-6- Vitamin $B_{12}$ biosynthesis in *Salmonella typhimurium*.

In *S. typhimurium*, the biosynthesis of vitamin $B_{12}$ occurs only under anaerobic growth conditions even though this bacterium is a facultative aerobe [Jeter *et al.* 1984]. This micro-organism contains four known $B_{12}$ dependent enzymes, methionine synthase (MetH), propanediol dehydratase, ethanolamine ammonia lyase and queuosine synthase. None of these reactions seems to justify the synthesis of the vitamin. For example, *S. typhimurium* can produce methionine in absence or presence of oxygen as the bacteria also contains an alternative way to produce methionine in the absence of methyl-cobalamin, the MetE route [Roth *et al.* 1993].

The biosynthesis of vitamin $B_{12}$ is divided into three parts named *cobI*, *cobII* and *cobIII*. The different classes have been determined by analysing the effect of cobinamide, dimethylbenzimidazole or vitamin $B_{12}$ on *S. typhimurium* $B_{12}$ mutant strains:

- *cobI* includes all genes encoding for the conversion of uro'gen III into cobinamide.
- *cobII* includes genes encoding for the synthesis of dimethylbenzimidazole.
- *cobIII* includes all the genes encoding for proteins involved in the assembly of dimethylbenzimidazole with cobinamide.

A single operon mapped at 41 min on the *S. typhimurium* genome has been isolated and found to contain the *cobI-III-II* genes. The *cob* operon is co-regulated with the 5′ propanediol operon (*pdu*) by pocR, which is encoded by a gene located between the two operons. The *cob* operon contains 20 genes (Figure 1-15) of which 13 are similar to *P. denitrificans cob* genes ($cbiA^{St}$ ~ $cobB^{Pd}$; $cbiB^{St}$ ~ $cobD^{Pd}$; $cbiC^{St}$ ~ $cobH^{Pd}$; $cbiE^{St}$ *and* $cbiT^{St}$ ~ $cobL^{Pd}$; $cbiF^{St}$ ~ $cobM^{Pd}$; $cbiH^{St}$ ~ $cobJ^{Pd}$; $cbiJ^{St}$ ~ $cobK^{Pd}$; $cbiL^{St}$ ~ $cobI^{Pd}$; $cbiP^{St}$ ~ $cobQ^{Pd}$ and $cobU^{St}$ ~ $cobP^{Pd}$; $cobS^{St}$ ~ $cobV^{Pd}$; $cobT^{St}$ ~ $cobU^{Pd}$) (Figure 1-15). It has been speculated that four of the genes might be involved in cobalt transport, *cbiM*, *-N*, *-Q* and *-O*. No similar genes to the *P. denitrificans cobC*, *-E*, *-F*, *-G*, *-N*, *-S*, *-T*, *-W* have been found in the *S. typhimurium cobI-III-II* operon. Conversely, three of the *S. typhimurium cbi*

genes, *cbiD, -G, -K,* have no equivalent in *P. denitrificans* [Roth *et al.* 1993; Raux *et al.* 1996].

Three other regions on the chromosome are known to be involved in vitamin $B_{12}$ biosynthesis:
(1) at 14 min, *cobC* and *cobD*. *cobC* encodes an α-ribazole-5'-phosphate phosphatase [O'Toole *et al.* 1994] and *cobD*, for a pyridoxal phosphate-dependent decarboxylase which synthesises (*R*)-1-amino-2-propanol *O*-2-phosphate (CobD is similar to the *P. denitrificans* CobC, 29.5% on 349 amino acids) [Brushaber *et al.* 1998];
(2) at 34 min, *cobA*, which encodes the adenosylation of the corrin molecules (similar to the *P. denitrificans* CobO);
(3) at 73 min, *cysG*, which is involved in the first two methylations of uro'gen III leading to precorrin-2 and also able to insert the cobalt (similar to the *P. denitrificans* CobA on the last 250 amino acids) [Spencer *et al.* 1993; Fazzio and Roth, 1996].

The results concerning the *S. typhimurium cobI-III-II* operon obtained previously to this thesis, by the team headed by Dr Alain Rambach in France are described in the introduction of Chapter 2.

## 1-7- Vitamin $B_{12}$ biosynthesis in *Bacillus megaterium.*

In *B. megaterium*, the biosynthesis of vitamin $B_{12}$ occurs when the cells are grown aerobically. Therefore, the presumed pathway used in this bacterium was thought to be similar to the *P. denitrificans* pathway.

In 1986, two studies on *B. megaterium* biosynthesis were published [Wolf and Brey, 1986; Brey *et al.* 1986]. They described the identification of thirty-four $B_{12}$ auxotroph strains of which twenty-eight were complemented by cobinamide (*cobI*) and six by cobalamin (*cobIII*). No mutants were found to be blocked in dimethylbenzimidazole or sirohaem biosynthesis. The mutations were localised in four different loci, named complementation groups A, B, C and D. Within group A, two distinct mutants accumulated cobalt free intermediates whereas

four mutants accumulated cobalt-containing intermediates. The three mutants constituting group B all produced cobalt free intermediates whereas five mutants from group C were found to accumulate cobalt-containing intermediates. All mutants from group A, B and C were complemented by cobinamide. In group D, two mutants were complemented by cobalamin only.

Subsequent to these studies, the *cobA* gene of *B. megaterium* was isolated and sequenced [Robin *et al.* 1991]. Unfortunately, it was not made clear where the gene was localised on the genetic maps or from which complementation group it came. Initially, *cobA* was isolated from a 4 kb DNA fragment, which was further restricted to a 2 kb and finally to a 1.3 kb DNA fragment. The 2 kb fragment was able to complement an *E. coli cysG* strain proving that methylase activity as well as dehydrogenase and iron-chelatase activities, essential for sirohaem synthesis, were present within the 2 kb fragment. In the smallest fragment, only the methylase activity was present and the gene product (M62881) corresponding to a CobA shared 43.5% identity with the *P. denitrificans* CobA. The dehydrogenase and iron-chelatase activities must be encoded by a gene flanking *cobA*.

In 1993, a *B. megaterium cob* operon containing fourteen genes was isolated and fully sequenced. Nine genes of this operon are similar to *P. denitrificans cob* and *S. typhimurium cbi* genes: $cbiH_{60}$ (only the 5' half of the gene), *cbiJ, -C, -ET* (one gene like for the *P. denitrificans cobL*), *cbiL, -F, -A*, $cysG^{A}$ (~ $cobA^{Pd}$) and *btuR* (~ $btuR^{Ec}$, $cobO^{Pd}$ and $cobA^{St}$). Two genes are similar to *S. typhimurium cbi* genes, *cbiD* and *cbiG*, and three genes have no equivalent, *cbiW*, *-X* and *-Y* (Figure 1-15).

The *cob* operon recently isolated, in comparison with the previous study on *B. megaterium* $B_{12}$ biosynthesis, is consistent with complementation group A in which cobalt-free and cobalt-containing intermediates were isolated. However, the $CysG^{A}$ sequence from this operon is only 63% identical to the CobA sequence published by Robin *et al.* The *B. megaterium* from which the *cob* operon have been isolated is the DSM509 strain equivalent to the ATCC10778 used by Brey *et al.* Identical $cysG^{A}$ sequences have been found from these two strain origins showing that this $cysG^{A}$ is a real *B. megaterium* gene [R. Beck PhD

thesis, 1997]. Analysis of the *B. subtilis* genome shows that this strain harbours two distinct genes, *ylnD* and *nasF*, encoding for uro'gen III methyltransferase. Therefore, *B. megaterium* might also contain two different *cobA* genes.

## 1-8- Vitamin $B_{12}$ biosynthesis in *Propionibacterium shermanii.*

*P. shermanii*, also known as *P. freudenreichii*, is an aerotolerant anaerobe and synthesises vitamin $B_{12}$. Two loci have been recently identified and isolated. One of these DNA fragments was found to contain *cbiM, -N, -Q, -O* and *cobA* (equivalent to the *P. denitrificans cobA*) whereas a second locus encoded an uncharacterised ORF, *cbiL, -F, -ET, -C, -D* [Roessner *et al.* 1996].

Previous experiments on crude cell lysates of *P. shermanii* revealed a major difference in the timing of cobalt insertion in the macrocycle in comparison with *P. denitrificans*. Indeed, a crude extract made from the soluble fraction of *P. shermanii* cells, together with all the required cofactors, can transform porphobilinogen into corrins. When the crude extract was made from cells which had been grown in the presence of cobalt, porphobilinogen was transformed into cobyrinic acid. This intermediate is not generated via the aerobic pathway. The equivalent to cobyrinic acid in *P. denitrificans* is hydrogenobyrinic acid which is still cobalt free (Figure 1-17). When the crude extract was made from cells which had been grown in the absence of cobalt, cobyrinic acid was not detected but three early intermediates were identified, precorrin-1, -2 and -3A. The absence of cobalt interrupted the pathway at precorrin-3A, implying that cobalt insertion occurs at or close to that step [Blanche *et al.* 1995].

These experiments on *P. shermanii* provided proof that the cobalamin pathway used by anaerobic organisms was divergent, in regards to the timing of cobalt insertion, to the aerobic pathway followed in *P. denitrificans*.

## 1-9- Two different pathways for cobalamin biosynthesis.

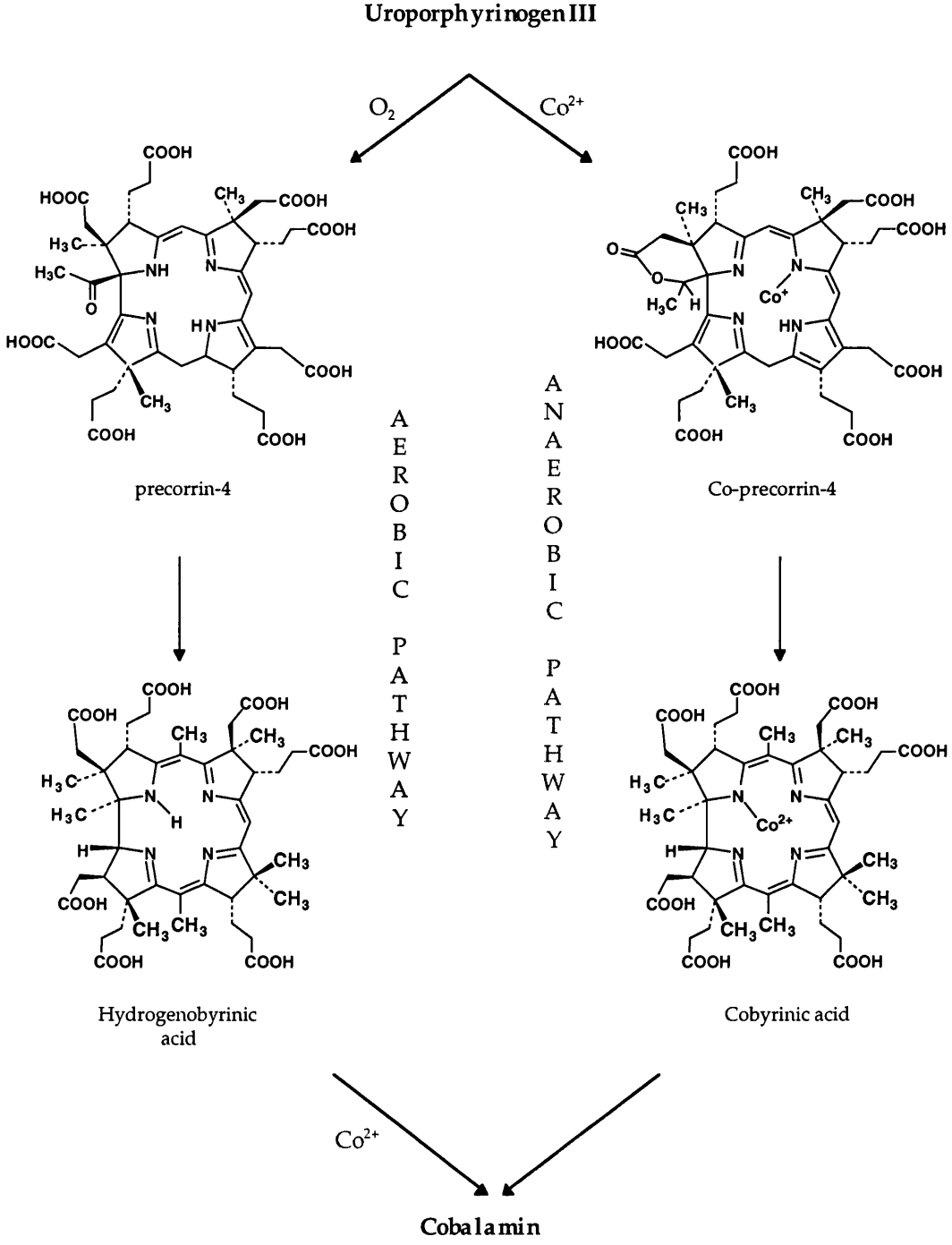
The first evidence that bacteria used different pathways for synthesising vitamin $B_{12}$ came not only from the *P. shermanii* experiments on the timing of cobalt insertion, described in the previous section, but also from the fact that molecular oxygen was required as a substrate in the *P. denitrificans* pathway. Indeed, in the anaerobe *P. shermanii*, cobalt is inserted at an early stage of the biosynthesis (precorrin-2 or precorrin-3A, prior to ring contraction [Muller *et al.* 1991]) whereas in *P. denitrificans* cobalt is added at least six steps after ring contraction. The ring contraction process in *P. denitrificans* is also dependent upon the presence of molecular oxygen, which is not available for strict anaerobes. To overcome the lack of molecular oxygen required for the ring contraction procedure, anaerobes may use the oxygen of the carboxyl group of ring A acetate to extrude C-20 [Scott *et al.* 1996]. The differences in the timing of cobalt insertion and oxygen requirement are also reflected in the genetic content of the various organisms. The monooxygenase CobG, which catalyses the formation of the γ-lactone initiating ring contraction, and the enzyme-complex responsible for cobalt insertion (CobN, -S and -T) in *P. denitrificans* are not present in *P. shermanii* or *S. typhimurium*.

Although the step-by-step biosynthesis of the *P. denitrificans* pathway has now been elucidated, little is known about the intermediates in the anaerobic pathway between precorrin-2 and cobyrinic acid. Recently, a new intermediate, cobalt factor IV, the oxidised form of cobalt precorrin-4, has been isolated and identified from *P. shermanii* cell extracts [Scott *et al.* 1996] and has also been produced by the incubation of precorrin-3A, SAM and cobalt in a cell extract obtained from a recombinant *E. coli* containing the *S. typhimurium cbiH* [Santander *et al.* 1997]. Although this new intermediate has been isolated in only vanishing quantities, it proves that the cobalt is inserted prior to the fourth methylation in the anaerobic pathway. A further difference with the aerobic pathway was noticed when the C-20 and its attached methyl group, were

found to be excluded as acetaldehyde in *P. shermanii*, whereas *P. denitrificans* released them as acetic acid [Wang *et al.* 1996]. Moreover, the protein which may catalyse the deacylation in *P. denitrificans*, the methylase CobF, is absent in *S. typhimurium*. In fact, it now appears that the two pathways are divergent until the amidation of cobyrinic acid occurs via the anaerobic pathway, and until the amidation of hydrogenobyrininc acid and cobalt insertion occur via the aerobic pathway (Figure 1-17).

The *B. megaterium cob* operon, which was initially proposed to be similar to that found in *P. denitrificans* due to its ability to produce vitamin $B_{12}$ aerobically, seems to contain none of the characteristics of the aerobic pathway; no *cobG*, *-F*, *-N*, *-S* and *-T*. Conversely, *B. megaterium* contains two genes similar to *S. typhimurium*, *cbiD* and *cbiG*. It appears therefore that *B. megaterium* contains all the hallmarks of the anaerobic pathway, an anaerobic pathway which can operate aerobically. Interestingly, *cbiD* is also found in *P. shermanii* and it seems likely that *cbiD* and *cbiG* may be specific genes of the anaerobic pathway.

Figure 1-17- Differences between the aerobic and anaerobic cobalamin pathways.



## 1-10- Purpose of this thesis.

The aim of this thesis is to study the biosynthesis of vitamin $B_{12}$ in *S. typhimurium* and *B. megaterium.*

These two organisms contain no equivalent of the *P. denitrificans* genes involved in the initiation of the ring contraction (*cobG*) and cobalt insertion (*cobN, -S and -T*) and have two genes in common, *cbiD* and *cbiG*, which are not present in *P. denitrificans.* The first question to address was to determine whether *cbiD* and *cbiG* are essential genes in cobalamin biosynthesis. This study was extended in order to determine if all the unique genes within the *S. typhimurium* and *B. megaterium cob* operons are involved in $B_{12}$ biosynthesis. This includes the *S. typhimurium cbiK* and the *B. megaterium* $cbiH_{60}$, *-W, -X* and *-Y*. Once this was accomplished, a characterisation of the corresponding gene product, if involved in vitamin $B_{12}$ biosynthesis, was attempted.

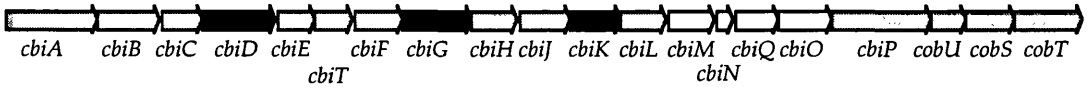
# CHAPTER 2,

## IDENTIFICATION OF THE GENES REQUIRED FOR VITAMIN $B_{12}$ BIOSYNTHESIS VIA THE ANAEROBIC PATHWAY

## 2-1- Introduction

In *S. typhimurium*, the genes which encode the enzymes responsible for the transformation of precorrin-2 into cobalamin are clustered in a single operon at 41 mins on the *S. typhimurium* chromosome. This operon contains 20 genes which are all translationally coupled, *cbiA, -B, -C, -D, -E, -T, -F, -G, -H, -J, -K, -L, -M, -N, -Q, -O, -P, cob U, -S* and *-T*. Thirteen of these genes are similar to *P. denitrificans cob* genes (Figure 2-1). The genes *cbiD, -G, -K, -M, -N, -Q* and *-O* are not similar to the *P. denitrificans cob* genes [Roth *et al.* 1993; Raux *et al.* 1996] but are found in other anaerobic cobalamin producers like *C. acetobutylicum, Synechocystis sp.* or *M. jannaschii* (Table 7-1).

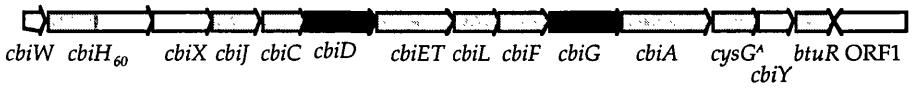
Figure 2-1- The *S. typhimurium cobI-III-II* operon.




The grey-filled boxes indicate genes, which are similar to the *P. denitrificans cob* genes. The dark boxes indicate genes, which are specific to *S. typhimurium*. The white-boxed genes are involved in cobalt transport.

In *B. megaterium*, most of the cobalamin biosynthetic genes are clustered in one main operon which contains 14 genes, *cbi W, -$H_{60}$, -X, -J, -C, -D, ET, -L, -F, -G, -A, cysG$^A$, cbiY* and *btuR*. Nine of these genes have a significant similarity with *P. denitrificans cob* genes (Figure 2-2). The genes *cbiW, -X, -D, -G, -Y* and the 3' extension of *cbi$H_{60}$* are not similar to any of the *P. denitrificans cob* genes [Raux *et al.* 1998].

Figure 2-2- The *B. megaterium cob* operon.




The grey-filled boxes indicate genes which are similar to the *P. denitrificans cob* genes. The dark boxes indicate genes which are found in the *S. typhimurium* and *B. megaterium cob* operons but not among the *P. denitrificans cob* genes. The white-boxed genes do not have a homologue in *S. typhimurium* and *P. denitrificans*.

Previously, a number of *S. typhimurium cob* mutants had been generated, isolated and characterised by complementation studies. In total, 29 *S. typhimurium cob* mutants were found to be complemented by plasmids containing *S. typhimurium cbiA, -B, -D, -F, -J, -L, -P, cobS, -T* and *cysG*, which

had been cloned individually. This established a functional role for these genes in cobalamin synthesis. As *cbiD* is not found in *P. denitrificans*, this mutant revealed a major functional difference between the *S. typhimurium* and *P. denitrificans* pathways [Raux *et al.* 1996].

Among the *S. typhimurium cob* mutants, five mutant strains, *cbiA, -F, -J, -L* and *cysG*, were also complemented with equivalent genes from the *B. megaterium cob* operon. One of two *S. typhimurium cbiD* mutants (AR3718) was not complemented by a plasmid carrying the *B. megaterium cbiD* gene whereas the second mutant (AR3728) complemented the strain only after 48 hours instead of the 24 hours normally required. This result suggested that even if the two species can complement each other, some steps may differ [Raux *et al.* 1998].

No *S. typhimurium cob* mutants were complemented by DNA fragments containing *cbiM, N, Q* and *O*. An *E. coli* strain harbouring the complete *S. typhimurium cob* operon but in which *cbiM, N, Q* and *O* genes had been deleted was still able to make cobyric acid indicating that these genes could be either located on the *E. coli* genome or, more likely, are not essential for corrin production in *E. coli.* In fact, *cbiN, Q* and *O* are thought to be involved in the active transport of cobalt. CbiN and CbiQ present typical multiple membrane-spanning domains whereas CbiO is similar to the ATP-dependent membrane transport proteins. For active transport of ions, a periplasmic binding protein is also required and CbiM is a potential candidate for this role [Roth *et al.* 1993].

The second half of CbiG is slighly similar to the *P. denitrificans* CobE (27.2% identity in a 103 amino acid overlap with the *S. typhimurium* CbiG and 28.1% in 121 amino acid overlap with the *B. megaterium* CbiG). A mutation in the *P. denitrificans cobE* leads to a cobalamin-negative phenotype although no function has been ascribed to the gene product [Crouzet *et al.* 1990]. This slight similarity may suggest that CbiG, too, plays an important role in cobalamin biosynthesis.

The *S. typhimurium cbiK* and the *B. megaterium cbiD, -W,- X,* and *-Y* do not have equivalent genes within *P. denitrificans* and have not been characterised for their role in the cobalamin synthesis. Both *cbiD* and *cbiG* are particularly interesting because they appear to be specific to the anaerobic cobalamin biosynthetic pathway.

This chapter investigates the functional role of *cbiD* and *cbiG* in cobalamin biosynthesis via the anaerobic pathway and also determines if *cbiK*, *-W*, *-X*, and *-Y* are required for vitamin $B_{12}$ biosynthesis. Moreover, a study of CysG, the first and common enzyme of the sirohaem, vitamin $B_{12}$ and coenzyme $F_{430}$ pathways, was undertaken. CysG had been characterised to be a multifunctional enzyme involved in sirohaem biosynthesis [Spencer *et al.* 1993]. It is therefore important to determine the exact role of CysG in cobalamin biosynthesis.

## RESULTS

### 2-2- Use of *E. coli* as host for the *cob* operons and for generating *cob* phenotypes.

Homologous proteins found in *S. typhimurium* and *E. coli* generally display more than 90% identity [Lawrence and Roth, 1995]. For example, the CysG proteins are 90.2% identical and the *E. coli* BtuR/*S. typhimurium* CobA, which encode for the cobalamin adenosylation, are 89.3% identical.

However, one of the main differences between these two species is that *E. coli* can not synthesise cobinamide *de novo* as it does not possess the genes required for the transformation of uro'gen III into cobinamide, but it can transform cobinamide into cobalamin. Three proteins are involved in the latter transformation, the CobU, -S, -T. Between these two bacteria, CobU, -S and -T are only 76 to 84% identical suggesting that the two *cob* operons have not evolved from a common ancestor. In fact, it has been proposed that *S. typhimurium* re-acquired the *cob* operon by horizontal gene transfer [Lawrence and Roth, 1995 and 1996].

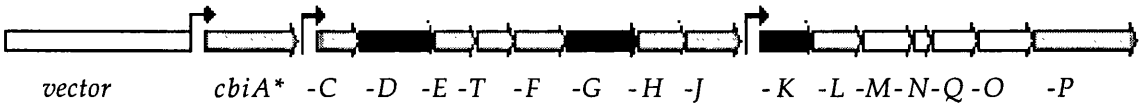
As *E. coli* does not possess a *cobI* operon, it can act as the perfect *S. typhimurium cobI* mutant and is therefore the ideal host for functional studies of the *S. typhimurium* and *B. megaterium cob* operons.

#### 2-2-1- *E. coli* as a host for the *S. typhimurium cob* operon.

An *E. coli* strain harbouring the *S. typhimurium cob* operon is able to synthesise high levels of cobalamin. Indeed, AR8981 (*E. coli* with the cosmid pAR3062, which carries the whole *S. typhimurium cobI-III-II* operon) produces cobalamin at a level of 10 pmol/$OD_{600}$ (the microbiological assay determines the amount

of cobalamin in terms of the optical density of the cells, measured at 600 nm). Cobyric acid is the immediate precursor of cobinamide and is also the first intermediate that can be detected by microbiological assay [Materials and Methods, section 8-4-3]. Therefore, all genes required for the synthesis of cobyric acid were cloned together creating plasmid pAR8827, which contains *cbiA-C-D-E-T-F-G-H-J-K-L-M-N-Q-O-P* under the control of three *tac* promoters and an improved ribosome binding site for *cbiA* as shown in Figure 2-3 [Raux *et al.* 1996].

Figure 2-3- Gene organisation in pAR8827.




The vector is derived from pACYC184 and harbours *lacI*$^q$. The arrows represent the position of the *tac* promoters. The grey-filled boxes indicate genes, which are similar to the *P. denitrificans cob* genes. The black boxes indicate genes, which are specific to *S. typhimurium*. The white-boxed genes are involved in cobalt transport.

After transformation of *E. coli* with pAR8827, the recombinant strain (AR8959) was able to produce cobyric acid at 200 pmol/$OD_{600}$. The synthesis of cobyric acid was increased further, up to 800 pmol/$OD_{600}$, when a plasmid containing *cysG* was added to the strain. This demonstrated the importance of channeling flux from the uro'gen III branch-point towards cobalamin synthesis and the need to understand the exact role of CysG in sirohaem and cobalamin biosynthesis [Raux *et al.* 1996].

## 2-2-2- Role of CysG in cobalamin synthesis.

CysG is a multifunctional protein known as sirohaem synthase and catalyses the transformation of uro'gen III to sirohaem. CysG methylates uro'gen III at C-2 and C-7, dehydrogenates the macrocycle and finally chelates ferrous iron [Spencer *et al.* 1993]. A *cysG* mutation in *S. typhimurium* leads to sirohaem as well as cobalamin deficiency phenotypes [Goldman and Roth, 1993]. With regard to cobalamin biosynthesis, CysG is responsible for the synthesis of precorrin-2, although it has also been suggested that CysG could insert cobalt into dihydrosirohydrochlorin (precorrin-2) for cobalamin biosynthesis [Spencer *et al.* 1993].

To investigate whether the three activities associated with CysG are essential for the production of cobalamin, a series of experiments were undertaken with

CysG or CysG analogues. Initially, an *E. coli cysG* mutant strain, 302Δa, [Griffiths and Cole, 1987] was cotransformed with plasmids pAR8827 (*S. typhimurium cobI* genes) and pER119 (harbouring the *P. denitrificans cobA* which encodes solely for uro'gen III methyltransferase activity). CobA is similar to the methyltransferase domain of *E. coli* and *S. typhimurium* CysG [Roessner *et al.* 1995; Blanche *et al.* 1989]. However, the *P. denitrificans* CobA was used in preference to a truncated *E. coli* CysG ($CysG^A$) as CobA does not produce large quantities of trimethylpyrrocorphin (precorrin-2 with an extra methyl group at C-12, which is associated with $CysG^A$ overexpression) [Warren *et al.* 1994].

Strain 302Δa (pAR8827) was also transformed with either a vector without a uro'gen III methylase gene to check that no cobyric acid was produced in the *cysG* mutant strain, or with a second plasmid carrying *cysG* as a positive control. A summary of the cobyric acid synthesis in the different strains is shown in Table 2-1.

Table 2-1- Comparison of cobyric acid synthesis with *cobA* or *cysG* as uroporphyrinogen III methyltransferase.

| strain | *E. coli cysG* mutant containing pAR8827 and | cobyric acid synthesis (pmol/$OD_{600}$) non-induced | cobyric acid synthesis (pmol/ $OD_{600}$) induced (ITPG) |
|---|---|---|---|
| ER99 | pKK223.3 (no insert) | 0 | 0 |
| ER123A | pER119 (*cobA*$^{Pd}$) | 210 | 550 |
| AR8957 | pAR8414 (*cysG*) | 510 | 540 |

*E. coli cysG* mutant is the strain 302Δa. pAR8827 contains all *S. typhimurium cbi* genes except *cbiB*.

No cobyric acid was detected in the absence of a uro'gen III methylase (CobA or CysG), proving the essential role of the C-2, C-7 methylase activity. Cobyric acid was restored only in the presence of a second plasmid containing either *cobA* or *cysG*. Therefore, only the methylase activity of CysG seems necessary for corrin synthesis. This implies that, in *S. typhimurium*, the role of CysG in cobalamin biosynthesis is only to produce precorrin-2.

However, one additional observation with regard to strain ER123A was that the cysteine auxotrophy of the 302Δa strain had also been overcome. Indeed the strain was able to grow on minimal medium without cysteine. Since it had been previously established that the *cobA* gene on its own was unable to complement the *E. coli* 302Δa cysteine auxotrophy, this indicated that a gene

within the *cobI* operon was able to act in concert with the *cobA* gene to produce sirohaem. The gene responsible for the $CysG^B$ like activities within the *cob* operon will be discussed in section 2-6.

## 2-2-3- *E. coli* as a host for the *B. megaterium cobI* operon.

The *B. megaterium cobI* operon was isolated from a *B. megaterium* genomic library for its ability to complement a *cbiE* mutant strain. The clone containing the plasmid with the *cob* region was named AR8766 and the plasmid, pAR8766. This plasmid was fully sequenced.

The *cobI* region was restricted with the enzymes *Bgl*I and *Sal*I and cloned into pAR8668 (pKK223.3 with only one *Bam*HI site in the multicloning site). The new plasmid was named pAR8877 and contains *cbiW-*$H_{60}$*-X-J-C-D-ET-L-F-G-A-*$cysG^A$*-cbiY-btuR* under the control of a *tac* promoter [Raux *et al.* 1998].

The *cbiP* gene product encodes for the amidation of the side chain at position b, d, e and g and is necessary for the transformation of cobyrinic acid a,c-diamide into cobyric acid. This gene is absent in the *B. megaterium cobI* operon and therefore the *S. typhimurium cbiP* gene had to be included with the *B. megaterium cobI* genes in *E. coli* in order to allow cobyric acid synthesis, the earliest intermediate which can be detected by microbiological assay [Blanche *et al.* 1991; Roth *et al.* 1993; Raux *et al.* 1996].

As shown in Table 2-2, when the appropriate strain of *E. coli* (LE392 $met^+$ with the *S. typhimurium cbiP* gene on a plasmid) was transformed with either pAR8766 (a promoterless plasmid harbouring the *B. megaterium cobI* operon) or pAR8877 (the *cobI* operon under the control of a *tac* promoter), the two new strains were able to produce cobyric acid at 400 to 660 pmol/$OD_{600}$ in the presence or absence of IPTG. No major difference in corrin synthesis was observed between the promoterless and the *tac* promoter constructs. These results indicate that the *cob* promoter(s) are included in the 1.7 kb upstream of *cbiW* and that all the genes necessary for cobyrinic acid a,c-diamide biosynthesis are present within the *B. megaterium cob* operon.

INSTITUTE OF OPHTHALMOLOGY
LIBRARY
BATH STREET EC1V 9EL

Table 2-2- Comparison of cobyric acid synthesis from the *B. megaterium cob* operon under the control of its own promoter or a *tac* promoter.

| strain | plasmid carrying the *B. megaterium cobI* operon | cobyric acid production pmol/$OD_{600}$ | |
|---|---|---|---|
| | | non-induced | induced (IPTG) |
| ER241 | pAR8766 (promoterless) | 550 | 600 |
| AR8937 | pAR8877 (*tac* promoter) | 400 | 660 |

The *E. coli* host strain is AR3730 (LE392 rendered $met^+$) with a plasmid harbouring the *S. typhimurium cbiP*.

## 2-2-4- Production of cobyric acid in *E. coli* occurs only under anaerobic conditions.

It had previously been demonstrated that *S. typhimurium* is only able to produce cobyric acid under anaerobic conditions. Thus it was not surprising to find that after transferring the *S. typhimurium cob* operon to *E. coli*, corrin synthesis was also restricted to anaerobic conditions [Raux *et al.* 1996].
When *B. megaterium* was grown aerobically, vitamin $B_{12}$ biosynthesis was observed. However, when the *B. megaterium cobI* operon was transferred into *E. coli*, no cobyric acid could be detected when the strain (AR8937) was grown aerobically. As with the synthesis of cobalamin in *S. typhimurium*, the *B. megaterium cob* operon when cloned in *E. coli* produces cobyric acid only anaerobically. Two reasons can be forwarded to explain why a cobalamin biosynthetic pathway from an apparent obligate aerobe only functions under anaerobic conditions in a facultative organism. Firstly, when *E. coli* is grown aerobically the oxidative environment may prove to be detrimental to the highly labile early corrin intermediates. In *B. megaterium* the labile intermediates may be protected by other ancillary protein factors. Alternatively, *E. coli* may encode a protein, expressed only under strictly anaerobic conditions, which may be involved in the oxidation-reduction of the central cobalt ion, a process which is thought to mediate the ring contraction process.

## 2-3- Role of *cbiD* and *cbiG* in cobalamin synthesis.

*cbiD* and *cbiG* are both found in the *S. typhimurium* and *B. megaterium cob* operons but have no equivalent in *P. denitrificans*. It appears that these genes are specific to the anaerobic pathway. Indeed, the essential role of *cbiD* in

cobalamin synthesis in *S. typhimurium* had already been established, as a *S. typhimurium cob* mutant was complemented by a plasmid harbouring *cbiD* [Raux *et al.* 1996]. To confirm the absolute requirement of *cbiD* to make vitamin $B_{12}$, and to create a simpler approach to study the role of *cbiD*, a mutation was introduced in the *cbiD*s of both *S. typhimurium* and *B. megaterium* plasmid-borne *cob* operons. This allowed the role of *cbiD* within *E. coli* to be studied. None of the *S. typhimurium* mutants were complemented by *cbiG*. A mutation in both *cbiG* genes of the plasmid-borne *S. typhimurium* and *B. megaterium cob* operons was therefore introduced to allow the role of *cbiG* to be investigated.

## 2-3-1- Introduction of separate mutations in *cbiD* and in *cbiG* of the *cob* operons.

The *cbiD* and *cbiG* mutations were introduced in two different plasmids, pAR8814 (pKK223.3 harbouring the *S. typhimurium cobI* genes, except *cbiB*, under the control of three *tac* promoters) and pAR8877 (pKK223.3 harbouring the *B. megaterium cobI* genes under the control of a *tac* promoter). Both plasmids contain all the genes necessary for the synthesis of either cobyric acid (pAR8814) or cobyrinic a,c-diamide (pAR8877).

### a- In the *S. typhimurium cob* operon.

A mutation was introduced into *cbiD* (located between nucleotides 4266 to 5405 of the *S. typhimurium cob* operon) on pAR8814 at position 4608 after restriction of this plasmid by *MluI*, end-filling and religation. This produced pAR8977 which contains a frameshift in *cbiD* such that only the first 114 codons are translated (the whole CbiD contains 379 amino acids).

The mutation in *cbiG* (located between nucleotides 7310 and 8365) was introduced by digestion of pAR8814 with *BstBI* at positions 7774 and 8211 and religation. This resulted in a deletion of 437 bp. The mutated plasmid was named pAR8976. These two plasmids were constructed by F. Levillayer, CNRS, Gif-sur-Yvette.

### b- In the *B. megaterium cob* operon.

A mutation in *cbiD* was obtained as follows: pAR8877 was partially restricted with *Acc*I, blunt ends were generated and subsequently religated. This

procedure generated a 2bp insertion into *cbiD* (position 6129 to 7230) resulting in a frameshift and a new stop codon at position 6444. The mutated plasmid was named pAR8909 and encodes for a truncated protein of 105 amino acids as opposed to the 367 amino acids found in the wild type enzyme.

A similar procedure was used to introduce a mutation into the *Cla*I site of *cbiG*. The 2 bp insertion into *cbiG* (position 9892 to 11014) led to a stop codon at position 10411 which was harboured in plasmid pAR8921. A truncated protein of 173 amino acids was expressed instead of the 374 amino acids, which constitute the full CbiG.

## 2-3-2- *cbiD* and *cbiG* are essential genes for the synthesis of cobalamin.

*E. coli* LE392 $met^+$ was initially transformed with two plasmids, one carrying the *E. coli cysG* and the other carrying the *S. typhimurium cob* operon with a mutation either in *cbiD* or in *cbiG*.

The same host strain was also transformed with two other plasmids, one carrying *S. typhimurium cbiP* and one carrying the *B. megaterium cob* operon with a mutation either in *cbiD* or in *cbiG*.

The amount of cobyric acid produced in these strains was analysed using bioassay plates. Table 2-3 shows the levels of cobyric acid obtained from the different strains. To ensure that the mutations did not cause polarity effects on the downstream genes, a "wild type" gene copy was transferred into the "mutant strains" on a separate, compatible, plasmid.

The mutation of *cbiD* (pAR8977) in the *S. typhimurium cob* operon led to a frameshift within the gene and produced a 114 amino acids truncated protein. After cloning this construct into *E. coli* together with the plasmid containing *cysG**, strain AR8985 was obtained. This strain contains all the functional genes required for cobyric acid biosynthesis except for *cbiD*. By examining extracts of this strain by microbiological assay, it was concluded that AR8985 was unable to produce cobyric acid. This result also confirms *cbiD* as a *cobI* gene.

Table 2-3- Role of *cbiD* and *cbiG* on cobyric acid synthesis.

| Strain | *cob* operon origin | gene of the *cob* operon mutated or deleted | gene(s) on a second plasmid | pmol/$OD_{600}$ non-induced | pmol/$OD_{600}$ induced (IPTG) |
|---|---|---|---|---|---|
| AR8830 | *S. typ.* | (pAR8827) | *cysG* | 800 | 500 |
| AR8985 | *S. typ.* | *cbiD* (pAR8977) | *cysG* | nd | nd |
| AR8984 | *S. typ.* | *cbiG* (pAR8976) | *cysG* | nd | nd |
| AR8937 | *B. meg.* | (pAR8877) | *cbiP* | 400 | 660 |
| ER164 | *B. meg.* | *cbiD* (pAR8909) | *cbiP* | nd | nd |
| ER167 | *B. meg.* | *cbiD* (pAR8909) | *cbiP, cbiD* | 40 | 320 |
| ER165 | *B. meg.* | *cbiG* (pAR8921) | *cbiP* | 0.5 | 0.3 |
| ER169 | *B. meg.* | *cbiG* (pAR8921) | *cbiP, cbiG* | 290 | 680 |

The *E. coli* host strain is AR3730 (LE392 rendered *met*$^+$). nd means not detectable.

A deletion of 417 bases from *cbiG* in the *S. typhimurium cob* operon gives pAR8976. When this plasmid was transformed into *E. coli* (generating AR8984) corrin production was undetectable by analysis of the cell extract on the vitamin $B_{12}$ microbiological assay.

Mutations were introduced separately into *cbiD* and *cbiG*, within the large 16kb *B. megaterium cob* operon harboured within pAR8877.

The derivative of pAR8877 with a mutation in *cbiD* was called pAR8909 and when this was introduced into *E. coli*, to give strain ER164, no cobyric acid was detected. When this strain (ER164) was transformed by an additional plasmid carrying wild type *B. megaterium cbiD* (pER156), the resulting strain (ER167) produced 40 pmol/$OD_{600}$ of cobyric acid when non-induced whereas 320 pmol/OD were produced when the strain was induced with IPTG.

The derivative of pAR8877, which contained the mutated *cbiG*, was called pAR8921 and produced barely detectable levels of cobyric acid (<1 pmol/OD) when transformed into *E. coli* (strain ER165) under either inducing or non-

inducing conditions. However, when the wild type *B. megaterium cbiG* was reintroduced back into the strain on plasmid pER160, the levels of cobyric acid production were restored to normal levels.

These results demonstrate that the *B. megaterium cbiD* and *cbiG* are necessary for cobyric acid production and are essential, specific, genes of the anaerobic pathway.

## 2-3-3- CbiD possesses a glycine rich motif.

In the aerobic pathway, the enzyme responsible for the sixth methylation at C-1 of the macrocycle occurs after the ring contraction process and produces precorrin-6. This step is catalysed by CobF in *P. denitrificans*. In the anaerobic pathway, no equivalent to the *P. denitrificans cobF* is found. The methyl group bound to the C-1 carbon is known to originate from S-adenosyl methionine (SAM) and not from the C-20 position which is released, with its attached methyl group, as an acetaldehyde molecule during the ring contraction process [Wang *et al.* 1996]. Therefore, an enzyme is required to transfer the methyl group from SAM to the C-1 position. In theory, this could be achieved by one of the other methyltransferases involved in corrin synthesis as both CysG/CobA and CbiE/CobL perform two methylations [Warren *et al.* 1990; Blanche *et al.* 1992]. An alternative explanation could be that an enzyme, different to a typical methyltransferase and specific to the anaerobic pathway, is responsible for the methylation at C-1.

CbiD has been shown to be essential and specific to the anaerobic pathway. Moreover, an alignment of nine CbiD amino acid sequences from different micro-organisms reveals that this enzyme contains a glycine rich motif typical of a nucleotide binding site at position 100 to 120. Such a motif can be found in proteins that bind $NAD^+$, $NADP^+$, FAD or SAM. If this glycine rich motif does represent a SAM binding site, then it would seem extremely likely that the role of CbiD would be to act as the C-1 methyltransferase. Figure 2-4 shows the sequence alignment of nine CbiDs from *Methanococcus jannaschii, Methanobacterium thermoautotrophicum, S. typhimurium, Clostridium acetobutylicum, B. megaterium, Synechocystis sp., Porphyromonas gingivalis, Pseudomonas aeruginosa* and *Archaeoglobus fulgidus*.

Figure 2-4 - Alignment of nine CbiDs.

```
          201                                                        250
Mj CbiD   .......... .......... .......... .......... ..MIYDFRKK
Mth CbiD  .......... .......... .......... .......... ......MNDE
St CbiD   .......... .......... .......... ....MSELSF DAPVWHHGKA
Ca CbiD   .......... .......... .......... ........ML EMYVNCDGKK
Bm CbiD   .......... .......... .......... .......... MKEVPKEPKK
Pa CbiD   .......... .......... .......... .........M REETPEQPAP
Pg CbiJD  YFREKIEAAR RMGIRIYAVV RPPLPPSFIP VGGPVGLRRA VERLVPGFFS
Syn CbiD  .......... .......... .........M IIALFFVINQ *RENLSQSIS
Af CbiD   .......... ...MLIDPIE LYRYPEKWIK DRDAEKKVRS GLYILTEDGY
Consensus ---------- ---------- ---------- ---------- ----------

          251                                                        300
Mj CbiD   SKFGYTTGSC AAAGAYSALY YLKFGKKLSY VEIE.NLNGD KLIIPIEKIE
Mth CbiD  RVFGITTGTA ATAAALASLL CLK.GCEASK VRVR.TPSG. ILEVDVEYVE
St CbiD   LRKGYTTGSC ATAAAKVAAL MVLRQHLIHQ VSIV.TPSGV TLCLNVESP.
Ca CbiD   LRCGYTTGSC AAGAAKAATY MLYNEEILDL IKID.TPKGI ELMLPIENI.
Bm CbiD   LREGYTTGAC ATAATRAALL TLISGEVQDE STIY.LPVGR FATFHLEEC.
Pa CbiD   LRSGYTTGSC ATATSLAAAR LLLGGTISDA VQIV.LPKGQ QVLMRLEFC.
Pg CbiJD  LRSGFTTGTT ATAAVVAAMY RLMGLGSLAE APVE.LPSGE IVSLPIAEI.
Syn CbiD  AQSGYTLPVF ACASAIAAVE TLLTSNCPDS VTLELLEPAR TAEIAIEQGA
Af CbiD   LRRGITTGTT ASAAAVAAIA SLK..EKVEK VKVS.TPAGV DVEVEVE...
Consensus ---G-TTG-- A-A----A-- -L-------- --I-----G- ------E---

          301                                                        350
Mj CbiD   K.CGNKAKAV VIKDAGG.DI DITNGIEIIT EVELKKGKKD VIIKGGEGVG
Mth CbiD  RISGDTARAC VIKRPYP.DP DVTVNLEIIS TVQI.TGTPE IMIRGGEGVG
St CbiD   HIEGQQAIAA IRKDGGD.DV DATHGMLIFA RVTLNDSG.E ITLTGGEGIG
Ca CbiD   KKGDGFVECS IIKDGGD.DP DITNGIEIWA RAEVKKEG.. YTLKGGIGVG
Bm CbiD   EYRTSSAVAS IIKDAGD.DP DATHGALIIS EVSWCNG.VD IIIDGGVGVG
Pa CbiD   RAWENGAEAG TLKDAGD.DP DVTHGALVFA RVRLSAE.PG VRFHAGPGVG
Pg CbiJD  REEEDAVVSA VLKDAGD.DP DVTNGMAVCA TIRLNPEHEE VRFLQGEGVG
Syn CbiD  LLGHHRALAI TRSEPGN.NL DLTRHTPVWA EVEFTPGEGK LIIQGGEGIG
Af CbiD   ...AEKGFAR VRKFSGDHEF DVTNGIIFEA EV....CETS GIFF.GRGVG
Consensus ---------- --K--G--D- D-T----I-- -V-------- -----G-GVG

          351                                                        400
Mj CbiD   IVTKNGLQVK KGEPAINPKP REMIRNNLLK LLND...DEV VEVTISIPKG
Mth CbiD  IVTKPGLQVP PGEAAINPVP RRMIEENLRE HLRD...DEG AVVTVSVPGG
St CbiD   TVTRKGIGLP LGSAAINRTP RHTIESAVRE AIGP...ARG ADVEIFAPEG
Ca CbiD   IVKSEGLYVE KGDYAINPVP RLMIEKEVKK VLPK...DNG VLITVFVPKG
Bm CbiD   RVTKPGLPVP VGEAAINPVP RKMLKETAQQ LLAEYNIQKG VKVVISVPEG
Pa CbiD   TVTRPGLTLA VGEPAINPVP RQMIERHLAQ LAAERGYAGG FEVAIGVEGG
Pg CbiJD  VVTLPGLGLE VGGPAINLVP RRMMTAEVRR LYA....QGG VDITISVPEG
Syn CbiD  KQLDRE.... .GQAAIYSYA QRLLRHHLQP YISG...DQT LMVSLILPLG
Af CbiD   ........VK AGEKAVSRSA KLQILENFIK ASREFNFSGG ..VRISVPDG
Consensus -----G---- -G--AI---- ---------- ---------- --V-I--P-G

          401                                                        450
Mj CbiD   KELAKKTLNP KLGIVGGLSI LGTTGIVRPM SNEAYMNSLA PQIDVALANG
Mth CbiD  ERIASRTMNP RLGIIGGISI LGTTGIARPM STEAYRESLA CQIDIATARG
St CbiD   EARAQKTYNS RLGILGGISI IGTTGIVTPM SEESWKRSLS LELEIKRASG
Ca CbiD   EEIAKKTFNP RLNIVGGISI LGTTGIVVPM SEEALQQSIK LEMNQKIKSG
Bm CbiD   EEMAKKTLNA RLGILGGISI LGTRGIVVPF STAAYKASIV QAISVAKASN
Pa CbiD   AELALKTMNP RLGILGGLSI LGTSGIVRPF SCSAYIASIH QGIDVARANG
Pg CbiJD  REAATQTFNP RLGIRDGISI IGTSGVVKPF SAEAFVGAIR KQVGIATALG
Syn CbiD  RTLATRTSNA AFGVVEGLSL LGTSGIAQPL SAPEQLAEFQ QHLTSA.AQQ
Af CbiD   EEVAKKTGNE KVGIKGGISI LGTTGFVEPW CKK......L VETKLKIAMQ
Consensus ---A--T-N- -LGI--G-SI -GT-GI--PM S--------- -------A--
```

```
          451                                                        500
Mj CbiD   YKRLIFVPGN IGTKYAKQLL NANDDE.IIE VSNFWGFMLD KAKE.KGVEE
Mth CbiD  FRELVFVPGN IGERFAREYF ESIEDERIIQ MANFPGYMLE EADG.RGVER
St CbiD   LTRVILVPGN HGERFVREQM .GVDTQAVVT MSNFVGYMIE EAVR.LGFCQ
Ca CbiD   IKNFTFLFGN MGEDKAKE.M .GIDPKGFVI MSNYVGFALN CCRE.NNIKK
Bm CbiD   CEHVVITTGG RSEKYGMKQF PELPEEAFIQ MGDFVGFTLK QCKK.QGMKK
Pa CbiD   VRHIAACTGN ASEDAMRRRY A.LPEIALIE MGDFAGAVLK HLRR.APVEK
Pg CbiJD  ANHIVLNSGA KSERYVKGAY PALIPQAFVQ YGNFVGESLS CVASFPSVRS
Syn CbiD  HQCLVFCLGE NGLDLARQW. .GVPLDQMVK TANWLGSLFV AAAAV.GVQE
Af CbiD   YHRIAITTGR KAWLYARKKF PEYQPFVF.. .....GVHID EALKHPGEKI
Consensus --------G- ---------- ---------- ---F-G---- ----------

          501                                                        550
Mj CbiD   ILIFGHAGKI IKLAGGIYNT HSKVADCRNE ILAAYSSLFI DDKEAIKKIL
Mth CbiD  IILMGHAGKL IKLSAGIFQT RSSDADARRE IMAAHAALMG ASRDTVTRIF
St CbiD   IVLVGHPGKL IKIAAGIFHT HSHIADARME TLVAHLALLG APLELLTLVS
Ca CbiD   ILMVGHIGKM CKIAAGCFNT HSRVCGVRLE ILALELALMG ADTNFIEKIY
Bm CbiD   VSLVGMMGKF SKVAQGVMMV HSKSAPIDFN FLAKAASESG ASAELVEEIK
Pa CbiD   LSLCGGFGKI SKLAGGHLDL HSRHSSIDLP QLAGWAAALG ASTALQDSMR
Pg CbiJD  VTVGIMLGKA VKLAEGYLDT HSKKVVMNRD FLHELARQAG CSEDIHAIID
Syn CbiD  ILLLGYHGKL IKLAGGIFHT HHHLADGRLE ILTAQAVQAG LPYPLVQELG
Af CbiD   IVGFPGLLKI WAGSRDRIEE RAREEGVRVV VIED...... ...DMDSWVW
Consensus I------GK- -K---G---- ---------- -L-------- ----------

          551                                                        600
Mj CbiD   YSNTTEEVIK IL......EE KGVLNDVFNF IAKRVVERLS ERWKGIKF..
Mth CbiD  NDGTVEEMIS EL......EG EGLTGRVFNS IADAIRERFN ERF.SFDT..
St CbiD   DCDTTEAAME HI......EA YGF.GHIYNH LARRICLRVM QMLR.FTKTP
Ca CbiD   REKTTEGAVK II........ ....GKEYKD IYRRIGVKIM KKIKDFTYGE
Bm CbiD   GANTASQVGD LM......TQ SGH.....HQ FFEKLCEYCC LSALKEVGDG
Pa CbiD   AANT.SQQAL AQ......AH AEG.....VA LGDAVCAHAL RFARGIVPTE
Pg CbiJD  SLNLARELWT MP......SA EDS.....DR LLRKIAERSW ETCRPSVPSA
Syn CbiD  QAPTTEAGLK LLRHWQTEQN CPWVSKIYQA MADTVDRRSE EYVYKVSQHQ
Af CbiD   DVQGTDH... .......... .......... .......... ..........
Consensus ---------- ---------- ---------- ---------- ----------

          601                                  636
Mj CbiD   ...SCIIINM KGEVLGSYIW NKK....... ......
Mth CbiD  ...DVLIFSL DGRILNSNFK TKRNHRFTEV F.....
St CbiD   PVCDAILFSF DNHILGSNRP VDEIAKELQC ......
Ca CbiD   VNADIVLYSM EKGILWDSRE VI........ ......
Bm CbiD   IDVDTSLYTL KGDFLGQAVQ HGN....... ......
Pa CbiD   VALEVFAIDR QGNLVGQACE ERR....... ......
Pg CbiJD  .ELELLLIDE SGAIRFRIGG E......... ......
Syn CbiD  LKVGSLLFDG DRQPVAISTQ GQAMADKLGM TIPET
Af CbiD   .......... .......... .......... ......
Consensus ---------- ---------- ---------- ------
```

The *Synechocystis sp.* CbiD sequence from Genebank starts at the first valine after the glycine, tyrosine and threonine (GYT) conserved residues, probably due a sequencing mistake. This explains why a stop codon is present in the N-terminus of the protein. St: *S. typhimurium;* Bm: *B. megaterium;* Syn. : *Synechocystis sp.;* Ca: *C. acetobutylicum;* Pg: *P. gingivalis;* Mj: *M. jannaschii;* Mth: *M. thermoautotrophicum;* Af: *A. fulgidus* and Pa: *P. aeruginosa.*

Figure 2-5 - Consensus obtained with different SAM binding sites.

| | 21 40 | 128 145 | 161 170 |
|---|---|---|---|
| Rc/CobF | LIGIGTGNPR HITGQAVDAM | RVA LLVWGDPSLY DSTLR | LTTKVIPGIT |
| Pd/CobF | IIGIGSGNPE HMTVQAINAL | TGA FLVWGDPMLY DSTIR | FAYDVIPGIT |
| Pd/CobI | GVGTGPGDPE LLTVKAVKAL | TVA VLSEGDPLFY GSYMH | FPVEVIPGIT |
| Mtub/CobI | GVGLGPGDPE LVTVKAARVI | NVA LLAEGDPLFY SSYMH | FNAVIVPGVT |
| Rc/CobI | CAGLGPGDPD LISVRSDRAI | DVV VLCEGDPFLY GSFMH | AEVEVIPGIP |
| Bm/CbiL | GLGVGPGDPE LITVKAFRKL | DVA FVTEGDPMLY STFIH | APIQVIPGIS |
| Syn/CbiL | GVSVGPGDPE LITLKGLKIL | DVA FACEGDISFY STFTY | LGINYIAGVS |
| Mth/CbiL | GVGVGPGDSE LLTLRAVNVL | DVA FITLGDPSIY STFSY | FKTEMVPGVT |
| Ca/CbiL | GIGVGPGDSE LLTIKAVKAI | NVG FLTIGDPFVY STYIY | YQTETIPGIT |
| Mj/CbiL | GVGVGVGDKK LLTLKALEVL | EVA IITIGDPTLY STFSY | VEVEIVNGIS |
| Sal/CbiL | ALSTGPGAPD LITVRAARIL | QVG FITLGDAMLF STWIF | W.LEIVPGVT |
| Af/CbiL | GVGIGLTKKH L.TLEAIDVI | TVA FCCIGDPVFY STFNH | LTVKVIPGIS |
| Rc/CobM | FIGAGPGAAD LITIRGRDLI | DVA RLHSGDLSIW SAMGE | IPYDVTPGVP |
| Pd/CobM | FIGAGPGAAD LITVRGRDLI | DVA RLHSGDLSVW SAVAE | IAYTMTPGVP |
| Mtub/CobM | FIGAGPGAAD LITVRGQRLL | DVA RLHSGDPSLY SALAE | IGYEIVPGVP |
| Mj/CbiF | IVGAGPGDPE LITIKGKKAI | KVV RLHTGDPSIY GAIKE | IDVEIIPGVS |
| Mth/CbiF | .MGAGPGDPE LITLKAVRAL | TVA RVHTGDPSIY GAIKE | IPYTIIPGVS |
| Sal/CbiF | FVGAGPGDRE LITLKGYRLL | TVV RLQTGDVSLY GSVRE | IDWQVVPGVS |
| Ca/CbiF | FIGAGPGDVD LITVKGRKIV | ITV RLHTGDPSIY GAIKE | IAYEVVPGVS |
| Bm/CbiF | IIGAGPGDPD LITVKGLKLL | MVV RVHTGDPAMY GAIME | VDIEIVPGVT |
| Af/CbiF | FVGFGPGNPE LLTLKAYKLL | RVV RVQSGDPSIF GAVWE | IDCEVVPGVS |
| Syn/CbiF | FVGAGPGDPD LLTIKGQKLL | IVV RLHSGDLTLY SAIHE | IPFVCVPGIS |
| Mj/CysGA | LVGAGPGDPE LITIKGLKAI | LVV RLKGGDPFVF GRGGE | IPYEVIPGIT |
| Mth/CysGA | LVGAGPGDPE LITLKAIRVL | TVV RLKGGDPFVF GRGGE | IPYRVIPGVT |
| Bm/CysGA | LVGAGPGDPK LITVYGMECI | VVT RLKGGDPFVF GRGAE | IPYEVVPGIT |
| Af/CysGA2 | IVGAGPGKKD LLTLRAYELI | TVV RLKGGDPCVF GRGGE | IPFEFVPGVT |
| Syn/CysGA | LVGAGPGDPG LITVKGKTLL | VVV RLKGGDPFIF GRGGE | IEVEVVPGIT |
| Sal/CysG | LVGAGPGDAG LLTLKGLQQI | RVV RLKGGDPFIF GRGGE | IPFSVVPGIT |
| Pd/CobA | LVGAGPGDPG LLTLHAANAL | RVL RLKGGDPFVF GRGGE | VPFRIVPGIT |
| Mtub/CysG2 | LVGGGPGDPE LITVRGRRLL | FVV RLKGGDPFVF ARGYE | IPVTVVPGVT |
| Ca/CysGAD | IIGAGPGEEE LITLKAIRKM | IVG RIKGGDPYIF GRGGE | IEFEVISGVT |
| Af/CysGA | IVGAGPGNFE LLTLKAYRLL | KIV RLKGGDPGIF GRMAE | IPFEVVPGVS |
| Rc/CobA | FVSSGPGDPE LLTLKAADRL | RVV RLKSGDSGLF GRLEE | IGYEIIPGVP |
| Mtub/CysG | FVGSGPGDPG LLTTRAAAVL | DVV RLVAGDPLTV DAVIS | LHIEIVPGLA |
| Rc/CobJ | VAGLGPGRED LVTPEVTAAL | RVV VVSSGDPGVF AMASA | TEIRILPGIT |
| Pd/CobJ | VVGTGPGSAK QMTPETAEAV | KVC MVSGGDPGVF AMAAA | VELVITPGVT |
| Mtub/CobJ | VVGLGPGDSD WMTPQSRREL | AVA VVSSGDPGVF AMATA | VRVRVIPAMT |
| Sal/CbiH | VIGIGPGSQA MMTMEAIEAL | NVA LISSGDAGIY GMAGL | VEVRLIPGMT |
| Ca/CbiH | VVGIGPGGLQ DMTLKARRAL | VVS IISTGDSGIY GMAGL | EIVEIIPGVT |
| Bm/CbiH60 | VIGFGPGSFE HITQRAREAI | TVA VISSGDAGVY GMAGL | VELEVIPGIS |
| Syn/CbiGH | LVGTGPGELS QISPAARTAI | TVA VLSSGDCGIY GMAGL | .AVEVFPGIS |
| Mj/CbiH | VVGIGSGNER HFTKEAEEIL | DVA LVSSGDATIY GLASL | VDIKVVPGIT |
| Af/CbiHC | VVGIGPGKEE LMTLKAKRAI | VVA LISGGDPSIY GILPL | VEIEAIPGVT |
| Mth/CbiH | IIGIGPAR.D DITIRALRAL | DVA LVSSGDPGVY GMANV | IEFEVIPGVT |
| Mj/CbiE | IVGIGPGDRE YLTLKAIKIV | KIA ILSTGDPCFS GLLKT | .DIEAISGIS |
| Mth/CbiE | IVGIGPGSPD YITPAAKNVV | EVS VLSTGDPGFS GVLKP | VRVKVVPGIS |
| Af/CbiE | IVGSG.TCRG QTTERAKEII | EVA VISTGDPMVA GLGRV | VEIKIEPAIS |
| Bm/CbiET | IIGIGDDGKL SLLPMYEQWI | NAV VLASGDPLFY GIGSY | .DVEIYPYLS |
| Syn/CbiET | VVGIGLNGAE GLTSSTLELI | TIV VLASGDPLYF GLGRL | .QLKFHPHLS |
| Sal/CbiE | VVGMGPAGRH LMTPAALEAI | GIV VLASGDPLFY GIGTR | .QVRIIPGIS |
| Rc/CobL | ILGLGEDGLD GLSLASRRAL | ATV VLASGDPFWF GAGSM | .EWQAFPVPG |
| Pd/CobL | VIGIGEDGVA GLGDEAKRLI | PVV VLASGDPFFF GVGVT | .EIRTLPAPS |
| Mtub/CobL | VVGIGADGMT GLSEHSRSEL | DLH VVASGDPLLH GIGST | .NVTVLPHVS |
| **Consensus** | -VG-GPG--- --T------- | -V- ----GDP--Y ----- | ----VIPGV- |

In order to compare the presumed CbiD GxGxG motif with the motif found in the biosynthetic methyltransferases, fifty-three different methylases involved in cobalamin biosynthesis from ten different bacteria were aligned. In the case of fusion proteins, only the methyltransferase domains were compared. The conserved-SAM binding site regions between all methyltransferases are displayed in Figure 2-5.

a- Comparison of the nucleotide binding sites.

The SAM binding site motif of all the precorrin biosynthetic methyltransferases is located in the first sixty amino acids of the protein (except for CysG, which is a fusion protein). The presumed CbiD SAM binding site is located about 100 amino acids downstream from the initial methionine, which suggests some divergence even if CbiD is a methyltransferase. The consensus domains of several glycine rich nucleotide binding domains are shown is Table 2-4.

Table 2-4- Consensus of different nucleotide binding sites.

| nucleotide binding site consensus | motif | reference |
|---|---|---|
| in CbiD proteins | GG-GVG-----G | This work |
| with SAM | VG-GPG-----T<br>I xG | This work |
| with $NAD^+$ | -G-G--G---G- | Scrutton *et al*, 1990 |
| with $NADP^+$ | -G-G--A---A- | Scrutton *et al*, 1990 |
| with FAD | -G-G--G---A- | Warren *et al*, 1994 |

By comparison with the different nucleotide binding site conserved residues, it is very difficult to decide whether the CbiD glycine rich motif is a SAM binding site, even if the CbiD motif (GGxGxG) appears more like a SAM binding site ((V/I)GxGxG) than an $NAD^+$ binding motif (GxGxxG). The second and the third conserved region of the methyltransferases are VxxxxxGDPxxY/F and VIPGV. None of these conserved residues were found within CbiD proteins.

A rapid SAM binding assay was recently developed to determine the importance of the different residues among the glycine rich motif in the CysG protein [Woodcock and Warren, 1996]. The simplest way to determine if the CbiD motif is a real SAM binding site is to analyse the ability of the protein to

bind SAM using the assay. The experiment was undertaken with the *B. megaterium* CbiD.

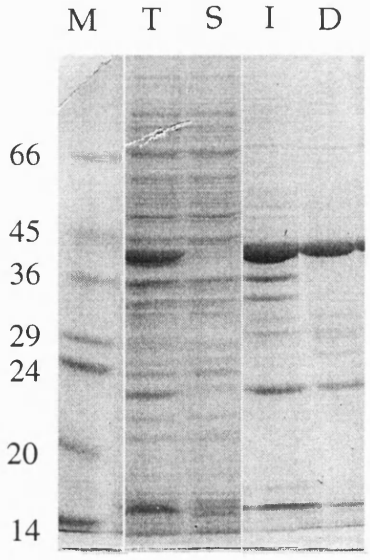
b- Cloning of the *B. megaterium cbiD*.

The *B. megaterium cbiD* was cloned after amplification of the gene by PCR. The primers used for this amplification, BmD-RBS and BmD-stop, are described in Materials and Methods, section 8-2-12. The ribosome binding site was changed to improve translation. The modified *cbiD*, with its improved ribosome binding site was cloned into the *Eco*RI-*Bam*HI sites of pAR8668 (pKK223.3 derived). After transformation into JM101, the recombinant strain was named ER222.

c- Purification of CbiD.

CbiD was overexpressed from strain ER222 and was found to produce a major protein of about 42 kD (Figure 2-6). Because CbiD was only located among the insoluble proteins, no trace of it was found in the soluble fraction, the protein was resolubilised after guanidine denaturation as described in Materials and Methods, section 8-5-2/e. CbiD was finally dialysed against Tris/HCl buffer, pH 7.8. The refolded protein remained in solution and was tested for its ability to bind SAM.

Figure 2-6- The different steps of CbiD purification.

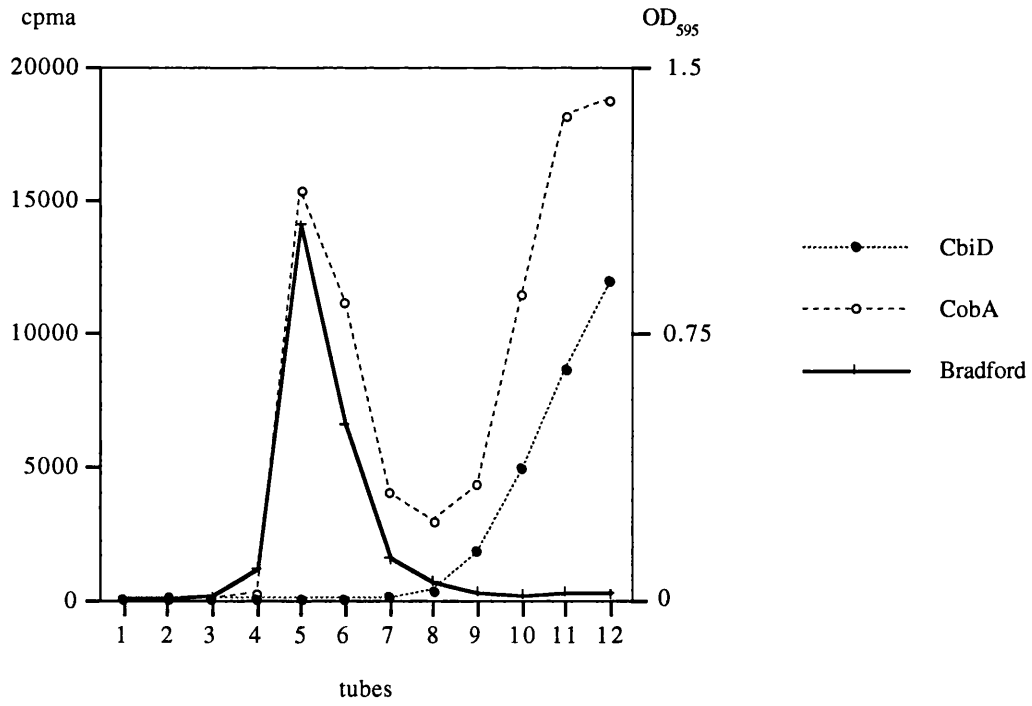



Lane "M" contains the molecular mass marker as indicated, in kD. Lane "T" corresponds to the total extact of the cells, lane "S" to the soluble protein fraction, lane "I" to the insoluble protein fraction, and finally lane "D" to the fraction after denaturation and resolubilisation.

d- SAM binding assay.

The assay was performed as described in Materials and Methods [section 8-5-4]. The gel filtration column was equilibrated with 50 mM Tris/HCl, pH7.8. Three separate SAM binding assays were undertaken; (*i*) with resolubilised CbiD alone (0.24 mg), (*ii*) with resolubilised CbiD incubated with [$^3$H-methyl]-SAM and finally (*iii*) with $CobA^{Pd}$ (0.24 mg) incubated with [$^3$H-methyl]-SAM. The last assay was used as a positive control. The purification of CobA is described in Chapter 3. A gel filtration column was used to separate [$^3$H-methyl]-SAM bound to the protein from unbound [$^3$H-methyl]-SAM. Figure 2-7 shows the results of these experiments.

Figure 2-7- [$^3$H-methyl]-SAM binding assay results after gel filtration.




The black plain line represents the presence of protein in each fraction; the double dotted line represents the [$^3$H-methyl]-SAM measurement after incubation with CobA and finally the single dotted line also represents the [$^3$H-methyl]-SAM measurement after incubation with CbiD.

Fractions (1 ml) were collected from the gel filtration column and were analysed for protein (Bradford reaction) and radioactivity. Protein was found to elute into fractions 5 to 7 (Figure 2-7). Two peaks of radioactivity can be observed with the CobA assay. The first peak of radioactivity, from tubes 5 to 7, is due to [$^3$H-methyl]-SAM tightly bound to the CobA protein and the second peak, from tubes 9 to 12, is due to the excess of free [$^3$H-methyl]-SAM.

No peak of radioactivity was observed with the CbiD fractions from tubes 5 to 7. The *B. megaterium* CbiD does not seem to bind SAM.

The alignment of CbiD sequences had revealed a conserved glycine rich motif, similar though not identical to the SAM binding site of the precorrin biosynthetic methyltransferases. The lack of SAM binding to CbiD does not necessarily mean that the CbiD glycine rich motif is not a SAM binding site. Indeed, CbiD was found insoluble and was refolded after denaturation by guanidine. There is no way of showing whether CbiD had refolded correctly. Moreover, some methylases implicated in cobalamin synthesis do not show a tight affinity for SAM, for example the *B. megaterium cbiF* methylase [Raux *et al.* 1998, Chapter 5, section 5-2-4].

## 2-4- Role of *S. typhimurium cbiK* in cobalamin synthesis.

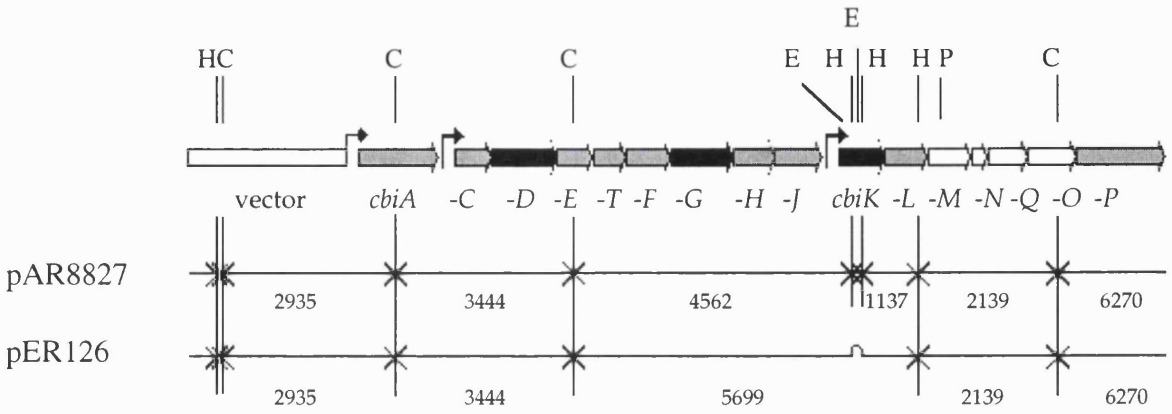
The *cbiK* gene is only found in the *S. typhimurium cob* operon. No equivalent was found in the *P. denitrificans cob* loci or in the *B. megaterium cob* operon. It was therefore important to investigate if *cbiK* was an essential gene in cobalamin biosynthesis and, if so, to elucidate its role.

### 2-4-1- Introduction of a deletion in *cbiK* of pAR8827.

Plasmid pAR8827 (carrying all the *S. typhimurium cobI* genes necessary for the production of cobyric acid from precorrin-2) contains four *Hind*III restriction sites, two of which are in the *cbiK* gene at positions 10028 and 10082. The DNA fragment between the two *Hind*III sites in *cbiK* site was deleted (Figure 2-8). The cloning details are outlined in Materials and Methods, section 8-3-3.

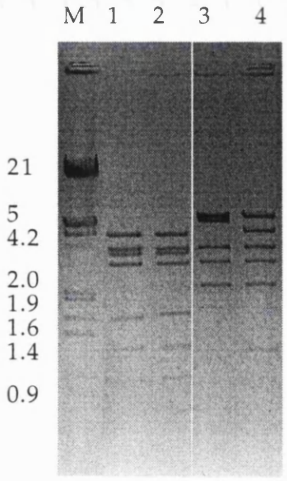
All the clones obtained following this procedure were analysed by restriction mapping to isolate plasmids of which the appropriate *Hind*III fragment had been removed. Two double digests were performed to compare the profiles of the parent plasmid pAR8827 with the mutated plasmid pER126; *Cla*I-*Hind*III to map the *Hind*III sites and *Eco*RI-*Pst*I (Figure 2-9). The plasmid containing the correct deletion was named pER126.

Figure 2-8- Comparison of pER126 and pAR8827 restriction maps.




"C" indicates a *Cla*I restriction site; "H", a *Hind*III site; "E" indicates the *Eco*RI site in 3' to the deletion and "P", the *Pst*I site in 5' of the deletion. The numbers underneath the figure correspond to the size of DNA fragment obtained after *Cla*I and *Hind*III double digest.

Figure 2-9- Profiles of pER126 and pAR8827 restriction digests.




After *Cla*I and *Hind*III digests, the difference observed between lane 4, the parent (pAR8827) and lane 3, the deleted-plasmid (pER126), is equivalent to that expected for the removal of the two *cbiK* *Hind*III sites. The *Cla*I-*Hind*III profile of pAR8827 exhibits three DNA fragments, *Cla*I-[4562 bp]-*Hind*III-[54 bp]-*Hind*III-[1137 bp]-*Hind*III. After the *cbiK* deletion, the resultant plasmid displays only one fragment,*Cla*I-[5699 bp]-*Hind*III. After *Eco*RI and *Pst*I digests, the 1784 bp fragment observed with the pER126 digest (lane 1) is 54 bp shorter than with the pAR8827 digest (lane 2) due to the deletion in *cbiK*.

## 2-4-2- The deletion in *cbiK* of pAR8827 has no effect on cobalamin synthesis in the presence of a genomic *cysG*.

The effect of the deletion of *cbiK* on cobyric acid biosynthesis in *E. coli* was investigated by transforming pER126 (plasmid containing all the *cobI* genes except *cbiB* and with a deletion in *cbiK*) into *E. coli* LE392 $met^+$. Moreover, in order to prove that the effect observed was due to the deletion in *cbiK*, *E. coli* was cotransformed with pER126 and a second plasmid harbouring a wild type copy of the *cbiK* gene. Similarly, *E. coli* was also transformed with pAR8827, the same plasmid as pER126 but with an unmodified *cbiK*. The three strains were

grown in minimum medium under anaerobic conditions and tested for cobyric acid production. The results are shown in Table 2-5.

Table 2-5- Role of *cbiK* in cobyric acid synthesis (in LE392 met$^+$).

| Strain | plasmid carrying the *S. typhimurium cob* operon | gene in the second plasmid | cobyric acid production pmol/$OD_{600}$ | |
|---|---|---|---|---|
| | | | non-induced | induced |
| AR8859 | pAR8827 | | 200 | 400 |
| ER126K$^\Delta$ | pER126 = pAR8827-[*cbiK*$^\Delta$] | | 590 | 500 |
| ER126K$^\Delta$K | pER126 = pAR8827-[*cbiK*$^\Delta$] | *cbiK* | 650 | 475 |

The *E. coli* host strain is AR3730 (LE392 rendered met$^+$). pAR8827 contains all the *S. typhimurium cobI* genes to produce cobyric acid from precorrin-2. pER126 is derived from pAR8827, such that a deletion in *cbiK* has been introduced.

Strain ER126K$^\Delta$ was found to produce similar levels of cobyric acid (around 500 pmol/OD) in the absence or presence of IPTG. This is akin to the amount of cobyric acid observed in AR8959, which produces between 200 and 400 pmol/OD. Transformation of ER126K$^\Delta$ with a second compatible plasmid harbouring the *cbiK* gene (pAR8580) gave strain ER126K$^\Delta$K which produces comparable levels of cobyric acid when grown in the absence or presence of IPTG (650 and 475 pmol/OD respectively).

The conclusion from these studies was that a deletion in *cbiK* had no detrimental effect on cobalamin synthesis in *E. coli*. Thus, on this basis, *cbiK* does not appear to be a *cob* gene, as it is not essential for cobyric acid synthesis in *E. coli* LE met$^+$. However, and as it will be described in section 2-6, the situation is not as simple as this.

## 2-5- Roles of *cbiW*, *cbiX*, *cbiY* and the 3' extension of $cbiH_{60}$ in cobalamin synthesis.

Four genes within the *B. megaterium cob* operon, *cbiW*, *cbiX*, *cbiY* and the 3' extension of the $cbiH_{60}$, have no homologue in any of the *P. denitrificans cob* loci nor in the *S. typhimurium cob* operon. Their roles in cobyric acid synthesis were investigated by creating (i) a deletion of 5' region of the operon including *cbiW*, (ii) a deletion of the *cysG*$^A$*-cbiY-btuR* region, (iii) a mutation in the *cbiX* gene and

at the start of the 3' extension of the *cbiH*$_{60}$ gene. The new mutated constructs were analysed for their abilities to promote cobyric acid synthesis within *E. coli*.

## 2-5-1- Effect of mutations or deletion introduced into *cbiW, cbiX, cbiY* and the 3' extension of *cbiH*$_{60}$ from the *B. megaterium cob* operon.

### a- Mutations introduced into *cbiX* and *cbiH*$_{60}$.

Both *cbiH*$_{60}$ and *cbiX* contain an *Acc*I restriction site, at position 3241 for *cbiH*$_{60}$ and at position 4088 for *cbiX*. Mutations in these genes were introduced as follows: pAR8877 (containing the whole *B. megaterium cob* operon) was partially restricted with *Acc*I and blunt ends were generated by the action of the Klenow fragment. The linearised plasmid was isolated and purified by agarose gel electrophoresis and was subsequently religated by the action of T4 DNA ligase.

When *Acc*I 3241 was filled in, a 2 bp insertion into *cbiH*$_{60}$ (position 2161 to 3781) was generated, creating a frameshift and a new stop codon at position 3295. The resulting plasmid was named pAR8916.

When *Acc*I 4088 was filled in, an extra 2 bp was introduced in *cbiX* (position 3787 to 4705) leading to a stop codon at position 4135. This mutation in *cbiX* generated pAR8911.

### b- Deletion of *cysG*$^{A}$*-cbiY-btuR*.

The *Eco*RI fragment from pAR8877, harbouring *cbiW, -H*$_{60}$*, -X, -J, -C, D, -ET, -L, -F, -G* and *-A*, was isolated and cloned into the *Eco*RI site of pKK223.3. The resultant plasmid, pAR8980, therefore contains the whole *B. megaterium cob* operon except for *cysG*$^{A}$*, cbiY* & *btuR*.

### c- Deletion of the 5' region of the *B. megaterium cob* operon and *cbiW*.

The deletion was achieved by restricting pAR8811 (pKK223.3 carrying the *Mlu*I fragment from pAR8766) with *Nco*I and *Sal*I and cloning the pAR8766 *Nco*I-*Sal*I fragment into it. The resulting plasmid, pAR8846, contains all the *B. megaterium cob* operon genes except for the first gene (*cbiW*) and the region including the putative *cob* promoter(s).

## 2-5-2- *cbiX* is necessary for cobyric acid production.

To investigate the role of *cbiX* in cobalamin biosynthesis, the derivative of pAR8877 harbouring the mutated *cbiX* (pAR8911) was transformed into *E. coli* giving strain AR8925. The level of cobyric acid produced was compared to AR8937, which contains pAR8877, and to AR8967, which contains the *cbiX* mutated plasmid (pAR8911) and a second plasmid harbouring a wild type copy of *cbiX*. The results are shown in Table 2-6.

Table 2-6- Role of *cbiX* in cobyric acid synthesis.

| Strain | plasmid carrying the *B. megaterium cob* operon | genes in the second plasmid | cobyric acid production pmol/$OD_{600}$ | |
|---|---|---|---|---|
| | | | non-induced | induced (IPTG) |
| AR8937 | pAR8877 | $cbiP^{St}$ | 400 | 660 |
| AR8925 | pAR8911 = pAR8877-[$cbiX^{M}$] | $cbiP^{St}$ | 0.1 | 2 |
| AR8967 | pAR8911 = pAR8877-[$cbiX^{M}$] | $cbiP^{St}$, *cbiX* | 110 | 440 |

The *E. coli* host strain is AR3730 (LE392 rendered $met^{+}$). pAR8877 contains all the *B. megaterium cobI* genes. With the *S. typhimurium cbiP*, pAR8877 is able to promote cobyric acid synthesis in *E. coli*. pAR8911 is derived from pAR8877 and contains a mutation within *cbiX*.

The strain harbouring the *cbiX* mutation within the *B. megaterium cob* operon produced 2 pmol/OD of cobyric acid, about 300 fold less than the control containing no mutation. When a copy of the wild type *cbiX*, on pAR8962, was reintroduced back into the strain (AR8967) the level of cobyric acid production under inducing conditions was 440 pmol/OD, some 200 fold higher than without *cbiX*.

These results established *cbiX* as an essential gene for cobalamin biosynthesis within the *B. megaterium cobI* operon.

The C-terminus of CbiX contains a polyhistidine tail (19 histidine residues on the last 36 amino acids) which suggest that the protein may be involved in metal chelation. Polyhistidine regions are known to be involved in binding nickel [Rey *et al.* 1994; Fu *et al.* 1995]. It was also shown that the *B. megaterium cob* operon was able to complement strain 302Δa [Beck *et al.* 1997] which is an *E. coli cysG* mutant strain [Griffiths and Cole, 1987]. As the *B. megaterium cob* operon encodes for $CysG^{A}$, the uro'gen III methylase, it was presumed that one

of the other genes was encoding for $CysG^B$-like activity. The gene responsible for the complementation was found to be *cbiX* and was described as a ferrochelatase involved in the synthesis of sirohaem [Beck *et al.* 1997].

In fact, knowing that CbiX is directly involved in cobalamin synthesis, its principal role is more likely to be that of a cobaltchelatase. The 2 pmol/OD production of cobyric acid observed when the *cob* operon contains a mutation within *cbiX*, could be due to the presence of a genomic *cysG*. The investigation about the synthesis of cobyric acid with the *B. megaterium cob* operon in the *E. coli* 302Δa (*cysG* minus background) will be described in Chapter 3.

## 2-5-3- *cbiW*, *cbiY* and the 3′ extension of $cbiH_{60}$ are not required for cobalamin biosynthesis in *E. coli*.

The various constructs described in section 2-5-1 in which mutations or deletions had been introduced into plasmid pAR8877 (carrying the *B. megaterium cob* operon) were analysed for their abilities to produce cobyric acid in *E. coli*. The results are shown in Table 2-7.

Table 2-7- Role of *cbiW*, *cbiY* and the 3′ extension of $cbiH_{60}$ in cobyric acid synthesis.

| Strain | plasmid carrying the *B. megaterium cob* operon | genes in the second plasmid | cobyric acid production (pmol/OD) | |
|---|---|---|---|---|
| | | | non-induced | induced (IPTG) |
| AR8937 | pAR8877 | $cbiP^{St}$ | 400 | 660 |
| AR8930 | pAR8916 = pAR8877-[$cbiH_{60}^{M}$] | $cbiP^{St}$ | 220 | 400 |
| ER144 | pAR8846 = pAR8877-[5′->*cbiW*]$^{\Delta}$ | $cbiP^{St}$ | 45 | 500 |
| ER176 | pAR8877 | $cbiP^{St}$, $cobA^{Pd}$ | 300 | 460 |
| ER177 | pAR8980 = pAR8877-[$cysG^{A}$-*cbiY*-*btuR*]$^{\Delta}$ | $cbiP^{St}$, $cobA^{Pd}$ | 0.4 | 390 |

The *E. coli* host strain is AR3730 (LE392 rendered met$^+$). pAR8877 contains all the *B. megaterium cobI* genes. With the *S. typhimurium cbiP*, pAR8877 is able to promote cobyric acid synthesis in *E. coli*. $cobA^{Pd}$ is the *P. denitrificans cobA* used to replace the deleted $cysG^{A}$.

The removal of the first 1914 bp of the *B. megaterium* 16 kb DNA fragment within pAR8877 gave rise to pAR8846. The deleted region spanned the whole

region situated upstream of the first gene of the *cob* operon, *cbiW*, as well as the first 50 codons of the 127 codons *cbiW* itself. When pAR8846 was transformed into *E. coli* together with *cbiP* it gave strain ER144, but the deletion of the 1914 bp had little effect on cobyric acid synthesis as the strain produces 500 pmol/OD when grown anaerobically in the presence of IPTG. The low level of cobyric acid observed when ER144 was grown in the absence of IPTG (45 pmol/OD) is probably a consequence of the deletion of the *cob* operon promoter.

A deletion of the distal 3′ end of the *cobI* operon, which includes *cysG*$^{A}$-*cbiY*-*btuR* resulted in the construction of pAR8980. When this plasmid was cotransformed into *E. coli* with a compatible plasmid harbouring the *P. denitrificans cobA* (the equivalent of the *cysG*$^{A}$ portion of the *E. coli cysG* gene) strain ER177 was generated. No major difference in cobyric acid production was observed between this strain and the strain containing the complete *B. megaterium cobI* operon and the *P. denitrificans cobA* (ER176) as both strains produce around 400pmol/OD of cobyric acid when grown anaerobically in the presence of IPTG.

When *cbiH*$_{60}$ was truncated within the *cobI* operon such that the mutated *cbiH*$_{60}$ encoded a truncated protein missing the last 162 amino acids (pAR8916), the mutation did not significantly reduce cobyric acid synthesis as the strain harbouring the mutant *cbiH*$_{60}$ (AR8930) still produces 400pmol/OD when grown anaerobically in the presence of IPTG.

The conclusion from these studies is that *cbiW*, *cbiY* and the 3′ extension of *cbiH*$_{60}$ do not seem to be required for cobalamin biosynthesis in *E. coli*.

## 2-5-4- Sequence comparisons of CbiW, CbiY and CbiH$_{60}$ with the database.

Whether or not these proteins are involved, directly or indirectly, in the biosynthesis of cobalamin, they are all part of the *B. megaterium cob* operon. Although they would not appear to be essential for cobalamin biosynthesis in *E. coli*, a comparative sequence analysis may give an inkling as to why they have been maintained within the *cob* operon.

a- CbiW compared to the database.

The *B. megaterium* CbiW protein was found to have some similarity with a *Clostridium pasteurianum* ferredoxin [Meyer, 1993] and, to a lesser extent, with a *Synechocystis sp.* ORF [Kaneko *et al.* 1996 and 1997]. CbiW might act as oxido-reductant, perhaps involved in the change in oxidation of the central cobalt ion. However, if this is the case, then *E. coli* must possess an equivalent protein as the engineered *E. coli* strains are able to synthesise cobyric acid in the absence of *cbiW*.

b- CbiY compared to the database.

The *B. megaterium* CbiY protein presented some similarity with a *Vibrio fischeri* NAD(P)-flavin oxido-reductase [Zenno *et al.* 1994] and to an *E. coli* ORF with an unknown function, part of a selenium operon [Sawers *et al.* 1991]. Selenocysteine is readily oxidised and is incorporated into proteins that are oxygen sensitive [Low and Berry, 1996]. It is interesting to note the anaerobicity connection between selenoproteins and cobalamin biosynthesis. However, it must be emphasised that, in *B. megaterium*, cobalamin biosynthesis does take place aerobically.

c- $CbiH_{60}$ compared to the database.

The *B. megaterium* C-terminal region of $CbiH_{60}$ is similar to the C- terminal region of a *Synechococcus sp.* ferredoxin-nitrite reductase (28.8% identity) [Luque *et al.* 1993] including 4 conserved cysteine residues in a region proposed as a 4Fe-4S/sirohaem domain found in nitrite and sulfite reductases [Crane *et al.* 1995]. $CbiH_{60}$, a fusion protein between precorrin-3 methylase and a reductase, could play a role in the level of oxidation during the ring contraction. CbiH is responsible for the addition of the methyl group at C-17 and is proposed as the protein which induces ring contraction for the anaerobic pathway [Santander *et al.* 1997].

## 2-6- Complementation of *E. coli* 302Δa to find the gene responsible for the $CysG^B$ activity.

As shown in section 2-2-2, one gene within the *S. typhimurium cob* operon is able to promote the synthesis of sirohaem from precorrin-2, or in other words, to mimic the $CysG^B$ function.

Within the *S. typhimurium cob* operon, 7 genes have no equivalent *in P. denitrificans*, *cbiD*, *-G*, *-K*, *-M*, *-N*, *-Q* and *-O*. These genes are the best candidates to encode for the $CysG^B$ like activity. In order to find the gene, the *cysG* deleted strain (302Δa) was transformed with a plasmid harbouring the *P. denitrificans cobA*. Different plasmids carrying candidate gene(s) had been transformed in the precorrin-2 producing strain. The results obtained after transformations with the various plasmids and restreaking onto minimum medium in the absence and presence of cysteine are shown in Table 2-8.

Table 2-8- Cysteine auxotrophy complementation of *E. coli cysG* deleted strain producing precorrin-2.

| Strain | genes on the second plasmid | growth on M9 medium after 24 hours at 37°C | |
|---|---|---|---|
| | | no cysteine | + cysteine |
| *E. coli* 302Δa, $cobA^{Pd}$ | pAR8827 | + | + |
| *E. coli* 302Δa, $cobA^{Pd}$ | pAR8977, pAR8814-$cbiD^M$ | + | + |
| *E. coli* 302Δa, $cobA^{Pd}$ | pAR8976, pAR8814-$cbiG^M$ | + | + |
| *E. coli* 302Δa, $cobA^{Pd}$ | pER126, pAR8827-$cbiK^{\Delta}$ | - | + |
| *E. coli* 302Δa, $cobA^{Pd}$ | pAR8603, *cbiM-N* | - | + |
| *E. coli* 302Δa, $cobA^{Pd}$ | pAR8606, *cbiF-G-H* | - | + |
| *E. coli* 302Δa, $cobA^{Pd}$ | pAR8358, *cbiN-Q-O* | - | + |
| *E. coli* 302Δa, $cobA^{Pd}$ | pAR8533, *cbiD* | - | + |

pAR8814 and pAR8827 contain the same insert, the *S. typhimurium cobI* operon (except *cbiB*), but their vectors are different; pAR8814 is derived from pKK223.3 and therefore has an OriC replication origin whereas pAR8827 is derived from pACYC184 and contains a p15A replication origin.

Mutations in either *cbiD* or *cbiG* within the *S. typhimurium cob* operon did not affect cysteine complementation. Moreover, combinations between *cobA* and either *cbiD* or *cbiG* or *cbiM* or *cbiN* or *cbiQ* or *cbiO* do not induce growth on minimum medium. None of these genes are therefore responsible, at least individually, for the $CysG^B$-like activity.

In contrast, pER126 (pAR8827 which contains a 54 bp deletion in *cbiK*) was unable to complement the cysteine auxotrophy indicating that *cbiK* is the gene encoding for the $CysG^B$-like activity. This was further confirmed when gene combinations of $cobA^{Pd}$ and *cbiK* alone were found to be able to complement 302Δa [Chapter 3, section 3-3-2].

## 2-7- Conclusion.

Within the 17 genes of the *S. typhimurium cob* operon, three genes (*cbiD*, *cbiG* and *cbiK*) have no similarity with *P. denitrificans cob* genes; their roles in cobalamin biosynthesis have been investigated.

Within the 14 genes of the *B. megaterium cob* operon, *cbiD* and *cbiG* are also present and three other genes, *cbiX*, *cbiW* and *cbiY*, have no similarity with any of the cobalamin biosynthetic genes found within *P. denitrificans or S. typhimurium*. Another gene, $cbiH_{60}$, had an unknown 3' extension fused to *cbiH*.

The two *cob* operons were cloned, transferred and expressed in *E. coli*. Indeed, the study of the biosynthetic genes essential for cobalamin production is much easier in *E. coli*, which has a *cobI* minus genotype and is therefore unable to transform precorrin-2 into cobinamide. The two *cob* operons were found to confer to the *E. coli* host cells the ability to produce corrins, but only in the absence of oxygen; cobinamide was synthesised from the *S. typhimurium cobI* operon and cobyrinic a,c-diamide from the *B. megaterium cob* operon.

Corrin production increased 4 fold when *cysG* was overexpressed in the presence of the *S. typhimurium cob* operon. Further studies revealed that the augmented 4-fold production was also obtained by overexpressing the *P. denitrificans cobA* instead of *cysG*: only the methylase part of CysG appeared necessary for the synthesis of cobalamin. However, and rather surprisingly, the *E. coli* strain containing the *S. typhimurium cob* operon and the *cobA* gene was able to produce sirohaem even in the absence of a $CysG^B$. It was deduced that a gene within the *cob* operon was responsible for the dehydrogenation and ferrochelation of precorrin-2. This gene was found to be *cbiK* although the gene displayed no sequence similarity to $CysG^B$. A deletion in the *cbiK* gene of the *S. typhimurium cob* operon does not prevent *E. coli* making cobyric acid. The role of CbiK in cobalamin synthesis will be described in the Chapter 3.

To investigate the role of *cbiD*, *cbiG*, *cbiX*, *cbiW*, *cbiY* and the extension of $cbiH_{60}$, mutations or deletions were introduced in these genes in the appropriate plasmids carrying the *cob* operons. The synthesis of cobyric acid by recombinant *E. coli* strains carrying the new constructs was analysed on bioassay plates.

As previously shown with the *S. typhimurium cbiD* mutant strain, *cbiD* was confirmed as essential for cobyric acid synthesis in both the *S. typhimurium* and *B. megaterium cob* operons. A glycine rich motif (GxGxG) conserved in several CbiD sequences suggests that CbiD might be the enzyme responsible for the methylation at C-1 but this hypothesis was not proven. The *B. megaterium* CbiD is an insoluble protein and although it was possible to resolubilise the protein, the refolded CbiD did not bind SAM.

As with *cbiD*, *cbiG* was also found to be necessary for both operons to synthesise cobyric acid. The protein is not similar to any other protein in the database (except for a slight similarity with the uncharacterised *P. denitrificans* CobE) and contains no obvious motifs. Thus, no role for CbiG in cobalamin synthesis is suggested.

A mutation in *cbiX* was found to reduce considerably the production of cobyric acid. CbiX was later discovered to have a $CysG^{B}$-like activity [Beck *et al.* 1997]. Therefore, it is not surprising that the bacteria were able to produce a low amount of cobyric acid as the host bacterium contained a genomic *cysG*. CbiX might be the *B. megaterium* equivalent protein to the *S. typhimurium* CbiK although they share no sequence similarity. These aspects are further investigated in Chapters 3 and 4.

CbiW and the C-terminal fused part of $CbiH_{60}$ were not found to be essential for the biosynthesis of cobyric acid in *E. coli*. They both share some similarity with ferredoxin proteins, which could suggest a role in an oxido-reduction system. Finally, a *cbiY* deletion had no effect on corrin synthesis. If CbiW, the C-terminus of $CbiH_{60}$ or CbiY play a role in cobalamin synthesis, their roles are not essential when the *cob* operon is expressed in *E. coli*.

In conclusion, within the *S. typhimurium cob* operon *cbiD* and *cbiG* are essential for cobyric acid synthesis despite the lack of similarity with the *P. denitrificans cob* genes. In regards to sirohaem biosynthesis, it was shown that a deletion in *cbiK* could be overcome by *cysG*. Within the *B. megaterium cob* operon the *cbiD*, *cbiG* and *cbiX* genes are essential for cobyrinic acid a,c-diamide synthesis whereas the *cbiW*, *cbiY* and the 3′ extension of $cbiH_{60}$ are not required for corrin synthesis, at least in *E. coli*. CbiK and CbiX are presumed to perform the same function since they both exhibit $CysG^{B}$-like activities.

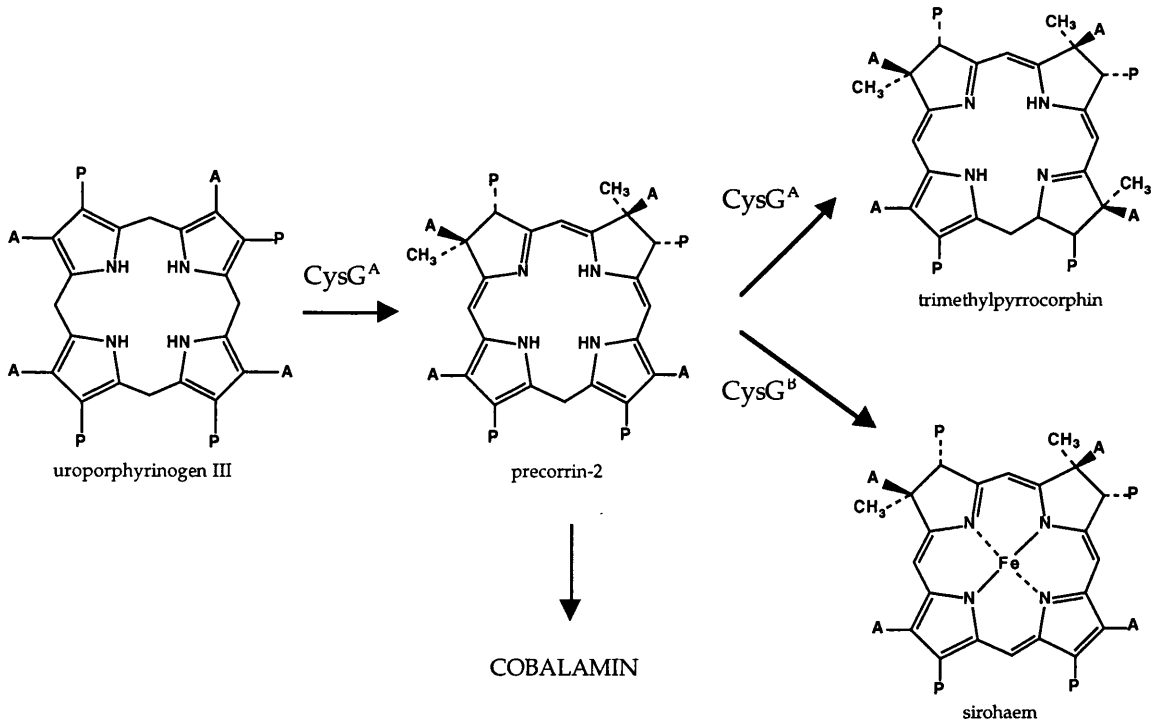
# CHAPTER 3,

## COBALT CHELATION, THE SECOND STEP OF VITAMIN $B_{12}$ BIOSYNTHESIS VIA THE ANAEROBIC PATHWAY AND SIMILARITIES WITH SIROHAEM SYNTHESIS

## 3-1- Introduction.

CysG is a protein of 457 amino acids and can be thought as a fusion of 2 functional domains, which are designated $CysG^B$ and $CysG^A$. The $CysG^A$ domain is responsible for the double methylation of uro'gen III at C-2 and C-7 which yields precorrin-2. The first half of *cysG*, encoding the $CysG^B$ domain, can be deleted and the resulting protein, $CysG^A$, is active and still performs the double methylation. Unfortunately, this enzyme can actually perform a third methylation at C-12 (Figure 3-1). No physiological role has been ascribed for the subsequent trimethylpyrrocorphin which is produced [Warren *et al.* 1990].

Figure 3-1- Overexpression of $CysG^A$ leads to synthesis of trimethyl-pyrrocorphin.




Some vitamin $B_{12}$ producers naturally possess a gene which is only similar to the $CysG^A$ domain of CysG. These proteins are called either $CysG^A$, as in *B. megaterium* [Raux *et al.* 1998], or CobA as in *P. denitrificans* or *P. shermanii* [Blanche *et al.* 1989 and Sattler *et al.* 1995]. In the work described in this chapter,

the *P. denitrificans cobA* was used to biosynthesise precorrin-2 instead of *cysG*$^A$ to avoid the overproduction of trimethylpyrrocorphin.

From precorrin-2 two further steps are required to make sirohaem, a dehydrogenation and iron chelation (Figures 3-1 and 1-14). These two steps are catalysed by the CysG$^B$ domain of CysG. In *E. coli* and *S. typhimurium*, a deletion in *cysG* leads to cysteine auxotrophy since sirohaem is required as the prosthetic group of sulfite reductase. If the bacteria are unable to catalyse the transformation of sulfite into sulfide, the synthesis of cysteine can not be completed (Figure 1-7) [Griffiths and Cole, 1987].

One of the main differences between the aerobic cobalamin pathway in *P. denitrificans* and the anaerobic pathway in *S. typhimurium*, *B. megaterium* or *P. shermanii*, is the timing of cobalt insertion. Indeed, in the anaerobic pathway, cobalt is added at an early stage, prior to ring contraction. Furthermore cobalt-precorrin-3 and cobalt-factor III have been isolated from *P. shermanii* extracts, but it was shown that the non-oxidised form was incorporated at a faster rate by cell-free extracts to produce cobyrinic acid [Müller *et al.* 1991].

In *S. typhimurium*, it has been shown that *cysG* is essential for cobalamin biosynthesis. However, it was not known whether cobalamin biosynthesis required only the activity of the CysG$^A$ domain or whether CysG$^B$ was also required, for instance to act as cobalt chelatase [Spencer *et al.* 1993].

In a recombinant *E. coli cysG* deleted strain, it was shown that only the methyltransferase activity of CysG, together with the proteins expressed from the *S. typhimurium cob* operon, was required to produce cobyric acid and that this strain was also able to produce sirohaem despite the absence of CysG$^B$. However, if the *cob* operon had a 54 bp deletion in the *cbiK* gene, sirohaem could no longer be synthesised. CbiK seems to be involved in the transformation of precorrin-2 into sirohaem, a reaction usually undertaken by CysG$^B$ (Chapter 2).

There are two more proteins which also have a $CysG^B$ type functionality;

(i) The *Saccharomyces cerevisiae* MET8p protein is described as a candidate for $CysG^B$ despite the lack of similarity between MET8p and CysG at the primary structure level. The only common feature shared by these proteins is a putative $NAD^+$ binding site at the N-terminus. The *MET8* mutant was identified in *S. cerevisiae* by its methionine requiring growth phenotype. Moreover that mutation could be complemented by the *S. typhimurium cysG* [Hansen *et al.* 1997].

(ii) The *B. megaterium cbiX* together with the *P. denitrificans cobA* was shown to complement the *E. coli cysG* deleted strain, 302Δa. CobA and CbiX are therefore able to mimic the CysG activities [Beck *et al.* 1997].

In total, four proteins have been identified as being able to perform the transformation of precorrin-2 into sirohaem: CysG, CbiK, CbiX and MET8p. CysG is involved in both cobalamin and sirohaem synthesis. The *cbiK* and *cbiX* genes are found in vitamin $B_{12}$ producers; *cbiK* within the *S. typhimurium cob* operon and *cbiX* within the *B. megaterium cob* operon. Finally, MET8p can be only involved in sirohaem synthesis as *S. cerevisiae* does not possess the genes required to make cobalamin.

This chapter examines the role of CysG, CbiK, CbiX and MET8p in cobalamin and sirohaem biosynthesis. It investigates whether these enzymes are able to insert cobalt into the macrocycle to produce corrinoid intermediates. Moreover, it also investigates whether the dehydrogenase activity (associated with $CysG^B$) is important for vitamin $B_{12}$ synthesis.

## RESULTS

### 3-2- Comparison of sequence, CysG, MET8p, CbiX and CbiK.

The sequences of CysG, MET8p, CbiX and CbiK were compared to determine any possible similarity between these proteins.

## 3-2-1- Comparison between CysG, MET8p, CbiX and CbiK.

The sequence comparison was undertaken using the Fasta program (Unix/GCG Wisconsin package). The best results are shown in Table 3-1.

Table 3-1- Sequence comparisons of CysG, MET8p, CbiX and CbiK.

| proteins | CbiK | CbiX | CysG |
|---|---|---|---|
| **Met8p** | 25.0% identity in 47 aa | 28.6% identity in 21 aa | 22.5% identity in 218 aa (gaps) |
| **CysG** | 24.6% identity in 57 aa | 30.0% identity in 50 aa | |
| **CbiX** | 23.0% identity in 74 aa | | |

The *E. coli* CysG Swissprot accession number is p11098; *S. typhimurium* CbiK is q05592 *and S. cerevisiae* Met8p is p15807.

The percentage of identity between CysG, CbiK, CbiX and Met8p is very low, under 30%, and only for short sections of the polypeptide chain. This indicates that the proteins share no significant similarity. In the case of the comparison between CysG and Met8p, the percentage of identity between the two proteins was 22.5 %, but the high number of gaps in the alignment implied that these proteins are not truly similar. Nevertheless, CysG and Met8p do share a common motif, the $NAD^+$ binding site at the N-terminus of both proteins [Hansen *et al.* 1997] (Figure 3-2).

Apart from the common $NAD^+$ binding site of CysG and Met8p, overall CbiK, CbiX, Met8p and CysG are not significantly similar.

Figure 3-2- CysG, Met8p and YlnF $NAD^+$ binding sites.


The amino acids written in grey boxes represent the consensus between 4 $NAD^+$ binding sites in CysG proteins [Woodcock *et al.* 1998].

3-2-2- Comparison of CysG, MET8p, CbiX and CbiK with the database.

A systematic Fasta search with CysG, CbiK, CbiX and Met8p was undertaken to find hypothetical similar proteins within Swissprot. Significant results were recorded and are listed in Table 3-2.

Table 3-2- CysG, MET8p, CbiX and CbiK compared to Swissprot.

| Proteins | Similar proteins found in the database and role | % identity/aa |
|---|---|---|
| CbiK | *Porphyromonas gingivalis* | 37.1% in 259 aa |
| | *Clostridium acetobutylicum* (contig 427) | 35.2% in 270 aa |
| | *Clostridium acetobutylicum* (contig 206) | 32.0% in 269 aa |
| | *Yersinia Enterocolitica* HemR (p31499) | 30.1% in 103 aa |
| CbiX | *Synechocystis sp.* sll10048 | 36.2% in 301 aa |
| | *B. subtilis* YlnE | 32.6% in 261 aa |
| CysG | *S. typhimurium* CysG (q05592) | 90.2% in 457 aa |
| | *Neisseria meningitidis* CysG | 52.9% in 461 aa |
| | *Clostridium josui* [HemA-$CysG^B$] (q59292) | 25.7% in 144aa |
| | *Vibrio anguillarium* [$CysG^B$ / ?] (jc4347) | 32.7% in 266 aa |
| | *B. subtilis* YlnF [$CysG^B$/dehydrogenase] | 27.1% in 144 aa |
| | *P. denitrificans* CobA (p21631) | 42.7% in 248 aa |
| | *B. subtilis* YlnD | 47.3% in 243 aa |
| | *B. megaterium* $CysG^A$ | 46.2% in 240 aa |
| | *Clostridium josui* [$CysG^A$-HemD] (q59294) | 49.4% in 247 aa |
| | *Saccharomyces cerevisiae* Met1p [? / $CysG^A$] (z28294) | 36.9% in 233 aa |
| Met8p | *Aquifex aeolicus* CysG (hyperthermophilic/o067282) | 26.4% in 193 aa |
| | *B. subtilis* YlnF | 25.7% in 202 aa |

The strict anaerobe *P. gingivalis* contains a CbiK which is 37.1% identical to the *S. typhimurium* CbiK. The *P. gingivalis* CbiK does not seem to be included within a *cob* operon whereas other $B_{12}$ genes in *P. gingivalis* are grouped in loci [Dr J. Roper, pers. Commun.]. Two hypothetical proteins similar to CbiK were found in *Clostridium acetobutylicum* in two different contigs (427 and 206) which both contain only vitamin $B_{12}$ biosynthetic genes. The *cbiK* found in contig 427 is part of the cobalamin locus also including *cbiC, cbiA, cbiP, cbiG* and *cbiB*. In contig 206, only 2 genes are present, the second *cbiK* and also a second *cbiB*. The two *C. acetobutylicum* CbiK share 42.9% identity (in 266 amino acids) indicating either that this bacterium harbours two vitamin $B_{12}$ biosynthetic genes from evolutionary different origins or that contig 206 was isolated and sequenced as

a contaminant from a different bacterium. Since the *C. acetobutylicum* data are from a genome sequencing project, no further information was available at the time of the search. However, it is now clear that *cbiK* is not specific to *S. typhimurium* and that other vitamin $B_{12}$ producers contain the same gene, suggesting a common role in cobalamin biosynthesis. A protein from *Yersinia Enterocolitica*, HemR, also shows some similarity with CbiK, but only over 100 amino acids. This slight similarity may reflect, for instance, a common substrate binding site. Indeed, HemR is a hemin receptor protein and is an outer membrane iron-regulated protein [Stojiljkovic and Hantke, 1992].

The *B. megaterium cbiX* gene product was found to be similar to a *Synechocystis sp.* sll10048 hypothetical protein. An interesting feature shared by CbiX and this homologue is a histidine-rich region at their C-terminal, which suggests that the protein may be involved in metal chelation. Indeed, poly-histidine motifs are implicated in nickel binding [Rey *et al.* 1994 and Fu *et al.* 1995]. The CbiX poly-histidine motif is more likely to be required in cobalt binding than nickel. CbiX is also 32.6% identical to the *B. subtilis* YlnE in 261 amino acids, but the poly-histidine tail is not present in YlnE. *ylnE* is the central gene of a group of three genes translationally coupled, including YlnD, identical to uro'gen III methyltransferases (CobA/$CysG^A$), YlnE and finally YlnF. This last enzyme is a 162 amino acid peptide, contains a conserved $NAD^+$ binding site (Figure 3-2) and shares 21.7% identity in 144 amino acids with the N-terminal of CysG. The YlnD, -E, -F proteins may represent the three functions performed by CysG. The fact that the YlnE does not possess the poly-histidine tail may indicate that this enzyme is involved in iron chelation for sirohaem biosynthesis rather than in cobalt chelation. Besides, *B. subtilis* does not produce cobalamin.

CysG encodes for a multifunctional protein and thus similarity searches in the database locate proteins corresponding to the different functions of the protein:
(1) The CysG homologues: *E. coli*, *S. typhimurium* and *Neisseria meningitidis* [Mac Veigh *et al.* unpublished data] CysGs overall share a high degree of similarity.

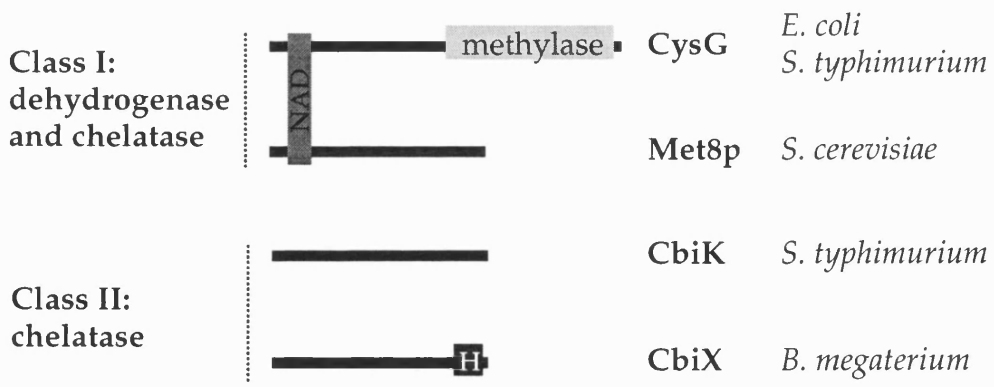
(2) The $CysG^B$ homologues: the N-terminus of *Vibrio anguillarium* jc4347 [Milton *et al.* 1995] is only similar to the $CysG^B$ domain. Its C-terminus remains uncharacterised. *Clostridium josuii* [HemA/$CysG^B$] (q59292) most likely

represents a fusion protein between HemA and $CysG^B$, but the percentage of identity with the $CysG^B$ alignment is low (25.7%). Nonetheless, the $NAD^+$ binding site is very similar to that found in CysG.

(3) The $CysG^A$ homologues: Some proteins are found to be only similar to the methyltransferase domain of CysG, for example the *P. denitrificans* CobA, *B. megaterium* $CysG^A$ or *B. subtilis* YlnD. Other proteins are the result of a fusion with a $CysG^A$ domain, like the *Clostridium josui* [$CysG^A$/HemD] (q59294) which is a fusion of $CysG^A$ and HemD or the *Saccharomyces cerevisiae* Met1p which contains an uncharacterised N-terminus fused to $CysG^A$.

CysG, CbiK, Met8p and CbiX can be separated in two classes; the first class contains a $NAD^+$ binding site (CysG and Met8p) whereas the second class which includes CbiK and CbiX do not have such a motif and therefore most probably do not have any dehydrogenase activity (Figure 3-3).

Figure 3-3- The two classes of $CysG^B$ like proteins.



## 3-3- Sirohaem synthesis in *E. coli cysG* mutant strain.

*E. coli* 302Δa, a *cysG* deleted strain [Griffiths and Cole, 1987], was used to study *in vivo* the role of CbiK and Met8p in sirohaem synthesis. The synthesis of sirohaem was analysed by the ability of the host strain (302Δa) to grow in minimum medium, reflecting complementation of the cysteine auxotrophy.

### 3-3-1- Construction of the different plasmids.

Three new plasmids were constructed for this study; pER119 which is a pKK223.3 derived plasmid containing the *P. denitrificans cobA*; pER170 which is the same vector but harbouring both the *P. denitrificans cobA* and *S. typhimurium cbiK*; and finally, pER250 which harbours the *P. denitrificans cobA* and *S. cerevisiae MET8*. All the cloning details are given in Materials and Methods (section 8-3). As a control, the final plasmid used for this study was pER108, which is a pKK vector harbouring the *E. coli cysG* [Raux *et al.* 1997].

### 3-3-2- Complementation of the cysteine auxotrophy of *E. coli* 302Δa and effect of the cobalt.

The *E. coli* strain 302Δa, containing plasmid pAR8086 (pACYC184 harbouring the $lacI^q$ repressor to repress the *tac* promoter) was transformed with the following plasmids: pER119 (*P. denitrificans cobA*), pER170 (*P. denitrificans cobA* and *S. typhimurium cbiK*), pER250 (*P. denitrificans cobA and S. cerevisiae MET8*) and pER108 (*E. coli cysG*). The resulting strains were tested for their ability to grow on minimum medium plates with either no supplement, 5μM cobalt or cysteine. Cobalt was added to the minimum medium plates to see if it interfered with the strains ability to catalyse the final step of sirohaem synthesis. The results are shown in the Table 3-3.

Table 3-3- Growth of strain ER172, ER173, ER251 and ER182 on minimum medium plates

| Strain | plasmid (Gene(s)) | Growth on minimum medium after 24 hours at 37°C | | |
|---|---|---|---|---|
| | | No addition | $CoCl_2$ (5 µM) | Cysteine |
| ER172 | pER119 (*cobA*) | – | – | +++ |
| ER173 | pER170 (*cobA* + *cbiK*) | ++ | – | +++ |
| ER251 | pER250 (*cobA* + *MET8*) | +++ | | +++ |
| ER182 | pER108 (*cysG*) | +++ | ++ | +++ |

The E. coli host strain is *E. coli* 302Δa containing pAR8086 (pACYC184 harbouring $lacI^q$). The symbol, –, means no growth on minimum medium plates; ++, slow growth and, +++, normal growth.

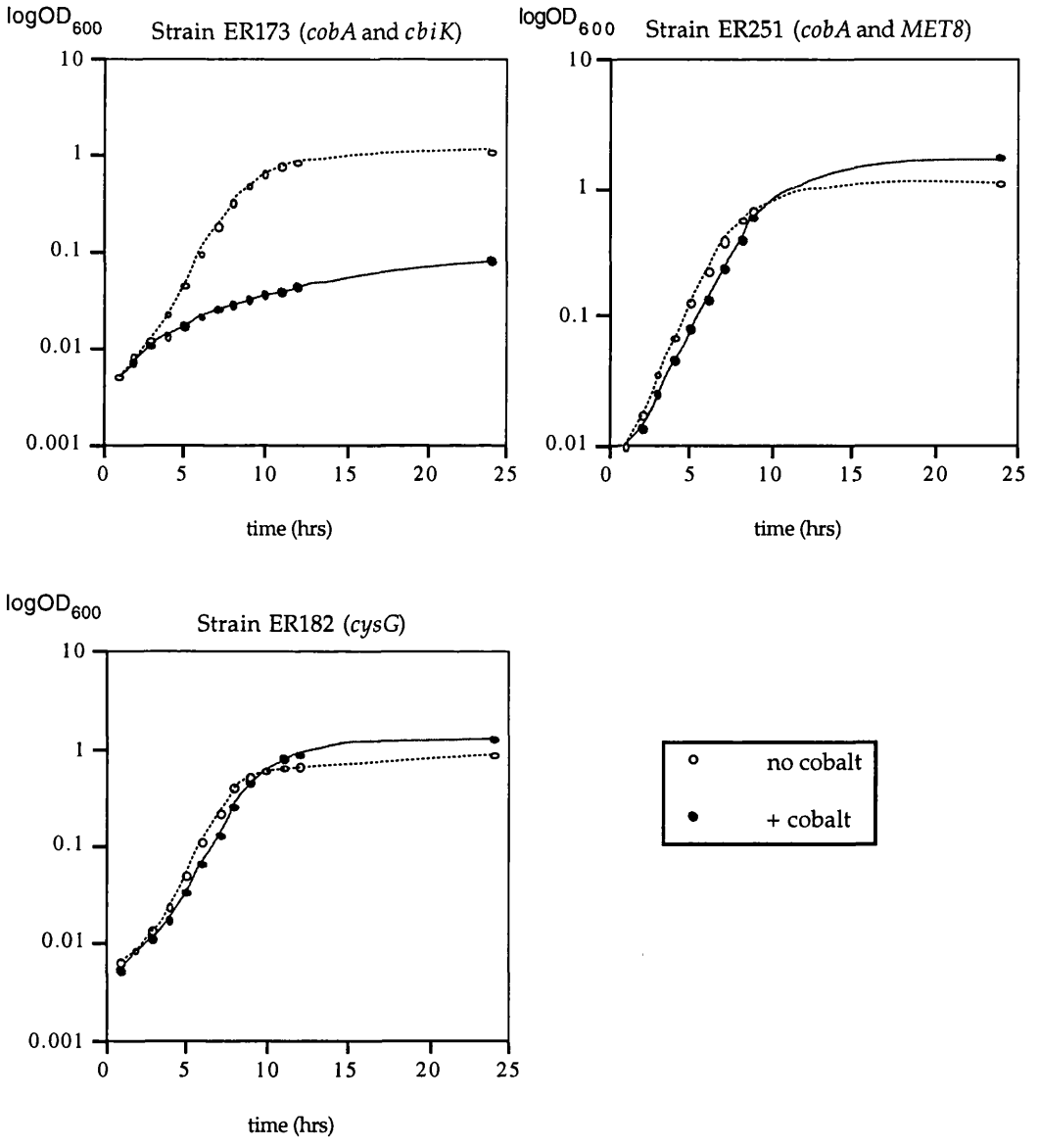
Strain ER172 (*cobA*) does not grow on minimum medium without cysteine. This confirms that CobA on its own is not able to produce sirohaem in a *cysG* deleted strain. The three other strains containing either *cobA-cbiK*, *cobA-MET8* or *cysG*, can complement the *E. coli* 302Δa *cysG* deleted strain. Since these three constructs confer upon 302Δa the ability to grow on minimum medium, CbiK, MET8p and CysG can therefore all perform the transformation of precorrin-2 to sirohaem. This transformation requires two separate reactions, a dehydrogenation and chelatase action to introduce ferrous iron in the modified tetrapyrrole.

To investigate whether the observed ferrochelation of sirohaem synthesis by CbiK, MET8p and CysG was a fortuitous surrogate reaction of a cobalt chelatase, exogenous cobalt was added to the minimum medium. This resulted only in a detrimental manner to ER173 (*cobA* and *cbiK*) as no growth could be observed on plates containing cobalt.

The effect of exogenous cobalt on the three *cysG* complementing strains, *cobA-cbiK*, *cobA-MET8* and *cysG*, was also investigated in liquid culture. The strains were grown in 100 ml of minimum medium, with either no supplement or with

5µM cobalt. Their growths were followed for 24 hours at 37°C. Figure 3-4 shows the growth curves of the different strains.

Figure 3-4- Growth curves of strain ER172, ER251 and ER182.




The curves in black dotted lines represent the growth obtained in minimum medium without cobalt whereas the curves in grey represent the growth curves obtained in presence of cobalt.

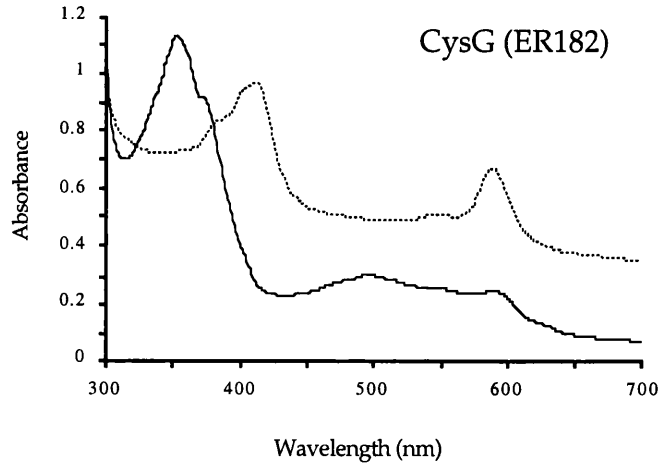
Cobalt had a little effect on the growth curve profiles of either ER251 (*cobA* and *MET8*) or ER182 (*cysG*), which confirms the results obtained on plates. MET8p and CysG therefore appear to be able to synthesise sirohaem in presence of cobalt.

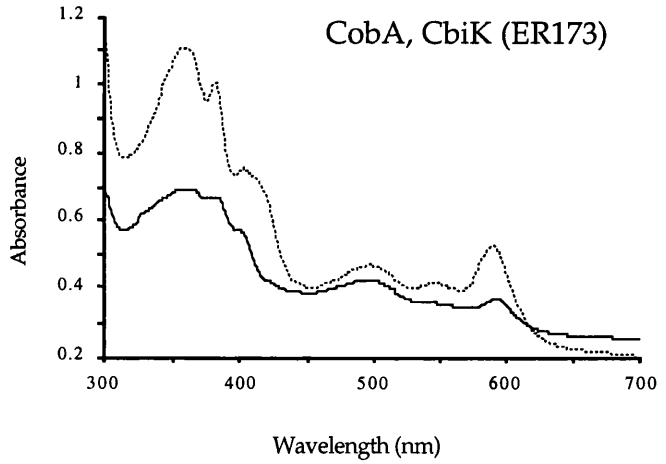
In contrast, exogenous cobalt retarded the growth of ER173 (*cobA* and *cbiK*) such that the strain reached only an $OD_{600}$ of 0.08 after 24 hours of culture. The inhibited complementation suggests that CbiK appears to have a preference for $Co^{2+}$ over $Fe^{2+}$. If this is the case, this implies that the major role of CbiK is to act as cobalt chelatase in cobalamin synthesis rather than as an iron chelatase in sirohaem synthesis. In Chapter 2 [section 2-4-2], it was demonstrated that in the presence of a genomic *cysG* background, *cbiK* is not essential for cobalamin synthesis. The role of CbiK in cobalamin synthesis therefore needs to be re-evaluated in a *cysG* minus background, such as in strain 302Δa (*E. coli cysG* mutant strain). This is investigated in section 3-4-1.

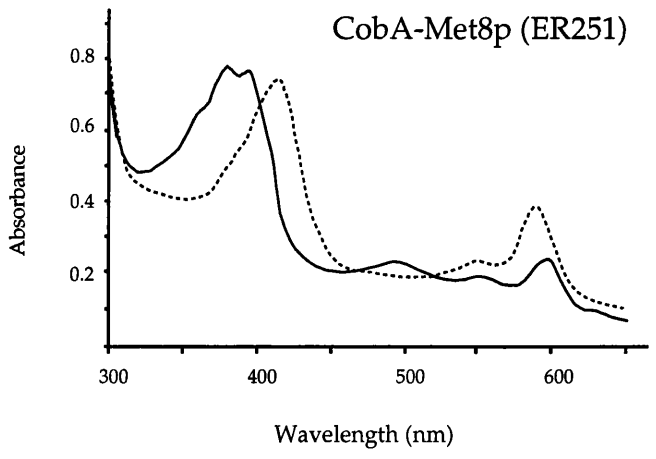
### 3-3-3- UV-visible spectra of accumulated compounds *in vivo*.

Cell extracts of ER173 (*P. denitrificans cobA* and *S. typhimurium cbiK*), ER251 (*P. denitrificans cobA and S. cerevisiae MET8*) and ER182 (*E. coli cysG*), after growth on minimum medium supplemented with ALA (10 mg/l) and in the absence or in the presence of $CoCl_2$ (5 µM) in anaerobic conditions, were analysed by UV-visible spectroscopy. The accumulated tetrapyrrole-derived material was extracted from the cell lysate by binding it onto an ion-exchange resin as described in Materials and Methods (section 8-5-5).

Figure 3-5- Spectra of tetrapyrrole-derived compounds from strain ER173, ER251 and ER182.



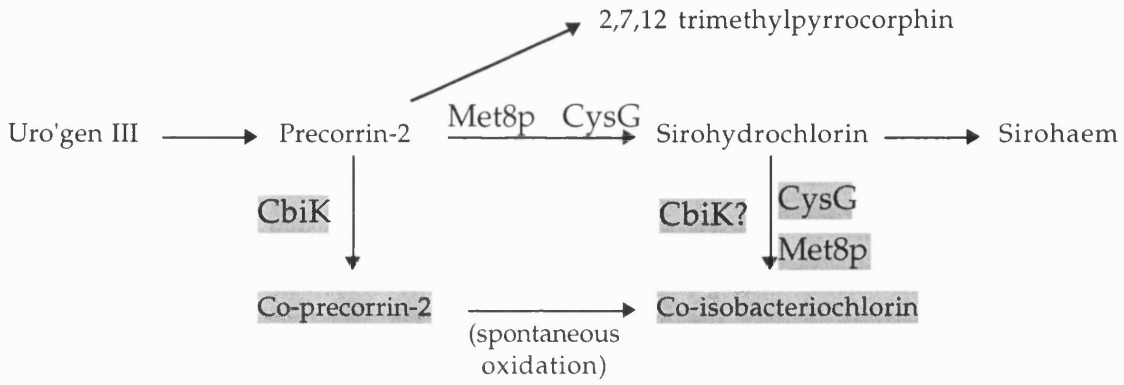

The black lines represent the spectra obtained from cobalt-free medium whereas the dotted lines represent the spectra obtained when cobalt is present in the medium.

In the absence of cobalt the material extracted from ER182 (CysG) gave a spectrum with an absorption maximum at 354 nm (Figure 3-5) which is indistinguishable from the spectrum of trimethylpyrrocorphyrin (precorrin-2 with an extra methyl at C-12 due to an overexpression of CysG) which has been reported as a non-physiological product of that reaction [Scott *et al.* 1990; Warren *et al.* 1990]. The spectrum of the material isolated from the same strain (CysG) but in presence of cobalt gave a different spectrum with an absorption maxima shifted from 354 nm to 413 nm and a second maxima at 590 nm, which is consistent with the presence of cobalt isobacteriochlorin [Battersby and Sheng, 1992]. Cobalt-isobacteriochlorin (or cobalt sirohydrochlorin) is the oxidised form of cobalt-precorrin-2 and is thought to act as the prosthetic group for sulphite reductase in a family of primitive sulphite reducing organisms [Hatchikian, 1981; Moura *et al.* 1980].

The spectrum obtained from ER173 (CobA and CbiK) indicates the presence of both trimethylpyrrocorphin (absorption maxima at 354 nm) and cobalt isobacteriochlorin (absorption maxima at 413 nm and 590 nm). Another maxima was observed at 382 nm which could suggest the presence of sirohydrochorin or sirohaem (the two molecules gave indistinguishable spectra with an absorption maxima at 378 nm [Warren *et al.* 1994]). In the absence of cobalt, the spectrum of the CobA/CbiK extract is difficult to interpret, and relates to the instability of the compounds.

Finally, the material isolated from ER251 (CobA and MET8p) gave an identical spectrum to CysG in the presence of cobalt, that of cobalt isobacteriochlorin. In the absence of cobalt, the absorbance maxima are at 376 nm and 390 nm and reflect the presence of sirohydrochorin or sirohaem. There is no major absorbance at 354 nm, and therefore, CobA and Met8p do not seem to overproduce the trimethylated compound in the same manner as observed with CysG.

Figure 3-6- Scheme interpreting the absorption spectrum results.




The grey colour indicates the routes used in presence of cobalt by CysG, Met8p or CbiK. The black arrows indicate the routes used in absence of cobalt.

The Met8p spectra are consistent with the ability of Met8p to act as a dehydrogenase and a chelatase, synthesising sirohydrochlorin or sirohaem from precorrin-2 in the absence of cobalt, and cobalt-sirohydrochlorin in its presence (Figure 3-6). Identical spectra have been attained with a multi-enzymatic assay containing purified CobA and Met8p [data not shown]. The spectrum obtained with CysG leads to a similar conclusion, but the accumulation of trimethylpyrrocorphin in the absence of cobalt, indicates a less efficient dehydrogenation of precorrin-2 as with Met8p. Indeed, it was observed, in Woodcock *et al.* 1998, that the dehydrogenation activity of CysG might stabilise the synthesis of precorrin-2 and therefore prevents the excessive accumulation of the trimethylated compound. The spectrum obtained with CbiK is in agreement with the role of CbiK as a cobalt chelatase at the level of precorrin-2 or sirohydrochlorin. The fact that the trimethylpyrrocorphin compound is also observed may reflect that the non-catalytic oxidation of precorrin-2 represents a limiting step in the synthesis of cobalt-sirohydrochlorin. This result is consistent with the fact that the true intermediate of the anaerobic cobalamin pathway is thought to be cobalt-precorrin-2 and not cobalt-sirohydrochlorin.

## 3-4- Essential genes required for the synthesis of cobalamin in the absence of *cysG*.

In section 2-2-2, it was shown that only the $CysG^A$ domain (methyltransferase part of CysG) or CobA were essential for cobalamin synthesis with the *S. typhimurium cob* system. This result was obtained in the presence of *cbiK* within the *S. typhimurium cob* operon. When the role of *cbiK* was investigated, it was not found to be essential for cobalamin synthesis in the presence of a genomic *cysG* background (Section 2-4-2). With the knowledge that both CbiK and CysG can undertake cobalt chelations, it became important to determine whether *cbiK* was essential for cobalamin synthesis in absence of a *cysG* genomic background, and then if a *cbiK* and *cysG* minus genotype could be complemented by *cbiX*, *MET8* and $cysG^{G21D}$. $CysG^{G21D}$ is a point-mutation introduced in *cysG* leading to the modification of the glycine residue 21 to aspartate, which disrupts the $NAD^+$ binding site from GxGxxA to GxDxxA and prevents the binding of $NAD^+$ [Woodcock *et al.* 1998]. The comparison of complementation of the *cysG* and *cbiK* double-mutant strain by *cysG* or $cysG^{G21D}$ was undertaken to study if the dehydrogenation step was a requirement for cobalamin synthesis.

Moreover, the synthesis of cobyric acid was investigated from the *B. megaterium cob* operon after it had been transferred to the *E. coli cysG* deleted strain (302Δa).

### 3-4-1- Effect of deletions in *cysG* and *cbiK*.

The genes responsible for the transformation of cobalt precorrin-2 into cobyric acid were introduced into *E. coli* 302Δa on the *cbiK* deleted *S. typhimurium cob* operon plasmid (giving strain ER185). This strain was transformed with a second plasmid, harbouring either *P. denitrificans cobA*, *P. denitrificans cobA* and *S. typhimurium cbiK*, *P. denitrificans cobA and S. cerevisiae MET8* , *P. denitrificans cobA* and *B. megaterium cbiX* [pER179], or *E. coli cysG and E. coli* $cysG^{G21D}$ [Woodcock *et al.* 1998].

The synthesis of cobyric acid was analysed by the ability of the different cell extracts to induce growth of the *S. typhimurium metE cysG* strain contained in the bioassay plates. The results are given in Table 3-4.

Table 3-4- Cobyric acid assays from strain ER185 (*cysG* deleted strain, *S. typhimurium cob* operon; *cbiK*$^{\Delta}$)

| Strain | *cob* operon / gene deletion | Gene(s) in the 2nd compatible plasmid | Cobyric acid (pmol/$OD_{600}$) under: | |
|---|---|---|---|---|
| | | | non-inducing conditions | inducing conditions |
| ER188 | *S. typ / cbiK*$^{\Delta}$ | *cobA* | 0 | 0 |
| ER190 | *S. typ / cbiK*$^{\Delta}$ | *cobA, cbiK* | 14 | 482 |
| ER191 | *S. typ / cbiK*$^{\Delta}$ | *cysG* | 536 | 544 |
| ER189 | *S. typ / cbiK*$^{\Delta}$ | *cobA, cbiX* | 82 | 466 |
| ER252 | *S. typ / cbiK*$^{\Delta}$ | *cobA, MET8* | 390 | 290 |
| ER278 | *S. typ / cbiK*$^{\Delta}$ | *cysG*$^{G21D}$ | 250 | 170 |

All the strains are derived from strain ER185: *E. coli cysG* deleted strain (302Δa) containing the *S. typhimurium cbiA-C-D-E-T-F-G-H-J-K*$^{\Delta}$*-L-M-N-Q-O-P* genes with a deletion in *cbiK*.

Strain ER185, *E. coli* 302Δa containing pER126 (harbouring the *S. typhimurium cbiA-C-D-E-T-F-G-H-J-K*$^{\Delta}$*-L-M-N-Q-O-P* with a deletion of 54 bp in *cbiK*) was initially transformed with a second compatible plasmid carrying only the uro'gen III methylase *cobA*. No cobyric acid could be detected from this strain (Table 3-4, strain ER188) on bioassay plate, proving that *cbiK* is an essential gene for cobalamin synthesis in the absence of a *cysG* background.

The inability to produce cobyric acid could be rectified by transforming ER185 with a plasmid carrying *cobA* and *cbiK* (strain ER190). When non-induced, cobyric acid synthesis was relatively low, 14 pmol/$OD_{600}$, which may reflect a poor expression of *cbiK*, cloned downstream of *cobA* in pER170. However, in the presence of IPTG, the *cobA/cbiK* strain produces cobyric acid up to 482 pmol/$OD_{600}$, confirming that the cobalamin minus phenotype observed in strain ER188 (*cobA* alone) was due to the absence of a functional *cbiK*. A similar

high level of cobyric acid could be obtained by transforming ER185 with plasmids harbouring either *cysG* (strain ER191, 544 pmol/$OD_{600}$), *cbiX* (strain ER189, 466 pmol/$OD_{600}$), *MET8p* (strain ER252, 290 pmol/$OD_{600}$) or $cysG^{G21D}$ (strain ER278, 170 pmol/$OD_{600}$). These four genes are able to complement the *cbiK* deletion and allow the synthesis of cobyric acid.

CbiK, CbiX, Met8p, CysG and $CysG^{G21D}$ can all perform the cobalt chelation. The fact that $CysG^{G21D}$ can also restore cobyric acid synthesis means that the dehydrogenase activity of CysG does not appear to be required for cobalamin synthesis in *S. typhimurium*.

## 3-4-2- *cbiK* can complement a *cbiX* mutation

The essential role of *cbiX* in cobalamin synthesis has been described in Chapter 2 (2-5-2). Furthermore, CbiX is able to complement the *E. coli cysG* deleted strain containing the *S. typhimurium cob* operon with a deletion in *cbiK*. However, it was not known whether *cbiK* would be able to complement the *B. megaterium cob* operon containing a mutation in *cbiX*.

*E. coli*, with a genomic *cysG*, was transformed with the *B. megaterium cob* operon harbouring a mutation in *cbiX* together with a second, compatible, plasmid carrying the *S. typhimurium cbiP* and *cbiK*.

Table 3-5- *cbiK* can complement a *cbiX* mutation within the *B. megaterium cob* operon

| Strain | *cob* operon/ gene mutation | Gene(s) in the 2nd compatible plasmid | Cobyric acid (pmol/$OD_{600}$) under: | |
|---|---|---|---|---|
| | | | Non-inducing conditions | Inducing conditions |
| AR8925 | *B. meg / cbiX*$^{M}$ | *cbiP* | 0.1 | 2 |
| AR8967 | *B. meg / cbiX*$^{M}$ | *cbiP, cbiX* | 110 | 440 |
| ER205 | *B. meg / cbiX*$^{M}$ | *cbiP, cbiK* | 106 | 563 |

All the strains are derived from *E. coli* (with a genomic *cysG*) containing the *B. megaterium cbiW-*$H_{60}$*-*$X^{M}$*-J-C-D-ET-L-F-G-A-cysG*$^{A}$*-cbiY-btuR* with a mutation in *cbiX*. As the first corrin intermediate that can be detected by bioassay is cobyric acid, the *S. typhimurium cbiP* was added to all strains to transform cobyrinic acid a,c-diamide into cobyric acid.

The results shown in Table 3-5 reveal that the reduced production of cobyric acid when the *B. megaterium cob* operon contains a mutation in *cbiX* could be overcome by *cbiK*. Indeed, the level of cobyric acid synthesised by strain ER205 (563 pmol/$OD_{600}$ when induced with IPTG) is similar to that produced by strain AR8967 which has a non-mutated copy of *cbiX* (440 pmol/$OD_{600}$). CbiK and CbiX would therefore appear to be isofunctional.

### 3-4-3- With the *B. megaterium cob* operon in *E. coli cysG* deleted strain.

A high level of cobyric acid is synthesised in *E. coli* from the *S. typhimurium cob* operon with a deletion in *cbiK* because of the presence of the genomic *cysG*, even though the synthesis of corrins relies upon the presence of *cbiK* in an *E. coli cysG* deleted strain (302Δa). In the absence of *cbiK* and *cysG*, corrin synthesis could then be restored by adding back either a wild type of *cbiK* or *cbiX*. Conversely, a very low level of corrin is synthesised in *E. coli* containing the *B. megaterium cob* operon with a mutation in *cbiX* despite the presence of the genomic *cysG*. Therefore, it became important to investigate the synthesis of cobyric acid from the *B. megaterium cob* operon in the *E. coli cysG* deleted strain.

Table 3-6- Cobyric acid assays from the *B. megaterium cob* operon and the *cbiX* mutated operon in an *E. coli cysG* deleted strain.

| Strain | plasmid carrying *the B. megaterium cob* operon | Gene(s) in the 2nd compatible plasmid | Cobyric acid (pmol/$OD_{600}$) under: | |
|---|---|---|---|---|
| | | | Non-inducing conditions | Inducing conditions |
| AR8948 | pAR8877 | *cbiP* | 0 | 0 |
| AR8949 | pAR8911 = pAR8877-[$cbiX^M$] | *cbiP* | 0 | 0 |
| ER203 | pAR8877 | *cbiP, cysG* | 7 | 204 |
| ER298 | pAR8911 | *cbiP, cysG* | 0 | 6 |

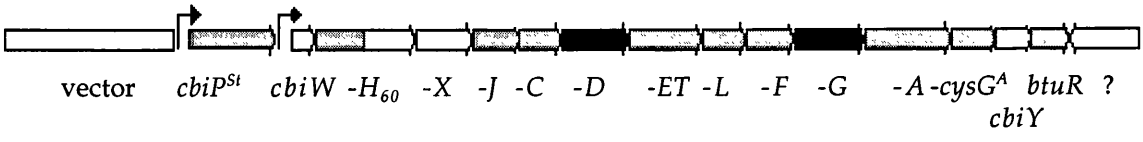
The host strain is strain *E. coli* 302Δa deleted in the *cysG* gene. pAR8877 contains the *B. megaterium cbiW-$H_{60}$-X-J-C-D-ET-L-F-G-A-$cysG^A$-cbiY-btuR* genes. pAR8911 contains the same genes but with a mutation in *cbiX*.

As shown in Table 3-6, no cobyric acid was synthesised in the *cysG* minus strain with the complete *B. megaterium cob* operon (strain AR8948) or with the same construct which contains a frameshift in *cbiX* (strain AR8949). Cobyric acid synthesis was produced up to 204 pmol/$OD_{600}$ only when a *cysG* together with the complete *B. megaterium cob* operon was present (strain ER203). Indeed, the *B. megaterium cob* operon with a mutation in *cbiX*, even in the presence of *cysG*, led to a cobyric acid synthesis of only 6 pmol/$OD_{600}$ (strain ER298).

The conclusion from these experiments is that *cysG* is essential for the synthesis of cobyric acid with the *B. megaterium cob* operon. Moreover, *cysG* is not able to fully complement a *cbiX* mutation in the *B. megaterium cob* operon. The cobalt chelatase activity of CysG is not able to act as an efficient surrogate for the activity of CbiX.

In order to determine if the observed *cysG* requirement for cobyric acid synthesis with the *B. megaterium* system was due to the CysG dehydrogenase activity, the *B. megaterium cob* operon was studied together with the two classes of gene coding for $CysG^B$-like activity (class I: *MET8, cysG* and class II: *cbiK, cbiX, cysG*$^{G21D}$). A third class of $CysG^B$ like-protein can be created with the *B. subtilis* YlnF which very likely catalyses the sirohydrochlorin dehydrogenation. To accomplish this, the *B. megaterium cob* operon was subcloned under the control of a *tac* promoter together with the *S. typhimurium cbiP* into a vector derived from pACYC184. The new plasmid was named pER270 (Figure 3-7) and was introduced in *E. coli cysG* deleted strain (302Δa) to give strain ER270.

Figure 3-7- Genetic map of plasmid pER270.




The cloning details are described in Materials and Methods (section 8-3-5).

Strain ER270 was transformed with compatible pKK223.3 derived plasmids harbouring either *P. denitrificans cobA*, *P. denitrificans cobA* and *S. typhimurium cbiK*, *P. denitrificans cobA and S. cerevisiae MET8*, *P. denitrificans cobA* and *B. megaterium cbiX*, *P. denitrificans cobA* and *B. subtilis ylnF*, *E. coli cysG* or *E. coli* $cysG^{G21D}$. The restriction maps of the different plasmids 302Δa are analysed in section 8-3-6.

Table 3-7- Cobyric acid assays from strain ER270 (*cysG* deleted strain, *B. megaterium cob* operon and *S. typhimurium cbiP*)

| Strain | *cob* operon | Gene(s) in the 2nd compatible plasmid | Cobyric acid (pmol/$OD_{600}$) under: | |
|---|---|---|---|---|
| | | | Non-inducing conditions | Inducing conditions |
| ER273 | *B. meg* | *cobA* | 0 | 0 |
| ER274 | *B. meg* | *cobA, cbiK* | 0 | 2 |
| ER276 | *B. meg* | *cobA, cbiX* | 0 | 1 |
| ER272 | *B. meg* | $cysG^{G21D}$ | 0 | 3 |
| ER271 | *B. meg* | *cysG* | 2 | 110 |
| ER275 | *B. meg* | *cobA, MET8* | 3 | 160 |
| ER315 | *B. meg* | *cobA, ylnF* | <1 | 140 |

All the strains are derived from strain ER270: *E. coli cysG* deleted strain (302Δa) containing *B. megaterium cbiW-*$H_{60}$*-X-J-C-D-ET-L-F-G-A-*$cysG^{A}$*-cbiY-btuR* genes and the *S. typhimurium cbiP*.

When *E. coli* 302Δa was transformed with plasmid pER270 harbouring the complete *B. megaterium cob* operon and the *S. typhimurium cbiP*, together with a second plasmid carrying the *P. denitrificans cobA*, no cobyric acid synthesis was observed (Table 3-7, strain ER273). This result is in agreement with the result obtained previously with the strain AR8948 (Table 3-6).

Strain ER270 was transformed with the genes encoding for $CysG^{B}$ like activity which do not possess a $NAD^{+}$ binding site and therefore have no

dehydrogenase activity. The *B. megaterium cob* operon together with *cbiK*, *cbiX* or *cysG*$^{G21D}$ produced cobyric acid at a very low level, no more than 3 pmol/$OD_{600}$ (strain ER274 (CbiK), ER276 (CbiX) and ER272 (CysG$^{G21D}$)).

When strain ER270 was transformed with the genes containing a precorrin-2 dehydrogenase activity, the synthesis of cobyric acid increased up to 110 pmol/$OD_{600}$ with *cysG* (ER271), 160 pmol/$OD_{600}$ with *MET8* (ER275) and 140 pmol/$OD_{600}$ with *ylnF* (ER315). YlnF was proposed to be the precorrin-2 dehydrogenase involved in sirohaem biosynthesis (based on sequence similarity only).

These results strongly suggest that the dehydrogenase activity is essential for corrin synthesis via the *B. megaterium* pathway. This supposition was effectively confirmed by strain ER272 (CysG$^{G21D}$), which contains a point mutation in the CysG $NAD^+$ binding site and therefore is deficient in dehydrogenase activity, but is still capable of chelation. The level of cobyric acid synthesis produced by strain ER272 was barely detectable, this result is similar to the results obtained with strains which have a class II CysG$^B$-like proteins (no dehydrogenase activity).

## 3-5- Conclusion

Together with the *P. denitrificans* uro'gen III methyltransferase, CobA, three proteins are able to mimic the CysG activities to complete the synthesis of sirohaem: the *S. typhimurium cbiK* (found within the *S. typhimurium cob* operon), the *B. megaterium cbiX* (found within the *B. megaterium cob* operon) and the *Saccharomyces cerevisiae MET8* (involved in sirohaem synthesis).

These four separate proteins share no sequence similarity but they can clearly be divided into 2 classes of chelatase:

The class I enzyme contains both a dehydrogenase and a chelatase activity. This class is easily identifiable as it contains an $NAD^+$ binding site, and includes CysG and Met8p. Exogenous cobalt has no effect on these enzyme abilities to

perform or finish sirohaem synthesis. Nevertheless, they can also accumulate cobalt sirohydrochlorin *in vivo*.

The class II enzyme is solely a precorrin-2 chelatase and includes CbiK and CbiX. The ability of these proteins to complement a sirohaem deficiency is easily prevented by the presence of exogenous cobalt suggesting that cobalt was able to out-compete ferrous ion at the active site. Cobalt sirohydrochlorin is also accumulated *in vivo* (data not shown for CbiX, R. Beck thesis) probably due to spontaneous oxidation of cobalt precorrin-2.

For cobalamin biosynthesis in *S. typhimurium*, the $CysG^{B}$ domain of CysG is able to function as a cobalt chelatase. In the absence of $CysG^{B}$, CbiK is also able to perform the same function. The fact that *S. typhimurium* has two proteins capable of undertaking the same reaction is consistent with the hypothesis forwarded by Lawrence and Roth, 1996, who proposed that the *S. typhimurium cob* operon had been reacquired after horizontal transfer. The cobalt chelatase function in *S. typhimurium* can also be substituted by the *B. megaterium* CbiX or the *S. cerevisiae* Met8p. Significantly, it was shown that no dehydrogenase is required in the early steps of corrin biosynthesis in *S. typhimurium*. This was effectively demonstrated by $CysG^{G21D}$ which has no dehydrogenase activity, but was able to produce cobyric acid with a *S. typhimurium cob* operon *cbiK* minus in a *cysG* deleted strain.

However, the situation with the *B. megaterium cob* pathway is different. A mutation in *cbiX* within the *B. megaterium cob* operon leads to a massive reduction of corrin synthesis in a presence of an *E. coli* genomic *cysG*, which can be restored by adding the *S. typhimurium cbiK*. CbiK and CbiX appear to be isofunctional. Besides, the synthesis of corrin in an *E. coli cysG* deleted strain from the *B. megaterium cob* operon, containing *cbiX*, requires either the presence of a class I enzyme, a cobalt chelatase also endowed with a dehydrogenase activity ($CysG^{B}$ or Met8p) or a protein only endowed with dehydrogenase activity, such as YlnF. If CysG is substituted by $CysG^{G21D}$, no dehydrogenase activity is present and cobyric acid is not synthesised.

A major difference between the *S. typhimurium* and *B. megaterium* cobalamin pathways is therefore observed: *B. megaterium* requires a dehydrogenation step

in the early stage of cobalamin biosynthesis. Another slight difference between the two pathways, previously described [Raux *et al.* 1996], was that the *B. megaterium cbiD* is unable to complement efficiently a *S. typhimurium cbiD* mutant strain. There may be an association between these differences.

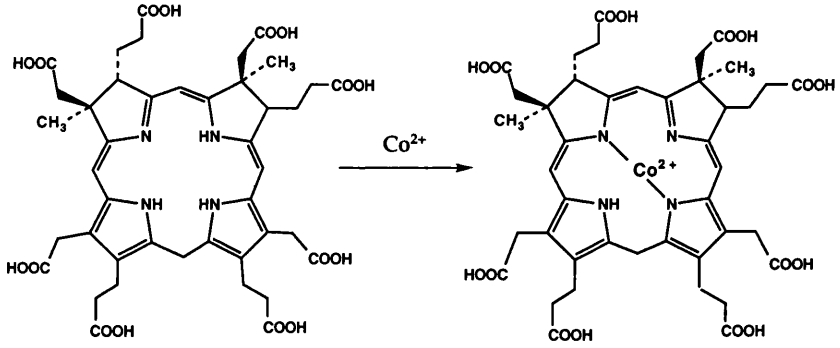
# CHAPTER 4,

## CLONING, EXPRESSION AND X-RAY STRUCTURE OF THE PRECORRIN-2 COBALT CHELATASE (CbiK) OF *S. typhimurium* AND COMPARISON WITH PROTOPORPHYRIN FERROCHELATASE.

## 4-1- Introduction.

In Chapter 2, *cbiK* was identified as the gene with $CysG^{B}$-like activity, able to complete the synthesis of sirohaem from precorrin-2 by chelating ferrous iron. In Chapter 3, the *S. typhimurium cbiK* was shown to be an essential component of the *cobI* pathway in the absence of a genomic *cysG*. When CbiK was overproduced with CobA, bacteria were found to accumulate cobalt-sirohydrochlorin indicating that CbiK acts as a cobalt chelatase. From these observations, it seems likely that the physiological role of CbiK is to insert cobalt into precorrin-2 (Figure 4-1) during the biosynthesis of cobalamin in bacteria such as *S. typhimurium* or *Clostridium acetobutylicum* . This latter micro-organism appears to rely only on *cbiK* since it does not seem to possess a *cysG* to perform this reaction.

Figure 4-1- The probable reaction catalysed by CbiK.

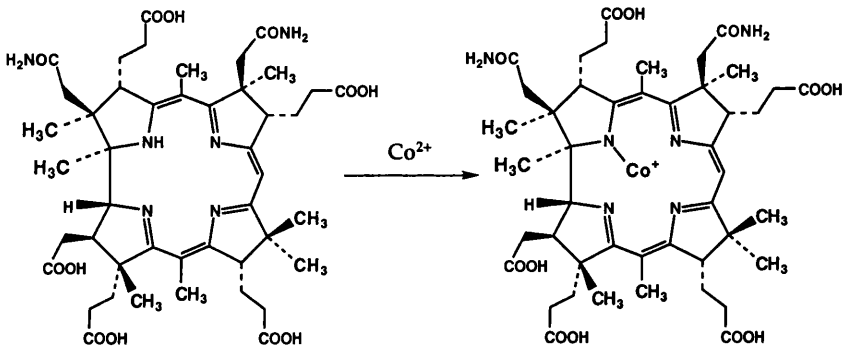



Precorrin-2 Co-precorrin-2

Cobalt insertion for the biosynthesis of cobalamin in *P. denitrificans* occurs after ring contraction, at the level of hydrogenobyrinic a,-c diamide, and requires the concerted action of a three protein-complex, Cob[N-S-T] (Figure 4-2). CobN is the largest subunit of this complex with a molecular mass of 140 kDa and is responsible for substrate binding whereas the exact function of CobS and CobT, the smaller subunits (38 kDa and 80 kDa respectively), has not been characterised. However, it is known that the cobaltochelatase complex is ATP-

dependent and that CobS possesses an ATP binding motif ($Gx_4GKSx_6A$) [Debussche *et al.* 1992].

Figure 4-2- Cobalt insertion in the *P. denitrificans* cobalamin pathway.




Hydrogenobyrinic acid a,c-diamide

Cob(II)yrinic acid a,c-diamide

Both ferrous iron and magnesium protoporphyrin IX chelations occur at the branch point of haem and chlorophyll biosynthesis (Figure 4-3). Indeed, protoporphyrin IX magnesium chelatase (Figure 4-4) is involved in the first specific step of chlorophyll biosynthesis [Walker and Willows, 1997] whereas protoporphyrin IX ferrochelatase (Figure 4-5) is involved in the final step of haem biosynthesis [Ferreira *et al.* 1995; Al-Karadaghi *et al.* 1997]. Another example of ferrochelatase is the sirohaem synthase CysG (Figure 1-14) [Spencer *et al.* 1993]. The enzyme responsible for the insertion of nickel during the biosynthesis of coenzyme $F_{430}$ is still unknown but is thought to occur at the level precorrin-2.

Figure 4-3- The pathways leading to Mg-protoporphyrin IX and protohaem from uro'gen III.

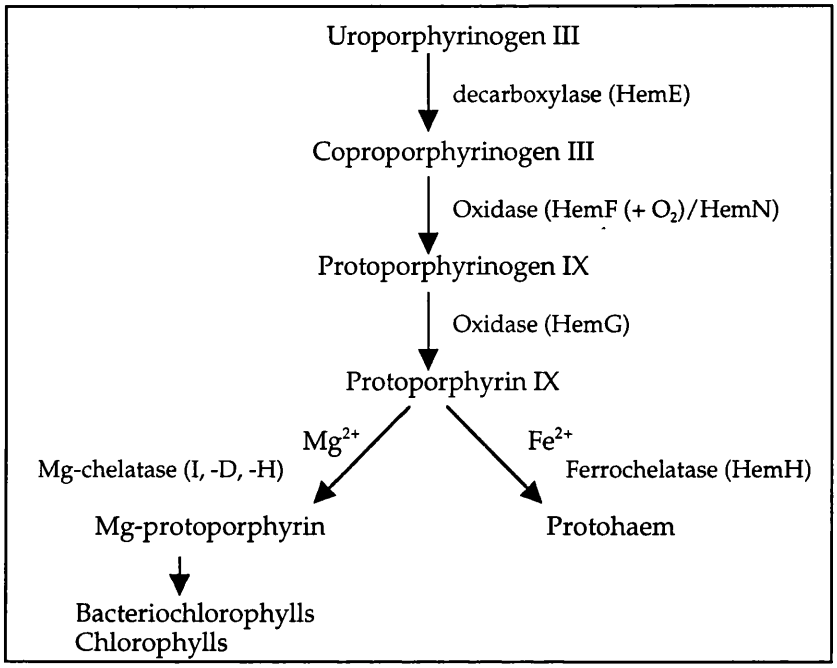
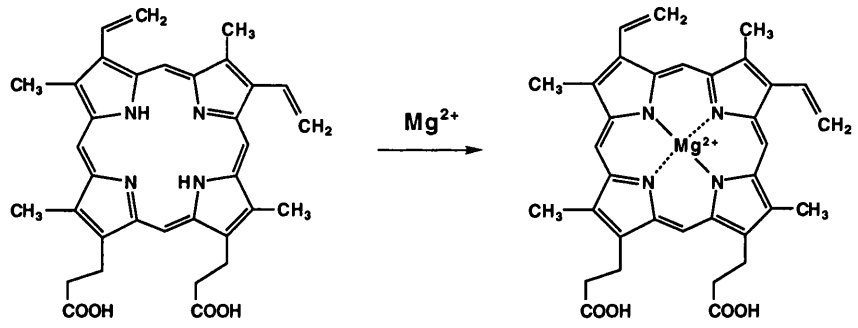



Figure 4-4- The reaction catalysed by protoporphyrin IX magnesium-chelatase.



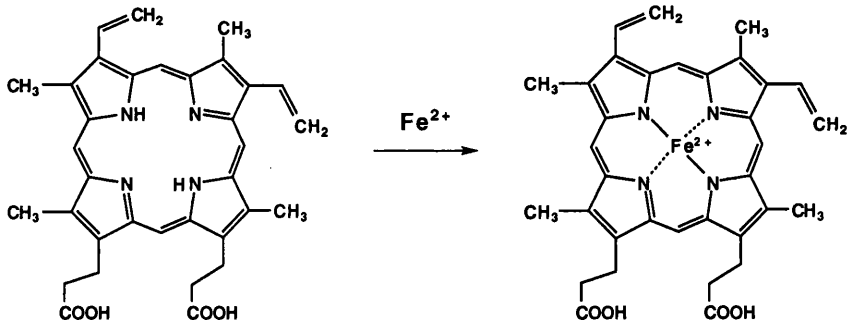

Protoporphyrin IX

Mg-protoporphyrin IX

The magnesium chelatase is the first step unique to the (bacterio)-chlorophyll-specific branch and catalyses the insertion of magnesium into protoporphyrin IX. It also consists of a complex enzyme with three subunits, (B)ChlI (about 40 kDa), (B)ChlD (60 to 87 kDa) and finally (B)ChlH (123 to 153 kDa). The largest subunit, (B)ChlH, shares some similarity with the CobN subunit of the cobaltochelatase complex and the (B)ChlI/CobS have two common motifs, an ATP and metal ion binding sites. No similarity was found between the (B)ChlD and CobT [Walker and Willows, 1997]. A more detailed study of the

magnesium chelatase from a chlorophyll *a*-synthesising bacteria (*Synechocystis* PCC6803) can be found in Jensen *et al.* 1998.

Figure 4-5- The reaction catalysed by protoporphyrin IX ferrochelatase.




Protoporphyrin IX Protohaem IX

Ferrochelatase is the terminal enzyme in haem biosynthesis and catalyses the insertion of ferrous iron into protoporphyrin IX, yielding protohaem. Many protoporphyrin IX ferrochelatases have been sequenced. They have, in general, a molecular mass of 35 to 40 kDa, exist as monomers and are membrane-associated. In a major difference between prokaryotic and most eukaryotic ferrochelatases, the mammalian ferrochelatases share a [2Fe-2S] cluster binding motif. The role of this cluster is not known but is thought to be involved in regulation. Ferrochelatases accept $Fe^{2+}$, $Co^{2+}$ and $Zn^{2+}$ as substrates, whereas $Mn^{2+}$, $Cd^{2+}$ and $Hg^{2+}$ are inhibitors [Ferreira *et al.* 1995]. In contrast to most other ferrochelatases, the *B. subtilis* (p32396) is a water-soluble protein, a property that has facilitated the crystallisation process. The X-ray structure had been subsequently determined at 1.9Å [Al-Karadaghi *et al.* 1997]. The ferrochelation is apparently catalysed by the binding of protoporphyrin to the enzyme in a ruffled conformation enabling insertion of the divalent metal into the distorted core of the macrocycle. No ATP is required for this reaction.

Among all modified tetrapyrroles, there appear to be two classes of metal-ion chelatases: the first type of chelatases does not require ATP and is catalysed by a single protein (ferrochelatases), whereas the second type requires ATP and consists of a multi-subunit complex of three subunits (cobalt and magnesium chelatases).

Since nothing was known about the cobalt chelation process of the anaerobic pathway, it was decided to study this reaction in more detail. For this, an *in vitro* assay has been developed, the three-dimension structure of CbiK has been determined at 2.4 Å and the amino acids proposed to play a direct role in the active site have been altered by site-directed mutagenesis.

The actual process of crystallisation and analysis of the data have been performed by Dr Heidi L. Schubert at the University of York.

## RESULTS

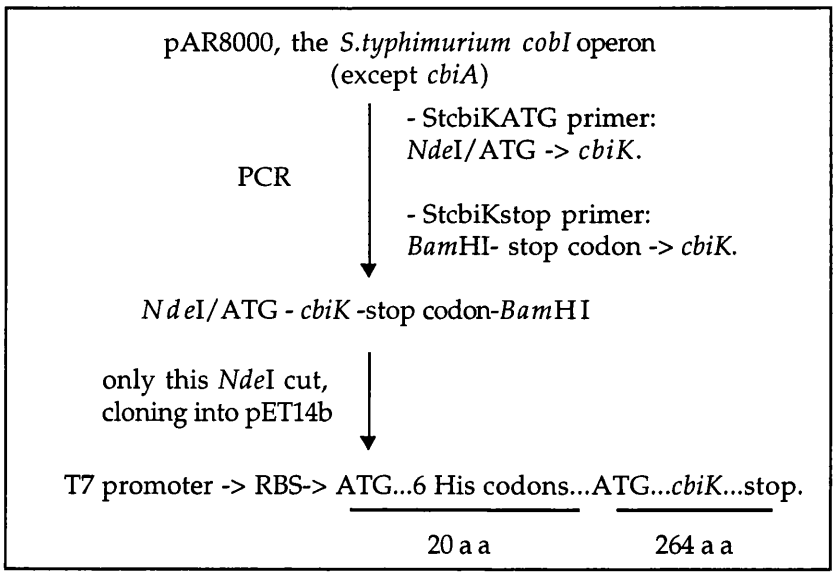
### 4-2- Cloning, expression, purification and characterisation of the *S. typhimurium* CbiK.

*S. typhimurium cbiK* was subcloned from the *cobI-III-II* operon into an expression vector (pET14b), in-frame with a twenty amino acid peptide which contains a stretch of six histidines. The resulting His-tagged recombinant CbiK permits a simple one step purification of the enzyme using metal chelate chromatography.

### 4-2-1- Cloning of *cbiK* in an expression vector.

*S. typhimurium cbiK* was amplified by PCR from plasmid pAR8000, which harbours all the *S. typhimurium cbi* genes except the first ORF, *cbiA*, using the primers StcbiKATG and StcbiKstop. The PCR protocols and primer sequences are indicated in Materials and Methods (Section 8-2-11 and 8-2-12). The PCR product was digested with *Bam*HI and partially with *Nde*I, allowing only the *Nde*I site introduced within the 5' primer to be cut and not the site within the gene. The digested DNA fragment was cloned into the corresponding restriction sites of pET14b (Figure 4-6), generating pER231. This plasmid was transformed into JM101 to maintain the plasmid (ER231) and into BL21(DE3)(pLysS) for expression studies (ER227).

Figure 4-6- The cloning procedure of *cbiK* into pET14b.




## 4-2-2- Expression and purification of *S. typhimurium* CbiK.

In strain ER227, the T7 RNA polymerase is under the control of a *lac* promoter (λDE3 lysogen) and is therefore induced by the action of IPTG. In the absence of IPTG, the production of T7 RNA polymerase is completely inhibited by the T7 lysosyme (pLysS). The induction of RNA polymerase involves the transcription of the His-tagged *cbiK* [pET system manual/Novagen].

The His-tagged CbiK was expressed and purified as described in Materials and Methods. The cell pellet from one litre of ER227 culture was resuspended in binding buffer and sonicated. The soluble fraction was loaded onto a 5 ml His-bind column previously charged with 50 mM $Co(NO_3)_2$. The column was washed with 10 volumes of binding buffer, 6 volumes of wash buffer (containing 100 mM imidazole) and then the proteins were subsequently eluted with elute buffer (containing 400 mM imidazole). Proteins were detected by the BioRad protein assay. Fractions containing the His-tagged CbiK were pooled and desalted by passing through a PD10 column, which had been previously equilibrated in 50mM Tris/HCl pH7.8. From one litre of culture, approximately 25 mg of pure His-tagged CbiK were reproducibly obtained. The purified protein was stored at -20°C and appears active even after several months.

When analysed by SDS/PAGE the purified protein ran as a single band with a molecular mass of about 31 kDa [Figure 4-7], which is in agreement with the expected molecular mass (29 kDa plus 2.2 kD corresponding to the His-tag). The His-tagged protein was sometimes found to be contaminated with the chloramphenicol acetyltransferase peptide (23 kDa) which also binds to the metal chelate column. The native molecular mass of CbiK was estimated from gel filtration studies using a Bioselect gel filtration column. The His-tagged CbiK eluted with a molecular mass of 83 kDa (+/- 10 kDa) suggesting that the protein exists as a homotrimer [Dr S. Woodcock, pers. commun.]. A native gel of the His-tagged CbiK is shown in Figure 5-6.

The His-tag could be removed from the recombinant CbiK by incubation with thrombin in 20 mM sodium citrate, pH6.5 and 100 mM NaCl, 70 mM Tris, pH8.5 and 2.5mM $CaCl_2$ at room temperature. The cleaved peptide was separated from CbiK by gel filtration [Dr H. Schubert, pers. comm.].

### 4-2-3- Functional characterisation of CbiK *in vitro*.

In order to characterise the function of CbiK *in vitro*, precorrin-2 has to be generated. However, since precorrin-2 is particularly unstable, the substrate has to be freshly prepared and used immediately. Unfortunately, the two preceding intermediates, uroporphyrinogen III and preuroporphyrinogen are also unstable. Therefore, a multienzymatic incubation was developed to generate precorrin-2 from the stable precursor porphobilinogen. This requires the action of three extra enzymes: PBG deaminase (HemC), uro'gen III synthase (HemD) and uro'gen III methyltransferase ($CobA^{Pd}$).

PBG deaminase and uro'gen III synthase were available in the lab and had been purified from overproducing recombinant *E. coli* strains by Dr S. Awan. As a source of uroporphyrinogen III methyltransferase, the *P. denitrificans cobA* [Blanche *et al.* 1989] was cloned in an expression vector such that it was expressed as a His-tagged recombinant enzyme to facilitate a single-step purification of the enzyme.

Precorrin-2 was thus produced by adding porphobilinogen and SAM to a mixture of the three enzymes. During the first 10-15 min at 37°C, the incubation developed a yellow colour, usually observed with the presence of precorrin-2, and which gave a UV/visible spectrum consistent with a dipyrrocorphin. However, during the next 15-30 minutes, the incubation mixture turned pink due to the oxidation of the labile product.

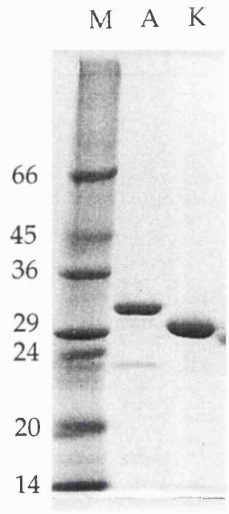
Incubation of CbiK with precorrin-2 and cobalt can be used as a qualitative assay, but it is not suitable for accurate kinetics studies.

a- Cloning, expression and purification of the *P. denitrificans* CobA.

The strategy used for the expression of *cbiK* was also applied to clone and express the *P. denitrificans cobA*. Two primers, PdcobA-ATG and PdcobA-stop, were used to amplify the gene by PCR (Materials and Methods, section 8-2-12), using the plasmid pCR395 (a gift from Prof. I. A. Scott) as DNA template. The amplified *cobA* was cloned into pET14b in-frame with the poly-histidine encoding sequence and transformed into BL21(DE3)(pLysS) to allow expression. The resultant strain was named ER243. The His-tagged CobA was purified using a similar procedure to that described for CbiK except that the wash buffer contained only 60 mM imidazole. About 20 mg of CobA were obtained from a 1 litre culture of ER243 (Figure 4-7). The protein was finally freeze-dried in 50 mM Tris-HCl at pH7.8 and stored at –20°C.

INSTITUTE OF OPHTHALMOLOGY
LIBRARY
BATH STREET EC1V 9EL

Figure 4-7- SDS/PAGE of the His-tagged CobA and CbiK.




Lane M represents standard protein molecular mass markers, whose sizes are indicated on the left side of the gel. Lane A represents the purified *P. denitrificans* CobA, and lane K the purified *S. typhimurium* CbiK. The His-tagged CobA and His-tagged CbiK have the same molecular mass of 31.4 kD. The difference of size observed on the gel presumably reflects some peculiarity in structure of the unfolded proteins.

b- Multienzymatic assay of CbiK.

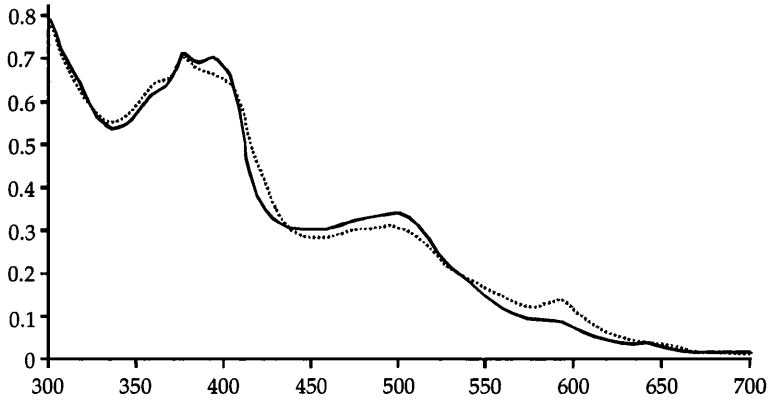
Due to the instability of the different modified tetrapyrrole intermediates, the CbiK assays were directly performed in Quartz cuvettes. In an attempt to keep the reaction anaerobic, the 50 mM Tris-HCl, pH 7.8 buffer was degassed for 20 minutes under vacuum and was used to resuspend all reagents and enzymes. The assays were prepared as indicated in Table 4-1 and incubated at either room temperature or 37°C. Spectra were recorded periodically (every 5 to 15 minutes) over a range of 300 to 700 nm. The spectra shown in Figure 4-8 were obtained after an incubation of 60 minutes at room temperature.

Table 4-1- *cbiK* assay *in vitro*.

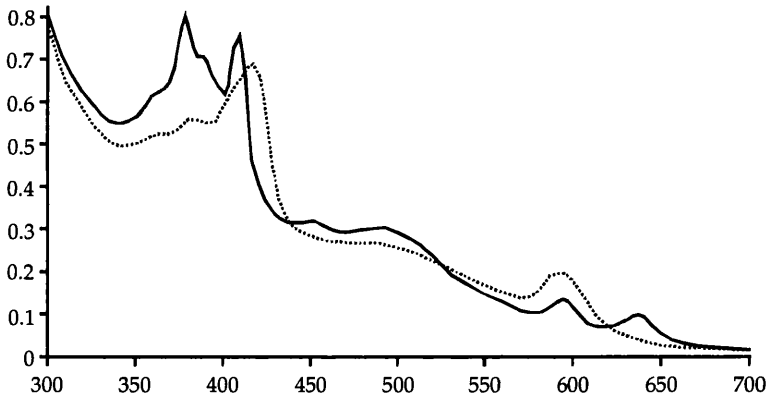
| **Reagent/Enzymes** | **CobA** | **CobA +cobalt** | **CobA & CbiK** | **CobA & CbiK + Cobalt** |
|---|---|---|---|---|
| PBG deaminase | 0.1 mg | 0.1 mg | 0.1 mg | 0.1 mg |
| UroIII cosynthase | 0.15 mg | 0.15 mg | 0.15 mg | 0.15 mg |
| $CobA^{Pd}$ | 0.5 mg | 0.5 mg | 0.5 mg | 0.5 mg |
| $CbiK^{St}$ | - | - | 0.5 mg | 0.5 mg |
| SAM | 0.75 mg | 0.75 mg | 0.75 mg | 0.75 mg |
| $CoCl_2$, 6 $H_2O$ | - | 0.1 mg | - | 0.1 mg |
| PBG | 0.1 mg | 0.1 mg | 0.1 mg | 0.1 mg |
| 50 mM Tris-HCl, pH7.8 | up to 1 ml | up to 1 ml | up to 1 ml | up to 1 ml |

Figure 4-8- Spectra obtained after the multienzymatic assays.

a- CobA +/- cobalt




b- CobA and CbiK +/- cobalt




The plain line represents spectra obtained in absence of cobalt whilst the dotted line represents spectra obtained in the presence of exogenous cobalt.

The two spectra obtained from the multienzymatic assay in the absence of CbiK (Spectra a) are similar to a spectrum of sirohydrochlorin (λ max at 378 nm) [Warren *et al.* 1994]. The addition of cobalt does not cause a significant alteration in the spectrum. Under these conditions, and although the solution turns pink after 15 minutes, the spectra do not provide evidence for the appearance of trimethylpyrrocorphin accumulation, which was observed *in vivo*.

The addition of CbiK to the mixture (spectra b) induces drastic changes in the spectrum and a complete shift of the maxima in the presence of cobalt. The spectrum obtained in the absence of cobalt has two maxima, at 378 nm

presumably due to the presence of sirohydrochlorin [Warren *et al.* 1994], and at 408 nm which could indicate the formation of a lactone form of sirohydrochlorin. The fact that the spectrum with CbiK in the absence of cobalt is more defined than the spectrum obtained without CbiK may also represent the stabilisation of one tautomeric form of sirohydrochlorin by CbiK. In the presence of CbiK and of an excess of cobalt, a clear shift in the spectrum is observed that results in the appearance of two major peaks at 416 and 590 nm, consistent with the presence of cobalt-sirohydrochlorin [Battersby and Sheng, 1982].

This experiment confirms, *in vitro* , the results obtained *in vivo* in Chapter 3, and indicates that CbiK is able to mediate the transformation of precorrin-2 into cobalt-sirohydrochlorin. However, it is more likely that CbiK catalysed the synthesis of cobalt-precorrin-2 and that spontaneous oxidation of this intermediate results in the production of cobalt-sirohydrochlorin. It is thus important to verify the synthesis of cobalt-precorrin-2 using some biophysical technique such as NMR.

## 4-3- X-ray structure of CbiK.

The crystallisation and X-ray structure determination of CbiK has been performed by Dr Heidi Schubert at the University of York. A summary of this work is described in sections 4-3-1- and 4-3-2.

### 4-3-1- Crystallisation data.

Both His-tagged CbiK and the cleaved-off CbiK proteins were concentrated to 7 mg/ml. Crystals could be grown from different solutions (Figure 4-9). The crystals obtained from a solution containing 10 to 15% polyethylene glycol (molecular mass 4 kDa), 0.1 M Tris pH8.5 and 0.2 M $Li_2SO_4$ induced the best diffraction. The structure was determined at 2.4 Å resolution, the crystals are space group $P6_322$ with cell dimension of a=b= 128.08Å, c= 85.44Å. No density was visible for the His-tagged residues as well as for the first and the last six amino acids of CbiK.

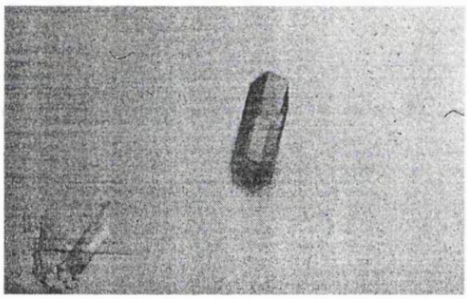
The crystals were soaked for one hour prior to data collection with 1 mM of

MeHgCl and (Me)3PbOAc which created two single site heavy atom derivative data sets.

Figure 4-9- Crystals of CbiK.

a/with DTT

b/with PEG

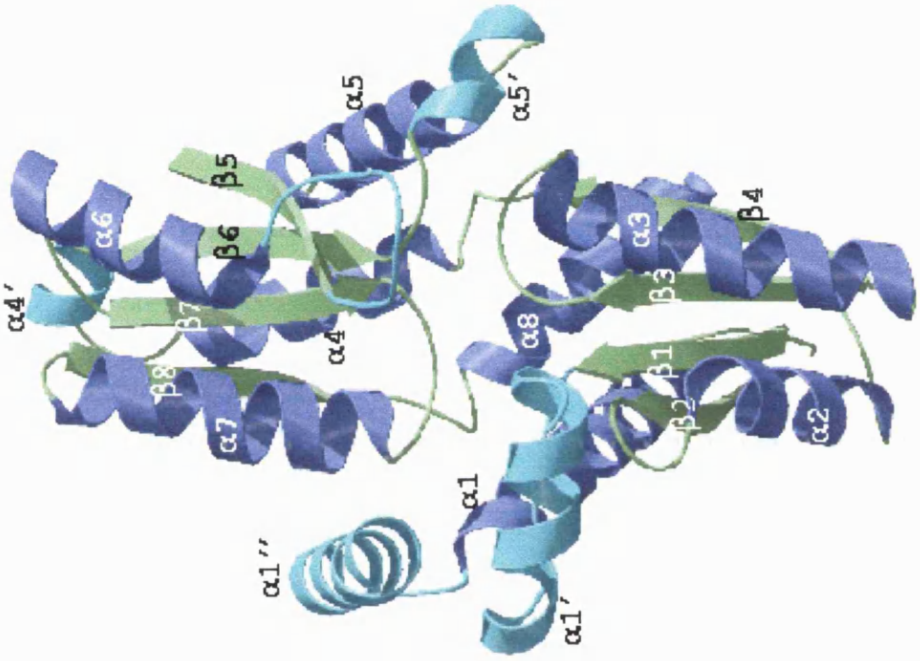


## 4-3-2- Structure of CbiK.

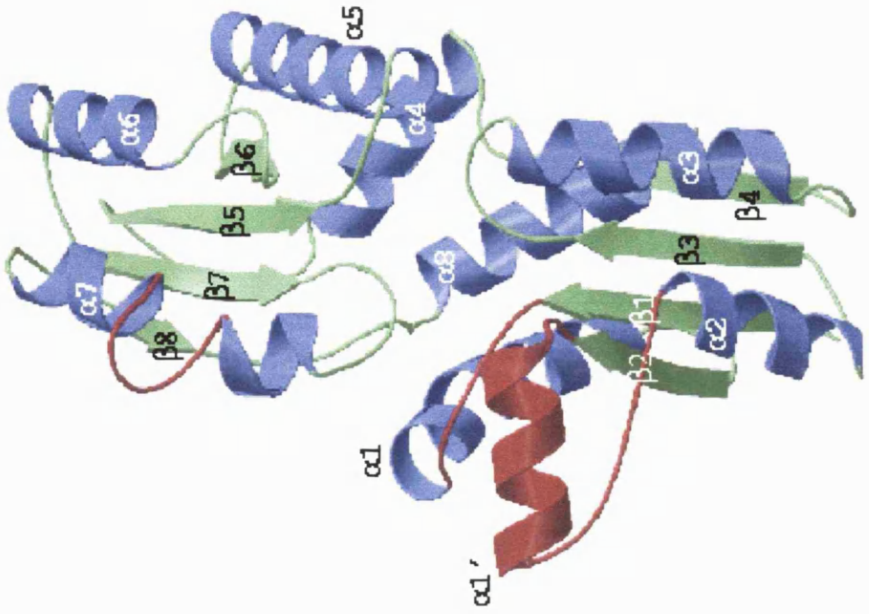
CbiK is a bilobal enzyme of the periplasmic-type protein fold family, which have two similar α/ β domains linked by a hinge region (Figure 4-10).

The two major domains both contain a four stranded parallel β-sheet surrounded by at least four helices. The N-terminal domain contains helices α1, α2, α3 and α8 (and α1′) and the C-terminal domain α4, α5, α6 and α7. The similarity of the two domains, except helix α1′ and the position of helix α1, suggests that an evolutionary gene duplication event has occurred. From the gel filtration studies, it was deduced that CbiK was a trimer, however, the crystallographic results suggest an equilibrium between a monomer and a trimer.

Figure 4-10- Three-dimensional representation of CbiK and HemH.




HemH, protoporphyrin IX ferrochelatase
(*Bacillus subtilis*)



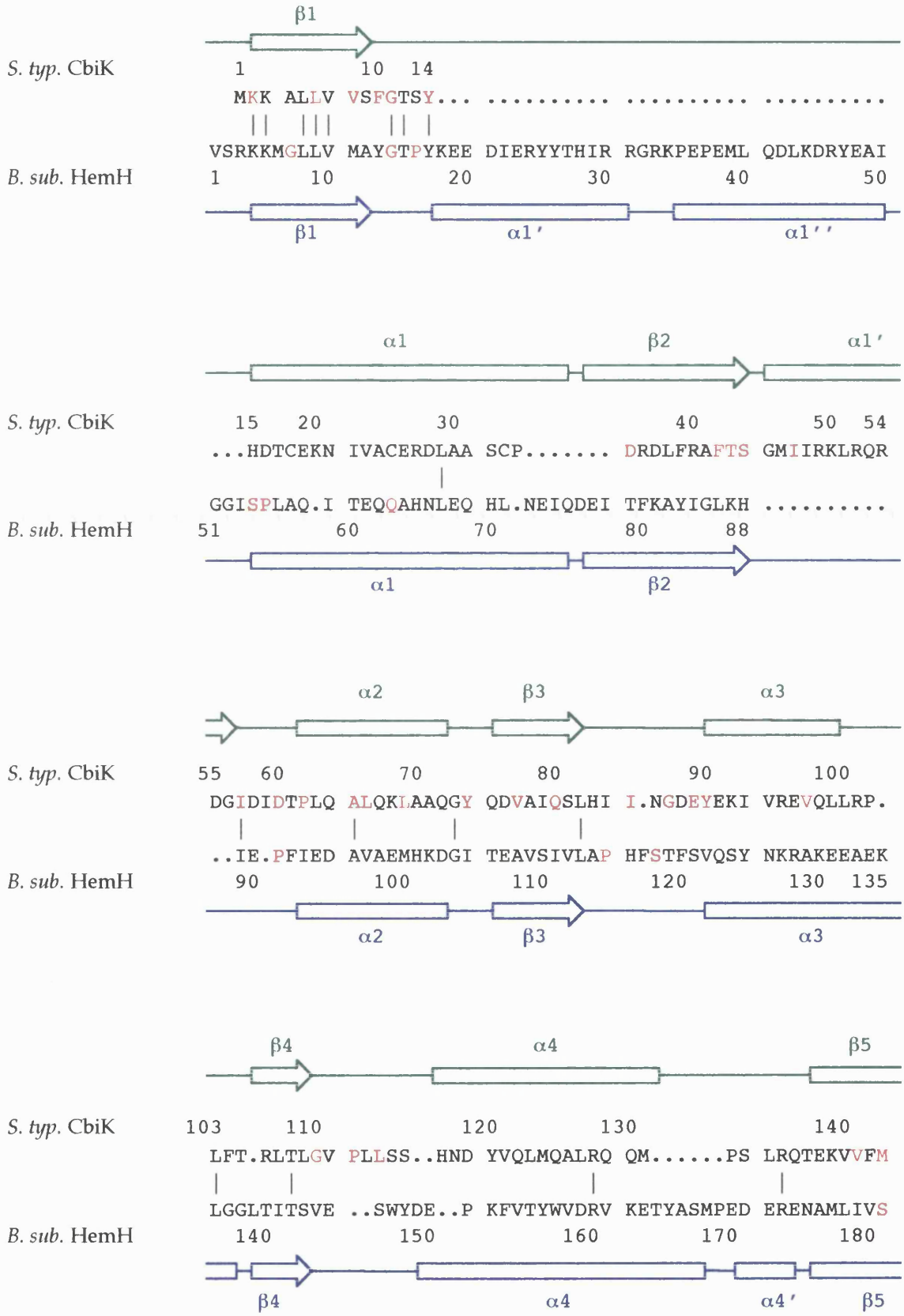

CbiK, precorrin-2 cobalt chelatase
(*Salmonella typhimurium*)

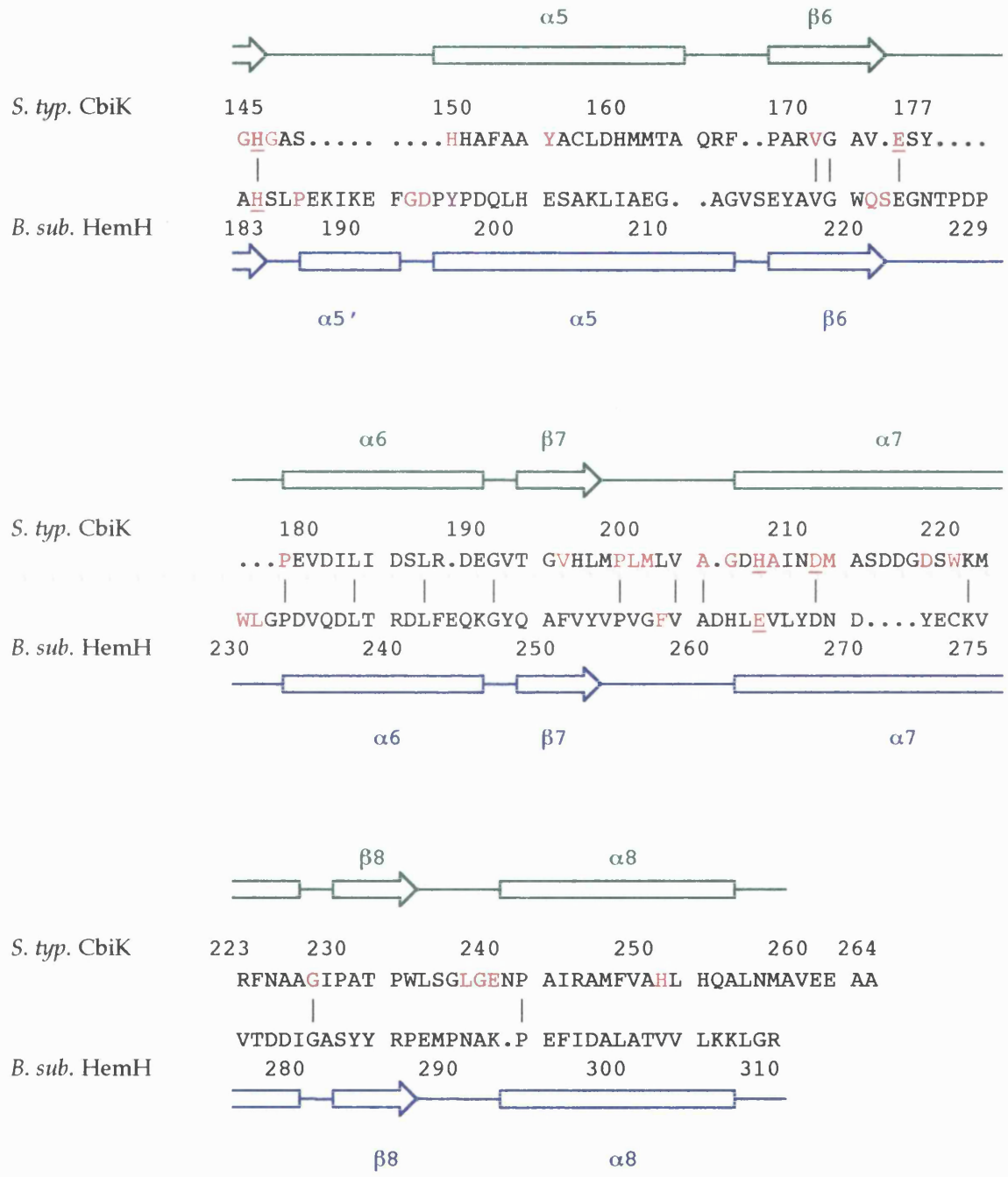
## 4-4- Comparison of CbiK with the *B. subtilis* ferrochelatase.

Although no significant sequence similarity can be noticed at the primary level between CbiK and the ferrochelatase HemH, the structure resolution of both proteins reveals common features (Figure 4-10) and the superimposition of both enzymes presents a *rms* deviation of 3.1Å. In HemH, α1′ and α1″ helices constitute a small subdomain, which conceals the active site. These two helices are found in all ferrochelatases with variable length and composition, but are absent in CbiK which might therefore have a more accessible active site. CbiK and HemH share thirty-two common amino acids of which only 14 are conserved in the 4 CbiK sequences (Figure 4-12).

The active site of CbiK lies in a deep rectangular cleft between the two symmetrical domains. In ferrochelatase, the ferrous iron is thought to bind via His 183 at the end of β-5, and Glu 264 on α-7. By comparison with the CbiK active site, a histidine (His 145) is found in the same configuration as His 183 whereas no glutamate residues are present in the α-7 corresponding helix in CbiK. However, the position of Glu 264 is substituted by another histidine in CbiK (His 207). This suggests that His 145 could be the final cobalt binding site and that His 207 could play a role in the ion selectivity as well as in cobalt binding site. Two other residues, Glu 175 and Asp 211, are thought to be involved indirectly in the metal chelation process by being negatively charged in the active site. Asp211 is also linked by a hydrogen bond to His 207. Finally, Glu 89 is well positioned to play a crucial catalytic role as a general acid due to its position within the active site (Figure 4-13).

Figure 4-11- Alignment of HemH and CbiK in terms of primary and secondary structure.



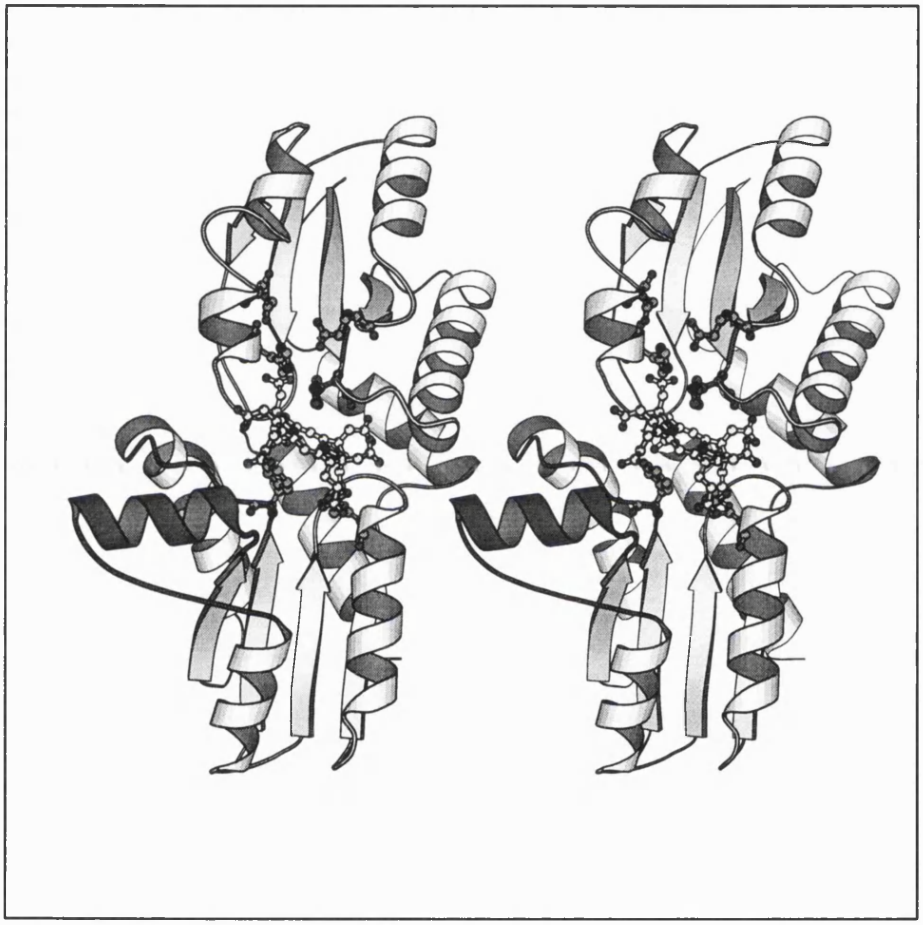



The amino acids coloured in red indicate the conserved residues within ferrochelatase or CbiK sequences. The underlined residues are thought to be involved in metal binding.

Figure 4-12- Alignment of all known CbiKs.

```
                1                                                   50
C.aceto. CbiK                                        MKKAILI ISFGTTYENT
C.aceto. CbiK2                                     MLLLKKGILI VGFGTIYKDA
S. typ. CbiK                                         MKKALLV VSFGTSYHDT
P. ging. CbiK          MKKL ILATLGLMAI AMLSCSSNNK DLENKGEATL LVTFGSSYKAP
Consensus              ----- ---------- ---------- ----K---L- V-FG--Y---

                1                                                   50
C.aceto. CbiK    RKLTIDAIEE KVKDEFKDYE IRRAFTSYKI IKVLKERDKI IVDTPGMALS
C.aceto. CbiK2   RACTIDKIKE EVCKQFPDYE VLQAFTSSKI IEKIKETEGI DFDTPQKALH
S. typ. CbiK     CEKNIVACER DLAASCPDRD LFRAFTSGMI IRKLRQRDGI DIDTPLQALQ
P. ging. CbiK    RE.TYAKIEK TFAAAYPDQR ISWTYTSSII RKKLAQ.QGI YIDAPDEALE
Consensus        ---------- -------D-- ----FTS--I ---------I --D-P--AL-

                51                                                  100
C.aceto. CbiK    RLKREGFEKV IVQPLHIIPG EEYDFVKRVV NEYSYDFKEI ..KLGRPVLY
C.aceto. CbiK2   KFKNKAYREV IIQPIYMIPG YEYEKLIGTV DKYLKDFERL ..KIGRPVLY
S. typ. CbiK     KLAAQGYQDV AIQSLHIING DEYEKIVREV QLLRPLFTRL T..LGVPLL.
P. ging. CbiK    KLARLGYKKI NVQSLHVIPG REYDEMIDFV NKFKAAHSDI TVKVGAPLF.
Consensus        -L----Y--V --Q----I-G -EY------V --Y------- ----G-P-L-

                101                                                 150
C.aceto. CbiK    YKGITGEFPD DYEIMVNSIK SIIPRDEIVI LMGHGSTHSS NACYSCLQLV
C.aceto. CbiK2   FKGNNKNEPN DYEIFLEKIK DMIPKDKITV FMCHGTMHES NEYYLKLDSF
S. typ. CbiK     ..SSHNDYVQ LMQALRQQM. PSLRQTEKVV FMGHGASHHA FAAYACLDHM
P. ging. CbiK    ..DTDEDMRE VAEILHKRFQ QTIEKGEAIV FMGHGTEHAA NDRYARINKI
Consensus        ---------- ----L----- ---------V FM-HG--H-- ---Y------

                151                                                 200
C.aceto. CbiK    LRDNDFGNVF VANVEGYPDL NNVLRYIENH ETR.....KV TLIPLMLVAG
C.aceto. CbiK2   IKKHGFTSVV IATLEGFPKV EDVIEYMEKN KFKEKYKGEV KIIPLLVTAG
S. typ. CbiK     MTAQRFP.AR VGAVESYPEV DILI.....D SLRDEGVTGV HLMPLMLVAG
P. ging. CbiK    MKNYSKF.MI VGTVESDPSI NDVI.....A ELKETGATAV TMMPLMSVAG
Consensus        ---------- V---E--P-- ---------- ---------V ---PLM--AG

                201                                                 250
C.aceto. CbiK    NHAVIDMAGN SEDSWKNILT SNGFEV.... ..TAYVHGLG EVKEFQDIYI
C.aceto. CbiK2   KHVKKDM... .QEEWVNIFK DKGYKV.... ..ETYEHGLG EIYDFRSIYI
S. typ. CbiK     DHAINDMASD DGDSWKMRFN AAGIP..... .ATPWLSGLG ENPAIRAMFV
P. ging. CbiK    DHATNDMAGD EDDSWKTLLT NAGYTVSIDK LDNGNFSALG DIEEIRNIWL
Consensus        -H---DM--- --D-W----- --G------- --------LG E-------Y-

                251
C.aceto. CbiK    EHIKDVMEGR YEKEVKEEIK CPEL
C.aceto. CbiK2   QHIKDAI... ....IRENYT AL..
S. typ. CbiK     AHLHQALNMA VEEAA..... ....
P. ging. CbiK    KHMK ATSAR .......... ....
Consensus        -H-------- ---------- ----
```

Figure 4-13- Stereo representation of the active site of CbiK.



The amino acids Phe 10,Glu 89, His 145, Glu 175, Asp 211, His 207 and precorrin-2 are displayed on the ribbon representation of CbiK as follows: from left to right, and top to bottom, Asp 211, Glu 175, His 207, His 145, precorrin-2, Phe 10 and Glu 89.

## 4-5- Site-directed mutagenesis of five CbiK active site amino acids.

By comparison of the three-dimensional structures of CbiK with *B. subtilis* ferrochelatase, five residues in CbiK have been proposed to act as key amino acids in the active site (section 4-4). These amino acids are histidine H145, histidine H207, glutamate E89, glutamate E175 and aspartate D211 (Figures 4-11 and 4-13). To confirm this hypothesis, all these residues were modified to alanine by site-directed mutagenesis, a double mutant H145A/H207A was also constructed. The effects of these alterations on the activity of CbiK were analysed in three ways. First, the complementation of the cysteine auxotrophy of *E. coli* 302Δa by the mutated CbiKs was observed to give an insight into the enzyme's ability to ferrochelate. Secondly, the ability of the modified CbiKs to restore cobyric acid biosynthesis in a cobalt chelatase deficient *E. coli* recombinant strain was measured. Finally, the UV-visible spectra of products that accumulated *in vivo* from strains overexpressing the mutated CbiKs and CobA were also recorded.

### 4-5-1- Site-directed mutagenesis.

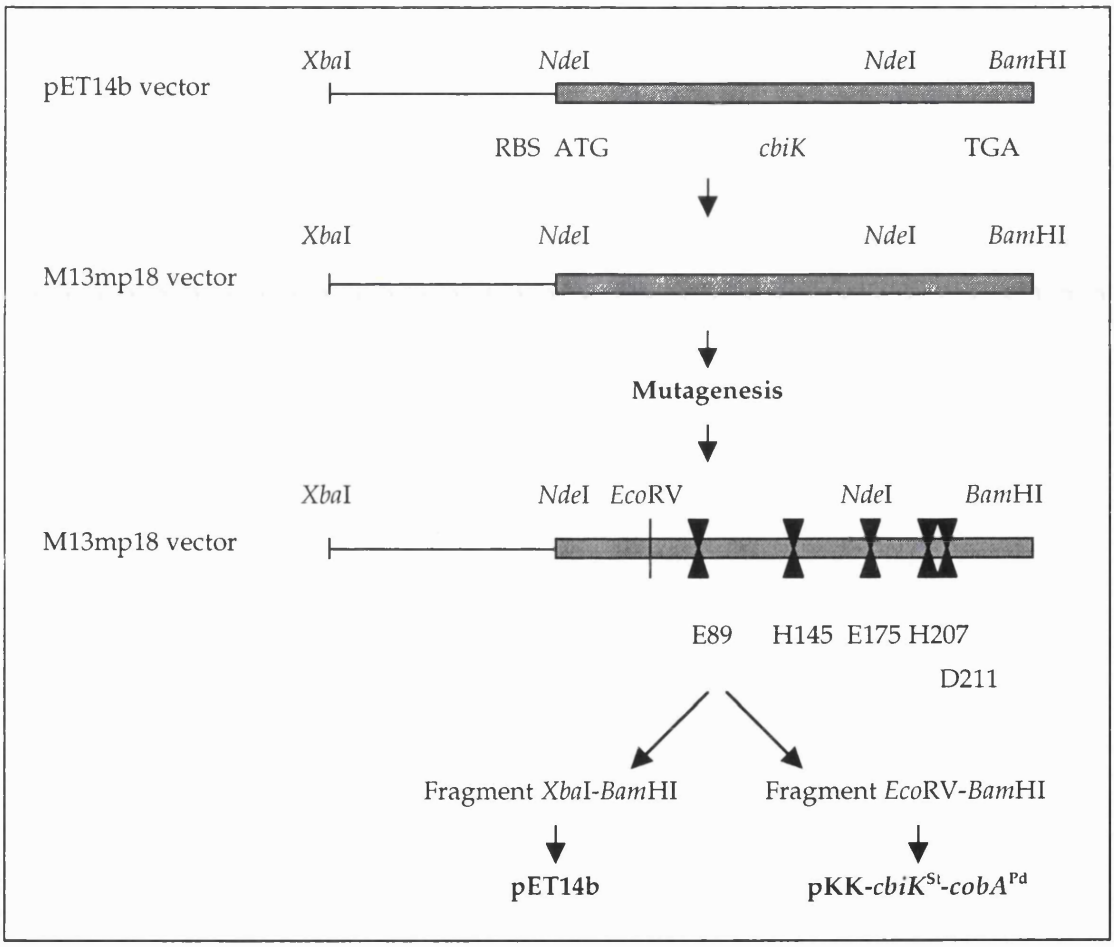
The necessity of single stranded DNA for the mutagenesis reactions required *cbiK* to be cloned into M13 and subsequently, after the codon substitutions, to be subcloned into two different expression vectors: (1) into pER119 (pKK harbouring the *P. denitrificans cobA*) to permit *cysG* complementation *in vivo*; (2) into pET14b for future *in vitro* studies.

#### a- Cloning into M13mp18.

The cloning of *cbiK* into pET14b, described in section 4-2-1, was technically demanding since it required a partial *Nde*I digest. Therefore, to facilitate cloning of mutated *cbiKs* into pET14b, the *Xba*I-*Bam*HI DNA fragment from pER231 (pET14b-*cbiK*) was isolated and cloned into M13mp18. This DNA fragment contains the ribosome binding site followed by the His-tagged encoding DNA of the pET14b with *cbiK* cloned in frame. The resultant M13 construct was named pER304.

After mutagenesis, a *Xba*I-*Bam*HI digest allowed the RBS-His tagged mutated *cbiK* to be subcloned easily into pET14b. Moreover, by exchange of the *Eco*RV-*Bam*HI fragment between the M13 recombinant DNA and a pKK-*cbiK-cobA* vector (pER303, section 4-5-2/a), the mutated *cbiKs* could rapidly be introduced into this second expression vector. The cloning strategy is outlined in Figure 4-14.

Figure 4-14- Cloning strategy of *cbiK* mutants.




b- Site-directed mutagenesis.

The site-directed mutagenesis was performed using the Sculptor™ *in vitro* mutagenesis kit RPN1526 (Amersham) according to the manufacturer's intructions. An outline of the mutagenesis is given in Materials and Methods, section 8-2-13. The single-stranded DNA used for the mutagenesis was obtained

from pER304, and from pER304$^{H145A}$ for the mutagenesis of the double mutant. The primers used to generate the codon substitutions are described in Figure 4-15.

Figure 4-15- Primers used for mutagenesis of *cbiK*.

| primer | primer sequence | codon change |
|---|---|---|
| E89A | GCACATTATTAACGGCGACGCATATGAAAAAATTGTCCGTG | GAA -> GCA |
| H145A | GTCGTTTTTATGGGCGCCGGCGCCAGCCATCAC | CAC -> GCC |
| E175A | CGGCGCCGTAGCAAGCTACCCGGAGG | GAA -> GCA |
| H207A | GTGGCGGGCGATGCAGCCATTAATGATATGGC | CAC -> GCA |
| D211A | CACGCCATTAATGCTATGGCTTCAGACG | GAT -> GCT |

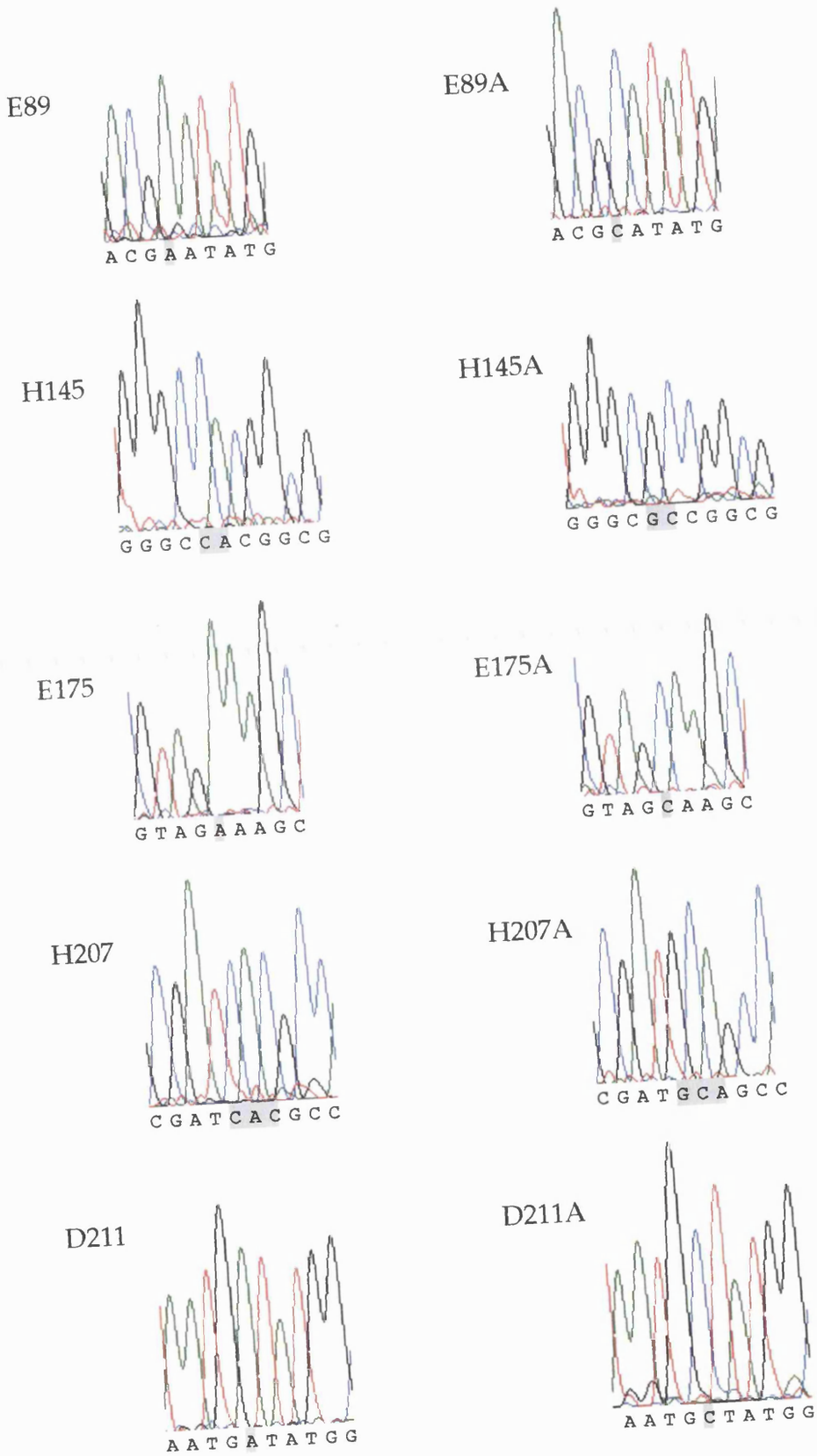
The modified bases are indicated with a grey box, and the codon changes are underlined.

c- Sequencing.

In order to verify that the mutations had been introduced, the single-stranded recombinant DNAs were purified from the plaques obtained after transformation. The sequencing was performed following the instructions given by the ABI PRISM® BigDye ™ terminator cycle sequencing ready reaction kit with AmpliTaq®DNA polymerase, FS, using a 373A Applied Biosystems DNA sequencer.

The five single mutations were all confirmed by sequencing as was the double mutation H145A/H207A. The electropherogram of each mutations is shown in Figure 4-16.

Figure 4-16- Electropherogram of the five *cbiK* mutations.



## 4-5-2- Activity of mutated CbiK variants *in vivo.*

As mentioned earlier in section 4-5-1-a, in order to analyse the effect of the different mutations introduced into *cbiK*, the mutated genes had to be cloned into a suitable expression vector. However, to analyse the effect of the different *cbiKs in vivo*, it was essential to clone all the *cbiK* variants such that they were co-expressed with the uro'gen III methyltransferase, $cobA^{Pd}$. When cloned in this manner, the resultant *cobA-cbiK* plasmids can be studied for sirohaem complementation in strain ER171 (*E. coli cysG* deleted strain carrying pACYC-$lacI^q$) and for corrin biosynthesis in strain ER185 (*E. coli cysG* deleted strain carrying all the *S. typhimurium cobI* operon with a deletion in *cbiK*).

### a- Construction of a vector to analyse the *cbiK* mutants.

The *cobA-cbiK* plasmid (pER170) used in section 3-3-1 does not contain the appropriate restriction sites for subcloning the mutant-M13-*cbiK* constructs. Therefore, a new plasmid harbouring *cobA* and *cbiK* was generated. This was achieved by amplification of *cbiK* with a 5′ primer designed to include a modified ribosome binding site (StcbiK-RBS primer) and a 3′ primer introducing a *Bam*HI site after the stop codon (StcbiK-stop primer, Materials and Methods, section 8-2-12). The modified *cbiK* was cloned into pGEM-T easy and subsequently subcloned into the *Eco*RI/*Bam*HI sites of pER119 (pKK harbouring the *P. denitrificans cobA*). The resulting plasmid pKK-*cbiK-cobA* was named pER303.

### b- Subcloning of *cbiK* mutants into pER303.

The codons that were subject to site-directed mutagenesis are all contained within the 510 bp *Eco*RV/*Bam*HI fragment of pER303. This fragment was removed from pER303 and replaced with the corresponding *Eco*RV/*Bam*HI fragment from the M13-*cbiK* mutant constructs. The recombinant plasmids were called $pER303^{E89A}$, $pER303^{H145A}$, $pER303^{E175A}$, $pER303^{H207A}$, $pER303^{D211A}$ and $pER303^{H145A/H207A}$.

c- Sirohaem complementation.

Strain ER171, *E. coli cysG* deleted strain carrying pACYC-*lacI*$^{q}$, was transformed with pER119 (pKK harbouring the *P. denitrificans cobA* only), with pER303 and the six pER303 derived plasmids respectively. The resulting strains were re-streaked on agar minimal medium plates, minimal medium plates containing cobalt chloride and minimal medium plates containing cysteine. Cobalt and cysteine were added as indicated in section 3-3-2. Table 4-2 shows the results obtained after 40 hours at 37°C.

Table 4-2- Complementation of *E. coli cysG* deleted strain for cysteine auxotrophy.

| Strain | plasmid pKK derived constructs | Growth on minimal medium after 40 hrs at 37°C | | |
|---|---|---|---|---|
| | | no addition | $CoCl_2$ | cysteine |
| ER316 | *cobA* | - | - | +++ |
| ER317 | *cobA-cbiK* | ++ | - | +++ |
| ER318 | *cobA-cbiK*$^{E89A}$ | +++ | - | +++ |
| ER319 | *cobA-cbiK*$^{H145A}$ | + | +/- | +++ |
| ER320 | *cobA-cbiK*$^{E175A}$ | +++ | - | +++ |
| ER321 | *cobA-cbiK*$^{H207A}$ | +++ | ++ | +++ |
| ER322 | *cobA-cbiK*$^{D211A}$ | ++ | - | +++ |
| ER323 | *cobA-cbiK*$^{H145A/H207A}$ | - | - | +++ |

The *E. coli* host strain is *E. coli* 302Δa containing pAR8086 (pACYC184 harbouring *lacI*$^{q}$). The symbol, -, means no growth on minimal medium plates, +, very slow growth and, +++, normal growth.

The mutants harbouring *cbiK*$^{E89A}$, *cbiK*$^{D211A}$ and *cbiK*$^{E175A}$ have the same phenotype as the wild type, since they complement the *cysG* minus strain only in the absence of exogenous cobalt.

Conversely the histidine mutants have very interesting phenotypes: although *cbiK*$^{H207A}$ efficiently complemented the *cysG* deletion strain, *cbiK*$^{H145A}$ was barely able to overcome the cysteine auxotrophy indicating that this enzyme had a greatly reduced activity. However, quite surprisingly both these mutants were able to complement in the presence of exogenous cobalt, suggesting that the two mutant proteins had altered cobalt affinities.

Finally, the double histidine mutation *cbiK*$^{H145A/H207A}$ was unable to restore growth to the *cysG* deletion strain. The double mutant protein is therefore inactive as ferrochelatase in sirohaem biosynthesis. It is likely that the removal of both histidine residues alters the metal-ion binding to such an extent that it is unable to bind. However, this does not preclude the possibility that the double mutation has caused some large scale change in the structure of the enzyme, rendering it inactive.

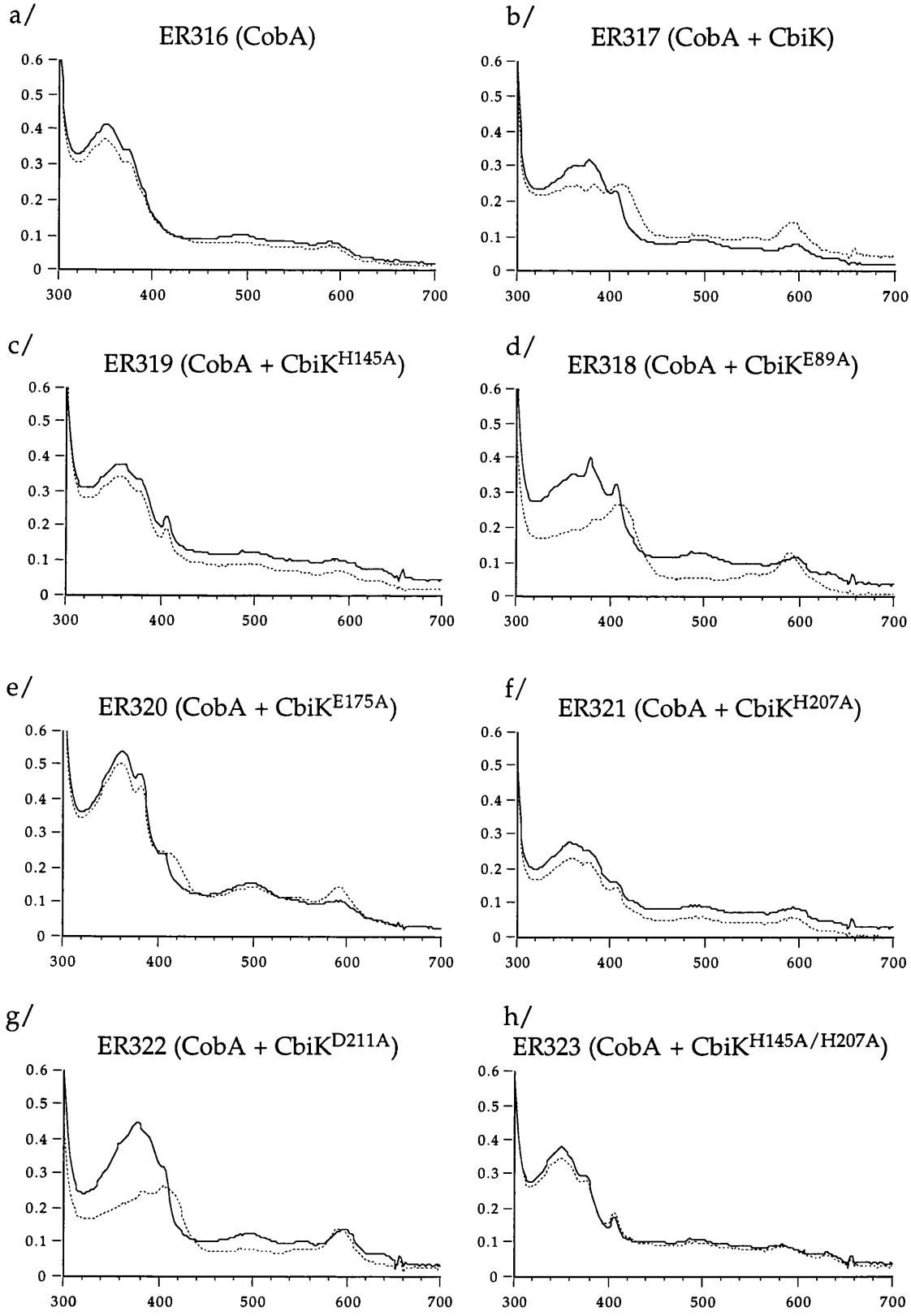
d- UV-visible spectra of accumulated compounds *in vivo.*

Cell extracts of the eight strains described in the previous section were grown anaerobically in minimal medium supplemented with cysteine, ALA and IPTG. All strains were grown in the presence or absence of exogenous cobalt. The accumulated modified tetrapyrroles were isolated on a DEAE-sephacel column (Materials and Methods, section 8-5-5) and a UV-visible spectrum was recorded immediately (Figure 4-17).

The spectrum obtained from the strain overproducing only the *P. denitrificans cobA* (Figure 4-17/a) is consistent with the presence of trimethylpyrrocorphin (maxima absorption at about 350 nm), although a second absorption at 380 nm suggests that some sirohydrochlorin is also produced. Exogenous cobalt does not modify the spectrum. It is known that trimethylpyrrocorphin accumulates in the presence of high concentrations of uro'gen III methyltransferases, so this is in agreement with previous results.

When *cbiK* is added to *cobA* in the absence of cobalt, the spectrum indicates that sirohydrochlorin is largely synthesised (Figure 4-17/b) and that a peak at 406 nm appears. This peak was also observed in section 4-2-3 during the CbiK *in vitro* study and may represent the formation of sirohydrochlorin-lactone. However, two new maxima at 414 and 594 appear in the presence of cobalt and the spectrum becomes very similar to that recorded previously for cobalt-sirohydrochlorin.

Figure 4-17- Spectra of tetrapyrrole-derived compounds.




The plain line indicates that the cells were grown in the absence of cobalt, and the dotted line, that they were performed in the presence of exogenous cobalt.

The spectroscopic analysis of the tetrapyrrole-derived compounds from the strains which contain *cobA* and the various *cbiK* mutant constructs can be classified into three categories:

The first class of *cbiK* mutant includes of $cbiK^{E89A}$ and $cbiK^{D211A}$, which both give similar spectra to the wild type *cbiK*. The two mutations do not appear to have an effect on the activity of the enzyme.

The second category contains the $cbiK^{E175A}$ which synthesises mainly trimethylpyrrocorphin but, in the presence of cobalt, can also produce cobalt-sirohydrochlorin at a low level. The E175A mutation in CbiK reduces the protein activity.

Finally, the third mutant category of effect includes the different histidine alterations, $cbiK^{H145A}$, $cbiK^{H207A}$ and $cbiK^{H145A/H207A}$. All these mutants give similar spectra to the strain containing *cobA* alone, producing largely trimethylpyrrocorphin. The three histidine mutants are unable to efficiently catalyse the cobalt chelation of precorrin-2.

e- Cobyric acid biosynthesis.

ER185, the *E. coli cysG* deleted strain containing all the genetic information to synthesise cobyric acid from cobalt-precorrin-2 was transformed with pER119, pER303 and its six *cbiK* mutant derivatives respectively. The resulting strains were grown in the presence of high, medium and low concentrations of exogenous cobalt and were assayed for cobyric acid production as indicated in section 8-4-3 and 8-4-4. The results are given in Table 4-3.

Table 4-3- Cobyric acid assays obtained with *cbiK* and the mutated constructs.

| Strain | *cob* operon /gene deletion | plasmid pKK derived constructs | Cobyric acid (pmol/$OD_{600}$) with $CoCl_2$, $6H_2O$ at | | |
|---|---|---|---|---|---|
| | | | 0 low | 0.01 mg/l medium | 1 mg/l high |
| ER306 | *S. typ / cbiK*$^{\Delta}$ | *cobA* | 0 | 0 | 0 |
| ER307 | *S. typ / cbiK*$^{\Delta}$ | *cobA-cbiK* | 9 | 94 | 100 |
| ER308 | *S. typ / cbiK*$^{\Delta}$ | *cobA-cbiK*$^{E89A}$ | 7 | 94 | 136 |
| ER309 | *S. typ / cbiK*$^{\Delta}$ | *cobA-cbiK*$^{H145A}$ | 0 | 5 | 26 |
| ER310 | *S. typ / cbiK*$^{\Delta}$ | *cobA-cbiK*$^{E175A}$ | 9 | 55 | 103 |
| ER311 | *S. typ / cbiK*$^{\Delta}$ | *cobA-cbiK*$^{H207A}$ | 1 | 55 | 111 |
| ER312 | *S. typ / cbiK*$^{\Delta}$ | *cobA-cbiK*$^{D211A}$ | 9 | 94 | 103 |
| ER313 | *S. typ / cbiK*$^{\Delta}$ | *cobA-cbiK*$^{H145A/H207A}$ | 0 | 0 | 0 |

All the strains are derived from strain ER185: *E. coli cysG* deleted strain (302Δa) containing the *S. typhimurium cbiA-C-D-E-T-F-G-H-J-K*$^{\Delta}$*-L-M-N-Q-O-P* genes with a deletion in *cbiK*.

As observed for the *cysG* complementation, the *cbiK*$^{E89A}$ and *cbiK*$^{D211A}$ mutants would also appear to have very similar properties to the wild type *cbiK* in regard to cobyric acid synthesis. Both mutants produce almost identical levels of cobyric acid under the different range of exogenous cobalt concentrations tested. Two other mutants, *cbiK*$^{E175A}$ and *cbiK*$^{H207A}$, produce identical levels of cobyric acid as wild type *cbiK* in the presence of high concentration of exogenous cobalt. In the presence of a medium containing only 0.01 mg/l of cobalt, both mutants have a slightly reduced synthesis of cobyric acid, but since this decrease is less than a two fold reduction, its significance is questionable. However, in the absence of exogenous cobalt, although the *cbiK*$^{E175A}$ mutant produced a similar quantity of cobyric acid to wild type, the *cbiK*$^{H207A}$ mutant barely produced any measurable product. This result reinforces the role of histidine 207 in cobalt binding. The most severe single mutation with respect to corrin biosynthesis is the histidine 145 alteration. This enzyme has greatly reduced cobyric acid synthesis under high and medium cobalt concentrations and does not produce cobyric acid at low cobalt levels. Finally, the *cbiK*$^{H145A/H207A}$ construct does not induce cobyric acid production, even at high cobalt concentration, confirming that the protein is completely inactive.

f- Summary of *cbiK* mutants.

Although positioned at the active site of CbiK, residues E89 and D211 do not appear to be crucial to the enzyme activity. When either one of these groups is mutated to alanine, the resultant protein appears to have a very similar phenotype to the wild type *cbiK* (Table 4-4). Clearly the criteria on which the mutants were analysed are not very sensitive but they do demonstrate that these two amino acids are not essential in the chelation process.

The other acidic amino acid change, E175A, also appeared to have little effect on either cobalt- or ferrochelatase activity. However, when the enzyme was overproduced with *cobA*, the strain only accumulated approximately 25% of the cobalt-sirohydrochlorin observed with wild type *cbiK*. Therefore, the mutation E175A does have an effect on the enzyme. More information on the effect of the mutation will become available once the purified enzyme will be assayed *in vitro*. The position of E175 within the protein indicates that this residue is quite likely to be involved in metal binding.

Finally, the two histidine residues H145 and H207 were both found to play an important role in cobalt binding and metal-ion selection. Out of the two, H145 appears to play the more essential role since its mutation renders the enzyme a very poor ferro- and cobalt chelatase. His 145 is also conserved in protoporphyrin ferrochelatase. Most notable about H207A is the observation that the enzyme appears to become a more competent precorrin-2 ferrochelatase, having apparently lost the ability to bind cobalt with the same affinity. It will be interesting to see if changing the histidine 207 to a glutamate (the conserved residue found in protoporphyrin ferrochelatase at the same position) makes the protein an even more efficient precorrin-2 ferrochelatase.

The combined importance of H145 and H207 becomes apparent from the results of the double mutant. In this case, the enzyme is completely inactive. It will be important to establish whether this double change has just altered the metal binding site or whether more drastic structural alterations have taken place.

The three-dimensional CbiK structure determination and the mutagenesis experiments give some information about the mechanism of cobalt chelation in

the cobalamin anaerobic pathway. Firstly, the similarity of the CbiK topology with the protoporphyrin ferrochelatase implies that the two enzymes utilise a common mechanism. Thus, it is likely that the tetrapyrrole-derived substrate binds to the enzyme in such a way that the four pyrrole nitrogens can accommodate the metal ion, probably through distortion of the substrate. Once bound in this way, the proximity of the metal ion, bound via His 145 and His 207, ensures that chelation takes place. The chelation then induces a conformation change in the modified tetrapyrrole such that it favours a more linear conformation, encouraging the metal-tetrapyrrole complex to vacate the active site.

Table 4-4- Summary of the *cbiK* mutation effects *in vivo*.

| CbiK mutation | sirohaem complementation | UV-visible spectra | cobyric acid synthesis | Effect |
|---|---|---|---|---|
| E89A | good | Co-SHC | 100% | no visible effect |
| H145A | slow, +/- $Co^{2+}$ | TMPC | 25% | reduction of activity and Co selectivity |
| E175A | good | low Co-SHC | 100% | reduction of activity |
| H207A | good, +/- $Co^{2+}$ | TMPC | 10% without $Co^{2+}$ added | reduction of Co selectivity |
| D211A | good | Co-SHC | 100% | no visible effect |
| H145A/ H207A | no | TMPC | 0 | complete lost of activity |

SHC is the abbreviation of sirohydrochlorin; TMPC, trimethylpyrrocorphin.

## 4-6- Comparison of CbiK with CbiX.

The *B. megaterium* CbiX has been classified in Chapter 3 as another precorrin-2 cobalt chelatase. The sequence similarity at the primary level between CbiX and CbiK is very low but an alignment between the two proteins can be forced (Figure 4-18). This alignment reveals some very interesting amino acid conservations. For instance, the amino acids which are known to form the catalytic site of CbiK, His 145, His 207, Glu 175 and Asp 211 are also present in CbiX. If this is significant, then the N-terminal domain of CbiX is very likely similar to the C-terminal domain of CbiK. A secondary structure alignment of CbiX with the known CbiK structure also indicates a similar overall topology [Dr H. Schubert, pers. comm.]. The mechanism of cobalt chelation employed by CbiX is likely to be similar to CbiK. A major difference found between CbiK and HemH was the substitution of Glu 264 (HemH) to His 207 (CbiK). This substitution was presumed to play a role in the selectivity of different metals, either ferrous iron or cobalt. To some extent, this hypothesis has been proved with the substitution of the histidine 207 in CbiK to alanine, resulting in the loss of selectivity of cobalt over ferrous-iron. Among the three CbiX homologous proteins, *B. megaterium* CbiX, *Synechocystis sp.* CbiX and *B. subtilis* YlnE, a conserved histidine is also found at that corresponding position (Figure 4-18). However, the function of YlnE is very likely to be a sirohydrochlorin iron chelatase since *B. subtilis* does not make cobalamin *de novo* and *ylnE* is part of a sirohaem operon. Moreover, YlnE does not possess the poly-histidine tail that is found in the two other CbiX proteins, which could explain the difference in metal preference.

Only the structure determination of CbiX will definitively prove whether the two enzymes, CbiX and CbiK, really do share a similar domain and utilise the same mechanism.

Figure 4-18- Alignment of CbiX, YlnE and CbiK.

```
                1                                                    50
Bm CbiX                                MG GHYMKSVLFV GHGSRDPEGN DREFISTMKH
Syn CbiX       MTLTSVPAPV SLFPELELPP LPYHRPLLMI GHGTRDEDGR QTFLDFVAQY
Bs YlnE                                  MKQAILYV GHGSRVKKAQ QEAAAFLEGC
CbiX Consensus ---------- ---------- -------L-- GHG-R----- ----------
St CbiK        LSSHNDYVQL MQALRQQMPS LRQTEKVVFM GHGASHHAFA AYACLDHMMT
                                                 145

                51                                                  100
Bm CbiX        D.WDASILVE TCFLEFERPN VSQGIDTCVA KGAQDVVVIP IMLLPAGHSK
Syn CbiX       QALDHSRPVI PCFLELTEPN IQAGVQQCVD QGFEEISALP ILLFAARHNK
Bs YlnE        KAHISVPVQE ISFLELQEPT IETGFEACVK QGATHIAVVP LLLLTAAHAK
CbiX consensus ---------- --FLE---P- ---G---CV- -G-------P --L--A-H-K
St CbiK        .AQRFPARVG AV..ESY PE VDILIDSLRD EGVTGVHLMP LMLVAGDHAI
                         175                                 207

              101                                                   150
Bm CbiX          IHIPAAIDEA KEKYPHVNFV YGRPIGVHEE ALEILKTRLQ ESGENLETP.
Syn CbiX         FDVTNELDRS RQAHPQINFF YGRHFGITPA ILDLWKARLN QLDSPEANPQ
Bs YlnE          HDIPEEIVRV ASRYPSVRIS YGKPIGIDEE VVKAVYHRMK DIGVPYENAR
CbiX consensus   ---------- ----P----- YG---G---- -------R-- ----------
St CbiK          NDMASDDGDS WKMRFNAAGI PATPWLSGLG ENPAIRAMFV AHLHQALNMA
              211

              151                                                   200
Bm CbiX          ...AEDTAVI VLGRGGSDPD ANSDLYKITR LLWEKTNYKI VETSFMGVTA
Syn CbiX         GIDRQDTVLL FVGRGSSDPD ANGDVYKMAR MLWEGSGYQT VETCFIGISH
Bs YlnE          ........VV LIGRGSSDPD VKRDVTGIAN LLQEMVPVKE VIPCFLTACG
CbiX consensus   ---------- --GRG-SDPD ---D------ -L-E------ V---F-----
St CbiK          VEEAA*

                201                                                  250
Bm CbiX          PLIDEGVERC LKLGAKKVVI LPYFLFTGVL IKRLEEMVKQ YKMQHENIEF
Syn CbiX         PRLEEGFRRA RLYQPKRIIV LPYFLFMGAL VKKIFTITEE QRATFPEIEI
Bs YlnE          PNYKEVFSEL EKDDGITTFI VPYLLFTGML MNEIEREVQK LKAHNPNV..
CbiX consensus   P---E----- ---------- -PY-LF-G-L ---------- ----------

                251                                                  300
Bm CbiX          KLAGYFGFHP KLQTILKERA EEGLE.GEVK MNCDTCQYRL GIMEHIDHHH
Syn CbiX         QSLSEMGIQP ELLALVRERE IETQL.GQVA MNCEACKFRL AF.....KNQ
Bs YlnE          YLSSYIGFHP HVKNAFLNRV RETAANSEGQ FDFDGGSYAS AAH*
CbiX consensus   ------G--P --------R- -E-------- ---------- ----------

                301                                              342
Bm CbiX          HHDHDHDHDH GHHHHDHHHD HHEDKVGELK *
Syn CbiX         GHGHDHGHGH HHHGHDHGHS HGEWVDTYIE PTAYHEKIWQ AP*
Bs YlnE
CbiX consensus   ---------- ---------- ---------- ---------- --
```

The amino acids from the St CbiK sequence in grey boxes indicates that they are conserved in all CbiK sequences.

## 4-7- Conclusion.

The *S. typhimurium cbiK* has been subcloned in-frame into the pET14b vector, allowing its expression as a recombinant protein which contains an extra twenty amino acids at its N-terminal. The introduction of the His-tag permits a simple, one step protein purification procedure.

The His-tagged CbiK was assayed *in vitro* in the presence of *E. coli* PBG deaminase (HemC), *E. coli* uro'gen III synthase (HemD), *P. denitrificans* uro'gen III methyltransferase (CobA), SAM and PBG. The incubated mixture transformed porphobilinogen (PBG) into sirohydrochlorin and into cobalt sirohydrochlorin in the presence of cobalt. These results confirm the previous data obtained *in vivo* (Chapter 3) and demonstrate that the His-tagged CbiK is active and acts as a cobalt chelatase.

The structure determination of CbiK by crystallisation and X-ray diffraction revealed a high degree of similarity with the *B. subtilis* protoporphyrin ferrochelatase despite the lack of similarity at the amino acid level. The conservation of structure between the two enzymes suggests a common mechanism for the ATP-independent metal chelation of the modified tetrapyrroles. Five amino acids have been highlighted and proposed to be important within the active site of CbiK, His 145, His 207, Glu 89, Glu 175 and Asp 211. A series of directed-site mutagenesis were performed such that these amino acids were changed to alanines. The CbiK mutants were subsequently analysed by complementation and spectroscopic studies (Table 4-4). It was shown that Glu 89 and Asp 211 did not play an essential role in the catalytic process, whereas Glu 175 had a significant role in cobalt binding. However His 145, presumed to be an essential component of the cobalt binding site, was found to be vital for the functioning of the enzyme since mutagenesis of this residue to alanine resulted in a protein with very little catalytic activity. In close proximity to His 145 is another histidine, His 207, which also seems important in metal binding. The equivalent position in protoporphyrin ferrochelatase is occupied by glutamate, but mutagenesis of His207 to alanine does produce an enzyme with an apparent greatly reduced cobalt affinity.

The *B. megaterium* precorrin-2 cobalt chelatase (CbiX) sequence was compared with CbiK. CbiX seems to share a common domain with CbiK (domain II). The structure determinations of CbiX (cobalt chelatase) and YlnE (ferrochelatase) will be important for understanding the mechanism of metal-ion chelation as well as metal-ion selectivity. Indeed, His 207 of CbiK has been shown to be a key residue for the cobalt selection over ferrous iron, but also to be conserved in YlnE.

Two other enzymes have been characterised to have a metal-ion chelation activity, CysG and Met8p. In both cases, there is no apparent homology with CbiX, CbiK or HemH (protoporphyrin ferrochelatase). This may indicate that the multifunctional CysG and Met8p have unique structures, representing the convergent evolution of these chelatases.

# CHAPTER 5

## CLONING, EXPRESSION AND X-RAY STRUCTURE OF THE COBALT-PRECORRIN-4 METHYLTRANSFERASE (CbiF) FROM *B. megaterium* AND COMPARISON OF ALL METHYLTRANSFERASES INVOLVED IN COBALAMIN BIOSYNTHESIS.

## 5-1- Introduction.

The biosynthesis of cobalamin is dependent upon eight *S*-adenosyl-L-methionine (SAM) transmethylations. These methylations occur during the construction of the corrin ring at position C-2 and C-7 ($CysG^A$/CobA), C-20 (CbiL/CobI), C-17 (CbiH/CobJ), C-11 (CbiF/CobM), C-1 (?/CobF), C-5 and C-15 (CbiE/CobL). The most striking feature between all these methyltransferases is the well conserved SAM binding site [VGxGPGxxxxxT], [VxxxxxGDPxxY/F] and [VIPGV] (Figures 2-5 and 5-1). Thus, the structure elucidation of anyone of these methyltransferases will not only provide an insight into the mechanism of precorrin methylation but will also generate a model onto which the structure of the other precorrin methyltransferases can be predicted.

Figure 5-1- CbiF and the conserved motifs associated with all methyltransferases.

```
          M KLYIIGAGPG DPDLITVKGL KLLQQADVVL YADSLVSQDL
IAKSKPGAEV LKTAGMHLEE MVGTMLDRMR EGKMVVRVHT GDPAMYGAIM
EQMVLLKREG VDIEIVPGVT SVFAAAAAAE AELTIPDLTQ TVILTRAEGR
TPVPEFEKLT DLAKHKCTIA LFLSATLTKK VMKEFINAGW SEDTPVVVVY
KATWPDEKIV RTTVKDLDDA MRTNGIRKQA MILAGWALDP HIHDKDYRSK
LYDKTFTHGF RKGVKSE
```

The three conserved motifs found in all methytransferases involved in cobalamin biosynthesis are indicated by grey boxes ((i) VGxGPGxxxxxT, (ii) VxxxxxGDPxxY/F and (iii) VIPGV).

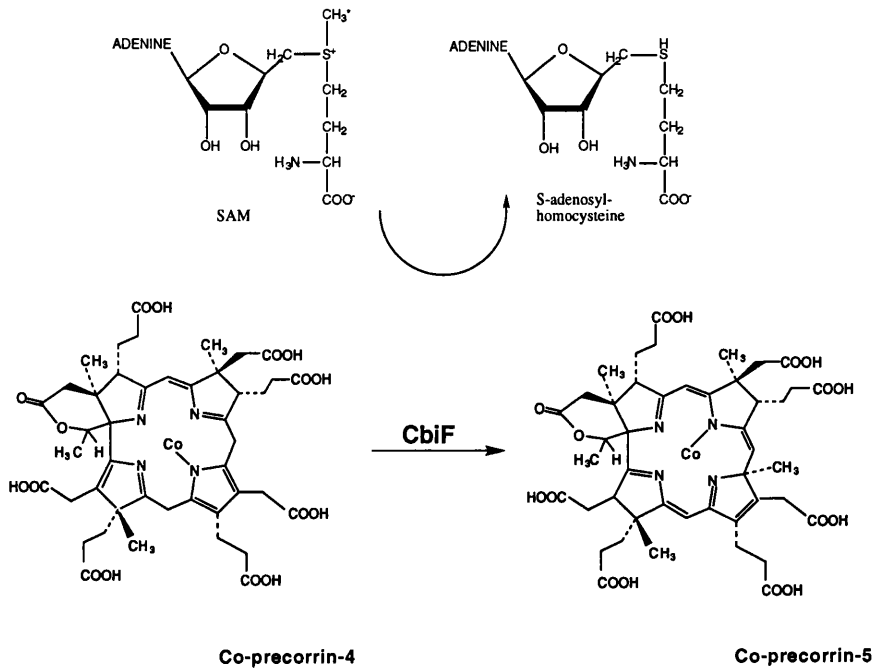
The anaerobic methyltransferases (Cbi prefix) have been identified and classified by their high degree of similarity with the previously characterised aerobic enzymes of *P. denitrificans* (Cob prefix). Except for the CobA/$CysG^A$ uro'gen III methyltransferase, the aerobic and anaerobic enzyme substrates are different since the anaerobic enzymes require a substrate with a central cobalt [Muller *et al.* 1991]. In the aerobic pathway, cobalt insertion occurs only on a fully methylated corrin, at the stage of hydrogenobyrinic a,c-diamide [Debussche *et al.* 1992].

Among all the methyltransferases involved in vitamin $B_{12}$ biosynthesis, the *B. megaterium* CbiF cobalt-precorrin-4 methyltransferase that performs the methylation at C-11 (Figure 5-2) was chosen for crystallisation trials since it could be expressed to high level as a soluble, active, His-tagged recombinant protein.

The X-ray structure of *B. megaterium* CbiF has now been determined to a resolution of 2.4 Å [Schubert *et al.* 1998]. The process of crystallisation and analysis of the data were performed by Dr Heidi L. Schubert at the University of York.

With the structure of CbiF in mind, a detailed analysis and comparison of all the other methyltransferases involved in cobalamin biosynthesis (from twelve different bacteria) was undertaken.

Figure 5-2- The reaction catalysed by *B. megaterium* CbiF.



## RESULTS

### 5-2- Cloning, expression and purification of *B. megaterium* CbiF.

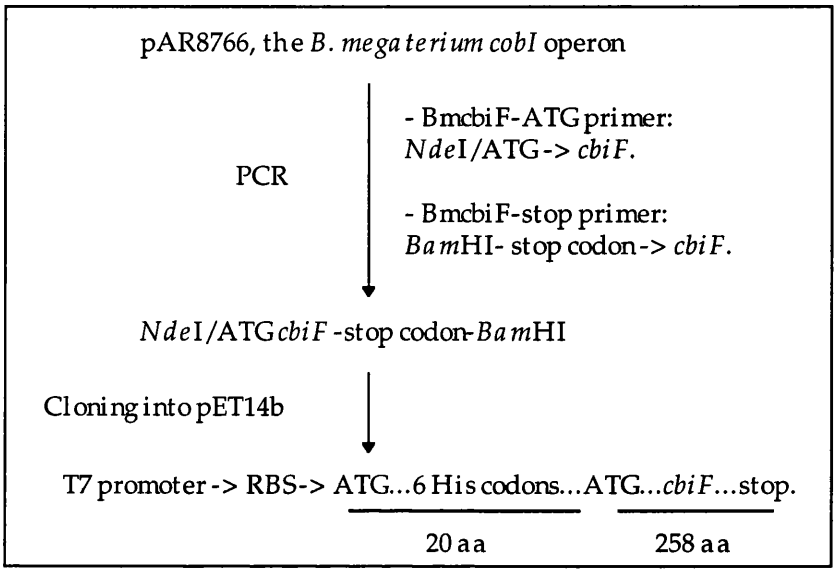
The *B. megaterium cbiF* was amplified by PCR from the main *B. megaterium cob* operon and cloned into an expression vector such that the gene was fused to a DNA fragment encoding for a hexa-histidine peptide. When transformed into *E. coli*, this construct led to an overproduction of His-tagged CbiF, which was easy to purify.

This recombinant protein was tested for activity *in vivo*, by complementation of a *S. typhimurium cbiF* mutant strain, and *in vitro*, by using the SAM binding assay already described in Chapter 2. Finally, the purified protein was used in crystallisation trials after the clone was sent to Dr H. Schubert (University of York).

### 5-2-1- Cloning of *cbiF* in an expression vector.

In order to attain high level expression of CbiF, the *cbiF* gene was amplified by PCR using the plasmid pAR8766, which contains the whole *B. megaterium cobI* operon, as template. The sequence of the 5′ and 3′ primers, BmcbiF-ATG and BmcbiF-stop, are given in section 8-2-12. The sequence of the 5′ primer contains an *Nde*I site thus allowing the amplified gene to be cloned into pET14b in-frame, thereby creating a fusion protein between a twenty amino acid peptide, containing a stretch of six histidines, and CbiF. The sequence of the 3′ primer contains the *cbiF* stop codon and a *Bam*HI site. After amplification of *cbiF*, the PCR product (the PCR procedure is described in Materials and Methods, section 8-2-11) was digested with *Nde*I and *Bam*HI and cloned to the corresponding sites in pET14b (Figure 5-3). JM101 was transformed with the resulting plasmid (pER229), which gives ER229. JM101 was used to maintain the plasmid.

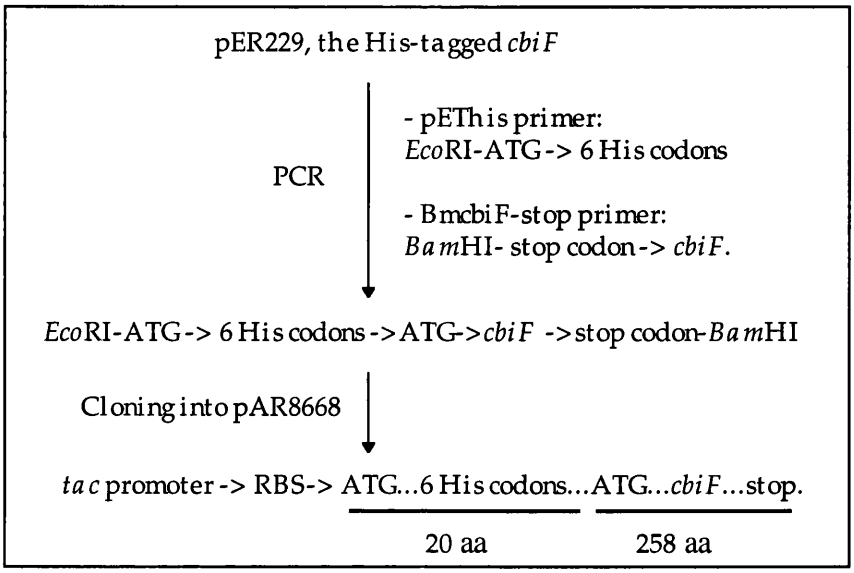
Figure 5-3- The cloning procedure of *cbiF* into pET14b.




## 5-2-2- Complementation of *S. typhimurium cbiF* mutant with the His-tagged *B. megaterium cbiF.*

In order to check that the His-tagged CbiF was physiologically active, the fused gene was subcloned under the control of a *tac* promoter (instead of the T7 promoter), since the *tac* promoter allows expression in the *S. typhimurium cbiF* mutant strain. The His-tagged *cbiF* was amplified by PCR using a primer designed to start at the ATG of the histidine stretch of plasmid pET14b. The 5′ primer (pEThis) contains an *Eco*RI site. The 3′ primer used in this amplification was the same as for the previous PCR (BmcbiF-stop). After amplification, the PCR product was digested with *Eco*RI and *Bam*HI and cloned into the corresponding sites in pAR8668 (pKK223.3 with only one *Bam*HI site) (Figure 5-4). The ribosome binding site of pKK223.3, located in 5′ of the multi-cloning site, can be used to initiate the translation of the His-tagged CbiF if the corresponding gene is cloned in the *Eco*RI site. In this case, the resulting sequence is AGGAAACCAGAATTCATG. The pKK-His-tagged *cbiF* plasmid was named pER279.

Figure 5-4- The cloning procedure of His-tagged *cbiF* into pAR8668.




*S. typhimurium cbiF* mutant strain (AR2720) was transformed with pER279 (the His-tagged *B. megaterium cbiF*), pAR8668 (negative control) and pAR8477 (positive control, *S. typhimurium cbiF*). The recombinant strains were subsequently tested for their ability to produce cobalamin as described in Materials and Methods, section 8-4-2. Both, pER279 and pAR8477 were found to complement the *S. typhimurium* mutant whereas the negative control, pAR8668, remained unable to produce cobalamin: the His-tagged CbiF is therefore active *in vivo*.

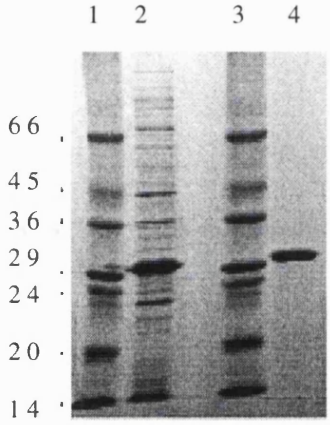
## 5-2-3- Expression and purification of His-tagged CbiF.

pET14b-*B. megaterium cbiF* (pER229) was transformed into BL21(DE3)(pLysS) for protein expression, resulting in strain ER225. The His-tagged CbiF was expressed and purified as described in section 4-2-2. From one litre of culture, approximately 20 mg of pure protein was reproducibly obtained. The protein was stored at –20°C in 50mM Tris HCl at pH 7.8.

The purified protein ran as a single band on SDS/PAGE with an apparent molecular mass of about 31 kDa (Figure 5-5), which is in agreement with the expected molecular mass (28.3 kDa plus 2.2 kD corresponding to the His-tag).

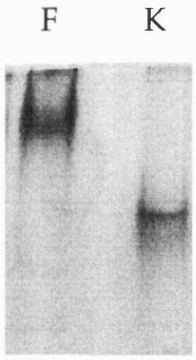
By non-denaturing gel electrophoresis, the purified protein also migrated as a single band which suggests that the protein has a relatively stable quaternary structure (Figure 5-6). On this basis, CbiF was a good candidate for crystallisation trials. The native molecular mass of CbiF was estimated from gel filtration studies using a Bioselect gel filtration column. The His-tagged CbiF eluted with a molecular mass of 62 kDa (+/- 5 kDa) suggesting that the protein exists as a homodimer [Dr S. Woodcock, pers. commun.].

Figure 5-5- SDS/PAGE of the His-tagged CbiF.




Lanes 1 and 3 contain molecular mass markers as indicated (in kDa). Lane 2 contains a crude cell extract of the recombinant *E. coli* that over-expressed the His-tagged-CbiF. Lane 4 contains the purified His-tagged CbiF.

Figure 5-6- Non-SDS PAGE of the His-tagged CbiF and His-tagged CbiK.



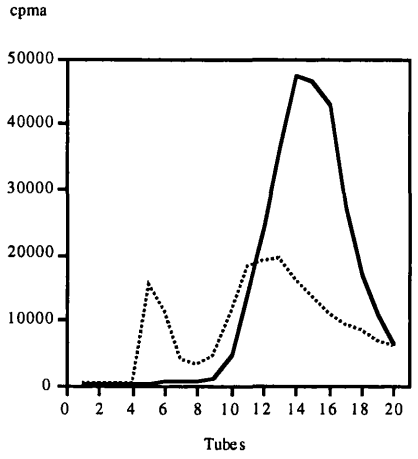

Lane F contains the native His-tagged CbiF and lane K, the native His-tagged CbiK.

The N-terminal His-tag peptide of CbiF could be proteolytically cleaved from CbiF by incubating the fusion protein overnight at 30°C with thrombin in 70 mM Tris/HCl at pH8.5, 100 mM NaCl and 2.5 mM $CaCl_2$. The cleaved CbiF was purified by gel filtration [Dr H. Schubert, pers. commun.].

## 5-2-4- SAM binding assay.

As with other methyltransferases involved in cobalamin biosynthesis, CbiF contains a highly conserved SAM binding site. To determine whether CbiF is able to bind [$^3$H-methyl]-SAM and retain the substrate upon gel filtration, a SAM binding assay was undertaken as described in Materials and Methods, section 8-5-4. Since $CobA^{Pd}$ is known to bind SAM very tightly, it was used as positive control in this experiment (Figure 5-7).

Figure 5-7- [$^3$H-methyl]-SAM binding assay results after gel filtration.




The dotted line represents the [$^3$H-methyl]-SAM measurement after incubation with CobA and gel filtration; the plain lines represents the equivalent [$^3$H-methyl]-SAM measurement with CbiF.

Fractions (1 ml) were collected and were analysed for protein content and radioactivity. Protein was found to elute in the early fractions (5 to 7) whereas the free-SAM eluted in the later fractions (12 to 20). Radioactivity in tubes 5 to 7 is only found when SAM binds a protein tightly, as it is observed with CobA but not with CbiF. The *B. megaterium* CbiF does not bind SAM tightly enough to retain any substantial radioactivity during gel filtration. The presence of the second substrate, cobalt-precorrin-4, may be important for the configuration of the protein for SAM binding. Furthermore, it has to be remembered that CobA

has been purified from an environment where uroporphyrinogen III was present.

## 5-3- X-ray structure of CbiF.

The crystallisation and X-ray structure determination of CbiF were performed by Dr H. Schubert at the University of York. A summary of this work is described in sections 5-3-1- and 5-3-2 [Schubert *et al.* 1998].

### 5-3-1- Crystallisation data.

The purified His-tagged CbiF was stored in 20 mM sodium acetate at pH 5.6, 100mM NaCl, and concentrated from 16 to 20 mg/ml. After an initial Hampton screen using the vapour diffusion technique, two crystal forms of the enzyme were observed (Figure 5-8).

The first crystal type, poorly formed rods, grew from the precipitate formed when an equal volume of protein was mixed with reservoir solution containing 2 to 3% poly (ethylene glycol) (molecular mass 8 kDa), 2 to 5% methyl pentane-diol and 100 mM Hepes, at pH7.5.

The second crystal form was obtained when the protein solution was mixed with an equal volume with a reservoir solution containing 1.0 to 1.2 M sodium potassium phosphate, 100 mM Hepes at pH 7.5. The addition of 4% dioxane eliminated severe twinning (Figure 5-8). Diffraction data were collected to 2.4Å. The crystals are trigonal, space group $P3_121$, with cell dimension of a=b=80.4Å, c=109.6Å.

Soaking the crystal in 500 mM MeHgCl for 1 hour prior to data collection resulted in a single-site heavy-atom derivative.

After thrombin cleavage of the His-tag, CbiF did not crystallise in the high phosphate medium used for the His-tagged-CbiF. However, crystals did grow from 25% monoethylether poly (ethylene glycol) (molecular mass 2 kDa), 200 mM $MgCl_2$ and 100mM Tris at pH8.5. The His-tagged-free form of CbiF was found to have cell dimension a=b=80.04Å, c=77.92Å. Diffraction data were collected to 3.1Å.

In all cases, crystals only grow in the presence of SAM or S-adenosyl-L-homocysteine [Schubert *et al.* 1998]. Within the crystal, only the S-adenosyl-L-homocysteine form was found.

## 5-3-2- Structure of CbiF.

CbiF is composed of two α/β domains linked by a single coil having the shape of a kidney-bean. Both domains contain a five-stranded β-sheet flanked by four α-helices. The entire structure follows a β-α repeating pattern with a single exception of a β-hairpin late in the C-terminal domain (β8- β9). From the gel filtration study to estimate the molecular mass, CbiF was found to be a dimer. There are 36 direct protein-protein hydrogen bonds in the dimer surface (Figure 5-9).

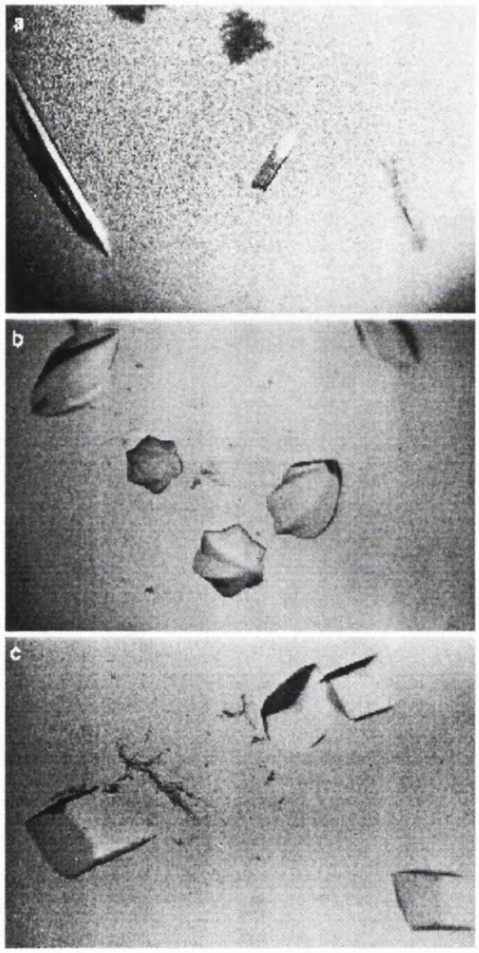
The absence of phosphate in the crystallisation medium after cleavage of the His-tagged peptide led to a new space group with a tighter crystal packing. The overall fold remains the same.

S-adenosyl-L-homocysteine is bound in one pocket of a large trough between the N- and C-terminal domains. S-adenosyl-L-homocysteine bound to CbiF is kinked between the sulfur atom and the sugar ring to place both the homocysteine backbone and the adenosine ring into pockets of the trough, similar to a two pronged plug in a socket (Figure 5-12). The binding of S-adenosyl-L-homocysteine to CbiF is quite different to those found in other SAM-binding proteins not involved in vitamin $B_{12}$ biosynthesis such as DNA and catechol transmethylases [Malone *et al.* 1995].

The most likely binding site for cobalt-precorrin-4 is the large trough in the N-terminal domain. Several loops make up the walls of the trough, loop β2-αB and β3-αC forms the left side of the trough and loop β4-αD and helix αD, the right side.

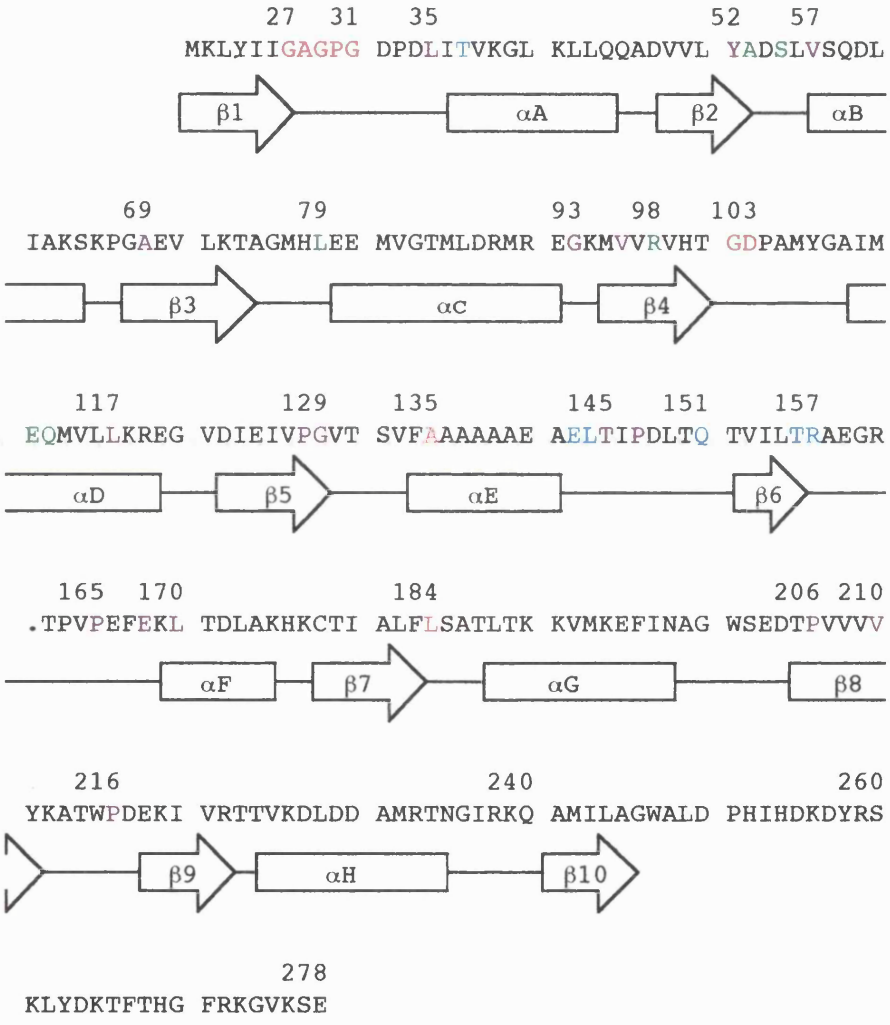
No residues were found to act as a cobalt ligand. The nearest residue which could be involved in such an act is histidine 100, located at the bottom of the substrate trough of the strand β4 [Schubert *et al.* 1998]. However, this histidine is also conserved in the four CobM sequences, suggesting that this hypothesis is unlikely. The C-terminus of all CbiFs possesses a conserved motif, SxLYxxxxxH, of which the histidine is not present in CobM. Unfortunately, this motif does not show up in the structure of the CbiF since this region has a high degree of flexibility.

Figure 5-8- Different crystal forms of CbiF.



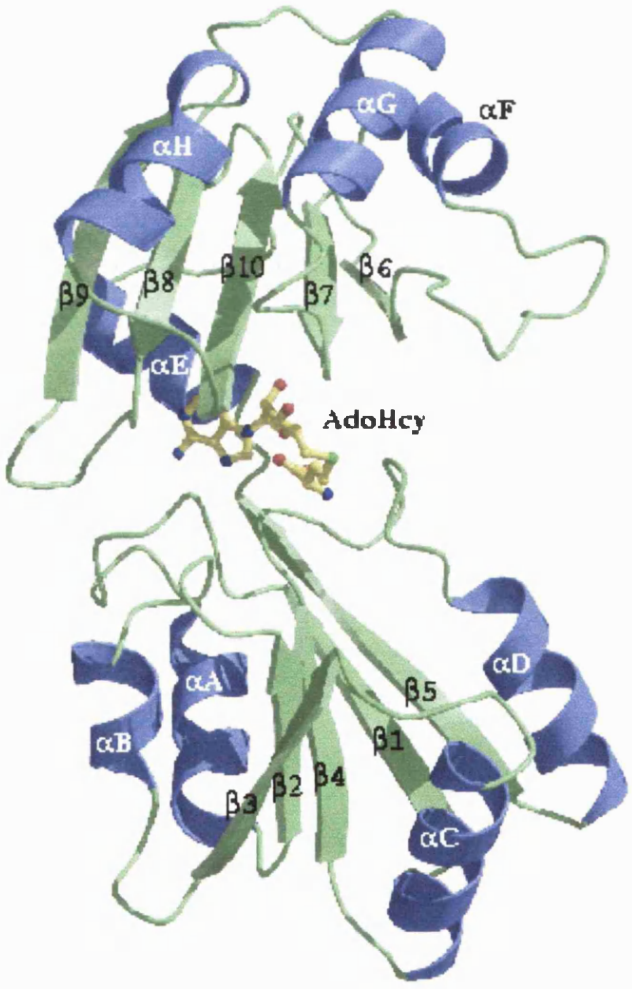

a) Poorly formed rods. b) CbiF crystals obtained after addition of sodium potassium phosphate. c) Addition of dioxane prevents the twinning observed in CbiF crystals.

Figure 5-9- Structural representation of CbiF.




The amino acids printed in red are conserved residues involved in SAM binding; in green, residues involved in precorrin binding; in blue, residues involved in the dimer interactions and finally, in purple, residues which make up either the hydrophobic core or tight turns. Arrows represent α-helices and rectangle, β-sheet. Loop β2-αB and β3-αC forms the left side of the precorrin trough and loop β4-αD and helix αD, the right side.

Figure 5-10- Three-dimensional representation of CbiF.



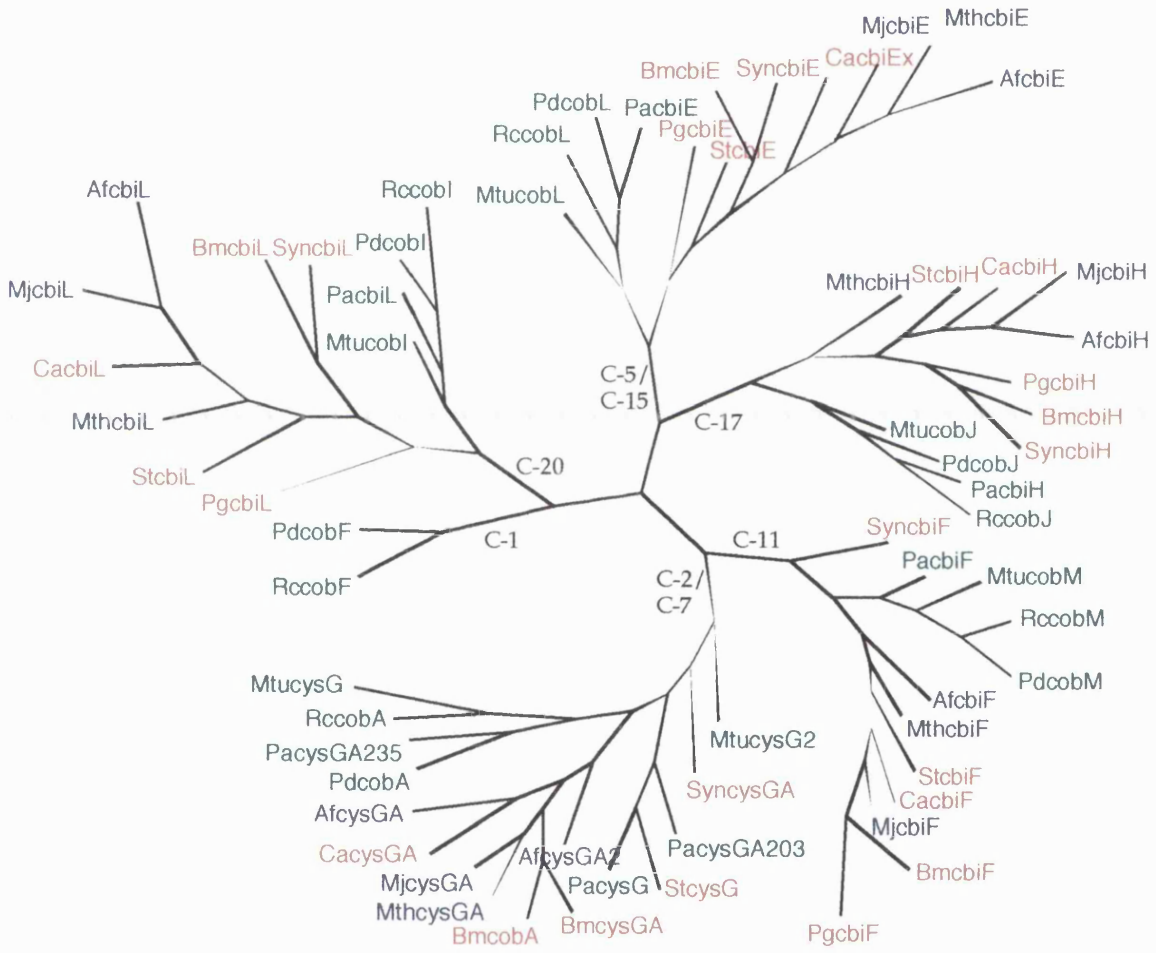
## 5-4- Comparison with other CbiFs, CobMs and other methyltransferases.

### 5-4-1- Comparison of 66 methyltransferases involved in vitamin $B_{12}$ biosynthesis.

The amino acid sequences of all the methyltransferases from twelve different bacteria involved in cobalamin biosynthesis are available on the databases. As well as the sequences of *P. denitrificans*, *B. megaterium* and *S. typhimurium*, the genome sequencing projects provide information about nine other vitamin $B_{12}$ producers; (i) three Archaeobacteria, *Methanobacterium jannaschii*, *Methanobacter thermoautotrophicum*, and *Archaeoglobus fulgidus*, with characteristic enzymes of an anaerobic pathway (CbiD and CbiG); (ii) an endospore forming bacteria, *Clostridium acetobutylicum*, a Cyanobacteria, *Synechocystis sp.* and a strict pathogenic anaerobe, *Porphyromonas gingivalis*, all of which appear to contain specific enzymes of the anaerobic pathway; (iii) two bacteria which contain some of the characteristic enzymes of an aerobic pathway, the phototrophic *Rhodobacter capsulatus* (CobF, -N, -S, -T but no CobG) and *Mycobacterium tuberculosis* (CobG, CobN but no CobF, -S, -T); and finally (iv) a bacterium which contains specific enzymes of both pathways, *Pseudomonas aeruginosa*.

The 66 methyltransferases include $CysG^A$/CobA, CbiL/CobI, CbiH/CobJ, CbiF/CobM, CobF and CbiE/CobL (the *C. acetobutylicum* CbiE sequence contains at least one sequencing error). Some bacteria were found to have duplicate genes; *M. tuberculosis*, *A. fulgidus*, *B. megaterium* and *P. aeruginosa* have two or even three $CysG^A$ like proteins each. In the case of fusion proteins, the peptides have been restricted to the methylase domain. The *M. tuberculosis* CobIJ has been separated into two distinct peptides. An alignment using the GCG packages (pileup, lineup and pretty) was performed and gave the consensus displayed in Figure 5-12 with a plurality of 40. The formatted alignment was transferred to PAUP 3.1.1 with the help of Dr J. Roper and Dr M. Van Dornum. The tree obtained after analysis of the data is shown in Figure 5-11.

Figure 5-11- Phylogenetic tree obtained from 66 methyltransferases involved in cobalamin biosynthesis.




All proteins from bacteria having an aerobic cobalamin biosynthetic pathway are indicated in green whereas proteins from bacteria having an anaerobic pathway are shown in red except for the Archaeobacteria, which are highlighted in blue. The gene prefix indicate the strains of origin: *P. denitrificans* (Pd), *B. megaterium* (Bm), *S. typhimurium* (St) *Methanobacterium jannaschii* (Mj), *Methanobacter thermoautotrophicum* (Mth), *Archaeoglobus fulgidus*, (Af), *Clostridium acetobutylicum* (Ca), *Synechocystis sp* (Syn), *Rhodobacter capsulatus* (Rc), *Mycobacterium tuberculosis* (Mtu), *Porphyromonas gingivalis* (Pg) and *Pseudomonas aeruginosa* (Pa).

The phylogenetic tree is mainly composed of five branches, each one corresponding to a different type of methyltransferase. The most interesting feature of the tree is the formation of a sub-division within each branch, where the proteins segregate according to aerobic/anaerobic pathway affiliation. From an evolutionary point of view this suggests that the different methyltransferases of corrin biosynthesis evolved from a common and basic methyltransferase ancestor before the aerobic and anaerobic pathways appeared, and therefore before the enzymes were able to discriminate between cobalt and cobalt-free intermediates.

A sub-branch is formed with four aerobic bacteria within the CobA/CysG$^A$ branch, but the segregation in this group is debatable due to the apparent gene duplication in several bacterial species. This may reflect the fact that these enzymes catalyse the synthesis of a key branch-point intermediate (Figure 1-2) and that the different *cobA/cysG*$^A$ genes could have been acquired from operons encoding for different pathways.

For example, the protein called MtuCysG represents a fusion between CobA and HemD whereas the protein MtuCysG2 is similar to the *S. typhimurium* CysG. One protein may have evolved for cobalamin biosynthesis whilst the other may be specific for sirohaem biosynthesis. However, since nothing is known about vitamin $B_{12}$ biosynthesis in *M. tuberculosis*, the reason why this bacterium has two CysG like proteins remains a conundrum. The same gene multiplication is observed within the *P. aeruginosa* genome, which harbours three *cysG* like genes.

The CbiL/CobI branch of the tree also contains an interesting sub-branch since it accommodates the CobF sequences. Only two CobF proteins (the aerobic C-1 methyltransferase) have so far been determined from *P. denitrificans* and *R. capsulatus*. Significantly, no CobF has been found in *P. aeruginosa* or *M. tuberculosis*, bacteria that are also thought to harbour aerobic-like pathways. As with the anaerobic pathway, it is unclear how these two bacterial species methylate at C-1.

Surprisingly, the three archaebacteria do not form a sub-division within the five branches. This could be viewed as further evidence of corrin synthesis being an ancient pathway, having already diversified within the archaebacteria prior to the appearance of the eubacteria.

The *B. megaterium* CbiF is located in the CobM/CbiF branch which is more related to the uro'gen III methyltransferase branch. This is not so surprising since it is known that, functionally, CobA can methylate at C-12, albeit to generate a physiological irrelevant 2, 7, 12 trimethylpyrrocorphin. Both CobA and CbiF enzymes must therefore bind their substrate in a similar manner to accommodate the regiospecific transmethylation in the south-eastern corner of the macrocycle. One can draw a similar conclusion to the inclusion of CobF into the CbiL branch, since both enzymes methylate in the western periphery of the macrocycle, between ring A and the inverted ring D. Pairs of enzymes like CobA/CbiF or CobF/CbiL have more similar protein sequences presumably because the sequences reflect more closely related molecular recognition site.

## 5-4-2- Comparison of some CbiF/CobM sequences and identification of conserved residues.

The crystal structure of CbiF provided a unique opportunity to study the role of a number of conserved residues, but unfortunately there is no simple assay for this enzyme. To overcome this problem, advantage of the similarity between CbiF and CysG was utilised by comparing the role of key residues in the uro'gen III methyltransferase through site-directed mutagenesis [Woodcock *et al.* 1998]. The conserved residues were identified from an alignment of seven CbiF sequences, four CobM sequences and CysG (Figure 5-12).

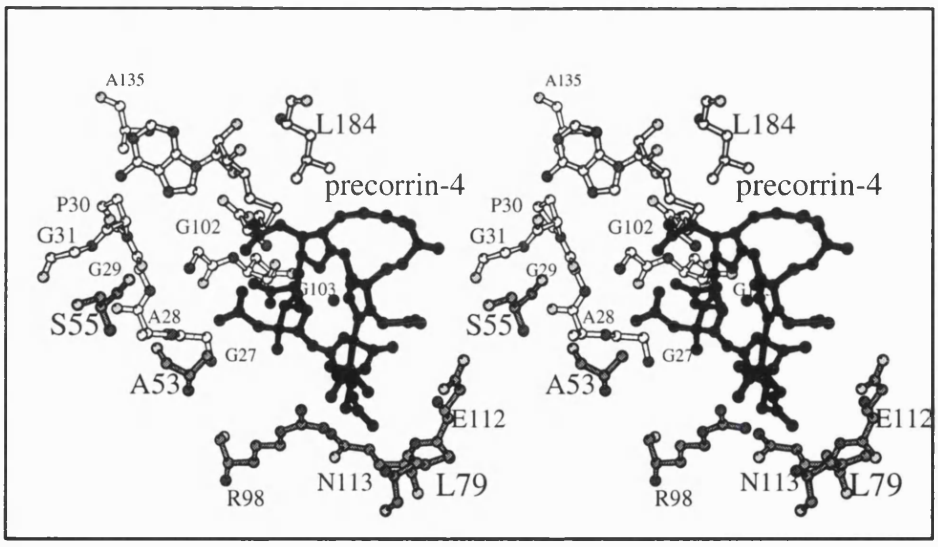
Figure 5-12- Alignment of CbiF with other methyltransferases involved in vitamin $B_{12}$ biosynthesis.

```
                                21         31         41         51         61
E. coli CysG                    ---------- -----GAGPG D--L-T-KGL ---QQADVV- Y-D-LVS-D-
B. meg CbiF                       M KLYIIGAGPG DPDLITVKGL KLLQQADVVL YADSLVSQDL
Consensus CbiF                  ---------- -----GAGPG D--L-T-K-- -----AD-V- YA-SLV----
Consensus CobM                  ---------- MTVFIGAGPG AADLIT-R-- ------PVCL YAGS-----L
Consensus methyltransferase     ----------VG-GPG -----T---- --------V- ----------

                                71         81         91         101        111
E. coli CysG                    -------A-V -K-AG-H-EE -----L---- -GK-VVR--- GDP---G---
B. meg CbiF                     IAKSKPGAEV LKTAGMHLEE MVGTMLDRMR EGKMVVRVHT GDPAMYGAIM
Consensus CbiF                  --------E- -------LE- ---------- -GK-VVR--- GD---YGA--
Consensus CobM                  L--CP-GA-- --T----L-I -------A-A -G-DVARLHS GD-S--SA--
Consensus methyltransferase     ---------- ---------- ---------- ----V----- GDP--Y----

                                121        131        141        151        161
E. coli CysG                    E----L---G -----VPG-T ------A--- --LT--D--Q -V-L------
B. meg CbiF                     EQMVLLKREG VDIEIVPGVT SVFAAAAAAE AELTIPDLTQ TVILTRAEGR
Consensus CbiF                  EQ---L---- I-----PGVS ---AAAA--- -E-T-P---Q ----TR--G-
Consensus CobM                  EQ-R-L---- --Y---PGVP -F-AAA--L --ELT-P-V-Q ---LTR----
Consensus methyltransferase     ---------- ----VIPGV- ---------- ---------- ----------

                                162                                        211
E. coli CysG                    -T---E---- --LA--K-T- ---------- ------I--G -----PV--V
B. meg CbiF                     .TPVPEFEKL TDLAKHKCTI ALFLSATLTK KVMKEFINAG WSEDTPVVVV
Consensus CbiF                  -T--P--E-L --LA------ ---L------ ---------- -----P-AV-
Consensus CobM                  ----P--E-L ---------- --HL------ -----L---- Y----P-A-V
Consensus methyltransferase     ---------- ---------- ---------- ---------- ----------

                                212                                        261
E. coli CysG                    ---T------ ---T---L-- ---------- --I--G---- ----------
B. meg CbiF                     YKATWPDEKI VRTTVKDLDD AMRTNGIRKQ AMILAGWALD PHIHDKDYRS
Consensus CbiF                  --A-W-DE-- ---TL-D--- -----GI--- A-I--G---- ---------S
Consensus CobM                  --AS-P---- -R-TL----- ---------T A---VG--L- ---F-----S
Consensus methyltransferase     ---------- ---------- ---------- ---------- ----------

                                262            291
E. coli CysG                    ---------- --------
B. meg CbiF                     KLYDKTFTHG FRKGVKSE
Consensus CbiF                  -LY-----H- -R------
Consensus CobM                  -LY------- --------
Consensus methyltransferase     ---------- --------
```

The amino acids printed in red are conserved residues involved in SAM binding; in green, residues involved in precorrin binding; in blue, residues involved in the dimer interactions and finally, in purple, residues which make up either the hydrophobic core or tight turns. The motifs SxLY only found in CbiF/CobM and the histidine residue specific to the CbiF is indicated in gold.

Thirty-eight amino acids are conserved between all CbiF primary sequences. Nine of them are involved in interactions with S-adenosyl-L-homocysteine (or SAM binding): Gly 27, Ala 28, Gly 29, Pro 30 and Gly 31 (the SAM binding site), Gly 102, Asp 103, Ala 135 and Leu 184. These residues are all conserved among the 53 methyltransferases except Ala 28, 135 and Leu 184. Six amino acids constitute a putative precorrin binding site and finally six amino acids are involved in dimer interactions (Figures 5-12, 5-13 and Table 5-1). Amino acids involved in forming the hydrophobic core or in tight turns are indicated in Figure 5-12.

Figure 5-13- Cobalt precorrin-4 and SAM binding sites (stereo representation).




Cobalt precorrin-4 is shown in back. The amino acids printed in grey are conserved residues involved in the precorrin binding site and in white, residues involved in the SAM binding site.

Table 5-1- Comparison of conserved residues in CbiF with the other methyltransferases.

| Interaction | CbiF | CobA/CysG$^{A}$ | CobI/CbiL | CobJ/CbiH | CobL/CbiE |
|---|---|---|---|---|---|
| SAM | Gly 27 | + | + | + | + |
| | Ala 28 | + | - | - | - |
| | Gly 29 | + | + | + | + |
| | Pro 30 | + | + | + | +/- |
| | Gly 31 | + | + | + | +/- |
| | Gly 102 | + | + | + | + |
| | Asp 103 | + | + | + | + |
| | Ala 135 | + | + | + | - |
| | Leu 184 | - | - | - | - |
| precorrin | Ala 53 | Asp | - | - | - |
| | Ser 55 | Leu | - | - | Arg |
| | Leu 79 | Gln | - | Glu | - |
| | Arg 98 | + | - | - | - |
| | Glu 112 | + | - | - | - |
| | Gln 113 | Glu | Leu | - | Leu |
| dimer | Thr 37 | + | + | + | + |
| | Glu 143 | - | Pro | Pro | Pro |
| | Leu 144 | Leu/Val | + | + | - |
| | Gln 151 | - | Leu | Ile | - |
| | Thr 156 | - | - | - | - |
| | Arg 157 | Thr | - | Leu | - |

The symbol, +, indicates that the amino acid is conserved in the different classes of methyltransferases. The symbol, -, means that the amino acids found at that position in other methyltransferases are different.

By comparison with the results obtained with CysG [Woodcock *et al.* 1998], it is known that the amino acid substitution of the equivalent of Gly 29 in CbiF (CysG$^{G224A}$) disrupts the SAM binding ability. The same phenotype is observed with the mutant CysG$^{R298L}$, equivalent to Arg 98 in CbiF. From the structure of CbiF, it is clear that Arg 98 does not interact with SAM but with precorrin. This

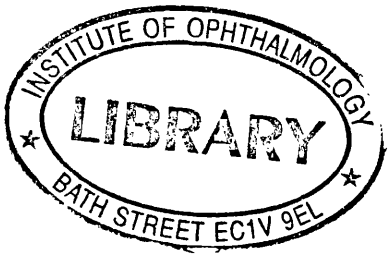
mutation could create a larger structural modification and therefore prevent SAM binding. More surprising is the observation that the CysG D303A mutation (corresponding to CbiF Asp 103) is still able to bind SAM. Finally, two CysG mutations D248A (CbiF 54) and R309L (CbiF 109) have a reduced activity, which is most likely due to the disruption of the cleft where precorrin binds.

## 5-5- Conclusion.

The *B. megaterium cbiF* has been subcloned from the main *B. megaterium cobI* operon into an expression vector, such that the gene is expressed fused to a histidine rich peptide. This allows the recombinant protein to be over-produced to a level of about 20 mg/l of culture, permitting its purification in one step through a metal affinity column. The His-tagged CbiF protein was shown to complement a *S. typhimurium cbiF* mutant strain, thereby proving that the His-tagged protein is physiologically active. The protein has also been tested for its ability to bind SAM, but this interaction is not tight enough to retain the substrate upon gel filtration. Two possible explanations can be forwarded for the lack of SAM binding; (i) the protein conformation is not optimum for SAM binding due to the absence of the second substrate, cobalt precorrin-4 or (ii) CbiF can catalyse the transformation of [$^3$H-methyl]-SAM into S-adenosyl-homocysteine, therefore loosing the labelled methyl group. In support of this latter suggestion, it was observed that the CbiF crystals were only found to contain bound S-adenosyl-homocysteine even though the protein was initially incubated in the presence of SAM.

The crystallisation of CbiF and its subsequent structure determination has provided an important insight into its structure/function relationship. Moreover, the sequence similarity shared between all the methyltransferases involved in $B_{12}$ biosynthesis means that the structure of CbiF can act as a frame onto which all these proteins can be modelled. Sequence alignments have identified a number of conserved amino acids also common with CysG which had been the object of extensive studies. Their functional roles within the structure of CbiF have been analysed.

Finally, a phylogenetic tree was generated from all available $B_{12}$ biosynthetic methyltransferases. Each type of methyltransferase segregates into one of two sub-groups, representing either the aerobic or anaerobic cobalamin pathway.



# CHAPTER 6

## PRODUCTION AND ANALYSIS OF A LATE COBALAMIN INTERMEDIATE, COBYRIC ACID, FROM GENETICALLY ENGINEERED *E. coli.*

## 6-1- Introduction.

The first intermediate along the cobalamin pathway that *S. typhimurium metE cysG*, the indicator strain used for the bioassay test, can take up and transform into methyl-cobalamin is cobyric acid [Raux *et al.* 1996]. This intermediate is comparatively easy to detect, which is important for the purification of this compound. Cobyric acid was therefore the obvious choice as the first compound to characterise in an attempt to systematically analyse all the vitamin $B_{12}$ pathway intermediates.

Cobyric acid also represents the point at which the aerobic and anaerobic pathways converge to a unique route, after the divergence of the pathways at precorrin-2, to accommodate the different processes of ring contraction and cobalt insertion.

In this chapter, the production and purification of cobyric acid produced from [$^{13}$C]-ALA are described. Cobyric acid was synthesised from an *E. coli hemA* mutant host strain, containing the *S. typhimurium cob* operon, in order to obtain a high incorporation of $^{13}$C label. The purified $^{13}$C-cobyric acid was characterised by Nuclear Magnetic Resonance (NMR) and Electron Paramagnetic Resonance (EPR).

## RESULTS

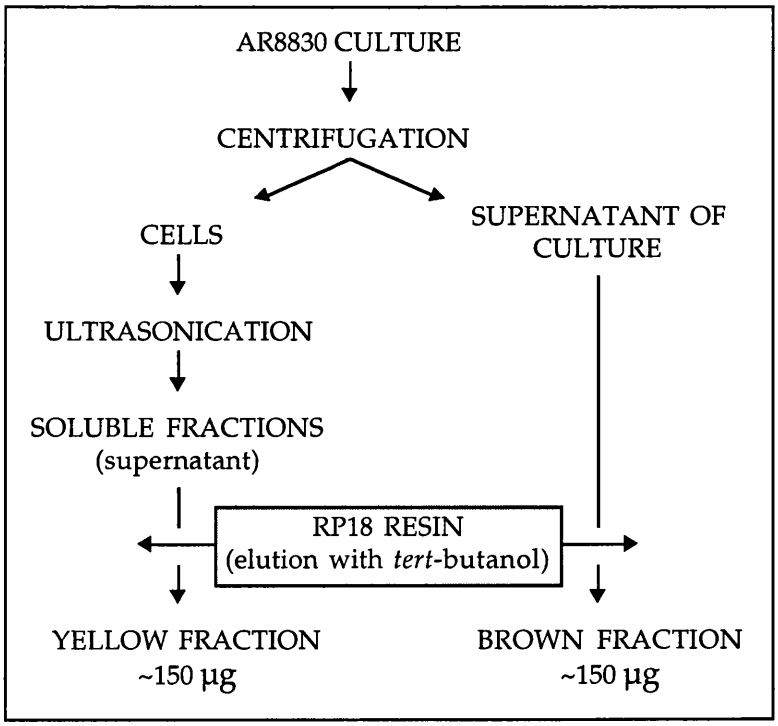
## 6-2- Production and isolation of cobyric acid from genetically engineered *E. coli*.

Cobyric acid was produced from *E. coli* that was initially transformed with pAR8827 (pACYC184 harbouring all the *S. typhimurium cobI* genes except *cbiB*) and subsequently with a second compatible plasmid harbouring *cysG*. The resultant strain was named AR8830. A starter culture of the strain was grown overnight in LB medium and was used to inoculate a 1 litre flask of minimal medium supplemented with 5% LB and 1 mg/l $CoCl_2$, $6H_2O$. Minimal medium was used to ensure that no exogenous vitamin $B_{12}$ from the tryptone

component of LB was taken up. After 5 hours at 37°C, ALA (10 mg/l) and IPTG (0.4 mM) were added to the culture. The cells were harvested after 24 hours anaerobic growth at 37°C and lysed by ultrasonication. The soluble fraction was applied to a 3 ml LiChroprep RP-18 column (N-dodecyloctylsilane modification) that had been previously equilibrated in distilled water. The preparation of the RP-18 resin is described in Materials and Methods (section 8-5-6). The column was then washed with 3 volumes of water and the corrin-containing fractions were eluted with 25% *tert*-butanol. All fractions were analysed for the presence of cobyric acid by microbiological assay. As shown in Figure 6-1, a yellow fraction, obtained from the RP18 column, was found to contain the majority of cobyric acid (about 150 µg). The yellow pigment was freeze-dried.

Cobyric acid was also found excreted into the medium in which the strain had been grown. Indeed, an equivalent amount of cobyric acid (150 µg) could be detected from the medium after the cells had been harvested.

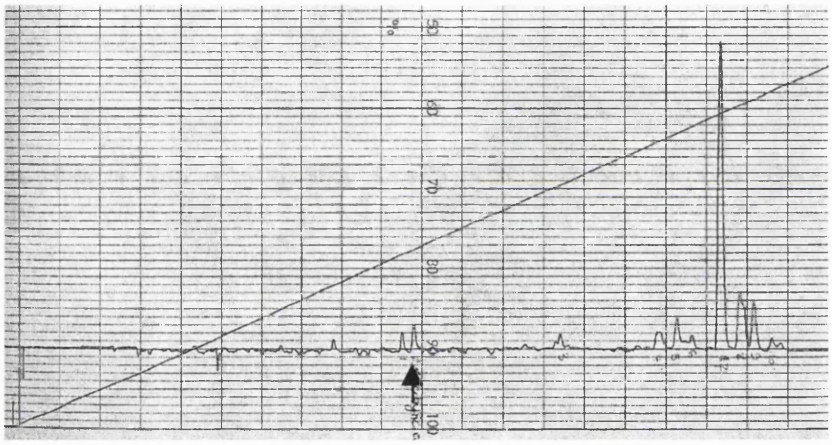
Figure 6-1- Isolation of cobyric acid by RP-18 from AR8830.



## 6-3- Purification of cobyric acid by High-Performance Liquid Chromatography (HPLC).

Cobyric acid could be further purified by HPLC using a reverse-phase C-18 column. The mobile phases were: buffer A, water with 0.1% TFA (trifluoroacetic acid) and buffer B, acetronitrile ($CH_3CN$) with 0.1% TFA. Cobyric acid extract was dissolved in buffer A. About 50 µg of sample were loaded on the C-18 column for each run. All the collected fractions were diluted 200 fold and analysed for the presence of cobyric acid (Figure 6-2).

Figure 6-2- HPLC profile and cobyric acid content from AR8830.




The diagonal line represents the gradient between buffer A and B. The arrow indicates the fraction containing cobyric acid. The absorbance was measured at 350 nm.

Cobyric acid was found to elute at about 22% $CH_3CN$/0.1% TFA. The HPLC trace obtained from the cell-retained cobyric acid sample was much purer than that obtained from the cell-excreted sample, which contained a large number of contaminants from the medium. Thus, in all further experiments only the cell-retained cobyric acid sample was used for analysis.

## 6-4- Optimisation of ALA incorporation in cobyric acid.

The next step in the characterisation of cobyric acid was the incorporation of $^{13}C$ in each carbon of the macrocycle. This required the availability of ALA with

$^{13}C$ at position C-3, C-4 or C-5 (Figures 6-4, 6-5 and 6-7). These three isotopes were generously provided by Prof. A. I. Scott, Texas, A & M University, U. S. A.

*E. coli* biosynthesises endogenous ALA by the C-5 pathway (Chapter 1, section 1-3). For the transformation of glutamate to ALA, three enzymes are required: glutamyl t-RNA synthetase, glutamyl t-RNA dehydrogenase (HemA) and glutamate 1-semialdehyde aminotransferase (HemL). To ensure that no endogenous unlabelled ALA was produced and that all [$^{13}C$]-ALA was incorporated in the tetrapyrrole, an *E. coli hemA* mutant, HU227 [Sasarman *et al.* 1968], was used as the host to express [$^{13}C$]-cobyric acid. HU227 was transformed with two plasmids, pER119 and pAR8827. These two plasmids contain all the genetic information for the synthesis of cobyric acid from uro'gen III, since they harbour the *P. denitrificans cobA* (pER119) and all *S. typhimurium cbi* genes except *cbiB* (pAR8827). This *E. coli hemA* derived strain was called ER210 and is auxotrophic for ALA and methionine, but is hemin permeable.

## 6-4-1- Culture conditions for strain ER210.

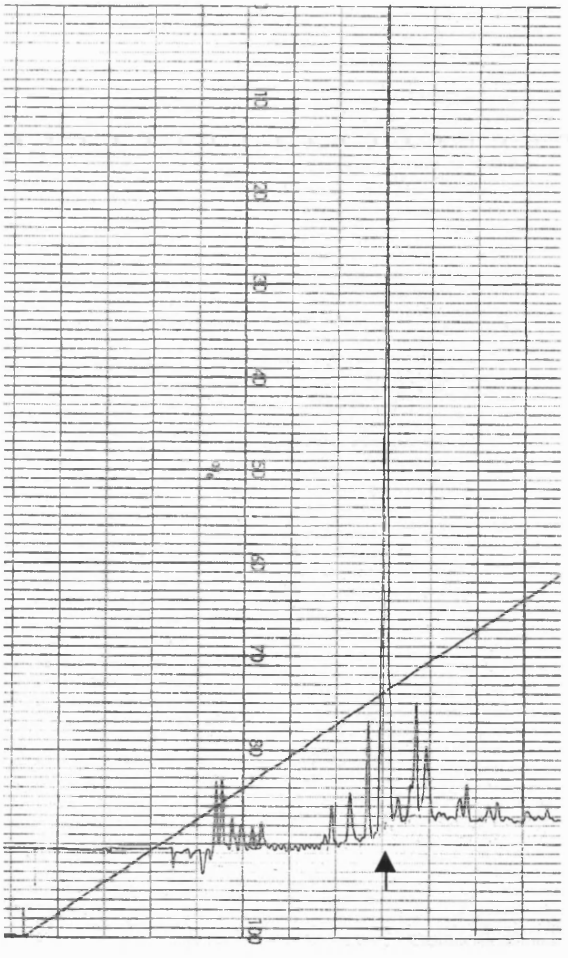
Strain ER210 was grown in 1 litre of minimal medium supplemented with 5% LB, 5 mg/l hemin, 50 mg/l methionine and 1 mg/l $CoCl_2$, $6H_2O$. The medium was inoculated from an overnight culture grown in LB with 10 mg/l ALA (ALA induces a better growth than hemin). After 5 hours at 37°C under anaerobic conditions, ALA was added at 10 mg/l and growth was continued for a further 24 hours.

When the *tac* promoter was induced by IPTG (which was added on the same time as ALA) to increase the level of transcription and thus the overexpression of the *cob* operon, the synthesis of cobyric acid decreased from 600 to 0.4 pmol/$OD_{600}$. Therefore, the production of cobyric acid from strain ER210 was undertaken in the absence of any inducer.

## 6-4-2- Purification of cobyric acid from strain ER210.

Cobyric acid was purified from strain ER210 in the same way as described for strain AR8830. In brief, the soluble fraction of the cell extract was passed through an RP-18 column and cobyric acid was eluted in a brown fraction and purified by HPLC using a C-18 column. As shown in Figure 6-3, the HPLC profile obtained from the brown fraction only has one major peak that contains cobyric acid. In fact, the brown fraction isolated after RP-18 is pure enough to allow a NMR study without requiring any further purification by HPLC.

Figure 6-3- HPLC profile and cobyric acid content from ER210.



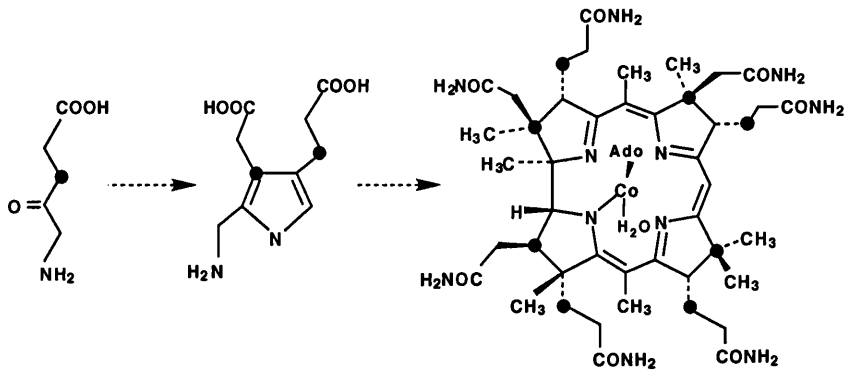
The diagonal line represents the gradient between buffer A and B. The arrow indicates the fraction containing cobyric acid. The absorbance was measured at 350 nm.

## 6-5- NMR studies of cobyric acid synthesised from strain ER210.

To analyse cobyric acid by NMR, the molecule was enriched with $^{13}C$. The incorporated $^{13}C$ behaves like a magnet which can be flipped from one energy state to another when under the influence of microwave radiation producing a resonance. If the resonance can be identified, it is possible to determine the conformation of a particular group or atom within cobyric acid.

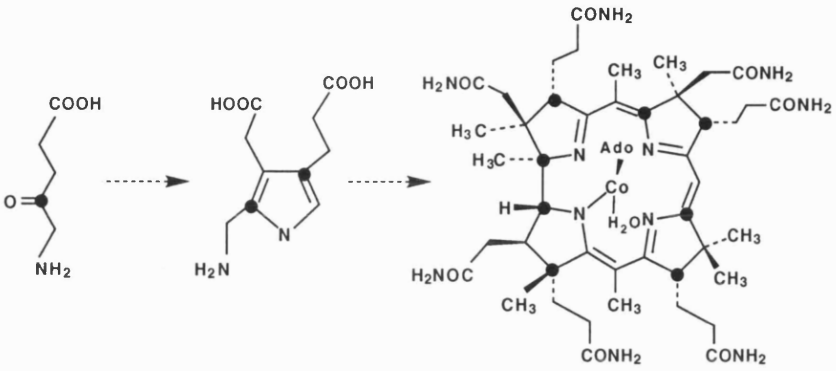
All the carbon atoms in the corrin macrocycle of cobyric acid can be enriched with $^{13}C$ by feeding strain ER210 with ALA labelled at positions C-3, C-4 and C-5. Figures 6-4, 6-5 and 6-7 illustrate how these labelled-ALA substrates are incorporated into cobyric acid. The NMR studies were performed by Dr Neal J. Stolowich, Texas A & M University, U. S. A. The actual NMR spectrum of cobyric acid derived from [4-$^{13}C$]-ALA is shown in Figures 6-6. The [3-$^{13}C$]-ALA and [5-$^{13}C$]-ALA NMR spectra are not shown.

Figure 6-4- Incorporation of $^{13}C$ into cobyric acid from [3-$^{13}C$]-ALA.



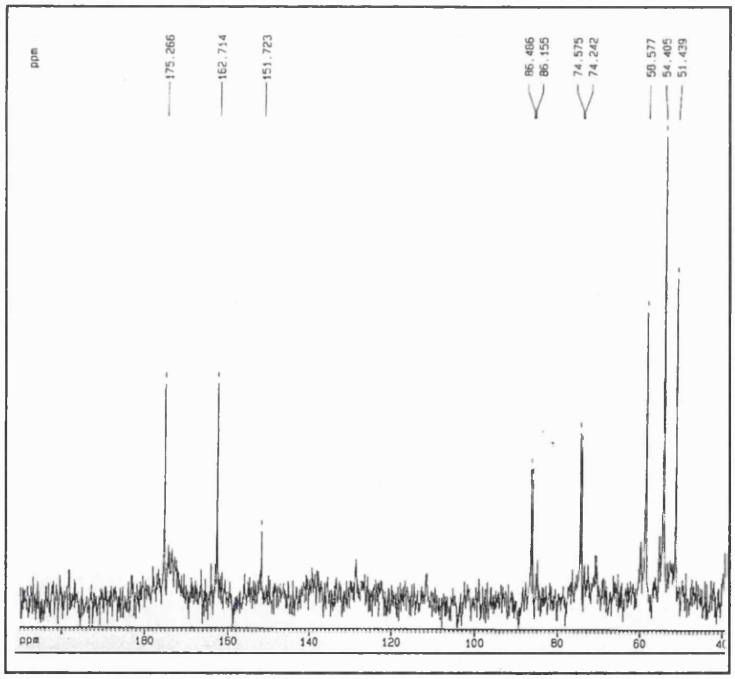

[3-$^{13}C$]-ALA PBG Cobyric acid

Figure 6-5- Incorporation of $^{13}C$ into cobyric acid from [4-$^{13}C$]-ALA.




[4-$^{13}C$]-ALA PBG Cobyric acid

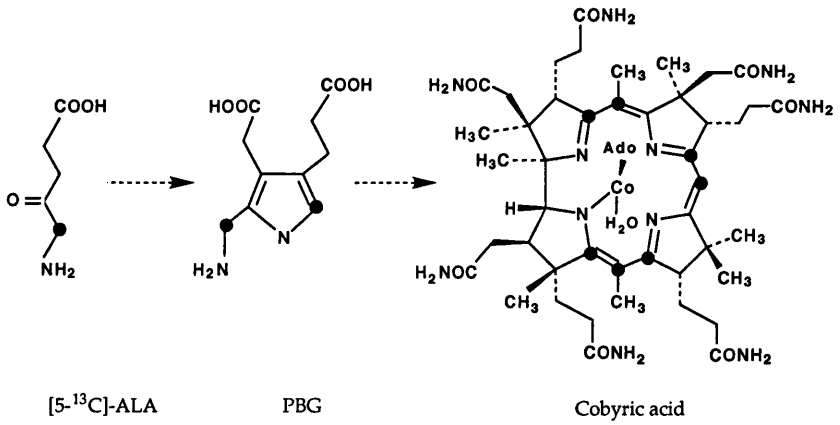
Figure 6-6- NMR spectra of cobyric acid synthesised from [4-$^{13}C$]-ALA.




Cobyric acid obtained from [4-$^{13}C$]-ALA gave a spectrum with two doublets at 74 and 86 ppm which confirms that C-1 and C-19 are directly bound and that the two carbons are in $sp^3$ conformations. This proves that the ring contraction had occurred. The peaks of resonance at 51 (C-8), 54 (C3 and C-13) and 58 (C17) ppm are consistent with $sp^3$ carbon conformations whereas the peaks at 162 (C-6) and 175 (C-11) are typical of $sp^2$ carbons. The fact that the C-17 is $sp^3$ indicates that the methylation at that carbon had occurred. The NMR spectrum analysis

confirmed that the C-1, C-3, C-6, C-8, C-11, C-13, C-17 and C-19 conformations proposed for cobyric acid are correct.

Figure 6-7- Incorporation of $^{13}C$ into cobyric acid from [5-$^{13}C$]-ALA.




[5-$^{13}C$]-ALA PBG Cobyric acid

The NMR studies of cobyric acid labelled with [3-$^{13}C$]-ALA and [5-$^{13}C$]-ALA completed the characterisation of the macrocycle, which is in agreement with the proposed structure of cobyric acid as isolated from *P. denitrificans* [Blanche *et al.* 1991].

## 6-6- EPR studies of cobyric acid synthesised from strain ER210.

Electron Paramagnetic Resonance (EPR), also called Electron Spin Resonance (ESR) or Electron Magnetic Resonance (EMR) is a technique that can detect spin-state changes in substances containing at least one unpaired electron in the presence of a static magnetic field. EPR is one of the primary tools to investigate metals in biological systems and can be used to quantify and identify paramagnetic transition metal ions and enzyme prosthetic groups containing these ions.

In organic compounds such as cobalamin, cobalt is present in the strongly-bound diamagnetic $Co^{3+}$ state which does not produce an EPR signal. However, after photolysis, a signal attributed to $Co^{2+}$ can be obtained due to the break of the bond between cobalt and the adenosyl ligand. In the case of the cyanide

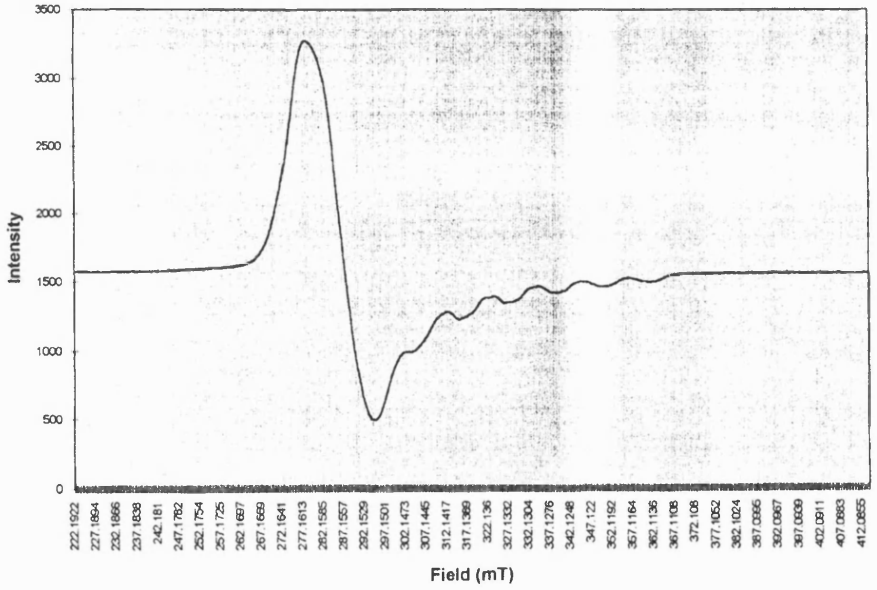
form, the bond is stronger and can not be disrupted by photolysis. Free inorganic cobalt does not induce an EPR signal. [Dr P. Heathcote, pers. commun.]. No information is available concerning the EPR signal of cobalt after photolysis of methyl cobalamin.

Cobyric acid, isolated from strain ER210 and purified by HPLC, was submitted to EPR analysis. This was performed by Dr P. Heathcote, Queen Mary and Westfield College, London.

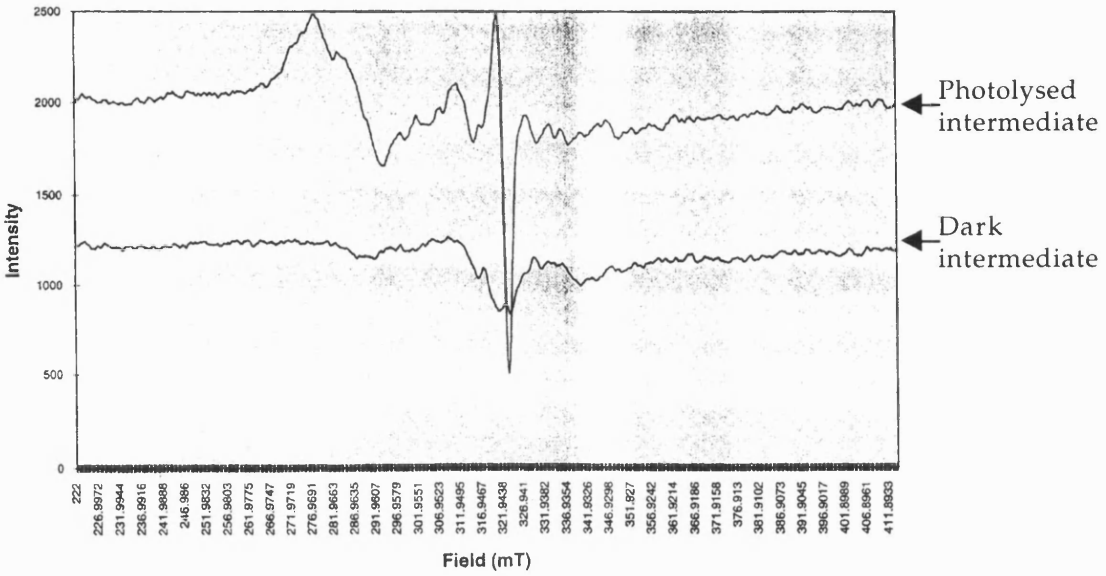
The EPR studies on cobyric acid after photolysis revealed an EPR signal consistent with the presence of $Co^{2+}$ in the molecule. This result clearly demonstrates that cobyric acid isolated from ER210 contains a cobalt-carbon bond that is susceptible to photolysis (Figure 6-8). In parallel, electrospray mass spectrometry was undertaken by Dr Russell Jones, at the University of Leicester, on the compound previously used for EPR. Cobyric acid gave a mass of 1181 g/mol, consistent with the expected mass of the adenosyl form (adenosyl cobyric acid has an expected molecular weight of 1183 g/mol).

Figure 6-8- EPR spectra.

a- EPR spectrum of adenosyl-cobalamin after photolysis.



b- EPR spectra of cobyric acid produced from strain ER210.




Cobyric acid samples extracted from strain ER210 were kept anaerobically either in the dark for 5 minutes and subsequently frozen in liquid nitrogen or illuminated for 2 minutes with a 1000 W halogen lamp and subsequently frozen. The EPR signal after photolysis changes drastically and gives a similar spectrum (Figure b, upper spectrum) than with adenosyl-cobalamin (Figure a).

## 6-7- Conclusion.

Cobyric acid has been produced from a genetically engineered *E. coli.* An *E. coli* host strain has been co-transformed with the *P. denitrificans cobA*, which increases the level of precorrin-2 in the cell, and with the *S. typhimurium cobI* operon, which contains the genetic information necessary to produce cobyric acid from precorrin-2.

The structure of cobalamin biosynthetic intermediates can be determined by the use of $^{13}C$ NMR. However, this requires the incorporation of $^{13}C$ into the molecule. To ensure a high incorporation of $^{13}C$-labelled ALA into cobyric acid, an *E. coli hemA* host strain was used. After analysis of the NMR spectra obtained with cobyric acid derived from [3-$^{13}C$]-ALA, [4-$^{13}C$]-ALA and [5-$^{13}C$]-ALA, the whole carbon ring of cobyric acid was investigated and found to be identical to the formula published by Blanche *et al.*, 1991. EPR study confirmed that cobyric

acid contained a carbon-bound central cobalt and mass spectral analysis indicated that the extracted intermediate was the adenosyl form of cobyric acid.

The first attempt to produce a vitamin $B_{12}$ intermediate, cobyric acid, from a genetically modified *E. coli* and also to isolate and characterise that compound by NMR was successful and seems to be an appropriate and powerful method to study other vitamin $B_{12}$ intermediates from the anaerobic pathway. However, cobyric acid was the easiest intermediate to study. The precursors of cobyric acid, via the anaerobic pathway are mostly unknown and their structure determination is likely to prove more technically challenging.

# CHAPTER 7

# GENERAL DISCUSSION

The biosynthesis of cobalamin is often referred to as the Mount Everest of biosynthetic problems [Battersby, 1994]. Up until 1989 little had been achieved on the understanding of corrin biosynthesis but in a remarkable period from 1989 to 1994, the step-by-step synthesis of cobalamin was almost completely elucidated for the industrial $B_{12}$ over-production strain, *Pseudomonas denitrificans*. To achieve this, a range of classic and modern technologies including genetics, molecular biology, biochemistry, chemistry and biophysics were employed. The research required the concerted action of three labs, those of Rhone-Poulenc Rorer in France, Sir A. Battersby in England and Prof. A. I. Scott in the U. S. A. The choice of organism proved to be critical since, and despite great efforts, no progress in the pathway between precorrin-3 and cobyrinic acid from other bacteria had been achieved during the same time [Blanche *et al.* 1995]. It is only recently that cobalt precorrin-4 has been isolated from *P. shermanii* [Scott *et al.* 1996]. The determination of the cobalamin pathway in *P. denitrificans* not only confirmed the complexity of the synthesis, it also highlighted that one step requires molecular oxygen and that cobalt is added at a late stage, after the completion of the ring contraction process. The results with *P. denitrificans* were such that it became apparent that other bacteria such as *S. typhimurium*, *P. shermanii* and the methanogens must employ an alternative way to synthesise their cobalamin. Indeed these bacteria are able to make $B_{12}$ *de novo* in the absence of oxygen and experiments with radio-labelled cobalt proved that cobalt was inserted at the level of precorrin-2 or -3, prior to the ring contraction step. The alternative way to synthesise cobalamin was therefore designated the "anaerobic pathway".

*Bacillus megaterium* synthesises $B_{12}$ under both aerobic and anaerobic environments (data not shown, in collaboration with Dr D. Jahn, Freiburg, Germany). However, when the *B. megaterium cobI* operon was transferred into *E. coli*, the host strain was only able to make corrins anaerobically. The genetic composition of the *B. megaterium cobI* operon and its ability to complement *S. typhimurium cob* mutants clearly suggested that it belonged to the anaerobic pathway class. Thus, in contrast to *S. typhimurium*, the anaerobic pathway of *B. megaterium* can operate aerobically and is not inhibited by the presence of oxygen.

Table 7-1- Genes involved in cobalamin biosynthesis in 12 micro-organisms.

| | *P. den* | *R. caps* | *M. tub* | *P. aer* | *S. typ* | *B. meg* | *Syn.* | *C. ace* | *P. gin* | *M. jan* | *M. th* | *A. ful* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Respira-tion | $O_2$ | AAE | $O_2$ | AAE | AAE | AAE | Ana | Ana | Ana | Ana | Ana | Ana |
| ALA | Shemin | Shemin | Glu | Glu | Glu | Glu | Glu | Glu | ? | Glu | Glu | Glu |
| C-2 & C-7 methylase | CobA p21631 | CobA rrc00657 | CysG Rv0511 | CysG | CysG P25924 | $CysG^A$ | $CysG^A$ sll0378 | $CysG^A$-HemD | | $CysG^A$ MJ0965 | $CysG^A$ mth0167 | $CysG^A$ AF0422 |
| | | | CysG2 Rv2847c | $CysG^A$ | | | | | | | | $CysG^A2$ AF1243 |
| Cobalt chelatase (no ATP) | | | | | CbiK q05592 | | | CbiK | CbiK | | | |
| | | | | | | | | CbiK2 | | | | |
| Cobalt chelatase (no ATP) | | | ? Rv2393 | | | CbiX | CbiX sll0037 | | | ? MJ0970 | ? mth1397 | |
| C-20 methylase | CobI p21639 | CobI rrc00651 | CobIJ Rv2066 | CobI | CbiL q05593 | CbiL | CbiL slr1879 | CbiL | CbiL | CbiL MJ0771 | CbiL mth1348 | CbiL AF0727 |
| mono-oxygenase | CobG p21637 | | CobG Rv2064 | CobG | | | | | | | | |
| C-17 methylase | CobJ p21640 | CobJ rrc00652 | CobIJ Rv2066 | CbiGH | CbiH q05590 | $CbiH_{60}$ | CbiGH slr 0969 | CbiH | CbiH | CbiH MJ0813 | CbiH mth1403 | CbiHC AF0724 |
| ? | CobE p21635 | CobE rrc00655 | | CobE | | | | | | | | |
| *cobI* | | | | CbiGH | CbiG q05631 | CbiG | CbiGH slr0969 | CbiG | CbiGF | CbiG MJ1144 | CbiG mth1408 | CbiG AF0725 |
| C-11 methylase | CobM p21922 | CobM rrc00656 | CobM Rv2071c | CobM | CbiF q05630 | CbiF | CbiF slr0239 | CbiF | CbiGF | CbiF MJ1578 | CbiF mth0602 | CbiF AF0726 |
| C-1 methylase | CobF p21636 | CobF rrc00658 | | | | | | | | | | |
| *cobI* | | | | CbiD | CbiD q05628 | CbiD | CbiD slr1438 | CbiD | CbiJD | CbiD MJ0022 | CbiD mth0808 | CbiD AF0723 |
| precorrin 6a reductase | CobK p21920 | CobK rrc00653 | CobK Rv2070c | CobK | CbiJ q05591 | CbiJ | CbiJ slr0252 | CbiJ | CbiJD | CbiJ MJ0552 | CbiJ mth1002 | ? AF0256 |
| C-5 & C-15 methylase | CobL p21921 | CobL rrc00654 | CobL Rv2072c | CobL | CbiE q05629 | CbiET | CbiET sll0099 | CbiE* | CbiET | CbiE MJ1522 | CbiE mth1514 | CbiE AF0722 |
| decarboxy-lase | | | | | CbiT q05632 | | CbiT slr1368 | CbiT | | CbiT MJ0391 | CbiT mth0146 | CbiT AF0732 |
| precorrin 8x isomerase | CobH p21638 | CobH rrc00650 | CobH Rv2065 | CobH | CbiC q05601 | CbiC | CbiC sll0916 | CbiC | CbiC | CbiC MJ0930 | CbiC mth0227 | CbiHC AF0724 |
| hydrogeno-byrinic a,c-diamide synthase | CobB p21632 | CobB rrc00666 | CobB Rv2848c | CobB | CbiA p29946 | CbiA | CbiA sll1501 | CbiA | CbiA | CbiA MJ1421 | CbiA mth1460 | CbiA AF2229 |
| | | | | | | | | | | | CbiA2 mth1497 | |

| | *P. den* | *R. caps* | *M. tub* | *P. aer* | *S. typ* | *B. meg* | *Syn.* | *C. ace* | *P. gin* | *M. jan* | *M. th* | *A. ful* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cobalt chelatase | CobN p29929 | CobN rrc00649 | CobN Rv2062c | CobN | | | CobN slr1211 | | | CobN MJ0908 | ? mth1363 | |
| | | | | | | | ? sll1890 | | | | | |
| Cobalt chelatase | CobS p29933 | CobS rrc01954 | | | | | | | | | | |
| Cobalt chelatase | CobT p29934 | CobT rrc01953* | | | | | ? sll1954 | | | | ? mth0354 | |
| Reductase | N.A. | | | | | | | | | | | |
| adenosyl trans-ferase | CobO p29930 | BtuR rrc00646 | BtuR Rv2849c | BtuR | CobA p31570 | BtuR | BtuR slr0260 | | | | | |
| cobyric acid synthase | CobQ p29932 | CobQ rrc00645 | CobQ Rv0255c | CobQ | CbiP q05597 | | CbiP slr0618 | CbiP | CbiP | CbiP MJ0484 | CbiP mth0787 | CbiP AF1338 |
| | | | CobQ2 Rv3713 | | | | | | | | | |
| cobina-mide synthase | CobD p21634 | CobD rrc00706 | CobD Rv2236c | CobD | CbiB q05600 | - | CbiB slr1925 | CbiB | CbiB | CbiB MJ1314 | CbiB mth1409 | CbiB AF1336 |
| cobina-mide synthase | CobC p21633 | CobC rrc00642 | CobC Rv2231c | CobC | CobD u90625 | | ? slr1713 | CobD | CobD | ? MJ0955 | CobD mth1587 | ? AF2002 |
| | | | | | | | | | | | | ? AF2024 |
| cobina-mide synthase | protein α | | | | | | | | | | | |
| cobina-mide kinase | CobP p29931 | CobP rrc00591 | CobP Rv0245c | CobP | CobU q05599 | | CobU slr0216 | CobU | CobU | | | |
| αribazole phosphate synthase | CobU p29935 | CobU rrc00590 | CobU Rv2207 | CobU | CobT q05603 | | | CobT | CobT | | | |
| αribazole phosphate phosphata-se | | | ? Rv2228c | CobC | CobC p39701 | | ? sll0395 | CobC | | | | |
| | | | | | | | ? slr1748 | | | | | |
| cobalamin synthase | CobV p29936 | CobV rrc00589 | CobV Rv2208 | CobV | CobS q05602 | | CobS slr0636 | CobS* | CobS | CobS MJ1428 | CobS mth1112 | CobS1 AF0037 |
| | | | | | | | | | | | | CobS2 AF2323 |
| ? | CobW p29937 | CobW rrc00648 | ? Rv0106 | CobW | | | CobW slr0502 | | | | | |

| | *P. den* | *R. caps* | *M. tub* | *P. aer* | *S. typ* | *B. meg* | *Syn.* | *C. ace* | *P. gin* | *M. jan* | *M. th* | *A. ful* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $Co^{2+}$ transport | | CbiM rrc00659 | | | CbiM q05594 | | CbiM sll0383 | CbiM | | CbiM MJ1091 | CbiM mth0130 | CbiM1 AF0728 |
| | | | | | | | | | | CbiM2 MJ1569 | CbiM2 mth1707 | CbiM2 AF1843 |
| $Co^{2+}$ transport | | CbiN rrc00660 | | | CbiN q05595 | | | | | CbiN MJ1090 | CbiN mth0131 | CbiN AF0729 |
| $Co^{2+}$ transport | | CbiO rrc00662 | ? Rv2564 | | CbiO q05596 | | ? sll0385 | CbiO | | CbiO MJ1088 | CbiO1 mth0133 | CbiO1 AF0731 |
| | | | | | | | | | | | CbiO2 mth1704 | CbiO2 AF1841 |
| $Co^{2+}$ transport | | CbiQ rrc00661 | | | CbiQ q05598 | | ? sll0384 | CbiQ | | CbiQ MJ1089 | CbiQ mth0132 | CbiQ1 AF0730 |
| | | | | | | | | | | | | CbiQ2 AF1842 |
| ? | | | | | | CbiW | CbiW sll1584 | | | | | |
| ? | | | ? Rv0306 | ? BluB | | CbiY | | | | | ? mth0120 | ? AF2267 |
| *cobII or III* | | BluF rrc00592 | ? Rv2135c | | | | | | | | | |
| phospho threonine synthase | | BluB rrc00643 | BluB Rv0306 | ? BluB | | | | | | | ? mth0120 | |
| *cobI* | | BluE rrc00593 | | | | | | | | | | |
| ? | | ? rrc00663 | ? Rv1817 | | | | | | | | | |

The respiration type is indicated as follows: $O_2$ means strict aerobe; AAE is aero-anaerobe and Ana, strict anaerobe. The row marked ALA indicates the way that the bacteria synthesise amino laevulinic acid. Shemin is the C-4 pathway and Glu (glutamate) is the C-5 pathway. A star indicates that the database contains sequencing errors. NA means not available in the database.

P. den: *P. denitrificans*; R. caps: *R. capsulatus*; M. tub: *M. tuberculosis*; S. typ: *S. typhimurium*; B. meg: *B. megaterium*; Syn. : *Synechocystis sp.*; C. ace: *C. acetobutylicum*; P. gin: *P. gingivalis*; M. jan: *M. jannaschii*; M. th: *M. thermoautotrophicum*; A. ful: *A. fulgidus*; P. aer: *P. aeruginosa*.

This table has been achieved using the following web pages:
http://www.ncbi.nlm.nih.gov/BLAST/unfinishedgenome.html
http://capsulapedia.uchicago.edu/capsulapedia/capsulapedia/capsulapedia.shtml
http://www.biosci.ohio-state.edu/~genomes/mthermo/mthermo_data/Porphyrin.html
http://www.sanger.ac.uk/Projects/M_tuberculosis/
http://www.pseudomonas.com/obtaining_data.html
http://www.kazusa.or.jp/cyano/cyano.orig.html
http://www.genomecorp.com/genesequences/clostridium/clospage.html
http://www.forsyth.org/pggp/
http://www.tigr.org/tdb/mdb/mjdb/mjdb.html

[Blanche *et al.* 89, Bult *et al.* 96, Cameron *et al.* 91, Cole *et al.* 98, Crouzet *et al.* 90 (3), Kaneko *et al.* 95 and 96, Klenk *et al.* 97, O'Toole *et al.* 94, Raux *et al.* 98b, Roth *et al.* 93, Smith *et al.* 97 and Vleck *et al.* 97]

## 7-1- *cbiD* and *cbiG* are the genetic hallmarks of the anaerobic pathway.

Two genes, *cbiG* and *cbiD*, found within the *cob* operons of *S. typhimurium* and *B. megaterium*, are significantly absent in *P. denitrificans* and have therefore been classified as the genetic hallmarks of the anaerobic pathway. Indeed in both the *S. typhimurium* and *B. megaterium* operons, their absence has proven to be detrimental to the biosynthesis of $B_{12}$, and are clearly involved in reactions between uro'gen III and cobyric acid. These two genes are found in all anaerobic $B_{12}$ producers, such as *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Porphyromonas gingivalis*, *Clostridium acetobutylicum*, *Archaeoglobus fulgidus* and *Synechocystis sp.* (Table 7-1).

## 7-2- Phylogenetic relationship of the methyltransferases.

Both aerobic and anaerobic cobalamin biosynthesis requires the peripheral addition of eight SAM derived methyl groups to the tetrapyrrole framework. This task is performed by the action of six methyltransferases, $CysG^A$/CobA (C-2 and C-7), CbiL/CobI (C-20), CbiH/CobJ (C-17), CbiF/CobM (C-11), ?/CobF (C-1), CbiE/CobL (C-5 and C-15). The final molecule, cobalamin, contains only seven peripheral methyl groups since the methyl group attached to C-20 is extruded during the ring contraction event. The sequences of all these methyltransferases have been used to generate a phylogenetic tree in order to understand how these enzymes are related. The methyltransferases are found to segregate into five distinct branches, where each branch represents a different type of methyltransferase. Moreover, within each branch, there is a clear division of aerobic and anaerobic enzymes. This can be explained on the basis of substrate specificity since the anaerobic pathway enzymes recognise cobalt-complexed intermediates whereas the aerobic enzymes utilise cobalt free intermediates. For the C-2/C-7 methyltransferases, uro'gen III is the unique and universal substrate via both pathways and four proteins from the aerobic pathway form a sub-branch. However, the gene duplications within this class do not allow a definite interpretation. Interestingly the methyltransferases that add a methyl group at similar sites on the modified tetrapyrrole framework appear more closely related. For instance, the enzymes that methylate at C1 (CobF) and C20 (CbiL) are part of the same branch whilst the enzymes that methylate at C2/C7 ($CysG^A$) and C11 (CbiF) are more closely rooted to each

other than to the other methyltransferases. These results emphasise the evolutionary separation of the two pathways and demonstrate that the methyltransferases have segregated according to their substrate specificity.

The crystal structure of the *B. megaterium CbiF*, the protein that catalyses the attachment of a methyl group to C-11 to generate cobalt-precorrin-5, has been determined. The three-dimensional structure of CbiF represents a new class of methyltransferases, the $B_{12}$ methyltransferases, since the topology is different to the methyltransferases that have been previously determined for DNA methylation [Schluckebier *et al.* 1995]. The protein contains two $\alpha/\beta$ domains, forming a cleft in which SAM and cobalt-precorrin-4 are thought to bind.

Since all the corrin biosynthetic methyltransferases are evolutionary related, they can be modelled onto the structure of CbiF. The SAM binding site of all methyltransferases is almost identical since they all share the common motif: $VGxGPGx_6T$, $Vx_5GDPx_2Y$ and VVPGV. Therefore, the SAM binding pocket must have a very similar configuration for all proteins and the tetrapyrrole-derived intermediates will be positioned by the different class of enzymes such as they can face the methyl group donor. Only during the C-17 methylation event is the precorrin molecule positioned in an alternative configuration, up side down, in comparison with the other precorrins. The structure of CbiF therefore gives the first proper structure/function study of the $B_{12}$ family of biosynthetic methyltransferases.

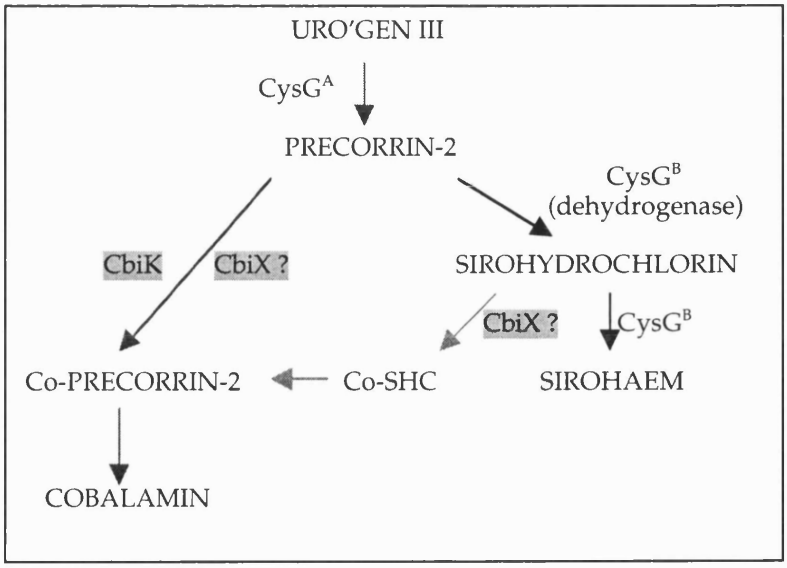
**7-3- Cobalt insertion.**

In the aerobic pathway, cobalt is inserted to hydrogenobyrinic a-c, diamide by an ATP-dependent enzyme complex formed between CobN, S and T. The complex shares similarities with the Bch/ChlH, I and D complex required for magnesium chelation of protoporphyrin IX for bacteriochlorophyll and chlorophyll biosynthesis. This first specific step for the biosynthesis of (bacterio)-chlorophyll is also ATP-dependent. None of the bacteria containing an anaerobic cobalamin pathway have proteins that display a high degree of similarity to the three peptides constituting the cobaltochelatase complex.

In *S. typhimurium* and *B. megaterium*, the anaerobic cobalt chelatases were identified as CbiK and CbiX respectively, and although they are both clearly isofunctional, the proteins display no similarity at the amino acid level. Both proteins were also shown to act as ferrochelatases in sirohaem synthesis. This reaction was considered a surrogate reaction since the sirohaem synthesis catalysed by CbiK/CbiX is inhibited by the presence of exogenous cobalt. Both activities, ferrochelation and cobalt chelation, can also be undertaken by *E. coli* sirohaem synthase, CysG [Spencer *et al.* 1993], which is found within the nitrite reductase operon of *S. typhimurium* [Griffiths and Cole, 1987]. Cobalt does not inhibit CysG in the synthesis of sirohaem.

In contrast to the aerobic route, cobalt chelation at the level of precorrin-2 does not require ATP. The determination of the X-ray structure of CbiK has shown that this protein shares a three-dimensional structure that is highly similar to the *B. subtilis* protoporphyrin IX ferrochelatase (HemH), the enzyme responsible for the final step of protohaem biosynthesis. CbiK and protoporphyrin IX ferrochelatase both consist of two domains and most probably utilise a similar mechanism for inserting divalent cations. It also seems likely that CbiX shares at least one of the two domains (the active site) with CbiK and HemH. However, CbiX contains an extra polyhistidine stretch at the C-terminal (19 histidines within the last 34 amino acids), which surely plays a role in metal chelation.

Figure 7-1- Cobalt chelation step via the anaerobic pathway.



In summary, cobalt insertion via the anaerobic pathway occurs at the level of precorrin-2 or sirohydrochlorin, does not require ATP (Figure 7-1) and is catalysed by either CbiK, CbiX or CysG. They probably all share a similar active site and mechanism, which is also found in the *B. subtilis* protoporphyrin IX ferrochelatase.

### 7-4- CbiD and CbiG involved in the ring contraction process?

The ring contraction in *P. denitrificans* is performed in two main steps by the monooxygenase CobG, which forms the γ-lactone, and by the methyltransferase CobJ, which methylates at C-17. Via the anaerobic pathway, the substrate and the mechanism of the ring contraction process are different since no oxygen is involved and early cobalt insertion is essential. It has been shown that cobalt-precorrin-3 could act as the substrate of the *S. typhimurium* CbiH (C-17 methyltransferase) and that this reaction results in δ-lactone formation [Santander *et al.* 1997]. However, the rate of this reaction was very low, suggesting that, *in vivo*, other enzymes are required. From a study of the fusion-genes involved in cobalamin biosynthesis, it was noticed that *Synechocystis sp.* contains a *cbiGH* whereas *P. gingivalis* harbours a *cbiGF*. Such fusion proteins are formed, in general, from two genes that encode sequential reactions (*cbiET*, *hemD-cysG*$^{A}$). However, this is not always the case since both *A. fulgidus* (*cbiHC*) and *M. tuberculosis* (*cbiIJ*) encode for non-sequential reactions (of course, sequencing errors cannot be completely disregarded). Nonetheless, the *P. gingivalis cbiGF* fusion has been confirmed by protein expression studies [Dr J. Roper, pers. comm.] and allows speculation that CbiG could play a role in the ring contraction process between the C-17 and C-11 methylation reactions.

Another important gene fusion observed within the *P. gingivalis* genome is *cbiJD*. The fusion connects the cobalt-precorrin-6 reductase and the *cobI* protein CbiD. An alignment of all CbiD proteins highlights a conserved motif, similar although not completely typical, of the SAM binding site of methyltransferases. Since the anaerobic C-1 methyltransferase has not yet been identified as no equivalent of the aerobic CobF exits, the ability of *B. megaterium* CbiD to bind SAM was investigated to see if its role is to methylate the C-1 carbon. Unfortunately CbiD was found to be insoluble and the SAM binding assay with refolded CbiD did not show any association between enzyme and substrate.

Nevertheless, unless one of the methyltransferases such as CbiF is bifunctional, CbiD remains the most likely candidate for the role of C-1 methyltransferase.

## 7-5- Comparison of cobalamin pathways between different micro-organisms.

An extensive search in the database for genes involved in cobalamin biosynthesis has been undertaken, resulting in Table 7-1. The search was based on similarity to the *P. denitrificans* and *S. typhimurium* Cob/Cbi proteins. The *B. megaterium* CbiW, CbiX and CbiY, and the *R. capsulatus* Blu proteins were also compared with the database. The searches were finally focused on *Rhodobacter capsulatus, Mycobacterium tuberculosis, Pseudomonas aeruginosa, Synechocystis sp., Clostridium acetobutylicum, Porphyromonas gingivalis,* and the archaebacteria *Methanococcus jannaschii, Methanobacterium thermoautotrophicum* and *Archaeoglobus fulgidus.* All these micro-organisms are part of genome sequencing projects, although some of these sequences have not been yet annotated (no ORF references).

### 7-5-1- The archaebacteria.

Three archaebacteria were classified within the bacteria that follow an anaerobic route to synthesise $B_{12}$ (CbiD and CbiG). In these organisms, there is no apparent clustering of the genes and, with the exception of *cbiM, -N, O* and *Q*, they appear to be distributed randomly throughout the genome. The cobalt insertion process in archaebacteria is totally unknown since no CbiK, CysG or CbiX homologues are found in these micro-organisms (*M. jannaschii* contains a putative CobN but no CobS or CobT). None of these bacteria possess a BtuR like protein, which suggests that only the methyl-cobalamin form is utilised. Even if a putative cobalamin synthase ($CobV^{Pd}/CobS^{St}$) is present, the pathway between cobinamide and cobalamin in archaebacteria is quite different to the eubacteria pathway. Indeed the enzyme performing the transformation of cobinamide to GDP-cobinamide (CobP) is absent as are the dimethylbenzimidazole phosphoribosyl-transferase (CobU) and phosphatase.

The comparison of all methyltransferases involved in cobalamin biosynthesis highlights two subgroups within each class of methyltransferase, the aerobic and anaerobic enzymes, but no subgroups are formed with the archaebacteria

enzymes. This indicates that their cobalamin pathways diverged prior to the differentiation with eubacteria.

7-5-2- The eubacteria following an anaerobic pathway.

The eubacteria that possess an anaerobic cobalamin pathway contain either CbiK or CbiX as a cobalt-chelatase. These enzymes can complement each other despite the lack of similarity at the amino acid level. Although similar sets of enzymes responsible for the transformation of cobalt-precorrin-2 to cobalamin are all present in bacteria with an anaerobic pathway, there are still specific differences to be found. *S. typhimurium* and *B. megaterium cob* operons appear equivalent, but two major differences have been nevertheless observed. (i) The *B. megaterium cobI* operon can synthesise corrins only in the presence of a precorrin-2 dehydrogenase whereas *S. typhimurium* does not require this enzyme; (ii) The *B. megaterium cbiD* either does not complement or complements very slowly *S. typhimurium cbiD* mutants. These two facts are most probably linked in some way, but the requirement for the presence of a dehydrogenase in the biosynthesis of corrins with the *B. megaterium cob* operon is not understood.

7-5-3- The eubacteria following an aerobic pathway.

The main genetic characteristics of the aerobic pathway (elucidated in *P. denitrificans*) are the presence of the mono-oxygenase, CobG, the C-1 methyltransferase, CobF and the cobaltochelatase complex, CobN-S-T. Two other bacteria, *R. capsulatus* and *M. tuberculosis*, have now been classified as possessing the genetic requirements for an aerobic pathway, but this interpretation is debatable since there are apparent genetic variations between these organisms.

For instance, although the *M. tuberculosis* genome contains none of the anaerobic pathway specific genes, it does contain CobG and CobN, but CobF, CobS and CobT are all absent. *M. tuberculosis* might therefore utilise an aerobic pathway which is similar to the *P. denitrificans* pathway but with some alterations linked to the cobalt insertion and to the C-1 methylation.

*R. capsulatus* is able to synthesise cobalamin in the absence and in the presence of oxygen (data not shown, in collaboration with Dr S. Gough). This bacterium contains CobF and the cobaltochelatase complex CobN-S-T but is significantly missing the mono-oxygenase, CobG. It can therefore be proposed that *R. capsulatus* does not use molecular oxygen as a substrate for the ring contraction process, but must use either another substrate such as sulfur or an altogether different mechanism. In *R. capsulatus*, all the genes involved in cobalamin biosynthesis are clustered into two operons. One operon contains *cobS* and *cobT* whereas the second operon contains twenty-seven ORFs (Figure 7-2) of which one, RC00663, has no homologue within the *cob* genes and could be the candidate for initiating the ring contraction, replacing the missing CobG. This hypothetical protein contains an interesting motif that is similar to iron-binding proteins (CxxCxxCxxxC-23x- CHxCxxCxHxC).

7-5-4- *Pseudomonas aeruginosa* cobalamin pathway.

*P. aeruginosa* is a pathogenic bacterium that is increasingly problematic due to its resistance to most modern antibiotics. It is also the most complex bacterium with regards to cobalamin biosynthesis. In fact, *P. aeruginosa* contains the same genetic hallmarks for an aerobic pathway as *M. tuberculosis* in that it possesses a CobG and a CobN but no CobF and CobS-T. However, *P. aeruginosa* also contains a CbiG (GH) and a CbiD, the hallmarks of the anaerobic pathway plus a CysG, which can act as the anaerobic cobalt chelatase.

The fact that *P. aeruginosa* can synthesise cobalamin under both aerobic and anaerobic conditions (data not shown, in collaboration with Dr D. Jahn), proves that molecular oxygen is not essential for the synthesis of the vitamin.

A comparison of the methyltransferases found in *P. aeruginosa* with equivalent enzymes from other bacteria reveals that all *P. aeruginosa* proteins segregate with the aerobic enzymes. Although *P. aeruginosa* contains a CysG-like protein, it also harbours two CobA-like sequences. None of these enzymes are within the main *cob* operons and it is not clear yet whether *P. aeruginosa* uses CysG or CobA for cobalamin biosynthesis or whether it switches between them depending on the environmental conditions. Therefore, this micro-organism is perhaps capable of switching between the two pathways, although this implies

that the methyltransferases and most of the enzymes must be able to accept both cobalt-containing and cobalt-free intermediates. Alternatively, *P. aeruginosa* may use a single pathway, with the genetic hallmarks of the second pathway that represent some evolutionary remnant or some recent genetic acquisition. However, this hypothesis is very unlikely as *cbiD*, *cbiG*, *cobG* are part of a unique locus (Figure 7-2).

## 7-6- Final thoughts on the evolution of cobalamin biosynthesis.

It is generally perceived that the biosynthesis of cobalamin is performed via one of two pathways: (i) the anaerobic pathway, which is characterised by early cobalt insertion, has no requirement for molecular oxygen and is absolutely dependent upon two enzymes that are specific to this pathway, CbiD and CbiG, and (ii) the aerobic pathway, which in contrast requires molecular oxygen and inserts cobalt at a late stage in an ATP-dependent manner (Figure 7-3).

If we go along with the two pathway hypothesis then clearly the more primeval one has to be the anaerobic route. This view is supported by the presence of the anaerobic pathway in the archaebacteria, although interestingly the cobalt chelatase remains to be identified in these bacteria since they do not contain sequences for CysG, CbiK or CbiX. However, in an ancestral world it is likely that cobalt was inserted at the level of precorrin-2 and that a pathway based on cobalt-precorrin complexes evolved. Moreover, when the sequences of the archaebacteria cobalamin biosynthetic enzymes are compared to all the known cobalamin sequences, although they segregate into the anaerobic pathway enzymes, the archaebacteria enzymes do not form a sub-division thereby establishing that the anaerobic pathway had already diversified by the time the Eubacteria appeared. However, following this argument it becomes very difficult to explain how, somewhere around 2 billion years ago, the aerobic pathway appeared. How can an anaerobic pathway, with its enzymes already imprinted with cobalt-containing substrates suddenly switch to using molecular oxygen and a new set of cobalt-free intermediates? The two pathways must have evolved at a much earlier stage, prior to any rigid recognition of the intermediates. Thus the first cobalt-free pathway must have evolved in the absence of oxygen and must have discovered a ring contraction system that operated in its absence. How appropriate, then, that *R. capsulatus*

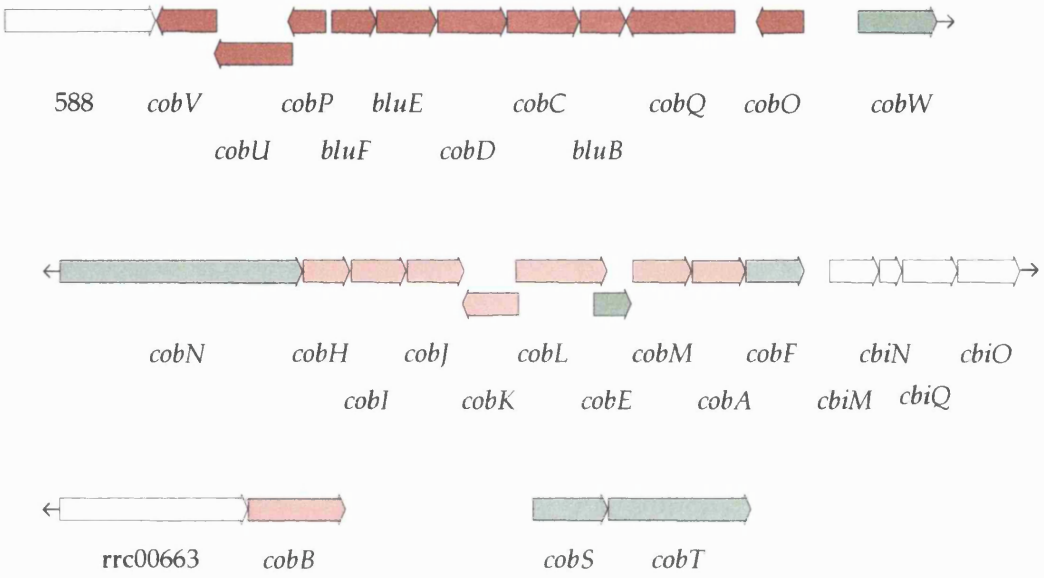
appears to have such a pathway - is it the missing link between the anaerobic and aerobic pathway? And what of *P. aeruginosa*, the bacterium that appears to carry the genetic hallmarks of both the aerobic and anaerobic pathway; does this represent a pathway that can use both cobalt-containing and cobalt-free intermediates? Clearly, there is still a great deal to be investigated before we can start to address these questions.

Finally, it becomes apparent that the terms "aerobic" and "anaerobic" are somewhat redundant, and frankly misleading. In *B. megaterium* the anaerobic pathway can operate efficiently in the presence of oxygen whilst, as discussed above, some aerobic-like pathways operate in the complete absence of oxygen. The main difference between the pathways is the timing of cobalt insertion, since this event preordains whether ring contraction occurs early or late. Thus a simpler, and more accurate, classification of the pathways should relate to the timing of cobalt insertion such that the aerobic pathway would be referred to as the late-cobalt insertion pathway whilst the anaerobic pathway would be termed the early-cobalt insertion pathway.

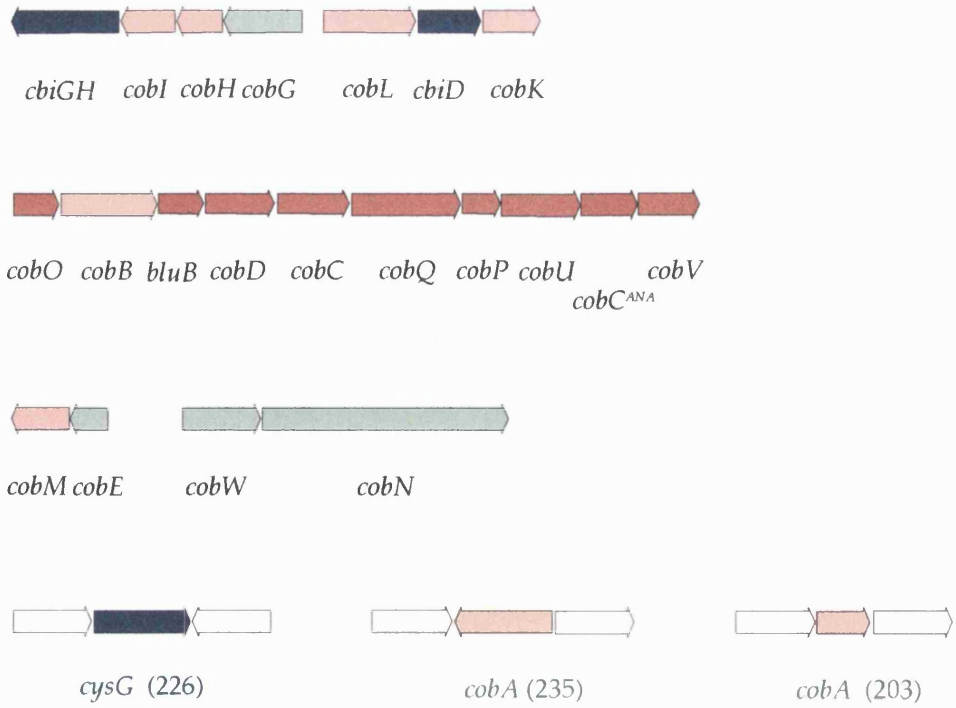
The remarkable study of the *P. denitrificans* cobalamin biosynthetic pathway has allowed researchers to claim that the Mount Everest of biosynthetic problems has been conquered. However, the work presented in this thesis has clearly shown that there are many routes to the summit. In fact, each bacterium probably has its own favoured trail with some unique features. In as much that we know about cobalamin biosynthesis, we have to acknowledge that we also know very little. Cobalamin biosynthesis will remain a challenge to biochemists and chemists for some time yet.

Figure 7-2- *R. capsulatus* and *P. aeruginosa cob* operons.

a- *Rhodobacter capsulatus.*

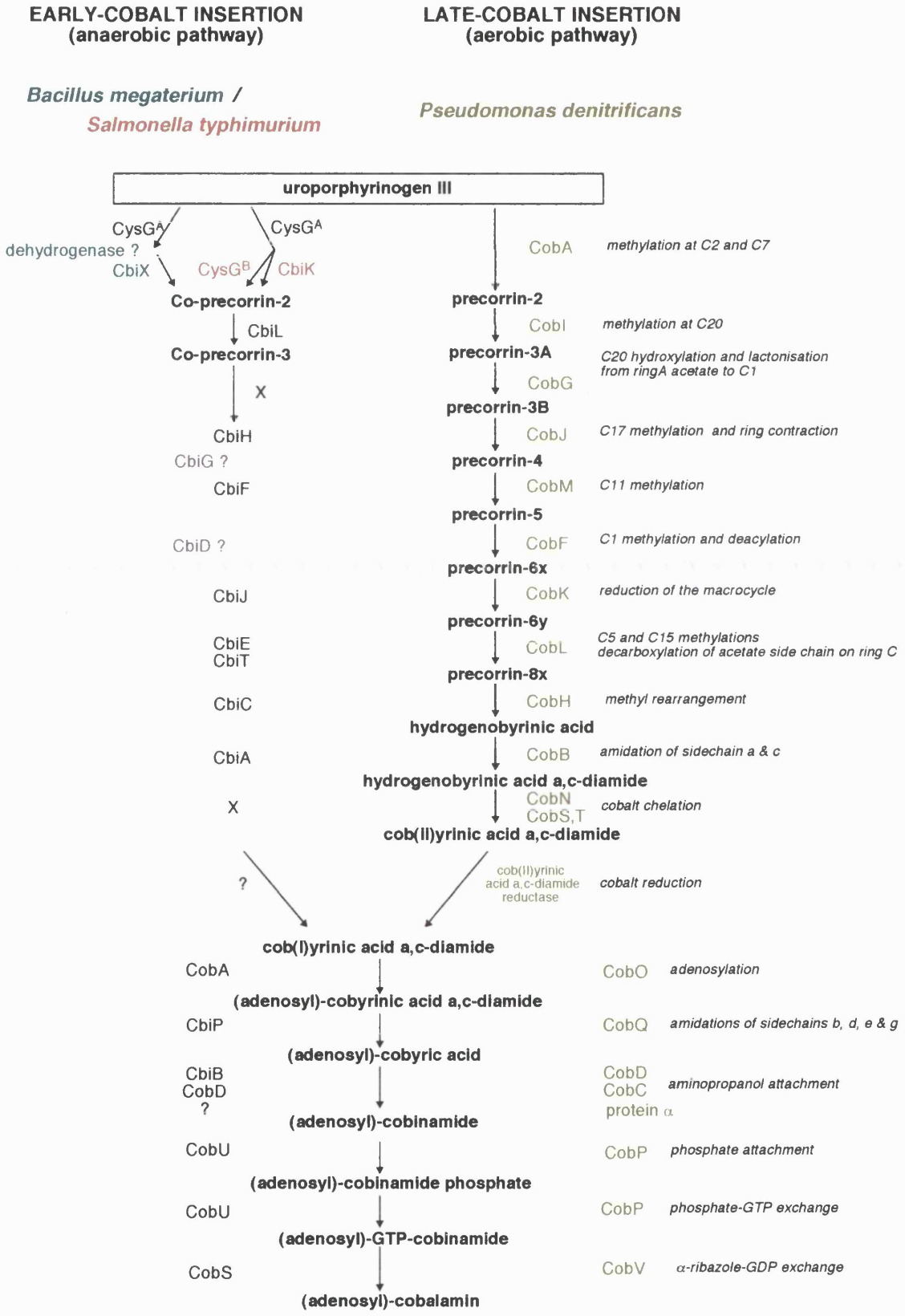



b- *Pseudomonas aeruginosa.*




The boxes filled in pink indicate that these genes are involved in cobyrinic acid a,c diamide synthesis. The boxes filled in red are involved in the transformation of cobyrinic acid a,c diamide into cobalamin. The boxes filled in green are genes specific to the aerobic pathway and finally, in blue, to the anaerobic pathway.

Figure 7-3- Reactions and genetic requirements for transformation of uro'gen III into cobalamin.



# CHAPTER 8,

# MATERIALS AND METHODS

## 8-1- Materials.

### 8-1-1- Chemicals.

Cyanocobalamin, antibiotics and most other chemicals were purchased from Sigma. Other materials were purchased from manufacturers as follows: restriction and modification enzymes from Promega, except *Nde*I which was purchased from Pharmacia; AnalaR water from BDH; Tryptone and yeast extract from Oxoid, PD10 columns, Sephadex G25 and DEAE Sephacel from Pharmacia; [methyl-$^3$H] SAM from American Radiolabeled Chemicals Inc; the Sculptor™ in vitro mutagenesis system (RPN 1526) from Amersham Life Science; ABI PRISM™ dye terminator cycle sequencing reaction kit from Applied biosystems.

### 8-1-2- Bacterial strains.

All strains (or plasmids) with an AR (or pAR) prefix were mainly constructed by either Anne Lanois, Florence Levillayer or myself between 1988 and 1995. They are all part of Dr Alain Rambach's collection. The use of all these strains and plasmids in this study was an invaluable asset and a generous gift from Dr Alain Rambach.

| Strains | Genotype and/or phenotype | Description | Reference |
|---|---|---|---|
| | ***Salmonella typhimurium*** | | |
| AR2720 | *S. typhimurium cbiF* mutant | | Raux/96 |
| AR3612 | *S. typhimurium Leu*$^+$ *Sm*$^R$ *cysG metE* | assays from cobyric acid | Raux/96 |
| | ***Bacillus megaterium*** | | |
| DSM509 | | | |
| ATCC10778 | | | |
| | ***Escherichia coli*** | | |
| 302Δa | *E. coli cysG*; Nir$^S$ Lac$^+$ CysG$^-$ | | Griffiths/ 87 |
| AR3730 | LE392 *met*$^+$ | rendered *met*$^+$ with P1 | Raux/96 |
| AR8766 | MC1061 (pAR8766) | *B. m. cobI* operon | This study |
| AR8830 | AR3730 (pAR8827)(pAR8414) | all *S.t. cbi* genes except *cbiB* | Raux/96 |

| | | | |
|---|---|---|---|
| AR8925 | AR3730 (pAR8849)(pAR8911) | *S.t cbiP* & all *B.m cbi* genes, a mutation in *cbiX* | Raux/98c |
| AR8930 | AR3730 (pAR8849)(pAR8916) | *S.t cbiP* & all *B.m cbi* genes except the 2nd part of $cbiH_{60}$ | Raux/98c |
| AR8937 | AR3730 (pAR8849)(pAR8877) | *S.t cbiP* & all *B.m cbi* genes | Raux/98c |
| AR8948 | 302Δa(pAR8849)(pAR8877) | *S.t cbiP* & all *B.m cbi* genes | This study |
| AR8949 | 302Δa(pAR8849)(pAR8911) | *S.t cbiP* & all *B.m cbi* genes, a mutation in *cbiX* | This study |
| AR8957 | 302Δa(pAR8827)(pAR8414) | all *S.t. cbi* genes but not *cbiB* and *E. coli cysG* | This study |
| AR8967 | AR3730 (pAR8962)(pAR8911) | *S.t cbiP* and *B.m cbiX* & all *B.m cbi* genes, a mutation in *cbiX* | Raux/98c |
| AR8959 | AR3730(pAR8827)(pKK223.3) | all *S.t. cbi* genes but not *cbiB* | Raux/96 |
| AR8981 | AR3730 (pAR3062) | *S. t. cobI-III-II* operon | Raux/96 |
| AR8984 | AR3730 (pAR8658)(pAR8976) | all *S. t. cbi* genes but not *cbiB* & *cbiG* | Raux/96 |
| AR8985 | AR3730 (pAR8658)(pAR8977) | all *S. t. cbi* genes but not *cbiB* & *cbiD* | Raux/96 |
| BL21(DE3) (pLysS) | $F^-$ *ompT* $hsdS_B(r_B^- m_B^-)$ *gal dcm (DE3)* pLysS | | Novagen |
| ER99 | 302Δa (pAR8827)(pKK223.3) | all *S. t. cbi* genes but not *cbiB* | This study |
| ER123A | *E. coli cysG* (pAR8827)(pER119) | all *S.t cbi* genes and *P.d cobA* | Raux/97 |
| ER126K$^{\Delta}$ | LE *met*$^+$ (pER126K$^{\Delta}$) | all *S.t cbi* genes with a deletion into *cbiK* | Raux/97 |
| ER126K$^{\Delta}$K | LE *met*$^+$ (pER126K$^{\Delta}$)(pAR8580) | all *S.t cbi* genes with a deletion into *cbiK* & *S.t cbiK* | Raux/97 |
| ER144 | AR3730 (pAR8849)(pAR8846) | *S.t cbiP* & all *B.m cbi* genes except *cbiW* | Raux/98c |
| ER164 | AR3730 (pAR8849)(pAR8909) | *S.t cbiP* & all *B.m cbi* genes except *cbiD* | Raux/98c |
| ER165 | AR3730 (pAR8849)(pAR8921) | *S.t cbiP* & all *B.m cbi* genes, a mutation in *cbiG* | Raux/98c |
| ER167 | AR3730 (pER156)(pAR8909) | *S.t cbiP* and *B.m cbiD* & all *B.m cbi* genes, a mutation in *cbiD* | Raux/98c |
| ER169 | AR3730 (pER160)(pAR8921) | *S.t cbiP* and *B.m cbiG* & all *B.m cbi* genes, a mutation in *cbiG* | Raux/98c |
| ER171 | 302Δa (pAR8086) | | Raux/97 |
| ER172 | 302Δa (pAR8086)(pER119) | *P.d cobA* | Raux/97 |
| ER173 | 302Δa (pAR8086)(pER170) | *P.d cobA* & *S.t cbiK* | Raux/97 |
| ER176 | AR3730 (pER146)(pAR8877) | *S.t cbiP* and *P.d cobA* & all *B.m cbi* genes | Raux/98c |

| | | | |
|---|---|---|---|
| ER177 | AR3730 (pER146)(pAR8980) | *S.t cbiP* and *P.d cobA* & all *B.m cbi* genes except *cysGA*, *cbiY* and *btuR* | Raux/98c |
| ER182 | 302Δa (pAR8086)(pER108) | *E. coli cysG* | Raux/97 |
| ER185 | 302Δa (pER126K$^{\Delta}$) | all *S.t cbi* genes with a deletion into *cbiK* | Raux/97 |
| ER188 | 302Δa (pER126K$^{\Delta}$)(pER119) | all S.t cbi genes with a deletion into cbiK and P.d cobA | Raux/97 |
| ER189 | 302Δa (pER126KΔ)(pER179) | all S.t cobI genes except cbiB but with a deletion in cbiK & P.d. cobA and B.m. cbiX | This study |
| ER190 | 302Δa (pER126KΔ)(pER170) | all S.t cbi genes with a deletion into cbiK &P.d cobAand cbiK. | Raux/97 |
| ER191 | 302Δa (pER126KΔ)(pER108) | all S.t cbi genes with a deletion into cbiK and E.c cysG | Raux/97 |
| ER203 | 302Δa (pER198)(pAR8877) | all B.m. cbi genes & S.t. cbiP and E.c. cysG | This study |
| ER205 | AR3730 (pER197)(pAR8911) | all B.m. cbi genes with a mutation in cbiX & S.t. cbiP and cbiK | This study |
| ER210 | HU227 (pAR8827)(pER119) | all S.t cbi genes except cbiB & P.d cobA | This study |
| ER222 | JM101 (pER222) | Bm cbiD (improved RBS) | This study |
| ER225 | BL21(DE3)(pLysS)(pER229) | His-tagged B.m cbiF | Raux/98a |
| ER227 | BL21(DE3)(pLysS)(pER231) | His-tagged S.t cbiK | This study |
| ER229 | JM101 (pER229) | His-tagged B.m cbiF | Raux/98a |
| ER231 | JM101 (pER231) | His-tagged S.t cbiK | This study |
| ER241 | AR3730 (pAR8849)(pAR8766) | S.t cbiP & all B.m cbi genes (no tac promoter) | Raux/98c |
| ER243 | BL21(DE3)(pLysS)(pER242) | His-tagged P.d cobA | This study |
| ER251 | 302Δa (pAR8086)(pER250) | *P.d cobA* and *S.c MET8* | This study |
| ER252 | 302Δa (pER126K$^{\Delta}$)(pER250) | all *S.t cbi* genes with a deletion into *cbiK* & *P.d cobA* and *S.c MET8* | This study |
| ER270 | 302Δa (pER270) | *S.t cbiP* and all *B.m cbi* genes | This study |
| ER271 | 302Δa (pER270)(pER108) | *S.t cbiP* and all *B.m cbi* genes & *E. c cysG* | This study |
| ER272 | 302Δa (pER270)(pSWG21D) | *S.t cbiP* and all *B.m cbi* genes & *E.c cysG*$^{G21D}$ | This study |
| ER273 | 302Δa (pER270)(pER119) | *S.t cbiP* and all *B.m cbi* genes & *P.d cobA* | This study |
| ER274 | 302Δa (pER270)(pER170) | *S.t cbiP* and all *B.m cbi* genes & *P.d cobA* and *S.t cbiK* | This study |

| | | | |
|---|---|---|---|
| ER275 | 302Δa (pER270)(pER250) | *S.t cbiP* and all *B.m cbi* genes & *P.d cobA* and *S.c MET8* | This study |
| ER276 | 302Δa (pER270)(pER179) | *S.t cbiP* and all *B.m cbi* genes & *P.d cobA* and *B.m cbiX* | This study |
| ER278 | 302Δa (pER270)(pSWG21D) | *S.t cbiP* and all *B.m cbi* genes & *P.d cobA* and *E.c cysG*$^{G21D}$ | This study |
| ER298 | 302Δa (pER198)(pAR8911) | all *B.m cbi* genes with a mutation in *cbiX*& *S.t cbiK* and *E.c cysG* | This study |
| ER306 | 302Δa (pER126K$^{\Delta}$)(pER119) | all *S.t cbi* genes with a deletion into *cbiK* & *P.d cobA* | This study |
| ER307 | 302Δa (pER126K$^{\Delta}$)(pER303) | all *S.t cbi* genes with a deletion into *cbiK* & *S. t cbiK* and *P.d cobA* | This study |
| ER308 | 302Δa (pER126K$^{\Delta}$)(pER303$^{E89A}$) | all *S.t cbi* genes with a deletion into *cbiK* & *S. t cbiK*$^{E89A}$ and *P.d cobA* | This study |
| ER309 | 302Δa (pER126K$^{\Delta}$)(pER303$^{H145A}$) | all *S.t cbi* genes with a deletion into *cbiK* & *S. t cbiK*$^{H145A}$ and *P.d cobA* | This study |
| ER310 | 302Δa (pER126K$^{\Delta}$)(pER303$^{E175A}$) | all *S.t cbi* genes with a deletion into *cbiK* & *S. t cbiK*$^{E175A}$ and *P.d cobA* | This study |
| ER311 | 302Δa (pER126K$^{\Delta}$)(pER303$^{H207A}$) | all *S.t cbi* genes with a deletion into *cbiK* & *S. t cbiK*$^{H207A}$ and *P.d cobA* | This study |
| ER312 | 302Δa (pER126K$^{\Delta}$)(pER303$^{D211A}$) | all *S.t cbi* genes with a deletion into *cbiK* & *S. t cbiK*$^{D211A}$ and *P.d cobA* | This study |
| ER313 | 302Δa (pER126K$^{\Delta}$)(pER303$^{H145A/H207A}$) | all *S.t cbi* genes with a deletion into *cbiK* & *S. t cbiK*$^{H145A/H207A}$ and *P.d cobA* | This study |
| ER315 | 302Δa (pER270)(pER315) | *S.t cbiP* and all *B.m cbi* genes & *P.d cobA* and *B.s ylnF* | This study |
| ER316 | 302Δa (pAR8086)(pER119) | *P.d cobA* | This study |
| ER317 | 302Δa (pAR8086)(pER303) | *S. t cbiK* and *P.d cobA* | This study |
| ER318 | 302Δa (pAR8086)(pER303$^{E89A}$) | *S. t cbiK*$^{E89A}$ and *P.d cobA* | This study |
| ER319 | 302Δa (pAR8086)(pER303$^{H145A}$) | *S. t cbiK*$^{H145A}$ and *P.d cobA* | This study |
| ER320 | 302Δa (pAR8086)(pER303$^{E175A}$) | *S. t cbiK*$^{E175A}$ and *P.d cobA* | This study |
| ER321 | 302Δa (pAR8086)(pER303$^{H207A}$) | *S. t cbiK*$^{H207A}$ and *P.d cobA* | This study |
| ER322 | 302Δa (pAR8086)(pER303$^{D211A}$) | *S. t cbiK*$^{D211A}$ and *P.d cobA* | This study |
| ER323 | 302Δa (pAR8086)(pER303$^{H145A/H207A}$) | *S. t cbiK*$^{H145A/H207A}$ and *P.d cobA* | This study |
| HU227 | *HfrPO$_2$A hemA41 metB1 relA1*, hemin permeable | | Sasarman / 68 |

| | | | |
|---|---|---|---|
| JM101 | *supE thi-1* Δ(*lac-proAB*) F' [*traD36 proAB*$^+$ *lacI*$^q$ *lacZ*Δ*M15* | | Sambrook / 89 |
| LE392 | F$^-$ e14$^-$ (McrA$^-$) *hsdR514* ($r_K^-$ $m_K^+$) *supE44 supF58 lacY1* or Δ(*lacIZY*)*6 galK2 galT22 metB1 trpR55* | | Sambrook / 89 |
| LE 392 *met*$^+$ | See AR3730 | | |
| TG1 | K12Δ(*lac-peoAB*) *supE thi hsdD5/F'traD36 proA*$^+$ *B*$^+$ *lacI*$^q$ *lacZ ΔM15* | | Sambrook / 89 |

*B.m: Bacillus megaterium and B.s: Bacillus subtilis; E.c : Escherichia coli; P.d : Pseudomonas denitrificans; S.t : Salmonella typhimurium; S.c : Saccharomyces cerevisiae.*
"All *S.t cbi* genes" means: *cbiA-C-D-E-T-F-G-H-J-K-L-M-N-Q-O-P* but does not contain *cbiB*.
"All *B.m cbi* genes" means: *B.m cbiW-H*$_{60}$*-X-J-C-D-ET-L-F-G-A-cysG*$^A$*-cbiY-btuR->*<-ORF1-ORF2

## 8-1-3- Plasmids.

| Plasmids | Genotype and/or phenotype | Description | Reference |
|---|---|---|---|
| M13mp18 | vector derived from the *E. coli* bacteriophage M13 | | Biolabs |
| pACYC184 | carries the p15A origin of replication | | Biolabs |
| pAR3062 | *S.t pduB-A-F-pocR-cbiA-B-C-D-E-T-F-G-H-J-K-L-M-N-Q-O-P-cobU-S-T* | 21.5 kb *S.typhimurium Sau3AI* fragment cloned into the *BglII* site of pLA2917 (cosmid) | Raux/96 |
| pAR8000 | *S.t cbiB-C-D-E-T-F-G-H-J-K-L-M-N-Q-O-P-cobU-S-T* | *Pst*I fragment of pAR3062 cloned into pKK223.3 | Raux/96 |
| pAR8086 | *E.coli lacI*$^q$ | pACYC184 with *lacIq* in *EcoRV* site | Raux/96 |
| pAR8358 | *S.t cbiN-Q-O* | *Pst*I 11445-15118 fragment of pAR3062 cloned into pKK223.3 | This study |
| pAR8414 | *E.coli cysG** | cloned into *EcoRI/HindIII* sites of pKK223.3 (modified RBS : AGGAGGTTTTACATG) | Raux/96 |
| pAR8477 | *S.t cbiF* | *ClaI* 6241/*EcoRI* 8207 fragment cloned into pKK223.3 | Raux/96 |
| pAR8531 | *S.t cbiP* | *SalI* 13423/*Eco*RI 15531 cloned into pKK223.3 | Raux/96 |
| pAR8533 | *S.t cbiD* | *StuI* 4261/*XmnI* 5676 fragment cloned into pKK223.3 | Raux/96 |
| pAR8580 | *S.t cbiK* | *PstI* 9594/*HindIII* 11219 fragment cloned into pKK223.3 | Raux/96 |
| pAR8603 | *S.t cbiM-N* | *HindIII* 11219/*BstXI* 13027 fragment cloned into pKK223.3 | Raux/96 |

| | | | |
|---|---|---|---|
| pAR8606 | *S.t cbiF-G-H* | *XmnI* 6400/*NsiI* 9308 fragment cloned into pKK223.3 | Raux/96 |
| pAR8658 | *E. coli cysG** | P*tac-E.c cysG** cloned into pAR8086 cut with *Bam*HI and *Nru*I | This study |
| pAR8668 | | variant of pKK223.3, contains 1 *Bam*HI site in the MCS only | Raux/98a |
| pAR8754 | *S.t cbiA- B.m cbiX-J-C-D-ET* | 6 kb *B. megaterium Alu*I fragment cloned into the *Sma*I site of pAR8720 (from position 3055 to 9063) | Raux/98b |
| pAR8766 | *B.m cbiW-$H_{60}$-X-J-C-D-ET-L-F-G-A-cysG$^A$-cbiY-btuR*-><-ORF1-ORF2 | 16.3 kb *B. megaterium Sau*3AI fragment cloned into *Bam*HI site of pKK223.3 (the *tac* promoter has been deleted) | Raux/98b |
| pAR8000 | *S.t cbiB-C-D-E-T-F-G-H-J-K-L-M-N-Q-O-P-cobU-S-T.* | partial *PstI* digest of pAR3062 (from *PstI* 1606) cloned into *PstI* site of pKK223.3 | Raux/96 |
| pAR8811 | *B.m cbi$H_{60}$-X-J* | 1914 to 5834 *MluI* fragment cloned into the *Sma*I site of pKK223.3 | Raux/98c |
| pAR8814 | *S.t cbiA*/cbiC-D-E-T-F-G-H-J/cbiK-L-M-N-Q-O-P* | partial *BamHI/ScaI* digest of pAR8600 cloned into pAR8720 cut with *BamHI/ScaI* | Raux/96 |
| pAR8827 | *S.t cbiA*/cbiC-D-E-T-F-G-H-J/cbiK-L-M-N-Q-O-P* | *SphI/ScaI* fragment of pAR8814 cloned into *SphI/NruI* sites of pAR8086 | Raux/96 |
| pAR8846 | *B.m cbi$H_{60}$-X-J-C-D-ET-L-F-G-A-cysG$^A$-cbiY-btuR*-><-ORF1-ORF2 | *Nco*I 2971-*Sal*I 16432 fragment of pAR8766 cloned into *Nco*I/*Sal*I of pAR8811; this plasmid does not contain *cbiW* | Raux/98c |
| pAR8849 | *S.t cbiP* | *Sph*I-*Sca*I fragment of pAR8531 (P*tac-S.t cbiP*) cloned into *Sph*I/*Nru*I sites of pAR8086 | Raux/98c |
| pAR8877 | *B.m cbiW-$H_{60}$-X-J-C-D-ET-L-F-G-A-cysG$^A$-cbiY-btuR*-><-ORF1-ORF2 | *Bgl*I 1218-*Sal*I 16432 fragment of pAR8766 cloned into *Sma*I/*Sal*I sites of pKK223.3 | Raux/98c |
| pAR8882 | *B.m cbiX* | *Ssp*I 3642-*Sna*BI 5041 fragment of pAR8766 cloned into the *Sma*I site of pKK223.3 | Raux/98c |
| pAR8909 | *B.m cbiW-$H_{60}$-X-J-C-D$^M$-ET-L-F-G-A-cysG$^A$-cbiY-btuR*-><-ORF1-ORF2 | mutation at the *Acc*I 6412 site into the *cbiD* gene of pAR8877 (TAG at position 6444) | Raux/98c |
| pAR8911 | *B.m cbiW-$H_{60}$-X$^M$-J-C-D-ET-L-F-G-A-cysG$^A$-cbiY-btuR*-><-ORF1-ORF2 | mutation at the *Acc*I 4088 site into the *cbiX* gene of pAR8877 (TAA at position 4135) | Raux/98c |

| | | | |
|---|---|---|---|
| pAR8916 | *B.m cbiW-$H_{60}^{M}$-X-J-C-D-ET-L-F-G-A-cysG$^{A}$-cbiY-btuR->*<-ORF1-ORF2 | mutation at the *Acc*I 3241 site into the 2nd part of *cbiH60* gene of pAR8877 (TGA at position 3295) | Raux/98c |
| pAR8921 | *B.m cbiW-$H_{60}$-X-J-C-D-ET-L-F-G$^{M}$-A-cysG$^{A}$-cbiY-btuR->*<-ORF1-ORF2 | mutation at the *Cla*I 10217 site into the *cbiG* gene of pAR8877 (TAA at position 10411) | Raux/98c |
| pAR8962 | *S.t cbiP-B.m cbiX* | *Eco*RI-*Sca*I digest of pAR8882 cloned into *Sal*I site of pAR8849 | Raux/98c |
| pAR8976 | *S.t cbiA*/cbiC-D-E-T-F-G$^{\Delta}$-H-J/cbiK-L-M-N-Q-O-P* | deletion of 417 bases between *BstBI* 7774/8211 sites into *cbiG* gene of pAR8814 | Raux/96 |
| pAR8977 | *S.t cbiA*/cbiC-D$^{M}$-E-T-F-G-H-J/cbiK-L-M-N-Q-O-P* | mutation at the *MluI* 4608 site into *cbiD* gene of pAR8814 => TAA at position 4619 | Raux/96 |
| pAR8980 | *B.m cbiW-$H_{60}$-X-J-C-D-ET-L-F-G-A* | *Eco*RI fragment of pAR8877 (from1218 to 12734) cloned into *Eco*RI site of pKK223.3 | Raux/98c |
| pCR395 | *P.d cobA* | | Roessner/95 |
| pER108 | *E.c cysG* | variant of pAR8414 | Raux/97 |
| •pER119 | *P.d cobA* | *Eco*RI-*Pst*I fragment of pCR395 cloned into pKK223.3 | Raux/97 |
| •pER126K$^{\Delta}$ | *S.t cbiA-cbiC-D-E-T-F-G-H-J-cbiK$^{\Delta}$-L-M-N-Q-O-P* | Deletion of 54 bp between *Hind*III sites (10028-10082) into *cbiK* gene of pAR8827 (TAA at position 10085) | Raux/97 |
| pER143 | *S.t cbiP* | variant of pAR8849 | |
| pER146 | *S.t cbiP-P.d cobA* | *Eco*RI-*Sal*I fragment of pCR395 cloned into *Sal*I site of pAR8849 | Raux/98c |
| pER156 | *S.t cbiP-B.m cbiD* | 5999 to 7755 *Pvu*II fragment of pAR8766 cloned into *Sal*I site of pAR8849 | Raux/98c |
| pER160 | *S.t cbiP-B.m cbiG* | *Pst*I 9471-*Kpn*I 11061 fragment of pAR8766 cloned into *Sal*I site of pAR8849 | Raux/98c |
| •pER170 | *P.d cobA & S.t cbiK* | *Eco*RI-*Sca*I fragment of pAR8580 cloned into pER119 cut partially by *Sal*I and *Sca*I | Raux/97 |
| pER179 | *P.d cobA & B.m cbiX* | *Eco*RI-*Sca*I fragment of pAR8882 cloned into pER119 cut partially by *Sal*I and *Sca*I | This study |
| pER197 | *S.t cbiP & cbiK* | *Eco*RI-*Sca*I fragment of pAR8580 cloned into pER143 cut partially by *Sal*I | This study |
| pER198 | *S.t cbiP & E.c cysG** | *Eco*RI-*Sca*I fragment of pER108 cloned into pER143 cut partially by *Sal*I | This study |

| | | | |
|---|---|---|---|
| pER222 | *B.m cbiD** | *Eco*RI/*Bam*HI PCR fragment from pAR8766 cloned into pAR8668 | This study |
| pER229 | His-tagged B.m cbiF | *Nde*I/*Bam*HI PCR fragment from pAR8766 cloned into pET14b | Raux/98a |
| pER231 | His-tagged *S.t cbiK* | partial *Nde*I/*Bam*HI PCR fragment from pAR8000 cloned into pET14b | This study |
| pER242 | His-tagged *Pd cobA* | *Nde*I/*Bam*HI PCR fragment from pCR395 cloned into pET14b | This study |
| •pER250 | *P.d cobA* & *S.,c MET8* | *Sal*I/*Pst*I PCR fragment from *S.c* genomic DNA cloned into pER119 cut partially by *Sal*I and *Pst*I | This study |
| pER263 | *B.m cbiW-$H_{60}$-X-J-C-D-ET-L-F-G-A-cysG$^A$-cbiY-btuR*-><-ORF1-ORF2 | *Bam*HI/*Sal*I fragment of pAR8877 cloned into pAR8086 (p15A origin) | This study |
| •pER270 | *S.t cbiP* & *B.m cbiW-$H_{60}$-X-J-C-D-ET-L-F-G-A-cysG$^A$-cbiY-btuR*-><-ORF1-ORF2 | *Bam*HI fragment of pAR8531 cloned into pER263 cut with *Bam*HI | This study |
| pER279 | His-tagged *B.m cbiF* | *Eco*RI/*Bam*HI PCR fragment from pER229 cloned into pAR8668 | Raux/98a |
| pER303 | *S.t cbiK* & *P.d cobA* | *Eco*RI/*Bam*HI fragment, from a PCR product of RBS-*S.t cbiK*-stop, cloned into pER119 | This study |
| pER303$^{E89A}$ | *S.t cbiK$^{E89A}$* & *P.d cobA* | *Eco*RV/*Bam*HI fragment of pER304$^{E89A}$ cloned into pER303 | This study |
| pER303$^{H145A}$ | *S.t cbiK$^{H145A}$* & *P.d cobA* | *Eco*RV/*Bam*HI fragment of pER304$^{H145A}$ cloned into pER303 | This study |
| pER303$^{E175A}$ | *S.t cbiK$^{E175A}$* & *P.d cobA* | *Eco*RV/*Bam*HI fragment of pER304$^{E175A}$ cloned into pER303 | This study |
| pER303$^{H207A}$ | *S.t cbiK$^{H207A}$* & *P.d cobA* | *Eco*RV/*Bam*HI fragment of pER304$^{H207A}$ cloned into pER303 | This study |
| pER303$^{D211A}$ | *S.t cbiK$^{D211A}$* & *P.d cobA* | *Eco*RV/*Bam*HI fragment of pER304$^{D211A}$ cloned into pER303 | This study |
| pER303$^{H145A/H207A}$ | *S.t cbiK$^{H145A/H207A}$* & *P.d cobA* | *Eco*RV/*Bam*HI fragment of pER304$^{H145A/H207A}$ cloned into pER303 | This study |
| pER304 | His-tagged *S.t cbiK* | *Xba*I/*Bam*HI fragment of pER231 cloned into M13mp18 | This study |
| pER304$^{E89A}$ | His-tagged *S.t cbiK$^{E89A}$* | E89A mutation obtained from pER304, (mutagenesis primer in Figure 4-15) | This study |
| pER304$^{H145A}$ | His-tagged *S.t cbiK$^{H145A}$* | H145A mutation obtained from pER304, (Figure 4-15) | This study |

| | | | |
|---|---|---|---|
| pER304$^{E175A}$ | His-tagged *S.t cbiK*$^{E175A}$ | E175A mutation obtained from pER304, (mutagenesis primer Figure 4-15) | This study |
| pER304$^{H207A}$ | His-tagged *S.t cbiK*$^{H207A}$ | H207A mutation obtained from pER304, (mutagenesis primer Figure 4-15) | This study |
| pER304$^{D211A}$ | His-tagged *S.t cbiK*$^{D211A}$ | D211A mutation obtained from pER304, (mutagenesis primer Figure 4-15) | This study |
| pER304$^{H145A/H207A}$ | His-tagged *S.t cbiK*$^{H145A/H207A}$ | H207A mutation obtained from pER304$^{H145A}$, (mutagenesis primer Figure 4-15) | This study |
| •pER315 | *B.s ylnF* & *P.d cobA* | *Eco*RI/*Bam*HI PCR fragment from *B.s* genomic DNA cloned into pER119 | This study |
| pET14b | His-tagged expression vector with T7 promoter | | Novagen |
| pGEM-T easy | vector for direct cloning of fragment generated by PCR | | Promega |
| pKK223.3 | overexpression vector derived from pBR322, with *tac* promoter, ampicillin resistant | | Pharmacia |
| pSWG21D | *E. coli cysG*$^{G21D}$ | mutant G21D of *cysG* cloned into *Eco*RI/*Hind*III sites of pKK223.3 | Woodcock /98 |
| pTAg | vector for direct cloning of fragment generated by PCR | | R&D |

The • symbol indicates that the cloning of these plasmids is described in section 8-3.

## 8-1-4- Media and solutions for bacterial work.

Luria-Bertani (LB) broth

| | |
|---|---|
| Tryptone | 10 g |
| Yeast extract | 5 g |
| NaCl | 5 g |

This was made up to 1 litre with distilled water and then autoclaved.

Luria-Bertani agar

15g of Bacto-agar were added to a litre of the LB Broth before autoclaving.

SOC medium

| | |
|---|---|
| Tryptone | 2 g |
| Yeast extract | 0.5 g |
| 5 M NaCl | 0.2 ml |
| 1 M KCl | 0.25 ml |
| 1 M $MgSO_4$ | 1 ml |

This was made up to 100 ml with distilled water and then autoclaved. 1.8ml of sterile 20% glucose solution was then added to constitute SOC medium.

10 x M9 salts

| | |
|---|---|
| $Na_2HPO_4$ | 60 g |
| $KH_2PO_4$ | 30 g |
| $NH_4Cl$ | 10 g |
| NaCl | 5 g |

This was made up to a litre with distilled water and then autoclaved

2x TY medium

| | |
|---|---|
| Tryptone | 16 g |
| Yeast extract | 10 g |
| NaCl | 5 g |

This was made up to 1 litre with distilled water and then autoclaved.

H-top agar

| | |
|---|---|
| Tryptone | 10 g |
| NaCl | 8 g |
| agar | 8 g |

This was made up to 1 litre with distilled water and then autoclaved.

L-plates

| | |
|---|---|
| Tryptone | 10 g |
| Yeast extract | 5 g |
| NaCl | 10 g |
| agar | 15 g |

This was made up to 1 litre with distilled water and then autoclaved.

Chloride calcium solution (0.1 M)

| | |
|---|---|
| $CaCl_2, 2H_2O$ | 0.74 g |

This was made up to 50 ml with distilled water. The solution was either autoclaved and used for minimum medium preparation, or filter sterilised and used to prepare competent cells.

Minimum medium broth and agar

The following reagents were autoclaved separately and subsequently mixed to constitute one litre of minimum medium.

| | |
|---|---|
| 10 x M9 salts | 100ml |
| 20% glucose | 20ml |
| 1 M $MgSO_4$ | 2ml |
| 0.1 M $CaCl_2$ | 1ml |
| $H_2O$ (+/- 15g bacto-agar) | 900ml |

IPTG solution (1 M)

2.38 g of ITPG were dissolved in 10 ml of water and sterilised by filtration (0.2 µm). IPTG was used at a final concentration of 0.4 mM.

X-gal solution (20 mg/ml)

X-gal was dissolved in dimethylformamide and stored in the dark at –20°C.

Hemin solution (5mg/ml)

30 g of hemin was slowly added to 400µl of 0.2 M KOH. 600µl of distilled water, 100µl of 1 M Tris HCl at pH4.8 and 4.8 ml of ethylene glycol were then added in that order. During the whole process, the solution was mixed by vortexing and was finally sterilised by filtration (0.2 µm).

## 8-1-5- Media and solutions for DNA work.

All buffers for DNA work, except electrophoresis buffers, were made with AnalaR water.

50 x TAE buffer

242g Trizma base
37.2g $Na_2EDTA$ pH 8
57.1ml glacial acetic acid

The pH was adjusted at 7.2 and the volume was brought up to 1 litre.

TE buffer

10 mM Tris HCl pH 8
1 mM EDTA pH8

20% PEG/2.5 M NaCl solution

| | |
|---|---|
| Polyethylene glycol 6000 | 20 g |
| NaCl | 14.6 g |

The solution was sterilised by filtration (0.2 µm) and stored at room temperature.

3 M sodium acetate, pH4.8

81.62 g of NaAc. 3 $H_2O$ was dissolved in 150 ml of AnalaR water. The pH was adjusted to 4.8 with glacial acetic acid. AnalaR water was then added to a final volume of 200ml. The solution was sterilised by autoclaving.

10 x One-Phor-All-Buffer + (Pharmacia)

100 mM Tris acetate pH 7.5
100 mM Magnesium acetate
500 mM Potassium acetate

10 x $NH_4$ buffer for PCR (Bioline)

160 mM $(NH_4)_2SO_4$
670 mM Tris-HCl pH8.8
0.1% Tween-20

DNA loading buffer

0.25% bromophenol blue
50% glycerol
50% TE buffer

10 mg/ml RNase A solution

0.1g of RNase A was dissolved in 10 ml of 10 mM Tris HCl pH8 and 15 mM NaCl and boiled for 15 min. This solution was allowed to cool slowly to room temperature and then stored in aliquots at -20°C.

Solutions for minipreparation of plasmids [Sambrook *et al*].

TEGL solution

25 mM Tris HCl pH8
5 mM EDTA
50 mM glucose
5 mg/l lysozyme

The stock solution of Tris HCl pH8, EDTA and glucose was autoclaved prior to mixing with lysozyme. The TEGL solution was kept in 1ml aliquots at –20°C.

NaOH/SDS solution

0.2 M NaOH
1% SDS

3 M potassium acetate, pH 4.8

| | |
|---|---|
| Potassium acetate | 58.88g |
| Acetic acid | ~ 124 ml |

Glacial acetic acid was added until pH 4.8 was attained, at which point water was added to a final volume of 200ml.

Solutions for minipreparation of plasmids (Wizard/Promega).

Solution I, cell resuspension solution

50 mM Tris HCl pH7.5

10 mM EDTA

100 µg/ml RNase A

Solution II, cell lysis solution (see NaOH/SDS solution)

Solution III, neutralisation solution

1.32 M potassium acetate pH4.8

Column Wash solution

200 mM NaCl

20 mM Tris HCl pH7.5

5 mM EDTA

170 ml of 95% ethanol were added to 125 ml of that solution.

## 8-1-6- Media and solutions for protein work.

a- Solutions for protein SDS-polyacrylamide gels.

| Solutions | Separation gel (12% acrylamide) | Stacking gel |
|---|---|---|
| 1.5 M Tris HCl, pH8.8 | 2.5 ml | - |
| 0.5 M Tris HCl, pH6.8 | - | 625 µl |
| 30% acrylamide/bis | 4 ml | 375 µl |
| Water | 3.35 ml | 1.45 ml |
| 10% SDS | 100 µl | 25 µl |
| 10% ammonium persulfate | 43 µl | 12.5 µl |
| TEMED | 10 µl | 2.5 µl |

2 x Sample disruption buffer

| | |
|---|---|
| 0.5 M Tris HCl, pH6.8 | 1 ml |
| Glycerol | 0.8 ml |
| 10% SDS | 1.6 ml |
| β-mercaptoethanol | 0.4 ml |
| Water | 4 ml |
| 0.05% bromophenol blue | 0.2 ml |

10 x Running buffer

| | |
|---|---|
| Trizma base | 30 g |
| Glycine | 144 g |
| SDS | 10 g |

This was made up to one litre with distilled water.

Coomassie blue stain

| | |
|---|---|
| Trichloroacetic acid | 250 ml |
| Coomassie blue R | 0.6 g |
| SDS | 0.1 g |
| Trizma base | 0.25 g |
| Glycine | 0.15 g |

This was made up to 500ml with distilled water.

b- Solutions for His•bind column.

1 x Binding buffer (pH7.9)

| | |
|---|---|
| Imidazole | 5 mM |
| NaCl | 0.5 M |
| Trizma base | 20 mM |

1 x Wash buffer (pH7.9)

| | |
|---|---|
| Imidazole | 60 to 100 mM |
| NaCl | 0.5 M |
| Trizma base | 20 mM |

1 x Elute buffer (pH7.9)

| | |
|---|---|
| Imidazole | 400 to 1000 mM |
| NaCl | 0.5 M |
| Trizma base | 20 mM |

1 x Strip buffer (pH7.9)

| | |
|---|---|
| EDTA | 100mM |
| NaCl | 0.5 M |
| Trizma base | 20 mM |

## 8-2- Molecular biology methods.

### 8-2-1- Isolation of plasmid DNA.

a- Large plasmids (~20kb).

For isolation of plasmids of a size of ~20kb, the classical method for minipreps described in Sambrook *et al.* was followed.

1.5 ml of an overnight culture of *E. coli* were harvested. The supernatant was discarded. The pellet was resuspended in TEGL buffer (100 µl) and lysed by adding NaOH/SDS (200µl). The solution was mixed gently until it became clear. This was incubated on ice for one or two minutes, and potassium acetate solution (150µl) was added. This was mixed gently again and left in ice for 5 minutes. The tube was then centrifuged for 5 minutes and the supernatant was transferred to a new tube and extracted with an equal volume of phenol/chloroform (equilibrated with TE buffer at pH8.0). The upper phase was transferred to a tube and the DNA precipitated by the addition of 1ml of ethanol. After 5 minutes at room temperature, the tube was centrifuged for 5 minutes, the supernatant discarded and the pellet washed with 500µl of 80% ethanol (v/v). The tube was centrifuged for 5 minutes. The pellet was then dried and resuspended in 50µl of TE buffer containing RNaseA at 20 mg/l.

b- Smaller plasmids (~3-10kb)

For smaller plasmid isolation, the rapid Wizard™ minipreps procedure was used. The first steps of this procedure are similar to the previously described method. 100 µl of TEGL is substituted by 200 µl of Solution I; 200 µl of NaOH/SDS is replaced by 200µl of Solution II and finally 150 µl of potassium acetate solution by 200 µl of Solution III. No incubation on ice or at room temperature was undertaken. After addition of Solution III and mixing, the tube was also centrifuged for 5 minutes. The supernatant was transferred into a new tube and 1 ml of Wizard™ resin was added. This mixture was applied to a mini-column connected to a syringe barrel, which allowed the resin to be separated from the buffer. The resin in the mini-column was washed with 2 ml of Column wash buffer. To remove all traces of this buffer, the mini-column was transferred to a new Eppendorf tube and centrifuged for 2 minutes. The mini-column was then transferred onto another tube. 55 µl of TE buffer, warmed to 70-80°C, was applied onto the resin. After 1 minute, the plasmid solution was released by centrifugation (15 seconds).

## 8-2-2- Restriction digest of DNA.

The optimal buffer for the enzyme or a combination of enzymes was chosen according to the Promega data sheet provided. The reactions were, in general, incubated for 2 hours at the temperature required by the enzyme. With *Nde*I, the digest was always performed in 2x One-Phor-All-Buffer + (Pharmacia).

## 8-2-3- Phenol/chloroform extraction and ethanol precipitation.

The DNA solution was diluted to a minimal volume of 50µl with TE buffer. One volume of phenol/chloroform equilibrated with TE buffer at pH8.0 was added to the DNA solution and vortexed. The tube was centrifuged for two minutes to separate the phases and the upper phase was transferred into a new tube. To remove all traces of phenol, a volume of chloroform was added, mixed with

the DNA solution and centrifuged. The upper phase was transferred into a final tube.

The DNA was precipitated by adding 0.1 volume of a 3M sodium acetate solution at pH4.8 and 2 volumes of ethanol. After 5 minutes incubation in a dry-ice/ethanol bath (or 15 minutes at -80°C or overnight at -20°C), the tube was centrifuged for 5 minutes, the supernatant was discarded and the pellet washed with 500 µl of 80% ethanol (v/v). The tube was centrifuged for 5 minutes and the pellet was finally dried and resuspended in an appropriate buffer.

## 8-2-4- Alkaline phosphatase treatment of linear DNA.

To prevent linearised plasmid from religating without an insert, alkaline phosphatase was used to remove the 5' termini phosphates of the vector. After ethanol precipitation, the DNA pellet was dissolved in 25 µl of 1 x One-Phor-All-Buffer + (Pharmacia). 0.05 units of Calf Intestine Phosphatase (CIP) per pmol of linear DNA was used to remove the 5' termini phosphates. For 5' protruding ends, the reaction was incubated for 30 minutes at 37°C whereas for 3' protruding ends or blunt-ended termini, the reaction was incubated for 15 minutes at 37°C, followed by 15 minutes at 56°C. The CIP was removed by digestion with Proteinase K (3 µl of 1 mg/ml) in SE buffer (3 µl of 5% SDS and 50 mM EDTA, pH8.0) for 30 minutes at 56°C.

## 8-2-5- Klenow fragment of polymerase I treatment.

Since the termini of the target fragments and vector may sometimes be incompatible, it is necessary to convert one or both termini of the DNA into blunt ends so they can ligate easily. Recessed 3' termini (for example *Eco*RI) can be filled by the polymerase activity of the Klenow fragment of *E. coli* DNA polymerase I in the presence of 50 µM dNTPs. This reaction is performed in One-Phor-All-Buffer +.

## 8-2-6- Electrophoresis of DNA

The electrophoresis of DNA through an agarose gel was used to separate, identify and purify DNA fragments.

### a- Agarose gel

The agarose gel concentration used was chosen according to the size of DNA fragments to separate as shown in Table 8-1.

Table 8-1- Agarose gel concentrations.

| Amount of agarose in gel (% w/v) | Efficient range of separation of linear DNA molecules (kb) |
|---|---|
| 0.3 | 5 / 60 |
| 0.6 | 1 / 20 |
| 0.7 | 0.8 / 10 |
| 0.9 | 0.5 / 7 |
| 1.2 | 0.4 / 6 |
| 1.5 | 0.2 / 4 |

The appropriate amount of agarose in 1 x TAE buffer was melted in a microwave oven. The solution was allowed to cool to 50°C and then poured onto a gel casting tray, the comb inserted and the gel allowed to set. When the gel had set, the comb was removed and the gel placed in a gel tank (Wide mini sub™ cell) filled with 1 x TAE buffer. The DNA samples, containing 20% loading buffer, were loaded into the wells and electrophoresis was carried out at 50 to 80 V (Biorad Power Pac 300).

When the electrophoresis was completed, the gel was immersed in 1 x TAE containing 1 mg/l ethidium bromide for 30 minutes. Ethidium bromide is a fluorescent dye that intercalates between base pairs.

b- Visualisation under UV

UV radiation (312 nm) was absorbed by the ethidium bromide binding DNA and re-emitted at 590 nm in the red orange wavelength range. Therefore, the DNA was visualised by placing the gel onto an UV transilluminator. The gel was photographed by a camera (Mitsubishi video copy processor) though a red filter.

c- Lambda/*Eco*RI -*Hind*III marker

| | |
|---|---|
| 21,226 bp | 1,584 bp |
| 5,548 bp | 1,375 bp |
| 4,973 bp | 947 bp |
| 4,268 bp | 831 bp |
| 3,530 bp | 564 bp |
| 2,027 bp | 125 bp |
| 1,904 bp | |

**8-2-7- Isolation and purification of a DNA fragment from an agarose gel.**

The DNA band of interest was carefully excised from an agarose gel using a sterile scalpel blade and transferred into a sterile tube. According to the DNA size or to the level of purity required for the next step, one of the two following methods were chosen. The best method for large DNA fragment isolation as well as better quality DNA purification is electroelution. The Geneclean™ method is quicker but it is not as efficient as electroelution.

a- By Geneclean™.

500 µl of sodium iodide were added and the tube incubated for 5 min at 50°C with frequent shaking to dissolve the agarose. 5 µl Glassmilk™ was added and the mixture vortexed, and incubated in ice for 5 minutes. The Glassmilk with bound DNA was centrifuged for 10 seconds and the supernatant discarded. The pellet was washed twice with 500 µl of ice-cold New Wash solution. Finally, the

tube was re-centrifuged and the remaining drop of New Wash solution removed using a fine pipette. The dry Glassmilk with bound DNA was resuspended in 10 µl of sterile AnalaR water, incubated for 5 minutes at 50°C and centrifuged for 30 seconds. The supernatant containing the released DNA was transferred to a new tube. The last step was repeated to ensure a high recovery of the DNA. The 20 µl DNA solution was stored at -20°C.

b- By electroelution into dialysis bags

The agarose slice containing DNA was transferred into a dialysis bag filled with sterile 0.5 x TAE buffer. The gel slice was allowed to sink to the bottom of the bag. The excess of buffer was removed and the bag closed. The bag was immersed in an electrophoresis tank containing 0.5 x TAE buffer. Power was applied at 80 to 100 V for 30 minutes to electroelute the DNA from the gel. The polarity of the current was reversed for 5 seconds to release the DNA from the wall of the bag. All buffer surrounding the gel slice was transferred into a new tube using a fine Pasteur pipette. The bag was washed twice with a small volume of sterile AnalaR water. Phenol, phenol-chloroform and chloroform extractions were undertaken to purify the DNA solution.

## 8-2-8- Ligation of DNA fragments.

After isolation and purification of the vector and insert by agarose gel electrophoresis, the DNA fragments were ligated together (Promega T4 DNA ligase used with its supplied buffer). For cohesive end ligation, a 1/2 ratio of vector (pmol)/insert (pmol) was used. For blunt ended ligation, a 2/1 ratio of vector (pmol)/insert (pmol) was used.

## 8-2-9- Preparation of competent cells (calcium chloride method).

Bacteria were grown to an $OD_{600}$ of 0.3. The cells were incubated in ice for 10 minutes, centrifuged at 4000 rpm for 10 minutes and washed with one volume of 100 mM $MgCl_2$. The cells were afterwards incubated in ice for 20 minutes in 1/2 volume of 100 mM $CaCl_2$. Finally, the cells were resuspended in 1/20

volume of 100 mM $CaCl_2$ and 15% glycerol and aliquoted in small fractions and stored at -80°C.

## 8-2-10- Transformation by the calcium chloride method.

Competent cells were defrosted in ice for 30 minutes. 1 to 7 µl of plasmid DNA solution was added to 100 µl of competent cells and the mixture was left in ice for 15 minutes. The cells were then heat shocked for 2 and a half minutes at 42°C. 200 µl of LB was immediately added and the cells were incubated, without shaking, at 37°C for 30 to 60 minutes for antibiotic resistance expression. This mixture was spread on LB agar plate containing the appropriate antibiotic.

## 8-2-11- PCR reactions.

The PCR machine generally used in this research was a Techne Cyclogene thermal cycler (this apparatus does not require the use of oil on the top of the sample).

### a- Basic PCR

All PCRs were performed as follows:

| | |
|---|---|
| 5 µl | $NH_4$ buffer (Bioline) |
| 2 µl | 4mM dNTPs |
| 1 µl | 100 µM 5′ primer |
| 1 µl | 100 µM 3′ primer |
| 3 to 5 µl | 25 mM $MgCl_2$ |
| 0.1 µl | BioTaq (Bioline) 0.5u/µl |
| 1 µl | template |
| up to 50 µl | $H_2O$ |

0.5 µg of *S. cerevisiae* or 5 to 50 ng of *B. subtilis* genomic DNAs were used as template for a PCR reaction. For plasmid based templates, the solution was diluted about 100 fold to reach a concentration of 1 ng/µl.

The cycles used were:

| | |
|---|---|
| 1 cycle: | 96°C for 3min |
| 30 to 35 cycles: | 96°C for 20 s |
| | 45°C for 20 s |
| | 72°C for 1 min/kb of amplified DNA |
| 1 cycle: | 72°C for 5 min. |

35 cycles were applied for genomic templates but only 30 cycles were required for plasmid templates.

b- Colony PCR:

A colony was touched with a sterile tip and the small quantity of bacteria was resuspended in 10µl of sterile AnalaR $H_2O$.

20 colonies were tested as follow:

| | |
|---|---|
| 160 µl | $H_2O$ |
| 21 µl | $NH_4$ buffer (Bioline) |
| 8.4 µl | 4mM dNTPs |
| 4.2 µl | 100 µM 5′ primer |
| 4.2 µl | 100 µM 3′ primer |
| 12.6 µl | 25 mM $MgCl_2$ |
| 0.5 µl | BioTaq (Bioline) 0.5u/µl |

0.5 µl of each bacterial suspension was placed at the bottom of an Eppendorf tubes and 10µl of the previous mixture was added to each tube. The same cycle program was used as for basic PCR (with 35 cycles).

## 8-2-12- List of primers used in this study.

| Primer | Site | Sequence |
|---|---|---|
| BmcbiF ATG | *Nde*I | C GCG CGC **CAT ATG** AAG TTA TAC ATA ATC GG |
| BmcbiF stop | *Bam*H I | CG **GGA TCC** TCA TTC CGA TTT CAC TCC |
| BmD RBS | *Eco*RI | C GCG C**GA ATT C**AG GAG GAA TTT AAA ATG AAG GAA GTG CCA AAA GAA CCT AAG AAA CTC |
| BmD stop | *Bam*H I | CG CGC **GGA TCC** TTA ATT GCC ATG TTG CAC CGC CTG TCC |
| BsylnF ATG | *Eco*RI | CG **GAA TTC** ATG GAG GTT CAT ATG CTT C |
| BsylnF stop | *Bam*H I | GA **GGA TCC** GCA TTT GTA TGC TCT GGC |
| CbiK E89A | - | G CAC ATT ATT AAC GGC GAC GCA TAT GAA AAA ATT GTC CGT G |
| CbiK H145A | - | GTC GTT TTT ATG GGC GCC GGC GCC AGC CAT CAC |
| CbiK E175A | - | C GGC GCC GTA GCA AGC TAC CCG GAG G |
| CbiK H207A | - | GTG GCG GGC GAT GCA GCC ATT AAT GAT ATG GC |
| CbiK D211A | - | CAC GCC ATT AAT GCT ATG GCT TCA GAC G |
| MET8-RBS | *Sal*I | GC **GTC GAC** GGA GGA TAA AAA ATG GTC AAA TCG CTA CAG CTA G |
| MET8-stop | *Pst*I | GC **CTG CAG** TCA AGA CGA GCA GTA CTC GC |
| PdcobA ATG | *Nde*I | C GCG CGC **CAT ATG** ATC GAC GAC CTC TTT G |
| PdcobA stop | *Bam*H I | CG **GGA TCC** TTA TGC CGG GTT CCT GAG |
| pEtHis | *Eco*RI | CG **GAA TTC** ATG GGC AGC AGC CAT CAT C |
| StcbiK ATG | *Nde*I | C GCG CGC **CAT ATG** AAA AAA GCG CTT CTG G |
| StcbiK-RBS | - | GC GAA GGA GTA AAA ATG AAA AAA GCG C |
| StcbiK stop | *Bam*H I | CG**G GAT CC**T CAT GCC GCC TCC TCT TAC |
| StcbiP ATG | *Nde*I | C GCG CGC **CAT AT G** ACG CAG GCA GTT ATG |
| StcbiP stop | *Bam*H I | CG **GGA TCC** TCA TAC CGG CTC CTG ATG |

The restriction site incorporated in the primers are indicated in bold. The ribosome binding site, methionine codon, stop codon and mutated codon are underlined. Mutated bases are indicates by grey boxes.

## 8-2-13- Site directed mutagenesis.

### a- M13 transformation.

The TG1 competent cells were prepared as described in section 8-2-9. The cells were aliquoted in 300 µl fractions and stored at –80°C. TG1 was transformed with either 1 µl of the M13-His tagged-*cbiK* miniprep or with 100 µl of the

repolymerisation solution which had been previously diluted 10 fold. The cell/DNA mixture was left on ice for 40 minutes. The tubes were heat shocked for 45 seconds at 42°C in a water-bath and returned to ice for 5 minutes. The cell/DNA mixture was then dispatched into three tubes containing 3 µl, 30 µl and the remainder of the 300 µl respectively. 200 µl of lawn cells were added to each tube (fresh culture of TG1 grown in 2x TY medium until the $OD_{600}$ reached about 0.3). The contents of each tube was then added to 4 ml H-top agar kept at 47°C (supplemented with 0.4 mM IPTG and 40 µl of X-gal at 2% for the M13-His tagged-*cbiK* transformation only). After a quick mix, the agar/cell/DNA was poured onto a L-plate and incubated overnight at 37°C.

b- Preparation of single- and double-stranded DNAs from M13.

5 ml of 2x TY medium were inoculated with 50 µl of a fresh culture of TG1. A single plaque was transferred into the medium using a Pasteur pipette. This was incubated with shaking at 37°C for about 5 hours.

The double-stranded DNA was prepared as described in section 8-2-1-a. A chloroform extraction was undertaken after the normal phenol-chloroform treatment to obtain purer DNA.

The single-stranded DNA was prepared as follows: The culture supernatant was re-centrifuged to discard all bacteria. 200 µl of PEG/NaCl solution was added to 1 ml of supernatant, shaken, and left at room temperature for 20 minutes. The phages were pelleted by centrifugation. The last traces of PEG/NaCl solution were removed by a second centrifugation and the pellet was resuspended in 100 µl of TE buffer. The DNA was extracted by adding 50 µl of TE-saturated phenol. The tube was vortexed and left at room temperature for 15 minutes. The upper phase was subsequently submitted to phenol/chloroform and chloroform extractions. The DNA solution was then precipitated. After washing the pellet with 80% ethanol, the single-stranded M13 was resuspended in 10 µl of TE buffer. An $OD_{260}$ was measured to estimate the concentration of the DNA (1 OD of single-stranded DNA is equivalent to 40 µg/ml).

c- Site-directed mutagenesis reactions.

Site-directed mutagenesis was undertaken as indicated in the Sculptor™ manual.

The primer, which encodes for a particular mutation (section 8-2-12), was phosphorylated by the action of the T4 polynucleotide kinase (the reaction was performed in the presence of ATP at a primer concentration of 1.6 pmol/µl).

The phosphorylated primer was annealed to the single-stranded recombinant M13 after denaturation at 70°C. This step was followed by an extension with a mix of deoxynucleotides containing dCTPαS instead of dCTP and Klenow DNA polymerase. The complementary strand was closed by the action of T4 DNA ligase. At this stage, circular double-stranded DNA was produced and the remaining single-stranded DNA or unreligated double-stranded DNA were digested by the T5 exonuclease.

One strand of the circular DNA, which contains dCTP (non-mutant strand), was then nicked by *Nci*I. For the mutation E175A, *Nci*I could not be used since the E175A primer contains that restriction site. *Pvu*I was used as a nicking enzyme for introducing the E175A mutation.

The dCTP strand was partially digested with exonuclease III such that it left a primer for the subsequent repolymerisation/ligation. These last two steps were catalysed by the DNA polymerase I and T4 DNA ligase.

Finally, TG1 was transformed with the new formed circular double-stranded M13 DNA (Figure 8-1) [RPN 1526 Sculptor™ manual, Amersham].

## Figure 8-1- Principle of the mutagenesis process.

primer E89A:

```
5' GCACATTATTAACGGCGACGCATATGAAAAAATTGTCCGTG 3'
     H  I  I  N  G  D  A  Y  E  K  I  V  R
```

M13-*cbiK* single stranded DNA:

```
3'--GTCAGCAACGTGTAATAATTGCCGCTGCTTATACTTTTTTAACAGGCACTTCATGT---- 5'
```

- **Step 1: Phosphorylation of the primer.**

- **Step 2: Annealing:**

```
             GCACATTATTAACGGCGACGCATATGAAAAAATTGTCCGTG
             |||||||||||||||||||| ||||||||||||||||||||
----GTCAGCAACGTGTAATAATTGCCGCTGCTTATACTTTTTTAACAGGCACTTCATGT----
```

- **Step 3: Extension with dCTPαS and ligation:**

```
          ligation
             ▼
--->CAGTCGTTGCACATTATTAACGGCGACGCATATGAAAAAATTGTCCGTGAAGTACA--->
|||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
----GTCAGCAACGTGTAATAATTGCCGCTGCTTATACTTTTTTAACAGGCACTTCATGT----
```

- **Step 4: Removal of all DNA that is non-closed circular double-stranded DNA.**

- **Step 5: Nicking the non- dCTPαS strand:**

An underlined and grey box C indicates a dCTPαS incorporation.

```
---CAGTCGTTGCACATTATTAACGGCGACGCATATGAAAAAATTGTCCGTGAAGTACA----CCGGG--
|||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
---GTCAGCAACGTGTAATAATTGCCGCTGCTTATACTTTTTTAACAGGCACTTCATGT----GGCCC--
                                                              ▲
                                                           Nicking
```

- **Step 6: Partial digest of the nicked strand:**

```
---CAGTCGTTGCACATTATTAACGGCGACGCATATGAAAAAATTGTCCGTGAAGTACA----CCGGG--
                                                          ||||
                                                          --GG
```

- **Step 7: Repolymerisation and ligation:**

```
---CAGTCGTTGCACATTATTAACGGCGACGCATATGAAAAAATTGTCCGTGAAGTACA----CCGGG--
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
<--GTCAGCAACGTGTAATAATTGCCGCTGCGTATACTTTTTTAACAGGCACTTCATGT----GGCCC<-
                                                              ▲
                                                           Ligation
```

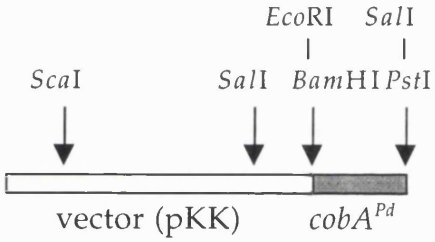
- **Step 8: Transformation of TG1.**

## 8-3- Cloning of different plasmids.

### 8-3-1- Cloning of pER119, *P. denitrificans cobA*.

The *cobA* gene of *P. denitrificans* was subcloned into pAR8668 (variant of pKK223.3, which has a single *Bam*HI site in the MCS) from plasmid pCR395 [Roessner *et al.* 1995] after it has been cut with *Eco*RI and *Pst*I. The plasmid harbouring the *P. denitrificans cobA* was called pER119 (Figure 8-2).

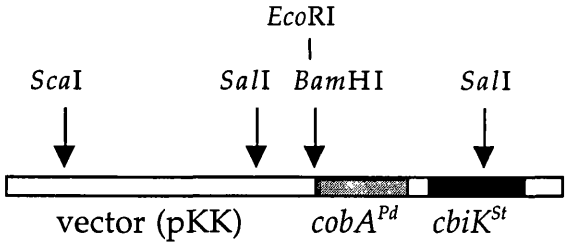
Figure 8-2- Restriction map of pER119:




### 8-3-2- Cloning of pER170, *P. denitrificans cobA* together with *S. typhimurium cbiK*.

pER119 was initially digested with *Sca*I. The linear plasmid was subsequently partially digested with *Sal*I such that it was cut 3' of the *cobA* gene. The 5' overhangs were filled in by the action of the klenow fragment of DNA polymerase I and then treated with alkaline phosphatase. The 4.6 kb fragment was isolated on an agarose gel and purified by Geneclean™. The pKK223.3 plasmid harbouring the *S. typhimurium cbiK*, pAR8580 [Raux *et al.* 1996], was digested with *Eco*RI and *Sca*I. The *Eco*RI overhangs were filled in by the action of the klenow. The 2.5 kb DNA fragment was isolated on an agarose gel and purified. The DNA fragment, including *cbiK*, was ligated into the pER119 vector. The resultant plasmid, with the right restriction map harbouring the two genes, was called pER170 (Figure 8-3).

Figure 8-3- Restriction map of pER170:



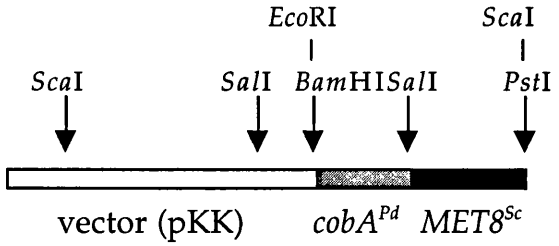

## 8-3-3- cloning of pER126K$^{\Delta}$, *S. typhimurium cob* operon with a deletion in *cbiK.*

pAR8827 was partially restricted with *Hind*III, the overhangs were filled in with the Klenow fragment, and the mixture was loaded on a 0.5% agarose gel. The linear form of the plasmid (about 20.6 kb) was isolated, purified and was ligated with T4 DNA ligase. Removal of the *Hind*III piece contained within *cbiK* resulted in a plasmid that is 54 bp shorter than the original plasmid (pAR8827). The restriction map of pER126K$^{\Delta}$ is shown in Figure 2-8.

## 8-3-4- Cloning of pER250, *P. denitrificans cobA* together with *S. cerevisiae MET8.*

*MET8* was cloned after amplification by PCR from the genomic DNA of *S. cerevisiae*. The primers (MET8-RBS, and MET8-stop) were designed according to the sequence deposited in the database under accession number p15807 (position 552 to 1376) [Hansen *et al.* 1997]. The PCR product obtained with these primers was cloned into a pTAg vector and subsequently subcloned after *Sal*I-*Pst*I digests into pER119 downstream of the *P. denitrificans cobA*. The pER119 vector was previously partially digested with *Sal*I such that a single cut was introduced 3' of the *cobA* gene. The linear plasmid was isolated on agarose gel and purified by electroelution. Finally, the linear plasmid was cut with *Pst*I and purified by Geneclean™ after agarose gel electrophoresis. The plasmid harbouring *cobA* and *MET8* was named pER250 (Figure 8-4).

Figure 8-4- Restriction map of pER250.




## 8-3-5- Cloning of pER270.

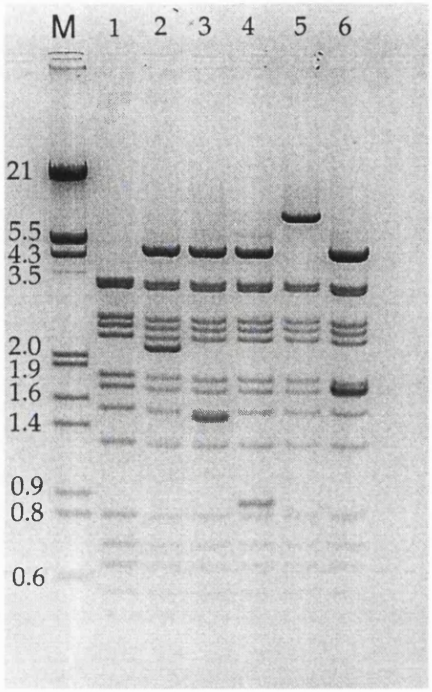
pAR8877, which harbours the whole *B. megaterium cobI* operon, was cut with *Bam*HI and *Sal*I. The DNA fragments obtained were separated on a 0.5% agarose gel and the larger fragment was electroeluted. The vector pAR8086 (pACYC184 with a *lacI*$^{q}$ gene) was also cut with *Bam*HI and *Sal*I and the major fragment was purified by geneclean™. The two purified DNA fragments were ligated together and the resulting plasmid was named pER263.

pER263 was cut with *Bam*HI, the 5′ termini phosphates were removed by the action of calf alkaline phosphatase and the linear plasmid was finally electroeluted. pAR8531 (pKK223.3 harbouring the *S. typhimurium cbiP*) was cut with *Bam*HI and the P*tac-cbiP* fragment was purified by electroelution. Fifteen clones were obtained from the ligation and were tested by colony PCR using StcbiP-ATG and StcbiP-stop primers. The correct resultant plasmid was named pER270. The genetic map is shown in Figure 3-7.

## 8-3-6- Analysis of ER270 derived strains.

*E. coli* 302Δa was transformed with pER270 and the recombinant strain (ER270) was rendered competent. ER270 was transformed with five different second compatible plasmids (pER119, *cobA*; pER170, *cobA* and *cbiK*; pER250, *cobA* and *MET8*; pER108, *cysG* and pSWG21D, *cysG*$^{G21D}$. The *Eco*RI and *Pst*I restriction maps of these plasmids are shown in Figure 8-5 and Table 8-2.

Figure 8-5- Profiles of plasmids from ER270 derived strains.




The marker is lambda cut with *Eco*RI and *Hind*III, the sizes of each fragment are indicated on the right side of the gel. All the other digests have been performed with *Eco*RI and *Pst*I. Lane 1 corresponds to pER270, lane 2 to pER270 and pER108, lane 3 to pER270 and pSWG21D, lane 4 to pER270 and pER119, lane5 to pER270 and pER170 and finally lane 6 to pER270 and pER250.

Table 8-2- Restriction map of plasmids from ER270 derived strains.

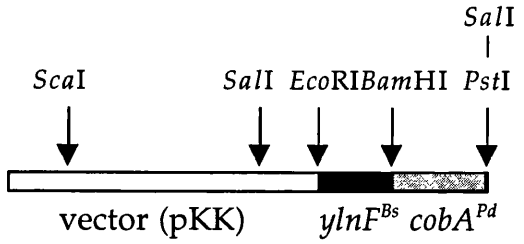
| pER270 | | pER108 | pSWG21D | pER119 | pER170 | pER250 |
|---|---|---|---|---|---|---|
| 3.26 | 1.26 | 4.86 | 4.86 | 4.86 | 7.05 | 4.86 |
| 3.21 | 0.82 | 2.20 | 1.37 | 0.84 | | 1.66 |
| 2.46 | 0.67 | | | | | |
| 2.40 | 0.61 | | | | | |
| 2.31 | 0.51 | | | | | |
| 1.81 | 0.14 | | | | | |
| 1.68 | 0.13 | | | | | |
| 1.51 | 0.60 | | | | | |

All DNA sizes are indicated in kb.

## 8-3-7- Cloning of pER315.

The *Bacillus subtilis* gene *ylnF* was amplified by PCR using the BsylnF-ATG and BsylnF-stop primers. The PCR product was isolated, purified and subsequently cloned into the pGEM-T easy vector. The gene was thus excised from pGEM after an *Eco*RI and *Bam*HI digest and cloned into pER119 that had been cut with the same restriction enzymes. The resultant plasmid was named pER315 (Figure 8-6).

Figure 8-6- Restriction map of pER315.




## 8-4- Microbiological methods.

### 8-4-1- Storage of a bacterial strain.

150 µl of glycerol were added to 850 µl of an overnight bacterial culture. The tube was immediately stored at -80°C.

### 8-4-2- Complementation studies of *S. typhimurium cob* mutant strain.

Competent cells of the *S. typhimurium cob* mutant of interest were made following the chloride calcium procedure and stored at -80°C. The cells were transformed with different plasmids and spread on LB agar plates containing the appropriate antibiotic. The plates were incubated for 24 hours at 37°C. A few colonies were picked from each plate and transferred to agar analytic plates [Bobik *et al.* 1992]. If the strain reacquires the ability to produce vitamin $B_{12}$, the colony develops a red colour.

The *S. typhimurium cysG* mutant could also be studied in minimum medium for cysteine auxotrophy complementation.

## 8-4-3- Quantification of cobyric acid by bioassay.

The cobyric acid quantitative assay is based on the fact that *S. typhimurium metE* cells are reliant upon their alternative $B_{12}$-dependent methionine synthase, encoded by *metH*, for the biosynthesis of methionine. *S. typhimurium metE cysG* (AR3612) was used as an indicator strain to estimate the unknown contents of various samples.

The bioassay plates containing the indicator strain were prepared as follows: bacterial growth was scraped from an overnight minimum medium plate containing methionine and cysteine, washed with 0.9% NaCl and finally mixed with 400 ml of minimum medium agar previously cooled to 47/48°C and containing cysteine. 10 µl sample droplets were placed on the surface of bioassay plates and the plates were incubated for 18 hours at 37°C. The extent of *S. typhimurium* mutant growth (in the bioassay plate agar) circumventing the sample application point is dependent upon the amount of cobyric acid present in the sample. To generate a standard curve, 10 µl samples containing 0.01, 0.1, 1, 10 and 40 pmol of vitamin $B_{12}$ were also assayed for each set of experiments (Figures 8-7 and 8-8).

Figure 8-7- Principle of the bioassay plates.

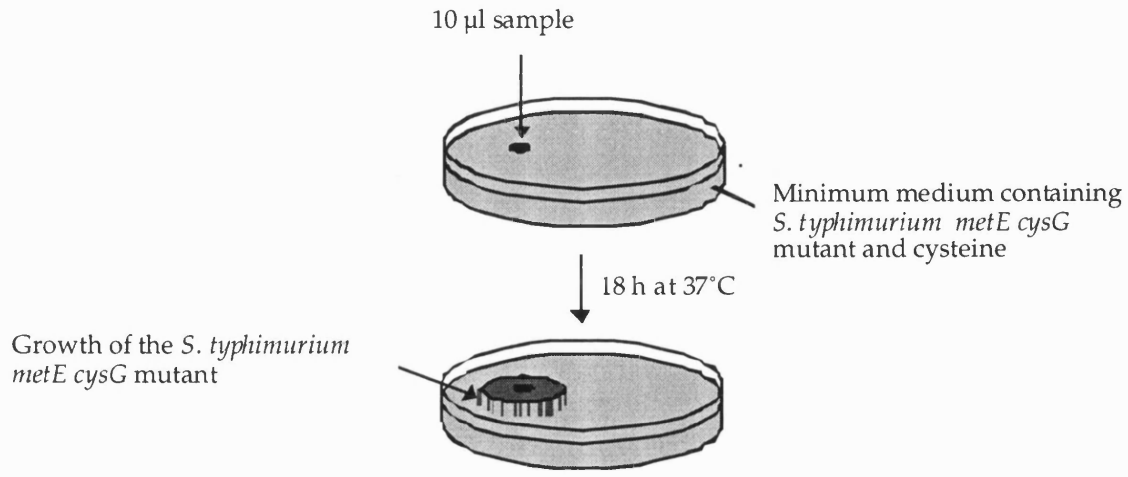
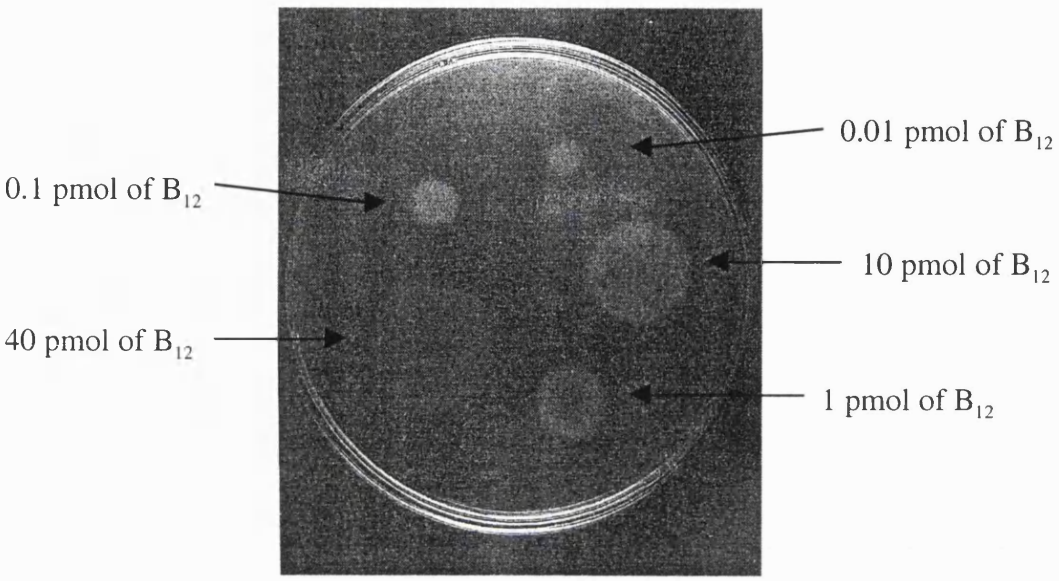
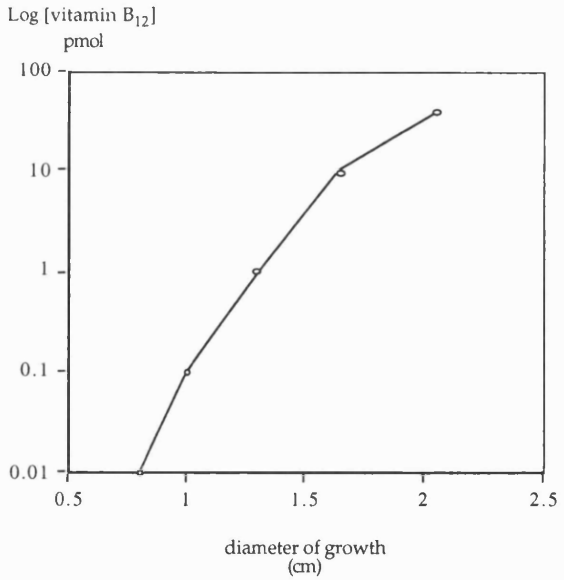


Figure 8-8- Standard curve obtained from the bioassay plates.

a- The biossay plate.




b- The Standard curve.




## 8-4-4- Preparation of bacteria for cobalamin intermediate production.

To estimate the level of cobalamin production in various strains, the bacteria of interest were grown under anaerobic conditions in minimum medium supplemented with antibiotics and 0.1 g/l yeast extract. After 6 hours at 37°C,

ITPG was added at 0.4 mM and the cultures were left in the incubator for 18 hours. The cells were then centrifuged, resuspended in water at an $OD_{600}$ of 8 and finally sonicated. A 10µl droplet of the sonicated extract was placed on the *S. typhimurium metE cysG* bioassay plate.

## 8-5- Biochemical methods

### 8-5-1- Protein assay (Bradford).

The Bradford assay relies on the dye coomassie blue G250 binding to proteins, and most readily to arginyl and lysyl residues. The absorbance maxima of the dye shifts from the red cationic form (470 to 650 nm) to the blue anionic form (590 to 620 nm) in presence of proteins. Thus, the quantity of protein can be estimated by measuring the amount of dye in the blue form at 595 nm [Bradford, 1976].

200 µl of dye protein assay from Biorad were added to 1 to 200 µl of protein solution diluted up to 800 µl of water. The reaction was left for 5 minutes at room temperature and the $OD_{595}$ was measured. A standard curve was generated with bovine serum albumin (5, 10, 15, 20 and 25 µg) on the same time.

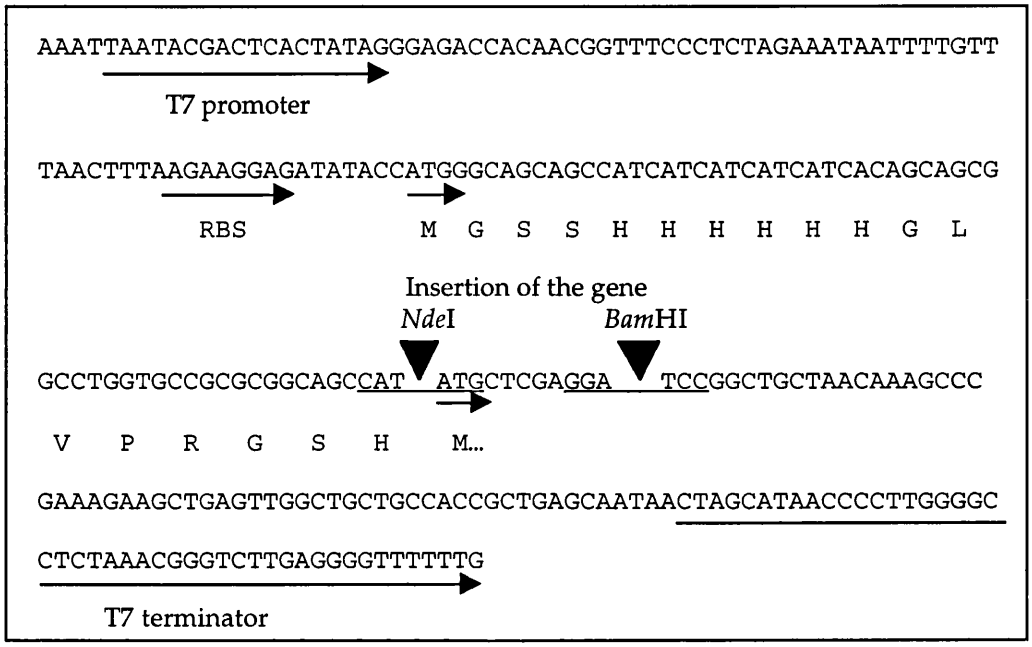
### 8-5-2- Purification of His-tagged proteins.

#### a- Recombinant plasmids and bacterial growth

The gene of interest is cloned into the pET14b vector so that (i) the xxxATG of the gene is substituted by a *Nde*I restriction site (CATATG) and (ii) a *Bam*HI site is introduced in 3′ of the stop codon. The PCR product is then either cloned directly into the *Nde*I and *Bam*HI sites of the pET14b (Figure 8-9) or into the pGEM-T easy vector and subsequently into pET14b.

The N-terminus of the recombinant protein is therefore fused to the C-terminus of a short peptide containing a stretch of six consecutive histidines.

Figure 8-9- pET14b.




BL21(DE3)(pLysS) was transformed with pET14b containing a gene cloned in frame. The recombinant strain was grown in LB medium up to an $OD_{600}$ of about 0.6. IPTG was then added to a final concentration of 0.4mM. The cells were left at 37°C for a further 2 hours.

b- Sonication of bacteria

Harvested cells were resuspended in 1 x Binding buffer. The cells were broken by sonication, using a Soniprep 150, at an amplitude of 8 to 10 microns in 3 bursts of 30 seconds.

c- His•bind affinity column

The His•Tag sequence of the fused protein binds to divalent cations ($Co^{2+}$, $Ni^{2+}$) immobilized on the His•bind resin. After unbound proteins are washed away,

the His•tag protein is eluted with imidazole. 5 ml of resin was used to purify proteins from 1 litre of culture.

The resin was initially rinsed with 15 ml water, charged with 25 ml of a 50 mM divalent cation solution (Charge Buffer) and equilibrated with 15 ml Binding Buffer. The soluble fraction of the sonicated extract was applied to the column. The column was then washed with 50 ml of binding buffer and with 30 ml of Wash Buffer. 30 ml of Elute buffer was applied to the column to recover the target protein. The resin was regenerated by applying a solution containing EDTA (Strip Buffer) [pET system manual, seventh edition, 1997, Novagen].

d- Protein desalting

The proteins were desalted by either dialysis against 50 mM Tris-HCl at pH 7.8 or by gel filtration using a sephadex G25 (PD10 column, Pharmacia) equilibrated in the same buffer.

e- Purification of an insoluble protein, the *B. megaterium* CbiD.

Strain ER222 (JM101 containing pKK harbouring the *B. megaterium cbiD*) was grown in LB supplemented with ampicillin. When the $OD_{600}$ reached 0.15, the synthesis of CbiD was induced by the action of IPTG and a further incubation for 90 minutes at 37°C followed. CbiD was only found among the insoluble protein fraction after sonication and centrifugation. The inclusion body fraction was resolubilised as follows. The pellet of sonication was washed with inclusion body wash buffer (50 mM Tris-HCl at pH7.8, 100 mM NaCl, 15 mM EDTA, 0.5% Triton X-100) and resuspended in 50 mM Tris-HCl at pH7.8, 6 M Guanidine, 0.5M NaCl, 5 mM β-mercaptoethanol. The protein was then left at room temperature for 30 minutes. The CbiD protein was subsequently dialysed at 4°C against the same solution but without guanidine. The dialysed protein was centrifuged to discard all remaining trace of precipitate.

## 8-5-3- SDS-polyacrylamide gel electrophoresis.

The composition of the gel mixes, buffers, stain and destain solutions were previously shown in section 2-1-6. The samples were denatured by adding an equal volume of disruption buffer and then boiled for 5 minutes. A molecular weight marker was loaded for each gel. Electrophoresis was performed at a constant voltage of 200 V until the blue dye reached the bottom of the gel (Dual Mini Slab AE6450, Genetics Research Instrumentation Ldt). The gel was stained with the coomassie blue stain and then destained with 7% acetic acid. The gel was finally rinsed in water and dried between two sheets of cellophane (BioRad).

Molecular weight markers for denaturing polyacrylamide gels (SDS7 from Sigma)

| Protein | kD |
|---|---|
| Bovine serum albumin | 66.0 |
| Ovalbumin | 45.0 |
| G3P dehydrogenase | 36.0 |
| Carbonic anhydratase | 29.0 |
| Trypsinogen | 24.0 |
| Soybean trypsin inhibitor | 20.1 |
| α-lactalbumin | 14.2 |

## 8-5-4- SAM binding assay/Gel filtration

0.25 to 1 mg of protein was incubated with 0.5 µCi of [methyl-$^3$H] SAM in a final volume of 400 µl for 30 min at 37°C. The mixture was applied to the Sephadex G-25 column (1 x 20cm) and 20 fractions of 1 ml were collected. All fractions were analysed for their radioactivity content by adding 1 ml of EcoScintA fluid (National Diagnostic) to each fraction. All tubes were counted for 1 min using a scintillation counter [Crouzet *et al.* 1990 and Woodcock *et al.* 1996].

## 8-5-5- UV-visible spectrophotometry.

The accumulated modified tetrapyrroles were extracted from sonicated cell extracts by binding the chromophore onto DEAE sephacel resin. The resin was washed with a solution of 0.1 M NaCl and 50 mM Tris-HCl at pH8.0. The chromophore was eluted by applying a solution of 1 M NaCl in 50 mM Tris-HCl at pH8.0 onto the column. The spectrum was immediately recorded in a Hewlett-Packard 8452A diode array spectrophotometer over a range of 300 to 700 nm.

## 8-5-6- LiChroprep RP18 column.

To extract cobyric acid from 1 litre of bacterial culture, 3 g of LiChroprep RP18 resin were used. The resin was initially resuspended in hexane and left to decant. Hexane was discarded and the resin washed twice with 95% ethanol. The resin was finally poured in a column and equilibrated with $H_2O$.

The bacterial extract was applied to the column and washed with two volumes of water. Cobyric acid was eluted from the column by application of 25% *tert*-butanol (v/v).

Cobyric acid-containing fractions were detected by bioassay and were stored at –20°C after lyophilisation.

# REFERENCES

Al-Karadaghi, S., Hansson, M., Nikonov, S., Jonsson, B. and Hedersted, L. (1997) Crystal structure of ferrochelatase: the terminal enzyme in heme biosynthesis. *Structure*. **5**: 1501-1510.

Battersby, A. R., and Sheng, Z. C. (1982) Preparation and spectroscopic properties of $Co^{III}$-isobacteriochlorins: relationship to the cobalt-containing proteins from *Desulfovibrio gigas* and *D. desulphuricans*. *Chem. Commun.* (J. Chem. Soc. Sect. **D**.): 1393-1394.

Battersby, A. R. (1994) How nature builds the pigments of life: the conquest of vitamin $B_{12}$. *Science*. **264**: 1551-1557. (Published erratum appears in *Science*. 1994: **265**, 1159).

Beale, S. I. and Weinstein, J. D. (1991) Biochemistry and regulation of photosynthetic pigment formation in plants and algae. In P. M. Jordan (Ed.) Biosynthesis of tetrapyrroles: 155-236.

Beck, R., Raux, E., Thermes, C., Rambach, A. and Warren, M. J. (1997) CbiX: A novel metal-binding protein involved in sirohaem biosynthesis *in Bacillus megaterium*. *Biochem. Soc. Trans*. **25**: 77S.

Beck, R. PhD thesis, University College London, 1997.

Blanche, F., Debussche, L., Thibaut, D., Crouzet, J. and Cameron, B. (1989) Purification and characterization of S-adenosyl-l-methionine-uroporphyrinogen-III methyltransferase from *Pseudomonas denitrificans*. *J. Bacteriol*. **171**: 4222-4231.

Blanche, F., Thibaut, D., Frechet, D., Vuilhorgne, M., Crouzet, J., Cameron, B., Hlineny, K., Traubeberhard, U., Zboron, M. and Muller, G. (1990) Hydrogenobyrinic acid - isolation, biosynthesis, and function. *Angew. Chem. Int. Ed. Engl*. **29**: 884-886.

Blanche, F., Couder, M., Debussche, L., Thibaut, D., Cameron, B. and Crouzet, J. (1991) Biosynthesis of vitamin-$B_{12}$ - stepwise amidation of carboxyl group-

b, group-d, group-e, and group-g of cobyrinic acid a,c-diamide is by one enzyme in *Pseudomonas denitrificans*. *J. Bacteriol.* **173**: 6046-6051.

Blanche, F., Debussche, L., Famechon, A., Thibaut, D., Cameron, B. and Crouzet, J. (1991) A bifunctional protein from *Pseudomonas-denitrificans* carries cobinamide kinase and cobinamide phosphate guanylyltransferase activities. *J. Bacteriol.* **173**: 6052-6057.

Blanche, F., Thibaut, D., Famechon, A., Debussche, L., Cameron, B. and Crouzet, J. (1992) Precorrin-6x reductase from *Pseudomonas-denitrificans* - purification and characterization of the enzyme and identification of the structural gene. *J. Bacteriol.*, **174**: 1036-1042.

Blanche, F., Famechon, A., Thibaut, D., Debussche, L., Cameron, B. and Crouzet, J. (1992b) Biosynthesis of vitamin $B_{12}$ in *Pseudomonas denitrificans*: the biosynthetic sequence from precorrin-6y to precorrin-8x is catalyzed by the *cobL* gene product. *J. Bacteriol.* **174**: 1050-1052.

Blanche, F., Maton, L., Debussche, L. and Thibaut, D. (1992c) Purification and characterization of cob(II)yrinic acid *a,c*-diamide reductase from *Pseudomonas denitrificans*. *J. Bacteriol.* **174**: 7452-7454.

Blanche, F., Cameron, B., Crouzet, J., Debussche, L., Thibaut, D., Vuilhorgne, M., Leeper, F. J. and Battersby, A. R. (1995) vitamin $B_{12}$: How the problem of its biosynthesis was solved. *Angew. Chem. Int. Ed. Engl.* **34**: 383-411.

Bobik, T., Ailion, A. M. and Roth, J. R. (1992) A single regulatory gene integrates control of vitamin $B_{12}$ synthesis and propanediol degradation. *J. Bacteriol.* **174**: 2253-2266.

Bradford, M. M. (1976) A rapid and sensitive method for quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**: 248-254.

Brey, R. N., Banner, C. D. B. and Wolf, J. B. (1986) Cloning of multiple genes involved with cobalamin (vitamin $B_{12}$) biosynthesis in *Bacillus megaterium. J. bacteriol.* **167**: 623-630.

Brushaber, K. R., O'Toole, G.A. and Escalante-Semerena, J. C. (1998) CobD, a novel enzyme with L-threonine-*O*-3-phosphate decarboxylase activity, is responsible for the synthesis of (*R*)-1-amino-2-propanol *O*-2-phosphate, a proposed new intermediate in cobalamin biosynthesis in *Salmonella typhimurium. J. Biol. Chem.* **273**: 2684-2691.

Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. -F., Aams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. M., Weidman, J. F., Fuhrmann, J. L., Presley, E. A., Nguyen, D., Utterback, T. R., Kelley, J. M., Peterson, J. D., Sadow, P. W., Hanna, M. C., Cotton, M. D., Hurst, M. A., Roberts, K. M., Kaine, B. P., Borodovsky, M., Klenk, H. -P., Fraser, C. M., Smith, H. O., Woese, C. R. and Venter J. C. (1996) Complete genome sequence of the Methanogenic archaeon, *Methanococcus jannaschii. Science.* **273**: 1058-1073.

Cameron, B., Blanche, F., Rouyez, M. C., Bisch, D., Famechon, A., Couder, M., Cauchois, L., Thibaut, D., Debussche, L., Crouzet, J. (1991) Genetic analysis, nucleotide-sequence, and products of 2 *Pseudomonas cob* genes encoding nicotinate-nucleotide-dimethylbenzimidazole phosphoribosyl-transferase and cobalamin (5'-phosphate) synthase. *J. Bacteriol.* **173**: 6066-6073.

Chang, C. K. (1993) Haem *d*1 and other haem cofactors from bacteria. In Ciba foundation Symposium 180. The biosynthesis of tetrapyrrole pigments: 228-246.

Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E. III., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M. A.,

Rajandream, M-A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, S., Squares, R., Sulston, J. E., Taylor, K., Whitehead, S. and Barrell, B. G. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*. **393**: 537-544.

Crane, B. R., Siegel, L. M.and Getzoff, E. D. (1995) Sulfite reductase structure at 1.6 angstrom - Evolution and catalysis for reduction of inorganic anions. *Science*. **270**: 59-67.

Crouzet, J., Cauchois, L., Blanche, F., Debussche, L., Thibaut, D., Rouyez, M. C., Rigault, S., Mayaux, J. F. and Cameron, B. (1990) Nucleotide sequence of *Pseudomonas denitrificans* 5,4-kilobase DNA fragment containing five *cob* genes and identification of structural genes encoding *S*-adenosyl-methionine:uroporphirinogen III methyltransferase and cobyrinic acid *a,c*-diamide synthase. *J. Bacteriol.* **172**: 5968-5979.

Crouzet, J., Cameron, B., Cauchois, L., Rigault, S., Rouyez, M. C., Blanche, F., Thibaut, D. and Debussche, L. (1990) Genetic and sequence analysis of an 8,7-kilobase *Pseudomonas denitrificans* fragment carrying eight genes involved in the transformation of precorrin-2 to cobyrinic acid. *J. Bacteriol.* **172**: 5980-5990.

Crouzet, J., Levy-Schil, S., Cameron, B., Cauchois, L., Rigault, S., Rouyez, M. C., Blanche, F. and Debussche, L. (1990) Nucleotide sequence and genetic analysis of a 13,1-kilobase pair *Pseudomonas denitrificans* DNA fragment containing five *cob* genes and identification of structural genes encoding cob(I)alamin adenosyltransferase, cobyric acid synthase, and a bifunctional cobinamide kinase-cobinamide phosphate guanylyltransferase. *J. Bacteriol.* **173**: 6074-6087.

Debussche, L., Thibaut, D., Cameron, B., Crouzet, J. and Blanche, F. (1990) Purification and characterization of cobyrinic acid a,c-diamide synthase from *Pseudomonas-denitrificans*. *J. Bacteriol.* **172**: 6239-6244.

Debussche, L., Couder, M., Thibaut, D., Cameron, B., Crouzet, J. and Blanche, F. (1991) Purification and partial characterization of

cob(I)adenosyltransferase from *Pseudomonas denitrificans*. *J. Bacteriol.* **173**: 6300-6302.

Debussche, L., Couder, M., Thibaut, D., Cameron, B., Crouzet, J. and Blanche, F. (1992) Assay, purification, and characterization of cobaltochelatase, a unique complex enzyme catalyzing cobalt insertion in hydrogenobyrinic *a,c*-diamide during coenzyme $B_{12}$ biosynthesis in *Pseudomonas denitrificans*. *J. Bacteriol.* **174**: 7445-7451.

Debussche, L., Thibaut, D., Cameron, B., Crouzet, J. and Blanche, F. (1993) Biosynthesis of the Corrin Macrocycle of coenzyme $B_{12}$ in *Pseudomonas denitrificans*. *J. Bacteriol.* **175**: 7430-7440.

Ermler, U., Grabarse, W., Shima, S., Goubeaud, M. and Thauer, R. K. (1997) Crystal structure of methyl-coenzyme M reductase: the key enzyme of biological methane formation. *Science.* **278**: 1457-1462.

Fazzio, T. G. and Roth, J. R. (1996) Evidence that the CysG protein catalyses the first reaction specific to $B_{12}$ synthesis in *Salmonella typhimurium*, insertion of cobalt. *J. bacteriol.* **178**: 6952-6959.

Fenton, W. A. and Rosenberg, L. E. (1989) Inherited disorders of cobalamin transport and metabolism. In Scriver, C. R. *et al.*, 6$^{th}$ edition. The metabolic basis of inherited disease. Chapter **82**: 2065-2082.

Ferreira, G. C., Franco, R., Lloyd, S. G., Moura, I., Moura, J. J. G. and Huynh, B. H. (1995) Structure and function of ferrochelatase. *J. Bioenergetics and biomembranes*. **27**: 221-229.

Friedmann, H. C., Klein, A. and Thauer, R. K. (1991) Biochemistry of coenzyme F430, a nickel porphinoid involved in methanogenesis. In P. M. Jordan (Ed.) Biosynthesis of tetrapyrroles: 139-155.

Fu, C. L., Olson, J. W. and Maier, R. J. (1995) HypB protein *of Bradyrhizobium japonicum* is a metal-binding GTPase capable of binding 18 divalent nickel ions per dimer. *Proc. Natl. Acad. Sci. USA*. **92**: 2333-2337.

Goldman, B. S. and Roth, J. R. (1993) Genetic structure and regulation of the *cysG* gene in *Salmonella typhimurium*. *J. bacteriol.* **175**: 1457-1466.

Griffiths, L. A., and Cole, J. A. (1987) Lack of redox control ot the anaerobically-induced *nirB*$^+$ gene of *Escherichia coli* K-12. *Arch. Microbiol.* **147**: 364-369

Hansen, J., Muldbjerg, M., Cherest, H., and Surdin-Kerjan, Y. (1997) Siroheme biosynthesis in *Saccharomyces cerevisiae* requires the products of both the *MET1* and *MET8* genes. *FEBS Letters*. **401**: 20-24.

Harm, U. and Thauer, R. K. (1996) The energy conserving methyltransferase complex from *Methanobacterium thermoautotrophicum*: reconstitution and characterisation of the corrinoid -containing 23kD subunit MtrA overproduced in *E. coli*. ). In the 4$^{th}$ European Symposium on vitamin $B_{12}$ and $B_{12}$-proteins. Abstract L21.

Hatchikian, E. C. (1981) A cobalt porphyrin containing protein reducible by hydrogenase isolated from *Desulfovibrio-desulfuricans* (Norway) *Biochem. Biophys. Res. Commun.* **103**: 521-530.

Ishida, T., Yu, L., Akuru, H., Azawa, K., Kawanishi, S., Seto, A., Inubushi, T. and Sano, S. (1998) A primitive pathway of porphyrin biosynthesis and enzymology in desulfovibrio vulgaris. *Proc. Natl. Acad. Sci.* USA. **95**: 4853-4858.

Jensen, P. E., Gibson, L. C. D. and Hunter, C. N. (1998) Determinants of catalytic activity with the use of purified I, D and H subunits of the magnesium protoporphin IX chelatase from Synechocystis PCC6803. *Biochem. J.* **334**: 335-344.

Jeter, R. M., Olivera, B. M., and Roth, J. R. 1984. *Salmonella typhimurium* synthesizes cobalamin (vitamin $B_{12}$) de novo under anaerobic growth conditions. *J. Bacteriol.* **159**: 206-213.

Jordan, P. M. (1994) The biosynthesis of uroporphyrinogen III mechanism of action of porphobilinogen deaminase. In Ciba foundation Symposium 180. The biosynthesis of tetrapyrrole pigments: 70-96.

Kaneko, T., Tanaka, A., Sato, S., Kotani, H., Sazuka, T., Miyajima, N., Sugiura, M. and Tabata, S. (1995) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis sp.* strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome. *DNA Res.* **2**: 153-166.

Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M. and Tabata, S. (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis sp.* strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3**: 109-136.

Kannangara, C. G, Andersen, R. V., Pontoppidan, B., Willows, R. and Von Wettstein, D. (1994) Enzymic and mechanistic studies on the conversion of glutamate to 5-aminolaevulinate. In Ciba foundation Symposium 180. The biosynthesis of tetrapyrrole pigments: 3-25.

Klenk, H. P., Clayton, R. A., Tomb, J. F., White, O., Nelson, K. E., Ketchum, K. A., Dodson, R. J., Gwinn, M., Hickey, E. K., Peterson, J. D., Richardson, D. L., Kerlavage, A. R., Graham, D. E., Kyrpides, N. C., Fleischmann. R. D., Quackenbush, J., Lee, N. H., Sutton, G. G., Gill, S., Kirkness, E. F., Dougherty, B. A., McKenney, K., Adams, M. D., Loftus, B., Venter, J. C., *et al.* (1997) The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature.* **390**: 364-70. (Published erratum appears in *Nature.* (1998) **394**: 101).

Kredich, N. M. (1987) In *Escherichia coli* and *Salmonella typhimurium* cellular and molecular biology (Neihardt, F.C., Ed.) *American Society for Microbiology*: 419-428.

Lawrence, J. G. and Roth, J. R. (1995) The cobalamin (Coenzyme $B_{12}$) biosynthetic genes of *Escherichia coli*. *J. bacteriol.* **177**: 6371-6380.

Lawrence, J. G. and Roth, J. R. (1996) Evolution of coenzyme B12 synthesis among enteric bacteria: Evidence for loss and reacquisition of a multigene complex. *Genetics*. **142**: 11-24.

Low, S. C. and Berry, M. J. (1996) Knowing when to stop: selenocysteine incorporation in eukaryotes. *TIBS* (reviews). **21**: 203-207.

Luque, I., Flores, E. and Herrero, A. (1993) Nitrite reductase gene from *Synechococcus sp* PCC-7942 - homology between cyanobacterial and higher-plant nitrite reductases. *Plant Molecular Biology*. **21**: 1201-1205.

Mac Veigh, T., Wilkie, S. E., Woodcock, S. C., Gulnahar, B., Mortuza, Peters, S. E. and Warren, M. J. Sirohaem biosynthesis in *Neisseria meningitidis.* Unpublished data.

Malone, T., Blumenthal, R. M. and Cheng, X. (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J Mol Biol*. **253**: 618-632.

Meyer, J. (1993) Cloning and sequencing of the gene encoding the [2Fe-2S] ferredoxin from *Clostridium-pasteurianum. Biochim. Biophys. Acta*. **1174**: 108-110.

Milton, D. L., Norqvist, A. and Wolf-Wartz, H. (1995) Sequence of a novel virulence-mediating gene, *virC*, from *Vibrio anguillarum. Gene*. **164**: 95-100.

Min, C. H., Atshaves, B. P., 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. American Chem. Soc.* **115**: 10380-10381.

Moura, J. J. G., Moura, I., Brushi, M., Le Gall, J. and Xavier A. V (1980) The cobalt containing protein isolated from *Desulfovibrio gigas*, a sulfate reducer. *Biochem. Biophys. Res. Commun.* **92**: 962-970.

Müller, G., Zipfel, F., Hlinemy, K., Savvidis, E., Hertle, R., Traub-Eberhard, U., Scott, A. I., Williams, H. J., Stolowich, N. J., Santander, P. J., Warren, M. J., Blanche, F. and Thibaut, D. (1991) Timing of cobalt insertion in vitamin $B_{12}$ biosynthesis. *J. Am. Chem. Soc.* **113**: 9893-9895.

O'Toole, G. A., Trzebiatowski, J. R. and Escalante-semerena, J. C. (1994) The *cobC* gene of *Salmonella typhimurium* codes for a novel phosphatase involved in the assembly of the nucleotide loop of cobalamin. *J. Biol. Chem.* **269**: 26503-26511.

Peakman, T., Crouzet, J., Mayaux, J. F., Busby, S., Mohan, S., Harborne, N., Wootton, J., Nicolson, R. and Cole, J. (1990) Nucleotide-sequence, organization and structural-analysis of the products of genes in the *nirB-cysG* region of the *Escherichia coli* K-12 chromosome. *Eur. J. Biochem.* **191**: 315-323.

Raux, E., Lanois, A., Levillayer, F., Warren, M. J., Brody, E., Rambach, A. and Thermes, C. (1996) *S. typhimurium* cobalamin (vitamin $B_{12}$) biosynthetic genes; functional studies in *S. typhimurium* and *E. coli*: *J. Bacteriol.* **178**: 753-767.

Raux, E., Thermes, C., Heathcote, P., Rambach, A., and Warren, M. J. (1997) The role of the *Salmonella typhimurium cbiK* in cobalamin (vitamin $B_{12}$) and siroheme biosynthesis. *J. Bacteriol.* **179**: 3202-3212.

Raux, E., Schubert, H. L., Woodcock, S. C., Wilson, K. S., and Warren, M. J. (1998 a) Cobalamin (vitamin $B_{12}$) biosynthesis: Cloning, expression and crystallisation of the *Bacillus megaterium* S-adenosyl-L-methionine dependent cobalt-precorrin-4 transmethylase CbiF. *Eur. J. Biochem.* **254**: 341-346.

Raux, E., Lanois, A., Warren, M. J., Rambach, A., and Thermes, C. (1998 b) Cobalamin (vitamin $B_{12}$) biosynthesis: Identification and characterisation of a *Bacillus megaterium cob* operon. *Biochem. J.* **335**: 159-166.

Raux, E., Lanois, A., Rambach, A., Warren, M. J., and Thermes, C. (1998 c) Cobalamin (vitamin $B_{12}$) biosynthesis: Functional characterisation of the *Bacillus megaterium cbi* genes required to convert uroporphyrinogen III into cobyrinic acid *a,c*-diamide. *Biochem. J.* **335**: 167-173.

Rey, L., Imperial, J., Palacios, J. M. and Ruizargueso, T. (1994). Purification of *Rhizobium leguminosarum* HypB, a nickel-binding protein required for hydrogenase synthesis. *J. bacteriol.* **176**: 6066-6073.

Robin, C., Blanche, F., Cauchois, L., Cameron, B., Couder, M. and Crouzet, J. (1991) Primary structure, expression in *Escherichia coli*, and properties of S-adenosyl-L-methionine:uroporphyrinogen III methyltransferase from *Bacillus megaterium*. *J. Bacteriol.* **173**: 4893-4896.

Roessner, C. A., Spencer, J. B., Ozaki, S., Min, C., Atshaves, B. P., Nayar P., Anousis, N., Stolowich, N. J., Holderman, M. T., Scott, A. I. (1995). Overexpression in *Escherichia coli* of 12 vitamin $B_{12}$ biosynthetic enzymes. *Protein Expr. Purif.* **6**: 155-163.

Roessner, C. A., Huang, K. and Scott, A. I. (1996) Isolation and characterization of genes required for vitamin $B_{12}$ biosynthesis in *Propionibacterium freudrenreichii* (*shermanii*). In the 4$^{th}$ European Symposium on vitamin $B_{12}$ and $B_{12}$-proteins. Abstract P5.

Roth, J. R., Lawrence, J. G., Rubenfield, M., Kieffer-Higgins, S. and Church, G. M. (1993) Characterization of cobalamin (vitamin $B_{12}$) biosynthetic genes of *Salmonella typhimurium*. *J. Bacteriol.* **175:** 3303-3316.

Sambrook, J., Fritsh, E. F. and Maniatis, T. (1989) Molecular Cloning. Cold Spring Harbour Laboratory Press.

Santander, P. J., Roessner, C. A., Stolowich, N. J., Holderman, M. T. and Scott, A. I. (1997) How corrinoids are synthesized without oxygen: nature's first pathway to vitamin $B_{12}$. *Chemistry & Biology*. **4**: 659-666.

Sasarman, A., Suredeanu, M. and Horodniceanu, T. (1968) Locus determining the synthesis of δ-aminolevulinic acid in *Escherichia coli* K-12. *J. Bacteriol.* **96**: 1882-1884.

Sattler, I., Roessner, C. A., Stolowich, N. J., Hardin, S. H., Harris-Haller, L. W., Yokubaitis, N. T., Murooka, Y., Hashimoto, Y. and Scott, A. I. (1995) Cloning, sequencing, and expression of the uroporphyrinogen III methyltransferase *cobA* gene of *Propionibacterium freudenreichii (shermanii)*. *J Bacteriol.* **177**: 1564-1569.

Sawers, G., Heider, J., Zehelein, E. and Bock, A. (1991) Expression and operon structure of the *sel* genes of *Escherichia-coli* and identification of a third selenium-containing formate dehydrogenase isoenzyme. *J. Bacteriol.* **173**: 4983-4993.

Schluckebier, G., O'Gara, M., Saenger, W and Cheng, X. (1995) Universal catalytic domain structure of AdoMet-dependent methyltransferases. *J. Mol. Biol.* **247**: 16-20.

Schubert, H. L., Wilson, F. S., Raux, E., Woodcock, S. C. and Warren, M. J. (1998) The X-ray structure of a cobalamin biosynthetic enzyme, cobalt-precorrin-4 methyltransferase. *Nature Structure Biology*. **5**: 585-592.

Scott, A. I., Warren, M. J., Roessner, C. A., Stolowich, N. J. and Santander, P J (1990) Development of an "Overmethylation" strategy for Corrin synthesis. Multi-enzyme preparation of pyrrocorphins. *J. Chem. Soc. Chem. Commun.* 593-597.

Scott, A. I., Stolowich, N. J., Wang, J., Gawatz, O., Fridrich, E. and Muller, G. (1996) Biosynthesis of vitamin $B_{12}$: factor IV, a new intermediate in the anaerobic pathway. *Proc. Natl. Acad. Sci.* U S A. **93**: 14316-14319.

Scrutton, N. S., Berry, A. and Perham, R. P. (1990) Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. *Nature.* **343**: 38-43.

Senior, N., PhD thesis, University College London, 1996.

Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Reeve, J. N., *et al.* (1997) Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. *J Bacteriol.* **179**: 7135-7155.

Smith, K. M. (1991) The structure and biosynthesis of bacteriochlorophylls. In P. M. Jordan (Ed.) Biosynthesis of tetrapyrroles: 237-255.

Spencer, J. B., Stolowich, N. J., Roessner, C. A. and Scott, A. I. (1993) The *Escherichia cysG* gene encodes the multifunctional protein, sirohaem synthase. *FEBS Lett.* **335**: 57-60.

Spencer, J. B., Stolowich, N. J., Roessner, C. A., Min, C. H., Scott, A. I. (1993) Biosynthesis of vitamin-$B_{12}$ - ring contraction is preceded by incorporation of molecular-oxygen into precorrin-3. *J. American Chem. Soc.* **115**: 11610-11611.

Stojiljkovic, I. and Hantke, K. (1992) Hemin uptake system of *Yersinia enterocolitica*: similarities with other TonB-dependent systems in Gram-negative bacteria. *EMBO J.* **11**: 4359-4367.

Thibaut, D., Couder, M., Crouzet, J., Debussche, L., Cameron, B., Blanche, F. (1990) Assay and purification of S-adenosyl-L-methionine: precorrin-2 methyltransferase from *Pseudomonas denitrificans. J Bacteriol.* **172**: 6245-6251.

Thibaut, D., Debussche, L. and Blanche, F. (1990) Biosynthesis of vitamin- $B_{12}$ - isolation of precorrin-6x, a metal-free precursor of the corrin macrocycle retaining 5 S-adenosylmethionine-derived peripheral methyl-groups. *Proc. Natl. Acad. Sci.* USA. **87**: 8795-8799.

Thibaut, D., Blanche, F., Debussche, L., Leeper, F. J.and Battersby, A. R. (1990) Biosynthesis of vitamin- $B_{12}$ - structure of precorrin-6x octamethylester. *Proc. Natl. Acad. Sci.* USA. **87**: 8800-8804.

Thibaut, D., Kiuchi, F., Debussche, L., Blanche, F., Kodera, M., Leeper, F. J.and Battersby, A. R. (1992) Biosynthesis of vitamin- $B_{12}$ - structural studies on precorrin-8x, an octamethylated intermediate and the structure of its stable tautomer. *J. Chem. Soc.-Chem. Com.* **14**: 982-985.

Thibaut, D., Couder, M., Famechon, A., Debussche, L., Cameron, B., Crouzet, J., Blanche, F. (1992) The final step in the biosynthesis of hydrogenobyrinic acid is catalyzed by the *cobH* gene-product with precorrin-8x as the substrate. *J. Bacteriol.* **174**: 1043-1049.

Thibaut, D., Debussche, L., Frechet, D., Herman, F., Vuilhorgne, M., Blanche, F. (1993) Biosynthesis of vitamin $B_{12}$ - the structure of factor-IV, the oxidized form of precorrin-4. *J. Chem. Soc.-Chem. Com.* **6**: 513-515.

Thibaut, D. (1996) Vitamin $B_{12}$ biosynthetic pathway in *Pseudomonas denitrificans*. In the 4$^{th}$ European Symposium on vitamin $B_{12}$ and $B_{12}$-proteins. Poster P48.

Vlcek, C., Paces, V., Maltsev, N., Paces, J., Haselkorn, R. and Fonstein, M. (1997) Sequence of a 189-kb segment of the chromosome of *Rhodobacter capsulatus* SB1003. *Proc Natl Acad Sci U S A.* **94**: 9384-9388.

Walker, C. J. and Willows, R. D. (1997) Mechanism and regulation of Mg-chelatase. *Biochem. J.* **327**: 321-333.

Wang, J., Stolowich, N. J., Santander, P. J., Park, J. H. and 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. Natl. Acad. Sci. USA*. **93**: 14320-14322.

Warren, M. J., Scott, A. I. (1990) Tetrapyrrole assembly and modification into the ligands of biologically functional cofactors. *Trends Biochem Sci.* **15**: 486-491.

Warren, M. J., Stolowich, N. J., Santander, P. J., Roessner, C. A., Sowa, B. A. and Scott, A. I (1990) Enzymatic synthesis of dihydrosirohydrochlorin (precorrin-2) and a novel pyrrocorphin by uroporphyrinogen III methylase. *FEBS Letts.* **261**: 76-80.

Warren, M. J., Roessner, C. A., Santander, P. J., and Scott, A. I (1990) the *Escherichia coli cysG* gene encodes S-adenosylmethionine-dependent uroporphyrinogen III methylase. *Biochem. J.* **265**: 725-729.

Warren, M. J., Bolt, E. L., Roessner, C. A., Scott, A. I., Spencer, J. B., and Woodcock, S. C. (1994) Gene dissection demonstrates that the *Escherichia coli cysG* gene encodes a multifunctional protein. *Biochem. J.* **302**: 837-844.

Wolf, J. B. and Brey, R. N. (1986) Isolation and characterizations of *Bacillus megaterium* cobalamin biosynthesis-deficient mutants. *J. bacteriol.* **166**: 51-58.

Woodcock, S. C. and Warren, M. J. (1996) Evidence for a covalent intermediate in the *S*-adenosyl-L-methionine-dependent transmethylation reaction catalysed by sirohaem synthase. *Biochem. J.* **313**: 415-421.

Woodcock, S. C., Raux, E., Levillayer, F., Thermes, C., Rambach, A., and Warren, M. J. (1998) Effect of mutations in the transmethylase and dehydrogenase/chelatase domains of sirohaem synthase (CysG) on sirohaem and cobalamin biosynthesis. *Biochem. J.* **330**: 121-129.

Wu, W. F., Urbanowski, M. L. and Stauffer, G. V. (1992) Role of the MetR regulatory system in vitamin $B_{12}$-mediated repression of the *Salmonella typhimurium metE* gene. *J. Bacteriol.* **174**: 4833-4837.

Zenno, S., Saigo, K., Kanoh, H.and Inouye, S. (1994) Identification of the gene encoding the major NAD(P)H-flavin oxidoreductase of the bioluminescent bacterium *Vibrio-fischeri* ATCC-7744. *J. Bacteriol.* **176**: 3536-3543.