



Article

Translational Frameshifting in the *chlD* Gene Gives a Clue to the Coevolution of the Chlorophyll and Cobalamin Biosyntheses

Stepan Kuznetsov ¹, Alexander Milenkin ¹ and Ivan Antonov ^{2,3,*}

¹ Moscow Institute of Physics and Technology, 141701 Dolgoprudny, Russia; stepan.v.kuznetsov@phystech.edu (S.K.); milenkin.aa@phystech.edu (A.M.)

² Institute of Bioengineering, Research Center of Biotechnology, Russian Academy of Science, 117312 Moscow, Russia

³ Laboratory of Bioinformatics, Faculty of Computer Science, National Research University Higher School of Economics, 101000 Moscow, Russia

* Correspondence: ivan.antonov@gatech.edu

Abstract: Today, hundreds of prokaryotic species are able to synthesize chlorophyll and cobalamin (vitamin B12). An important step in the biosynthesis of these coenzymes is the insertion of a metal ion into a porphyrin ring. Namely, Mg-chelatase ChlIDH and aerobic Co-chelatase CobNST are utilized in the chlorophyll and vitamin B12 pathways, respectively. The corresponding subunits of these enzymes have common evolutionary origin. Recently, we have identified a highly conserved frameshifting signal in the *chlD* gene. This unusual regulatory mechanism allowed production of both the small and the medium chelatase subunits from the same gene. Moreover, the *chlD* gene appeared early in the evolution and could be at the starting point in the development of the chlorophyll and B12 pathways. Here, we studied the possible coevolution of these two pathways through the analysis of the chelatase genes. To do that, we developed a specialized Web database with comprehensive information about more than 1200 prokaryotic genomes. Further analysis allowed us to split the coevolution of the chlorophyll and B12 pathway into eight distinct stages.

Keywords: chlorophyll; cobalamin; vitamin B12; frameshifting; evolution; chelatase; chlIDH; cobNST



Citation: Kuznetsov, S.; Milenkin, A.; Antonov, I. Translational Frameshifting in the *chlD* Gene Gives a Clue to the Coevolution of the Chlorophyll and Cobalamin Biosyntheses. *Microorganisms* **2022**, *10*, 1200. <https://doi.org/10.3390/microorganisms10061200>

Academic Editors: Robert Blankenship and Matthew Sattley

Received: 9 April 2022

Accepted: 2 June 2022

Published: 11 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The early evolution of photosynthesis has been an intriguing but complex research subject for many years [1,2]. Clearly, the appearance of photosynthesis required the ability of the ancient prokaryotes to synthesize chlorophyll (CHL). Chlorophyll is a relatively complex organic compound that includes Mg²⁺ ion in its tetrapyrrole ring [3]. Modern organisms utilize a number of different enzymes to produce chlorophyll from a metal-free tetrapyrrole intermediate Uroporphyrinogen III (Uro) [4]. On the other hand, other species utilize several alternative pathways that can turn Uro into cobalamin (vitamin B12), heme, or coenzyme F430 (Figure 1). These coenzymes contain different metal ions in their tetrapyrrole rings [5]. Namely, cobalamin (vitamin B12), haem, and coenzyme F430 include Co²⁺, Fe²⁺ and Ni²⁺, respectively. The enzymes that insert those ions into the corresponding organic rings are called chelatases.

The biosynthesis of Mg²⁺-containing chlorophylls and bacteriochlorophylls requires the magnesium chelatase that performs the insertion of the magnesium ion (Mg²⁺) into protoporphyrin IX [6]. This enzyme consists of small (I), medium (D), and large (H) subunits [7,8]. The corresponding genes are called *chlI*, *chlD*, and *chlH* in the genomes of chlorophyll-producing organisms and *bchI*, *bchD*, and *bchH* in bacteriochlorophyll producers. To reduce the number of different gene names in this work, we will use the *chlI*, *chlD*, and *chlH* terminology for both groups of genes.

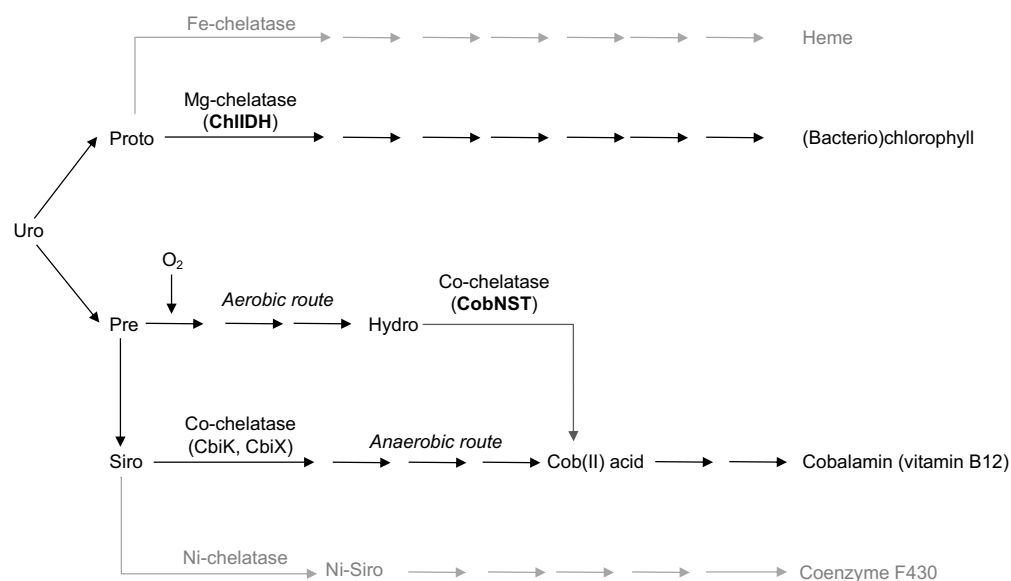


Figure 1. Overview of the existing biochemical pathways that produce (Bacterio)chlorophyll, Heme, Cobalamin, and Coenzyme F430 from Uroporphyrinogen III (Uro). The magnesium-chelatase ChlIDH inserts Mg^{2+} ion in the Protoporphyrin IX (Proto). There are two alternative pathways to synthesize Cobalamin. The CobNST chelatase inserts Co^{2+} in the aerobic path. It has been suggested that ChlIDH and CobNST have common evolutionary origin.

Interestingly, the cobalamin biosynthesis pathway has two alternative routes—the anaerobic and aerobic ones (Figure 1) [9]. Strict anaerobes only possess the anaerobic B12 pathway. On the other hand, cobalamin-producing facultative anaerobes or aerobic organisms usually possess both routes and switch between them depending on the current living conditions [10]. From the evolutionary point of view, it is logical to assume that the aerobic route could only evolve when free oxygen was available in the Earth’s atmosphere, e.g., only after the CHL pathway and oxygenic photosynthesis have already existed. Moreover, the interplay between the CHL pathway and the aerobic B12 route is not limited to the oxygen itself. It has been well established that the cobalt chelatase CobNST from the aerobic route resembles the magnesium chelatase ChlIDH as it also consists of the small, medium and large subunits encoded by *cobS*, *cobT*, and *cobN* genes, respectively [11]. Additionally, the *chlI* and *chlD* from the chlIDH chelatase are likely to be homologous to the *cobS* and *cobT* genes of the cobNST chelatase [12]. The similarity between the large subunit genes (i.e., *chlH* and *cobN*) has also been reported [13].

Later studies have revealed additional existing strategies to encode the cobNST enzyme. First, Rodionov et al. [14] have suggested that in some species, the products of the *chlI* and *chlD* can replace the *cobS* and *cobT* genes and function as the small and medium subunits of the aerobic Co-chelatase. Moreover, a functional programmed ribosomal frameshifting (PRF) found in some *chlD* genes may allow the production of both the small and the medium subunits of the cobNST chelatase from the same gene [15,16]—see Supplementary Figure S1. Altogether, this indicates a close connection between the CHL pathway and the aerobic B12 route suggesting their possible coevolution [17].

Here, we attempted to organize all the different strategies to encode cobNST and chlIDH enzymes and suggest a logical model of the CHL and aerobic B12 pathways coevolution. Our strategy was to pinpoint possible prokaryotic genomes that may still be using the ancient strategies to encode Mg- and/or Co-chelatases. Such “living fossil” species can be used to reconstruct the early stages of the CHL and B12 pathway evolution.

2. Materials and Methods

The Chelatase DB was developed using the Django framework implemented in Python. To populate the database the tBLASTn [18] tool was used to identify the genomes containing *chlD* gene(s). The translations of the three *chlD* genes with the validated frameshifting signals were used as queries [16]. To allow the identification of the other frameshifted *chlD* genes, two separate tBASTn searches (E-value threshold = 10^{-6}) were performed using the N-terminal (i.e., before the frameshift) or the C-terminal (after the frameshift) parts of each query protein. The genomic regions with adjacent hits (in the correct order) with up to one frameshift were classified as *chlD* genes.

In order to identify the small and the large chelatase subunit genes as well as the other genes from the chlorophyll and cobalamin biosynthesis pathways, tBLASTn search was performed for a set of reference query proteins. Given the similarities between different chelatase subunits, we imposed additional constraints on their lengths. Namely, we required that the small chelatase subunits were between 250 aa and 500 aa, medium were between 500 aa and 800 aa, and large were between 1000 aa and 2000 aa (these values corresponded to the lengths of the annotated chelatase subunits). The identified genes were automatically annotated (i.e., assigned a gene name such as *chlD*, *bchD*, or *cobT*) by the best reciprocal hit approach.

We used the identified representative sets of the genes from the CHL and B12 pathways to predict whether a given prokaryotic genome possessed the corresponding biochemical pathway. Additionally, the automatically assigned genotype of the species mentioned in the text was double-checked with the information from the KEGG database (the “Porphyrin metabolism” section) [19]. The phylogenetic tree was constructed using RAxML version 8.2.12 [20]. Subsequent phylogenetic analyses and visualizations were performed using the “ape” R package [21]. The RNA secondary structures were predicted using the RNAfold web server [22] available at <http://ma.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi> (accessed on 1 April 2022).

All the data files and scripts used to generate figures and tables from the article are available at <https://github.com/vanya-antonov/article-chelatase-db> (accessed on 1 April 2022).

3. Results

Our recent analysis [16] as well as earlier studies [8,14,23] have suggested that there were several different strategies to encode aerobic cobalt chelatase cobNST. Interestingly, in a number of species, the products of the Mg-chelatase genes (*chlD* and/or *chlI*) functioned as corresponding subunits of the cobNST complex (Supplementary Figure S1). To explore the diversity of the strategies utilized by different prokaryotes to encode chelatases, we developed a specialized Web database called Chelatase DB (<http://ivanya.com/chelatase/> (accessed on 1 April 2022)). For each prokaryotic genome, the Chelatase DB provided information about the number of the small, medium, and large subunit genes, the existence of the chlorophyll and/or B12 biosynthesis pathway and reference to the KEGG database for detailed analysis. Additionally, the database contained a collection of the frameshifting signals predicted in some *chlD* genes.

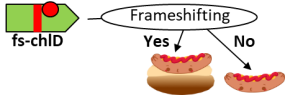
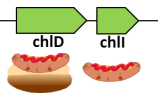
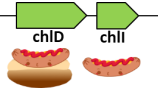

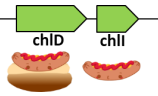

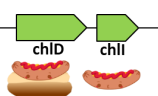
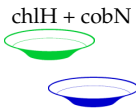
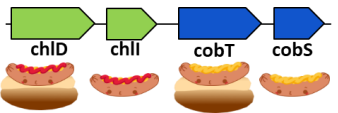
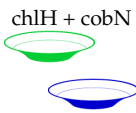
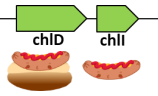

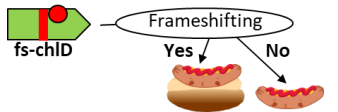

To identify the possible starting point in the evolution of the CHL pathway, the *chlD* gene encoding medium subunit of the Mg-chelatase was analyzed. We chose this gene because it was likely to appear early in the evolution, since it may have been present in the genome of the last universal common ancestor (LUCA) [24]. Thus, it was possible that the modern genes encoding medium chelatase subunits were descendants of the putative *chlD* gene from LUCA. Our previous analysis [16] of the frameshifted *chlD* (*fs-chlD*) genes has demonstrated that such genes were present in all major prokaryotic phyla (including archaea, proteobacteria, and actinobacteria)—Supplementary Figure S2. Indeed, on the phylogenetic tree of 286 medium subunit genes all the *fs-chlD* genes were located near the root of the tree (Supplementary Figure S3). Thus, the modern *chlD*, *bchD*, and *cobT* genes may have originated from the ancient *fs-chlD* gene.

Our next goal was to use the functionality of the Chelatase DB to reveal a possible evolutionary trajectory of the CHL and B12 pathways with respect to the corresponding chelatase genes. Namely, we inspected different chlIDH/cobNST genotypes and matched them with the corresponding predicted phenotypes (i.e., the ability to synthesize CHL and/or B12). This allowed us to define eight stages in the evolution of CHL and B12 pathways as well as *chlD* gene—see Table 1. Since *fs-chlD* gene had an ancient origin, it was logical to assume that it appeared in the evolution before the complete CHL and B12 pathways existed. Therefore, we search the Chelatase DB for the genomes with *fs-chlD* genes that did not have the genes from the CHL and B12 pathways. Indeed, we identified several such species including *Spirochaeta thermophyla* and *Brevefilum fermentans* (Table 1, stage 1). All of these bacteria that corresponded to the conditions of the prechlorophyll era were anaerobic [25]. Moreover, according to the global prokaryotic phylogenetic tree, both Spirochaetes and Chloroflexi were among the most ancient prokaryotic phyla that are present today [26]. The presence of the frameshifting signals in the *fs-chlD* genes from *S. thermophyla* and *B. fermentans* suggested that they produced two proteins from the same mRNA (Supplementary Figure S4). In the *chlD* genes from various taxa -1 frameshifts have been observed more frequently than $+1$ frameshifts [16]. The recoding signals in these genes usually consisted of AT-rich “slippery sites” together with a stimulatory RNA secondary structures located downstream. More info about the frameshifting mechanism can be found elsewhere [27]. Due to the stochastic nature of the programmed frameshifting, ribosomes reading the same mRNA can produce two different proteins that allows cotranslational assembly of protein complexes. Consequently, it was likely that the two proteins produced from *fs-chlD* mRNA interacted with each other—similarly to the CobS and CobT subunits [28]. However, the function of this protein complex was unclear since the genomes of the corresponding bacteria completely lacked the CHL and B12 pathways as well as the large chelatase subunit gene (either *chlH* or *cobN*). Together, these observations indicated that the *S. thermophyla* and *B. fermentans* could be living prokaryotic fossils and their genotypes can be considered as the starting point in the evolution of the CHL and B12 pathways.

It has previously been reported that the *chlI* gene encoding the small chelatase subunit was homologous to the N-terminal part of the *chlD* gene [29]. Therefore, the next putative evolutionary event was duplication of the N-terminal part of the *fs-chlD* gene and formation of a separate *chlI* gene. In this case, the frameshifting signal in the *fs-chlD* was likely to disappear since there was no need to produce both proteins from a single gene. Accordingly, we found another prokaryotic group that had *chlD* and *chlI* genes but still lacked the CHL and B12 pathways (Table 1, stage 2). Interestingly, the ability of the cotranslational interaction between the products of these proto-*chlI* and proto-*chlD* genes was possible since they were located next to each other on the genome and can be present in the same polycistronic mRNA. Although the function of the proteins produced from these two genes was still unclear, the corresponding genotypes can be considered as the next stage in the early evolution of CHL biosynthesis. Moreover, we identified several Chloroflexi species (from the Anaerolineaceae group) with a potential intermediate genotype between Stage 1 and 2. Namely, *Anaerolinea thermophila* UNI-1, *Ornatilinea apprima*, *Longilinea arvooryzae*, and *Levilinea saccharolytica* possessed the *fs-chlD* genes with characteristic frameshifting signals but also had a separate *chlI* gene. Interestingly, these species were also strict anaerobes and thermophiles suggesting their ancient origin [30]. Thus, there were a number of living prokaryotes that provided insights into the early evolution of the modern Mg-chelatase enzyme.

A good representative case of the stage 3 was photosynthetic proteobacteria *Polynucleobacter duraquae* that had the *chlH* gene encoding the large subunit of the Mg-chelatase (Table 1, stage 3). Similar to the genomes from stage 2, all three chelatase subunit genes were located at the same genomic locus, possibly allowing cotraslational assembly of the Mg-chelatase from a polycystronic mRNA. According to the KEGG database, *P. duraquae* possessed all the major genes required for CHL biosynthesis. It was likely that this bacteria was only capable of anoxygenic photosynthesis [31] supporting its relatively early position in our proposed evolutionary system.

Table 1. The possible stages of the evolution of the genes encoding Mg-/Co-chelataes as well as CHL and B12 pathways.

Stage	Example Species (Phyla)	Synthesis of the Medium and Small Proteins	Function of the Medium and Small Proteins	B12 Pathway (Anaerobic)	Large Subunit (chlH/cobN)	CHL Pathway	B12 Pathway (Aerobic)
1	<i>Spirochaeta thermophila</i> (Spirochaetes) <i>Brevefilum fermentans</i> (Chloroflexi)		Unknown	-	-	-	-
2	<i>Thermodesulfobacterium commune</i> (Thermodesulfobacteria) <i>Thermodesulfobivrio yellowstonii</i> (Nitrospirae)		Unknown	-	-	-	-
3	<i>Polynucleobacter duraquae</i> (Proteobacteria)		Mg-chelatae (small and medium subunits)	-		Yes	-
4	<i>Halorhodospira halophila</i> (Proteobacteria) <i>Thiocystis violascens</i> (Proteobacteria)		Mg-(Co-?)chelatae (small and medium subunits)	Yes		Yes	Yes?
5	<i>Synechococcus elongatus</i> (Cyanobacteria) <i>Prochlorococcus marinus</i> (Cyanobacteria)		Mg- and Co- chelataes (small and medium subunits)	Yes		Yes	Yes
6	<i>Rhodospirillum rubrum</i> (Proteobacteria) <i>Dinoroseobacter shibae</i> (Proteobacteria) <i>Rhodovulum sulfidophilum</i> (Proteobacteria)		Mg- and Co- chelataes (small and medium subunits)	Yes		Yes	Yes
7	<i>Mycobacterium tuberculosis</i> (Actinobacteria) <i>Prauserella marina</i> (Actinobacteria)		Small and medium subunits of Co- chelataes	Yes		-	Yes
8	<i>Pseudomonas aeruginosa</i> (Proteobacteria) <i>Methanocaldococcus fervens</i> (Archaea)		Small and medium subunits of Co- chelataes	Yes		-	Yes

At stage 4, we placed anoxygenic proteobacteria [32] that in addition to chlorophyll biosynthesis also had the B12 pathway (Table 1, stage 4). These bacteria possessed anaerobic B12 pathway as well as many genes from the aerobic route. However, it was unclear if they were able to synthesize cobalamin using oxygen since these species were anaerobic. Additionally, their genome did not encode the large subunit of the aerobic Co-chelatase (the *cobN* gene). Thus, if the B12 biosynthesis was possible through the aerobic route, then the Mg-chelatase from these species might be able to function as a Co-chelatase as well.

Modern Cyanobacteria (such as *Synechococcus elongatus*) were at the next stage in our model because they were able to perform oxygenic photosynthesis, encoded large subunits for both Mg- and Co-chelatases, and could produce cobalamin via anaerobic and aerobic routes (Table 1, stage 5). Interestingly, Cyanobacteria did not have the *cobS* and *cobT* genes that encode the small and the medium subunits of the cobNST chelatase. It has been suggested that the corresponding subunits from the Mg-chelatase (*chlI* and *chlD*) could substitute them [14].

A number of oxygenic phototrophic Proteobacteria had the most advanced genotype with respect to the Mg- and Co-chelatases (Table 1, stage 6). In comparison to Cyanobacteria, their genomes contained *cobS* and *cobT* genes that were likely duplicated from the *chlI* and *chlD* and subsequently specialized to function as subunits of the cobNST chelatase. Since many of these bacteria (e.g., *Rhodospirillum rubrum*) were able to live under both aerobic and anaerobic conditions, their genomes encoded both routes of B12 biosynthesis [33].

Starting from stage 1 to stage 6, we observed gradual development of the gene sets encoding chlIDH and cobNST chelatases. On the other hand, a number of modern heterotrophic Actinobacteria (such as *Mycobacterium tuberculosis*) did not have the CHL pathway but encoded both pathways of B12 biosynthesis (Table 1, stage 7). However, they utilized the *chlI* and *chlD* genes, rather than *cobN* and *cobS*, to encode the small and the medium subunits of their aerobic Co-chelatase. This suggested that their chelatase genotypes originated from the Cyanobacteria where *chlI* and *chlD* genes performed similar function. Thus, we hypothesized that Actinobacteria from stage 7 adopted some genes from the Cyanobacterial CHL pathway (stage 5) and lost the rest of the CHL biosynthesis genes. Finally, stage 8 represented further simplification of the chelatase genotype (Table 1, stage 8) where a frameshifting signal in the *fs-chlD* gene allowed the production of both the small and the medium subunits of the Co-chelatase for the aerobic route of B12 biosynthesis [16].

4. Discussion

In this paper, we presented a model for the possible coevolution of the CHL and B12 (aerobic route) pathways (Figure 2). To develop this model, we assumed that among the thousands of the sequenced prokaryotic genomes, there were examples of so-called “living fossils”, i.e., the direct descendants of the species from the early days of cellular life on Earth. Our analysis was based on the *chlD* gene due to its likely early appearance in the evolution [24]. We hypothesized that the frameshifted version of this gene was the starting point in the early evolution of CHL pathway. Namely, we identified this gene in several bacteria that could be considered as living fossils (stage 1). We also identified several other bacteria where a part of *chlD* gene was probably duplicated producing the proto-*chlI* gene (stage 2). Importantly, the stage 1 and 2 phyla (Spirochaetes, Chloroflexi, Thermodesulfobacteria, and Nitrospirae) that appeared in the evolution before the Proteobacteria from the stage 3 and 4 [26] supporting our suggested evolutionary order. Interestingly, species from stage 1 and 2 did not have B12 or CHL pathways as well as either large subunit genes. On the other hand, the *fs-chlD*, *chlI* and *chlD* genes from these species can produce two proteins that may interact with each other and form a separate complex [28]. To our knowledge, the function of the *chlD* + *chlI* complex (without the a subunit protein) has not been studied and its function is unknown. Thus, further studies of these genomes may reveal the new function for the corresponding proteins.

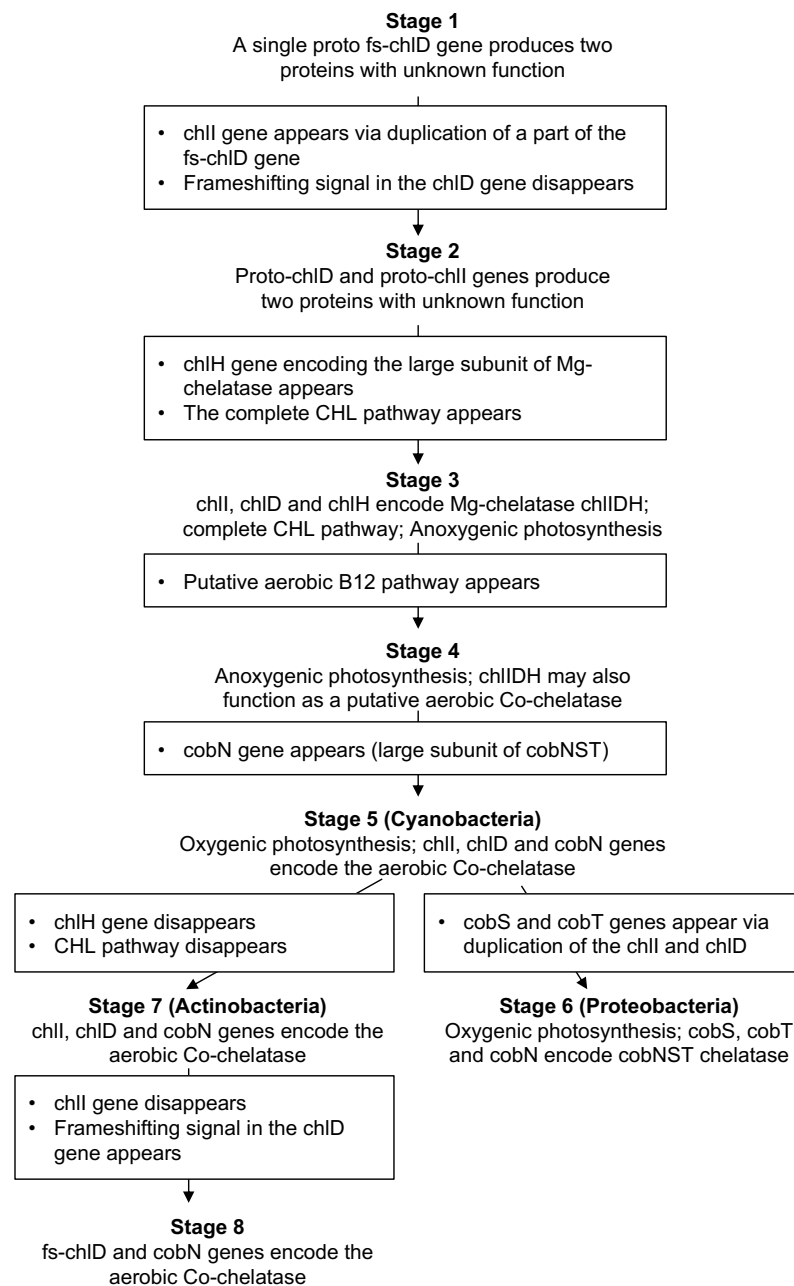


Figure 2. A putative chain of events that could happen between different stages during the evolution of the CHL and aerobic B12 pathways.

The appearance of the CHL and the aerobic B12 pathways as well as oxygenic photosynthesis was observed in Cyanobacteria at stage 5 of our system. Clearly, the aerobic B12 pathway could only appear after the oxygenic photosynthesis; however, we were not able to find any oxygenic phototroph without the aerobic B12 pathway. Although we believe that such putative bacteria existed at some evolutionary point, they probably adopted the more efficient aerobic B12 pathway after its appearance and therefore could not be found today. Proteobacteria at stage 6 represented the most advanced chelatase genotypes where all the subunits of the chlIDH and cobNST enzymes were encoded by specialized genes. On the other hand, a possible reduction of the Cyanobacterial genomes could give rise to the modern Actinobacteria that utilized *chlI* and *chlD* genes to encode small and medium cobNST subunits. Indeed, according to the global phylogenetic tree, Actinobacteria had a common ancestor with Cyanobacteria which appeared in the evolution first [26]. Finally, we observed that some prokaryotes utilized the ancient strategy to encode both the small

and medium subunits of the aerobic Co-chelatase using frameshifting signal in the *fs-chlD* gene (stage 8). The *chlD* genes with frameshifting signals in these genomes likely originated via horizontal gene transfer rather than the signals evolving once again in the conventional *chlD/bchD* genes. This is supported by the fact the *fs-chlD* genes do not group with the other genes without frameshifts (Supplementary Figure S3).

Another interesting question was related to the anaerobic cobalamin biosynthesis pathway—e.g., whether it appeared before or after the CHL biosynthesis pathway. Interestingly, the Chelatase DB included several archaeal species (such as *Methanobacterium paludis* and *Thermoplasma acidophilum*) that contained the *chlD* gene (without a frameshift) as well as the anaerobic B12 pathway. Again, the function of the *chlD* gene in these organisms was unclear, but the presence of the anaerobic B12 route may suggest that this pathway existed before the development of the CHL pathway.

To conclude, our analysis suggested a logical model for the coevolution of CHL and B12 pathways. We pinpointed specific prokaryotic species that can be considered as living fossils and their further study may shed a new light on the early evolution of these pathways. To our knowledge, the function of the *chlI* and *chlD* proteins without the large subunit (*chlH* or *cobN*) remained unexplored, and here, we suggested some model organisms where this could be studied.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/microorganisms10061200/s1>, Figure S1. The main strategies utilized by different prokaryotic species to encode Mg and Co-chelatases; Figure S2. The number of the identified *chlD* genes in the genomes from different prokaryotic phyla; Figure S3. The reference phylogenetic tree of the annotated medium chelatase subunit proteins and as well as the long products of the *fs-chlD* genes; Figure S4. Frameshifting signals identified in the *fs-chlD* genes from *Spirochaeta thermophila* DSM 6578 (A) and *Brevefilum fermentans* (B); Figure S5. A putative chain of events that could happen between different stages during the evolution of the CHL and aerobic B12 pathways.

Author Contributions: S.K., A.M. and I.A. formulated and evaluated the ideas and designed the website and the database structure. S.K. and A.M. conducted the research and implemented the Web database. I.A. supervised the research and wrote the original draft. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the Russian Science Foundation grant number 20-74-00128 to IA.

Informed Consent Statement: Not applicable.

Data Availability Statement: The website of the database with all the information about 1207 prokaryotic genomes with the *chlD*, *bchD*, and *cobT* genes together with the corresponding information about the predicted frameshifts and frameshifting signals are available at <http://ivanya.com/chelatase/> (accessed on 1 April 2022).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

B12	vitamin B12 or cobalamin (contains Co ²⁺)
BCHL	bacteriochlorophyll (contains Mg ²⁺)
CHL	chlorophyll (contains Mg ²⁺)
F430	coenzyme F430 (contains Ni ²⁺)
<i>fs-chlD</i> gene	frameshifted <i>chlD</i> gene
Proto	Protoporphyrin IX
Uro	Uroporphyrinogen III

References

1. Blankenship, R.E.; Hartman, H. The origin and evolution of oxygenic photosynthesis. *Trends Biochem. Sci.* **1998**, *23*, 94–97. [[CrossRef](#)]
2. Blankenship, R.E. Early evolution of photosynthesis. *Plant Physiol.* **2010**, *154*, 434–438. [[CrossRef](#)]
3. Willows, R.; Hansson, M. Mechanism, structure, and regulation of magnesium chelatase. *Porphy. Handb. II* **2003**, *113*, 1–48.
4. Thweatt, J.L.; Canniffe, D.P.; Bryant, D.A. Biosynthesis of chlorophylls and bacteriochlorophylls in green bacteria. *Adv. Bot. Res.* **2019**, *90*, 35–89.
5. Guillard, R.; Kadish, K.M.; Smith, K.M.; Guillard, R. *The Porphyrin Handbook*; Academic Press: New York, NY, USA, 2003; Volume 18.
6. Walker, C.J.; Willows, R.D. Mechanism and regulation of Mg-chelatase. *Biochem. J.* **1997**, *327 Pt 2*, 321–333. [[CrossRef](#)]
7. Bollivar, D.W.; Suzuki, J.Y.; Beatty, J.T.; Dobrowolski, J.M.; Bauer, C.E. Directed mutational analysis of bacteriochlorophyll a biosynthesis in *Rhodobacter capsulatus*. *J. Mol. Biol.* **1994**, *237*, 622–640. [[CrossRef](#)]
8. Gibson, L.; Willows, R.D.; Kannangara, C.G.; von Wettstein, D.; Hunter, C.N. Magnesium-protoporphyrin chelatase of *Rhodobacter sphaeroides*: Reconstitution of activity by combining the products of the bchH, I, and-D genes expressed in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 1941–1944. [[CrossRef](#)]
9. Fang, H.; Kang, J.; Zhang, D. Microbial production of vitamin B12: A review and future perspectives. *Microb. Cell Factories* **2017**, *16*, 1–14. [[CrossRef](#)]
10. Crespo, A.; Blanco-Cabra, N.; Torrents, E. Aerobic vitamin B12 biosynthesis is essential for *Pseudomonas aeruginosa* class II ribonucleotide reductase activity during planktonic and biofilm growth. *Front. Microbiol.* **2018**, *9*, 986. [[CrossRef](#)]
11. Fodje, M.; Hansson, A.; Hansson, M.; Olsen, J.; Gough, S.; Willows, R.; Al-Karadaghi, S. Interplay between an AAA module and an integrin I domain may regulate the function of magnesium chelatase. *J. Mol. Biol.* **2001**, *311*, 111–122. [[CrossRef](#)]
12. Escalante-Semerena, J.C.; Warren, M.J. Biosynthesis and use of cobalamin (B12). *EcoSal Plus* **2008**, *3*. [[CrossRef](#)]
13. Lundqvist, J.; Elmlund, D.; Heldt, D.; Deery, E.; Söderberg, C.A.; Hansson, M.; Warren, M.; Al-Karadaghi, S. The AAA+ motor complex of subunits CobS and CobT of cobaltochelatase visualized by single particle electron microscopy. *J. Struct. Biol.* **2009**, *167*, 227–234. [[CrossRef](#)]
14. Rodionov, D.A.; Vitreschak, A.G.; Mironov, A.A.; Gelfand, M.S. Comparative genomics of the vitamin B12 metabolism and regulation in prokaryotes. *J. Biol. Chem.* **2003**, *278*, 41148–41159. [[CrossRef](#)]
15. Antonov, I.; Coakley, A.; Atkins, J.F.; Baranov, P.V.; Borodovsky, M. Identification of the nature of reading frame transitions observed in prokaryotic genomes. *Nucleic Acids Res.* **2013**, *41*, 6514–6530. [[CrossRef](#)]
16. Antonov, I.V. Two Cobalt Chelatase Subunits Can Be Generated from a Single chlD Gene via Programed Frameshifting. *Mol. Biol. Evol.* **2020**, *37*, 2268–2278. [[CrossRef](#)]
17. Dailey, H.A. Illuminating the black box of B12 biosynthesis. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 14823–14824. [[CrossRef](#)]
18. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [[CrossRef](#)]
19. Kanehisa, M.; Araki, M.; Goto, S.; Hattori, M.; Hirakawa, M.; Itoh, M.; Katayama, T.; Kawashima, S.; Okuda, S.; Tokimatsu, T.; et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **2007**, *36*, D480–D484. [[CrossRef](#)]
20. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [[CrossRef](#)]
21. Paradis, E.; Schliep, K. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **2018**, *35*, 526–528. [[CrossRef](#)]
22. Lorenz, R.; Bernhart, S.H.; Zu Siederdisen, C.H.; Tafer, H.; Flamm, C.; Stadler, P.F.; Hofacker, I.L. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **2011**, *6*, 26. [[CrossRef](#)]
23. Debussche, L.; Couder, M.; Thibaut, D.; Cameron, B.; Crouzet, J.; Blanche, F. Assay, purification, and characterization of cobaltochelatase, a unique complex enzyme catalyzing cobalt insertion in hydrogenobyrinic acid a, c-diamide during coenzyme B12 biosynthesis in *Pseudomonas denitrificans*. *J. Bacteriol.* **1992**, *174*, 7445–7451. [[CrossRef](#)]
24. Weiss, M.C.; Sousa, F.L.; Mrnjavac, N.; Neukirchen, S.; Roettger, M.; Nelson-Sathi, S.; Martin, W.F. The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* **2016**, *1*, 16116. [[CrossRef](#)]
25. Angelov, A.; Liebl, S.; Ballschmiter, M.; Bomeke, M.; Lehmann, R.; Liesegang, H.; Daniel, R.; Liebl, W. Genome sequence of the polysaccharide-degrading, thermophilic anaerobe *Spirochaeta thermophila* DSM 6192. *J. Bacteriol.* **2010**, *192*, 6492–6493. [[CrossRef](#)]
26. Zhu, Q.; Mai, U.; Pfeiffer, W.; Janssen, S.; Asnicar, F.; Sanders, J.G.; Belda-Ferre, P.; Al-Ghalith, G.A.; Kopylova, E.; McDonald, D.; et al. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. *Nat. Commun.* **2019**, *10*, 1–14. [[CrossRef](#)]
27. Atkins, J.F.; Loughran, G.; Bhatt, P.R.; Firth, A.E.; Baranov, P.V. Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. *Nucleic Acids Res.* **2016**, *44*, 7007–7078. [[CrossRef](#)]
28. Cameron, B.; Guilhot, C.; Blanche, F.; Cauchois, L.; Rouyez, M.; Rigault, S.; Levy-Schil, S.; Crouzet, J. Genetic and sequence analyses of a *Pseudomonas denitrificans* DNA fragment containing two cob genes. *J. Bacteriol.* **1991**, *173*, 6058–6065. [[CrossRef](#)]
29. Sousa, F.L.; Shavit-Grievink, L.; Allen, J.F.; Martin, W.F. Chlorophyll biosynthesis gene evolution indicates photosystem gene duplication, not photosystem merger, at the origin of oxygenic photosynthesis. *Genome Biol. Evol.* **2013**, *5*, 200–216. [[CrossRef](#)]

30. Sekiguchi, Y.; Yamada, T.; Hanada, S.; Ohashi, A.; Harada, H.; Kamagata, Y. *Anaerolinea thermophila* gen. nov., sp. nov. and *Caldilinea aerophila* gen. nov., sp. nov., novel filamentous thermophiles that represent a previously uncultured lineage of the domain Bacteria at the subphylum level. *Int. J. Syst. Evol. Microbiol.* **2003**, *53*, 1843–1851. [[CrossRef](#)]
31. Hahn, M.W.; Schmidt, J.; Pitt, A.; Taipale, S.J.; Lang, E. Reclassification of four Polynucleobacter necessarius strains as representatives of Polynucleobacter asymbioticus comb. nov., Polynucleobacter duraquae sp. nov., Polynucleobacter yangtzensis sp. nov. and Polynucleobacter sinensis sp. nov., and emended description of Polynucleobacter necessarius. *Int. J. Syst. Evol. Microbiol.* **2016**, *66*, 2883.
32. Challacombe, J.F.; Majid, S.; Deole, R.; Brettin, T.S.; Bruce, D.; Delano, S.F.; Detter, J.C.; Gleasner, C.D.; Han, C.S.; Misra, M.; et al. Complete genome sequence of Halorhodospira halophila SL1. *Stand. Genom. Sci.* **2013**, *8*, 206–214. [[CrossRef](#)]
33. Schultz, J.; Weaver, P. Fermentation and anaerobic respiration by *Rhodospirillum rubrum* and *Rhodopseudomonas capsulata*. *J. Bacteriol.* **1982**, *149*, 181–190. [[CrossRef](#)]