Convergent Acquisition of a Novel Protein Domain

Membrane Anchoring of Aminoacyl-tRNA Synthetases by Convergent Acquisition of a Novel Protein Domain

Elvira Olmedo-Verd‡, Javier Santamaría-Gómez‡, Jesús A. G. Ochoa de Alda, Lluis Ribas de Pouplana, and Ignacio Luque

From the Instituto de Bioquímica Vegetal y Fotosíntesis, C.S.I.C. and Universidad de Sevilla, Avda América Vespuccio 49, E-41092 Seville, the School of Biology, IE University, Campus de Santa Cruz la Real, c/Cardenal Zúñiga 12, E-40003 Segovia, the Institute for Research in Biomedicine (IRB Barcelona), c/Baldri Reixac 15-21, 08028 Barcelona, Catalonia, and the ICREA, Passeig Lluís Companys 1, Barcelona 08010, Catalonia, Spain

Background: Aminoacyl-tRNA synthetases (aaRS) are enzymes that couple tRNAs to their cognate amino acids.

Results: aaRSs from cyanobacteria containing the novel CAAD protein domain are localized in the thylakoid membrane.

Conclusion: Confinement of aaRSs to the membrane is advantageous for cyanobacteria under specific conditions.

Significance: These findings provide the first description of membrane-bound aaRSs.

Four distinct aminoacyl-tRNA synthetases (aaRSs) found in somecyanobacterial species contain a novel protein domain that bears two putative transmembrane helices. This CAAD domain is present in glutamyl-, isoleucyl-, leucyl-, and valyl-tRNA synthetases, the latter of which has probably recruited the domain more than once during evolution. Deleting the CAAD domain from the valyl-tRNA synthetase of Anabaena sp. PCC 7120 did not significantly modify the catalytic properties of this enzyme, suggesting that it does not participate in its canonical tRNA-charging function. Multiple lines of evidence suggest that the function of the CAAD domain is structural, mediating the membrane anchorage of the enzyme, although membrane localization of aaRSs has not previously been described in any living organism. Synthetases containing the CAAD domain were localized in the intracytoplasmic thylakoid membranes of cyanobacteria and were largely absent from the plasma membrane. The CAAD domain was necessary and apparently sufficient for protein targeting to membranes. Moreover, localization of aaRSs in thylakoids was important under nitrogen limiting conditions. In Anabaena, a multicellular filamentous cyanobacterium often used as a model for prokaryotic cell differentiation, valyl-tRNA synthetase underwent subcellular relocation at the cell poles during heterocyst differentiation, a process also dependent on the CAAD domain.

Aminoacyl-tRNAs (aa-tRNAs) are substrates for protein synthesis at the ribosomes and they perform several other cellular functions (1, 2). These aa-tRNAs are synthesized by aminoacyl-tRNA synthetases (3), which couple each tRNA with its cognate L-amino acid. Thus, the rules of the genetic code are dictated by the double specificity of aaRSs for L-amino acids and tRNAs with particular anticodons. The reaction catalyzed by aaRSs occurs in two successive steps: activation of the amino acid with ATP, forming an aminoacyl adenylate intermediate; followed by the transfer of the amino acid moiety to the 3’ acceptor end of the tRNA (3). Despite the common catalytic mechanism, aaRSs have been categorized into two unrelated classes based on sequence and structure criteria (4–7). Enzymes of the same class are thought to be derived from a common ancestral protein that arose very early in evolution, and that diversified through divergence and multiple gene duplication events (8). Class I aaRSs are often monomeric and exhibit a characteristic nucleotide-binding Rossmann-fold in their catalytic domain. By contrast, class II enzymes are dimeric or tetrameric with a catalytic domain formed by anti-parallel β strands surrounded by α helices (5). As modular proteins, aaRSs have evolved by the acquisition (and/or loss) of protein domains, which have been appended to the catalytic domain, and are involved in a variety of functions (see below and Refs. 9 and 10). Some domains are present in one or multiple aaRSs and are universally distributed in nature. Such domains must have been recruited very early in evolution. Other domains are restricted to one or a few aaRSs in specific phyla, suggesting that they were acquired more recently. A domain appended to a particular aaRS in some organisms may exist in other organisms as a stand-alone polypeptide operating in trans to perform the same function (11, 12). The number of appended domains in a particular aaRS tends to be greater in more complex organisms, leading to the proposal that domain recruitment by aaRSs is an accretive and progressive phenomenon during evolution (13, 14).

The function of some appended domains may be related to the canonical aminoacylation activity of aaRSs. Thus, some
Membrane-anchored Aminoacyl-tRNA Synthetases

domains are involved in tRNA binding, augmenting their affinity (and in some cases specificity) for the tRNA (15, 16), whereas other domains may participate in editing functions, i.e. the hydrolysis of ester bonds mistakenly established by the synthetase between the tRNA and a noncognate amino acid (17). Some other domains participate in cellular functions unrelated to the aminoacylation reaction (13, 14, 18). For instance, the WHEP domain of eukaryotic GluProRS is involved in translational control of genes encoding proinflammatory proteins by directly interacting with the GAIT element in the 3’-UTR of target mRNAs (19, 20).

Convergent recruitment of a particular protein domain by distinct aaRSs has been described, for instance, the internal editing domain of AlaRS is homologous to the N-terminal editing domain of bacterial/eukaryotic ThrRS (21, 22). Furthermore, in eukaryotes GST, WHEP, or EMAP II domains are present in different aaRSs (13, 14).

We have recently described that several cyanobacterial genomes contain genes of anomalous length encoding some class I aaRS, including glutamyl-tRNA synthetase (GluRS), valyl-tRNA synthetase (ValRS), leucyl-tRNA synthetase (LeuRS), and isoleucyl-tRNA synthetase (IleRS). These aaRSs contained a foreign sequence of 100–200 amino acids with two putative transmembrane helices, which we termed the CAAD domain (for cyanobacterial aminocyl-tRNA synthetases appended domain) (23). The presence of CAAD-containing aaRSs is not universal in the phylum but rather, it is restricted to certain species, indicating that multiple acquisition events probably occurred during the diversification of the different lineages. In the corresponding genomes, genes encoding these aaRSs are found in a single copy, indicating that their products are functional. Here, we characterize the CAAD domain at the functional and present evidence demonstrating the structural role of CAAD in anchoring aaRSs to the membrane.

EXPERIMENTAL PROCEDURES

Organisms and Growth Conditions—Anabaena sp. PCC 7120 and derivative strains were cultured in BG11 medium (24) under continuous illumination (75 µE m⁻² s⁻¹, unless otherwise indicated), at 30 °C in shaken liquid cultures or bubbled with a mixture of CO₂ and air (1% v/v). Bubbled cultures were supplemented with 10 mM NaHCO₃. Solid medium was prepared by the addition of 1% Difco agar. Antibiotics for the selection of manipulated strains were used at the following concentrations: neomycin, 10–50 µg ml⁻¹; streptomycin 2–5 µg ml⁻¹; and spectinomycin 2–5 µg ml⁻¹. To induce heterocysts, bubbled cultures of Anabaena grown in BG11 medium were harvested, washed twice with BG11₀ medium (similar to BG11 but lacking NaN₃), inoculated in BG11₀ medium supplemented with 10 mM NaHCO₃, and cultured for 24 h at 30 °C under continuous illumination. For growth tests, cultures were supplemented with different inhibitors at the following concentrations: 1–methionine sulfoximine, 1–7.5 µM; sulfometuron methyl, 0.01–0.1 µM; chloramphenicol, 1–5 µg/ml; and hydrogen peroxide, 1–5 mM.

Escherichia coli was routinely grown in LB medium supplemented with antibiotics at standard concentrations when necessary (25). DH5α and XL1-blue strains were used for standard cloning and the C41(DE3) strain for the overexpression of Anabaena ValRS::His and ValRSΔCAAD::His proteins under control of the T7 promoter. Expression of the T7 RNA polymerase in C41(DE3) cells was induced by addition of IPTG (isopropyl β-d-thiogalactopyranoside) at a final concentration of 0.4 mM.

Cell Fractionation—Cyanobacterial cell fractionation was carried out and membrane preparations were obtained according to the protocol described by Sobotka et al. (26), with some modifications. Cells from 600–700-ml cultures of cyanobacteria were harvested by centrifugation, washed with 50 mM Tris-HCl buffer (pH 7.5), and resuspended in buffer T (20 mM HEPES-NaOH, pH 7.5, 10 mM MgCl₂, 5 mM CaCl₂, and 20% (v/v) glycerol) at a ratio of 5 ml of buffer T/g of cell pellet (wet weight). Cells were disrupted in the French press at 9000 p.s.i. in the presence of 1 mM PMSF (phenylmethylsulfonyl fluoride) as a protease inhibitor. The cell extracts were centrifuged at 32,000 × g for 10 min at 4 °C to eliminate unbroken cells and cell debris. The supernatant (referred to as the “cell extract”) was ultracentrifuged at 100,000 × g for 1 h at 4 °C. To avoid contamination with the membrane fraction only the top portion (about 75% of the total volume) of the supernatant containing the soluble fraction was saved, whereas the bottom portion in direct contact with the pellet containing the membranes was discarded. The pellet containing the membranes was washed with buffer T and resuspended in the same buffer supplemented with 1% (w/v) n-dodecyl-β-d-maltoside. In experiments where membrane purity was paramount, pellets were resuspended in buffer T and ultracentrifuged a second time at 100,000 × g for 1 h at 4 °C. The supernatant was discarded and the pellet was resuspended in the same buffer supplemented with 1% (w/v) n-dodecyl-β-d-maltoside. The chlorophyll concentration in cyanobacterial cultures or cell fractions were determined as previously described (27), whereas the protein content was determined by the modified Lowry procedure (28).

E. coli cell fractionation was achieved by resuspending cells from IPTG-induced cultures in buffer A (0.05 M phosphate buffer, pH 8, 0.15 M NaCl, and 10% (v/v) glycerol) and then disrupting them in the French press at 12,000 p.s.i. in the presence of 1 mM PMSF. Extracts were centrifuged at 21,000 × g and the supernatant was further centrifuged at 100,000 × g. The supernatant containing the soluble fraction was saved and the pellet was washed extensively with buffer A and resuspended in buffer A supplemented with 1% (w/v) n-dodecyl-β-d-maltoside.

Strain Construction—The plasmid and strain construction are described under supplemental data. The oligonucleotides used are listed under supplemental Table S1.

Protein Purification—Anabaena ValRS::His was purified from E. coli C41(DE3, pC50) membrane preparations. Membranes from 100-ml cultures induced with IPTG (see above) were resuspended in 1 ml of buffer containing 20 mM HEPES-NaOH (pH 7.5), 10 mM MgCl₂, 5 mM CaCl₂, 20% (v/v) glycerol, 0.04% (w/v) n-dodecyl-β-d-maltoside, and 2 mM imidazole, mixed with 1 ml of HIS-Select nickel affinity gel (Sigma), and incubated for 1 h at 4 °C with gentle agitation. The resin was decanted and the supernatant containing nonbound proteins was discarded. The resin was subsequently washed in successive steps with buffer containing 0, 5, and 10 mM imidazole.
Bound proteins were recovered by step elution with buffer containing 100, 150, and 200 mM imidazole. The fractions containing ValRS::His were pooled and subjected to gel filtration in PD-10 columns (GE Healthcare) for buffer exchange. Purified protein was stored in buffer containing 20 mM HEPES-NaOH (pH 7.5), 10 mM MgCl2, 5 mM CaCl2, 20% (v/v) glycerol, and 1% (w/v) n-dodecyl-β-d-maltoside. ValRS::His was purified from the soluble fraction of E. coli C41(DE3, pCE53) cells induced with IPTG (see above) by nickel affinity chromatography using His-Trap HP columns (GE Healthcare).

**Protein Electrophoresis, Western Blotting and In-gel Fluorescence**—Proteins were resolved by SDS-PAGE in the Miniprotein III apparatus (Bio-Rad) and transferred to Hybond-P PVDF membranes (GE Healthcare). Western blots were performed as described previously (29). Antibodies against an internal peptide (EYDRKKDESPSVEAC) of *Anabaena* ValRS were obtained from rabbits and affinity purified (Genscript, Piscataway, NJ). Penta-His antibodies (Qiagen) were used to detect His-tagged proteins. In-gel fluorescence in SDS-PAGE gels was registered in an Ettan DIGE Imager scanner (GE Healthcare).

**tRNA Overexpression and Purification**—E. coli XL1-Blue strains containing the pCE13, pCE14, or pCE41 plasmids for tRNA<sup>Val</sup> overexpression and the pCE47 plasmid for that of tRNA<sup>Leu</sup> were induced with 1 mM IPTG and the total tRNA was extracted as described previously (30). The content of *Anabaena* tRNA<sup>Val</sup> or *Lyngbya* tRNA<sup>Leu</sup> was determined in plateau aminoacylation assays.

**ATP-PP<sub>i</sub> Exchange**—The active concentration of functional aaRSs was determined in active site titration assays (31, 32). The *Kₘ* for ATP and L-Val were determined by the ATP-PP<sub>i</sub> exchange reaction as described previously (32). Standard reaction mixtures containing 50 mM HEPES (pH 7.6), 20 mM MgCl₂, 1 mM DTT, 2 mM ATP, 5 mM L-Val, 2 mM sodium orthovanadate (5 Ci mol⁻¹), and 5–20 nm enzyme were incubated at 30 °C. To determine the *Kₘ* for the concentration of ATP or L-Val was varied over a range of 0.2–20-fold the *Kₘ* value. Initial velocities of the reactions were plotted against substrate concentration and fitted to the Michaelis-Menten equation to deduce the *Kₘ*.

**Aminoacylation Reactions**—Aminoacylation assays were performed as described previously (32). Standard tRNA charging reactions were incubated at 30 °C and contained 100 mM HEPES (pH 7.2), 20 mM KCl, 30 mM MgCl₂, 5 mM ATP, 0.1 mg ml⁻¹ BSA, 0.5 mM DTT, 20 μM ¹⁴C-labeled L-Val (250 Ci mol⁻¹), 10 μM tRNA, and 5–15 nm enzyme. Reactions were stopped on filters soaked with 5% trichloroacetic acid, washed with 5% trichloroacetic acid, and the radioactivity retained was counted by scintillation.

**Bioinformatics**—Homologous CAAD domains were retrieved from a PSI-BLAST search until convergence (33) using the CAAD domain of ValRSs from *Nostoc punctiforme* as a seed (8 iterations, inclusion threshold <0.005). The resulting sequences were aligned using COBALT (34), exported to Jalview (35) to remove the nonhomologous regions, and realigned using MAFFT (36). This procedure resulted in a multiple alignment of 197 sequences that were used to build a Hidden Markov model (HMM) profile, which in turn was used to search (HMMsearch) against the nr-protein data base (November, 2010). HMMBuild and HMMsearch were implemented into the HMMPer3.0 package (37), which was run using Institut Pasteur Mobyle framework (38). Removal of 100% redundancy resulted in 205 unique sequences. Weblogo was used to construct sequence logos from the HMM profile of these sequences (39).

We used the JNet method (40, 41) implemented in Jalview (35) to infer the likely secondary structure of the aligned CAAD domains. The results were confirmed using other secondary structure predicting programs such as Quick2D (42), PSIPRED (43), and PRALINE (44).

**Microscopy**—Confocal microscopy of *Anabaena* filaments on solidified medium was performed using a Leica HCX PLAN-APO ×63 1.4 NA oil immersion objective attached to a Leica TCS SP2 confocal laser-scanning microscope. GFP was excited by 488 nm irradiation from an argon ion laser. Fluorescent emission of GFP and cyanobacterial autofluorescence was monitored by collection across windows of 500–540 and 630–700 nm, respectively.

**RESULTS**

**Sequence and Structural Features of the CAAD Domain**—CAAD was found in 4 distinct class I synthetases (GluRS, ValRS, LeuRS, and IleRS) from certain cyanobacterial species but not in the aaRSs from any other organisms (Fig. 1A). In GluRS, CAAD appears as a C-terminal extension connected to the C-terminal anticodon-binding domain by a repetitive sequence. By contrast the CAAD in ValRS, LeuRS, and IleRS are inserted in the C-terminal half of the sequence, in domains that contact the anticodon stem-loop of the tRNA, far from the catalytic domain (Fig. 1A and supplemental Fig. S2).

CAAD domains from the distinct aaRSs exhibit weak sequence identity (about 25% in pairwise sequence comparisons: supplemental Fig. S1) but they invariably share the following common secondary structural features: (i) two putative transmembrane helices at similar positions (although their boundaries may vary slightly depending on the prediction program used) separated by a short connecting peptide of 6–8 amino acids and (ii) an α-helix at the C terminus. Conserved residues concentrated in the second and third helices, and some sequence features are invariant or highly conserved indicating that they are functionally important (Fig. 1B and supplemental Fig. S1). These include the presence of acidic residues in the connecting peptide; a Pro residue at the end of the connecting peptide; Gly, Glu, and Arg residues, and several hydrophobic amino acids in the second helix; an absolutely conserved Arg and a Lys-rich sequence in the third helix; and a Gly residue at the C terminus of the domain (Fig. 1B and supplemental Fig. S1).

** Evolutionary Recruitment of the CAAD Domain by Class I aaRSs**—In GluRS, IleRS, and LeuRS, the CAAD was found at a single insertion point, whereas in ValRS it was found at two
Membrane-anchored Aminoacyl-tRNA Synthetases

TABLE 1
Kinetic parameters of Anabaena ValRS and ValRSΔCAAD

|                | tRNAVal | tRNAVal-2 | tRNAVal-3 | l-Val | ATP  | kcat |
|----------------|---------|-----------|-----------|-------|------|------|
| ValRS          | 2.08 ± 0.18 | 1.08 ± 0.32 | 0.61 ± 0.09 | 192.6 ± 16.8 | 200.45 ± 29.05 | 0.074 ± 0.002 |
| ValRSΔCAAD     | 1.01 ± 0.13  | 1.25 ± 0.09  | 0.47 ± 0.03  | 164.3 ± 3.4   | 123.35 ± 14.55  | 0.107 ± 0.013  |

FIGURE 1. Cyanobacterial species containing CAAD-bearing aaRSs and features of the CAAD domain. A, species containing aaRSs with the CAAD insertion, indicating the approximate position and length of the insertion. HFC, heterocyst-forming cyanobacteria. B, sequence logos from the HMM alignment of CAAD domains of cyanobacterial aaRSs (top) and the entire CAAD protein domain family (including CAAD domains of cyanobacterial aaRS and CAAD-like proteins from cyanobacteria and plants). The height of each letter (vertical axis) is proportional to the frequency (the most common is shown on top). The horizontal axis represents the position of residues used in the analysis. Secondary structural features are shown over each weblogo. The position of transmembrane helices (TM) is indicated by bars with discontinuous ends, indicating that the limits of the helices may vary depending on the prediction program used. C, phylogenetic guide tree of cyanobacterial species based on 16S ribosomal sequences (CAT + GTR + I + 6TI evolutionary model). Highlighted species contain the CAAD domain in GluRS (green), ValRS (pink), LeuRS (yellow), or IleRS (orange).
type strain in which ValRS activity was observed predominantly in the membrane fraction, in the SR10 mutant virtually all of this activity was detected in the soluble fraction (Fig. 3B).

Membrane-anchored Aminoacyl-tRNA Synthetases

FIGURE 2. ValRS and LeuRS activity in fractionated extracts from cyanobacteria. A, incorporation of $^{14}$C-labeled L-Val was measured in aminoacylation assays of cell extracts (open circles) or membrane preparations (squares) containing 0.5–1.5 $\mu$g of chlorophyll, or an equivalent volume of the soluble fraction (open triangles) from the indicated cyanobacterial strains. Anabaena tRNA$^{Val}_{UAC}$Val was used as the substrate for Anabaena and Synechococcus, and Anabaena tRNA$^{Val}_{GAC}$Val for Cyanothece. B, LeuRS activity was measured in cell extracts (open circles) or membrane preparations (squares) containing 0.06 $\mu$g of chlorophyll, or an equivalent volume of the soluble fraction (open triangles) from the indicated cyanobacterial strains. Lyngbya tRNA$^{Leu}_{CAG}$Leu was used as a substrate.

FIGURE 3. Localization of ValRS in fractionated extracts from Anabaena strains. A, aliquots of Anabaena cell extracts (CE) or membrane fractions (M) containing 1.1 $\mu$g of chlorophyll (75 $\mu$g of protein of the cell extract), or a similar volume of the soluble fraction (SF), were analyzed in Western blots probed with antibodies against Anabaena ValRS. Aliquots of Anabaena cell extracts or membrane fractions containing 0.15 $\mu$g of chlorophyll (10 $\mu$g of protein for the cell extract), or a similar volume of soluble fraction, were analyzed in Western blots probed with antibodies against GluRS, glutamine synthetase (GS), or the D1 protein of photosystem II. B, plots represent ValRS activity in fractionated extracts of wild-type Anabaena and the SR10 mutant: cell extracts (open circles), membrane fraction (squares), and the soluble fraction (open triangles). Other details are as described in the legend to Fig. 2. C, aliquots of cell extracts or the membrane fraction from 7120E5 and 7120E6 strains containing 0.57 or 0.023 $\mu$g of chlorophyll, or a similar volume of the soluble fractions, were analyzed in Western blots probed with commercial antibodies against polyhistidine. Aliquots of cell extracts or the membrane fraction from 7120E5 and 7120E6 strains containing 0.4 $\mu$g of chlorophyll, or a similar volume of the soluble fraction, were analyzed in Western blots probed with antibodies against GS or the D1 protein.
To further confirm these findings, two strains derived from *Anabaena*, 7120E5 and 7120E6, were constructed by inserting an extra copy of a recombinant *valS* gene into a replicative plasmid, encoding ValRS or ValRSΔCAAD fused to C-terminal histidine tags, respectively. The presence of ValRS was determined in each cell fraction using commercial antibodies against the histidine tag. Although ValRS::His was localized in the membrane fraction, ValRSΔCAAD::His was detected in the soluble fraction (Fig. 3C), which further indicated that the CAAD domain is required for the localization of ValRS to the membrane.

Cyanobacteria are photosynthetic bacteria that contain two independent membrane systems, the plasma membrane and the thylakoid membrane (50–54), the latter of which contains the photosynthetic apparatus. To determine whether CAAD-containing aaRSs are found in one or in both membrane systems, we constructed strains derived from *Anabaena* expressing ValRS or ValRSΔCAAD fused to the GFP marker protein (strains 7120E7 and 7120E8, respectively) and they were analyzed by confocal microscopy. The fluorescence of the ValRS::GFP fusion protein (Fig. 4A, top center panel) co-localized perfectly with the red autofluorescence of the photosynthetic pigments of the thylakoid membranes (Fig. 4A, top left and right panels). By contrast, ValRSΔCAAD::GFP fluorescence was confined to the central part of the cytoplasm, and it was largely excluded from regions where thylakoids were detected, exhibiting a similar distribution to the soluble GFP protein in strain 7120E2 (Fig. 4, A, middle panels, B, bottom panels). Strain CSVT15 expressing a fusion protein of the ammonium transporter Amt1 with GFP (55) was used as a control for plasma membrane localization. In this strain, the green fluorescence of GFP was observed surrounding the thylakoids (Fig. 4A, bottom panels). Thus, our results indicate that *Anabaena* ValRS is localized in the thylakoid membrane but not to the plasma membrane, and further demonstrates that CAAD is required for membrane anchoring of these enzymes.

**FIGURE 4.** Subcellular localization of GFP-fused proteins in *Anabaena* filaments. A and B, filaments of the *Anabaena* strains (left) expressing the proteins indicated in parentheses were analyzed by confocal microscopy. Left panels show the red autofluorescence of photosynthetic pigments and the central panels show the green fluorescence of GFP. The panels on the right show the merged images of the left and central panels.
A GFP protein fused to the CAAD domain was also found to colocalize with the red fluorescence of the thylakoids, although in this case the GFP::CAAD pattern exhibited a patchy distribution (Fig. 4B, top panels). To further corroborate the localization of this protein in the thylakoid membrane, extracts of the 7120E1 strain expressing GFP::CAAD were fractionated by ultracentrifugation, resolved by SDS-PAGE, and the fluorescent proteins were detected by in-gel fluorescence in an Ettan DIGE scanner. The GFP::CAAD protein co-localized with chlorophyll in the membrane fraction (not shown), confirming its presence in the thylakoids. These results indicate that the CAAD domain is sufficient to direct an unrelated protein (GFP) to the thylakoid membrane. However, as the subcellular distribution of the GFP::CAAD protein is not identical to that of ValRS::GFP, we cannot rule out a potential contribution of sequences outside CAAD in conferring the localization.

Under certain nutritional conditions, Anabaena filaments are composed of two metabolically and morphologically distinct cell types (56–58). Cell differentiation is aimed to separate two incompatible metabolic functions, namely oxygenic photosynthesis and nitrogen fixation, by the confinement of each process in different cell types (57, 59). The differentiation of the cells called heterocysts is triggered by deprivation of nutritional sources of chemically combined nitrogen. These cells appear semi-regularly interspersed in the filament and, unlike the intervening vegetative cells that perform oxygenic photosynthesis, they are mostly devoted to nitrogen fixation (59). To analyze the localization of the ValRS::GFP in heterocysts, filaments from the 7120E7, 7120E8, 7120E1, and 7120E2 strains were deprived of combined nitrogen for 24 h, which was sufficient to induce heterocyst differentiation in our laboratory conditions. Heterocysts were identified by their thicker morphology and weaker autofluorescence as compared with vegetative cells (Fig. 5). Surprisingly, ValRS::GFP and CAAD::GFP were confined to the poles of the cell in 7120E7 and 7120E1 heterocysts, respectively, whereas some areas of the thylakoid membranes occupying nonpolar positions were devoid of these proteins. The localization of these proteins in distinct cell fractions of the 7120E7 and 7120E1 filaments, containing an average of 1 heterocyst per 10–15 vegetative cells, was analyzed by cell fractionation and in-gel fluorescence following SDS-PAGE. Both proteins were detected exclusively in the membrane fraction, indicating that they are also anchored to membranes in heterocysts (not shown). By contrast, ValRS::GFP and CAAD::GFP were detected in the central portion of the cytoplasm of heterocysts of the 7120E8 and 7120E2 strains, respectively (Fig. 5).

**Impact of Membrane Localization of ValRS on Cell Fitness**—To assess the physiological impact of the presence of aaRSs in the thylakoid membrane, we analyzed the wild-type strain and
Membrane-anchored Aminoacyl-tRNA Synthetases

Evolution of the CAAD Domain—Our results indicate that the CAAD module has been recruited on multiple occasions by distinct aaRSs at different time points in evolution. CAAD-containing LeuRS or GluRS are each found in only one species, suggesting a recent recruitment event. By contrast, CAAD-containing ValRS is present in all species of the heterocyst-forming clade, suggesting a single event early in the divergence of this group, which is estimated to have occurred over 2 billion years ago (64). The mechanism by which CAAD domains were acquired is unknown. The existence of genes encoding single-domain CAAD-like proteins in cyanobacterial genomes suggests that the aaRSs hereby described originated from the insertion of one of these genes into the ORF of a class I aaRS. If this event were random, we would expect to find this domain inserted in other regions of the genome. However, we observed the recurrent acquisition of the domain by proteins with a related structure-function. This can be envisaged as a case of evolutionary convergence and suggests that adaptive pressure contributed to the formation of this combination of functional domains.

Sequence and Structural Features of the CAAD Domain—A major feature of the CAAD domain and CAAD-like proteins is the presence of three α-helices, including two transmembrane helices. Conserved features of CAAD also include an invariant Arg residue and a Lys-rich sequence immediately C-terminal to the second helix, which are probably topological determinants for correct insertion into the membrane. Indeed, positively charged residues are often found in the cytosolic side of membrane proteins flanking transmembrane helices. Conserved features of CAAD-bearing aaRSs and the position of this domain in the C-terminal portion of these proteins indicate that insertion in the membrane is most likely post-translational, as the CAAD domain has to exit the ribosome before insertion.

ValRS, LeuRS, and IleRS are closely related class I aaRSs, exhibiting a relatively high degree of conservation in both sequence and structure. In addition to the Rossmann-fold catalytic domain, these proteins have several common accessory domains, including the CP-1 domain for editing that is inserted in the catalytic domain and the helix bundle anticondon-binding domain (supplemental Fig. S2) (69–72). C-terminal to the helix bundle domain, ValRS and IleRS both contain an anticondor stem-binding domain (supplemental Figs. S2 and S3), which is absent in LeuRS (70–72). The C-terminal domain is idiosyncratic for each aaRS. Accordingly, ValRS contains a coiled-coil domain essential for aminoacylation, which contacts the external corner of the L-shaped tRNA, whereas IleRS contains a zinc-binding domain that contacts the anticondor wobble base, an identity determinant for tRNA\textit{\r}\textsuperscript{16}, and LeuRS has an α,β-domain linked by a flexible tether that is required for its activity in bacteria (69, 75, 76). Insertion of CAAD occurs at domains C-terminal to the Rossmann-fold, far from the catalytic site in the primary and tertiary structure.

DISCUSSION

Domain shuffling, involving the gain or loss of individual domains, is one of the major driving forces behind the molecular innovation of modular proteins, including aaRSs (62, 63). Novel appended domains may fuse or insert into pre-existing domains, provided that they do not disrupt the canonical function of the aaRS. The present findings indicate that the novel CAAD protein domain recruited by 4 distinct class I aaRSs is essential for membrane anchoring. To the best of our knowledge, these are the only known cases of membrane-bound aaRSs.

FIGURE 6. Growth of wild-type \textit{Anabaena} and the SR10 mutant in the presence of MSX. A, curves represent growth as the chlorophyll content of cultures of wild-type \textit{Anabaena} (closed symbols) and the SR10 mutant (open symbols), in the absence (circles) or presence of 1 (squares), 2.5 (triangles), or 5 μM (diamonds) of MSX. A representative plot of three independent experiments is shown. B, serial dilutions of cultures of wild-type \textit{Anabaena} and the SR10 mutant were spotted onto BG11-agar plates containing the concentrations of MSX indicated on the left.
Membrane-anchored Aminoacyl-tRNA Synthetases

(supplemental Figs. S2 and S3). In LeuRS, the CAAD is inserted into the helix bundle domain, whereas in IleRS the insertion splits the anticodon stem-binding domain. In ValRS, one of the insertions occurs in the loop connecting the helix bundle to the anticodon stem-binding domain, whereas the second insertion splits the latter domain in a manner similar to that observed in IleRS (supplemental Fig. S2). The site of CAAD insertion into the reported crystal structures of bacterial ValRS, IleRS, and LeuRS enzymes can be ascertained (69–72), although with some degree of ambiguity, due to sequence divergence (supplemental Figs. S2 and S3). In all cases the insertion site corresponds to superficial residues of the protein, on surfaces opposing the docking face for tRNA (supplemental Fig. S3). Thus, it is conceivable that in these three enzymes CAAD protrudes out of the main body of the protein, enabling the insertion of its two helices into the membrane, causing minimal disturbance to the overall structure of the synthetase, the folding of adjacent domains, and the docking of tRNA. The cytoplasmic residues of the CAAD probably act as flexible linkers connecting the transmembrane helices with the rest of the protein. In the case of GluRS, folding of the synthetase portion is not likely to be affected by CAAD due to the existence of a long (270 amino acids) intervening linker sequence. This is consistent with our observation that deletion of CAAD of *Anabaena* ValRS does not alter the kinetic parameters of the enzyme, including the $k_{cat}$ and $K_m$ values for tRNA$^\text{Val}$, l-Val, and ATP. Indeed, this suggests that CAAD does not participate in the aminoaacylation reaction and that its insertion does not significantly change the overall conformation of the rest of the protein.

Function of the CAAD Domain—Our deletion experiments strongly suggest that CAAD is required for membrane anchoring of aaRSs. Although residues of ValRS outside CAAD may also be involved, at least part of the information required for membrane insertion must be contained in the CAAD sequence, because its fusion to an unrelated protein, such as GFP, is sufficient to target it to the membrane. Moreover, proteins containing CAAD, either aaRSs or the CAAD::GFP fusion protein, are specifically localized to the thylakoid membrane, and are absent or largely excluded from the plasma membrane. Hence, CAAD sequences appear to contain determinants for protein sorting to this structure. The localization of plant and cyanobacterial CAAD-like proteins in the thylakoid membrane further supports the proposed role of CAAD as a protein-sorting domain. The *Arabidopsis thaliana* protein, named TMP14 or PSI-P, is found at the stromal side of the thylakoid membrane (70, 77). A cyanobacterial homologue of TMP14 has been shown to interact with the Psal subunit of photosystem I and with HIC, an integral thylakoid membrane protein induced under light stress (78, 79). Although the subcellular localization of 2 CAAD-like proteins has been established, other functional information pertaining to these proteins could be useful to infer the putative metabolic or signaling function of the CAAD domain.

Cell Differentiation-induced Relocalization—In differentiated heterocysts of *Anabaena* sp. PCC 7120, ValRS concentrates in polar thylakoids (i.e. those next to the junction with a neighboring cell), whereas thylakoids in other regions of the cytoplasm are largely devoid of the enzyme. In mature heterocysts of the *Nostoc* and *Anabaena* genera, polar accumulation of highly curved membranes, named honeycomb thylakoids, has been observed, whereas thylakoids in other parts of the cytoplasm exhibit a normal, less curved appearance (80). Our findings demonstrate the heterogeneous protein composition of thylakoid membranes in heterocysts, in turn suggesting the existence of sorting signals governing the localization of newly synthesized ValRS. Accumulation at the poles depends on the presence of CAAD, which is sufficient to direct an unrelated protein such GFP to this region. The poles of heterocysts are probably regions of high metabolic activity, as a significant portion of the exchange of metabolites with neighboring vegetative cells occurs through the septum (81), which may account for the observed localization of ValRS at heterocyst poles.

Impact of aaRS Membrane Anchoring on Cell Fitness—Domains appended to the catalytic domain of aaRSs mediate diverse activities, including tRNA binding, editing, or interaction with other proteins. The results of the present study ascribe a function to CAAD distinct from those previously described and suggest a dual role: a structural role as a membrane anchor and a sorting role, whereby proteins may be directed to specific subcellular localizations in different cell types. Membrane confinement of a single aaRS (or two in *T. erythraeum* cells) is intriguing and difficult to interpret intuitively. Cells of all three kingdoms of life are compartmentalized, with the most complex degree of compartmentalization observed in eukaryotic cells. In general, cell compartmentalization serves to localize cell components in environments appropriate for their function, to bring together functionally related molecules and thereby favor their interaction, or to separate and prevent the interaction of other molecules. The cyanobacterial aaRSs described here are compartmentalized in the thylakoid membrane, and undoubtedly prevented from diffusing in the cytoplasm. Nevertheless, membrane-anchored aaRSs appear to efficiently mediate translation given the success of many cyanobacteria that have carried these enzymes for billions of years. Our results indicate that the fitness of the SR10 *Anabaena* strain (containing a ValRS from which the CAAD domain is deleted) is similar to that of the wild-type in most conditions, except upon treatment with MSX. MSX impairs glutamine synthetase activity, provoking a shortage of nitrogenous compounds in the cell, including amino acids. Therefore, under conditions of nitrogen limitation, the presence of ValRS in the thylakoid membrane appears to be advantageous for the cell, permitting faster growth. It is tempting to speculate that the advantage of a defined localization in the membrane might be due to the proximity to other molecules required for aminoacylation or alternative functions. By contrast, a soluble enzyme (e.g. ValRS::CAAD) with a more dispersed localization would not have such ready access to functional partners. Although this difference may not be critical in most situations, it may be of great importance under particular conditions of stress associated with shortages of specific molecules or nutrients. In nature, nitrogen-fixing organisms such as *Anabaena* may be subjected to transient periods of nitrogen limitation similar to those induced by MSX treatment. One such example is when nitrogen limitation occurs concomitant with conditions that
compromise the synthesis of active nitrogenase enzyme (e.g. during shortages of iron or molybdenum).

The findings presented here do not rule out the possibility that membrane-bound aaRSs play additional metabolic or signaling roles in addition to their canonical aminoaacylation function. Further studies are currently underway to address this issue.

Acknowledgments—We are indebted to A. M. Muro-Pastor for providing the pCSAM200 plasmid and E. Flores, A. Herrero, and V. Merino-Puerto for the CSV1T5 strain. We thank A. Lindahl and M. I. Muro-Pastor for antibodies against D1 and GS proteins, providing the pCSAM200 plasmid and E. Flores, A. Herrero, and

REFERENCES

1. Schimmel, P., and Ribas De Pouplana, L. (2000) Trends Biochem. Sci. 25, 207–209
2. Banerjee, R., Chen, S., Dare, K., Gilreath, M., Praetorius-Ibba, M., Raina, M., Ibba, M., Reynolds, N. M., Rogers, T., Roy, H., Yadavalli, S. S., and Ibba, M. (2010) FEBS Lett. 584, 387–395
3. Iba, M., and Soll, D. (2000) Annu. Rev. Biochem. 69, 617–650
4. Eddy, S. R. (2008) PLoS Comput. Biol. 4, e1000069
5. Néron, B., Ménager, H., Mauvais, C., Joly, N., Maupetit, J., Letort, S., Carrere, S., Tuffery, P., and Letondal, C. (2009) Bioinformatics 25, 3005–3011
6. Schneider, T. D., and Stephens, R. M. (1990) Nucleic Acids Res. 18, 6907–6100
7. Cole, C., Barber, J. D., and Barton, G. J. (2008) Nucleic Acids Res. 36, W197–201
8. Cuff, J. A., and Barton, G. J. (2000) Proteins 40, 502–511
9. Biegert, A., Mayer, C., Remmert, M., Söding, J., and Lupas, A. N. (2006) Nucleic Acids Res. 34, W335–339
10. Broyson, R., McLean, S., Bhat, G. J., and Marko, M. (2007) Bioinformatics 23, 1073–1079
11. Cusack, S., Berthet-Colominas, C., Härtlein, M., Nassar, N., and Leberman, R. (1990) Nature 347, 249–255
12. Burbaum, J. J., and Schimmel, P. (1991) J. Biol. Chem. 266, 16965–16968
13. Ribas de Pouplana, L., and Schimmel, P. (2000) Cell Mol. Life Sci. 57, 865–870
14. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Nucleic Acids Res. 25, 2369–2372
15. Schimmel, P., and Ribas De Pouplana, L. (2000) Trends Biochem. Sci. 25, 207–209
16. Schimmel, P., and Ribas De Pouplana, L. (2000) J. Biol. Chem. 278, 52857–52864
17. Schimmel, P., and Ribas De Pouplana, L. (2000) Cell 102, 449–454
Membrane-anchored Aminoacyl-tRNA Synthetases

Res. 9, 689–710
64. Tomitani, A., Knoll, A. H., Cavanaugh, C. M., and Ohno, T. (2006) Proc. Natl. Acad. Sci. U.S.A. 103, 5442–5447
65. von Heijne, G. (1986) EMBO J. 5, 3021–3027
66. von Heijne, G., and Gavel, Y. (1988) Eur. J. Biochem. 174, 671–678
67. Yuan, J., Zweers, J. C., van Dijl, J. M., and Dalbey, R. E. (2010) Cell Mol. Life Sci. 67, 179–199
68. Facey, S. J., and Kuhn, A. (2004) Biochim. Biophys. Acta 1694, 55–66
69. Tukalo, M., Yaremchuk, A., Fukunaga, R., Yokoyama, S., and Cusack, S. (2005) Nat. Struct. Mol. Biol. 12, 923–930
70. Fukai, S., Nureki, O., Sekine, S., Shimada, A., Tao, J., Vassylyev, D. G., and Yokoyama, S. (2000) Cell 103, 793–803
71. Fukai, S., Nureki, O., Sekine, S., Shimada, A., Vassylyev, D. G., and Yokoyama, S. (2003) RNA 9, 100–111
72. Silvian, L. F., Wang, J., and Steitz, T. A. (1999) Science 285, 1074–1077
73. Senger, B., Auxilien, S., Englisch, U., Cramer, F., and Fasiolo, F. (1997) Biochemistry 36, 8269–8275
74. Nureki, O., Vassylyev, D. G., Tateno, M., Shimada, A., Nakama, T., Fukai, S., Konno, M., Hendrickson, T. L., Schimmel, P., and Yokoyama, S. (1998) Science 280, 578–582
75. Hsu, J. L., and Martinis, S. A. (2008) J. Mol. Biol. 376, 482–491
76. Hsu, J. L., Rho, S. B., Vannella, K. M., and Martinis, S. A. (2006) J. Biol. Chem. 281, 23075–23082
77. Hansson, M., and Vener, A. V. (2003) Mol. Cell. Proteomics 2, 550–559
78. Khrouchtchova, A., Hansson, M., Paakkarinen, V., Vainonen, J. P., Zhang, S., Jensen, P. E., Scheller, H. V., Vener, A. V., Aro, E. M., and Haldrup, A. (2005) FEBS Lett. 579, 4808–4812
79. Wang, Q., Jantaro, S., Lu, B., Majeed, W., Bailey, M., and He, Q. (2008) Plant Physiol. 147, 1239–1250
80. Lang, N. J., and Fay, P. (1971) Proc. R. Soc. Lond. B Biol. 178, 193–203
81. Mullineaux, C. W., Mariscal, V., Nenninger, A., Khanum, H., Herrero, A., Flores, E., and Adams, D. G. (2008) EMBO J. 27, 1299–1308