Structure
Article

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 Coles[1,*]
[1]Department of Protein Evolution, Max-Planck-Institute for Developmental Biology, 72076 Tübingen, Germany
[2]Computational Chemistry Group, GlaxoSmithKline, Collegeville, PA 19426, USA
[3]Department of Biochemistry, Max-Planck-Institute for Developmental Biology, 72076 Tübingen, Germany
[4]These authors contributed equally to this work.
[5]Present 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 βαβ-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, *Methanocaldococcus jannaschii* Mj0056, which is an archaeal riboflavin kinase. Mj0056 is unusual in utilizing CTP rather than ATP as the donor nucleotide, and sequence conservation in the relevant residues suggests that this is a general feature of archaeal riboflavin kinases.

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, 1Q9S [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 *Methanocaldococcus 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 Ef-Tu and related translation factors, ribosomal protein L3, the N-domain of the F_1 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,

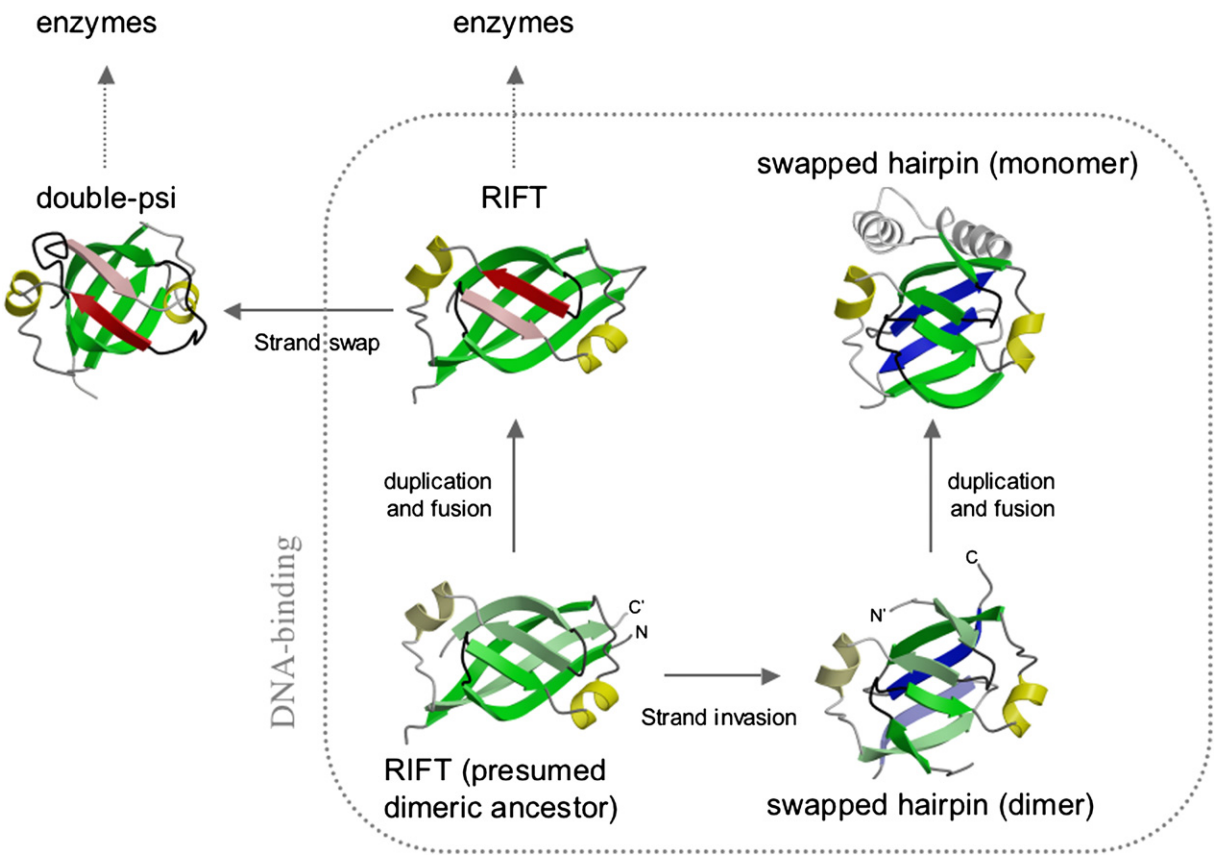

Figure 1. A Scenario for Cradle-Loop Barrel Evolution

Examples are shown for the three folds for which homologous relationships have been established. For dimeric examples, the monomers are distinguished by light and dark colors. The invading strands of the swapped-hairpin barrel are shown in blue, while the swapped strands of the double-psi barrel are in red. The example proteins are: VatN-N (double psi), PhS018 (RIFT), MraZ (monomeric swapped hairpin), and AbrB-N (dimeric swapped hairpin). The structure of the presumed dimeric RIFT barrel ancestor is a homology model of Orf5 from the pME2200 plasmid of *Methanothermobacter thermautotrophicus* (MTpME2200 Orf5) based on PhS018. For details of this scenario see (Coles et al., 2006).

PhS018. 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 (SENSER) (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

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

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

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).

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 assayed 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.

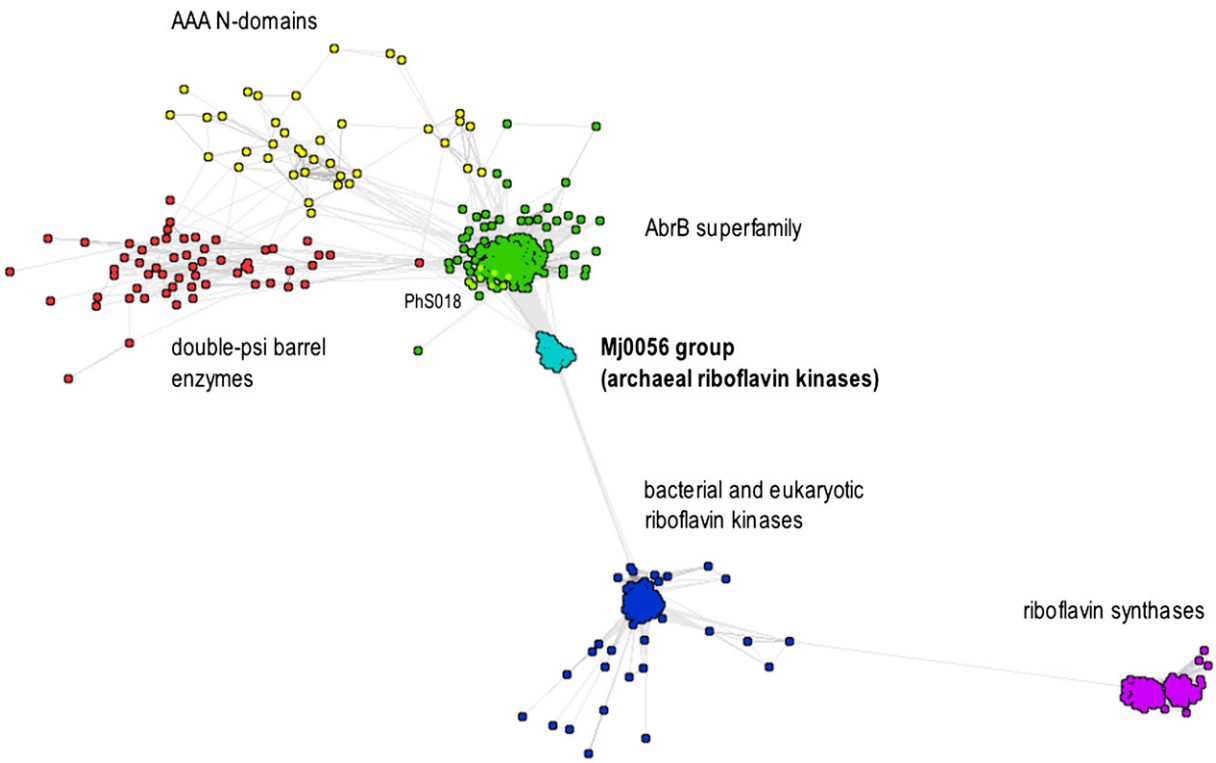

Figure 2. Cluster Map of Mj0065 Homologs
The map, obtained with CLANS (see Experimental Procedures), shows Mj0056 in the context of double-psi barrels (AAA N domains, double-psi barrel enzymes), swapped-hairpin barrels (AbrB superfamily), and RIFT barrels (PhS018 group, riboflavin kinases, riboflavin synthases).

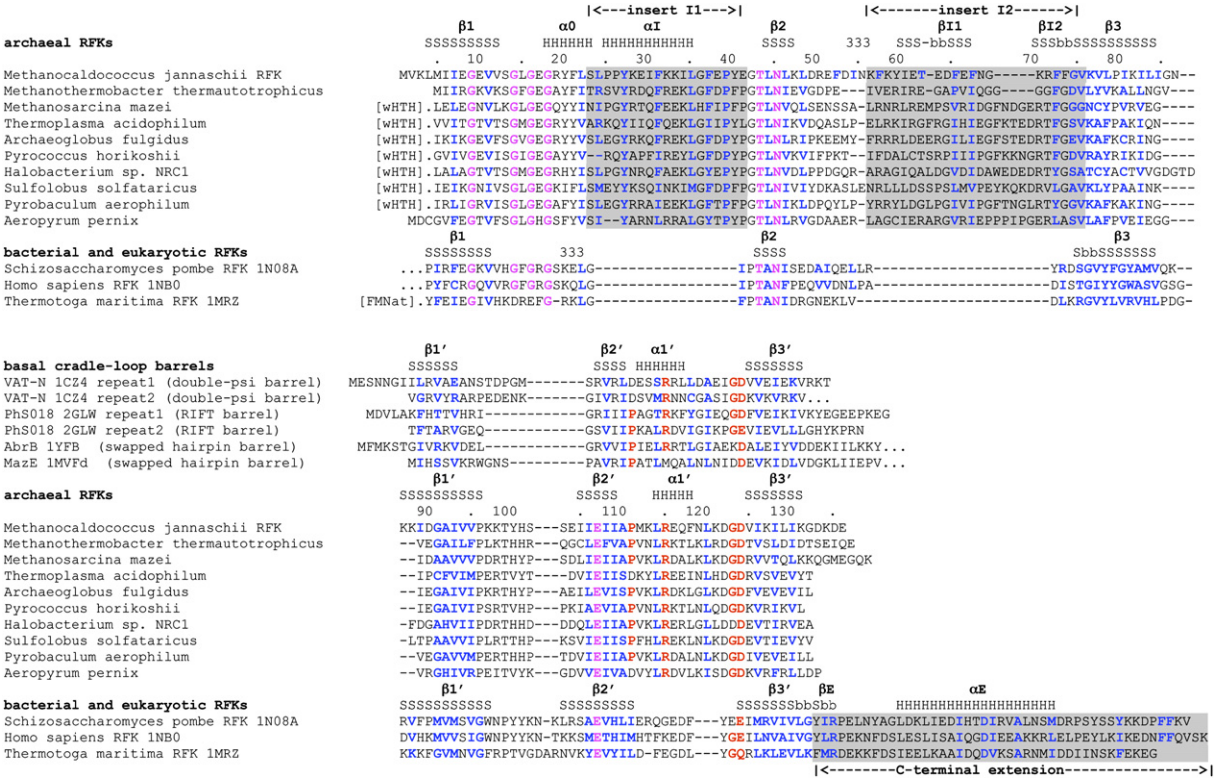

Figure 3. Multiple Alignment of Archaeal Riboflavin Kinases and Their Homologs

The N- and C-terminal halves of the cradle-loop barrel fold are shown in the upper and lower panel, respectively. The alignment shows a phylogenetically representative selection of archaeal riboflavin kinases (RFKs), as well as the three bacterial and eukaryotic RFKs of known structure; sequence motifs shared by sequences of all three kingdoms are colored magenta. Many archaeal RFKs contain an N-terminal winged helix-turn-helix DNA-binding domain, indicated by the abbreviation: wHTH. Thermotoga RFK contains an N-terminal FMN adenylyltransferase domain, denoted by FMNat, and both eukaryotic and bacterial RFKs contain a C-terminal extension, which is structurally equivalent to the two insert regions of archaeal RFKs. The C-terminal half of archaeal RFKs is most similar in sequence to a group of cradle-loop barrels that we have described previously (Coles et al., 2006); these are shown above the C-terminal half, and shared sequence motifs are colored red. In all sequences, residues buried in the core of the structures, as judged by their relative solvent accessibility computed by the SARIG server (http://www.weizmann.ac.il/SARIG/), are colored blue. The secondary structure is shown above the sequences (H, helix; S, strand; b, beta bulge; 3, 3_{10} helix).

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 electospray 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 $^{3}J_{HNH\alpha}$ and $^{3}J_{NH\alpha i-1}$ 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

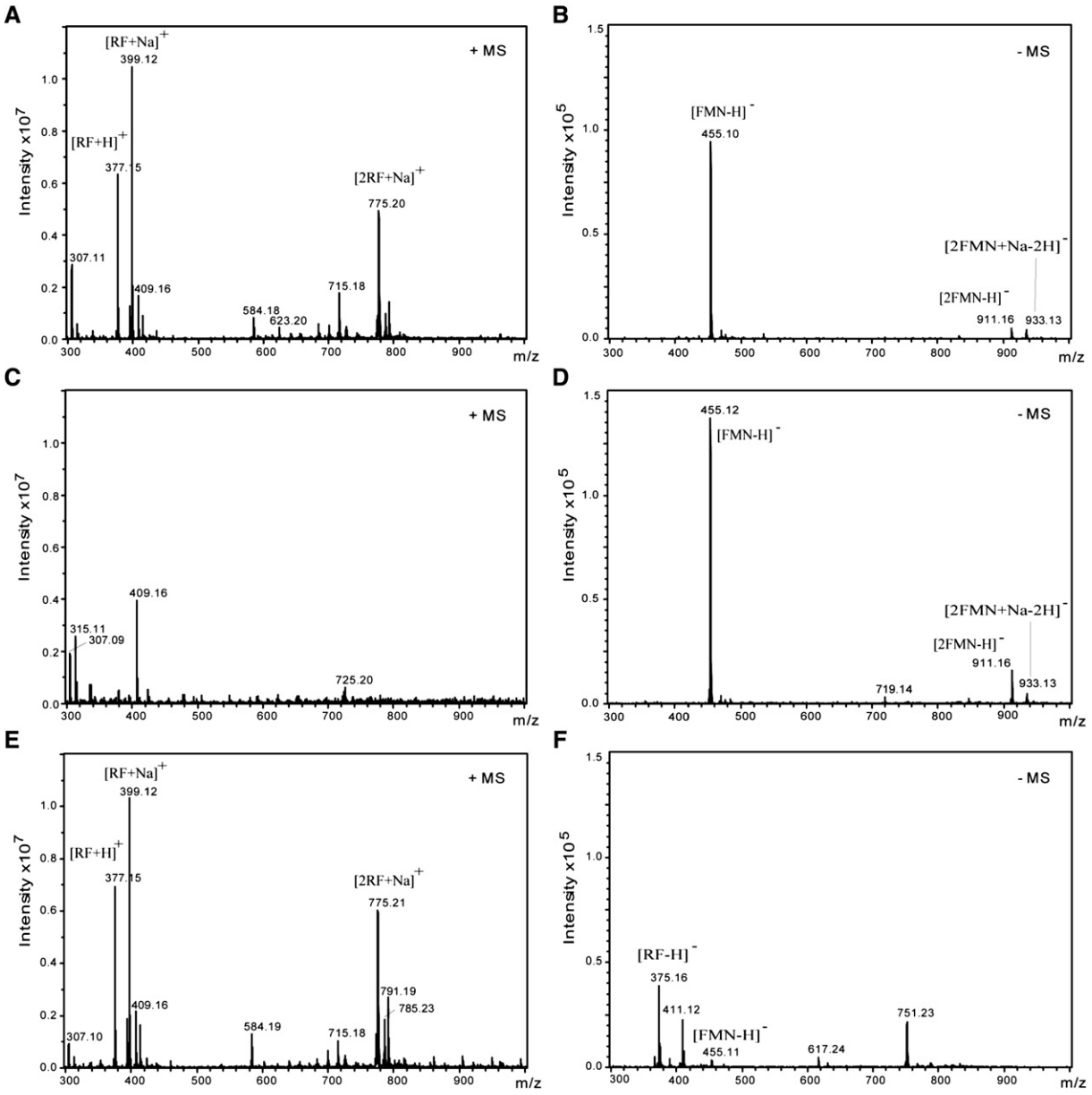

Figure 4. MJ0056 Is a CTP-Specific Riboflavin Kinase

(A and B) MS spectra of the riboflavin and FMN controls measured in positive and negative ion mode, respectively. Adducts of riboflavin and FMN are annotated in angular brackets next to the corresponding peaks.

(C and D) The positive and negative ion mode spectra of samples containing enzyme, riboflavin, and CTP. Complete turnover of riboflavin to FMN is observed (60 min, 50°C).

(E and F) The positive and negative ion mode MS spectra of samples containing enzyme, riboflavin, and ATP.

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 βαβ 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

Table 2. Apo-Mj0056 Solution Structure Statistics and Atomic Rmsds

Structural Statistics				
Rmsd from distance restraints (Å)[a]	SA		<SA>$_r$	
All (951)	0.012 ± 0.001		0.011	
Intraresidue (145)	0.000 ± 0.001		0.000	
Interresidue sequential (299)	0.009 ± 0.001		0.008	
Medium range (80)	0.020 ± 0.003		0.017	
Long range (315)	0.016 ± 0.001		0.015	
H bond (112)	0.000 ± 0.000		0.000	
Rmsd from dihedral restraints (399)	0.094 ± 0.002		0.092	
Rmsd from J-coupling restraints (Hz) (180)	0.387 ± 0.004		0.386	
H bond restraints; averages (Å/deg.)[b] (58)	2.18 ± 0.12/11.9 ± 5.8		2.13 ± 0.13/11.9 ± 6.2	
H bond restraints; min-max (Å/deg.)	1.93–2.45/1.3–32.3		1.88–2.43/0.72–33.6	
Deviations from Ideal Covalent Geometry				
Bonds (Å × 10^{-3})	7.17 ± 0.01		7.14	
Angles (deg.)	0.78 ± 0.01		0.77	
Impropers (deg.)	1.46 ± 0.04		1.51	
Structure Quality Indicators[c]				
Ramachandran map regions (%)	98.0/100.0/0.0		98.5/100.0/0.0	
Steric clashes >0.4 Å per 1,000 atoms	0.0		0.0	
Atomic Rmsd (Å)[d]				
	SA versus <SA>		SA versus <SA>$_r$	
	Backbone	All	Backbone	All
All residues	0.99 ± 0.32	1.44 ± 0.323	1.28 ± 0.32	1.80 ± 0.32
Secondary structure[e]	0.25 ± 0.10	0.61 ± 0.08	0.39 ± 0.08	0.73 ± 0.09
<SA> versus <SA>$_r$[f]	0.30	0.48		

Structures are labeled as follows: SA, the set of 19 final simulated annealing structures; <SA>, the mean structure calculated by averaging the coordinates of SA structures after fitting over secondary structure elements; <SA>$_r$, the structure obtained by regularizing the mean structure under experimental restraints.
[a] Numbers in brackets indicate the number of restraints of each type.
[b] Hydrogen bonds were restrained by treating them as pseudocovalent bonds (see Experimental Procedures section). Deviations are expressed as the average distance/average deviation from linearity for restrained hydrogen bonds.
[c] Determined with the program MolProbity (Lovell et al., 2003). Percentages are for residues in favored (98%), allowed (99.8%), and disallowed regions of the Ramachandran map.
[d] Based on heavy atoms superimpositions.
[e] Defined as residues L4-S13, L24-G96, and E104-G132.
[f] Rms difference for superimposition over ordered residues.

conserved hydrophobic and hydrogen bonding interactions. One significant difference to the basal βαβ element is in the entrance to the α 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 β 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 β2′, 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 α1 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 α-helix (αI; P25-L35) and contributing to a structured connector (G36-T43), which leads into β2. The second insertion consists of a short 3_{10} helix and a β-hairpin (βI1-βI2), which continues unbroken into strand β3.

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

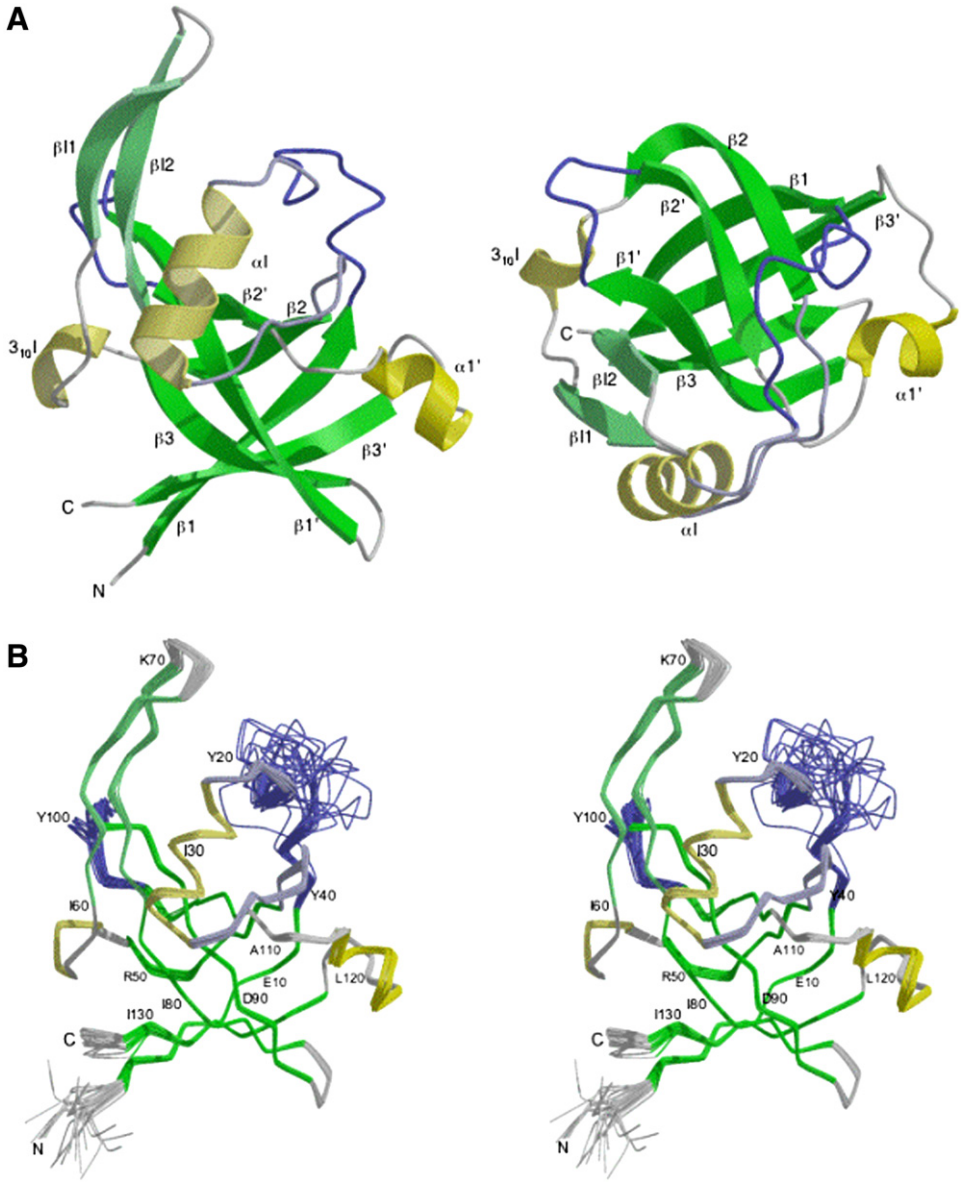

Figure 5. The Solution Structure of Apo-Mj0056

(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).

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-PO_4, 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 Mg^{2+} 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

Table 3. Summary of Mj0056 Structures

Name	First Cradle Loop[a]	Second Cradle Loop	Oligomer[b]	Space Group	Resn. (Å)	PDB[c]
Apo-Mj0056	flexible	flexible	monomer	solution	-	2P3M
Mj0056-PO_4	unstructured	closed	dimer	$P4_32_12$	3.0	2VBS
Mj0056-MgCDP[d]	α0 helix	open	monomer	$I4_1$	1.7	2VBU
"	"	"	dimer	$P4_12_12$	2.6	-
Mj0056-NaCDP-PO_4	α0 helix	closed	dimer	$P4_32_12$	2.7	2VBT
"	"	"	"	$P3_221$	3.3	-
Mj0056-MgCDP-FMN	α0 helix	closed	dimer	$P2_12_12_1$	2.4	2VBV

Full structural statistics for solution and crystal structures are shown in Tables 2 and 4, respectively.
[a] Indicates the state of first cradle loop in the region of the transient α0 helix.
[b] All dimeric forms show a similar dimerization via pairing of the β1 strands.
[c] Where structures were solved in two space groups, the higher resolution structure was chosen for deposition.
[d] Similar to the New York SGC Research Center for Structural Genomics structure for Mj0056-NaCDP (2OYN).

structure of the New York SGC Research Center for Structural Genomics (2OYN).

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 β-barrel, the transient helix α0, 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 π-stacking interactions with the side chain of a conserved aromatic residue (F21) on α0, and hydrogen bonds to the backbone of F73 on βl2 and to the side chain of Y27 on αl via a bridging water molecule. The 4′ and 5′ oxygens of FMN contact the two carboxyl oxygens of the invariant glutamate on β2′ (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 α0 is formed

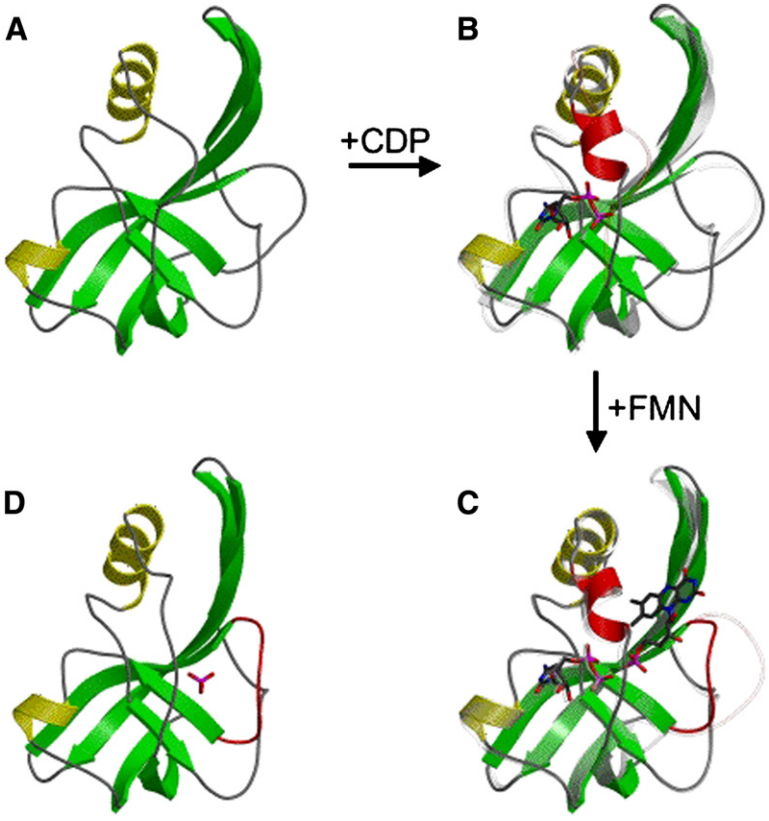

Figure 6. The Proposed Sequential Substrate Binding Cycle of Mj0056 and Concerted Conformational Changes
The binding of the nucleotide to the apo structure (A and B) leads to the formation of the transient helix α0 in the first cradle loop; with the subsequent binding of riboflavin/FMN (B and C), the 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 α0. In the structure Mj0056-NaCDP-PO_4 (not shown), the cradle loop conformations are similar to those shown in (C).

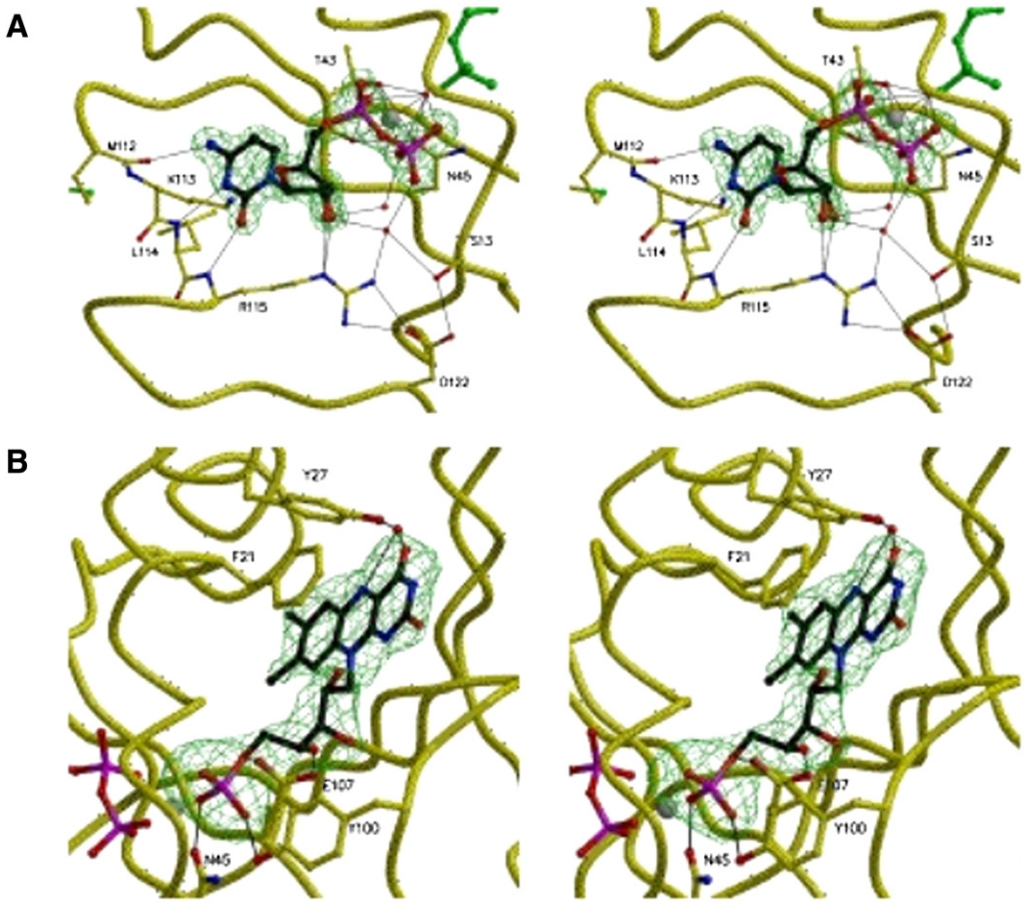

Figure 7. Ligand Binding in Mj0056

(A) The binding mode of MgCDP in the Mj0056-MgCDP structure (1.7 Å). Three hydrogen bonds from the cytosine moiety to the backbone illustrate the high specificity for this nucleotide. Superposed in green is the phosphate group of FMN from Mj0056-MgCDP-FMN, which has one of its oxygens in the position of a water molecule in the first magnesium coordination sphere. The *Fo-Fc* omit map for CDP is contoured at 5σ.

(B) Binding details of FMN in the Mj0056-MgCDP-FMN structure (2.4 Å). The *Fo-Fc* omit map for FMN is contoured at 2.5σ. Selected hydrogen bonds are shown as thin black lines, water molecules as red spheres, Mg^{2+} is gray.

only in Mj0056-NaCDP-PO_4. Thus, we observe two major conformational changes upon ligand binding, both affecting the FMN binding site: formation of helix α0 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 β1 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 *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 β 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 $\alpha1'$ and the arginine coordinating the ribose hydroxyl groups. Instead, ATP-dependent RFKs use small side chains and a wider loop in place of $\alpha1'$ 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 isoalloxazine 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 $\alpha0$ of the former is present as a shorter 3_{10} 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 $\alpha0$ in all crystal forms containing the 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 aI-$\beta2$ 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 $\beta2'$. This bond creates a shift in the position of $\alpha1'$, 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 $\beta2'$ 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 $\alpha1'$, 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 (1KZL_A: 1–92), and fission yeast riboflavin kinase (1N08_A) by using HHsenser

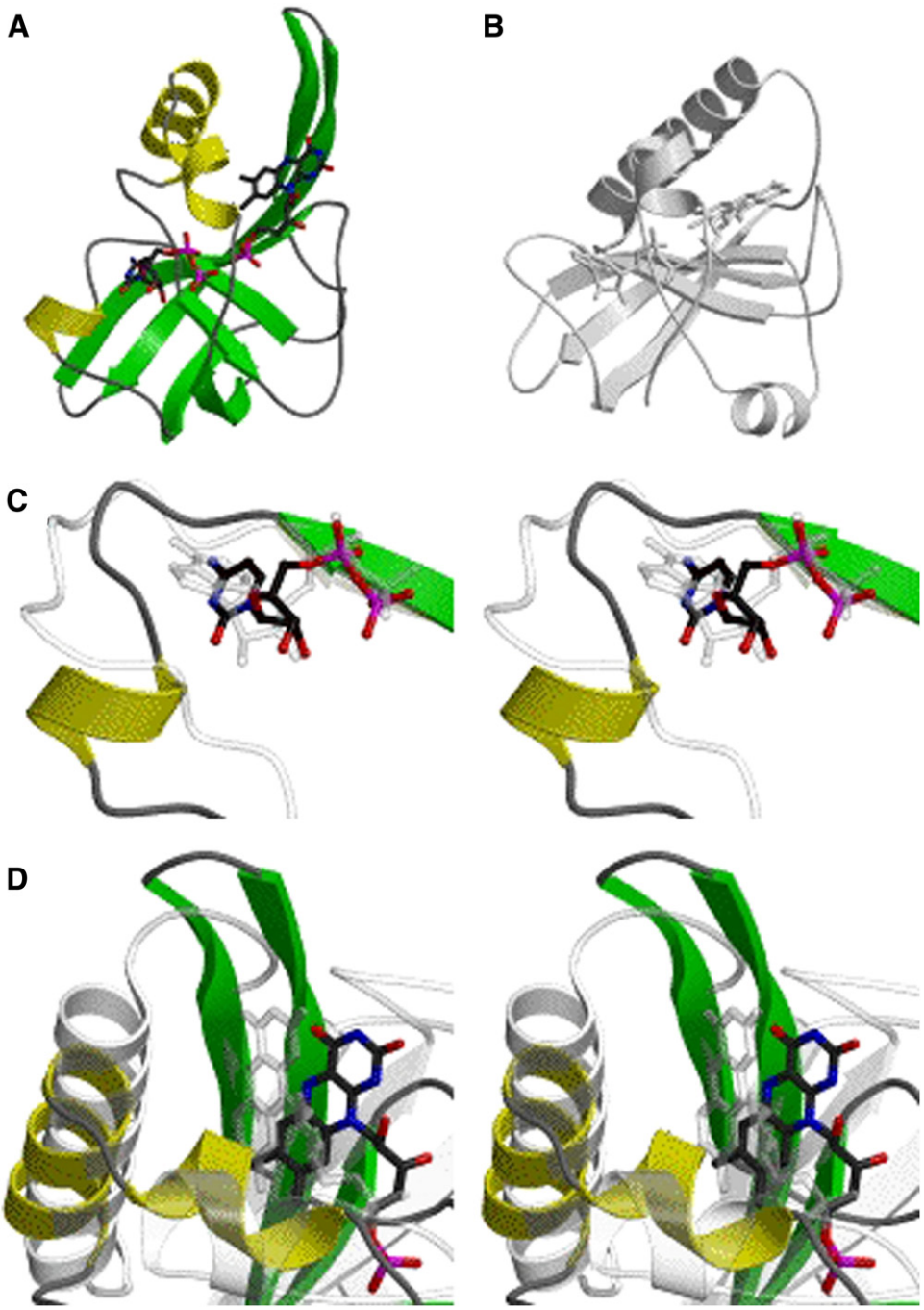

Figure 8. Structural Comparison of Archaeal and Bacterial/Eukaryotic RFKs

Mj0056-MgCDP-FMN (A) is compared to HsRFK-MgADP-FMN (1Q9S) (B). A superposition of the nucleotide binding sites (C) illustrates how the absence of $\alpha 1'$ in HsRFK accommodates the larger adenosine moiety. The superimposition in (D) shows the FMN binding site, highlighting the structural equivalence of $\alpha 1$ in Mj0056 to the C-terminal extension in HsRFK.

(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

Table 4. Structural Statistics for Mj0056 Crystal Structures

	Mj0056-MgCDP	Mj0056-MgCDP-FMN	Mj0056-NaCDP-PO_4	Mj0056-PO_4
Data Collection Statistics[a]				
Space group	$I4_1$	$P2_12_12_1$	$P4_32_12$	$P4_32_12$
Wavelength (Å)	0.9762	1.2141	0.9762	1.0716
Resolution (Å)	33.15–1.70 (1.80–1.70)	38.58–2.4 (2.55–2.4)	38.32–2.70 (2.87–2.70)	19.99–3.00 (3.20–3.00)
Unique reflections	20274 (3209)	12903 (2091)	9448 (1457)	6951 (1183)
Redundancy	7.82 (6.30)	5.50 (5.62)	5.50 (5.60)	4.28 (4.43)
Completeness (%)	99.8 (99.4)	98.6 (98.5)	99.5 (98.1)	99.3 (100)
R_{sym}	5.1 (46.8)	13.1 (68,7)	7.6 (91.8)	13.4 (32.6)
$I/\sigma(I)$	23.1 (4.35)	8.67 (2.25)	16.57 (2.02)	8.43 (4.39)
Refinement Statistics				
Space group	$I4_1$	$P2_12_12_1$	$P4_32_12$	$P4_32_12$
Resolution (Å)	20–1.7	20–2.4	20–2.7	20–3.0
R_{cryst} (%)	16.1	22.7	20.9	23.0
R_{free} (%)	21.2	32.3	26.6	28.2
Nonhydrogen atoms	1283	2336	1167	1078
Mean B value ($Å^2$)	22.2	42.4	60.5	49.9
Rmsd bond length (Å)	0.012	0.015	0.014	0.014
Rmsd bond angle (deg.)	1.55	1.70	1.48	1.54
Crystallization Conditions				
Protein preparation[b]	Prep. C (uncolored)	Prep. C (uncolored)	Prep. B (yellow fraction)	Prep. A
Protein concentration	10 mg/ml	10 mg/ml	20 mg/ml	10 mg/ml
Protein solution additives	10 mM MgCDP, 10 mM MgADP, saturated with riboflavin	10 mM MgCDP, 10 mM FMN	none[c]	none
Reservoir solution	35% v/v MPD, 0.1 M imidazole	20% w/v PEG 8000, 0.2 M sodium iodide	40% v/v ethylene glycol, 0.1 M phosphate-citrate (pH 4.2), 0.2 M NH_4SO_4	2 M NH_4SO_4, 10 mM Zn^{2+}

[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.

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 ^{15}N- and ^{15}N, ^{13}C-labeled Mj0056 was made by growing bacteria in M9 minimal medium by using $^{15}NH_4Cl$ (0.7 g/l) or $^{13}C_6$-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) NaN_3 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) NaN_3 (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

solution was used both for crystallizations trials with various additives and for enzymatic assays. The oligomeric state of pure uncolored Mj0056 obtained by Prep. C was analyzed on a calibrated analytical gel-sizing 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 $MgCl_2$, 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°, 37°, 50°, 70°, 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 MS 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 electrospray 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-TOCSY and CCH-COSY experiments, while aromatic assignments were made by linking aromatic spin systems to the respective $C^{\beta}H_2$ protons in a 2D-NOESY spectrum. Stereospecific assignments and the resulting χ^1 rotamer assignments were determined for 54 of 93 prochiral $C^{\beta}H_2$ protons and for the $C^{\gamma}H_3$ groups of 7 of 8 valine residues. Assignments of χ^1 rotamers were also available all isoleucine residues and 3 of 4 threonine residues. Assignments of χ^2 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 $C^{\delta}H_3$ groups.

Distance data were derived from a set of five 3D-NOESY spectra, including the heteronuclear edited NNH-, CCH-, and CNH-NOESY spectra (Diercks et al., 1999) in addition to conventional ^{15}N- and ^{13}C-HSQC-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 H^N-H^N crosspeaks with a minimum distance of 3.2 Å and medium intensity or weaker sequential and intraresidue H^N-H^{α} crosspeaks with a minimum distance of 2.7 Å. Allowances for the use of pseudoatoms (by using r^{-6} averaging) were added for methyl groups and nonstereospecifically assigned methylene groups. Dihedral angle restraints were derived for backbone ϕ and ψ angles based on C^{α}, C^{β}, C′, and H^{α} 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 during stereospecific assignment with a tolerance of 30°, with the exception of proline residues where the χ^1 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 $^3J_{HNH\alpha}$ coupling constants measured from an HNHA experiment and for 90 backbone ψ angles based on $^3J_{NH\alpha}$ 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 ^{15}N-HSQC-NOESY spectrum and where donor-acceptor pairs were consistently identified in preliminary calculations. The restraints were applied via inclusion of pseudocovalent bonds as described by Truffault 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/rad^2. 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 ^{15}N-HSQC-NOESY, CNH-, and NNH-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 device (Formulatrix) and visually inspected. The crystals of the Mj0056-MgCDP-FMN complex were cryoprotected in a solution containing 10% w/v PEG400 in addition to the reservoir solution; all other crystals were loop mounted directly from the crystallization plates and shock frozen in liquid nitrogen. All datasets were collected at beamline PXII at the SLS (Swiss Light Source, Villigen, Switzerland) under cryo conditions at 100 K on a MAR225 detector (mar research).

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-PO_4 complex by molecular replacement with the program MOLREP (Vagin and Teplyakov, 2004). The structure Mj0056-NaCDP-PO_4 was solved on the basis of the refined coordinates of Mj0056-PO_4 and subsequently used as a search model for Mj0056-MgCDP and Mj0056-MgCDP-FMN. Structures were built and refined with the programs COOT (Emsley and Cowtan, 2004) and REFMAC5 (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, UTP, ATP, and GTP at 85°C and are available at http://www.structure.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.; 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

Bauer, S., Kemter, K., Bacher, A., Huber, R., Fischer, M., and Steinbacher, S. (2003). Crystal structure of *Schizosaccharomyces pombe* riboflavin kinase reveals a novel ATP and riboflavin-binding fold. J. Mol. Biol. *326*, 1463–1473.

Cheek, S., Ginalski, K., Zhang, H., and Grishin, N.V. (2005). A comprehensive update of the sequence and structure classification of kinases. BMC Struct. Biol. *5*, 6.

Coles, M., Diercks, T., Liermann, J., Groger, A., Rockel, B., Baumeister, W., Koretke, K.K., Lupas, A.N., Peters, J., and Kessler, H. (1999). The solution structure of VAT-N reveals a "missing link" in the evolution of complex enzymes from a simple βαββ-element. Curr. Biol. *9*, 1158–1168.

Coles, M., Djuranovic, S., Söding, J., Frickey, T., Koretke, K., Truffault, V., Martin, J., and Lupas, A.N. (2005). AbrB-like transcription factors assume a swapped hairpin fold that is evolutionarily related to double-ψ β barrels. Structure *13*, 919–928.

Coles, M., Hulko, M., Djuranovic, S., Truffault, V., Koretke, K., Martin, J., and Lupas, A.N. (2006). Common evolutionary origin of swapped-hairpin and double-ψ β barrels. Structure *14*, 1489–1498.

Cornilescu, G., Delaglio, F., and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J. Biomol. NMR *13*, 289–302.

Diercks, T., Coles, M., and Kessler, H. (1999). An efficient strategy for assignment of cross-peaks in 3D heteronuclear NOESY experiments. J. Biomol. NMR *15*, 177–180.

Diercks, T., Daniels, M., and Kaptein, R. (2005). Extended flip-back schemes for sensitivity enhancement in multidimensional HSQC-type out-and-back experiments. J. Biomol. NMR *33*, 243–259.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Frickey, T., and Lupas, A. (2004). CLANS: a Java application for visualizing protein families based on pairwise similarity. Bioinformatics *20*, 3702–3704.

Holm, L., and Sander, C. (1993). Protein structure comparison by alignment of distance matrices. J. Mol. Biol. *233*, 123–138.

Hooft, R.W., Vriend, G., Sander, C., and Abola, E.E. (1996). Errors in protein structures. Nature *381*, 272.

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Crystallogr. *26*, 795–800.

Karthikeyan, S., Zhou, Q., Mseeh, F., Grishin, N.V., Osterman, A.L., and Zhang, H. (2003a). Crystal structure of human riboflavin kinase reveals a β barrel fold and a novel active site arch. Structure *11*, 265–273.

Karthikeyan, S., Zhou, Q., Osterman, A.L., and Zhang, H. (2003b). Ligand binding-induced conformational changes in riboflavin kinase: structural basis for the ordered mechanism. Biochemistry *42*, 12532–12538.

Koretke, K.K., Russell, R.B., and Lupas, A.N. (2002). Fold recognition without folds. Protein Sci. *11*, 1575–1579.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *26*, 283–291.

Lovell, S.C., Davis, I.W., Arendall, W.B., de Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S., and Richardson, D.C. (2003). Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. Proteins *50*, 437–450.

Lupas, A.N., Ponting, C.P., and Russell, R.B. (2001). On the evolution of protein folds: are similar motifs in different protein folds the result of convergence, insertion, or relics of an ancient peptide world? J. Struct. Biol. *134*, 191–203.

Murshudov, G.N., Vagin, A.A., Lebedev, A., Wilson, K.S., and Dodson, E.J. (1999). Efficient anisotropic refinement of macromolecular structures using FFT. Acta Crystallogr. D Biol. Crystallogr. *55*, 247–255.

Rappsilber, J., Ishihama, Y., and Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. Anal. Chem. *75*, 663–670.

Schuler, G.D., Altschul, S.F., and Lipman, D.J. (1991). The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. Proteins *9*, 180–190.

Söding, J., and Lupas, A.N. (2003). More than the sum of their parts: on the evolution of proteins from peptides. Bioessays *25*, 837–846.

Söding, J., Remmert, M., Biegert, A., and Lupas, A.N. (2006). HHsenser: exhaustive transitive profile search using HMM-HMM comparison. Nucleic Acids Res. *34*, W374–W378.

Susin, S., Abian, J., Sanchez-Baeza, F., Peleato, M.L., Abadia, A., Gelpi, E., and Abadia, J. (1993). Riboflavin 3′- and 5′-sulfate, two novel flavins accumulating in the roots of iron-deficient sugar beet (*Beta vulgaris*). J. Biol. Chem. *268*, 20958–20965.

Truffault, V., Coles, M., Diercks, T., Abelmann, K., Eberhardt, S., Luttgen, H., Bacher, A., and Kessler, H. (2001). The solution structure of the N-terminal domain of riboflavin synthase. J. Mol. Biol. *309*, 949–960.

Vagin, A., and Teplyakov, A. (2004). An approach to multi-copy search in molecular replacement. Acta Crystallogr. D Biol. Crystallogr. *56*, 1622–1624.

Wang, A.C., and Bax, A. (1995). Reparametrization of the karplus relation for $^{3}J_{(H\alpha\text{-}N)}$ and $^{3}J_{(HN\text{-}C')}$ in peptides from uniformly $^{13}C/^{15}N$-enriched human ubiquitin. J. Am. Chem. Soc. *117*, 1810–1813.

Wang, W., Kim, R., Jancarik, J., Yokota, H., and Kim, S.H. (2003). Crystal structure of a flavin-binding protein from *Thermotoga maritima*. Proteins *52*, 633–635.

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-PO_4 (2VBS), Mj0056-MgCDP (2VBU), Mj0056-NaCDP-PO_4 (2VBT), and Mj0056-MgCDP-FMN (2VBV).

