THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 278, No. 42, Issue of October 17, pp. 41148–41159, 2003

Printed in U.S.A.

Comparative Genomics of the Vitamin B_{12} Metabolism and Regulation in Prokaryotes*[S]

Received for publication, June 3, 2003
Published, JBC Papers in Press, July 17, 2003, DOI 10.1074/jbc.M305837200

Dmitry A. Rodionov‡§¶, Alexey G. Vitreschak‡‖, Andrey A. Mironov‡§, and Mikhail S. Gelfand‡§

From the ‡State Scientific Center GosNIIGenetika, Moscow 113545, Russia, the ‖Institute for Problems of Information Transmission, Moscow 101447, Russia, and §Integrated Genomics-Moscow, P.O. Box 348, Moscow 117333, Russia

Using comparative analysis of genes, operons, and regulatory elements, we describe the cobalamin (vitamin B_{12}) biosynthetic pathway in available prokaryotic genomes. Here we found a highly conserved RNA secondary structure, the regulatory *B12* element, which is widely distributed in the upstream regions of cobalamin biosynthetic/transport genes in eubacteria. In addition, the binding signal (*CBL*-box) for a hypothetical B_{12} regulator was identified in some archaea. A search for *B12* elements and *CBL*-boxes and positional analysis identified a large number of new candidate B_{12}-regulated genes in various prokaryotes. Among newly assigned functions associated with the cobalamin biosynthesis, there are several new types of cobalt transporters, ChlI and ChlD subunits of the CobN-dependent cobaltochelatase complex, cobalt reductase BluB, adenosyltransferase PduO, several new proteins linked to the lower ligand assembly pathway, L-threonine kinase PduX, and a large number of other hypothetical proteins. Most missing genes detected within the cobalamin biosynthetic pathways of various bacteria were identified as nonorthologous substitutes. The variable parts of the cobalamin metabolism appear to be the cobalt transport and insertion, the CobG/CbiG- and CobF/CbiD-catalyzed reactions, and the lower ligand synthesis pathway. The most interesting result of analysis of *B12* elements is that B_{12}-independent isozymes of the methionine synthase and ribonucleotide reductase are regulated by *B12* elements in bacteria that have both B_{12}-dependent and B_{12}-independent isozymes. Moreover, B_{12} regulons of various bacteria are thought to include enzymes from known B_{12}-dependent or alternative pathways.

Cobalamin (CBL),[1] along with chlorophyll, heme, siroheme, and coenzyme F_{430}, constitute a class of the most structurally complex cofactors synthesized by bacteria. The distinctive feature of these cofactors is their tetrapyrrole-derived framework with a centrally chelated metal ion (cobalt, magnesium, iron, or nickel). Methylcobalamin and Ado-CBL, two derivatives of vitamin B_{12} (cyanocobalamin) with different upper axial ligands, are essential cofactors for several important enzymes that catalyze a variety of transmethylation and rearrangement reactions. Among the most prominent vitamin B_{12}-dependent enzymes in bacteria and archaea are the methionine synthase isozyme MetH from enteric bacteria; the ribonucleotide reductase isozyme NrdJ from deeply rooted eubacteria and archaea; diol dehydratases and ethanolamine ammonia lyase from enteric bacteria involved in anaerobic glycerol, 1,2-propanediol, and ethanolamine fermentation; glutamate and methylmalonyl-CoA mutases from clostridia and streptomycetes; and various CBL-dependent methyltransferases from methane-producing archaea (1–5).

Most prokaryotic organisms as well as animals (including humans) and protists have enzymes that require CBL as cofactor, whereas plants and fungi are thought not to use it. Among the CBL-utilizing organisms, only some bacterial and archaeal species are able to synthesize CBL *de novo* (6). To our knowledge, there are two distinct routes of the CBL biosynthesis in bacteria (Fig. 1): the well studied oxygen-dependent (aerobic) pathway studied in *Pseudomonas denitrificans* and the oxygen-independent (anaerobic) pathway that was partially resolved in *Salmonella typhimurium*, *Bacillus megaterium* and *Propionibacterium shermanii* (7).

The biosynthesis of Ado-CBL from Uro′III, the last common precursor of various tetrapyrrolic cofactors, requires about 25 enzymes (6) and can be divided into two major parts. The first part, the corrin ring synthesis, is different in the anaerobic and aerobic pathways; the former starts with the insertion of cobalt into precorrin-2, whereas in the latter, this chelation reaction occurs only after the corrin ring synthesis. The second part of the Ado-CBL pathway is common for both anaerobic and aerobic routes and involves adenosylation of CR, attachment of the aminopropanol arm, and assembly of the nucleotide loop that bridges the lower ligand dimethylbenzimidazole and CR (4). The corresponding CBL genes from *S. typhimurium* and *P. denitrificans* have different traditional names, mainly using prefixes *cbi* and *cob*, respectively (Fig. 1). For example, *S. typhimurium* has two separate genes, *cbiE* and *cbiT*, that encode precorrin methyltransferase and decarboxylase, respectively, whereas in *P. denitrificans* these functions are encoded by one gene, *cobL*. For consistency, we use the *S. typhimurium* names whenever possible. In particular, we assign gene names to experimentally uncharacterized genes using analysis of orthology.

The anaerobic and aerobic pathways contain several pathway-specific enzymes. First, the cobalt insertion is performed by the ATP-dependent aerobic cobalt chelatase of *P. denitrificans*, which consists of CobN, CobS, and CobT subunits, and two distinct, ATP-independent, single subunit cobalt chelatases, CbiK from *S. typhimurium* and CbiX from *B. megaterium*, which are associated with the anaerobic pathway (8–10). Sec-

* This work was supported in part by Howard Hughes Medical Institute Grant 55000309 and Ludwig Institute for Cancer Research Grant CRDF RBO-1268. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.jbc.org) contains an additional table.

¶ To whom correspondence should be addressed. Fax: 7-095-315-05-01; E-mail: rodionov@genetika.ru.

[1] The abbreviations used are: CBL, cobalamin; Ado-CBL, adenosylcobalamin; Uro′III, uroporphyrinogen III; TMSs, transmembrane segments; CoA, coenzyme A; CR, corrin ring.

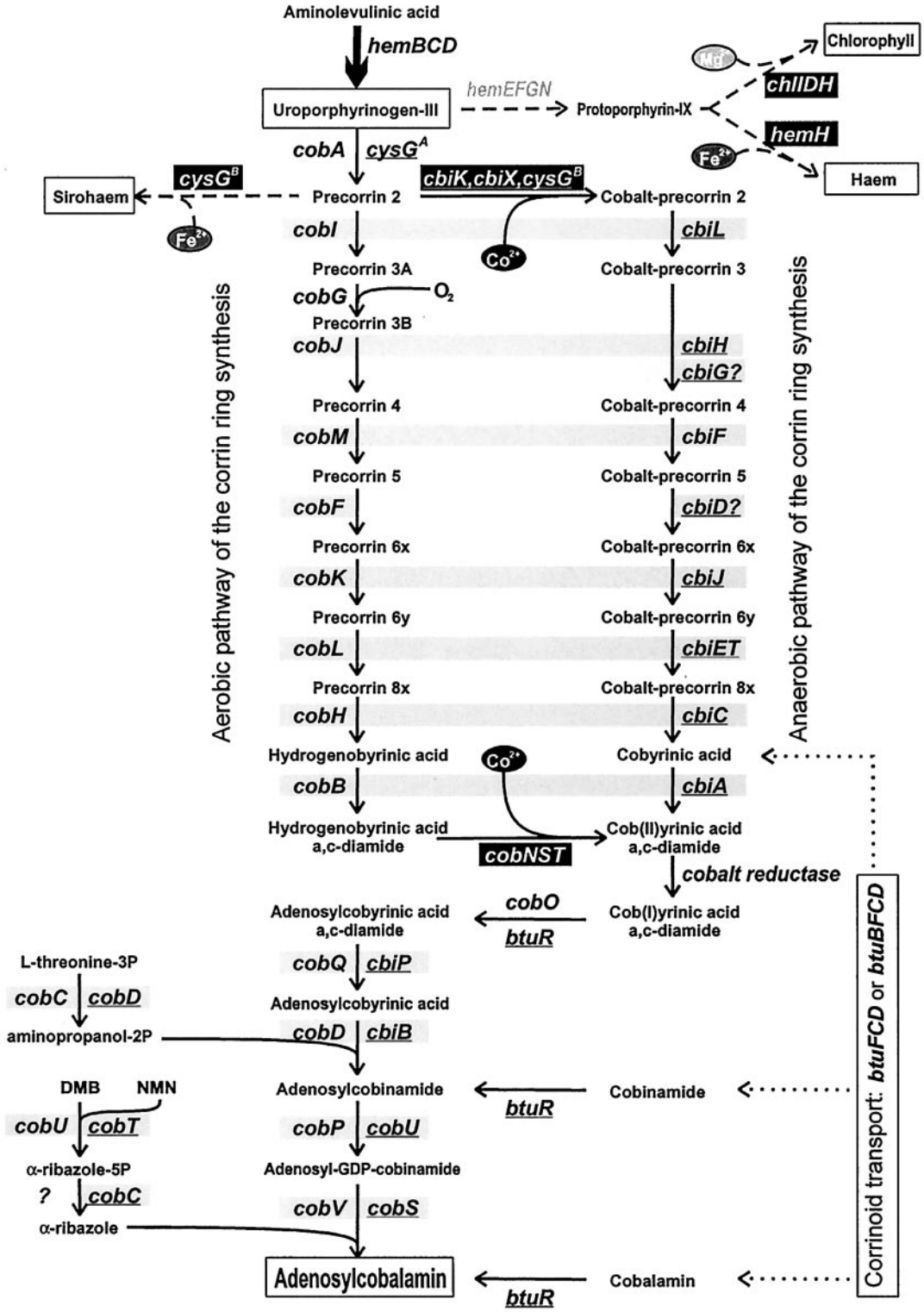

FIG. 1. **Biosynthetic pathways for adenosylcobalamin and other tetrapyrrolic cofactors in bacteria.** The anaerobic and aerobic Ado-CBL pathways are characterized by the early and late cobalt insertions, respectively. In bacteria with the anaerobic pathway, cobalt is inserted into the macrocycle using either the CbiK (as in *S. typhimurium*) or CbiX chelatases. (as in *B. megaterium*). *S. typhimurium* gene names are *underlined* and used throughout this work. Similar genes of *S. typhimurium* and *P. denitrificans* are arranged within same *gray block* (see the introduction for explanation). Various chelatases are in *black blocks*. The vitamin B_{12} and cobalt transport routes are shown by *lines with arrows*.

ond, since the majority of the intermediates of the anaerobic, but not aerobic, pathway have the cobalt ion inserted into the macrocycle, the pathways could use enzymes with different substrate specificities. CobG from *P. denitrificans* requires molecular oxygen to oxidize precorrin 3A and is specific for the aerobic pathway (11). The respective CR oxidation of anaerobic route is probably mediated via the complexed cobalt ion, which can assume different valence states. In summary, CbiD, CbiG, and CbiK are specific to the anaerobic route of *S. typhimurium*, whereas CobE, CobF, CobG, CobN, CobS, CobT, and CobW are unique to the aerobic pathway of *P. denitrificans.*

In most bacteria, cobalt and other heavy metal ions are

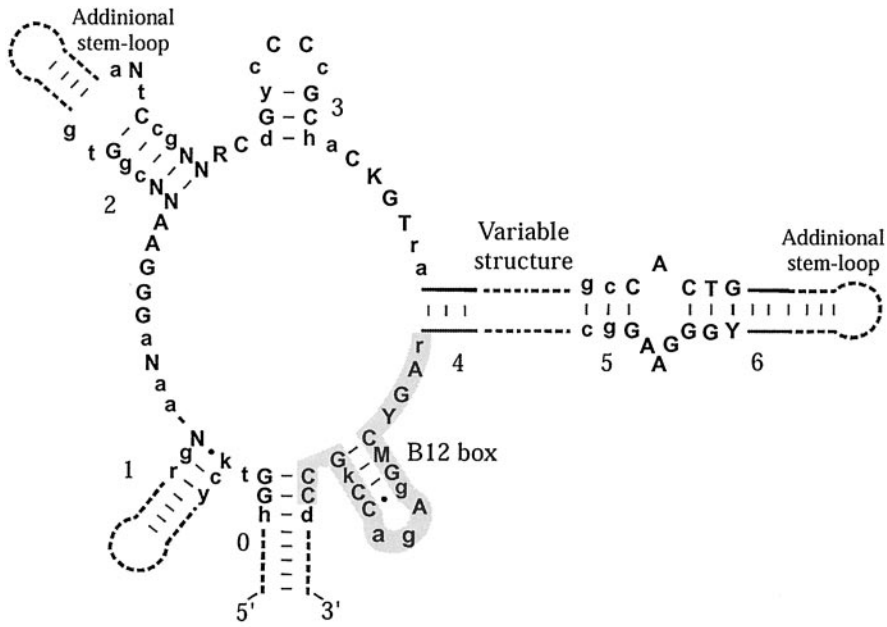

FIG. 2. **The conserved structure of the *B12* element.** *Capital letters* indicate invariant positions. *Lowercase letters* indicate strongly conserved positions. Degenerate positions are as follows: *R*, A or G; *Y*, C or U; *K*, G or U; *M*, A or C; *H*, not G; *D*, not C; *N*, any nucleotide.

mainly accumulated by the fast and unspecific CorA transport system (12). An additional cobalt transporter, a part of the cobalt-dependent nitrile hydratase gene cluster, was identified in *Rhodococcus rhodochrous* and, together with some nickel-specific transporters, belongs to the HoxN family of chemiosmotic transporters (13). Further, the ATP-dependent transport system CbiMNQO, encoded by the CBL biosynthetic operon in *S. typhimurium*, probably mediates high affinity transport of cobalt ions for the B_{12} synthesis (14). Vitamin B_{12}, cobinamide, and other corrinoids are actively transported in enteric bacteria using the TonB-dependent outer membrane receptor BtuB in the complex with the ABC transport system BtuFCD (15).

Vitamin B_{12} is known to repress expression of the *btuB* genes of *Escherichia coli* and *S. typhimurium* (16) and the *cob* operon in *S. typhimurium* (17). No B_{12}-regulatory genes were identified in bacteria, but it was shown that Ado-CBL is an effector molecule involved in the regulation of CBL genes in enterobacteria (18). The evolutionarily conserved B12-box, a *cis*-acting translational enhancer element, contains a stem-loop structure that would mask the ribosome binding site as well as several additional RNA structural elements. This element is found in the 5′-untranslated regions of the CBL operons and is absolutely required for their regulation, which is conferred mainly at the translational level (17). Recently, it was shown that the *btuB* mRNA leader sequence can directly bind an effector molecule, Ado-CBL, and consequently undergo conformational changes in the secondary and tertiary structure of the RNA and that the likely mechanism of regulation involves formation of two alternative RNA structures (18).

Combination of the comparative analysis of gene regulation, positional clustering of genes, and phylogenetic profiling, when applied to a metabolic pathway in a variety of bacterial species, is a powerful approach to the search of missing genes within the pathway as well as identification of specific metabolite transport genes (19–21). Here we use this combined comparative approach for the analysis of the CBL biosynthetic pathway in prokaryotes. The expression of genes involved in the CBL biosynthesis and vitamin B_{12} transport in eubacteria was predicted to be regulated mainly by a conserved RNA regulatory element, the *B12* element. In four archaeal genomes, a new DNA-type regulatory signal was observed upstream of the CBL-related genes. After reconstruction of the B_{12} regulon and the CBL pathway in most bacterial and archaeal genomes, we identified several new enzymes and transporters related to the CBL biosynthesis. In particular, numerous new cobalt transporters and chelatases, as well as new CR methyltransferases, were found. Furthermore, the vitamin B_{12} transporters are widely distributed in bacteria and archaea and mostly B_{12}-regulated. Finally, the *B12* element was predicted to regulate B_{12}-independent methionine synthase and ribonucleotide reductase isozymes in bacteria that also have corresponding B_{12}-dependent isozymes.

EXPERIMENTAL PROCEDURES

Complete and partial sequences of bacterial genomes were downloaded from GenBank™ (22). Preliminary sequence data were also obtained from the World Wide Web sites of the Institute for Genomic Research (www.tigr.org), the University of Oklahoma's Advanced Center for Genome Technology (www.genome.ou.edu/), the Wellcome Trust Sanger Institute (www.sanger.ac.uk/), the DOE Joint Genome Institute (jgi.doe.gov), and the ERGO Data base (ergo.integratedgenomics.com/ERGO) (23). Gene identifiers from the ERGO data base and GenBank™ are used throughout. The amino acid sequences of uncharacterized genes predicted here to be involved in the CBL metabolism have been collected in one FASTA file that is available upon request.

The RNA-PATTERN program (24) was used to search for conserved RNA regulatory elements. The input RNA pattern included both the RNA secondary structure and the sequence consensus motifs. The RNA secondary structure was described as a set of the following parameters: the number of helices, the length of each helix, the loop lengths, and description of the topology of helix pairs. The initial RNA pattern of the *B12* element was constructed using a training set of upstream regions of the *btuB* orthologs from proteobacteria. Each genome was scanned with the *B12* element pattern, resulting in detection of approximately 200 *B12* elements.

A protein similarity search was done using the Smith-Waterman algorithm implemented in the *GenomeExplorer* program (25). Multiple sequence alignments were constructed using ClustalX (26). Orthologous proteins were initially defined by the best bidirectional hits criterion (27) and, if necessary, confirmed by construction of phylogenetic trees. Note that the fact of gene absence used in phylogenetic profiling is reliable only for complete genomes. The phylogenetic trees were created by the maximum likelihood method implemented in PHYLIP (28) and drawn using the GeneMaster program.[2] Distant homologs were identified using PSI-BLAST (29). Transmembrane segments (TMSs) were predicted using the TMpred program (www.ch.embnet.org/software/TMPRED_form.html).

[2] A. Mironov, unpublished results.

RESULTS

Conserved Structure of the B12 Element—Previously, we have described two highly conserved RNA elements, *RFN* and *THI*, involved in the regulation of the riboflavin and thiamin biosynthetic genes in bacteria (20, 21). Vitamin B_{12}-dependent regulation of the *btuB* and *cbiA* genes in enterobacteria requires their upstream regions and occurs via a post-transcriptional mechanism involving formation of alternative RNA structures. Several recent studies describe possible secondary structures of the *E. coli btuB* and *S. typhimurium cbiA* 5′-untranslated leader sequences, but the proposed structures have a limited number of conserved elements (17, 18). Using the comparative analysis of nearly 200 regulatory regions of vitamin B_{12}-related genes in bacteria, we derived a highly conserved RNA structure named here the *B12* element (30). Similarly to the *RFN* and *THI* elements, the *B12* element has a set of unique stem-loops closed by a single base stem and highly conserved sequence regions, including the previously known B12-box (Fig. 2). In addition to seven conserved stem-loops, the *B12* element has three additional facultative stem-loops and one internal variable structure. Since direct binding of Ado-CBL to the *btuB* mRNA leader was recently shown (18), it is interesting that all internal loops of the *B12* element are highly conserved on the sequence level and, therefore, may be involved in Ado-CBL binding. By analogy to the model of regulation for riboflavin and thiamin regulons (20, 21), a model of regulation of B_{12}-related genes based on formation of alternative RNA structures involving the *B12* elements is suggested (30).

B_{12} Regulon: Identification of Genes and Regulatory Elements—Initially, orthologs of the cobalamin biosynthetic and transport genes ("CBL genes" below) in all available prokaryotic genomes were identified by similarity search (Table I). For further analysis, positional clusters (including possible operons) of the CBL genes are also described in Table I. The multifunctional gene *cysG* of *E. coli*, which encodes URO′III methyltransferase ($CysG^A$) and precorrin-2 oxidase/ferrochelatase ($CysG^B$) activities and is partially shared by the CBL and siroheme biosynthesis, was considered only if it was co-localized with other CBL genes.

Then we scanned nearly 100 genomic sequences using the RNA-PATTERN program and the pattern of a novel, B_{12}-specific RNA element (30) and found approximately 200 *B12* elements unevenly distributed in 66 eubacterial genomes (Table I). All genomes with *B12* elements, except *Bacillus cereus*, contain CBL biosynthesis and/or transport genes. Most obligate pathogenic bacteria (see below) as well as *Aquifex aeolicus* have neither CBL genes nor *B12* elements. *Staphylococcus aureus*, *Corynebacterium glutamicum*, *Bordetella pertussis*, *Magnetococcus*, and all archaeal genomes lack *B12* elements but have CBL genes. The detailed phylogenetic and positional analysis of the CBL genes and the *B12* elements is given below.

In attempt to analyze potential cobalamin regulons in archaea, a large phylogenetic group without *B12* elements, we collected upstream regions of all CBL genes and applied the signal detection procedure to each archaeal genome (31). The same strongest signal, a 15-bp palindrome with consensus 5′-TGGATAantTATCCA-3′, was observed in candidate cobalamin regulons in three *Pyrococcus* genomes (Table I). To find new members of the regulon, the derived profile (named *CBL*-box) was used to scan the genomes. The cobalamin regulon in the pyrococci appears to include all CBL biosynthesis and transport genes except *btuR*. In addition, conserved *CBL*-boxes were identified upstream of the *P. horikoshii* genes *PH0021*, *PH1306*, *PH0275*, *PH1928*, and *PH0272* and their orthologs in two other pyrococci. These genes are predicted to encode anaerobic ribonucleotide reductase NrdDG, two subunits of methylmalonyl-CoA mutase MutB, succinyl-CoA synthase SucS, and methylmalonyl-CoA epimerase MmcE, respectively. All of these genes are unrelated to the CBL biosynthesis or transport, but their co-regulation with the CBL genes seems to be rational because of their direct or indirect association with B_{12}-dependent enzymes (see below). The same CBL-specific profiles were obtained for two other archaea, *Aeropyrum pernix* and *Sulfolobus solfataricus*, but not for the remaining archaeal species. The predicted CBL regulon of *A. pernix* again contains the B_{12} transport system and methylmalonyl-CoA mutase. Among all archaea in this study, only pyrococci and *A. pernix* are likely to be unable to synthesize CBL *de novo* but may uptake and transform CBL precursors to Ado-CBL. The CBL regulon of *S. solfataricus* includes, in addition to the *cobT* and *btuFCD* genes, the *cbiGECHDTLF* genes for the *de novo* CBL synthesis and predicted cobalt transporter *hoxN*.

To select bacterial species that potentially require coenzyme B_{12} for their metabolism, we carried out a similarity search for all known B_{12}-dependent enzymes in prokaryotic genomes. As a result, *Chlamydia* spp., *Rickettsiae* spp., *Neisseria* spp., *Streptococcaceae*, *Mycoplasmataceae, Pasteurellaceae*, ϵ-proteobacteria, *Borellia burgdorferi*, *Treponema pallidum*, and *Xylella fastidiosa* (obligate pathogenic bacteria) as well as *A. aeolicus* were found to have no B_{12}-dependent enzymes (Supplementary Table VI). This finding is in agreement with the absence of the CBL biosynthetic and transport genes as well as with the absence of *B12* elements in these microorganisms. However, two other bacteria without any known B_{12}-dependent enzyme, *Bacillus subtilis* and *S. aureus*, were predicted to have the B_{12} transport system BtuFCD. Interestingly, *btuFCD-pduO* is the only *B12* element-regulated operon in *B. subtilis*. This shows that other, currently unknown, B_{12}-dependent enzymes may be present in these bacteria.

Vitamin B_{12} Transporters—Nearly one-fourth of the B_{12}-utilizing bacteria appear to have no complete pathway for the CBL biosynthesis and, therefore, should actively transport vitamin B_{12} or some precursor from the external medium. The only known transport system for vitamin B_{12} is the ABC transporter BtuFCD of enteric bacteria, which consists of periplasmic substrate-binding protein BtuF, two transmembrane subunits BtuC, and two peripheral ATP-binding subunits BtuD. In Gram-negative bacteria, the translocation of vitamin B_{12} across the outer membrane involves B_{12}-specific receptor BtuB and the periplasmic energy-coupling proteins TonB, ExbB, and ExbD, which are shared between various TonB-dependent receptors. Thus, the B_{12}-specific components of the transporters are BtuBFCD and BtuFCD in Gram-negative and Gram-positive bacteria, respectively. The corresponding components of ABC transporters involved in the uptake of ferric siderophores, heme, and vitamin B_{12} are similar and belong to the same families (32). Therefore, a similarity search is not sufficient to dissect the B_{12} and ferric transporters in species distant from enteric bacteria.

We combined a similarity search with identification of highly specific regulatory *B12* elements and with positional analysis of genes. The phylogenetic trees for the protein families formed by various components of the B_{12} and ferric transporters revealed B_{12}-specific subfamilies within each family (data not shown). The predicted transporters for vitamin B_{12} were found to be widely distributed in prokaryotes; among B_{12}-utilizing bacteria with complete genomes, they were not found only in four cyanobacterial and three archaeal species, in *Mycobacterium* spp., and in *Bacillus cereus* (Supplementary Table VI). In most cases, components of B_{12} transporters are encoded by clusters of co-localized genes that are regulated by the *B12*

TABLE 1

Cobalamin biosynthesis and transport genes and B12-elements in bacteria

The standard *S. typhimurium* names of genes, which are common for the aerobic and anaerobic CBL biosynthetic pathways, are used throughout (see Fig. 1 for the *P. denitrificans* equivalents). Genes of the first parts of the pathway involved in the corrin ring synthesis are shown in magenta; other CBL-biosynthetic genes are in green. Genes encoding transport proteins and chelatases subunits are shown in blue and orange, respectively. Parentheses denote gene fusions. Genes forming one candidate operon (with spacer less than 100 bp) are separated by dashes. Larger spacers between genes are marked by -//-. The direction of transcription in divergons is shown by angle brackets. For example, *cobD<>cbiB* denotes divergently transcribed genes, whereas *cbiB>><cbiP indicates convergently transcribed genes. Ampersands denote B12*-elements. Predicted *CBL*-boxes in archaea are denoted by dollar signs. Operons from different loci, if shown in one column, are separated by semicolons. Genes not related to the CBL biosynthesis are shown as *X.* The contig ends are marked by square brackets. The number of *B12*-elements per genome is given in the fourth column. The names of taxonomic groups in the first column, α, β, γ, ε, δ, Cya, CFB, T/D, B/C, Act, SP, and A, stand for α-, β-, γ-, ε-, and δ-proteobacteria, cyanobacteria, the CFB group, the *Thermus/Deinococcus* group, the *Bacillus/Clostridium* group, actinomycetes, spirochetes, and archaea, respectively. The genome abbreviations are given in the third column with unfinished genomes marked by #. Additional genome abbreviations are as follows: RP, *Rickettsia prowazekii;* RCO, *R. conorii;* NM, *Neisseria meningitidis;* NG, *N. gonorrhoeae;* YP, *Yersinia pestis;* EO, *Erwinia carotovora;* HI, *Hemophilus influenzae;* VK, *Pasteurella multocida;* AB, *Actinobacillus actinomycetemcomitans;* HP, *Helicobacter pylori;* CJ, *Campylobacter jejuni;* LL, *Lactococcus lactis;* PMA, *Prochlorococcus marinus;* SN, *Synechococcus sp.;* CY, *Synechocystis sp.;* BB, *B. burgdorferi;* TP, *T. pallidum;* PH, *P. horikoshii;* PF, *P. furiosus;* PO, *P. abyssii.*

α	*Mesorhizobium loti*	MLO	6	*ardX2<>&cbtAB-cobU-btuR><cbiB<>cobD-X; &cfrX-cobW-cobN-cobG-cbiCLH><cbiJ<>cbi(ET)-cobE-cbiF-cobA-cbiA-cobS-cobT1; cobF; X-cbiP; cobT2<>cblXY; &cfrX-btuDFC; &ardX-frdX; &metE; &bluB*
	Bradyrhizobium japonicum	BJA	5	*bluB-cobSU<>cbiB-cobD><cbiP-btuR<>&cfrX-cobW-cobN-cobG-cbiCLH><cbiJ<>cbi(ET)-cobE-cbiF-cbiA-cobA-&cbtC-cobF-cobT><cobC<>&btuB3-btuN; &btuB; &metE*
	Sinorhizobium meliloti, Pseudomonas denitrificans #	SM PD	5	*cbiP; &cobU-cobW-cobN-btuR-//-cobE-cobA-cbiA-cobD-cbiB; cobF-cobG-cbiCLH><cbiJ<>cbi(ET)-cbiF; cobT<>cobS-cblXY; &bluB; &btuFCD; &ardX-frdX; &cbtC*
	Brucella melitensis	BME	4	*cbiP-&cbtAB-cobU-cobW-cobN-btuR-//-cobE-cbiF><cbiJ-cbiD<>cobA-cbiA-cobD-cbiB; cbi(ET)<>cobG-cbiCL-cbi(G*H); cobT<>cobS-cblXY; &btuFCD2; &btuBFCD; &nrdHIAB; bluB*
	Agrobacterium tumefaciens	AU	6	*cobD<>cbiB><cbiP-btuR<>&cfrX-cobW-cobN-cobG-cbiCLH><cbiJ<>cbi(ET)-cobE-cbiF><cbiD<>cobA-cbiA; cobU; cobT<>cobS-cblXY; &btuFCD; &nrdHIAB; &cbtAB; &bluB; metE-X<>&achX-metR*
	Rhodopseudomonas palustris	RPA	9	*bluB-cobT1<>&cfrX-cobU-cobW-cobN-btuR-cbiP1><cbiB-btuF<>cobD><btuDC-hoxN&; &metE-~btuB-btuS-btuT; cobF-cobP2-cobS-cobT2&<>ORF663-&cbiCLH><cbiJ<>cbi(ET)-cobE-cbiF-cbiA-cobA><btuF2&; &btuB; &btuB3-btuN; cobC-&crdX*
	Rhodobacter capsulatus #	RC	13	*cobSTU<>cobC-bluE-cbiB-cobD-bluB><cbiP-btuR<>&cfrX-cobW-cobN-cbiCLH><cbiJ<>cbi(ET)-cobE-cbiF-cobA-cobF-&cbiMNQO-ORF663-cbiA; &btuFCD; &btuD; frdX-ardX&<metH<cbiP2-cbiP3&<>&cbtF&btuBFCD-cobX; &exbBD-tonB; &crdX; RC04759&<>&nrdDG; &cnoABCD*
	Rhodobacter sphaeroides #	RS	6	*[cobT; &X-bluE-cobD; &bluB; cbiB; cbiP; btuR-X<>&cfrX-cobW-cobN-cbiCLH><cbiJ<>cbi(ET)-cobE-cbiF-cobA-cobF; &cbtC-cbiA; &btuBFCD-cobX; &btuFC]*
	Sphing. aromaticivorans #	SAR	3	*hoxN-&cobW-cobN-cobG-cbiCLH><cbiJ<>cbi(ET)-cobE-cbiF-cbiA-cobF-btuR-bluB-cfrX-cobT-cobC-cobS-cobD-cbiB-cbiP-cobU; &btuBFCD; &~btuB*
	Rickettsia spp. (RP, RCO)		0	*no*
	Caulobacter crescentus	CO	2	*&btuB-cbiP; X-btuFCD; cobT; &metE*
β	*Bordetella pertussis*	BP	0	*metH<>btuBF; btuB3-btuN*
	Burkholderia pseudomallei	BPS	4	*cbi(G*H)-cbiLC-cobG&<>cbi(ET)-cbiDJF; cbiA-btuR-cobE&<>&hoxN-cobW-cobN--chlID; &btuBCD-cobTSC><btuF-cobD-cbiB-cobU<>cbiP; bluB; pduX*
	Neisseria spp. (NM, NG)	NM	0	*no*
	Nitrosomonas europaea	NE	1	*&btuB-btuM-btuR><btuT-btuS; ~btuB-btuS-btuT*
	Methylobacillus flagellatus #	MFL	3	*&btuB-btuM-btuR--cbiA-btuS-btuT-bluB-btuF-cobU; cbiP<>cbiB-cobD><cobC-cobST; &btuB3-btuN; &nrdAB*
	Ralstonia eutropha #	REU	1	*&btuBCD-btuR-cobTS-cobC><btuF-cobD-cbiB-cobU<>cbiP; cbiA; bluB*
	Ralstonia solanacearum	RSO	2	*&btuBCD-cobTS-cobC><btuF-cobD-cbiB-cobU<>cbiP; &hoxN-cbiW-cobN-chl(ID)--cbiE-(cbiX-cbiC)-cbiDLFG-cbi(HJ)-cobW-btuR-cbiA-bluB*
γ	*Escherichia coli*	EC	1	*&btuB; btuCDE; X-btuF-X; X-btuR; cobC; cobUST*
	Salmonella typhimurium	SY	2	*&btuB; btuCDE; X-btuF-X; X-btuR; cobC<>cobD; pdu cluster-&cbiABCDETFGHJKLMNQO-cbiP-cobUST*
	Klebsiella pneumoniae #	KP	2	*&btuB; btuC-//-btuED; X-btuF-X; X-btuR; cobC; cobD; pdu cluster-&cbiABCDETFGHJKLMNQO-cbiP-cobU-cobT2; cobS-cobT1*
	Yersinia enterocolitica	YE	2	*&btuB; btuCED; X-btuF-X; X-btuR; pduX-cobD-cobA<-pdu cluster-&cbiABCDETFGHJKLMNQO-cbiP-cobUS-cobC-cobT; cobA2-pduX2*
	Y. pestis; E. carotovora (YP, EO)		1	*&btuB; btuCDE; X-btuF-X; X-btuR*
	Vibrio cholerae	VC	1	*&btuB; btuCD; X-cbiB-btuF-X; btuR; cbiP-X; cobTSU-cobC*
	Pasteurellaceae (HI,VK,AB)		0	*no*
	Pseudomonas aeruginosa	PA	5	*&btuB-btuR-cbiA-bluB-cbiB-cobD-cbiP-cobUTCS;chlD-//-cobN-cobW&<>&cbtAB-cobE-cbiF; cbi(G*H)-cbiLC-cobG&<>cbi(ET)-cbiDJ; &bruB2-btuDFC; ~btuB-btuS-btuT; btuB3-btuN*
	Pseudomonas putida	PP	5	*&btuR-cbiA-bluB-cbiB-cobD-cbiP-cobUTCS; &cobW-cobN-chlID; &cbtAB-cobE-cbiF; cbi(G*H)-cbiLC-cobG<>cbi(ET)-cbiDJ; cobF; &btuB2-X-&btuFCD; btuF><btuB*
	Pseudomonas fluorescens #	PU	3	*&btuR-cbiA-bluB-cbiB-cobD-cbiP-cobUTCS; chlDI-cobN-cobW&<>&cbtAB-cobE-cbiF; cbi(G*H)-cbiLC-cobG<>cbi(ET)-cbiDJ; cobC2-cobF; btuB2-btuDFC; btuB3; btuF><btuB]*
	Pseudomonas syringae #	PY	4	*&btuR-cbiA-bluB-cbiB-cobD-cbiP-cobUTCS; chlDI-cobN-cobW&<>&cbtAB -cobE-cbiF; cbi(G*H)-cbiLC-cobG<>cbi(ET)-cbiDJ; cobF; &btuFCD; bruB2-btuDFC*
	Shewanella oneidensis	SON	2	*&btuB-X-metR<>metE; metH<>cobC-&btuDC--cobTSU-cbiP-btuR; cbiB; X-btuF*
	Azotobacter vinelandii #	AV	1	*&(btuB-btuM-btuR)-cbiA-bluB-cbiB-cobD-cbiP-cobUTCS; btuB; X-btuF*
	Xanthomonas axonopodis	XAX	1	*&btuB-btuM-btuM-btuR-cbiB-cobD-cbiP-cobUT-cobC-cobS; btuB3-btuN*
	Xylella fastidiosa	XFA	0	*no*
ε	*H. pylori, C. jejuni* (HP,CJ)		0	*no*
δ	*Geobacter metallireducens #*	GME	1	*&cobUTSC-cbiA-X-cbiMNQO -cbiX-cbiCD-X-cbi(ET)-cbiLFGH-cbiPB-cobD; [btuR-mutB*
	Magnetococcus #	MCO	0	*[cbiO-cbiA-cbiF-cbiX-cbiCD-cbi(TE)-cbiLG-cobE-cbiH><cbiP<>cobD; cbiB; X-btuR; bluB; cobC; cobTS; cobU*
T/D	*Deinococcus radiodurans*	DR	3	*XX-cobTS-X; X-cobU; &btuFCD; &btuF-btuR-cbiA-cbiB-cobD-cbiP; X-cobC; &achX-nrdIBA*
	Fusobacterium nucleatum	FN	2	*cbiP1-X-cbiB-X-cobD-cbiA-X-cbiC-//-cbiDE-X-cbiT-//-cbiL-X-cbiF-//-cbiGHJ; btuR; cobUS-cobC-cobT;cbiK; cbtF; &btuFCD; &btuBFCD; btuB<>btuFCD*
	Aquifex aeolicus	AA	0	*no*
	Chlamydia spp.	CT	0	*no*
B/C	*Bacillus subtilis*	BS	1	*&btuF-btuC-(btuD-cobX)-pduO*
	Bacillus cereus	ZC	1	*pduO; &metE*
	Bacillus megaterium #	BI	1	*[&cbiW-cbi(H-?)-cbiX-cbiJCD-cbi(ET)-cbiLFGA-cobA-cbiY-btuR]*
	Bacillus halodurans	HD	5	*&btuF-btuC-(btuD-cobX)-cbiB-cobDU-cbiP-cobSC-~cobU-pduO; &cobT; cbiA; cblS; &nrdAB; &metE; &achX*
	Bacillus stearothermophilus #	BE	3	*&btuF-btuC-(btuD-cobX)-cbiB-cobDUSC-~cobU-pduO-cblT-cblS; [&cbiW-cbiMNQO-cbi(H-?)-cbiX-cbiJCD-cbi(ET)-cbiLFG-cbiAP--cobT-cobA-btuR; &nrdAB*
	Staphylococcus aureus	SA	0	*btuFC*
	Listeria monocytogenes	LMO	2	*cblS-cobUSC>&<pocR<>pduABCDEGHKJL-eutJ-pduMNOPQFW-cobD-pduX; cblT-&cbiABCDETFGHJ-(cobA-hemD)-cbiK-cbiLMNQO-cbiP-pduO; btuF*
	Clostridium acetobutylicum	CA	2	*&cbiMNQO-cobD-cbiG-pduX-cobT-cbiK-cbiP-cbiACDTLFJH-cobUS-cobC; cbiB -cbiK2-cbiE; &btuFCD- crdX*
	Clostridium perfringes	CPE	4	*cobTSUC-cbiB-cobD-&btuFCD-CPE01087-cbiP; &cbiK-cbiCDETLFG--cbiHJ-btuFCD; btuR-X; &cbiMNQO; &cblT-cblS*
	Clostridium botulinum #	CB	1	*&cbiPB-cobD-cbiMNQ--hemC-(cobA-hemD)--hemB-cbiAC; cbiD-cbiE-X--cbiT; cbiLFGHJ-cblT-cbiK-pduX-btuR; cobC; cobU; cobS--cblS; btuFCD*
	Clostridium difficile #	DF	3	*cobTSU-cobC-&cbiP-cbiA-cbiB-cobD-pduX-cbiCDETFGHJ-(cbiL-cbiK)-hemC-(cobA-hemD)-hemB-cysG^D; &cbiMNQO; &btuFCD--nrdEF; crdX*
	T. tengcongensis	THT	2	*&btuR-&btuFCD-cbiA-cbiPB-cblT-cblS-cobD-cobUS*
	Enterococcus faecalis	EF	1	*&btuFCD-pduV-pduO--pdu cluster*
	Streptococcaceae(ST,PN,MN,LL)		0	*no*
	Mycoplasmataceae(MG,MP,UU)		0	*no*
	Heliobacillus mobilis #	HMO	5	*&cbiD-cbi(ET)-cbiLFGHJ-cbiX-cbiC-//-cobU-cbiPBA-btuR-cobTS-pduX-cobD-~cobC; &cbiMJ; &cbiQO; &cblS]; cblT-cbiO; &HMO01408*
	Desulfitobacterium halfniense #	DHA	6	*cbiD-&cbi(ET)-cbiLFGHJ-cbiX-cbiC-cbiPB--cbiA-btuR-pduX-cobD]; [cobTUS]; cobC]; &btuFCD; &cnoAB]; &cblS; [cbiMN]; [cbiQO]; &DHA05379; &nrdD]*
Act	*Corynebacterium glutamicum*	CGL	0	*cblZ-cobUTS; pduO; btuCFD*
	Corynebacterium diphtheriae	DI	1	*cblZ-cobUTS; chlID-btuR-cbiA-//-cobA; cbiP; cobF; cbiB-//-cobD; cobN<>cobG-cbiC-cbi(LH)><cbiJ-cbiF-cbi(ET); &btuCDF*
	Mycobacterium tuberculosis	MT	2	*cblZ-cobTS; chl(ID)-btuR-cbiA--cysG; &cbtG-cbiP-cobU; cbiB-//-cobD; cobN<>cobG-cbiC-cbi(LH)><metZ-X<>X><cbiJ-cbiF-cbi(ET); &metE; bluB*
	Mycobacterium leprae	ML	2	*cblZ-cobTS; btuR-cbiA; &cbtG; &metE*
	Thermobifida fusca #	TFU	3	*cblZ-//-cobUS]; cbiB<>cobD-(bluB-cobT)--cysG; cobF-cobN&<>cobG-cbiC-cbi(LH)><cbiF-cbi(ET)<>&chl(ID)-btuR-cbiA><cbiJ-cbiP-cbtE&*
	Rhodococcus str. #	RK	4	*cblZ--cobT]; cobU; [cobD]; &cbtE-cbiP]; [cobF];[cobN&<>cobG-cbiC-cbi(LH)];[cbiF];[cbi(ET)<>&chl(ID)];[btuR];[cbiA];[cbiJ><cbiB]; &btuFCD; [bluB*
	Streptomyces coelicolor	SX	7	*cblZ-//-cobU--cobT-//-cobS; (bluB-cobT)--cysG; cobF; cbiBP-cobN-chl(ID)-btuR-cbiA-cbiL><SX03279><cbiF-cbi(ET)-cbi(GH)-cbiX-cobD; cbiC><cbiJ; &cbiMNQO; &btuFCD; &btuCDF; &pduX; &SX12454; &nrdAB; &metE*
	Propionicibacterium shermanii #	PI	3	*mutBA&<>&cbiLF-cbi(EGH)-(cbiX-cysG^D)><cbiDCTJ; &cbiBP-btuR]; cbiMNQO-cobA*
Cya	(PMA,CY,SN)		1	*cbiJ; cbiF; cbiC; cbiL; cbi(GH); cbi(ET); cbiD; cbiB; cbiA; cobU; cobS; btuR; cbiP; cobD; cobN; cbiX; &hupE*
	Anabaena sp.	AN	3	*&cfrX-cbiJ; cbiF; &cobG-cbiCL-X-cbi(G*H); cbi(ET); cbiD; cbiB; cbiA; cobU; cobS; btuR; cbiP; cobD; cobN; &cfrX-btuB-(X-cbiW)-btuFCD; cbiMNQO*
	T. elongatus	TEL	2	*cbiJ; cbiF; cbiC; cbiL; cbi(GH); cbi(ET); cbiD; cbiB; cbiA; cobU; cobS; btuR; cbiP; cobD; cobN; &cbiX; cbiMNQO; &metE*
	Chloroflexus aurantiacus #	CAU	2	*&btuF-cbiW-btuCD-X]; [X-cbiCD-cbi(ET)-cbiLF]; [cbiG]; [(cbiH-cysG^D)-cobA-cbiA-cobUD-cbiP; &btuR-cobN-cobT-cbiMNQO-cbiB; cobCS; chl(ID)*

TABLE I—*continued*

CFB	*Porphyromonas gingivalis*	PG	5	*cobUTSC; cbiA-pduO-cbiP-cobD-cbiB; &cbtD; &btuB4-cbiK-btuFCD1; cbi(HC)-cbi(ET)-cbi(GF)-cbi(JD); cbiL; &btuFCD2; hmuY-hmuR-btuS-X-btuT; &nrdDG; &butD-butA-4hbD-sucD*
	Bacteroides fragilis #	BX	7	*cobUTSC><cbiB-cobD-cbiP; cbiA-pduO-//-cbtD&<>&btuB4-cbiK-btuB3-btuN-btuS-X-btuT-cbi(HC)-cbi(ET)-cbi(GF)-cbi(JD); cbiL<>btuFCD &btuB; btuFC; hmuY-hmuR-btuS-X-btuT-&nrdAB; ~btuB-&metE^C; &nrdDG; &pccCAB*
	Chlorobium tepidum	CL	5	*&btuBF-btuB-cbiP-cobD-cbiB-btuR-btuCD-X-cobUTS; &cbiMNQO-cobA-cbiK-cbiL-cbi(HC)-(ET)-(GF)-(JD); chl(ID)-cobN-&btuB2; cbiA-btuF2; &nrdJ; &frd*
	Thermotoga maritima	TM	1	*&btuFCD; btuR*
SP	*Treponema denticola #*	TDE	3	*&btuFCD; cbiK-cbiLA; cbiG-cbiF; cbiHJ-btuD3; cbiET; X-cbiC; cbiD; cobUSC; btuR; cbiB; X-cbiP; cobD; chlID-cobN--btuFCD2; &cbtF]; &rocG*
	Leptospira interrogans	LI	2	*&btuB; btuF; &(cbiX-cbiW)-X-frd-cbiDC-cbi(ET)-cbiLG-cbi(H?)-cbiF-btuR-cbiA-cobX-cobU-cbiPB; cobTSC*
	B.burgdorferi,T.pallidum (BB,TP)		0	*no*
A	*Thermoplasma volcanicum*	TVO	0	*cysG^B-cobA-cbiCHDTLF-cbi(GE); cbiPB; X-cobT; X-cbiA-XX; cobDS; btuR; cobC; cobX cobY><btuD-X; hoxN; cobX; btuF; btuC*
	Methanosarcina acetivorans	MAC	0	*cbiTLFGHC; cbiDE; cbiMNQO; cbiA1; cbiA2; cbiP; (cobD-crdX)-cbiB--cobZ-cobS-cobY; btuR; cobT; btuW-btuT-btuS; cobX; btuFC-X-btuD*
	Halobacterium sp.	HSL	0	*cbiTLFG-cbiH1-H2-X-(cbiX-cbiW)-X><chl(ID)<>cobN-cbiC-cbiE><cbiD<>cobT-cbiA-btuR-cbiP-HSL01294-cbiB-cobSYD-cobX; btuF<>btuCD*
	Archaeoglobus fulgidus	AG	0	*cbiT-cbiMNQO<>cbiLFG-cbi(HC)-cbiDE-cbiX-cblA; cbiB-X><cbiP; cbiA; cobY-X-cobS1; cobS2; ??-cobD-X; cobT; cobX; btuCD-XX-btuF*
	Aeropyrum pernix	AP	0	*cobT-(cobX-cobZ)-cobY-cobD-cobS-cbiB-X-cblA; $~btuFCD btuF<$>mutBA-ygfD*
	Methanopyrus kandleri	MK	0	*X-cbiC; X-cbiD; cbiT<>cbiL; cbiFGH; cbiE; cbiA1; cbiA2-X; cbiB; cbiP-X; ??-cobT; (cobX-cobZ)-cobS-cobY; ?-cobD*
	Methanococcus jannaschii	MJ	0	*cbiC; cbiD; cbiE; cbiF; cbiG; X-cbiH; cbiJ; cbiL; cbiT; cbiMNQO; cbiP; cbiA; X-cbiB-X; cobS-X; X-cobT; cobD; cobY; cobZ; cobX; crdX; btuF<>btuCD*
	Methanobacterium thermoaut.	TH	0	*cbiC-X; cbiD; cbiE; cbiF; cbiB-cbiG; (cbiH-crdX); X-cbiJ; cbiL; cbiT; cbiMNQO; cbiP; cbiA; (cobX-cobZ)-cobS-X; cobD; cobT-X; X-cobY;cobN2-btuW-btuT*
	Pyrobaculum aerophilum	PK	0	*cbiCHDTLF; cbiGE; X-cobD-cobS-cbiB; cbiA; cobT-(cobX-cobZ)-btuR;*
	Pyrococcus spp. (PH,PO,PF)		0	*cobD-cbiB<$>cblA-cobZ--cobS-cobY; $cobT; btuR; cobX; $nrdDG; $mutB*-$ygfD-mmcE; $mutB^; $sucS; $btuF; $btuCD*
	Sulfolobus solfataricus	SS	0	*$cbiGECHDTLF; cbiP<>cobS-cbiB; $cobT; X-cobD; cobY; cobC; cbiA; $hoxN; cobX; $btuFCD*

element (Table I). Interestingly, the regulatory *B12* element was found upstream of the *exbBD-tonB* operon from *Rhodobacter capsulatus* encoding common components of the TonB-dependent receptors for ferric siderophores and vitamin B_{12}.

Various variants of incomplete B_{12} transport systems were revealed in some bacteria. The *btuFCD* genes were absent in *Nitrosomonas europaea* and *Xanthomonas axonopodis*, and the *btuCD* genes were absent in *B. pertussis*, *Methylobacillus flagellatus*, *Azotobacter vinelandii, Listeria monocytogenes*, and *Leptospira interrogans.* The *btuB* gene of *N. europaea*, *M. flagellatus*, *A. vinelandii*, and *X. axonopodis* is located within the *btuB-btuM-btuR* cluster, which is a single fused gene in the latter bacterium. The hypothetical protein BtuM is not similar to any known protein and has five predicted transmembrane segments, indicating that, in these bacteria, BtuM may be a new type of transmembrane component of the B_{12} transporter, substituting the BtuC and BtuD proteins. The *btuB-btuN* cluster, one more example of the conserved positional linkage between BtuB and another hypothetical transmembrane protein (BtuN has four predicted TMSs), was found in BtuCD-deficient *B. pertussis*, *M. flagellatus*, and *X. axonopodis*. Similarly to BtuM, BtuN may be involved in the BtuCD-independent transport of vitamin B_{12}. The BtuFC system of *S. aureus* is another example of an incomplete transporter that does not include a specific ATPase, suggesting that it can share the ATPase component with some other ABC transport system.

Cobalt Transporters—The *cbiMNQO* locus encoding an ATP-dependent transport system for cobalt was identified in the CBL-producing microorganisms from different taxonomic groups including enterobacteria, ϵ- and δ-proteobacteria, the *Bacillus/Clostridium* group, cyanobacteria, actinobacteria, chloroflexi, and archaea (Table II). In most cases, the *cbiMNQO* genes were found either within large CBL operons or as separate operons and were preceded by regulatory *B12* elements. However, among 56 CBL-producing bacteria in this study, only 24 possess this high affinity cobalt transport system. This indicates the existence of other cobalt-specific transporters required for the CBL biosynthesis. Analysis of possible operon structures and regulatory *B12* elements allowed us to identify new candidate cobalt transporters (Table II).

We assign cobalt specificity to seven uncharacterized transporters from the HoxN family in various proteobacteria and archaea. Notably, most characterized members of this family are specific for nickel ions, but only one HoxN-type transporter was known as a cobalt transporter associated with Co^{2+}-depedent nitrile hydratase (13). Genes for the predicted HoxN-type transporters of cobalt are *B12*-regulated and co-localized with CBL-biosynthetic genes in eubacteria. Predicted co-regulation of the *hoxN* gene with CBL genes in *S. solfataricus* argues for the cobalt specificity of archaeal HoxN transporters as well (Table I and see above).

TABLE II
Differences in the predicted cobalt transporters required for cobalamin synthesis in prokaryotes

Tax	Genome	Candidate cobalt transporters
α	MLO, BME, AU	CbtAB
	BJA, SM, PD, RS	CbtC
	RPA, SAR	HoxN
	RC	CbiMNQO
β	BPS, RSO	HoxN
γ	TY, KP, YE	CbiMNQO
	PP, PU, PY, PA	CbtAB
B/C	BE, BI, HMO, DHA LMO, CA, CPE, CB, DF	CbiMNQO
Act	TFU, RK	CbtE
	DI	?
	MT	CbtG
	SX, PI	CbiMNQO
Cya	PMA, CY, SN	HupE
	AN, TEL	CbiMNQO
CFB	PG, BX	CbtD
	CL	CbiMNQO
SP	TDE	CbtF
	LI	?
A	TVO, STO	HoxN
	MAC, MJ, TH, AG	CbiMNQO
	HSL, MK, PK	?
Other	GME, MCO, CAU	CbiMNQO
	FN	CbtF

Two other *B12*-regulated genes, *cbtA* and *cbtC*, detected in various α-proteobacteria and pseudomonades (one per genome), possibly encode cobalt transporters with five predicted TMSs. These genes are not similar to any known protein and have only *B12*-regulated homologs, the majority of which are positionally linked to CBL genes. In addition, *cbtA* is always co-localized (or fused in *P. aeruginosa*) with a short gene, *cbtB*, which encodes one TMS followed by a histidine-rich motif probably involved in metal binding. In result, α-proteobacteria are predicted to possess at least four different types of cobalt transporters (CbiMNQO, HoxN, CbtAB, and CbtC).

In three cyanobacterial species that do not have the CbiMNQO transporter, the only member of the B_{12} regulon is the hypothetical transmembrane protein HupE with a histidine-rich metal-binding motif at its N terminus (TrEMBL accession number P73671). Other proteins from the HupE family are required for activities of Ni^{2+}-dependent hydrogenases and ureases and thought to be involved in nickel transport (33). Analysis of *B12* elements allows us to assign cobalt specificity to HupE-type transporters in cyanobacteria.

A candidate cobalt transporter of *Porphyromonas gingivalis* and *Bacteroides fragilis*, encoded by the *B12*-regulated gene *cbtD*, contains 10 predicted TMSs and a ligand-binding TrkA-like domain between TMS V and VI. In mycobacteria, the *B12*-regulated gene *cbtG* encodes a predicted cobalt transporter

TABLE III
Differences in the cobalamin biosynthetic pathways of prokaryotes

Tax	Genome	CobG or CbiG	Cobalt chelatases
α	MLO, BJA, SM, PD, SAR, AU, BME	CobG	CobN + CobST
	RPA, RC	ORF663	CobN + CobST
	RS	?	CobN + CobST
β	BPS	CobG	CobN + ChlID/CobST
	RSO	CbiG	CbiX; CobN + ChlID
γ	TY, KP, YE	CbiG	CbiK
	PP, PU, PY, PA	CobG	CobN + ChlID
B/C	BE, BI, HMO, DHA	CbiG	CbiX
	CA, CPE, CB, DF, LMO	CbiG	CbiK
Act	TFU, RK, DI, MT	CobG	CobN + ChlID
	SX	CbiG	CbiX; CobN + ChlID
	PI	CbiG	CbiX; $CysG^B$
Cya	PMA, CY, SN, TEL	CbiG	CbiX
	AN	CobG	?
SP	TDE	CbiG	CbiK; CobN + ChlID
	LI	CbiG	CbiX
CFB	PG, BX	CbiG	CbiK
	CL	CbiG	CbiK; CobN + ChlID
A	TVO	CbiG	$CysG^B$
	MJ, TH, MAC, MK, PK, STO	CbiG	?
	HSL, AG	CbiG	CbiX; CobN + ChlID
Other	MCO, GME	CbiG	CbiX
	FN	CbiG	CbiK
	CAU	CbiG	$CysG^B$; CobN + ChlID

with seven TMSs. In two other actinobacteria, *Thermobifida fusca* and *Rhodococcus* str., we identified one more candidate cobalt transporter, encoded by the *B12*-regulated gene *cbtE*, which has six possible TMSs and a histidine-rich loop between transmembrane segments I and II.

A new member of the B_{12} regulon in *Treponema denticola*, CbtF, has a predicted signal peptide cleavage site on its N terminus. In *Brucella melitensis*, *cbtF* lies in one locus with Ni^{2+}-dependent urease immediately upstream of the *cbiMQO* genes and is thought to be involved in the nickel transport. In addition, *Rhodobacter capsulatus* and *F. nucleatum* have a single *cbtF* gene, which is *B12*-regulated in the former genome. In view of the existence of mixed cobalt/nickel families of transporters (see above), CbtF, in conjunction with systems homologous to CbiMQO, could be involved in cobalt transport.

Cobalt Chelatases—Insertion of cobalt ions into CR at early and late stages of the CBL biosynthesis is mediated by different cobalt chelatases, termed here as "early" and "late." There are at least two distinct early chelatases, CbiK and CbiX, and one late chelatase composed of CobN, CobS, and CobT subunits. In contrast to the CobN component of the late chelatase, which is widely distributed in bacteria, the CobS and CobT components were found only in α-proteobacteria and *Burkholderia pseudomallei*, where the *cobST* cluster is always separated from other CBL genes and never has an upstream *B12* element. Thus, the CobST components of cobalt chelatase are missing in actinobacteria and pseudomonades. Nevertheless, genes similar to the Mg-chelatase subunits, namely *chlID*, were found in clusters with various CBL genes, most often with *cobN*. Notably, similar to CobNST, ATP-dependent Mg-chelatase involved in the bacteriochlorophyll biosynthesis consists of three subunits, ChlH, ChlI, and ChlD, and ChlH is a close homolog of CobN (34). The hypothesis that ChlI and ChlD are the missing components of the late cobalt chelatase was proved by phylogenetic analysis. Indeed, proteins associated with the chlorophyll and CBL biosynthesis form separate branches on the phylogenetic trees for both CobN/ChlH and ChlI families (data not shown). Based on these facts, we suggest that, in contrast to α-proteobacteria, the late cobaltochelatase complex of actinobacteria, pseudomonads, and β-proteobacteria consists of the CobN, ChlI, and ChlD subunits (Table III).

CBL gene clusters of proteobacteria possessing late cobalt chelatase contain the hypothetical gene *cobW*. This gene is always located immediately upstream of the *cobN* chelatase component (Table I). Interestingly, the N-terminal part of CobW has a P-loop nucleotide-binding motif and is similar to the urease/hydrogenase accessory proteins UreG and HypB, which are involved in the GTP-dependent incorporation of Ni^{2+} into the metallocenters of target enzymes. In addition, the variable loop between the conserved N- and C-terminal domains of CobW contains a histidine-rich motif possibly involved in metal binding. Finally, *cobW* is co-localized with predicted cobalt transporters in some genomes. We suggest that CobW is required for the cobalt chelation during CBL biosynthesis and it is possible that the histidine-rich region of CobW is used to store the cobalt ions within of the cell prior to their delivery to the chelatase complex.

Another predicted member of the B_{12} regulon, the *cfrX* gene, was found within the conserved gene cluster *cfrX-cobW-cobN* in all CBL-producing α-proteobacteria except *Sinorhizobium meliloti* and *B. melitensis*. CfrX is weakly similar to various ferredoxin proteins including the CBL-related ferredoxin CbiW of *B. megaterium* (35) and could act as an oxidoreductant during cobalt insertion.

Cobalt Reductases—Reduction of the cobalt ion of corrinoids is the least studied stage the CBL biosynthesis. It is a prerequisite for further corrinoid adenosylation. Although the NADH-dependent flavoprotein with cobalt reductase activity was purified in *P. denitrificans*, the gene encoding this activity has not been identified (36). In *S. typhimurium*, however, *in vitro* studies showed that flavodoxin FldA can catalyze the co(II)rrinoid reduction when the latter is bound to the adenosyltransferase enzyme (37). Using a similarity search, the *fldA* gene was found in enterobacteria, *Pasteurellaceae*, ϵ-proteobacteria, and cyanobacteria as well as in the *Bacillus/Clostridium* and CFB groups of bacteria. Since FldA was found in some CBL pathway-deficient bacteria, its function is unlikely to be restricted to the B_{12} synthesis. It is consistent with experimental facts that the FldA protein, shared by several metabolic pathways, is an essential enzyme in *E. coli* (38). Nevertheless, FldA was not found in bacteria with the aerobic CBL biosynthetic pathway.

In this work, a candidate cobalt reductase associated with the CBL biosynthesis was identified in most bacteria with the aerobic CBL pathway (α-proteobacteria, burkholderia, pseudo-

TABLE IV

Predicted B12-element-mediated regulation of bacterial genes, which are not involved in the CBL biosynthesis and transport

The cases of the predicted *CBL*-box-mediated regulation in pyrococci are underlined.

Gene	Function	Genome
B_{12}-independent isozymes of B_{12}-dependent enzymes		
&metE	Methionine synthase	α-Proteobacteria (MLO,BJA,RPA,AU,CO); bacilli (ZC,HD); actinobacteria (MT,ML,SX), cyanobacteria (TEL), CFB group (BX)
&nrdAB	Aerobic ribonucleotide reductase	α-Proteobacteria (BME,AU); β-proteobacteria (MFL); bacilli (HD,BE); actinobacteria (SX); CFB group (BX)
&nrdDG	Anaerobic ribonucleotide reductase	α-Proteobacteria (RC), *Bacillus*/*Clostridium* group (DF,DHA); CFB group (PG,BX); pyrococci (PH,PO,PF)
B_{12}-dependent or alternative metabolic pathways		
&rocG	Glutamate DHG (glutamate fermentation)	*T. denticola* (TDE)
&butDA . . .	Succinate fermentation	*P. gingivalis* (PG)
&mutB, sucS, mmcE	Succinate-propionate fermentation	Pyrococci (PH,PO,PF)
Predicted enzymes of unknown pathway		
&ardX-frdX	Hypothetical dioxygenase	α-Proteobacteria (MLO,SM,RC)
&achX	Hypothetical acyl-CoA hydrolase	α-Proteobacteria (AU); bacilli (HD), *D. radiodurans* (DR)

monads, and actinobacteria), as well as in several other species (Table I). The corresponding gene, *bluB*, was previously described in *R. capsulatus* as a gene of unknown function essential for the CBL synthesis (39). We found that *bluB* orthologs are predominantly regulated by *B12* elements and often co-localized with various CBL genes. The BluB proteins are similar to various FMN-dependent reductases from the nitroreductase family, including oxygen-insensitive NADH nitroreductase and NADH-flavin oxidoreductase, that catalyze the electron transfer from NADH to various electron acceptors. The *bluB* and *fldA* genes never co-occur in CBL-producing proteobacteria. The majority of these bacteria have only *bluB*, whereas it is not present in the *fldA*-containing group of enterobacteria. Thus, we propose that BluB functions in cobalt reduction for the CBL biosynthesis. Moreover, *bluB* was also found in three proteobacteria with incomplete CBL pathways, *M. flagellatus*, *A. vinelandii*, and *Ralstonia eutropha*. The CBL pathways in these bacteria include all genes for the conversion of cobyrinic acid to Ado-CBL and are possibly involved in the assimilation of exogenous corrinoids. We believe that predicted cobalt reductase BluB, which presumably acts on cob(II)yrinic acid *a*,*c*-diamide, is necessary for these incomplete pathways as well.

ATP:Corrinoid Adenosyltransferases—The active form of coenzyme B_{12}, Ado-CBL, can be obtained either by *de novo* synthesis or by assimilation of exogenous corrinoids. Both routes require ATP:corrinoid adenosyltransferase encoded by the *btuR* gene. BtuR adenosylates either CBL or an intermediate prior to CBL. The search for *btuR* in bacterial genomes showed that this widely distributed gene is usually co-localized with other CBL genes and sometimes is *B12*-regulated (Table I). However, among nearly 80 B_{12}-utilizing species, BtuR was not found in 18 genomes. It is known that, besides BtuR, enterobacteria possess two other CBL adenosyltransferases, PduO and EutT, which are associated with the CBL-dependent 1,3-propanediol dehydratase and ethanolamine ammonia lyase encoded by the *pdu* and *eut* gene clusters, respectively (40, 41). Strikingly, the BtuR, PduO, and EutT adenosyltransferases show no sequence similarity. Homologs of the *eutT* gene were found only within *eut* gene clusters, strongly suggesting that its only function is in the ethanolamine utilization. In contrast, *pduO* appears to be widely distributed in prokaryotes and, in particular, in most BtuR-deficient bacteria, where it would fill the gaps in the CBL pathways (Supplementary Table VI). The *pduO* genes may reside within the *pdu* and CBL gene clusters or be single genes. Notably, the phylogenetic tree of the PduO family contains distinct branches corresponding to the CBL- and PDU-associated genes and to the single genes (data not shown).

In summary, adenosyltransferases were found in all B_{12}-utilizing prokaryotes except two clostridia and three methanogenic archaea. However, the latter have only one B_{12}-dependent enzyme, a methyl-CBL-dependent methyltransferase, and thus do not require the adenosylated form of CBL. Interestingly, *pduO* from *Archaeoglobus fulgidus* and *btuR* from *Geobacter metallireducens* appear in one putative operon with B_{12}-dependent methylmalonyl-CoA mutase. Overall, it seems that particular types of adenosyltransferases are specialized for particular B_{12}-dependent enzymes or for the *de novo* CBL biosynthesis.

Nucleotide Loop Assembly Pathway—The pathways for the lower ligand synthesis of CBL (also known as the nucleotide loop assembly) are thought to vary between bacterial groups (42). Some bacterial genomes have neither *cobT* nor *cobC* genes required for the synthesis of α-ribazole from dimethylbenzimidazole or have only one of these genes. In contrast, two other genes of the nucleotide loop assembly pathway, *cobU* and *cobS*, are conserved in all CBL-synthesizing bacteria with the exception of *cobU* in archaea (see below).

Three Gram-positive bacteria, *L. monocytogenes*, *Clostridium botulinum*, and *Thermoanaerobacter tengcongensis*, lack the *cobT* gene but have all other genes for nucleotide loop assembly, including *cobC*. Instead, the CBL gene clusters of these bacteria contain two hypothetical genes, named *cblT* and *cblS*, which are not similar to any known protein. However, these two genes were found in several other Gram-positive bacteria simultaneously with *cobT*. The *cblTS* operon of *Clostridium perfringes* and single *cblS* genes of *D. halfniense* and *Heliobacillus mobilis* are preceded by regulatory *B12* elements. In addition, the *B12*-regulated CBL operon of *B. stearothermophilus* contains the *cblTS* genes. The hypothetical protein CblT has five predicted transmembrane segments. These facts allow us to propose a possible role of new CblT and CblS proteins in uptake of dimethylbenzimidazole and its subsequent transformation into α-ribazole-5P, respectively.

A case of nonorthologous displacement of the CBL genes was previously found in archaea, where the bacterial-type nucleotidyltransferase CobU is replaced by a new nucleotidyltransferase named CobY (43). Here we extend this analysis using 13 instead of 6 complete archaeal genomes. All of these genomes appear to lack the *cobU* gene and possess the *cobY* gene, thus confirming the nonorthologous displacement. The only exception is *Pyrobaculum aerophilum*, which lacks both nucleotidyltransferases. In 9 of 12 archaeal species, *cobY* is positionally linked to other CBL genes.

TABLE V
Predicted functions for B_{12}-related genes

Various techniques of genome analysis were used to infer these functional predictions: clustering on the chromosome (C); protein fusion events (F); co-occurrence phylogenetic profiles (O); analysis of regulatory sites (R); similarity search (S); motif search (M); and prediction of transmembrane segments (T). NOD, nonorthologous gene displacement. Asterisks mark new gene names given in this study. Genome abbreviations are listed in Table I.

Protein	Suggested function	Genomes	Reasons/comments
BtuM*	B_{12} transporter component	NE,MFL,AV,XAX	T, R, O, C, F
BtuN*	B_{12} transporter component	BP,MFL,XAX,RPA,BJA,PA,BX	T, R, O, C
CbtAB*	Cobalt transporter	AU,MLO,BME,PA,PU,PY,PP	T, R, O, C
CbtC*	Cobalt transporter	BJA,SM,RS	T, R, O, C
HupE*	Cobalt transporter	PMA,CY,SN	T, R, O, M
CbtD*	Cobalt transporter	PG, BX	T, R, O, M
CbtE*	Cobalt transporter	TFU,RK	T, R, O, M
CbtF*	Cobalt transporter	TDE,FN,RC	T, R, O, M
CbtG*	Cobalt transporter	MT,ML	T, R, O
CnoABCD*	Cobalt transporter	RC, DHA	T, R
ChlID	Cobalt chelatase components	BPS,PP,PA,PU,PY,RSO,TFU,RK,DI,MT,SX,CAU,TDE,CL,HSL	S, R, C, O
CobW	Possibly involved in cobalt chelation	MLO,BJA,SM,BME,AU,RPA,RC,RS,SAR,PA,PU,PP,PY,BPS	S, R, C, M
CfrX*	Putative ferredoxin	MLO,BJA,AU,RPA,RC,RS,SAR,AN	S, R, C
BhiB	Cobalt reductase	MLO,BJA,SM,BME,AU,RPA,RC,RS,SAR,BPS,PA,PU,PP,PY, MFL,AV,REU,RSO,MT,TFU,RK,SX,MCO,CL	S, R, C, O
CblT*	DMB transporter	BE,LMO,CPE,CB,THT,HMO	T, R, C
CblS*	α-ribazole-5P synthesis	BE,LMO,CPE,CB,THT,HMO,DHA,HD	R, C
CobY	NOD for CobU	TVO,MAC,HSL,AG,AP,MK,MJ,PH,PO,PF,SS	S, O, C
CobZ*	NOD for CobC	MAC,AP,MK,MJ,TH,PK,PH,PO,PF	S, O, C
HSL01294	NOD for CobC	HSL	S, O, C
CblXY*	NOD for CobC	MLO,SM,BME,AU	O, C
CblZ*	NOD for CobC	CGL,DI,MT,ML,TFU,RK,SX	O, C
PduX	L-Threonine kinase	CA, CB, DF, HMO, DHA, SX, SY, YE	S, R, C
BtuS*	Chelatase for metalloporphyrine salvage	NE,MFL, RPA, PA, PG,BX,MAC,TH	S, O, C
BtuT*	Transport of various metalloporphyrines	NE,MFL,RPA, PA,PG,BX,MAC,TH	T, O, C
BtuW*	Transport of various metalloporphyrines	MAC,TH	T, C
CbiW	Putative ferredoxin	BME,BE,RSO,CAU,AN,LI,HSL	S, R, O, C, F
MetZ*	NOD for CobF	MT	S, O, C
Frd*	Ferredoxin	LI,CL	S, R, C

Here we identified one more case of nonorthologous gene displacement of CBL genes in archaea. The *cobC* gene encoding α-ribazole-5P phosphatase was found in only two archaea, *S. solfataricus* and *Thermoplasma*. In all other archaea, except *A. fulgidus* and *Halobacterium* sp., the hypothetical gene *cobZ* (PF0294 in *P. furiousus*) was found within the CBL gene clusters, being most often linked with *cobS*. CobZ is weakly similar to bacterial phosphatidylglycerophosphatases but not to CobC. Thus, we predict that CobZ is the missing α-ribazole-5P phosphatase replacing CobC. In *Halobacterium* sp., the CBL locus contains the hypothetical gene *HSL01294*, another predicted nonorthologous displacement of CobC, which is weakly similar to the phosphoglycolate phosphatases from proteobacteria but has no orthologs in other archaea.

Among CBL-synthesizing α-proteobacteria, only four rhizobacteria, *Mesorhizobium loti*, *S. meliloti*, *B. melitensis*, and *A. tumefaciens*, lack the *cobC* gene. Instead, we have found a pair of genes, named *cblXY*, which is clustered with the *cobT* and *cobS* genes. Similarly, all CBL-synthesizing actinobacteria lack the *cobC* gene but have a new hypothetical gene, named *cblZ*, which is mainly clustered with *cobT*, *cobS*, and *cobU* genes. The hypothetical proteins CblX, CblY, and CblZ are not similar to any known protein; nevertheless, the CblX, containing a zinc ribbon motif, comprises a small metal-binding protein of about 60 amino acids. Thus, we propose that *cblXY* and *cblZ* are nonorthologous replacements of the *cobC* gene in rhizobacteria and actinobacteria, respectively.

Other New Members of the B_{12} Regulon Related to the CBL Biosynthesis—The 1,2-propanediol utilization operon *pdu* of *S. typhimurium* includes the *pduX* gene of unknown function (44). Here we found that some Gram-positive bacteria, namely three clostridia, *L. monocytogenes*, *D. halfniense*, and *H. mobilis*, contain the *pduX* genes located within the CBL gene clusters, usually adjacent to the *cobD* gene (Table I). Most of these clusters are predicted to be *B12*-regulated. In addition, the single *pduX* gene preceded by a *B12* element was found in *S. coelicolor*. Finally, *Y. enterocolitica* has two copies of *pduX*, and one of them is co-localized with the *cobD* gene. The PduX protein belongs to the GHMP kinase family and is weakly similar to galactokinase, L-homoserine kinase, and mevalonate kinase. Since the CBL biosynthesis requires L-threonine-3P as a substrate for the CobD aminotransferase, and the positional analysis shoes that PduX is probably CBL-related, we propose the L-threonine kinase function to PduX.

The phylogenetic tree of CobN-related proteins has three main branches (data not shown). The first two branches comprise the *B12*-regulated cobalt chelatases CobN involved in the CBL synthesis and the magnesium chelatases ChlH required for the bacteriochlorophyll biosynthesis. The third branch, named here BtuS, includes hypothetical proteins of unknown function from diverse bacterial and archaeal genomes. In all cases, the *btuS* genes are clustered with a new gene, named *btuT*, which encodes a hypothetical transporter with four predicted TMSs. In addition, the *btuST* clusters of *P. aeruginosa*, *R. palustris*, and *N. europaea* are co-localized with hypothetical outer membrane receptor genes encoding proteins homologous to the vitamin B_{12} receptor BtuB. Moreover, in *P. gingivalis* and *B. fragilis*, the *btuST* genes form a candidate operon with the iron-induced hemoglobin transport genes *hmuYR* (45). The BtuB-like HmuR receptor was recently found to bind hemoglobin, hemin, various porphyrins, and metalloporphyrins (46). In archaea, the *btuST* genes are linked to the *btuW* gene encoding a hypothetical transporter with seven predicted TMSs. These observations allow us to propose that the hypothetical transporter BtuT, chelatase BtuS, and homologs of the BtuB/HmuR receptors (or BtuW in archaea) are involved in the transport and salvage of various metalloporphyrines rather than in the CBL biosynthesis.

The first gene of the *B. megaterium cbi* operon, *cbiW*, encodes a hypothetical ferredoxin and could be involved in the CBL

biosynthesis, possibly acting as oxidoreductant during the ring contraction process under anaerobic conditions (35). Homologs of *cbiW* are widely distributed in prokaryotes, but only some of them are clustered with CBL genes and regulated by *B12* elements. The *cbiW* genes are co-localized with cobalt transporters and CBL biosynthetic genes in *Ralstonia solanacearum* and *B. stearothermophilus* and with B_{12} transport systems in *Chloroflexus aurantiacus* and *Anabaena* sp. In *L. interrogans* and *Halobacterium* sp., CbiW and the cobalt chelatase CbiX are encoded by a single fused gene. Thus, the B_{12}-related ferredoxins CbiW occur only in bacteria with anaerobic CBL pathways (see below).

M. tuberculosis, in contrast to most actinobacteria, lacks the *cobF* gene, but the CBL cluster of this bacterium contains another gene, named *metZ*, which is not similar to *cobF* but is similar to various methyltransferases. We predict that *metZ* is the possible nonorthologous gene displacement of *cobF*.

The CBL gene cluster of *R. solanacearum*, a bacterium without an ortholog of the bifunctional methyltransferase/decarboxylase Cbi(ET), contains a distant homolog of the CbiE methyltransferases from archaea. This exemplifies a possible xenologous gene displacement, whereby CbiE is displaced by a horizontally transferred homolog from another lineage. However, the CbiT-associated activity is still missing in this bacterium.

Differences in Prokaryotic Cobalamin Biosynthetic Pathways—Identification of known CBL genes and new *B12*-regulated genes allows us to reconstruct and compare the CBL pathways in various organisms. In addition to cobalt transporters and chelatases (see above), the enzymatic step of ring contraction during the CR biosynthesis is highly variable in bacteria (Table III). This reaction is mediated by the CobG or CbiG proteins and determines the aerobic or anaerobic type of the CBL pathway, respectively, since CobG, in contrast to CbiG, is an oxygen-dependent enzyme (7). Although *B. melitensis*, *B. pseudomallei*, *Anbaena* sp., and *Pseudomonas* species have both CobG and fused CbiG-CbiH proteins, the CbiG domains in these bacteria contain a large deletion and, therefore, may be nonfunctional. In contrast to other α-proteobacteria, *R. capsulatus* and *R. palustris*, lacking the CobG mono-oxygenase, have another enzyme, ORF663, involved in the ring contraction during CBL biosynthesis (47).

Analyzing genomes of 56 CBL-producing bacteria, we detected a correlation between the time of cobalt insertion and the oxygen dependence of the CBL pathway. The CobG mono-oxygenases were found only in bacteria with the ATP-dependent CobN-CobST/ChlDI cobaltochelatase complexes corresponding to the late cobalt insertion. With the exception of several archaeal genomes, where we could not detect cobaltochelatase genes, the CbiG proteins co-occur with ATP-independent CbiK/CbiX chelatases corresponding to the early cobalt insertion. In addition, *C. aurantiacus* and *T. volcanicum* are predicted to have early cobalt chelatases of another type, which are similar to the ferrochelatase $CysG^B$. The remaining question is the function of the additional *cobN-chlID* chelatases in the genomes of *R. solanacearum*, *S. coelicolor*, *C. aurantiacus*, *C. tepidum*, *T. denticola*, and *Halobacterium* sp., where they co-occur with cobalt chelatases *cbiK/cbiX* and *cbiG* and are co-localized with CBL genes (Table III).

In contrast to CobG, the exact biochemical role of CbiG in the CBL biosynthesis is unknown (7). Identification of a pair of genes from one genome that appear to be fused into a single gene within another genome represents strong evidence that the functions implemented by these genes may be closely related (48). In an attempt to identify the CbiG-catalyzed reaction in the CBL pathway, we summarized all CbiG-related protein fusion events. The CbiG-CbiH fusion proteins appear in *Pseudomonas* species, *B. melitensis, B. pseudomallei*, *S. coelicolor*, and cyanobacteria, whereas the CbiG-CbiF fusions were found in the CFB group of bacteria. Thus, we place CbiG between CbiH and CbiF on the pathway of CBL biosynthesis (Fig. 1).

B12-regulated Genes Not Involved in the CBL Biosynthesis and Transport—Analysis of the regulatory *B12* elements in bacterial genomes allowed us to detect *B12*-regulated genes that are not involved in the CBL biosynthesis. An unexpected result was that most of these genes appear to belong to B_{12}-dependent metabolic pathways (Table IV).

First, in some α-proteobacteria, actinobacteria, and *Bacillus* species, as well as in *B. fragilis* and *T. elongatus*, *B12* elements were found upstream of the *metE* gene encoding the B_{12}-independent methionine synthase. On the other hand, genes encoding the NrdDG and NrdAB ribonucleotide reductases are preceded by *B12 element*s in three α-proteobacteria, two *Bacillus* species, the CFB and *Thermus/Deinococcus* groups, *M. flagellatus*, *S. coelicolor*, *C. difficile*, and *D. halfniense*. To our knowledge, there are only two B_{12}-dependent enzymes, methionine synthase MetH and ribonucleotide reductase isozyme NrdJ, that are known to have B_{12}-independent isozymes, MetE and NrdAB/NrdDG, respectively (4, 6). To put these scattered observations into a more general context, we scanned bacterial genomes for the presence of both B_{12}-dependent and -independent isozymes and found that the B_{12}-independent isozymes are regulated by *B12* elements in most bacteria that have both isozymes. Although arhaeal genomes lack regulatory *B12* elements, in three *Pyrococcus* genomes with both NrdJ (B_{12}-dependent) and NrdDG (B_{12}-independent) isozymes, the *nrdDG* genes are predicted to be co-regulated with CBL biosynthetic genes via conserved *CBL*-boxes (see above). Thus, we propose that when vitamin B_{12} is present in the cell, expression of B_{12}-independent isozymes is inhibited, and only relatively more efficient B_{12}-dependent isozymes are used.

The *rocG* gene, encoding a catabolic glutamate dehydrogenase, has an upstream *B12* element in *T. denticola*. Further, this bacterium has an ortholog of the B_{12}-dependent glutamate mutase MutSL, which is known to catalyze the first step of the B_{12}-dependent pathway of glutamate catabolism (49). Moreover, MutSL is the only B_{12}-dependent enzyme found in *T. denticola*. These findings allow us to propose that, first, *T. denticola* has two alternative pathways of glutamate utilization, and second, an excess of vitamin B_{12}, repressing expression the *rocG* gene, would inhibit the B_{12}-independent glutamate pathway in this bacterium.

The predicted B_{12} regulon in *Pyrococcus* species includes the *mutB*, *sucS*, and *mmcE* genes, which are thought to be involved in the B_{12}-dependent succinate-propionate fermentation pathway. In *B. fragilis*, a *B12 element* precedes the *pccCAB* operon encoding propionyl-CoA carboxylase, an enzyme from the same B_{12}-dependent pathway. In *P. gingivalis*, a hypothetical *B12* element-regulated operon, named *butD-butA-4hbD-sucD*, encodes enzymes of the B_{12}-independent pathway of succinate fermentation, namely 4-hydroxybutanoyl-CoA dehydratase, 4-hydroxybutyrate coenzyme A transferase, NAD-dependent 4-hydroxybutyrate dehydrogenase, and succinate-semialdehyde dehydrogenase.

As demonstrated above, several genes for B_{12}-dependent and alternative pathways are often members of the vitamin B_{12} regulons both in eubacteria and archaea. This raises the possibility of identifying previously unknown B_{12}-dependent enzymes based on analysis of regulatory *B12* elements. In this vein, we identified a new member of the B_{12} regulon in *B. halodurans*, hypothetical acyl-CoA hydrolase AchX, which belongs to the thioesterase superfamily. This family includes

4-hydroxybenzoyl-CoA thioesterase, which catalyzes the final step in the catabolism of 4-hydroxybenzoate in *Pseudomonas* CBS-3 (50), and various cytosolic long-chain acyl-CoA thioester hydrolases. The *achX* gene was found in one *B12*-regulated operon with B_{12}-independent ribonucleotide reductase *nrdBA* in *Deinococcus radiodurans*. The candidate *achX-metR* operon of *A. tumefaciens* is also preceded by a *B12* element. Another new member of the B_{12} regulon, named the *ardX-frdX* operon, was found in three α-proteobacteria, *S. meliloti*, *M. loti*, and *R. capsulatus*. The hypothetical ArdX and FrdX proteins are highly similar to the alpha and ferredoxin-like subunits of various bacterial ring-hydroxylating dioxygenases. However, ArdX-FrdX and AchX orthologs from several other bacteria have no upstream *B12* elements. The only possible explanation of observed *B12* element-dependent regulation of the hypothetical *achX* and *ardX-frdX* genes is that they could encode B_{12}-independent analogs of yet unidentified B_{12}-dependent enzymes.

DISCUSSION

The biosynthesis of coenzyme B_{12} (Ado-CBL) is a metabolic pathway widely distributed in bacteria and archaea, but it is not found in eukaryotes. In addition, many prokaryotes have active transport systems for vitamin B_{12} and related compounds. Identification of the B_{12}-specific regulatory elements allows us to identify new genes related to the CBL biosynthesis. As a result, we reconstructed and compared the CBL biosynthesis pathways in various organisms. The most variable parts of the CBL pathway are the CobG/CbiG-mediated reaction of the corrin ring synthesis and cobalt chelation that could occur at either early or late stage of the pathway. The CobG and CbiG proteins determine the aerobic or anaerobic types of the CBL pathway. The type of a cobalt chelatase corresponds to the time of cobalt insertion and seems to be correlated with oxygen dependence of the CBL pathway (Table III). Furthermore, we observed two major corrinoid adenosyltransferases and nine different cobalt transport systems in various prokaryotes (Table II).

Identification of all known B_{12}-dependent enzymes in prokaryotic genomes allowed us to select bacterial species requiring coenzyme B_{12} for their metabolism. Not surprisingly, all of these genomes are capable of either *de novo* synthesis or transport of this vitamin or both. The only exception is the complete genome of *B. cereus*, which has neither CBL biosynthetic nor known transport genes but has B_{12}-dependent methionine synthase *metH*. On the other hand, there are bacteria (*e.g. B. subtilis* and *S. aureus*) that possess a vitamin B_{12} transporter but lack any known B_{12}-dependent enzyme. This indicates that previously unknown B_{12}-dependent enzymes may exist in these bacteria.

The metabolic reconstruction techniques reveal a large number of missing genes in the CBL biosynthetic pathways of various bacteria. Simultaneous analysis of gene clusters on the chromosome, protein fusion events, phylogenetic profiles, and regulatory *B12* elements allowed us to make functional assignment for several new genes related to the CBL biosynthesis (Table V). About half of them encode various transporters, whereas the remaining ones are enzymes involved in the CBL synthesis. In particular, we tentatively identified eight additional cobalt transporters, two vitamin B_{12} transporters, one 5,6-dimethylbenzimidazole transporter, and two possible transporters for various metalloporphyrines. Among new enzymes, we ascribed cobalt reductase function to BluB, cobalt chelatase function to ChlDI, and L-threonine kinase function to PduX as well as the involvement of the CobW, CfrX, and CbiW proteins in oxidation-reduction processes during the corrin ring synthesis. In addition, most functions corresponding to missing genes in several genomes were assigned to nonorthologous genes. Most remarkably, we identified the nonorthologous gene displacements for the *cobC* gene in archaea, α-proteobacteria, and actinobacteria. However, among complete genomes, still missing functions in the CBL pathway are CbiA in *C. perfringes*, CobD in *Shewanella oneidensis* and *L. interrogans*, CobU and CbiP in *P. aerophilum*, CobC in *C. tepidum*, and CbiJ in *L. interrogans* and in almost all archaeal genomes.

Using the global analysis of the *B12* elements in available bacterial genomes, we have found that this conserved RNA regulatory element is widely distributed in eubacteria and regulates most CBL genes. The *B12* elements do not occur in archaea, but we identified candidate B_{12}-regulatory operator sites in several archaeal genomes. Among all bacterial genes related to the CBL biosynthesis, only cobalt transporter genes, both known and predicted, are always *B12*-regulated. The only exceptions are the *cbiMNQO* operon in two cyanobacteria, *Anabaena* sp. and *T. elongatus*, and the *cbtF* gene in *F. nucleatum*. Most vitamin B_{12} transport systems as well as cobalt chelatases are also regulated by *B12* elements.

In this work, we for the first time demonstrated that *B12* elements regulate not only genes related to the CBL biosynthesis and transport but also several genes from B_{12}-dependent pathways. It appears that in most cases, the B_{12}-independent isozymes of methionine synthase and ribonucleotide reductase are regulated by *B12* elements in the genomes possessing both B_{12}-dependent and B_{12}-independent isozymes. Although the repression of B_{12}-independent enzymes by the excess of coenzyme B_{12} looks rational, this regulatory strategy was not previously known. This finding, together with identification of other *B12* element-regulated enzymes not related to the CBL biosynthesis and mostly hypothetical, opens an intriguing possibility to reveal new B_{12}-dependent pathways. In particular, the *ardX-frdX* gene pair, existing in most α-proteobacteria, has an upstream *B12* element in three bacterial species. Therefore, we predict the existence of a novel, alternative to ArdX-FrdX, B_{12}-dependent enzyme in these three α-proteobacteria.

From the practical standpoint, this work once again demonstrates the power of comparative genomics for functional annotation of genomes, especially when experimental data are limited. In particular, analysis of regulatory elements is a powerful tool for prediction of missing transport genes, as demonstrated here and in our analyses of other vitamin regulons (20, 21).

Acknowledgments—We are grateful to Andrei Osterman for attention, advice, and encouragement.

REFERENCES

1. Banerjee, R. (2001) *Biochemistry* **40,** 6191–6198
2. Daniel, R., Bobik, T. A., and Gottschalk, G. (1999) *FEMS Microbiol. Rev.* **22,** 553–566
3. Jordan, A., Torrents, E., Jeanthon, C., Eliasson, R., Hellman, U., Wernstedt, C., Barbe, J., Gibert, I., and Reichard, P. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 13487–13492
4. O'Toole, G. A., Rondon, M. R., Trzebiatowski, J. R., Suh, S.-J., and Escalante-Semerena, J. C. (1994) *Escherichia coli and Salmonella: Cellular and Molecular Biology* (Neidhardt, F. C., ed) pp. 710–720, American Society for Microbiology, Washington, D. C.
5. Sauer, K., and Thauer, R. K. (1999) *Eur. J. Biochem.* **261,** 674–681
6. Martens, J. H., Bargv, H., Warren, M. J., and Jahn, D. (2002) *Appl. Microbiol. Biotechnol.* **58,** 275–285
7. Scott, A. I., and Roessner, C. A. (2002) *Biochem. Soc. Trans.* **30,** 613–620
8. Debussche, L., Couder, M., Thibaut, D., Cameron, B., Crouzet, J., and Blanche, F. (1992) *J. Bacteriol.* **174,** 7445–7451
9. Raux, E., Thermes, C., Heathcote, P., Rambach, A., and Warren, M. J. (1997) *J. Bacteriol.* **179,** 3202–3212
10. Raux, E., Leech, H. K., Beck, R., Schubert, H. L., Santander, P. J., Roessner, C. A., Scott, A. I., Martens, J. H., Dahn, D., Thermes, C., Rambach, A., and Warren, M. J. (2002) *Biochem. J.* **370,** 505–516
11. Debussche, L., Thibaut, D., Cameron, B., Crouzet, J., and Blanche, F. (1993) *J. Bacteriol.* **175,** 7430–7440
12. Smith, R. L., Banks, J. L., Snavely, M. D., and Maguire, M. E. (1993) *J. Biol. Chem.* **268,** 14071–14080
13. Komeda, H., Kobayashi, M., and Shimizu, S. (1997) *Proc. Natl. Acad. Sci.*

U. S. A. **94,** 36–41
14. Roth, J. R., Lawrence, J. G., Rubenfield, M., Kieffer-Higgins, S., and Church, G. M. (1993) *J. Bacteriol.* **175,** 3303–3316
15. Cadieux, N., Bradbeer, C., Reeger-Schneider, E., Koster, W., Mohanty, A. K., Wiener, M. C., and Kadner, R. J. (2002) *J. Bacteriol.* **184,** 706–717
16. Nou, X., and Kadner, R. J. (1998) *J. Bacteriol.* **180,** 6719–6728
17. Ravnum, S., and Andersson, D. I. (2001) *Mol. Microbiol.* **39,** 1585–1594
18. Nahvi, A., Sudarsan, N., Ebert, M. S., Zou, X., Brown, K. L., and Breaker, R. R. (2002) *Chem. Biol.* **9,** 1043–1049
19. Gelfand, M. S., Novichkov, P. S., Novichkova, E. S., and Mironov, A. A. (2000) *Brief Bioinform.* **1,** 357–371
20. Rodionov, D. A., Vitreschak, A. G., Mironov, A. A., and Gelfand, M. S. (2002) *J. Biol. Chem.* **277,** 48949–48959
21. Vitreschak, A. G., Rodionov, D. A., Mironov, A. A., and Gelfand, M. S. (2002) *Nucleic Acids Res.* **30,** 3141–3151
22. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Wheeler, D. I. (2003) *Nucleic Acids Res.* **31,** 23–27
23. Overbeek, R., Larsen, N., Walunas, T., D'Souza, M., Pusch, G., Selkov, E. Jr., Liolios, K., Joukov, V., Kaznadzey, D., Anderson, I., Bhattacharyya, A., Burd, H., Gardner, W., Hanke, P., Kapatral, V., Mikhailova, N., Vasieva, O., Osterman, A., Vonstein, V., Fonstein, M., Ivanova, N., and Kyrpides, N. (2003) *Nucleic Acids Res.* **31,** 164–171
24. Vitreschak, A. G., Mironov, A. A., and Gelfand, M. S. (2001) *Procedings of the Third International Conference on "Complex Systems: Control and Modeling Problems," Samara, Russia, September 4–9, 2001*, pp. 623–625, The Institute of Control of Complex Systems, Samara, Russia
25. Mironov, A. A., Vinokurova, N. P., and Gelfand, M. S. (2000) *Mol. Biol.* **34,** 222–231
26. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) *Nucleic Acids Res.* **25,** 4876–4882
27. Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. (2000) *Nucleic Acids Res.* **28,** 33–36
28. Felsenstein, J. (1981) *J. Mol. Evol.* **17,** 368–376
29. Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. (1997) *Nucleic Acids Res.* **25,** 3389–3402
30. Vitreschak, A. G., Rodionov, D. A., Mironov, A. A., and Gelfand, M. S. (2003) *RNA* **9,** 1084–1097
31. Gelfand, M. S., Koonin, E. V., and Mironov, A. A. (2000) *Nucleic Acids Res.* **28,** 695–705
32. Koster, W. (2001) *Res. Microbiol.* **152,** 291–301
33. McMillan, D. J., Mau, M., and Walker, M. J. (1998) *Gene* (*Amst.*) **208,** 243–251
34. Schubert, H. L., Raux, E., Wilson, K. S., and Warren, M. J. (1999) *Biochemistry* **38,** 10660–10669
35. Raux, E., Lanois, A., Rambach, A., Warren, M. J., and Thermes, C. (1998) *Biochem. J.* **335,** 167–173
36. Blanche, F., Maton, L., Debussche, L., and Thibaut, D. (1992) *J. Bacteriol.* **174,** 7452–7454
37. Olczak, T., Dixon, D. W., and Genco, C. A. (2001) *J. Bacteriol.* **183,** 5599–5608
38. Fonseca, M. V., and Escalante-Semerena, J. C. (2001) *J. Biol. Chem.* **276,** 32101–32108
39. Gaudu, P., and Weiss, B. (2000) *J. Bacteriol.* **182,** 1788–1793
40. Pollich, M., and Klug, G. (1995) *J. Bacteriol.* **177,** 4481–4487
41. Johnson, C. L., Pechonick, E., Park, S. D., Havemann, G. D., Leal, N. A., and Bobik, T. A (2001) *J. Bacteriol.* **183,** 1577–1584
42. Kofoid, E., Rappleye, C., Stojiljkovic, I., and Roth, J. (1999) *J. Bacteriol.* **181,** 5317–5329
43. Cheong, C. G., Escalante-Semerena, J. C., and Rayment, I. (2001) *J. Biol. Chem.* **276,** 37612–37620
44. Thomas, M. G., and Escalante-Semerena, J. C. (2000) *J. Bacteriol.* **182,** 4227–4233
45. Bobik, T. A., Havemann, G. D., Busch, R. J., Williams, D. S., and Aldrich, H. C. (1999) *J. Bacteriol.* **181,** 5967–5975
46. Simpson, W., Olczak, T., and Genco, C. A. (2000) *J. Bacteriol.* **182,** 5737–5748
47. McGoldrick, H., Deery, E., Warren, M., and Heathcote, P. (2002) *Biochem. Soc. Trans.* **30,** 646–648
48. Enright, A. J., Iliopoulos, I., Kyrpides, N. C., and Ouzounis, N. C. (1999) *Nature* **402,** 86–90
49. Buckel, W. (2001) *Appl. Microbiol. Biotechnol.* **57,** 263–273
50. Benning, M. M., Wesenberg, G., Liu, R., Taylor, K. L., Dunaway-Mariano, D., and Holden, H. M. (1998) *J. Biol. Chem.* **273,** 33572–33579