A CTP-Dependent Archaeal Riboflavin Kinase Forms a Bridge in the Evolution of Cradle-Loop Barrels

Moritz Ammelburg,1,4 Marcus D. Hartmann,1,4 Sergej Djuranovic,1,5 Vikram Alva,1 Kristin K. Koretke,2 Jörg Martin,1 Guido Sauer,3 Vincent Truffault,1 Kornelius Zeth,1 Andrei N. Lupas,1,* and Murray Coles1,*

1Department of Protein Evolution, Max-Planck-Institute for Developmental Biology, 72076 Tübingen, Germany
2Computational Chemistry Group, GlaxoSmithKline, Collegeville, PA 19426, USA
3Department of Biochemistry, Max-Planck-Institute for Developmental Biology, 72076 Tübingen, Germany
4These authors contributed equally to this work.
5Present address: Howard Hughes Medical Institute, Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA.
*Correspondence: murray.coles@tuebingen.mpg.de (M.C.), andrei.lupas@tuebingen.mpg.de (A.N.L.)
DOI 10.1016/j.str.2007.09.027

SUMMARY

Proteins of the cradle-loop barrel metafold are formed by duplication of a conserved βxβ-element, suggesting a common evolutionary origin from an ancestral group of nucleic acid-binding proteins. The basal fold within this metafold, the RIFT barrel, is also found in a wide range of enzymes, whose homologous relationship with the nucleic acid-binding group is unclear. We have characterized a protein family that is intermediate in sequence and structure between the basal group of cradle-loop barrels and one family of RIFT-barrel enzymes, the riboflavin kinases. We report the structure, substrate-binding mode, and catalytic activity for one of these proteins, Methano- caldococcus jannaschii Mj0056, which is an archaeal riboflavin kinase. Mj0056 is unusual in utilizing CTP rather than ATP as the donor nucleotide, and therefore represents a rare CTP-dependent kinase.

INTRODUCTION

Riboflavin kinases (RFKs) from bacteria and eukaryotes catalyze the phosphorylation of riboflavin to form flavin mononucleotide (FMN). Systematically, they are classed as ATP:riboflavin 5’-phosphotransferases (EC 2.7.1.26). All examples known to date are closely related in sequence, and the available structures of the enzyme (Homo sapiens, 1NB0, 1O9S [Karthikeyan et al., 2003a, 2003b]; Schizosaccharomyces pombe, 1NO8, [Bauer et al., 2003]; and Thermotoga maritima, 1MRZ [Wang et al., 2003]) show a RIFT barrel fold, belonging to the cradle-loop barrel metafold of small β-barrels (Coles et al., 2006). In the course of a study into the evolution of this metafold, we identified a family of proteins exemplified by Methano- caldococcus jannaschii Mj0056, whose sequence and gene environment suggested a RIFT barrel with a role in riboflavin biosynthesis. Here, we characterize Mj0056 as an archaeal riboflavin kinase, structurally similar but topologically distinct from bacterial and eukaryotic examples. Surprisingly, Mj0056 utilizes CTP rather than ATP as the donor nucleotide, and therefore represents a rare CTP-dependent kinase.

The cradle-loop barrel metafold comprises three distinct topologies, all with (pseudo-) 2-fold symmetry: the double psi, the swapped hairpin, and the RIFT barrel (Coles et al., 1999, 2005, 2006). We have shown that a basal group of proteins spanning these three folds resemble each other at a level indicative of homology and have proposed an evolutionary scenario (Figure 1) in which an ancestral homodimeric RIFT barrel gave rise to swapped hairpin barrels by strand invasion and to double-psi barrels by fusion and strand swapping (Coles et al., 2006). This scenario is underpinned by the hypothesis that folded proteins evolved from an ancestral pool of peptides, which had themselves evolved as cofactors of RNA-based catalysis and replication (the “RNA world”) (Lupas et al., 2001; Söding and Lupas, 2003). In this case, the ancestral peptide consisted of a βαβ element that encloses an orthogonal turn with a conspicuous Gly-Asp motif (the GD box). Basal cradle-loop barrels retain the ability to bind nucleic acids, although in the case of the double-psi barrels found at the N terminus of AAA proteins, this activity is vestigial and superseded by polypeptide binding.

We proposed that the RIFT barrel is the ancestral form of cradle-loop barrels because of its simple topology and widespread occurrence in ancient proteins, such as El-Tu and related translation factors, ribosomal protein L3, the N-domain of the F1 ATPase, and enzymes involved in riboflavin synthesis, including riboflavin kinases (Coles et al., 2006). This proposal remained inconclusive, due to the lack of evidence for the homologous origin of these proteins from the basal RIFT barrel we characterized,
Here, we show that Mj0056 has sequence properties intermediate between basal cradle-loop barrels and ATP-dependent riboflavin kinases. We propose that it represents an evolutionary bridge between the two groups of proteins.

RESULTS

Bioinformatics

We first noticed the sequence similarity between Mj0056 and cradle-loop barrels in our initial bioinformatic characterization of these proteins (Coles et al., 1999) by using a sequence search tool based on reciprocal PSI-Blast searches (SENSE) (Koretke et al., 2002). At the time, we considered Mj0056 to be an archaeal transcription factor, due to its sequence similarity to the N-terminal DNA-binding domains of AbrB-like transcription factors and to the fact that most of its close homologs (but not Mj0056 itself) carry a winged-helix HTH DNA-binding domain at their N terminus. Indeed, in the COG database, Mj0056 homologs are annotated as COG1339—transcriptional regulator of a riboflavin/FAD biosynthetic operon. However, subsequent experiments designed to show an affinity of Mj0056 for DNA, including the sequence from the upstream region of its own gene, failed completely.

Revisiting this analysis more recently with a new search tool based on HMM-HMM comparisons (HHsenser) (Söding et al., 2006), we found that the top matches for Mj0056 not only included the expected basal cradle-loop barrels, but also two riboflavin kinases (Table 1), providing a new functional hypothesis. We therefore decided to analyze the sequence relationships between these proteins with a clustering procedure based on the Fruchterman-Reingold algorithm (CLANS) (Frickey and Lupas, 2004). In the cluster map, Mj0056 appeared at an intermediate position between AbrB-like transcription factors and riboflavin kinases (Figure 2). Although closer to the AbrB N-domain (AbrB-N), a multiple alignment showed that Mj0056 shared with riboflavin kinases three regions with residues important for the catalytic activity of riboflavin kinase (Karthikeyan et al., 2003a), distributed over its entire length (Figure 3). In contrast, the similarity with AbrB-N and other basal cradle-loop proteins only covered...
the C-terminal part of Mj0056 and rested mainly on the conservation of the structural residues required to form the extended GD-box (Figure 3).

**Mj0056 Is a CTP-Dependent Archaeal Riboflavin Kinase**

On the basis of our new functional hypothesis, we asayed Mj0056 for riboflavin kinase activity, as well as for nucleotide and riboflavin binding. These experiments were all done with ATP as the donor nucleotide and proved unsuccessful. We were aided out of this impasse by the crystal structure of Mj0056, which had just been deposited in the Protein Data Bank by the New York SGC Research Center for Structural Genomics (2OYN). At that point, we had had the solution structure of apo-Mj0056, described below, for almost 2 years, but 1OYN was in complex with a nucleotide product and the nucleotide was not ADP, but CDP.

![Figure 2. Cluster Map of Mj0056 Homologs](image)

The map, obtained with CLANS (see Experimental Procedures), shows Mj0056 in the context of double-psi barrels (AAA N domains, double-psi barrel enzymes), swapped-hairpin barrels (AbrB superfamily), and RIFT barrels (PhS018 group, riboflavin kinases, riboflavin synthases).

### Table 1. Best Sequence Matches to Mj0056 in the Protein Data Bank Using HMM-HMM Comparisons

<table>
<thead>
<tr>
<th>Rank</th>
<th>PDB</th>
<th>Protein</th>
<th>Prob.</th>
<th>E value</th>
<th>P value</th>
<th>Query HMM</th>
<th>Template HMM</th>
<th>Fold[a]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2OYN_A</td>
<td>Mj0056</td>
<td>100.0</td>
<td>0</td>
<td>0</td>
<td>1–136</td>
<td>3–138 (146)</td>
<td>RIFT barrel (self hit)</td>
</tr>
<tr>
<td>2</td>
<td>1YFB_A</td>
<td>AbrB-N</td>
<td>92.5</td>
<td>0.088</td>
<td>6e–06</td>
<td>107–134</td>
<td>23–50 (59)</td>
<td>swapped-hairpin barrel[b]</td>
</tr>
<tr>
<td>3</td>
<td>2GLW_A</td>
<td>PhS018</td>
<td>85.9</td>
<td>0.49</td>
<td>3.3e–05</td>
<td>108–132</td>
<td>62–86 (92)</td>
<td>RIFT barrel[a]</td>
</tr>
<tr>
<td>4</td>
<td>1MRZ_A</td>
<td>Riboflavin kinase/FMN adenyl transferase</td>
<td>71.2</td>
<td>4.6</td>
<td>0.00032</td>
<td>6–127</td>
<td>161–267 (293)</td>
<td>RIFT barrel[a]</td>
</tr>
<tr>
<td>5</td>
<td>1N08_A</td>
<td>Riboflavin kinase</td>
<td>66.7</td>
<td>7.9</td>
<td>0.00054</td>
<td>5–123</td>
<td>24–129 (163)</td>
<td>RIFT barrel</td>
</tr>
<tr>
<td>6</td>
<td>1WLF_A</td>
<td>PEX1-N</td>
<td>60.4</td>
<td>7.7</td>
<td>0.00053</td>
<td>93–133</td>
<td>59–97 (179)</td>
<td>double-psi barrel[b]</td>
</tr>
<tr>
<td>7</td>
<td>1MVF_D</td>
<td>MazE</td>
<td>52.5</td>
<td>7.1</td>
<td>0.00049</td>
<td>108–133</td>
<td>15–40 (82)</td>
<td>swapped-hairpin barrel[b]</td>
</tr>
</tbody>
</table>

The search was done with the program HHsearch 1.5.0 (http://toolkit.tuebingen.mpg.de/hhpred/) in default settings on the Protein Data Bank release of April 10, 2007, filtered for a maximum of 70% pairwise sequence identity (pdb70).

[a] Jointly, these three folds form the cradle-loop barrel metafold.

[b] We have shown previously that these proteins are similar at a level indicative of homology (Coles et al., 2006).
We therefore reanalyzed the ability of Mj0056 to convert riboflavin to FMN at various temperatures by mass spectrometry, this time with a range of donor nucleotides (see Experimental Procedures). In electrospray ionization mass spectrometry, riboflavin (MW = 376 Da) exhibits a strong response in positive ion mode at mass/charge ratios of 377, 399, and 775 (Figure 4A), whereas FMN (MW = 456 Da) shows an intense signal at 455 in negative ion mode and characteristic signals at 911 and 933 (Figure 4B) (Susin et al., 1993). Under the chosen assay conditions, we obtained riboflavin kinase activity with both CTP and UTP as phosphate donors, with UTP being at least one order of magnitude less efficient. At reaction temperatures of up to 85°C—the temperature of the natural habitat of M. jannaschii—riboflavin was completely converted to FMN (Figures 4C and 4D and see the Supplemental Data available with this article online). ATP and GTP did not support the production of FMN at any reaction temperature (Figures 4E and 4F and Supplemental Data).

Solution Structure of Apo-Mj0056

We determined the solution structure of apo-Mj0056 (136 residues, 15.7 kDa) in the earliest stages of this project. High quality spectra led to largely complete resonance assignments by using standard methods (see Experimental Procedures), with the notable exception of two larger segments, G14-S23 and T99-S102, where the backbone amide signals were not observed. Subsequent secondary structure analysis showed that these regions correspond to the two cradle loops expected in the RIFT barrel fold. We obtained a detailed structure with a combination of distance restraints derived from several 2D- and 3D-NOESY spectra, chemical-shift-derived backbone torsion angle restraints, and 3JHNH and 3JNH_i-1 coupling constants (see Experimental Procedures). The final set of experimental restraints is described in Table 2, and the ensemble of 19 structures is shown in Figure 5. The ensemble is well defined; the rmsd for superimposition over structured residues is 0.25 Å for backbone atoms and 0.61 Å for all heavy atoms (Table 2). The restraint violations...
are also very low, with no persistent distance restraint violations over 0.1 Å (4.0 violations > 0.05 Å per structure) and no dihedral restraint violation greater than 0.6°.

Mj0056 shows a six-stranded β-barrel fold (Figure 5) with the expected RIFT barrel topology (for an explanation of secondary structure notation, see the caption to Figure 5). In keeping with its sequence properties, the structure can be divided into a degenerate N-terminal and a canonical C-terminal half. The latter is structurally similar to the basal cradle-loop barrels. This similarity is centered on a homologous βzβ element, the hallmark of which is the GD box sequence motif (118-FnLkdGDvI-126 in Mj0056) (Figure 3). This motif forms a very similar structure in all examples known to date, as we have previously described in detail (Coles et al., 2006). The β strands are orthogonal and linked by the helix and GD-box, which cross over an open end of the barrel. The GD box itself contains a type II β turn positioned within the barrel architecture by
conserved hydrophobic and hydrogen bonding interactions. One significant difference to the basal $\beta\beta\beta$ element is in the entrance to the $\alpha$ helix, despite the conservation of the PxxxR sequence motif (P111-R115) observed in a wide range of cradle-loop barrels. In contrast to most other examples, two residues in extended conformation separate the proline from the helix, with the result that the $\beta$ strand is shifted outward. This conformation is stabilized by a noncanonical hydrogen bond to Y40 on the N-terminal half of the protein and creates a pocket behind $\beta'2$, later identified as the cytosine binding site.

In contrast to the C-terminal half, the degenerate N-terminal half of Mj0056 deviates significantly from the basal RIFT-barrel fold and lacks both helix $\alpha1$ and the GD box. The main differences are provided by two large insertions (Figure 3). One insertion elongates the first cradle loop (S13-T43), forming an $\alpha$-helix ($\alpha1$; P25-L35) and contributing to a structured connector (G36-T43), which leads into $\beta2$. The second insertion consists of a short $3_{10}$ helix and a $\beta$-hairpin ($\beta1$-$\beta2$), which continues unbroken into strand $\beta3$.

Crystal Structures of Mj0056

The first crystals we obtained for Mj0056 were cocrystals with inorganic phosphate (Mj0056-PO$_4$), and we solved this structure by molecular replacement with a preliminary
solution structure. After it became clear that Mj0056 is a CTP-dependent riboflavin kinase, we focused on obtaining cocrystals with substrates and products (Table 3). We solved two structures, Mj0056-MgCDP and Mj0056-MgCDP-FMN, in complex with natural reaction products and a third, Mj0056-NaCDP-PO₄, with inorganic phosphate bound in a similar position as the FMN phosphate (Figure 6).

Mj0056-MgCDP (Figure 7A) shows a nucleotide binding site centered on a conserved motif at the N terminus of β₂ (40-YegTLN-45). The cytosine ring packs between Y40 and L44, with further contacts to the pocket at the junction of β₂' and α₁', while the ribose interacts primarily with R115 on α₁'. The α- and β-phosphates of CDP are coordinated by T43 and N45 via the intermediate Mg²⁺ and interact with a glycine-rich sequence motif (G14-G18) in the first cradle loop. The induction of a transient helix beginning at the last residue of this glycine-rich motif (α₀, G18-S23) represents the major difference between the nucleotide-bound and apo forms of the protein (Figure 6). In the solution structure, only a weak helical tendency or nascent helix was detected for these residues. Our structure for Mj0056-MgCDP resembles closely the Mj0056-NaCDP structure.

**Figure 5. The Solution Structure of Apo-Mj0056**

(A) A secondary structure cartoon; β strands are in green, helices are yellow, and the two cradle-loops are blue. Secondary structure elements corresponding to the basal RIFT barrel fold are given conventional notation, while inserted elements are denoted with I and shown in light colors. The right view represents the left view rotated by 90° around the horizontal axis.

(B) A stereoview of the final set of 19 structures superimposed over ordered residues (defined in Table 2). Coloring is as in (A).
The ternary complex, Mj0056-MgCDP-FMN, contains the CDP in the same location as Mj0056-MgCDP. The FMN is enclosed on three sides by the $\beta$-barrel, the transient helix $a_0$, and the second cradle loop (K98-S103) (Figure 7B). The latter is in a closed conformation, making contacts to all moieties of FMN. This is in contrast to the solution structure, where it is flexible, and to the Mj0056-MgCDP structure, where it is in an open conformation (Figure 6). The isoalloxazine ring forms $\pi$-stacking interactions with the side chain of a conserved aromatic residue (F21) on $\alpha_0$, and hydrogen bonds to the backbone of F73 on $\beta I 2$ and to the side chain of Y27 on $\alpha I$ via a bridging water molecule. The 4' and 5' oxygens of FMN contact the two carboxyl oxygens of the invariant glutamate on $\beta I 2$ (E107) (Figure 3), suggesting that this residue acts as a base in activating the 5' hydroxyl of riboflavin for nucleophilic attack.

Two structures, Mj0056-PO$_4$ and Mj0056-NaCDP-PO$_4$, contain a PO$_4$ ion bound in the place of the FMN phosphate, albeit shifted away from the active site by approximately 2 Å. In both these structures, the second cradle loop is in the closed conformation, while helix $\alpha_0$ is formed.

**Figure 6. The Proposed Sequential Substrate Binding Cycle of Mj0056 and Concerted Conformational Changes**

The binding of the nucleotide to the apo structure (A and B) leads to the formation of the transient helix $\alpha 0$ in the first cradle loop; with the subsequent binding of riboflavin/FMN (B and C), the second cradle loop adopts a closed conformation. As depicted in (D), this closure can also be induced by binding of inorganic phosphate, which is however not sufficient for the formation of $\alpha 0$. In the structure Mj0056-NaCDP-PO$_4$ (not shown), the cradle loop conformations are similar to those shown in (C).
only in Mj0056-NaCDP-P\textsubscript{2}O\textsubscript{4}. Thus, we observe two major conformational changes upon ligand binding, both affecting the FMN binding site: formation of helix \( \alpha_0 \) in the first cradle loop, induced by CDP binding, and the closure of the second cradle loop, induced by phosphate binding (Figure 6 and Table 3).

With the exception of the Mj0056-MgCDP, which is monomeric in one crystal form, all crystal structures presented here, plus Mj0056-NaCDP (2OYN), form dimers via antiparallel pairing of their \( \beta_1 \) strands. This dimerization is independent of substrate binding (Table 3) and does not affect functional regions of the protein. Analytical gel sizing showed the solutions used for both crystallization and enzyme assays to have some small dimeric component (<5%). Thus, Mj0056 appears to have a weak tendency to dimerize that is accentuated in crystallization but is irrelevant to primary riboflavin kinase activity. This may not be the case for many of its close homologs that carry a winged helix HTH DNA-binding domain at their N terminus and are therefore likely to be active in a dimeric form.

Structure-based searches on Mj0056-MgCDP-FMN using DALI (Holm and Sander, 1993) show the expected similarity to bacterial and eukaryotic RFKs, with the enzyme from \textit{T. maritima} being among the top three matches (Z score/rmsd = 5.1/2.7 Å). Also among the best matches are other RIFT-barrel enzymes, such as siderophore-interacting protein (2GPJ, 5.3/3.4 Å), flavodoxin reductase (1FDR, 4.6/2.7 Å), and yeast riboflavin synthase (1KZL, 4.6/3.4 Å). However, the next group of matches are double-psi \( \beta \) barrels from AAA ATPase N domains, e.g., VAT (1CZ4, 3.9/2.8 Å) and NSF (1QCS, 3.8/3.4 Å), underlining the structural similarity of members of the cradle-loop barrel metafold.

DISCUSSION

We have identified a family of proteins with properties intermediate between basal cradle-loop barrels and riboflavin kinases. Detailed structural and biochemical analysis of one of these proteins, Mj0056, showed it to be an
archaeal riboflavin kinase, with a specificity for CTP as the phosphate donor. This specificity is highly unusual; of the 25 families in the kinase classification of Grishin and coworkers (Cheek et al., 2005), only one—dolichol kinase—is CTP specific. As an all-helical integral membrane protein, dolichol kinase clearly represents an analogous development to MJ0056.

**Mechanistic Implications**

Despite its different nucleotide specificity, MJ0056 clearly resembles bacterial and eukaryotic RFKs at several levels. Both share the RIFT barrel fold and similar overall structures (Figure 8A). In the active site, both have the glycine rich loop and the TxN motif, which coordinate the phosphates of the donor nucleotide, and the glutamate residue, which is thought to activate the 5'-hydroxyl of riboflavin, initiating the phosphate transfer (Bauer et al., 2003; Karthikeyan et al., 2003a).

Outside the phosphate transfer site, there are considerable differences in the nucleotide binding mode. For the donor nucleotide, the two large hydrophobic residues that sandwich the cytosine ring in MJ0056 are absent from bacterial and eukaryotic RFKs, as is helix α1’ and the arginine coordinating the ribose hydroxyl groups. Instead, ATP-dependent RFKs use small side chains and a wider loop in place of α1’ to accommodate the larger adenine moiety (Figure 8B). The flavin binding site also shows considerable differences. The elaborations to the RIFT barrel fold, which enclose the isalloxazine ring, have striking structural similarity but originate in entirely different ways; in MJ0056, they are found in two insertions into the N-terminal half of the barrel, while in bacterial/eukaryotic kinases, they form an extension to the C-terminal half (Figure 3). Also, the transient helix α0 of the former is present as a shorter 310 helix in the latter, where it lacks the π-stacking aromatic residue. This leads to significant differences in the position and mode of flavin binding in the two groups (Figure 8C).

Comparisons between the different structures we have determined allow us to make inferences regarding the mechanism of archaeal RFKs (Figure 6). On one hand, binding of CDP induces formation of the transient helix α0 in all crystal forms containing the nucleotide (data not shown). On the other, NMR binding studies with riboflavin and FMN showed no measurable affinity in the absence of the donor nucleotide (Table 3). On the other, NMR binding studies with riboflavin and FMN showed no measurable affinity in the absence of the donor nucleotide (data not shown). We conclude that substrate binding is sequential, with CTP binding first and inducing the conformation required for flavin binding in the first cradle loop. The interactions between the flavin and the second cradle loop induce the closed conformation in the latter, initiating the transfer reaction. Our structural data do not indicate in which order the products dissociate from the kinase.

**Evolutionary Implications**

The structure of MJ0056 provides a number of clues in devising a scenario for the evolution of riboflavin kinases from the basal cradle-loop fold. The differences to a basal RIFT barrel necessary for nucleotide binding are concentrated in the N-terminal half of the protein. Acquisition of an aromatic residue in the al-β2 loop allows this residue to form the sides of the cytosine binding pocket in conjunction with the hydrophobic anchoring residue of the strand (YxGTLN motif). The buried orientation of these two residues induces a γ turn between them, resulting in the formation of a backbone hydrogen bond between the aromatic residue and the end of β2. This bond creates a shift in the position of α1’, thus providing the last adjustment necessary to accommodate the cytosine moiety, without requiring any but conformational changes in the C-terminal half of the protein. Correspondingly, archaeal RFKs have a divergent N-terminal half relative to basal cradle-loop barrels, whereas their C-terminal half is indistinguishable from these in its conservation patterns. From these observations we conclude that CTP binding was an ancestral property of riboflavin kinases.

Once nucleotide binding was established, the presence of a glutamate in β2’ would have allowed transfer of the γ-phosphate to a range of substrates, dependent on the ability of the second cradle loop to assume a closed conformation. This cradle loop may also have conferred initial specificity toward substrates such as riboflavin, as judged by its interactions with flavins in present-day structures. The lineage of archaeal RFKs would have diverged at this point from that of bacteria and eukaryotes. Subsequently the latter altered their nucleotide specificity to ATP, resulting in the divergence of the C-terminal half. The wider space for the adenine moiety was obtained by mutating the two large hydrophobic residues to smaller residues and by a deletion in α1’, which abolished this helix and converted the region into an extended loop. Both lineages evolved convergently toward higher specificity for riboflavin. We conclude this from the considerable differences in the geometry of flavin binding and from the fact that the concomitant structural elaborations are superficially similar but show no sequence similarity and have an entirely different topological origin.

In conclusion, archaeal RFKs show the properties we would expect for an evolutionary bridge between basal RIFT barrels and one of the ancient enzyme families with this fold. This relationship also links basal RIFT barrels that utilize a DNA-binding site between the two cradle loops, with enzymes that additionally or exclusively use a binding site between the first cradle loop and the capping α-helix. At present, there is no evidence for a homologous relationship to other flavin-dependent RIFT barrel enzymes, such as riboflavin synthase. However, this does not imply analogy but may simply reflect the absence of supporting data. Our study emphasizes the importance of evolutionary intermediates in tracing the origins of structural and functional diversity in proteins.

**EXPERIMENTAL PROCEDURES**

**Bioinformatics**

We searched the nonredundant protein sequence database at NCBI, nr, for homologs of MJ0056, AbrB-N (1YFB_A), PhS018 (2GLW), VatN (1CZ5_A, 1-91), fission yeast riboflavin synthase (1K2L_A, 1-92), and fission yeast riboflavin kinase (1N08_A) by using HHsenser.
Structure
Structure of the Archaeal Riboflavin Kinase Mj0056

(Söding et al., 2006), a tool for exhaustive transitive profile searches with Hidden Markov Model (HMM) comparisons. The searches were done in default settings. HHsenser returns two sets of alignments, a strict and a permissive one. We pooled the strict sets to obtain 597 sequences, which we clustered in CLANS (Frickey and Lupas, 2004) with BLAST 2.2.16 as a comparison tool on a 2 GHz 32-bit Intel CPU. Clustering was done to equilibrium in 2D at a P-value cut off of 1.0e-04 with default settings, except for attract value = 20 and attract exponent = 2 (Figure 2). The multiple alignment in Figure 3 was generated interactively with MACAW (Schuler et al., 1991).

Sample Preparation
The DNA sequence encoding Mj0056 (gi:2128102) was amplified from genomic DNA of M. janaschii by polymerase chain reaction (PCR) and cloned into the pET-30b expression vector (Novagen) with Nde I and Hind III restriction sites. The identity of the construct was confirmed.

Figure 8. Structural Comparison of Archaeal and Bacterial/Eukaryotic RFKs
Mj0056-MgGDP-FMN (A) is compared to HsRFK-MgADP-FMN (1Q9S) (B). A superposition of the nucleotide binding sites (C) illustrates how the absence of α1' in HsRFK accommodates the larger adenosine moiety. The superimposition in (D) shows the FMN binding site, highlighting the structural equivalence of α1 in Mj0056 to the C-terminal extension in HsRFK.
by DNA sequencing. The protein was expressed in *E. coli* C41 (DE3) or C41 (DE3) RIL cells, which were grown at 37°C in LB medium containing 75 mg/l Kanamycin, induced at an OD 600 of 0.6 with 1 mM isopropylthiogalactoside (IPTG), and harvested after 4 hr. Uniformly 15N- and 15N, 13C-labeled Mj0056 was made by growing bacteria in M9 minimal medium by using 15NH4Cl (0.7 g/l) or 13C6-glucose (2 g/l) as the sole nitrogen and carbon sources, respectively. For NMR studies, samples were purified with a combination of anion (Mono-Q, Amersham) and cation exchange (SP Sepharose FF, Amersham) chromatography with 20 mM Tris-HCl (pH 7.0) buffer and salt gradient from 50 mM to 1 M NaCl. Fractions containing protein of interest were pooled and applied to a gel-sizing chromatography (Sephadex G-75, Amersham) equilibrated in buffer containing 30 mM sodium phosphate, 150 mM NaCl (pH 7.4). Purified labeled protein was concentrated to 10 mg/ml by ultrafiltration with Vivaspin 10 kDa membranes, and 0.02% (w/v) NaN3 was added to the sample. Monomeric protein was confirmed by NMR diffusion measurements.

Three protein preparations were used for crystallization trials (Table 4). The first employed the purification strategy used to obtain NMR samples, followed by dialysis against a buffer containing 20 mM MOPS (pH 7.25), 120 mM NaCl, and 0.02% (w/v) NaN3 (Prep. A). A second purification strategy was applied after the function of Mj0056 became clear, which met the requirement of low phosphate conditions for enzyme assays. In this strategy, soluble fractions of cellular extracts were subjected to an anion exchange (Mono Q, Amersham) and followed by a cation exchange chromatography (SP Sepharose FF, Amersham). Bound protein was eluted by a linear sodium chloride gradient from 50 mM to 1 M in Tris buffer (pH 6.8). Monitoring by SDS-PAGE indicated the presence of the target protein in a yellow and an uncolored fraction, which were pooled separately. Both pools were heated to 80°C for 20 min to precipitate thermolabile *E. coli* proteins, cooled to 4°C, and centrifuged. The yellow fraction was concentrated by ultrafiltration by using Vivaspin 10 kDa membranes and used directly for crystallization trials without additives (Prep. B). The uncolored fraction was applied to a Superdex G-75 preparative column that had been equilibrated in 25 mM HEPES buffer (pH 7.4) containing 100 mM NaCl. Eluted fractions were tested by SDS-PAGE, combined, and concentrated with Vivaspin 10 kDa concentrators (Prep. C). The resulting

---

**Table 4. Structural Statistics for Mj0056 Crystal Structures**

<table>
<thead>
<tr>
<th></th>
<th>Mj0056-MgCDP</th>
<th>Mj0056-MgCDP-FMN</th>
<th>Mj0056-NaCDP-PO4</th>
<th>Mj0056-PO4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Collection Statistics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>I4,1</td>
<td>P2,2,2,1</td>
<td>P4,2,2</td>
<td>P4,2,2</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9762</td>
<td>1.2141</td>
<td>0.9762</td>
<td>1.0716</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>33.1 (1.70)</td>
<td>38.5 (2.4)</td>
<td>38.3 (2.70)</td>
<td>19.9 (3.0)</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>20274 (3209)</td>
<td>12903 (2091)</td>
<td>9448 (1457)</td>
<td>6951 (1183)</td>
</tr>
<tr>
<td>Redundancy (%)</td>
<td>7.82 (6.30)</td>
<td>5.50 (5.62)</td>
<td>5.50 (5.60)</td>
<td>4.28 (4.43)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.8 (99.4)</td>
<td>98.6 (98.5)</td>
<td>99.5 (98.1)</td>
<td>99.3 (100)</td>
</tr>
<tr>
<td>R_sym (%)</td>
<td>5.1 (46.8)</td>
<td>13.1 (68.7)</td>
<td>7.6 (91.8)</td>
<td>13.4 (32.6)</td>
</tr>
<tr>
<td>I/σ(I)</td>
<td>23.1 (4.35)</td>
<td>8.67 (2.25)</td>
<td>16.57 (2.02)</td>
<td>8.43 (4.39)</td>
</tr>
<tr>
<td><strong>Refinement Statistics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>I4,1</td>
<td>P2,2,2,1</td>
<td>P4,2,2</td>
<td>P4,2,2</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>20–1.7</td>
<td>20–2.4</td>
<td>20–2.7</td>
<td>20–3.0</td>
</tr>
<tr>
<td>R_cryst (%)</td>
<td>16.1</td>
<td>22.7</td>
<td>20.9</td>
<td>23.0</td>
</tr>
<tr>
<td>R_free (%)</td>
<td>21.2</td>
<td>32.3</td>
<td>26.6</td>
<td>28.2</td>
</tr>
<tr>
<td>Nonhydrogen atoms</td>
<td>1283</td>
<td>2336</td>
<td>1167</td>
<td>1078</td>
</tr>
<tr>
<td>Mean B value (Å²)</td>
<td>22.2</td>
<td>42.4</td>
<td>60.5</td>
<td>49.9</td>
</tr>
<tr>
<td>Rmsd bond length (Å)</td>
<td>0.012</td>
<td>0.015</td>
<td>0.014</td>
<td>0.014</td>
</tr>
<tr>
<td>Rmsd bond angle (deg.)</td>
<td>1.55</td>
<td>1.70</td>
<td>1.48</td>
<td>1.54</td>
</tr>
<tr>
<td><strong>Crystallization Conditions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein concentration</td>
<td>10 mg/ml</td>
<td>10 mg/ml</td>
<td>20 mg/ml</td>
<td>10 mg/ml</td>
</tr>
<tr>
<td>Protein solution additives</td>
<td>10 mM MgCDP, 10 mM MgADP, saturated with riboflavin</td>
<td>10 mM MgCDP, 10 mM FMN</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>Reservoir solution</td>
<td>35% v/v MPD, 0.1 M imidazole</td>
<td>20% v/v PEG 8000, 0.2 M sodium iodide</td>
<td>40% v/v ethylene glycol, 0.1 M phosphate-citrate (pH 4.2), 0.2 M NH4SO4</td>
<td>2 M NH4SO4, 10 mM Zn²⁺</td>
</tr>
</tbody>
</table>

---

a Figures in parenthesis refer to the highest resolution shell.
b See Experimental Procedures.
c The crystals from this preparation contained endogenous CDP. The yellow color could be tentatively attributed to a flavin by UV/VIS spectroscopy (data not shown), but no flavin was observed in the resulting structure.
solution was used both for crystallizations with various additives and for enzymatic assays. The oligomeric state of pure uncoupled Mj0056 obtained by Prep. C was analyzed on a calibrated analytical gel-sieving column (Superose 12, Amersham).

Riboflavin Kinase Assays
Riboflavin kinase activity was assayed in reaction mixtures containing 40 mM Tris/HCl (pH 8) buffer, 50 mM NaCl, 5 mM MgCl₂, 1 mM DTT, 50 μM riboflavin, 3 mM nucleotide (ATP, CTP, GTP, or UTP), and 1 μM Mj0056. Reaction mixtures were incubated at various temperatures (25°C, 37°C, 50°C, 70°C, and 85°C) for 60 min and subsequently cooled to 4°C. Controls were processed identically but in the absence of enzyme. FMN controls contained 50 μM FMN instead of riboflavin. Riboflavin, FMN, ATP, CTP, and GTP were obtained from Sigma, UTP from Roth. One hundred microliters of reaction mixtures were desalted prior to NMR analysis with C18 extraction tips (Rappsilber et al., 2003) and eluted in 50 μL 50% acetonitrile/0.1% formic acid. MS data was acquired on an HCT Ultra ion trap (Bruker-Daltonics, Bremen) by electro-spray ionization in alternating positive and negative ion mode.

Solution Structure of Mj0056
All spectra were recorded at 300 K on Bruker DMX600, DMX750, and DMX900 spectrometers. Backbone sequential assignments were completed with standard triple resonance experiments implemented by using selective proton flipback techniques for fast pulsing (Diercks et al., 2005). Aliphatic side-chain assignments were completed by a combination of HCCH-TOSY and CCH-COSY experiments, while aromatic assignments were made by linking aromatic spin systems to the respective C²H₂ protons in a 2D-NOESY spectrum. Stereospecific assignments and the resulting Cα rotamer assignments were determined for 54 of 93 prochiral C²H₂ protons and for the C²H₄ groups of 7 of 8 valine residues. Assignments of Cα rotamers were also available all isoleucine residues and 3 of 4 threonine residues. Assignments of Cα rotamers were made for all isoleucine and 9 of 13 leucine residues by consideration of patterns of intraresidue NOE connectivities, leading to stereospecific assignment of the prochiral leucine C²H₄ groups.

Distance data were derived from a set of five 3D-NOESY spectra, including the heteronuclear edited NH²-, C²H²-, and CH²-NOESY spectra (Diercks et al., 1999) in addition to conventional 1H²- and 1H²-CH²-QQ-NOESY spectra and a 2D-NOESY spectrum recorded on an unlabelled sample. NOESY crosspeaks were converted into distance ranges after rescaling of intensities in the 3D spectra according to corresponding HSQC intensities. Crosspeaks were divided into four classes: strong, medium, weak, and very weak, which resulted in restraints on upper distances of 2.7, 3.2, 4.0, and 5.0 Å, respectively.

Lower distance restraints were also included for very weak or absent sequential H²-H³ crosspeaks with a minimum distance of 3.2 Å and medium intensity or weaker sequential and intraresidue H²-H³ crosspeaks with a minimum distance of 2.7 Å. Allowances for the use of pseudotautomers (by using ψ ± 5°) were added for methyl groups and nonstereospecifically assigned methylene groups. Dihedral angle restraints were derived for backbone φ and ψ angles based on Cα, C², C³, and H² chemical shifts with the program TALOS (Cornilescu et al., 1999). Restraints were applied for the 94 high-confidence predictions found by the program with the calculated range ±5°. Dihedral restraints were also applied for side-chain rotamers identified through stereospecific assignment with a tolerance of ±3°, with the exception of proline residues where the Cα rotamer was restrained to ±30° with a tolerance of ±15°. Direct coupling constant restraints were included for the backbone φ angles of 91 residues based on ¹³JHNH coupling constants measured from an HNHA experiment and for 90 backbone ψ angles based on ¹³JHNH coupling constants measured from an HNHB experiment (Wang and Bax, 1995). Hydrogen bond restraints were applied for 58 residues in secondary structure with low water exchange rates, as judged by the strength of water exchange crosspeaks in the ¹H-NOESY spectrum and where donor-acceptor pairs were consistently identified in preliminary calculations. The restraints were applied via inclusion of pseudovacovalent bonds as described by Truffaut et al. (2001).

Structures were calculated with XPLOR (NIH version 2.9.3) by standard protocols. Structures calculated in an initial simulated annealing protocol were refined in two further slow cooling stages, the first including a conformational database potential and the second with the force constant on peptide bond planarity relaxed to 50 kcal/mol/Å². For the final set, 50 structures were calculated and 21 chosen on the basis of lowest restraint violations. An average structure was calculated and regularized to give a structure representative of the ensemble (used here for all figures). Structures were validated with PROCHECK (Laskowski et al., 1993), WHATCHECK (Hooft et al., 1996), and MOLPROBITY (Lovell et al., 2003). Refinement was carried out by comparison of experimental and back-calculated ¹H-NOESY, CH²-CH²-NOESY, NH²-NOESY spectra (in-house software). This process resulted in adjustment of side-chain rotamers for several residues. As crystals structures were available by this stage, back calculation was simultaneously used to justify any differences between the solution and crystal structures.

X-Ray Crystallography
In all crystallization trials, 400 nl of protein solution were mixed with 400 nl of reservoir solution in 96-well Corning 3550 plates with 75 μl reservoir volume by using the honeybee 961 crystallization robot (Genomic Solutions). Drop images were obtained with the RockImager 54 (Nonob Consulting). Diffraction images were integrated and scaled with the XDS program package (Kabsch, 1993). A preliminary NMR model was used as an initial search model for structure solution of the Mj0056-Po₄ complex by molecular replacement with the program MOLREP (Vagin and Teplyakov, 2004). The structure Mj0056-NA²-Po₄ was solved on the basis of the refined coordinates of Mj0056-Po₄ and subsequently used as a search model for Mj0056-Mg²⁺-CDP and Mj0056-Mg²⁺-FMN. Structures were built and refined with the programs COOT (Emsley and Cowtan, 2004) and REFMACS (Murshudov et al., 1999). Validation with PROCHECK (Laskowski et al., 1993) and WHATCHECK (Hooft et al., 1996) showed good geometries for all structures. Refinement statistics are summarized in Table 4.

Supplemental Data
Supplemental Data show the positive and negative ion mode MS spectra of the riboflavin kinase reaction in the presence of the donor nucleotides CTP, ATP, and GTP at 85°C and are available at http://wwwstructure.org/cgi/content/full/15/12/1577/DC1/.

ACKNOWLEDGMENTS
The authors thank Prof. Horst Kessler and the staff of the Bavarian Nuclear Magnetic Resonance Centre at the Technical University, Munich, for access to spectrometers and technical support. We also thank Johannes Söding for discussion of bioinformatics aspects. Protein purification and assays were performed by M.A., S.D., and J.M.; mass spectrometry measurements by G.S. and M.A.; bioinformatic analyses by V.A., K.K.K., and A.N.L.; and NMR structure determination by M.C. and V.T.; and crystallography by M.D.H. and K.Z. The authors declare they have no competing financial interests.

Received: August 22, 2007
Revised: September 26, 2007
Accepted: September 26, 2007
Published: December 11, 2007
REFERENCES


Accession Numbers

The coordinates for the NMR ensemble (accession codes 2P3M) and the following crystal structures have been deposited in the Protein Data Bank: Mj0056-PO4 (2VBS), Mj0056-MgCDP (2VBU), Mj0056-NaCDP-PO4 (2VBT), and Mj0056-MgCDP-FMN (2VBV).