A novel bifunctional transcriptional regulator of riboflavin metabolism in Archaea

Irina A. Rodionova[1], Matthew W. Vetting[2], Xiaoqing Li[1], Steven C. Almo[2], Andrei L. Osterman[1] and Dmitry A. Rodionov[1,3,*]

[1]Sanford-Burnham-Prebys Medical Discovery Institute, La Jolla, CA 92037, USA, [2]Department of Biochemistry, Albert Einstein College of Medicine, Bronx, NY 10461, USA and [3]A.A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, 127051 Russia

ABSTRACT

Riboflavin (vitamin B2) is the precursor of flavin mononucleotide (FMN) and flavin adenine dinucleotide, which are essential coenzymes in all free-living organisms. Riboflavin biosynthesis in many Bacteria but not in Archaea is controlled by FMN-responsive riboswitches. We identified a novel bifunctional riboflavin kinase/regulator (RbkR), which controls riboflavin biosynthesis and transport genes in major lineages of Crenarchaeota, Euryarchaeota and Thaumarchaeota. RbkR proteins are composed of the riboflavin kinase domain and a DNA-binding winged helix-turn-helix-like domain. Using comparative genomics, we predicted RbkR operator sites and reconstructed RbkR regulons in 94 archaeal genomes. While the identified RbkR operators showed significant variability between archaeal lineages, the conserved core of RbkR regulons includes riboflavin biosynthesis genes, known/predicted vitamin uptake transporters and the *rbkR* gene. The DNA motifs and CTP-dependent riboflavin kinase activity of two RbkR proteins were experimentally validated *in vitro*. The DNA binding activity of RbkR was stimulated by CTP and suppressed by FMN, a product of riboflavin kinase. The crystallographic structure of RbkR from *Thermoplasma acidophilum* was determined in complex with CTP and its DNA operator revealing key residues for operator and ligand recognition. Overall, this study contributes to our understanding of metabolic and regulatory networks for vitamin homeostasis in Archaea.

INTRODUCTION

Vitamin B_2 (riboflavin) is a precursor to the coenzymes flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN), which are essential components of basic metabolism ([1]). Many characterized micro-organisms including Bacteria and Archaea, as well as plants and fungi, synthesize riboflavin using *de novo* biosynthetic pathways; however, many other microorganisms (and mammals, including human) are B_2 auxotrophs that salvage exogenous riboflavin ([2–5]).

The riboflavin biosynthesis pathway (RBP) includes a set of committed enzymes that produce one riboflavin molecule from one molecule of Guanosine triphosphate (GTP) and two molecules of ribulose 5-phosphate (Figure [1]) ([6]). In Bacteria, these enzymes include RibA, RibB, the bifunctional deaminase/reductase RibD, as well as RibH and RibE, involved in the final two steps of RBP (Table [1]). The archaeal RBP, which has been described in *Methanocaldococcus jannaschii*, involves several alternative enzymatic steps and non-orthologous gene displacements and is characterized by a reverse order of the deamination and reduction reactions. In *M. jannaschii*, ArfA hydrolyzes the imidazole ring of GTP but generates a different intermediate product that undergoes further hydrolysis by ArfB and reduction by ArfC ([7–9]). The ArfC reductase is orthologous to a reductase domain of bacterial RibD, whereas ArfA and ArfB are archaeal-specific enzymes. Orthologs of a pyrimidine deaminase domain of RibD are apparently missing in Archaea. A candidate for an archaeal-specific pyrimidine deaminase (termed PyrD) was proposed based on its frequent chromosomal clustering and phylogenomic co-occurrence with other RBP genes ([10]), however this functional prediction remains to be tested. In contrast to the RibB and RibH enzymes that have recognizable orthologs in archaeal genomes, RibC belong to a unique class of pentameric riboflavin synthases of Archaea that are non-homologous to bacterial RibE enzymes ([11]).

Downstream cofactor biosynthetic enzymes catalyzing phosphorylation of riboflavin to yield FMN and its subsequent adenylylation to FAD in Archaea, are non-orthologous to their bacterial counterparts (Figure [1]). Bac-
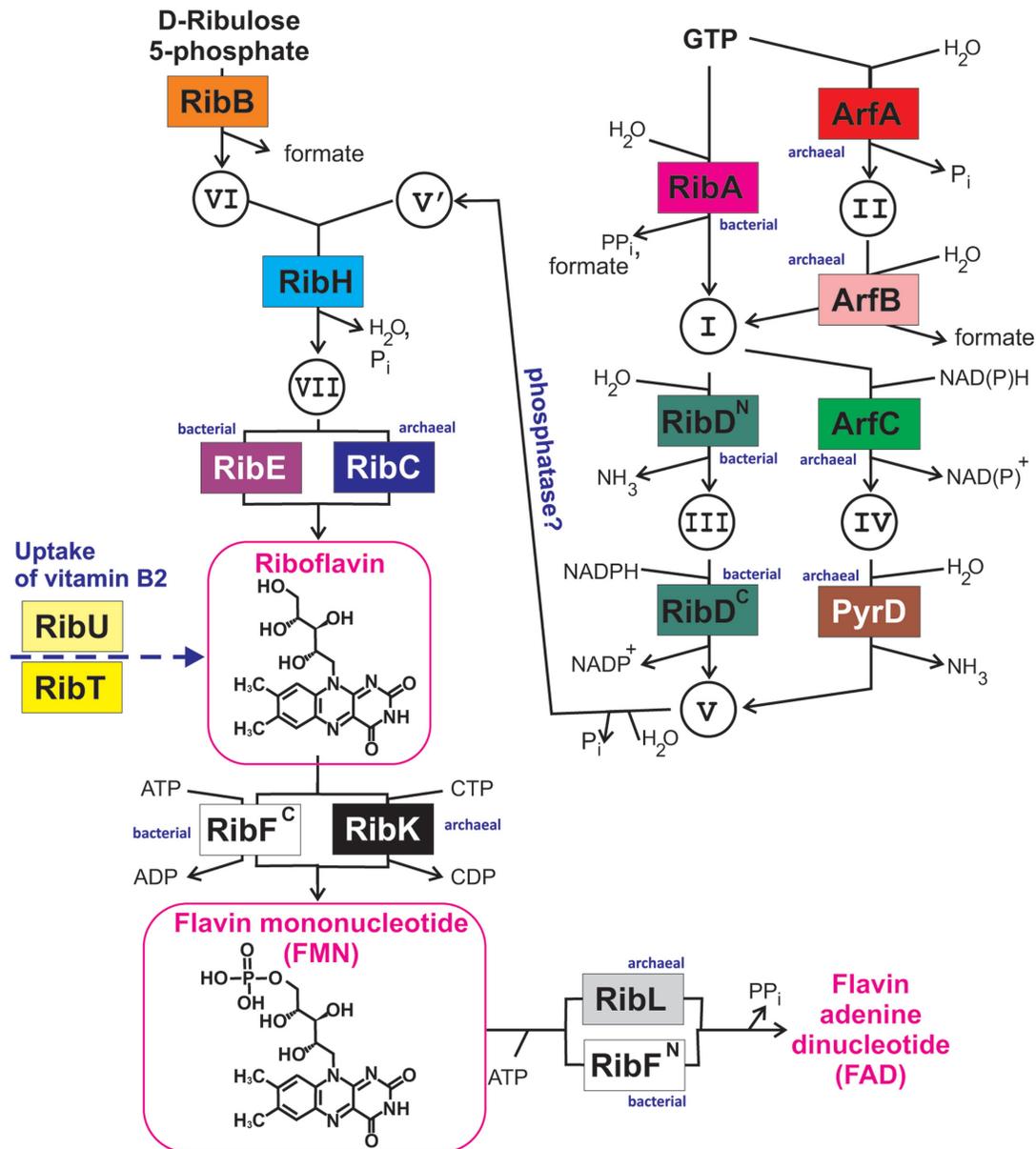
Figure 1. Pathways of riboflavin, FMN and FAD biosynthesis in Archaea and Bacteria. Alternative 'archaeal' and 'bacterial' enzymes are shown. Boxes in shades of grey show enzymes/transporters that were identified in at least one studied in this work archaeal species. White boxes denote bacterial enzymes that lack orthologs in any archaeal genome. Metabolic pathway intermediates: I, 2,5-diamino-6-(5-phospho-D-ribosylamino)pyrimidin-4(3H)-one; II, 2-amino-5-formylamino-6-(5-phospho-D-ribosylamino)pyrimidin-4(3H)-one; III, 5-amino-6-(5-phosphoribosylamino)uracil; IV, 2,5-diamino-6-(5-phospho-D-ribosylamino)pyrimidin-4(3H)-one; V, 5-amino-6-(D-ribitylamino)uracil; V', 5-amino-6-(D-ribitylamino)uracil; VI, 3,4-dihydroxy-2-butanone 4-phosphate; VII, 6,7-dimethyl-8-(D-ribityl)lumazine. Bacterial protein/gene names are provided as in *Escherichia coli* (note that in some other bacteria such as *Bacillus subtilis* some of these genes were historically named differently). Superscript 'N' and 'C' letters denote N- and C-terminal domains in the bifunctional RibD and RibF enzymes. The standard *Methanocaldococcus jannaschii* protein/gene names are used for archaeal enzymes. Functional roles of riboflavin biosynthetic enzymes are listed in Table 1. The RibU transporter from the ECF family was previously characterized in a number of Gram-positive bacteria. The RibT transporter from the MFS family is predicted in this work as a part of RbkR regulons in Archaea.

teria utilize a single bifunctional enzyme containing the adenosine triphosphate-dependent riboflavin kinase and the FMN adenylyltransferase domains (e.g. RibF in *Escherichia coli*) (12). In contrast, two individual and structurally distinct enzymes, Cytidine triphosphate (CTP)-dependent riboflavin kinase (RibK) and FAD synthetase (RibL), catalyze the last two steps of the pathway in *M.*

jannaschii (13–15) and their orthologs are present in all Archaea (but not in Bacteria or Eukaryota).

The RBP genes in Bacteria are controlled by *cis*-regulatory RNAs (the so-called *RFN* elements) that were first discovered by comparative genomics (4) and subsequently attributed to a class of FMN-sensing riboswitches that govern gene expression through formation of alternative leader mRNA structures (16–18). FMN riboswitches

Table 1. Functional roles of enzymes and transporters involved in flavin metabolism

Protein name	Functional role of protein (Enzyme classification #) [Protein family]	RbkR regulon[a]
ArfA	GTP cyclohydrolase III (EC 3.5.4.29), archaeal	25
ArfB	2-amino-5-formylamino-6-ribosylaminopyrimidin-4(3H)-one 5′-monophosphate deformylase (EC 3.5.1.102), archaeal	8
RibA	GTP cyclohydrolase II (EC 3.5.4.25)	12
RibB	3,4-dihydroxy-2-butanone 4-phosphate synthase (EC 4.1.99.12)	80
ArfC	5-amino-6-(5-phosphoribosylamino)uracil reductase (EC 1.1.1.193), archaeal	17
PyrD	Pyrimidine deaminase (EC 3.5.4.26), archaeal	15
RibD	diaminohydroxyphosphoribosylaminopyrimidine deaminase (EC 3.5.4.26) / 5-amino-6-(5-phosphoribosylamino)uracil reductase (EC 1.1.1.193)	8
RibH	6,7-dimethyl-8-ribityllumazine synthase (EC 2.5.1.78)	27
RibE	Riboflavin synthase (EC 2.5.1.9) [PF00677]	17
RibC	Riboflavin synthase (EC 2.5.1.9) [PF00885], archaeal	13
RbkR	CTP-dependent riboflavin kinase (EC 2.7.1.161) / Transcriptional regulator of riboflavin metabolism [PF12802], archaeal	36
RibK	CTP-dependent riboflavin kinase (EC 2.7.1.161), archaeal	
RibL	FAD synthetase (EC 2.7.7.2), archaeal	3
RibF	ATP-dependent riboflavin kinase (EC 2.7.1.26) / FAD synthetase (EC 2.7.7.2), bacterial	
RibU	Riboflavin transporter, ECF family [COG3601]	18
RibT	Predicted riboflavin transporter, MFS family [PF00087], archaeal	13

[a]Number of genomes in the analysed archaeal taxa where the target gene is preceded by a predicted RbkR-binding site.

provide a negative feedback loop for the control of RBP genes that are often clustered into *rib* loci in bacterial genomes. In *E. coli* and related proteobacteria, only a single RBP gene, *ribB*, is subject to FMN riboswitch-dependent regulation, whereas in *Bacillus subtilis* and other B$_2$-prototrophic Firmicutes the entire *ribDE(B/A)H* transcriptional unit is preceded by a FMN riboswitch (19). Analysis of distribution of FMN riboswitches in bacterial genomes allowed prediction of multiple families of putative riboflavin transporters, many of which were experimentally confirmed to contribute to B$_2$ uptake (2–4,20–22). These include the B$_2$-specific permease component RibU from the energy coupling factor (ECF) transporter family (23). RibU and other permease components of ECF transporters utilize the shared energizing module composed of the two adenosine triphosphatase subunits and the anchor membrane subunit (24).

Previous genomic analyses of FMN riboswitches confirmed that their occurrence is restricted to Bacteria (4,19,25) and the mechanisms of transcriptional regulation of RBP genes in Archaea have not been described. Our earlier genomic searches suggested at least one regulatory mechanism for some archaeal genomes that contain a version of the RibK protein fused with an N-terminal winged helix-turn-helix (wHTH) domain (10). Here we analyzed putative regulons controlled by these bifunctional riboflavin kinase/regulator proteins, termed RbkR, using the comparative genomics approach, which was previously successfully applied for large-scale reconstructions of transcriptional regulons in Bacteria (26–28) and Archaea (29,30). The detailed reconstruction of RbkR regulons in archaeal genomes revealed significant variation in the inferred regulon content and RbkR-binding DNA motifs between archaeal lineages. To evaluate the accuracy of the genomic reconstruction, we experimentally evaluated the RbkR regulators in representative species from three diverse lineages

of Archaea. *In vitro* binding assays with purified RbkR proteins confirmed their predicted DNA operators and revealed that CTP enhanced the formation of the RbkR–DNA complex, while FMN demonstrated a negative effect. We also determined the structure of RbkR from *Thermoplasma acidophilum* in complex with its specific DNA operator and CTP providing new insights into the mechanism of DNA and ligand recognition. Combining bioinformatics-based observations with these experimental data allowed for the construction of mechanism of the RbkR-mediated transcriptional regulation of riboflavin metabolism in Archaea.

MATERIALS AND METHODS

Genomic resources and bioinformatics tools

The analyzed archaeal genomes were downloaded from Genbank (31). Genomes of two Yellowstone National Park isolates, *Pyrobaculum yellowstonensis* WP30 (32) and *Metallosphaera yellowstonensis* MK1 (33), were downloaded from the IMG ER database at Joint Genomic Institute (34). Orthologs were identified as bidirectional best hits using protein BLAST (with default search parameters) and the Ortholog search tool in the Genome Explorer software (35) and additionally confirmed via construction of phylogenetic trees using PhyML with 100 bootstrap replicates (36). To perform reconstruction of the RBPs in archaeal genomes we utilized the subsystem-based comparative genomics approach (10) and the set of experimentally characterized enzymes and transporters from Bacteria and Archaea (Table 1). The gene neighborhood analysis was performed in the IMG ER, PubSEED and MicrobesOnline databases (34,37,38). The closely-related gene orthologs with experimentally determined function were identified via BLAST searches (with default parameters) in UniProtKB/SwissProt (39). The taxonomic distribution of

FMN riboswitches was taken from the Rfam database for the RF00050 family (40).

For identification of candidate RbkR-binding site motifs and regulon reconstructions, we used a previously established comparative genomics approach (41) implemented in the Genome Explorer software (35) and the RegPredict Web server tool (42). We started regulon reconstruction from collection of initial training sets of known genes involved in RBP and/or vitamin B_2 transport in each archaeal lineage containing a RbkR ortholog (Supplementary Table S1). For each training set of genes, we determined orthologs in related genomes, analyzed the potential transcriptional units and collected their upstream DNA regions up to 300 nucleotides upstream of the translation start site (excluding the coding regions of any upstream gene if the intergenic region was <300 nt). The prepared training sets for each RbkR-containing archaeal lineage were used for a search of lineage-specific conserved DNA motifs.

The Discover Profile tool in RegPredict uses the expectation-maximization methods and the palindromic or direct repeat symmetry of motifs, a feature that is typical of many prokaryotic transcription factors that interact with their binding sites as symmetric dimers (43). The algorithm identifies all weak palindromes (direct repeats) in an input group of sequences, iteratively clusters them to convergence and outputs several putative motifs ranked by their average positional information content. A search for palindromic DNA motifs of 16–20 bp of both odd and even length was carried out using the following parameters: min number of GC pair = 0; number of palindromic positions = 4, size of training set = 50. Motifs were further validated by the construction of multiple alignments of orthologous DNA fragments using MUSCLE (44) (Supplementary Figure S1). Conservative palindromic sites were selected as potential binding sites for RbkR and used for construction of positional weight matrices (PWMs) which describe the probability of each possible letter at each position in the pattern. Positional nucleotide weights in PWMs and site scores were defined as previously described (43). Further PWM-based searches for additional sites identified a second copy of putative RbkR site in majority of DNA upstream regions from the initial training set in the *Halobacteria* and *Methanomicrobia* lineages as well as in the *Metallosphaera* and *Pyrobaculum* spp. The tandem RbkR sites were added to the final training sets used for PWM construction.

The constructed PWMs were used to search for additional binding sites and regulon members in the analysed genomes using the Genome Explorer tool. The genome scan parameters were set up to reduce the chance of nonfunctioning sequences from being detected. Specifically, positions of candidate regulatory sites were set between 300 nt upstream and 25 nt downstream of a gene start codon. The score thresholds for site searches were selected as the lowest site score in the training set (Supplementary Table S2 in the supplemental material). Weaker sites (with scores 15% less than the threshold) were also taken into account if their positions were similar to positions of stronger sites upstream of orthologous genes. False-positive site predictions were eliminated using the consistency check approach (43). New candidate members were attributed to the regulon if they were preceded by candidate site in more than 50% genomes in each taxonomic group. The reconstructed regulons were extended to include all genes in putative transcriptional units. Genes were considered to belong to a transcriptional unit if they were transcribed in the same direction and, with intergenic distances not exceeding 50 nt. Sequence logos for DNA motifs were drawn using the WebLogo package (45). All reconstructed regulons including predicted sites, regulators and regulated genes are available in Supplementary Tables S1 and S2.

Gene cloning and protein purification

The *rbkR* genes from *M. yellowstonensis* MK1 and *P. yellowstonensis* WP30 (Genbank accessions EHP68448.1 and AKT35049.1) were synthesized by GenScript Inc. with optimized codons for expression in *E. coli*. The synthesized gene fragments were cloned into the pSMT3 expression vector under T7 promoter (46). Each obtained vector encodes a fusion between the target protein and N-terminal hexa-histidine Smt3 polypeptide (a yeast SUMO ortholog), which enhances protein solubility. The resulting constructs were transformed into *E. coli* BL21/DE3. Cells were grown in LB medium (1 l) at 37°C to an OD_{600} of ~0.8 and protein expression was induced with 0.2 mM isopropyl-β-D-thiogalactopyranoside and harvested after shaking at 20°C overnight. The recombinant RbkR proteins were purified to homogeneity using Ni^{2+} chelation chromatography. The harvested cells were resuspended in 20 mM Hepes (pH 7) containing 100 mM NaCl, 0.03% Brij-35, 2 mM β-mercaptoethanol and 2 mM phenylmethylsulfonyl fluoride (Sigma-Aldrich). Cells were lysed by incubation with lysozyme (1 mg/ml) for 30 min at 4°C, followed by a freeze-thaw cycle and sonication. For purification of the soluble fraction after centrifugation, Tris-HCl Buffer (pH 8) was added to the supernatant (50 mM), which was loaded onto Ni-nitrilotriacetic acid agarose (Qiagen) minicolumn. After washing with starting buffer containing 1 M NaCl and 0.3% Brij-35, the bound proteins were eluted with 250 mM imidazole in the same buffer. The protein concentration was determined by Quick Start Bradford Protein Assay Kit from Bio-Rad. Both RbkR proteins were obtained with high purity (>90% as determined by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)). The His_6-Smt3-tag was then cleaved from the purified RbkR proteins by digestion with Ulp1 protease at 4°C overnight.

The *rbkR* gene from *T. acidophilum* (Uniprot accession Q9HJA6) cloned into a pSGX3 vector was obtained from the New York SGX Research Center for Structural Genomics. The *T. acidophilum* RbkR protein (*ta*RbkR) was expressed with a C-terminal His_6-Tag. All growth media contained 200 μg/ml kanamycin and 35 μg/ml chloramphenicol. The vector was transformed into the Rosetta 2(DE3)pLysS Competent Cells (EMD Millipore). *ta*RbkR was expressed in 3 l of autoinduction media at 25°C, cells were sonicated in 200 ml of lysis buffer (50 mM Hepes pH 7.5, 150 mM NaCl, 20 mM imidazole, 0.5% Tween-20). The supernatant was applied to a 15 ml His60 Ni-Superflow column (Clonetech) equilibrated against lysis buffer (50 mM Hepes pH 7.5, 150 mM NaCl) supplemented with 20 mM imidazole and proteins were step-eluted with the same

buffer supplemented with 300 mM imidazole. The eluted protein was concentrated to 30 mg/ml and applied in 5 ml injections onto a size exclusion column (HiLoad 16/600 Superdex 200 prep grade, GE Healthcare). Fractions containing *ta*RbkR as observed by SDS were pooled; solid ammonium sulfate was added to 1.5 M; and the protein was applied to a 15 ml phenyl sepharose fast flow column (GE Healthcare) equilibrated against buffer containing 50 mM MES pH 5.5, 1.5 mM ammonium sulfate pH 5.5 and 5 mM Dithiothreitol (DTT). The column was washed with equilibration buffer until all yellow color (associated with bound FMN cofactor) was eluted from the column. Fractions with over 95% purity by SDS-PAGE analysis were pooled, buffer exchanged by dilution and centrifugal ultrafiltration into storage buffer (20 mM MES pH 6.0, 50 mM NaCl, 5 mM DTT), snap frozen in liquid N_2 and stored at $-80°C$.

DNA binding assays

Interaction between the purified recombinant RbkR proteins from the three archaeal species and their predicted DNA binding sites was assessed using two techniques: electrophoretic mobility shift assay (EMSA), and fluorescence polarization assay (FPA). The single stranded labeled and unlabeled DNA oligos were synthesized by Integrated DNA Technologies (IDT). The double-stranded DNA fragments were obtained by annealing synthesized oligonucleotides at a 1:10 ratio of 5′-labeled with 6-carboxyfluorescein (for FPA) or biotin (for EMSA) to unlabeled complementary oligonucleotides.

Using the FPA assay, we tested the *my*RbkR, *py*RbkR and *ta*RbkR proteins with five DNA fragments containing their corresponding binding sites in *M. yellowstonensis*, *P. yellowstonensis* and *T. acidophilum*, as well as with three control DNA fragments that lack RbkR sites (Supplementary Table S3). The obtained 6-carboxyfluorescein-labeled double-stranded DNA fragments (10 nM) were incubated at 24°C with the increasing concentrations of RbkR in a 100 μl reaction mixture in 96-well black plates (VWR, Radnor, PA) for 20 min. The binding buffer contained 50 mM Tris-acetate buffer (pH 6.0), 0.5 mM ethylenediaminetetraacetic acid, 10 mM $MgSO_4$, glycerol 2.5% and 50 mM KCl. Herring sperm DNA (1 μg) was added to the reaction mixture as a non-specific competitor DNA to suppress non-specific binding. The fluorescence-labeled DNA was detected with the FLA-5100 fluorescent image analyzer. The effect of CTP (100 μM) on *my*RbkR was tested by its addition to the incubation mixture. The effect of FMN (100 μM) on *my*RbkR was tested by pre-incubation with protein for 3 h at 24°C and its further removal by ultrafiltration, followed by incubation with DNA fragments. The effect of FMN, FAD and riboflavin (200 nM) on the fixed concentration of *py*RbkR (60 nM) was tested by their addition to the incubation mixture.

Using the EMSA assay, we tested interaction between the *py*RbkR protein and two 104-bp DNA fragments containing predicted binding sites upstream of the *arfA* and *ribB* genes from *P. yellowstonensis* (Supplementary Table S3). The biotin-labeled DNA fragments (0.2 nM) were incubated with increasing concentrations of *py*RbkR (2.5– 10 nM) in a total volume of 20 μl. The binding buffer contained 50 mM Tris-HCl (pH 8.0), 0.2 M NaCl, 5 mM $MgCl_2$, 1 mM DDT, 0.05% NP-40, 2.5% glycerol and 1 μg herring sperm DNA. After 25 min of incubation at 62°C, the reaction mixtures were separated by electrophoresis on a 1.5% agarose gel (85 min, 90 V, room temperature). The DNA was transferred by electrophoresis onto a Hybond-N^+ membrane and fixed by UV-cross-linking. The biotin-labeled DNA was detected with the LightShift chemiluminiscent EMSA kit (Thermo Fisher Scientific Inc, Rockford, IL, USA).

Riboflavin kinase assay

The recombinant *my*RbkR and *py*RbkR proteins were tested for CTP-dependent riboflavin kinase activity using the HPLC-based end-point assay with detection of FMN product by absorbance at 450 nm. Reaction mixture (0.2 ml) contained 0.15 M HEPES (pH 8), 0.1 M NaCl, 5 mM $MgCl_2$, 0.2 mM CTP and 0.6 μg of RbkR. Reaction was stopped after 20 min of incubation at 60°C. For K_m measurements, concentration of riboflavin was varied from 0.016 to 1.3 mM. The enzyme kinetics fits the Michaelis–Menten equation and K_m and V_{max} were calculated in Prism 6.

Protein crystallization and structure determination

Desalted single stranded 19-bp oligos (SS-Oligo 1, 5′-ATTACTAATTCACGAGTAA-3′; SS-Oligo 2, 5′-TTTACTCGTGAATTAGTAA) were synthesized by IDT. The SS-oligos (2.5 mM in 100 mM potassium acetate, 30 mM HEPES, pH 7.5) were annealed by heating to 94°C for 2 min, cooled at the range of 1°C per minute to 4°C and stored at $-80°C$. *ta*RbkR was crystallized by sitting drop vapor diffusion in 96-well Intelliplates (Art Robbins Instruments) at 18°C. Crystallization mixtures contained 0.5 μl of protein (25 mg ml^{-1}, 1.25 mM ds-DNA, 5 mM CTP) combined with 0.5 μl of reservoir (0.1 M phosphate citrate pH 4.2, 40% ethanol, 5% (w/v) PEG1000). Crystals were mounted on nylon loops and flash-cooled by plunging directly into liquid nitrogen. Diffraction data were collected on the Lilly Research Laboratories Collaborative Access Team (LRL-CAT) beamline at Sector 31 of the Advanced Photon Source (Eli Lilly Company) from single crystals at 100 K using a wavelength of 0.9793 Å and a Rayonix 225 HE detector. Data were integrated in MOSFLM (47) and scaled in AIMLESS (48). Phases were determined within MOLREP (49) using the unliganded structure of *ta*RbkR (PDB ID 3CTA) as a molecular replacement model. The model was improved with iterative cycles of model building with COOT (50) and refinement with PHENIX (51). The final refinement cycles included TLSD parameterization (52) with TLSD boundaries determined by PHENIX. Data collection and refinement statistics are given in Supplementary Table S4.

RESULTS AND DISCUSSION

Genomic reconstruction of RbkR regulons

Riboflavin kinase is an essential enzyme required for synthesis of FMN cofactor from vitamin B_2. The archaeal riboflavin kinase RibK is a CTP-dependent enzyme, which

was characterized in *M. jannaschii*, an archaeon from the *Methanococci* class (13,15). In contrast to the RibK proteins from the *Methanococci*, their orthologs in many other classes of Archaea contain an additional N-terminal wHTH domain from the MarR family of DNA-binding transcription factors (53). The identified chimeric proteins are the only orthologs of RibK in 96 out of 119 studied genomes that belong to three major archaeal phyla (Table 2). We proposed that the identified two-domain proteins are bifunctional riboflavin kinases and transcriptional regulators, termed RbkR. Within the *Crenarchaeota* phylum (the *Thermoprotei* class), RbkR is present in all lineages except the *Desulfurococcales* and *Acidilobales* that have the monofunctional RibK proteins. Among the *Euryarchaeota* phylum, RbkR is ubiquitous in most taxa except the *Methanococci*, *Methanobacteria* and *Methanopyri* classes that have RibK. In addition, all studied genomes from the *Thaumarchaeota* phylum encode the RbkR proteins.

In order to reconstruct putative RbkR regulons, we first identified orthologs of known RBP enzymes in archaeal genomes (Table 1 and Supplementary Table S1). The complete sets of RBP genes were identified in 97 genomes (B_2 prototrophs), while the remaining 22 genomes lack all or most RBP genes (B_2 auxotrophs) but often contain an ortholog of the riboflavin transporter *ribU* from Bacteria (Table 2). To find putative RbkR-binding motifs, we searched upstream regions of the identified RBP and *ribU* genes for conserved DNA motifs across related archaeal species in each lineage containing *rbkR* orthologs (Supplementary Table S2). The identified 16–20 bp palindromic DNA motifs were refined by construction of multiple alignments of orthologous upstream regions (Supplementary Figure S1). In most genomes from the *Halobacteria*, *Methanomicrobia* and *Thaumarchaeota* classes, the predicted RbkR operators contain two adjacent palindromic sites, whereas in most other lineages a single palindromic RbkR site was identified in each target regulatory region, suggesting cooperative binding of two RbkR dimers to adjacent DNA operators.

The candidate linage-specific RbkR motifs are highly variable between some taxonomic groups of Archaea (Figure 2 and Supplementary Figure S2). For instance, the motifs in the *Methanomicrobia*, *Halobacteria* and *Thermococci* have a common consensus (TnnCA-N_4-TGnnA), whereas the motifs in the *Thermoplasmata* and *Thaumarchaeota* are characterized by a different consensus (TTAC-N_{10}-GTAA). In most genera from the *Thermoproteales*, the RbkR motifs conform to the consensus AAACT-N_2-AGTTT, while the consensus motif in *Thermofilum* spp. is different (TAAAA-N_6-TTTTA). Finally, the predicted RbkR motifs in the *Sulfolobales* and *Archaeoglobales* share common consensus TnAG-N_6-CTnA, although these two taxa belong to different phyla.

We analyzed relative positions of the identified RbkR binding sites and known transcription start sites (TSSs) from previously reported transcriptomes of *Sulfolobus solfataricus* (54), *Pyrococcus furiosus* (55) and *Thermococcus kodakaraensis* (56) (Supplementary Table S2). The predicted RbkR sites in *S. solfataricus* are located one or 12 nt upstream of putative promoter elements that are linked to the known TSSs, whereas the distance between RbkR op-

erators and TSS-associated TATA boxes in *P. furiosus* and *T. kodakaraensis* is around 5 nt (Supplementary Figure S1). This analysis suggests that RbkR likely acts as a transcriptional activator of its target genes, however further experiments are required to confirm this inference.

The reconstructed regulons demonstrate large differences between the lineages in the sets of predicted RbkR-regulated genes (Figure 3). Most analyzed genomes contain one or two gene loci preceded by candidate RbkR-binding sites. All RbkR-controlled transcriptional units can be classified into two major groups. The largest group includes 117 target transcriptional units in 80 genomes encoding from one to five RBP enzymes per genome. These include: (i) RBP enzymes that are unique to Archaea; (ii) the alternative RBP enzymes that are characteristic of bacterial RBP; and (iii) enzymes that are universally conserved between Archaea and Bacteria (Table 1). In 36 genomes, the regulon includes the putative transcriptional unit *rbkR-ribB*, suggesting autoregulation of RbkR. In addition, the RbkR-regulated RBP loci in *Aciduliprofundum* and *Methanomassiliicoccus* include the FAD synthetase *ribL*. The second major group of 30 RbkR-controlled loci includes putative riboflavin uptake transporters of two types. The bacterial-type transporter *ribU* is ubiquitously regulated by RbkR in the genomes of *Thermococci*, *Thermofilum* spp. and *Caldivirga maquilingensis*. A putative MFS-type transporter (termed *ribT*) was found within the RbkR regulons in the *Sulfolobales* and *Thermoplasmatales*. The *ribT* gene is the only identified target of RbkR in the *Thermoplasma* species that are B_2 auxotrophs suggesting it encodes a new type of riboflavin transporter in Archaea.

Experimental assessment of dual function of RbkR proteins

In order to validate the computationally predicted RbkR-binding DNA motifs in the *Sulfolobales* and *Thermoproteales* orders, we obtained the recombinant RbkR proteins from *M. yellowstonensis* (*my*RbkR) and *P. yellowstonensis* (*py*RbkR). The purified ligand-free RbkR proteins were assessed using FPA DNA binding assays with synthetic DNA fragments containing the predicted RbkR-binding sites from the corresponding genomes (Figure 4).

Two synthetic fluorescence-labeled DNA fragments containing the predicted RbkR-binding sites upstream of the *ribB* and *ribT* genes in *M. yellowstonensis* were assessed for their specific interaction with *my*RbkR. The 60-bp DNA fragment of the *ribT* gene upstream region contains two tandem putative RbkR sites, while the 38-bp DNA fragment of the *ribB* promoter region represents a single RbkR site. Both DNA fragments demonstrated the *my*RbkR concentration-dependent binding by FPA (Figure 4A). Addition of 100 μM of CTP, which is a co-substrate of the archaeal riboflavin kinase RibK, improved the binding of *my*RbkR to its target DNA sites. The apparent affinity estimated as effective concentrations (EC_{50}) for *my*RbkR protein interacting with both DNA fragments in the presence of CTP were in the range of 100 to 110 nM. To investigate potential influence of FMN (a product of riboflavin kinase) on protein–DNA interaction, we incubated the ligand-free *my*RbkR protein (16 μM) with FMN (100 μM) for 3 h and then removed unbound ligands by ultrafiltration. Then we

Table 2. Distribution of RbkR and RibK proteins in major classes of Archaea

Taxonomic Phyla/Class/*Order*	# genomes	RbkR/RibK	B2 prototrophs/auxotrophs
Crenarchaeota (Thermoprotei)			
Acidilobales	2	kinase	prototrophs
Desulfurococcales	13	kinase	prototrophs/auxotrophs
Fervidicoccales	1	regulator/kinase	auxotroph
Sulfolobales	8	regulator/kinase	prototrophs
Thermoproteales	17	regulator/kinase	prototrophs/auxotrophs
Euryarchaeota			
Aciduliprofundum	2	regulator/kinase	prototrophs
Archaeoglobi	5	regulator/kinase	prototrophs
Halobacteria	17	regulator/kinase	prototrophs
Methanobacteria	3	kinase	prototrophs
Methanococci	4	kinase	prototrophs
Methanomicrobia			
- *Methanomicrobiales*	9	regulator/kinase	prototrophs
- *Methanosarcinales*	9	regulator/kinase	prototrophs
- *Methanocellales*	3	regulator/kinase	prototrophs
Methanopyri	1	kinase	prototroph
Thermococci	15	regulator/kinase	prototrophs/auxotrophs
Thermoplasmata	5	regulator/kinase	prototrophs/auxotrophs
Thaumarchaeota			
Cenarchaeales	1	regulator/kinase	prototroph
Nitrosopumilales	3	regulator/kinase	prototrophs
Nitrososphaeria	1	regulator/kinase	prototroph
Korarchaeota			
Candidatus *Korarchaeum*	1	regulator/kinase	auxotroph
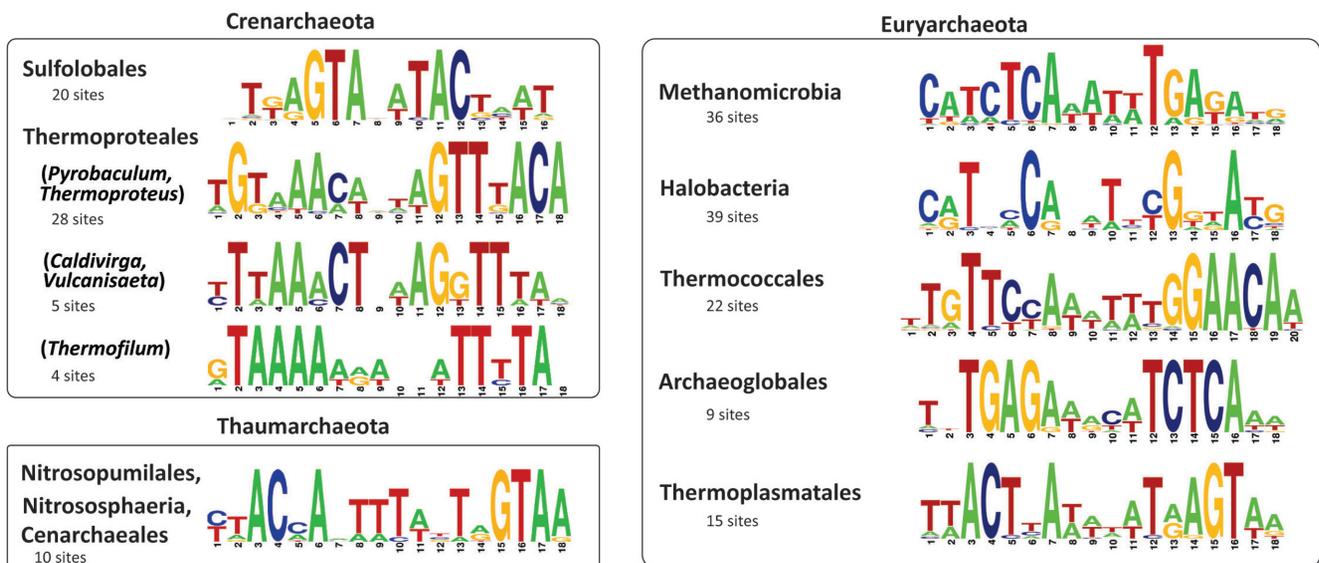
Figure 2. DNA binding motifs of RbkR regulators in various taxonomic groups of Archaea. Sequence logos representing the consensus binding site motifs were build using all candidate sites in the respective group of genomes. Number of sites in each training set is indicated.

conducted DNA-binding assays with either free or FMN-bound *my*RbkR to assess their interaction with both target DNA fragments. The FMN-bound protein did not show the *my*RbkR concentration-dependent binding with any tested DNA fragment, suggesting that FMN has a disruptive effect on DNA binding (Figure 4B).

The FMN-free *py*RbkR protein was tested by FPA for binding with two DNA fragments containing candidate regulatory sites upstream of the *arfA* and *ribB* genes in *P. yellowstonensis*. Both DNA fragments demonstrated specific interaction with *py*RbkR, with apparent EC_{50} values in the range of 20–40 nM (Figure 4C). Additional EMSA experi-

ments confirmed specific binding between *py*RbkR and intergenic regulatory regions of the *arfA* and *ribB* genes (Figure 4F). To test the influence of flavins on the interaction between *py*RbkR and the *ribB* DNA fragment, increasing concentrations of FMN, riboflavin or FAD were added to the incubation mixture. FAD did not show any significant effect on protein–DNA interaction, while riboflavin and FMN substantially diminished the binding affinity (Figure 4D). Half-maximal EC_{50} of riboflavin and FMN were calculated at 144 ± 17 nM and 41 ± 5 nM, respectively, showing that FMN has a greater negative effect on the DNA–protein interactions.
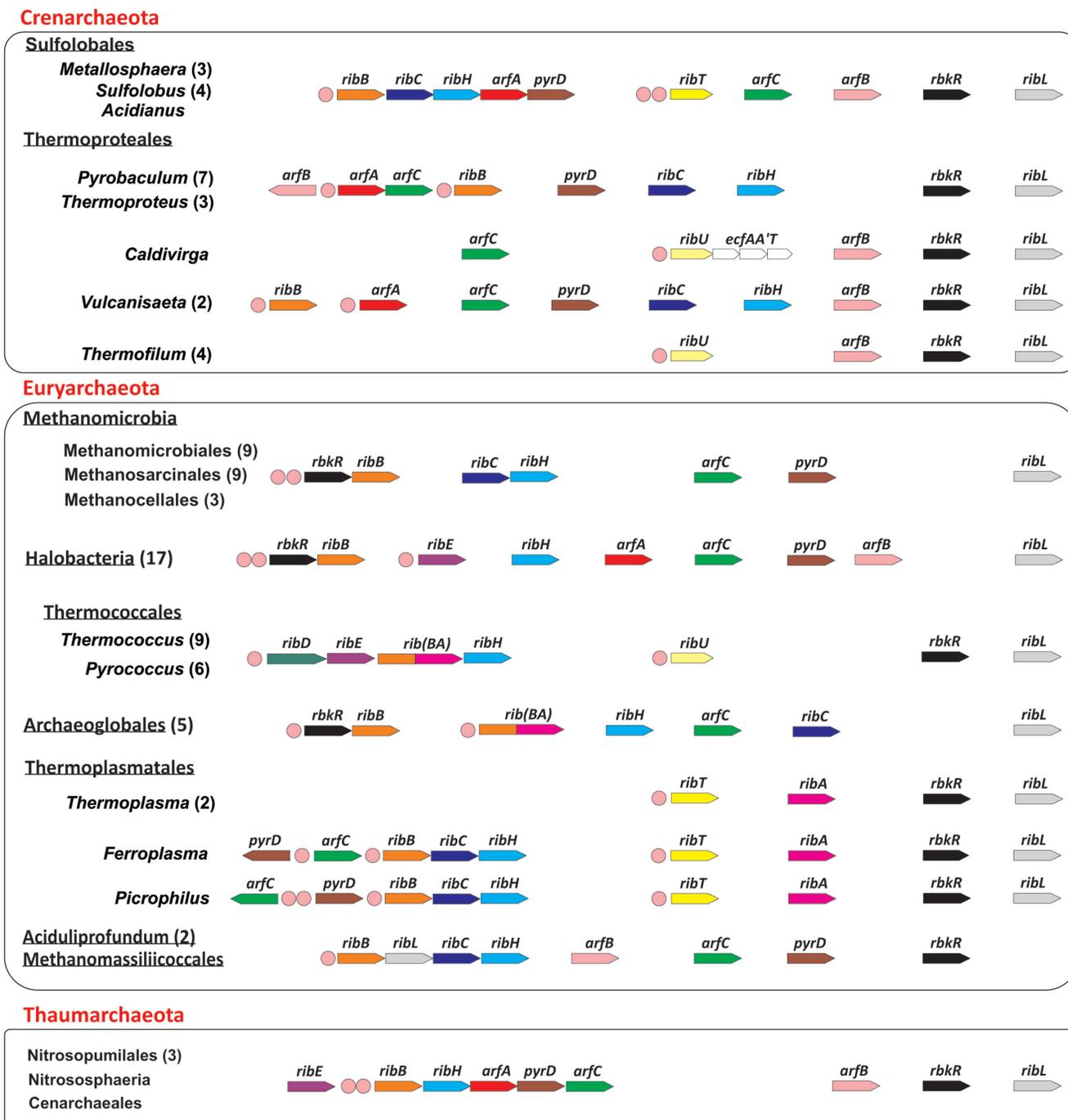
Figure 3. Genomic organization of riboflavin biosynthesis and uptake genes and reconstructed RbkR regulons in Archaea. Total number of analyzed genomes in each taxonomic group of species is shown in parenthesis. Genes are shown by arrows with the same functional role marked by matching shade of gray; abbreviations for genes correspond to those used in Figure 1. Gene fusions are shown in parenthesis. Candidate RbkR-binding sites are shown by pink circles.

To confirm that RbkR is a bifunctional protein, which functions both as DNA-binding transcriptional regulator and an enzyme, we measured the CTP-dependent riboflavin kinase activities of the recombinant RbkR proteins from *M. yellowstonensi* and *P. yellowstonensis* and compared them with the previously determined kinetic parameters of RibK from *M. jannaschii* (15). Both *my*RbkR and *py*RbkR pro-

teins displayed high riboflavin kinase activities, with maximum activity at 60°C (Supplementary Figure S3). The kinetics of both RbkR kinases was determined under steady-state conditions in the presence of saturating concentration of CTP. The measured apparent K_m values of *my*RbkR and *py*RbkR for riboflavin were 90 ± 37 μM and 73 ± 26 μM,
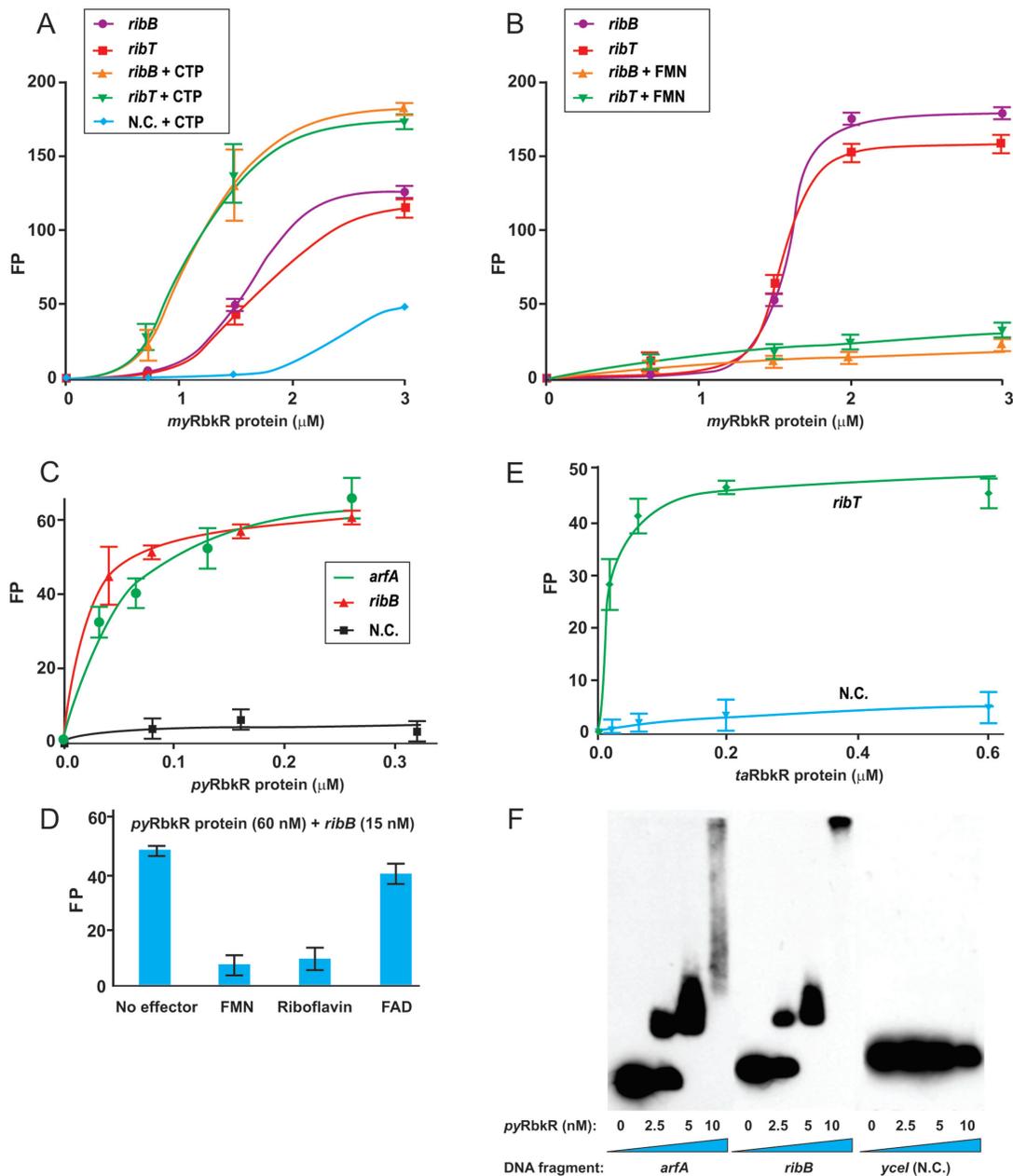
Figure 4. Experimental validation of the RbkR regulons by *in vitro* DNA binding assays. (**A–E**) Fluorescence polarization binding assay of RbkR proteins to cognate DNA sites in *Metallosphaera yellowstonensis*, *Pyrobaculum yellowstonensis* and *Thermoplasmaacidophilum* and the influence of effectors on protein–DNA binding. DNA–protein complex formation was monitored by an increase in the fluorescence polarization (FP) value. The error bars indicate standard deviations of triplicate values. (**A**) Increasing concentrations of *my*RbkR were mixed with 38- or 60-bp fluorescence-labeled DNA fragments of the *M. yellowstonensis ribB* and *ribT* gene promoter regions (10 nM). 0.1 mM of CTP added to preincubation mixture improved the protein–DNA interaction. As a negative control (N.C.), a 28-bp DNA fragment of the *M. yellowstonensis trxA* gene from the autotrophic regulon HhcR was used (30). (**B**) *my*RbkR was pre-incubated with 0.1 mM of FMN, and then residual cofactor was removed by ultrafiltration. Increasing concentrations of the FMN-treated and untreated *my*RbkR were mixed with target DNA fragments (10 nM). Treatment of *my*RbkR with FMN resulted in disruption of DNA–protein interaction with both fragments. (**C**) Increasing concentrations of *py*RbkR were mixed with 60- or 62-bp fluorescence-labeled DNA fragments of the *P. yellowstonensisarfA* and *ribB* gene promoter regions (10 nM). As a negative control (N.C.), a 62-bp mutated DNA fragment of the *arfA* gene was used. (**D**) Saturating concentrations of various flavins (200 nM) were added to the incubation mixture containing *py*RbkR (60 nM) and the *ribB* fragment (15 nM). (**E**) Increasing concentrations of *ta*RbkR were mixed with a 35-bp fluorescence-labeled DNA fragment of the *T. acidophilum ribT* gene promoter regions (10 nM). As a negative control (N.C.), a 35-bp scrambled DNA fragment was used. (**F**) EMSA with *py*RbkR and target DNA fragments from *P. yellowstonensis*. The biotin-labeled 102-bp DNA fragments (0.2 nM) of *arfA* and *ribB* were incubated with increasing concentrations of *py*RbkR (2.5–10 nM) in the absence of effectors. As a negative control (N.C.), a 243-bp DNA fragment of upstream region of the niacin transporter *yceI* from *P. yellowstonensis* was used.

respectively, similar to the obtained K_m value of RibK (159 μM) (15).

In summary, the results of *in vitro* biochemical assays confirm that RbkRs are bifunctional proteins catalyzing CTP-dependent phosphorylation of riboflavin and acting as DNA-binding transcriptional factors that are activated by CTP and inhibited by accumulation of FMN product.

Structure of specific *ta*RbkR–DNA complex

The tertiary structure of unliganded RbkR protein from *T. acidophilicum* (*ta*RbkR) was previously determined to 2.2 Å resolution by New York SGX Research Center for Structural Genomics (PDB ID: 3CTA) before the current project. Here we confirmed specific binding of *ta*RbkR to its single DNA site (5′-CTACTAATTCACGAGTAA) and determined the structure of *ta*RbkR in complex with DNA and CTP to highlight key residues for operator recognition and catalysis. We first tested the DNA-binding ability of the *ta*RbkR protein using FPA. A DNA fragment containing the predicted RbkR-binding site from the *T. acidophilicum ribT* gene upstream region demonstrated concentration-dependent binding to *ta*RbkR (Figure 4E). The apparent affinity estimated as EC_{50} of *ta*RbkR protein interacting with the tested DNA fragment was ∼16 nM.

In this work we have determined the structure of *ta*RbkR in complex with the tested DNA fragment and CTP (inclusion of CTP was critical to obtaining diffraction quality crystals). The 18-bp duplex co-crystallized with *ta*RbkR was formed by annealing of the 19-nt single stranded oligonucleotides (5′-**A**TTACTAATTCACGAGTAA, 5′-**T**TTACTCGTGAATTAG̅TAA̲) with single base pair complimentary overhangs (shown in bold), which is equivalent to the native *ta*RbkR operator except for a single C to T substitution (underlined). This change results in greater similarity to the palindromic consensus RbkR motif in the *Thermoplasmatales* (Figure 2). The *ta*RbkR–DNA–CTP complex crystallized in the monoclinic space group P21 with one protein dimer bound to a single DNA duplex per asymmetric unit. The structure contains the entire oligo sequence, the nucleotide CTP and the majority of the protein polypeptide, except four N-terminal residues and the hexahistidine C-terminal tag. The model was refined to R/R_{free} of 0.174/0.217 at 1.85 Å. The model has excellent stereochemistry with 98% of the backbone in the favored sections of the Ramachandran plot and a clash score (3.34) and molprobity score (1.18) >98% of deposited structures to a similar resolution (57). Data collection and refinement statistics are in Supplementary Table S4.

The overall structure of *ta*RbkR monomer contains two domains (Figure 5A). The N-terminal DNA-binding domain (residues 1–90) consists of a two-helical bundle (α2 and α3) composing the wHTH DNA-binding motif, a wing motif including three antiparallel β-strands and their connecting loop (β1, β2 and β3), and a pair of helices (α1 and α4) forming the dimerization interface. A search for similar structure using DALI (58) identified the MarR-family transcriptional regulator SlyA from *Salmonella enterica* (59) as the closest structural neighbor of the *ta*RbkR DNA-binding domain (19% identity, 41% similarity; Z score/rmsd = 7.1/3.4 Å). Although the relative positioning

of secondary structure elements in the *ta*RbkR and SlyA wHTH domains are similar, there are significant variations in their dimerization interfaces (Figure 5C; Supplementary Figure S6B).

The riboflavin kinase domain of *ta*RbkR determined in the complex with CTP (residues 91–220) contains four α-helices and seven β-strands forming a cradle-loop barrel fold similar to that determined for RibK kinase from *M. jannaschii* (13) (*mj*RibK, Figure 5B). Each *ta*RbkR subunit contains one bound CTP molecule. Pairwise structural similarity between the *mj*RibK and *ta*RbkR kinase domains, using DALI, yielded a Z-score of 18.0 and rmsd of 1.9 Å indicative of proteins with similar three dimensional structures but moderate sequence identity (∼40%) (Supplementary Figure S6A). The *ta*RbkR–CTP complex shows a nucleotide binding site formed by three amino acid motifs, 101-SGMGEGR-107, 128-YxGTLN-133 and 210-KYLR-203 (Supplementary Figure S4). Residues 199–212 form a loop and helix (α8) connecting β9 and β10 and contact the cytidine base portion of CTP. Residues 128–133 form a loop that contacts each phosphate of CTP either through a direct amide backbone hydrogen bond or indirect interactions involving an intervening sodium/magnesium ion. In the present structure, the sodium ion (likely derived from the protein buffer) was modeled on the basis of refined thermal factors and average oxygen/metal distances of 2.2–2.5Å, though it is assumed that in the physiological ion would be a magnesium. Residues 101–114 form a loop and a helix (α5) connecting β4 to α6 where six consecutive amide backbone atoms (Gly102-Arg107) form hydrogen bonds to the polyphosphate of CTP. The CTP-binding site of *ta*RbkR is similar to the previously described CDP-magnesium binding site of *mj*RibK (13). Multiple alignment of RbkR/RibK proteins from diverse archaeal lineages revealed that Gly102, Gly104, Gly106, Thr131, Leu132, Asn133, Leu202 and Arg203 from the CTP-binding site are highly conserved residues. In contrast, Tyr128 is conservatively replaced with phenylalanine in nearly half of riboflavin kinase domains (Supplementary Figure S5A).

The FMN-binding site previously identified in *mj*RibK–CDP–FMN (13) is conserved in *ta*RbkR (Tyr115, Phe165, Pro185, Tyr190 and Glu195), with only one analogous substitution of phenylalanine by tyrosine (Tyr109 in *ta*RbkR). Of these six FMN-contacting residues, Tyr115 and Glu195 are absolutely conserved in all RibK/RbkR domains, while Tyr109 and Phe165 are mostly represented by conservative substitutions by phenylalanine and tyrosine, respectively (Supplementary Figure S5A). Pro185 is occasionally replaced with isoleucine. The loop between β8 and β9 of *ta*RbkR is in an 'open' conformation, which was not observed in the *mj*RibK structures, suggesting a pathway for riboflavin access to the catalytic site. Residues Pro185 and Tyr190 (*ta*RbkR numbering) from this loop cover the flavin and ribose moiety in the *mj*RibK structure and therefore would need to close upon the riboflavin prior to catalysis. Using the FMN from the *mj*RibK–CDP–FMN structure as a model for the position of riboflavin in a *ta*RbkR–riboflavin–CTP complex, the ribose hydroxyl is 3.4Å from, and in the correct orientation for inline attack on the gamma phos-
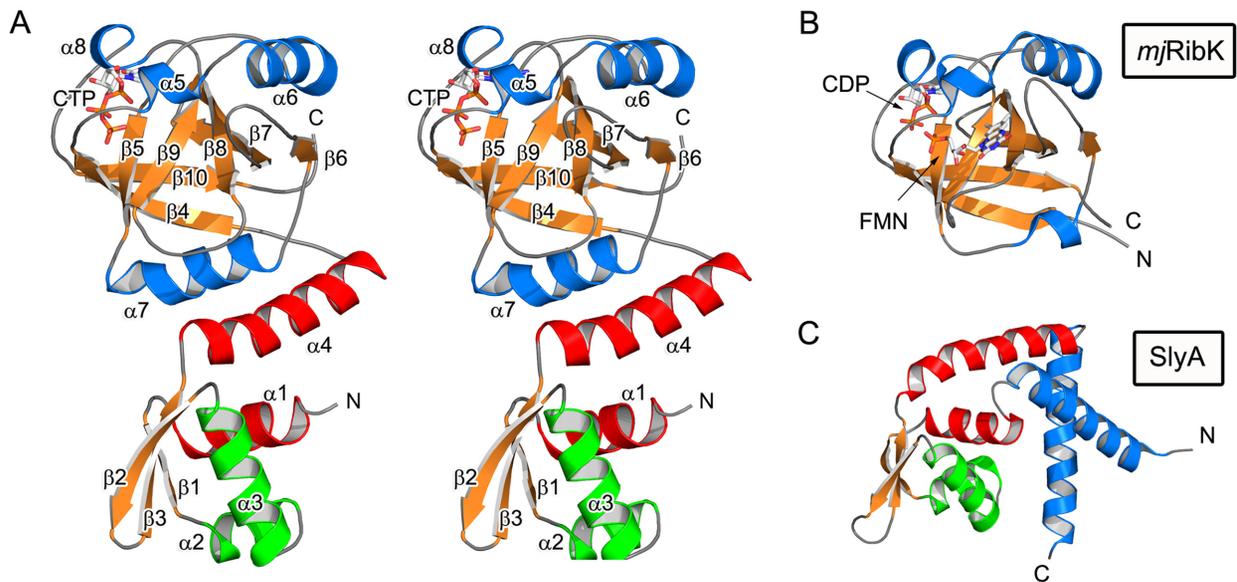
Figure 5. Structure of *ta*RbkR in complex with DNA and effector CTP. (**A**) Stereo ribbon diagram of a monomer of *ta*RbkR colored by secondary structure. HTH helicies α2 and α3 which bind the major groove colored in green while the dimer interface helices α1 and α4 colored in red. CTP binding to the riboflavin kinase domain is show in sticks colored by atom type. (**B**) *ta*RbkR has a kinase domain with a cradle-loop barrel fold similar to RibK from *Methanocaldococcus jannaschii* shown above (PDB ID 2VBV). (**C**) The DNA-binding domain of *ta*RbkR shows distant similarity to the MarR-family transcriptional regulator SlyA (PDB ID: 1LJ9). Structural elements of *mj*RibK and SlyA are colored by matching colors as *ta*RbkR in panel A.

phate of CTP, with Glu195 acting as a potential active site base.

The asymmetric unit of the *ta*RbkR–DNA–CTP complex structure contains a *ta*RbkR dimer bound to an 18-bp RbkR site DNA (Figure 6A). The conformations of the two monomers in the DNA-bound *ta*RbkR dimer are similar (rmsd of 0.735 Å over 215 common Cα atoms), with most of the differences arising from loop movements rather than domain movements. Comparison of the *ta*RbkR–DNA complex with the previously determined structure of apo-*ta*RbkR (PDB ID 3CTA) revealed larger differences (rmsd of 1.2 Å over 187 common Cα atoms). The largest differences between the apo state and the liganded structures are (i) an expansion (1–3 Å) of the interatomic distance between the HTH (α2 and α3) of one monomer and the corresponding structures of the opposing monomer, presumably due to optimization of their interactions with the major groove and (ii) a conformational change in the wing (β2/β3 and loop) where in the apo-*ta*RbkR structure this segment is 'flipped up', while in the DNA bound structure it is 'flipped down' to be in close proximity to the minor groove. While these changes can be explained by the new interactions formed between *ta*RbkR and DNA, there are no apparent changes in the relative orientation of the DNA binding domain and the kinase domain. Since FMN acts as a negative effector that causes dissociation of *ta*RbkR from DNA, we propose that a major conformational change in the dimer would appear upon binding of FMN ligand. Further structural characterization of RbkR in complex with FMN and comparison with the *ta*RbkR–DNA complex is required to achieve a better understanding of the RbkR-operated molecular switch.

The 18-bp DNA duplex in the complex with RbkR shows slight distortion from B-DNA conformation with an ∼30° bend in the middle, causing the ends of the DNA to move toward the protein (Figure 6A). Although the sequence of DNA is not fully palindromic (5′-TTACTaaTtcAcgAGTAA, palindromic positions are capitalized), the wHTH domain in each *ta*RbkR subunit makes a nearly identical set of contacts with the two halves of the DNA operator. *ta*RbkR interacts with DNA primarily through helix α3, termed the 'DNA recognition helix' and the β2–β3 hairpin wing of the wHTH domain. The residues Ser38, Glu39, Gln40 and Ser41 of the DNA recognition helix and Ser28 of helix α2 interact with DNA bases in the major groove (Figure 6B). The side chain of Gln39 hydrogen bonds with N6 and N7 of adenine 4 of one strand, while the side chain of Gln40 hydrogen bonds to O6 of guanine 16 on the opposite strand. Oγ of Ser28 approaches the phosphate of thymine 3, while Ser38 and Ser41 interact with the phosphate group of guanine 14 on the complementary strand. Residues Arg57 and Gln64 of the β-hairpin contact the phosphate groups of thymine 3 and adenine 4, respectively. Finally at the tip of the wHTH wing, Lys61 and Arg62 point into solvent, however they would likely interact with the phosphates of the minor groove if the oligo was extended.

Using a multiple sequence alignment of DNA-binding domains from archaeal RbkR proteins, we analyzed the conservation of residues involved in DNA recognition in *ta*RbkR (Supplementary Figure S5B). Noteworthy, RbkRs in the *Thermoproteales* lineage lack β1- and β2-strands suggesting a different mode of DNA recognition. Among eight residues forming hydrogen bonds with DNA, only Gln40 and Ser38 are conserved in most RbkR proteins (with a few exceptions). Gln40 is substituted with tryptophan and lysine in *Thermofilum* spp. and the *Methanomicrobia*, respectively, while Ser38 is replaced with threonine in some
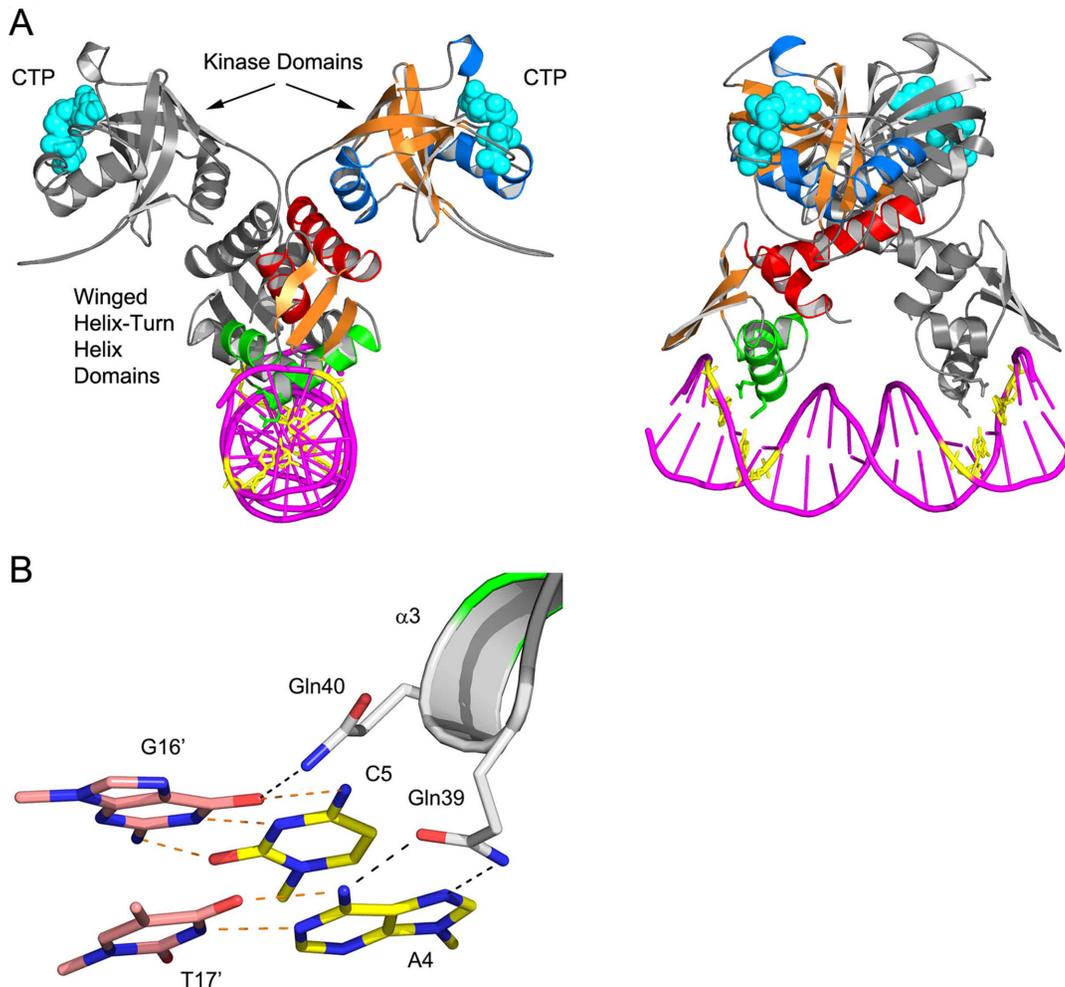
Figure 6. Structure of *ta*RbkR in complex with DNA. (**A**) Two views, rotated 90°, showing the interaction of *ta*RbkR dimer with a DNA oligonucleotide containing its cognate binding site. The protein is presented as ribbons. One monomer is colored as displayed in Figure 3, while the opposing monomer is colored gray. CTP effector is shown as CPK spheres in cyan. DNA is shown as sticks in magenta, with G16 and A4 shown in yellow. (**B**) Interactions between Gln39 and Gln40 of the recognition helix (α3) and G16 and A4 of the DNA duplex. Hydrogen bonds are indicated by dotted lines.

RbkRs from the *Thermoproteales*. Other DNA-contacting amino acids in *ta*RbkR are variable between various lineages of Archaea. Most strikingly, Gln39 directly contacting two nitrogen atoms in adenine is only conserved in the *Thermoplasmata*, *Thaumarchaeota* and *Sulfolobales*, while it is substituted with proline, arginine, asparagine, serine or branch chain amino acids in other taxonomic groups. Overall, the observed variabilities in the wHTH domain of RbkRs are consistent with the highly variable RbkR-binding DNA motifs predicted in various lineages of Archaea (Figure 2).

Evolutionary and functional implications

The identified RbkR regulators in Archaea represent a novel class of metabolite-sensing transcription factors emerging via fusion between DNA-binding and catalytic domains. The riboflavin kinase domain of RbkR proteins is orthologous to the RibK kinase previously characterized in *M. jannaschii* (13,15). Three lines of evidence suggest that all identified RbkR proteins in Archaea are bifunctional

kinases/regulators: (i) biochemical assays confirmed the riboflavin kinase activity of two analyzed representatives from the RbkR family; (ii) key residues involved in binding of FMN and CTP substrates are generally conserved across various RbkR orthologs; and (iii) the RbkR- and RibK-encoding genes never co-occur in a single genome representing a complementary phyletic pattern. Our results suggest that the riboflavin kinase domain of RbkRs serves not only as an essential function in the flavin biosynthesis but also as a sensor domain of DNA-binding transcription factor. A few other proteins combining DNA-binding transcriptional regulator with metabolic enzyme functions have been previously characterized in Bacteria. The biotin repressor/ligase BirA, which was previously characterized in Bacteria but that also has presumably bifunctional orthologs in some archaeal genomes (60), is composed of a DNA-binding domain fused to a functional biotin ligase domain (61). The NAD metabolism repressor NadR is a trifunctional protein with a DNA-binding domain fused to the nicotinamide ribose kinase and adenylyltransferase domains (62). The proline utilization protein PutA functions as a transcriptional

repressor and represents a fusion of a DNA-binding domain and two dehydrogenase domains that catalyze the first two reactions of proline degradation (63). All these multifunctional proteins, termed as 'trigger enzymes' (64), are involved in vitamin/cofactor and amino acid metabolism.

The observed genomic distribution of the *rbkR* and *ribK* genes in diverse lineages of Archaea suggests that a bifunctional enzyme/regulator likely existed in the last universal common ancestor of all Archaea. Later in speciation, some RbkR proteins could have lost their DNA-binding domains in several lineages of Archaea including the *Desulfurococcales*, *Methanococcales* and *Methanobacteriales*. Interestingly, the monofunctional RibK kinase in *M. jannaschii* still retain a weak tendency to dimerize (13). For the structures of *mj*RibK that crystallized as dimers, the subunits exhibit high spatial and orientation similarity to the kinase domains of *ta*RbkR, with *ta*RbkR losing the direct interactions between the kinase domains. Dimerization is an essential feature of DNA-binding transcriptional regulators, including RbkR, allowing them to bind cooperatively to palindromic operators.

A possible mechanism for the RbkR-mediated transcriptional regulation of riboflavin metabolism in Archaea can be proposed based on a combination of bioinformatics and experimental findings. In the unliganded form or in the complex with CTP, RbkR binds to the operator regions in the target genes and activates their transcription, thus promoting the biosynthesis or salvage of riboflavin. The synthesized or transported into the cell riboflavin is converted to FMN by the CTP-dependent riboflavin kinase constituting a C-terminal domain of RbkR. The subsequent accumulation of FMN in the cell causes RbkR to dissociate from the DNA and deactivate the transcription of the riboflavin biosynthesis/salvage genes. In contrast with the proposed 'activation-deactivation' mechanism in Archaea, regulation of the riboflavin metabolism genes in Bacteria is based on the alternative 'repression-derepression' mechanism utilizing the FMN-responsive RNA riboswitches (17,65).

Concluding remarks

In summary, by combining the comparative genomics-based regulon reconstruction with the focused biochemical and structural characterization, we identified a novel bifunctional transcriptional regulator/enzyme in Archaea. This regulator, termed RbkR, functions both as a CTP-dependent riboflavin kinase in the archaeal flavin synthesis pathway and as a DNA-binding transcription factor, likely activator, that controls the riboflavin biosynthesis and uptake genes in response to FMN. Moreover, RbkR is the first example of a transcription factor that controls the riboflavin metabolism. Cognate RbkR-binding sites were experimentally validated in three representative archaeal species, and the negative effect of FMN and positive effect of CTP on the RbkR–DNA complex formation was established. However, the predicted positive regulatory role of RbkR on target gene expression remains to be elucidated using the genetic and/or transcriptomics approaches. We further determined the structure of RbkR from *T. acidophilum* in complex with a specific DNA site and the CTP effector and compared it with the structure of the riboflavin kinase RibK from *M.*

jannaschii. The solved structure provides a first view of the structural mechanism underlying the regulatory function of RbkR. Highly variable sequences of the wHTH DNA-binding domains in RbkR proteins are consistent with substantial variations in the RbkR-binding DNA motifs identified in diverse archaeal lineages. On the other hand, conservation of CTP- and FMN-binding residues in the kinase domains of RbkR proteins indicates their conserved enzymatic function, and similar mechanism of response to CTP and FMN ligands. Identification of a novel family of DNA-binding regulators for the riboflavin metabolism fills a substantial gap in the knowledge of transcriptional regulation of the key metabolic pathways in Archaea.

ACCESSION NUMBER

The atomic coordinates and experimental data for *ta*RbkR in complex with DNA and CTP were deposited to the Protein Data Bank (PDB entry code 5TRD).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Fischer,M. and Bacher,A. (2010) Riboflavin Biosynthesis. In: Mander,LN and Liu,H (eds). *Comprehensive Natural Products II: Chemistry and Biology*. Elsevier Ltd., Kidlington, UK, Vol. 7, pp. 3–36.
2. Gutierrez-Preciado,A., Torres,A.G., Merino,E., Bonomi,H.R., Goldbaum,F.A. and Garcia-Angulo,V.A. (2015) Extensive identification of bacterial riboflavin transporters and their distribution across bacterial species. *PLoS One*, **10**, e0126124.
3. Rodionova,I.A., Li,X., Plymale,A.E., Motamedchaboki,K., Konopka,A.E., Romine,M.F., Fredrickson,J.K., Osterman,A.L. and Rodionov,D.A. (2015) Genomic distribution of B-vitamin auxotrophy and uptake transporters in environmental bacteria from the *Chloroflexi* phylum. *Environ. Microbiol. Rep.*, **7**, 204–210.

4. Vitreschak,A.G., Rodionov,D.A., Mironov,A.A. and Gelfand,M.S. (2002) Regulation of riboflavin biosynthesis and transport genes in bacteria by transcriptional and translational attenuation. *Nucleic Acids Res.*, **30**, 3141–3151.

5. Magnusdottir,S., Ravcheev,D., de Crecy-Lagard,V. and Thiele,I. (2015) Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes. *Front Genet.*, **6**, 148.

6. Fischer,M. and Bacher,A. (2011) Biosynthesis of vitamin B2: a unique way to assemble a xylene ring. *Chembiochem*, **12**, 670–680.

7. Graham,D.E., Xu,H. and White,R.H. (2002) A member of a new class of GTP cyclohydrolases produces formylaminopyrimidine nucleotide monophosphates. *Biochemistry*, **41**, 15074–15084.

8. Graupner,M., Xu,H. and White,R.H. (2002) The pyrimidine nucleotide reductase step in riboflavin and F(420) biosynthesis in archaea proceeds by the eukaryotic route to riboflavin. *J. Bacteriol.*, **184**, 1952–1957.

9. Grochowski,L.L., Xu,H. and White,R.H. (2009) An iron(II) dependent formamide hydrolase catalyzes the second step in the archaeal biosynthetic pathway to riboflavin and 7,8-didemethyl-8-hydroxy-5-deazariboflavin. *Biochemistry*, **48**, 4181–4188.

10. Osterman,A.L., Overbeek,R. and Rodionov,D.A. (2010) The use of subsystems to encode biosynthesis of vitamins and cofactors. In: Mander,LN and Liu,H (eds). *Comprehensive Natural Products II: Chemistry and Biology*. Elsevier Ltd., Kidlington, UK, Vol. 7, pp. 141–159.

11. Fischer,M., Romisch,W., Illarionov,B., Eisenreich,W. and Bacher,A. (2005) Structures and reaction mechanisms of riboflavin synthases of eubacterial and archaeal origin. *Biochem. Soc. Trans.*, **33**, 780–784.

12. Haase,I., Grawert,T., Illarionov,B., Bacher,A. and Fischer,M. (2014) Recent advances in riboflavin biosynthesis. *Methods Mol. Biol.*, **1146**, 15–40.

13. Ammelburg,M., Hartmann,M.D., Djuranovic,S., Alva,V., Koretke,K.K., Martin,J., Sauer,G., Truffault,V., Zeth,K., Lupas,A.N. *et al.* (2007) A CTP-dependent archaeal riboflavin kinase forms a bridge in the evolution of cradle-loop barrels. *Structure*, **15**, 1577–1590.

14. Mashhadi,Z., Xu,H., Grochowski,L.L. and White,R.H. (2010) Archaeal RibL: a new FAD synthetase that is air sensitive. *Biochemistry*, **49**, 8748–8755.

15. Mashhadi,Z., Zhang,H., Xu,H. and White,R.H. (2008) Identification and characterization of an archaeon-specific riboflavin kinase. *J. Bacteriol.*, **190**, 2615–2618.

16. Breaker,R.R. (2011) Prospects for riboswitch discovery and analysis. *Mol. Cell*, **43**, 867–879.

17. Winkler,W.C., Cohen-Chalamish,S. and Breaker,R.R. (2002) An mRNA structure that controls gene expression by binding FMN. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 15908–15913.

18. Mironov,A.S., Gusarov,I., Rafikov,R., Lopez,L.E., Shatalin,K., Kreneva,R.A., Perumov,D.A. and Nudler,E. (2002) Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell*, **111**, 747–756.

19. Sun,E.I., Leyn,S.A., Kazanov,M.D., Saier,M.H. Jr, Novichkov,P.S. and Rodionov,D.A. (2013) Comparative genomics of metabolic capacities of regulons controlled by cis-regulatory RNA motifs in bacteria. *BMC Genomics*, **14**, 597.

20. Burgess,C.M., Slotboom,D.J., Geertsma,E.R., Duurkens,R.H., Poolman,B. and van Sinderen,D. (2006) The riboflavin transporter RibU in *Lactococcus lactis*: molecular characterization of gene expression and the transport mechanism. *J. Bacteriol.*, **188**, 2752–2760.

21. Garcia Angulo,V.A., Bonomi,H.R., Posadas,D.M., Serer,M.I., Torres,A.G., Zorreguieta,A. and Goldbaum,F.A. (2013) Identification and characterization of RibN, a novel family of riboflavin transporters from *Rhizobium leguminosarum* and other proteobacteria. *J. Bacteriol.*, **195**, 4611–4619.

22. Vogl,C., Grill,S., Schilling,O., Stulke,J., Mack,M. and Stolz,J. (2007) Characterization of riboflavin (vitamin B2) transport proteins from *Bacillus subtilis* and *Corynebacterium glutamicum*. *J. Bacteriol.*, **189**, 7367–7375.

23. Jaehme,M. and Slotboom,D.J. (2015) Diversity of membrane transport proteins for vitamins in bacteria and archaea. *Biochim. Biophys. Acta*, **1850**, 565–576.

24. Rodionov,D.A., Hebbeln,P., Eudes,A., ter Beek,J., Rodionova,I.A., Erkens,G.B., Slotboom,D.J., Gelfand,M.S., Osterman,A.L., Hanson,A.D. *et al.* (2009) A novel class of modular transporters for vitamins in prokaryotes. *J. Bacteriol.*, **191**, 42–51.

25. Barrick,J.E. and Breaker,R.R. (2007) The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol.*, **8**, R239.

26. Leyn,S.A., Kazanov,M.D., Sernova,N.V., Ermakova,E.O., Novichkov,P.S. and Rodionov,D.A. (2013) Genomic reconstruction of the transcriptional regulatory network in *Bacillus subtilis*. *J. Bacteriol.*, **195**, 2463–2473.

27. Rodionov,D.A., Novichkov,P.S., Stavrovskaya,E.D., Rodionova,I.A., Li,X., Kazanov,M.D., Ravcheev,D.A., Gerasimova,A.V., Kazakov,A.E., Kovaleva,G.Y. *et al.* (2011) Comparative genomic reconstruction of transcriptional networks controlling central metabolism in the Shewanella genus. *BMC Genomics*, **12**(Suppl 1), S3.

28. Rodionov,D.A., Rodionova,I.A., Li,X., Ravcheev,D.A., Tarasova,Y., Portnoy,V.A., Zengler,K. and Osterman,A.L. (2013) Transcriptional regulation of the carbohydrate utilization network in *Thermotoga maritima*. *Front. Microbiol.*, **4**, 244.

29. Leyn,S.A. and Rodionov,D.A. (2015) Comparative genomics of DtxR family regulons for metal homeostasis in Archaea. *J. Bacteriol.*, **197**, 451–458.

30. Leyn,S.A., Rodionova,I.A., Li,X. and Rodionov,D.A. (2015) Novel transcriptional regulons for autotrophic cycle genes in Crenarchaeota. *J. Bacteriol.*, **197**, 2383–2391.

31. Benson,D.A., Clark,K., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J. and Sayers,E.W. (2015) GenBank. *Nucleic Acids Res.*, **43**, D30–35.

32. Jay,Z.J., Beam,J.P., Dohnalkova,A., Lohmayer,R., Bodle,B., Planer-Friedrich,B., Romine,M. and Inskeep,W.P. (2015) *Pyrobaculum yellowstonensis* strain WP30 respires on elemental sulfur and/or arsenate in circumneutral sulfidic geothermal sediments of yellowstone national park. *Appl. Environ. Microbiol.*, **81**, 5907–5916.

33. Jennings,R.M., Whitmore,L.M., Moran,J.J., Kreuzer,H.W. and Inskeep,W.P. (2014) Carbon dioxide fixation by *Metallosphaera yellowstonensis* and acidothermophilic iron-oxidizing microbial communities from Yellowstone National Park. *Appl. Environ. Microbiol.*, **80**, 2665–2671.

34. Markowitz,V.M., Mavromatis,K., Ivanova,N.N., Chen,I.M., Chu,K. and Kyrpides,N.C. (2009) IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics*, **25**, 2271–2278.

35. Mironov,A.A., Vinokurova,N.P. and Gel'fand,M.S. (2000) [Software for analyzing bacterial genomes]. *Mol. Biol. (Mosk)*, **34**, 253–262.

36. Guindon,S., Dufayard,J.F., Lefort,V., Anisimova,M., Hordijk,W. and Gascuel,O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.*, **59**, 307–321.

37. Overbeek,R., Olson,R., Pusch,G.D., Olsen,G.J., Davis,J.J., Disz,T., Edwards,R.A., Gerdes,S., Parrello,B., Shukla,M. *et al.* (2014) The SEED and the rapid annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.*, **42**, D206–D214.

38. Dehal,P.S., Joachimiak,M.P., Price,M.N., Bates,J.T., Baumohl,J.K., Chivian,D., Friedland,G.D., Huang,K.H., Keller,K., Novichkov,P.S. *et al.* (2010) MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res.*, **38**, D396–D400.

39. UniProt,C. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.

40. Nawrocki,E.P., Burge,S.W., Bateman,A., Daub,J., Eberhardt,R.Y., Eddy,S.R., Floden,E.W., Gardner,P.P., Jones,T.A., Tate,J. *et al.* (2015) Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.*, **43**, D130–D137.

41. Rodionov,D.A. (2007) Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. *Chem Rev.*, **107**, 3467–3497.

42. Novichkov,P.S., Rodionov,D.A., Stavrovskaya,E.D., Novichkova,E.S., Kazakov,A.E., Gelfand,M.S., Arkin,A.P., Mironov,A.A. and Dubchak,I. (2010) RegPredict: an integrated system for regulon inference in prokaryotes by comparative genomics approach. *Nucleic Acids Res.*, **38**, W299–W307.

43. Mironov,A.A., Koonin,E.V., Roytberg,M.A. and Gelfand,M.S. (1999) Computer analysis of transcription regulatory patterns in completely sequenced bacterial genomes. *Nucleic Acids Res.*, **27**, 2981–2989.

44. Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.

45. Crooks,G.E., Hon,G., Chandonia,J.M. and Brenner,S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.

46. Mossessova,E. and Lima,C.D. (2000) Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol. Cell*, **5**, 865–876.

47. Battye,T.G., Kontogiannis,L., Johnson,O., Powell,H.R. and Leslie,A.G. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 271–281.

48. Evans,P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 72–82.

49. Vagin,A. and Teplyakov,A. (2010) Molecular replacement with MOLREP. *Acta Cryst.*, **66**, 22–25.

50. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Cryst.*, **D60**, 2126–2132.

51. Adams,P.D., Gopal,K., Grosse-Kunstleve,R.W., Hung,L.W., Ioerger,T.R., McCoy,A.J., Moriarty,N.W., Pai,R.K., Read,R.J., Romo,T.D. *et al.* (2004) Recent developments in the PHENIX software for automated crystallographic structure determination. *J. Synchrotron. Rad.*, **11**, 53–55.

52. Winn,M.D., Isupov,M.N. and Murshudov,G.N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Cryst.*, **D57**, 122–133.

53. Iyer,L.M. and Aravind,L. (2012) Insights from the architecture of the bacterial transcription apparatus. *J. Struct. Biol.*, **179**, 299–319.

54. Wurtzel,O., Sapra,R., Chen,F., Zhu,Y., Simmons,B.A. and Sorek,R. (2010) A single-base resolution map of an archaeal transcriptome. *Genome Res.*, **20**, 133–141.

55. Yoon,S.H., Reiss,D.J., Bare,J.C., Tenenbaum,D., Pan,M., Slagel,J., Moritz,R.L., Lim,S., Hackett,M., Menon,A.L. *et al.* (2011) Parallel evolution of transcriptome architecture during genome reorganization. *Genome Res.*, **21**, 1892–1904.

56. Jager,D., Forstner,K.U., Sharma,C.M., Santangelo,T.J. and Reeve,J.N. (2014) Primary transcriptome map of the hyperthermophilic archaeon *Thermococcus kodakarensis*. *BMC Genomics*, **15**, 684.

57. Chen,V.B., Arendall,W.B. 3rd, Headd,J.J., Keedy,D.A., Immormino,R.M., Kapral,G.J., Murray,L.W., Richardson,J.S. and Richardson,D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12–21.

58. Holm,L. and Laakso,L.M. (2016) Dali server update. *Nucleic Acids Res.*, **44**, W351–W355.

59. Dolan,K.T., Duguid,E.M. and He,C. (2011) Crystal structures of SlyA protein, a master virulence regulator of *Salmonella*, in free and DNA-bound states. *J. Biol. Chem.*, **286**, 22178–22185.

60. Rodionov,D.A., Mironov,A.A. and Gelfand,M.S. (2002) Conservation of the biotin regulon and the BirA regulatory signal in Eubacteria and Archaea. *Genome Res.*, **12**, 1507–1516.

61. Wood,Z.A., Weaver,L.H., Brown,P.H., Beckett,D. and Matthews,B.W. (2006) Co-repressor induced order and biotin repressor dimerization: a case for divergent followed by convergent evolution. *J. Mol. Biol.*, **357**, 509–523.

62. Grose,J.H., Bergthorsson,U. and Roth,J.R. (2005) Regulation of NAD synthesis by the trifunctional NadR protein of *Salmonella enterica*. *J. Bacteriol.*, **187**, 2774–2782.

63. Zhou,Y., Zhu,W., Bellur,P.S., Rewinkel,D. and Becker,D.F. (2008) Direct linking of metabolism and gene expression in the proline utilization A protein from *Escherichia coli*. *Amino Acids*, **35**, 711–718.

64. Commichau,F.M. and Stulke,J. (2008) Trigger enzymes: bifunctional proteins active in metabolism and in controlling gene expression. *Mol. Microbiol.*, **67**, 692–702.

65. Pedrolli,D., Langer,S., Hobl,B., Schwarz,J., Hashimoto,M. and Mack,M. (2015) The *ribB* FMN riboswitch from *Escherichia coli* operates at the transcriptional and translational level and regulates riboflavin biosynthesis. *FEBS J.*, **282**, 3230–3242.