Comparative Genomics of the Vitamin B₁₂ Metabolism and Regulation in Prokaryotes

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 B₁₂) 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 B₁₂ regulator was identified in some archaea. A search for B₁₂ elements and CBL-boxes and positional analysis identified a large number of new candidate B₁₂-regulated genes in various prokaryotes. Among newly assigned functions associated with the cobalamin biosynthesis, there are several new types of cobalt transporters, ChiI and ChiD subunits of the CobN-dependent cobalt cleavase, cobalt reductase BluB, adenosyl-transferase 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 B₁₂-independent isozymes of the methionine synthase and ribonucleotide reductase are regulated by B₁₂ elements in bacteria that have both B₁₂-dependent and B₁₂-independent isozymes. Moreover, B₁₂ regulons of various bacteria are thought to include enzymes from known B₁₂-dependent or alternative pathways.

Cobalamin (CBL), along with chlorophyll, heme, siroheme, and coenzyme F₄₃₀, 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 B₁₂ (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 B₁₂-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 ethanamine ammonia lyase from enteric bacteria involved in anaerobic glycerol, 1,2-propanediol, and ethanamine 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 UroIII, 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, chiE and chiT, 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, ChiK from S. typhimurium and ChiX from B. megaterium, which are associated with the anaerobic pathway (8–10). Sec-
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
mainly accumulated by the fast and unspecific CorA transport system (12). An additional cobalt transporter, a part of the cobalt-dependent nitrite 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 ChiMNQO, encoded by the CBL biosynthetic operon in *S. typhimurium*, probably mediates high affinity transport of cobalt ions for the B12 synthesis (14). Vitamin B12, cobinamide, and other corrinoids are actively transported in enteric bacteria using the TonB-dependent outer membrane receptor BtuB and other corrinoid transporters (18). The evolutionarily conserved B12-box, a molecule involved in the regulation of CBL genes in enterobacteria (17), is a powerful approach to the search of missing genes within a metabolic pathway in a variety of bacterial species, applied to a metabolic pathway in a variety of bacterial species, applied to a metabolic pathway in a variety of bacterial species, applied to a metabolic pathway in a variety of bacterial species, applied to a metabolic pathway in a variety of bacterial species, applied to a metabolic pathway in a variety of bacterial species, applied to a metabolic pathway n 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 B12 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 B12 transporters are widely distributed in bacteria and archaea and mostly B12-regulated. Finally, the B12 element was predicted to regulate B12-independent methionine synthase and ribonucleotide reductase isozymes in bacteria that also have corresponding B12-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. 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).

---

1 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\textsubscript{12}-dependent regulation of the \textit{btaB} and \textit{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 \textit{E. coli btaB} and \textit{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\textsubscript{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 \textit{btaB} 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\textsubscript{12}-related genes based on formation of alternative RNA structures involving the B12 elements is suggested (30).

\textit{B12} 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 \textit{cysG} of \textit{E. coli}, which encodes UROIII methyltransferase (CysG\textsuperscript{A}) and precorrin-2 oxidase/ferrochelatase (CysG\textsuperscript{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\textsubscript{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 \textit{Bacillus cereus}, contain CBL biosynthesis and/or transport genes. Most obligate pathogenic bacteria (see below) as well as \textit{Aquifex aeolicus} have neither CBL genes nor B12 elements. \textit{Staphylococcus aureus}, \textit{Corynebacterium glutamicum}, \textit{Bordetella pertussis}, \textit{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 palindromic with consensus 5’-TGGAAnTATCCCA-3’, was observed in candidate cobalamin regulons in three \textit{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 \textit{btaR}. In addition, conserved CBL-boxes were identified upstream of the \textit{P. hortikoshii} genes PH0021, PH1306, PH0275, PH1928, and PH0272 and their orthologs in two other pyrococci. These genes are predicted to encode an-aerobic ribonucleotide reductase NrdDG, two subunits of methyllumalonyl-CoA mutase MutB, succinyl-CoA synthase SucS, and methyllumalonyl-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 B12-dependent enzymes (see below). The same CBL-specific profiles were obtained for two other archaea, \textit{Aeropyrum pernix} and \textit{Sulfolobus solfataricus}, but not for the remaining archaeal species. The predicted CBL regulon of \textit{A. pernix} again contains the B12 transport system and methyllumalonyl-CoA mutase. Among all archaea in this study, only pyrococci and \textit{P. horikoshii} are likely to be unable to synthesize CBL de novo but may uptake and transform CBL precursors to Ado-CBL. The CBL regulon of \textit{S. solfataricus} includes, in addition to the cobT and \textit{btuFCD} genes, the \textit{chiGECHDTLF} genes for the de novo CBL synthesis and predicted cobalt transporter \textit{hoxN}.

To select bacterial species that potentially require coenzyme B12 for their metabolism, we carried out a similarity search for all known B12-dependent enzymes in prokaryotic genomes. As a result, \textit{Chlamydia spp., Rickettsiae spp., Neisseria spp., Streptococcoceae, Mycoplasmataceae, Pasteurellaceae, \varepsilon-proteobacteria, Borellia burgdorferi, Treponema pallidum, and Xylella fastidiosa} (obligate pathogenic bacteria) as well as \textit{A. aeolicus} were found to have no B12-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 B12-dependent enzyme, \textit{Bacillus subtilis} and \textit{S. aureus}, were predicted to have the B12 transport system \textit{BtuFCD}. Interestingly, \textit{btuFCD-pduO} is the only B12 element-regulated operon in \textit{B. subtilis}. This shows that other, currently unknown, B12-dependent enzymes may be present in these bacteria.

Vitamin B12 Transporters—Nearly one-fourth of the B12-utilizing bacteria appear to have no complete pathway for the CBL biosynthesis and, therefore, should actively transport vitamin B12 or some precursor from the external medium. The only known transport system for vitamin B12 is the ABC transport system 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 B12 across the outer membrane involves B12-specific receptor BtuB and the periplasmic energy-coupling proteins TonB, ExbB, and ExbD, which are shared between various TonB-dependent receptors. Thus, the B12-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 B12 are similar and belong to the same families (32). Therefore, a similarity search is not sufficient to dissect the B12 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 B12 and ferric transporters revealed B12-specific subfamilies within each family (data not shown). The predicted transporters for vitamin B12 were found to be widely distributed in prokaryotes; among B12-utilizing bacteria with complete genomes, they were not found only in four cyanobacterial and three archaeal species, in \textit{Mycobacterium spp.}, and in \textit{Bacillus cereus} (Supplementary Table VI). In most cases, components of B12 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 divergents is shown by angle brackets. For example, cobD/H11021/H11022 denotes divergently transcribed genes, whereas chib/H11022/chip 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, /H9251/, /H9252/, /H9253/, /H9255/, and /H9254/-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; EO, Erwinia carotovora; HI, Hemophilus influenzae; VK, Pasteurella multocida; AB, Actinobacillus actinomycetemcomitans; HP, Helicobacter pylori; CJ, Campylobacter jejuni; LL, Lactococcus lactis; FMA, Prochlorococcus marinus; SN, Synechococcus sp.; CY, Synechocystis sp.; BB, B. burgdorferi; TP, T. pallidum; PH, P. horikoshii; PF, P. furiosus; PO, P. abyssi.
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 B₁₂.

Various variants of incomplete B₁₂ 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*, *Methyllobacillus 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 does not include a specific ATPase, suggesting that it can share a 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/Clorstridium* 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 *BtuB*-regulated 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 Co²⁺-dependent nitrile hydratase (13). Genes for the predicted HoxN-type transporters of cobalt are *BtuFCD*-regulated and co-localized with CBL-biosynthetic genes in *Bacillus* and *Listeria* species. 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).

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 (*ChibMNQO*, HoxN, CbtAB, and CbtC).

In three cyanobacterial species that do not have the *ChibMNQO* transporter, the only member of the B₁₂ 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 similar to HupE-type transporters in cyanobacteria. Two other B12-regulated genes, *cbtD* 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 (*ChibMNQO*, HoxN, CbtAB, and CbtC).

In three cyanobacterial species that do not have the *ChibMNQO* transporter, the only member of the B₁₂ 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 Ni²⁺-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.
were found only in widely distributed in bacteria, the CobS and CobT components contrast to the CobN component of the late chelatase, which is composed of CobN, CobS, and CobT subunits. In at least two distinct early chelatases, CbiK and CbiX, and one single late chelatase (see above), CbtF, in conjunction with systems homologous to the chelatase complex.

In view of the existence of mixed cobalt/nickel families of transporters (see above), CbtF, in conjunction with systems homologous to ChlI/ChlD, was found within the conserved gene cluster cfrX-cobW-cobN (35) and could act as an oxidoreductant during cobalt insertion.

Another predicted member of the B12 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 CobW 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 in the CBL biosynthesis. It is a prerequisite for further cobinamide 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 protein, shared by several metabolic pathways, 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.

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

**TABLE III**

Differences in the cobalamin biosynthetic pathways of prokaryotes

<table>
<thead>
<tr>
<th>Tax</th>
<th>Genome</th>
<th>CobG or ChlG</th>
<th>Cobalt chelatases</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>MLO, BJ, SM, PD, SAR, AU, BME</td>
<td>CobG</td>
<td>CobN + CobST</td>
</tr>
<tr>
<td></td>
<td>RPA, RC</td>
<td>ORF663</td>
<td>CobN + CobST</td>
</tr>
<tr>
<td></td>
<td>RS</td>
<td>?</td>
<td>CobN + CobST</td>
</tr>
<tr>
<td>β</td>
<td>BPS</td>
<td>CobG</td>
<td>CobN + CobST</td>
</tr>
<tr>
<td></td>
<td>RSO</td>
<td>ChiG</td>
<td>ChiX; CobN + ChlID</td>
</tr>
<tr>
<td>γ</td>
<td>TY, KP, YE</td>
<td>ChiG</td>
<td>ChiK</td>
</tr>
<tr>
<td></td>
<td>PP, FU, PY, PA</td>
<td>CobG</td>
<td>CobN + ChlID</td>
</tr>
<tr>
<td>B/C</td>
<td>BE, BI, HMO, DHA</td>
<td>ChiG</td>
<td>ChiX</td>
</tr>
<tr>
<td></td>
<td>CA, CPE, CB, DF, LMO</td>
<td>ChiG</td>
<td>ChiK</td>
</tr>
<tr>
<td>Act</td>
<td>TFU, RK, DI, MT</td>
<td>CobG</td>
<td>CobN + ChlID</td>
</tr>
<tr>
<td></td>
<td>SX</td>
<td>ChiG</td>
<td>ChiX; CobN + ChlID</td>
</tr>
</tbody>
</table>
|     | PI     | ChiG | ChiX; CysG

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 cfrX, which has six possible TMSs and a histidine-rich loop between transmembrane segments I and II.

A new member of the B12 regulon in Treponema denticola, CbtF, has a predicted signal peptide cleavage site on its N terminus. In Brucella melitensis, cfrF lies in one locus with Ni²⁺-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 cfrF 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 CobW, 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 ChlI, 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 has an upstream B12 element. Thus, the CobST components of cobalt chelatase are missing in actinobacteria and pseudomonads. 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 CobH (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 cobaltchelatase 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 Ni²⁺ 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.
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 *fdA* genes never co-occur in CBL-producing proteobacteria. The majority of these bacteria have only *bluB*, whereas it is not present in the *fdA*-containing group of enterobacteria. Thus, we propose that BluB functions in cobalt reductase for the CBL biosynthesis. Moreover, *bluB* was also found in three proteobacteria with incomplete CBL pathways, *M. flagellatus*, *A. vinelandii*, and *Ratlstonia 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 cobyrinic acid a,c-diamide, is necessary for these incomplete pathways as well.

ATP:Corrinoid Adenosyltransferases—The active form of cobenzyme B12, 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 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 B12-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 B12-utilizing prokaryotes except two clostridia and three methanogenic archaea. However, the latter have only one B12-dependent enzyme, a methyl-CBL-dependent methytransferase, and thus do not require the adenosylated form of CBL. Interestingly, *pduO* from *Archaeoglobus fulgidus* and *btuR* from *Geo bacter metallireducens* appear in one putative operon with B12-dependent methylmalonyl-CoA mutase. Overall, it seems that particular types of adenosyltransferases are specialized for particular B12-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 genomes 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 perfringens* and single *cblS* genes of *D. hali anesi* and *Helio bacterium mobilis* are preceded by regulatory B12 elements. In addition, the B12-regulated CBL operon of *B. steaothermophilus* 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 nucleotidyldtransferase CobU is replaced by a new nucleotidyldtransferase 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 nucleotidyldtransferases. In 9 of 12 archaeal species, *cobY* is positionally linked to other CBL genes.

### Table IV

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

<table>
<thead>
<tr>
<th>Gene</th>
<th>Function</th>
<th>Genome</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>BtuR</em></td>
<td>Adenosyltransferase</td>
<td>α-Proteobacteria (MLO,BJA,RPA,AU,CO); bacilli (ZC,HD); actinobacteria (MT,ML,SX); cyanobacteria (TEL); CF B group (BX)</td>
</tr>
<tr>
<td><em>AtoU</em></td>
<td>Aerobic ribonucleotide reductase</td>
<td>α-Proteobacteria (BME,AU); β-proteobacteria (MFL); bacilli (HD,HE); actinobacteria (SX); CFB group (BX)</td>
</tr>
<tr>
<td><em>PduO</em></td>
<td>Anaerobic ribonucleotide reductase</td>
<td>α-Proteobacteria (RC); Bacillus/Clostridium group (DF,DHA); CFB group (PG,BX); <em>pyrococci</em> (PH,PO,PF)</td>
</tr>
<tr>
<td><em>B12</em>-dependent or alternative metabolic pathways</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>AceG</em></td>
<td>Glutamate DHIG (glutamate fermentation)</td>
<td><em>T. denticola</em> (TDE)</td>
</tr>
<tr>
<td><em>ButDA</em></td>
<td>Succinate fermentation</td>
<td><em>P. gingivalis</em> (PG)</td>
</tr>
<tr>
<td><em>MutB, sucS, mmcE</em></td>
<td>Succinate-propionate fermentation</td>
<td><em>Pyrococci</em> (PH,PO,PF)</td>
</tr>
<tr>
<td>Predicted enzymes of unknown pathway</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>aroK-aroF, arX</em></td>
<td>Hypothetical dioxygenase</td>
<td>α-Proteobacteria (MLO,SM,RC)</td>
</tr>
<tr>
<td><em>achX</em></td>
<td>Hypothetical acyl-CoA hydrolase</td>
<td>α-Proteobacteria (AU); bacilli (HD), <em>D. radiodurans</em> (DR)</td>
</tr>
</tbody>
</table>

The cases of the predicted CBL-box-mediated regulation in pyrococci are underlined.
Here we identified one more case of nonorthologous gene displacement of CBL genes in archaea. The cobC gene encoding a-ribose-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* *cobS* genes, being most often linked with and are nonorthologous replacements of the cblZ protein of about 60 amino acids. Thus, we propose that similar to any known protein; nevertheless, the CblX, contain-genes. The hypothetical proteins CblX, CblY, and CblZ are not *S. solfataricus* and BtuT* Transport of various metalloporphyrines NE,MFL,RPA, PA,PG,BX,MAC,TH T, O, C
Frd* Ferredoxin LI,CL S, R, C

<table>
<thead>
<tr>
<th>Protein</th>
<th>Suggested function</th>
<th>Genomes</th>
<th>Reasons/comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>BtuM*</td>
<td>B12 transporter component</td>
<td>NE,MFL,AV,XAX</td>
<td>T, R, O, C, F</td>
</tr>
<tr>
<td>BtuN*</td>
<td>B12 transporter component</td>
<td>BP,MFL,XAX,RPA,BJA,PA,BX</td>
<td>T, R, O, C</td>
</tr>
<tr>
<td>ChtAB*</td>
<td>Cobalt transporter</td>
<td>AU,MLO,BME,PA,PU,PP,PP</td>
<td>T, R, O, C</td>
</tr>
<tr>
<td>ChtC*</td>
<td>Cobalt transporter</td>
<td>BJA,SM,RS</td>
<td>T, R, O, C</td>
</tr>
<tr>
<td>HupP*</td>
<td>Cobalt transporter</td>
<td>PMA,CY,SN</td>
<td>T, R, O, M</td>
</tr>
<tr>
<td>ChtD*</td>
<td>Cobalt transporter</td>
<td>PG, BX</td>
<td>T, R, O, M</td>
</tr>
<tr>
<td>ChtE*</td>
<td>Cobalt transporter</td>
<td>TFU,RK</td>
<td>T, R, O, M</td>
</tr>
<tr>
<td>ChtF*</td>
<td>Cobalt transporter</td>
<td>TDE,FN,RC</td>
<td>T, R, O, M</td>
</tr>
<tr>
<td>ChtG*</td>
<td>Cobalt transporter</td>
<td>MT,ML</td>
<td>T, R, O</td>
</tr>
<tr>
<td>CwaABC*</td>
<td>Cobalt transporter</td>
<td>RC, DHA</td>
<td>T, R, O</td>
</tr>
<tr>
<td>CobW</td>
<td>Possibly involved in cobalt chelation</td>
<td>MLO,BJA,SM,BME,AU,RPA,RC,RS,SAR,PA,PU,PP,PP,PPS</td>
<td>S, R, C, M</td>
</tr>
<tr>
<td>ChlX*</td>
<td>Putative ferredoxin</td>
<td>MLO,BJA,AU,RPA,RC,RS,SAR,AN</td>
<td>S, R, C</td>
</tr>
<tr>
<td>BhiB</td>
<td>Cobalt reductase</td>
<td>MLO,BJA,SM,BME,AU,RPA,RC,RS,SAR,BS,PP,PA,PP,PP</td>
<td>S, R, C, O</td>
</tr>
<tr>
<td>ChlT*</td>
<td>DMB transporter</td>
<td>BE,LMO,CPE,CB,THT,HMO</td>
<td>T, R, C</td>
</tr>
<tr>
<td>ChlS*</td>
<td>a-ribose-5P synthesis</td>
<td>BE,LMO,CPE,CB,THT,HMO,PHA,HD</td>
<td>R, C</td>
</tr>
<tr>
<td>CobY</td>
<td>NOD for CobU</td>
<td>TVO,MAC,HSL,AG,AP,PK,MP,HF,PO,PP,SSF</td>
<td>S, O, C</td>
</tr>
<tr>
<td>CobZ*</td>
<td>NOD for CobC</td>
<td>MAC,AP,PK,MP,HF,PF,PO,PP</td>
<td>S, O, C</td>
</tr>
<tr>
<td>HSLO1294</td>
<td>NOD for CobC</td>
<td>HSL</td>
<td>S, O, C</td>
</tr>
<tr>
<td>ChlXY*</td>
<td>NOD for CobC</td>
<td>MLO,SM,BME,AU</td>
<td>O, C</td>
</tr>
<tr>
<td>ChlZ*</td>
<td>NOD for CobC</td>
<td>CGL,DI,MT,TFU,RF,SK</td>
<td>O, C</td>
</tr>
<tr>
<td>PduX</td>
<td>l-Threonine kinase</td>
<td>CA, CB, DF, HMO, DHA, SX, SY, YE</td>
<td>S, R, C</td>
</tr>
<tr>
<td>BtuS*</td>
<td>Chelatase for metalloporphyrine salvage</td>
<td>NE,MFL, RPA, PA, PG, BX, MAC, TH</td>
<td>S, O, C</td>
</tr>
<tr>
<td>BtuT*</td>
<td>Transport of various metalloporphyrines</td>
<td>NE,MFL, RPA, PA, PG, BX, MAC, TH</td>
<td>T, O, C</td>
</tr>
<tr>
<td>BtuW*</td>
<td>Transport of various metalloporphyrines</td>
<td>MAC,TH</td>
<td>T, C</td>
</tr>
<tr>
<td>ChlB</td>
<td>Putative ferredoxin</td>
<td>BME,BE,RSO,CAU,AN,LI,HS</td>
<td>S, R, O, C, F</td>
</tr>
<tr>
<td>MeZ*</td>
<td>NOD for CobF</td>
<td>MT</td>
<td>S, O, C</td>
</tr>
<tr>
<td>Frd*</td>
<td>Ferredoxin</td>
<td>LL,CL</td>
<td>S, R, C</td>
</tr>
</tbody>
</table>

Here we found that some Gram-positive bacteria, namely *S. typhimurium* usually adjacent to the cobD gene. Among CBL-synthesizing actinobacteria, *Mesorhizobium loti* and *N. europaea* are 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 shows that PduX is probably CBL-related, we propose the l-threonine kinase 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 B12 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
Cobalamin Biosynthesis in Bacteria

41517

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 B12 transport systems in Chloroflexus aurantius and Anabaena sp. In L. interogans and Halobacterium sp., CbiW and the cobalt chelatase CbiX are encoded by a single fused gene. Thus, the B12-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 ChiET, contains a distant homolog of the ChiE methyltransferases from archaea. This exemplifies a possible xenologous gene displacement, whereby CbiI is displaced by a horizontally transferred homolog from another lineage. However, the ChIT-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, Anabaena 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/CbiD cobaltchelatase complexes corresponding to the late cobalt insertion. With the exception of several archaeal genomes, where we could not detect cobaltchelatase genes, the CbiG proteins co-occur with ATP-dependent CbiK/CbiX chelatases corresponding to the early cobalt insertion. In addition, C. aurantius and T. vulcanicum are predicted to have early cobalt chelatases of another type, which are similar to the ferrochelatase CysGQ. The remaining question is the function of the additional cobN-chiDID chelatases in the genomes of R. solanacearum, S. coelicolor, C. aurantius, 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 B12-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 B12-dependent methionine synthase. On the other hand, genes encoding the NrdDG and NrdAB ribonucleotide reductases are preceded by B12 elements in three α-proteobacteria, two Bacillus species, the CFB and Thermus/Deinococcus groups, M. flagellatus, S. coelicolor, C. difficile, and D. haliniiense. To our knowledge, there are only two B12-dependent enzymes, methionine synthase MetH and ribonucleotide reductase isozyme NrdJ, that are known to have B12-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 B12-dependent and -independent isozymes and found that the B12-independent isozymes are regulated by B12 elements in most bacteria that have both isozymes. Although arhaela genomes lack regulatory B12 elements, in three Pyrococcus genomes with both NrdJ (B12-dependent) and NrdDG (B12-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 B12 is present in the cell, expression of B12-independent isozymes is inhibited, and only relatively more efficient B12-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 B12-dependent glutamate mutase MutSL, which is known to catalyze the first step of the B12-dependent pathway of glutamate catabolism (49). Moreover, MutSL is the only B12-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 B12 repressing expression of the rocG gene, would inhibit the B12-dependent glutamate pathway in this bacterium.

The predicted B12 regulon in Pyrococcus species includes the mutB, sucS, and mmcE genes, which are thought to be involved in the B12-dependent succinate-propanoate fermentation pathway. In B. fragilis, a B12 element precedes the pccCAB operon encoding propionyl-CoA carboxylase, an enzyme from the same B12-dependent pathway of glutamate catabolism (49). Moreover, MutSL is the only B12-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 B12 repressing expression of rocG gene, would inhibit the B12-dependent glutamate pathway in this bacterium.

As demonstrated above, several genes for B12-dependent and alternative pathways are often members of the vitamin B12 regulons both in eu- and archaebacteria. This raises the possibility of identifying previously unknown B12-dependent enzymes based on analysis of regulatory B12 elements. In this vein, we identified a new member of the B12 regulon in B. halodurans, a hypothetical acyl-CoA hydrolase AehX, 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 _B12-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 _B12_ 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 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 _B12-independent analogs_ of yet unidentified _B12-dependent enzymes.

**Discussion**

The biosynthesis of coenzyme _B12_ (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 _B12_ and related compounds. Identification of the _B12-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 _B12-dependent_ enzymes in prokaryotic genomes allowed us to select bacterial species requiring coenzyme _B12_ 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 _B12-dependent methionine synthase _metH_. On the other hand, there are bacteria (e.g. _B. subtilis_ and _S. aureus_) that possess a vitamin _B12_ transporter but lack any known _B12-dependent enzyme_. This indicates that previously unknown _B12-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 _B12_ transporters, one _5,6-dimethylbenzimidazole_ transporter, and two possible transporters for various metalloporphyrines. Among new enzymes, we ascribed cobalt reductase function to BhuB, cobalt chelatase function to ChIDL1, 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. perfringens_, 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 genomes. The _B12_ elements do not occur in archaea, but we identified candidate _B12-regulatory_ operons 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 _cblMNQO_ operon in two cyanobacteria, _Anabaena_ sp. and _T. elongatus_, and the _chf_ gene in _F. nucleatum_. Most vitamin _B12_ transport systems as well as cobalt chelatases are also regulated by _B12_ elements. In this work, we found for the first time demonstrated that _B12_ elements regulate not only genes related to the CBL biosynthesis and transport but also several genes from _B12-dependent pathways_. It appears that in most cases, the _B12-independent enzymes_ are regulated by _B12_ elements in the genomes possessing both _B12-dependent_ and _B12-independent_ isoforms. Although the repression of _B12-independent enzymes_ by the excess of coenzyme _B12_ 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 _B12-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, _B12-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**

U. S. A. 94, 36–41