Chapter 2

A Novel Class of Modular Transporters for Vitamins in Prokaryotes


2.1 Abstract

This chapter describes the discovery of the ECF (Energy Coupling Factor)-type transporters, which were introduced in the previous chapter.

The specific and tightly controlled transport of numerous nutrients and metabolites across cellular membranes is crucial to all forms of life. However, many of the transporter proteins involved have yet to be identified. Comparative analysis of the ever-growing collection of microbial genomes coupled with experimental validation enables the discovery of such transporters. Here, this approach was used to discover an abundant class of vitamin transporters in prokaryotes with an unprecedented architecture. The transporters have energy-coupling modules (also named ECF modules, for Energy Coupling Factor), comprised of a conserved
transmembrane protein and two nucleotide-binding proteins typical for ATP binding cassette (ABC) transporters. These ECF-type ABC transporters do not make use of extracellular substrate-binding proteins (SBPs) to capture substrates. Instead, they use small integral membrane proteins (S-components) to capture specific substrates. 21 families of these S-components were identified, which are not related in amino acid sequence and with different specificities predicted by genome context analyses. Roughly half of the S-components (335 cases) have a dedicated energizing module (deduced from the organization of the genes in operons), but in 459 cases distributed among almost 100 gram-positive bacteria, including numerous human pathogens, different and unrelated S-components share the same energy-coupling module (data from the 365 prokaryotic genomes that were analyzed in 2008). The shared use of energy-coupling modules was experimentally confirmed for folate, thiamine and riboflavin transporters (this chapter) and later also for others (chapter 3). This class of membrane transporters is named energy-coupling factor (ECF) transporters and is a new class of ABC importers.

2.2 Introduction

Transport proteins residing in the cytoplasmic membrane allow the selective uptake and efflux of solutes and are essential for cellular growth and metabolism (Konings, 2006). Reflecting the importance of transporters, between 3% and 16% of the genes in prokaryote genomes are predicted to encode transporter proteins (Ren & Paulsen, 2007). These transporters form numerous families that are diverse in structure, energy-coupling mechanisms, and substrate specificities (see chapter 1). As only a small fraction of predicted transporter proteins have known substrates, the functional prediction and annotation of the specificities of transporter proteins in the rapidly growing number of sequenced genomes represent a substantial challenge (Ren & Paulsen, 2005; Saier, 1999). The use of computational comparative genomic techniques including gene co-localization, co-occurrence, and co-regulation analyses combined with experimental assays is a powerful approach to identify novel transporters and to uncover their cellular role (M. Gelfand & Rodionov, 2008). The starting point for the analysis presented here was the discovery of multicomponent transport systems for the vitamin biotin (BioYNM) and the transition metals nickel (NikMNQO) and cobalt (CbiMNQO) (Hebbeln et al., 2007; D. A. Rodionov et al., 2008). These transporters all have substrate-specific components (S-components), which are integral membrane proteins, and energy-
coupling factor (ECF-) modules. The S-components of the biotin transporter (BioY) and the metal transporters (NikMN and CbiMN) are dissimilar in sequence, but the ECF modules contain similar proteins. These modules consist of an ATPase typical of the ATP binding cassette (ABC) superfamily (generally called Ecfa, but specific names have been used for different systems: BioM/NikO/CbiO) and a characteristic transmembrane protein (generally named Ecft, specific names such as BioN/NikQ/CbiN). In many prokaryotes, genes encoding ECF modules are not linked to either nikMN, cbiMN, or bioY. This observation prompted us to hypothesize the existence of ECF module-dependent transporters for additional substrates, unrelated to transition metal ions and biotin. In this study, comparative genomics and experimental techniques were combined to identify new transporters with an ECF module plus S-component architecture that are specific for various vitamins and related substrates. The majority of these systems are predicted to share a single ECF module, a unique design among membrane transporters. The predicted modular design, substrate specificities of representative members, and shared use of the ECF module by various substrate capture components were confirmed by biochemical analyses of vitamin transport systems of the Firmicutes.

2.3 Materials and methods

2.3.1 Bioinformatics analysis and data sources.

Prokaryotic genome sequences used for the comparative analysis were obtained from GenBank (Benson et al., 2007). Metabolic reconstruction, genome context analysis, and functional gene annotation were performed using the SEED comparative genomics resource as described previously (M. Gelfand & Rodionov, 2008; Overbeek et al., 2005; D. A. Rodionov, 2007). The results were captured in the ECF class transporters subsystem (http://seed-viewer.theseed.org/seedviewer.cgi). Candidate DNA regulatory motifs were identified using Genome Explorer (M. S. Gelfand et al., 2000; Mironov et al., 2000; D. A. Rodionov, 2007). Candidate RNA regulatory elements such as riboswitches were identified with the RNA-Pattern program (Vitreschak et al., 2008) using input profiles from the RNA Families (Rfam) database (Griffiths-Jones et al., 2005). The Protein Families (Pfam) database was used to identify conserved functional domains (Finn et al., 2008). Transmembrane domains were predicted using the TMPred server (www.ch.embnet.org) (Hofmann & Stoffel, 1993).
2.3.2 Cloning and expression of *L. casei* transporters.

*Lactobacillus casei* folT and thiT were amplified using Taq DNA polymerase (Invitrogen, Carlsbad, CA) with *L. casei* ATCC 334 genomic DNA as the template and ligated into vector pNZ8048. The resulting constructs were used as templates for the amplification of fragments containing the *nisinA* promoter, folT or thiT, and a terminator using Pfu Ultra DNA polymerase (Invitrogen). Amplicons were ligated into vector pIL252 between the BamHI and XhoI sites. The ecfAA'T operon was amplified using Pfu Ultra DNA polymerase and inserted between the NcoI and SstI sites of vector pNZ8048. The recombinant plasmids were cloned into *Lactococcus lactis* strain NZ9000. Cells were grown at 30 °C in supplemented M17 medium (Difco), and for expression, nisin (0.1% (v/v) of a culture supernatant of the nisin A-producing strain NZ9700) was added.

2.3.3 *B. subtilis* disruption strains.

*Bacillus subtilis* ypaA (ribU), ybaF (ecfT), yuaJ (thiT), and yceI (niaP) disruption strains were obtained from the joint Japanese and European *Bacillus subtilis* consortium (Vagner et al., 1998).

2.3.4 Vitamin uptake assays.

For [3H]5-formyltetrahydrofolate and [3H]thiamine uptake assays, *L. lactis* cells were washed once with cold phosphate-buffered saline (PBS) and resuspended in PBS at an OD(600nm) of 20. Assays were performed at 30 °C with stirring. Cells (500 µL) were pre-incubated for 5 min with glucose or 2-deoxyglucose (25 mM final concentration). Assays were started by adding 500 µL of PBS containing 2.2 µM [3H]5-formyltetrahydrofolate (specific activity of 0.4 µCi/nmol) or 2.3 µM [3H]thiamine (0.33 µCi/nmol) to the mixture. At intervals, 150 µL aliquots were passed through a Whatman cellulose nitrate membrane filter (0.45 µm). Filters were washed twice with 2 mL of ice-cold PBS, and cell-bound radioactivity was quantitated by liquid scintillation counting.

For [3H]riboflavin uptake assays, *Bacillus subtilis* cells were cultivated in mineral salts medium at 37 °C with vigorous shaking. At an OD(600nm) of 0.5, [3H]riboflavin was added (16.6 nM final concentration; 588,000 dpm). Timed aliquots were mixed with 2 mL of ice-cold PBS, and cell-bound radioactivity was quantitated by liquid scintillation counting.
filters were washed twice with 2 ml of ice-cold buffer and dried. Radioactivity was determined by liquid scintillation counting.

2.3.5 Cloning and expression of L. mesenteroides transporters.

*Leuconostoc mesenteroides* strain ATCC 8293 *folT, panT*, and *ribU* were amplified and inserted between the *NeoI* and *BglII* sites of pARCV (an expression vector containing an ampicillin resistance marker) (D. A. Rodionov et al., 2006). The resulting plasmids, pLmFolT, pLmPanT, and pLmRibU, encode the respective membrane proteins with a C-terminal FLAG tag. The *ecfAA’T* operon was likewise amplified and inserted into a variant of pARCV that harbored a streptomycin resistance gene and 10 histidine codons upstream of the insertion site. Plasmid pLmEcf encodes His10-EcfA (with a deca-His tag at the N-terminus), EcfA, and EcfT-FLAG (with a C-terminal FLAG tag). An ampicillin resistance-conferring variant of pLmEcf (pLmEcf-Amp) was constructed by the insertion of the *ecfAA’T* fragment from pLmEcf into pARCV.

2.3.6 Purification of transport protein complexes.

*E. coli* BL21 cells containing pLacI-RARE2 (encoding a Lac repressor and the tRNAs for rare codons) were used as the host for the heterologous production of *L. mesenteroides* proteins. Cells harboring pLmFolT, pLmPanT, pLmRibU, or pLmEcf-Amp individually or pLmEcf in combination with pLmFolT, pLmPanT, or pLmRibU were grown in 2 liters of Luria-Bertani broth supplemented with the appropriate antibiotics and 1 mM isopropyl-β-D-thiogalactopyranoside at 37 °C with shaking to an OD(578nm) of ~2, harvested by centrifugation, washed in 35 mM sodium-potassium phosphate buffer (pH 7.0), resuspended in the same buffer containing a mixture of protease inhibitors (Roche), and disrupted by three passages through a French pressure cell. Membranes were pelleted by ultracentrifugation, resuspended and homogenized in 50 mM Tris-HCl (pH 8.0), and solubilized by agitation for 1 h in the presence of a solution containing 2 % (w/v) dodecyl-β-D-maltoside, 5 % (v/v) glycerol, 300 mM NaCl, and protease inhibitors at 4 °C. Non-solubilized material was pelleted by ultracentrifugation. Imidazole was added to the supernatant to a final concentration of 20 mM, and the solution (10 mL) was mixed with 0.5 mL Ni-nitrilotriacetic acid Superflow resin (Qiagen) and incubated for 30 min at 4 °C with rotation. After transfer to an empty column, the resin was washed with
50 mM Tris-HCl (pH 7.5) containing 0.05% dodecyl-β-D-maltoside, 5 % glycerol, 300 mM NaCl, and 100 mM imidazole. Bound protein was eluted with 4 mL of this buffer containing 500 mM imidazole. The protein solution was concentrated eightfold by Amicon concentrators (30 kDa cutoff), and samples (10 µg protein) were subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Proteins were blotted onto nitrocellulose membranes and probed with anti-penta-His (Qiagen) or anti-FLAG (Sigma) antibodies by standard Western blotting protocols.

2.3.7 Chemicals.

\[3',5',7,9^3\text{H}(\text{N})\]-\(6\text{S}\)-5-Formyltetrahydrofolic acid diammonium salt (10 Ci/mmol), \[^3\text{H}(\text{G})\]thiamine hydrochloride (10 Ci/mmol), and \[^3\text{H}(\text{G})\]riboflavin (24 Ci/mmol) were obtained from Moravek Biochemicals (Brea, CA). (6R,6S)-5-Formyltetrahydrofolic acid calcium salt was obtained from Schircks Laboratories (Jona, Switzerland).

2.4 Results

2.4.1 Bioinformatics analysis

A bioinformatics analysis of 365 prokaryotic genomes using the SEED comparative genomics platform (Overbeek et al., 2005) revealed that the abundance and functional diversity of transporters with the novel ECF architecture extends far beyond the few cases of metal and biotin transporters noted above. 432 gene cassettes encoding EcfA and EcfT components were found in 238 genomes (numbers from the available genomes in 2008), with the _ecfA_ genes being very often duplicated (_ecfA_ and _ecfA’_, which are similar but not identical in sequence) (see http://jb.asm.org/content/191/1/42/suppl/DC1). These _ecfAA’T_ or _ecfAT_ gene cassettes fall into two groups based on their genomic organizations (figure 2.1).

There are 335 operons (data from 2008) that resemble the previously described nickel, cobalt, and biotin transporters (Hebbeln et al., 2007; D. A. Rodionov et al., 2006) in that they occur next to genes encoding small integral membrane proteins. These membrane proteins are candidate S-components, but as fewer than half are related to the S-components of the nickel, cobalt, or biotin transporters, it is likely that the majority of them have novel substrate specificities (see table 2.1).
Figure 2.1. Distribution and comparative genomic analysis of the ECF transporters, a new class of prokaryotic transporters. A Classification and abundance of dedicated and shared ECF transporters. Dedicated transporters have a substrate-specific S-component and a dedicated ECF module whose genes co-localize on the genome in an operon. Shared transporters have individual S-components and shared ECF modules that are unlinked to S-components. Composite bar colors indicate the contributions of transporters found in different taxa to the total transporter number. Note that the S-components BioY, CbrT, HtsT, and QrtT (and, to a lesser extent, RibU, PanT, HmpT, ThwT, QueT, and CblT) occur in both groups and can thus function with either a dedicated or a shared ECF module. B Comparative genomic analysis of the identified transporter families including their domain compositions, names, predicted substrate specificities, and example gene identifications. Substrate-specific integral membrane components (S-components: S) are shown by black rectangles, conserved transmembrane components (EcfT: T) are shown by blue rectangles, and ATPase domains (EcfA: A / EcfA': A') are shown by red circles. Examples of genome context evidence (e.g., gene co-regulation or co-localization) supporting the predicted transporter function are shown on the right. Note that HmpT is currently annotated as PdxU2 and predicted to be bind a pyridoxine related compound. PdxT is currently named PdxU, but still predicted to bind pyridoxine.
The operons containing both a gene coding for a putative S-component and the ecfAA'T or ecfAT genes were named dedicated, and are found in diverse microbes. The remaining 97 (data from 2008) ecfAT and ecfAA'T gene cassettes (found mainly in the Firmicutes, the Thermotogales, and some members of the Archaea) do not have adjacent candidate S-component genes. Almost all these cassettes are accompanied by various candidate S-component genes (459 in total) scattered elsewhere in the genome, with some genomes having as many as 12 distinct S-component genes but only one or two copies of the ECF module. We predict that in such cases, multiple S-components can use the same ECF module to form an active transporter complex. The vast majority of these candidate S-components are again unrelated to S-components of the nickel, cobalt, or biotin transporters, and hence, most of them probably act on new substrates.

The S-components were classified into 21 protein families (table 2.1). Experimental data have shown that several of these are indeed involved in the transport of various substrates: the BioY (Hebbeln et al., 2007) (see also chapter 4), CbiMN (Siche et al., 2010), and NikMN (D. A. Rodionov et al., 2006) families mentioned above, one family involved in riboflavin uptake (RibU [YpaA] in Bacillus subtilis and Lactococcus lactis) (Burgess et al., 2006; Duurkens et al., 2007; Vogl et al., 2007), another involved in tryptophan uptake (TrpP in B. subtilis) (Yakhnin et al., 2004). More recently, specificity for other substrates has been experimentally verified: NiaX has been shown to be involved in niacin transport (chapter 3), ThiT in thiamin transport (this chapter and (Eudes et al., 2008; Erkens & Slotboom, 2010; Erkens et al., 2011)), FolT in folate transport (this chapter and (Eudes et al., 2008)) and PanT in transport of pantothenate (Neubauer et al., 2009) (see table 2.1 for an overview). For all other families except one, substrate specificities were predicted by combining genome context analysis and metabolic reconstruction (M. Gelfand & Rodionov, 2008; D. A. Rodionov, 2007; D. Rodionov et al., 2009). The results show a notable preference for substrates that are vitamins or their precursors, namely, niacin, folate, pantothenate, pyridoxine, lipoate, thiamine, and its precursors, and the precursors of cobalamin, methionine, and queuosine (figure 2.1 and table 2.1).
Table 2.1. Functional roles of the different groups of S-components

<table>
<thead>
<tr>
<th>S-component</th>
<th>Pfam accession no.</th>
<th>Substrate</th>
<th>Evidence for specificity</th>
<th>Evidence for EcfAA'T dependence</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>BioY</td>
<td>PF02632</td>
<td>Biotin</td>
<td>E,R,C,O</td>
<td>E,C,O</td>
<td>(Hebbeln et al., 2007; D. A. Rodionov, 2007; D. A. Rodionov &amp; Gelfand, 2006; D. A. Rodionov, Mironov &amp; Gelfand, 2002) and see chapter 4</td>
</tr>
<tr>
<td>NikMN</td>
<td>PF01891</td>
<td>Nickel</td>
<td>R</td>
<td>C</td>
<td>(D. A. Rodionov et al., 2006)</td>
</tr>
<tr>
<td>CbiMN</td>
<td>PF01891</td>
<td>Cobalt</td>
<td>E</td>
<td>C</td>
<td>(D. A. Rodionov et al., 2006, 2003)</td>
</tr>
<tr>
<td>YkoE</td>
<td></td>
<td>Thiamin precursor</td>
<td>R,C</td>
<td>C</td>
<td>(D. A. Rodionov, Vitreschak et al., 2002)</td>
</tr>
<tr>
<td>MtsT</td>
<td>PF07155</td>
<td>Methionine precursor</td>
<td>R,C</td>
<td>C</td>
<td>(D. A. Rodionov et al., 2004)</td>
</tr>
<tr>
<td>HtsT</td>
<td>PF09605</td>
<td>Queuosine precursor</td>
<td>R,C,O</td>
<td>C,O</td>
<td>(Burgess et al., 2006; Duburkens et al., 2007; M. S. Gel-fand et al., 1999; Kreneva et al., 2000; Vitreschak et al., 2002; Vogl et al., 2007; D. Rodionov et al., 2009; Zhang et al., 2010)</td>
</tr>
<tr>
<td>QtT</td>
<td></td>
<td>Riboflavin</td>
<td>E,R,O,S</td>
<td>E,C,O</td>
<td>(D. Rodionov et al., 2009; Eudes et al., 2008)</td>
</tr>
<tr>
<td>FolT</td>
<td>PF06177</td>
<td>Folate</td>
<td>E,C,O</td>
<td>E,O</td>
<td>(D. Rodionov et al., 2009; Neubauer et al., 2009)</td>
</tr>
<tr>
<td>QueT</td>
<td>PF09515</td>
<td>Thiamin</td>
<td>E,R,O,S</td>
<td>E,O</td>
<td>(D. Rodionov, Vitreschak et al., 2002)</td>
</tr>
<tr>
<td>PanT</td>
<td></td>
<td>Pantothenate</td>
<td>C,O</td>
<td>E,C,O</td>
<td>(D. A. Rodionov et al., 2009), see chapter 3</td>
</tr>
<tr>
<td>ThiT</td>
<td></td>
<td>Thiazole</td>
<td>R,C,O</td>
<td>C,O</td>
<td>(D. A. Rodionov, Vitreschak et al., 2002)</td>
</tr>
<tr>
<td>NiaX</td>
<td></td>
<td>Niacin</td>
<td>R,O</td>
<td>E,O</td>
<td>(D. A. Rodionov et al., 2008), see chapter 3</td>
</tr>
<tr>
<td>TrpP</td>
<td></td>
<td>Tryptophan</td>
<td>E,R</td>
<td>E,C,O</td>
<td>(Sarsaero et al., 2000; Vitreschak et al., 2008)</td>
</tr>
<tr>
<td>PdxU</td>
<td></td>
<td>Pyridoxine</td>
<td>C</td>
<td>E,O</td>
<td>This chapter, unpublished</td>
</tr>
<tr>
<td>PdxU2</td>
<td></td>
<td></td>
<td>C</td>
<td>E,O</td>
<td></td>
</tr>
<tr>
<td>CblT</td>
<td></td>
<td>Cobalamin precursor (DMB?!)</td>
<td>R,C,O</td>
<td>C,O</td>
<td>(D. A. Rodionov et al., 2003)</td>
</tr>
<tr>
<td>LipT</td>
<td></td>
<td>Lipoate</td>
<td>C</td>
<td>O</td>
<td></td>
</tr>
</tbody>
</table>

\(a\): Identifiers of protein families are from the Pfam database (http://pfam.sanger.ac.uk/). \(b\): Evidences for the function of S-components include experimental analysis (E), co-regulation of their genes in a regulon (R), co-localization on the chromosome (C), co-occurrence (O) in the genome with metabolic genes from the same pathway, or structural data from X-ray crystallography (S). \(c\): Evidences for the ECF module dependence of S-components. E, experimental; C, co-localization on the chromosome (C), or co-occurrence (O) in the genome with ECF module gene cassettes. \(d\): Initially named PdxT. \(e\): Initially annotated as HmpT and being involved in the transport of a thiamin precursor (HMP, hydroxymethylpyrimidine). \(f\): DMB, dimethylbenzimidazole.
2.4.2 Experimental evidence

The transport systems for folate, thiamine, and biotin in Lactobacillus casei were postulated by Henderson et al. in the 1970s to comprise individual substrate-binding membrane proteins plus a common energy-coupling component (named energy-coupling factor [ECF]) (Henderson et al., 1979). Below, we provide experimental evidence that these Lactobacillus casei transporters belong to the class of transporters reported here. In recognition of the pioneering work the name ECF transporters was given to this new class of transporters consisting of energy-coupling modules and S-components.

To validate the bioinformatics analysis, we tested three of its predictions, namely, that (i) the previously characterized S-component RibU (specific for vitamin B$_2$, riboflavin) depends on the ECF module; (ii) the S-components FolT and ThiT are specific for folate and thiamine, respectively, and are ECF module dependent; and (iii) the multiple S-components of shared transporters (figure 2.1) share a common ECF module.

(i) RibU interacts with the ECF module in B. subtilis. The dependence of riboflavin transport on the S-component (RibU) and the ECF module was analyzed in B. subtilis disruption strains. $[^3]$HRiboflavin transport required functional ribU ($ypaA$) and ecfT ($ybaF$) genes; the disruption of either gene abolished the uptake of the vitamin (figure 2.2).

(ii) Substrate specificity and ECF module dependency of FolT and ThiT. The predicted substrate and the ECF module dependence of FolT were tested first using proteins from Leuconostoc mesenteroides. The $folT$ and $ecfAA'T$ genes were expressed in a folate-auxotrophic Escherichia coli strain ($pabA$ $abgT$), which cannot produce the folate precursor 4-aminobenzoate or take up folates. When analyzed for growth on minimal agar medium (Klaus et al., 2005), only transformants containing both $folT$ and the $ecfAA'T$ cassette were able to utilize 5-formyltetrahydrofolate (figure 2.3a). This establishes that folate is transported and that transport requires the energy-coupling module. Then, FolT, ThiT, and the EcfAA'T proteins from L. casei were produced in Lactococcus lactis in various combinations and $[^3]$Hthiamin (in case of ThiT) or $[^3]$H5-formyltetrahydrofolate uptake (for cells expressing FolT) was assayed in energized or de-energized cells (figure 2.3b and c). De-energized cells acquired neither thiamin nor 5-formyltetrahydrofolate regardless of which
proteins they expressed. Among energized cells, only those co-expressing ThiT and EcfAA'T acquired significant amounts of thiamin. The situation was similar for 5-formyltetrahydrofolate, except that cells expressing FolT alone acquired some label; this slow uptake may be due to the functional interaction of FolT with the endogenous *L. lactis* EcfAA'T module or possibly to the basal activity of the solitary FolT protein (the question whether S-components can transport without the ECF module is still under debate, see chapter 7). These data confirm activity with the predicted FolT and ThiT substrates and demonstrate the dependence of both thiamin and folate uptake on the same ECF module, as inferred in the classical work of Henderson et al. (Henderson et al., 1979).

(iii) Physical interaction of an ECF module with various S-components. Interaction of multiple, unrelated S-components with the shared ECF module was also tested by a series of pullout experiments (see also chapter 3). His-tagged EcfA, untagged EcfA', and FLAG-tagged EcfT from *L. mesenteroides* were co-produced in *E. coli* with or without FLAG-tagged FolT, PanT, or RibU. Membranes of the recombinant strains were solubilized and subjected to nickel-chelate affinity chromatography, followed by SDS-PAGE and Western blotting. The three EcfAA'T proteins co-

Figure 2.2. Riboflavin uptake in *Bacillus subtilis* disruption strains. Shown are data for the effect of disrupting *ribU* (triangles), *efcT* (circles), or *yecI* (squares) on [*H*]riboflavin uptake by *B. subtilis*. (The *yecI* gene served as a control; it encodes a protein unrelated to ECF transporters.) Cells were grown without riboflavin to an OD$_{600nm}$ value of 0.5, and [*H*]riboflavin was added (17 nM final concentration). At the times indicated, cells were harvested by filtration and washed, and their radioactive contents were determined. Values are means of duplicates; error bars indicate ranges.
Folate and thiamin uptake is mediated via FolT and ThiT when they are overexpressed in E. coli or L. lactis cells and is dependent on the co-expression of the ECF module. A Folate uptake by E. coli cells co-expressing L. mesenteroides folT and ecfAA'T. Recombinants containing empty vector or expression plasmids for the production of FolT, EcfAA'T, or FolT plus EcfAA'T were spotted (10 µL) after serial 10-fold dilutions onto non-supplemented minimal medium and onto minimal medium containing 4-aminobenzoate (3.6 µM) or 5-formyltetrahydrofolate (11 µM). Plates were incubated for 48 h at 37°C. B and C Uptake of [³H]thiamin (B) and [³H]5-formyltetrahydrofolate (C) by L. lactis containing empty vectors (triangles) or carrying L. casei thiT or folT (circles), ecfAA'T (diamonds), or thiT or folT co-expressed with ecfAA'T (squares). Cells were energized with glucose (black symbols) or de-energized with 2-deoxyglucose (open symbols).

purified during chromatography (figure 2.4), as indicated by immunodetection (for His-tagged EcfA and FLAG-tagged EcfT) and peptide mass fingerprint analysis (for EcfA') (not shown). As expected, FLAG-tagged FolT, PanT, and RibU did not bind to the affinity resin in the absence of the EcfAA'T module (figure 2.4, six right-hand lanes). However, each of these S-components co-purified with the EcfAA'T complex. The EcfA, EcfA', and EcfT components thus form a stable tripartite complex and can form quadripartite complexes with each of the three different S-components. The binding of eight different S-components to the EcfAA'T complex from L. lactis will be shown in chapter 3.

### 2.4.3 Predictions

The domain fusions in various ECF transport systems give some clues about the stoichiometry of the complex (see figure 2.5). The nik, cbi, and bio gene cassettes encode a single EcfA (ATPase), but as noted above, dual ATPase (EcfAA') are more common. In some cases ecfAA' genes are fused (figure 2.1b and 2.5). In a few cases the gene encoding an S-component has been fused with the ecfAA' genes, fusions between the ecfT and ecfAA' genes have taken place, or the gene encoding
Figure 2.4. Physical interaction between S and EcfAA'T components from *L. mesenteroides*. Membranes of recombinant *E. coli* cells producing the proteins indicated in the upper six lines were solubilized with n-dodecyl-β-d-maltoside and subjected to Nickel-chelate affinity chromatography prior to SDS-PAGE and Western blotting or were separated by SDS-PAGE without chromatography. The bottom shows strips of the blots probed with anti-penta-His or anti-FLAG antibodies.

an S-component has fused with *ecfT* (figure 2.5). On this basis, and because shared EcfAA'T components and specific S-components formed quadripartite complexes (figure 2.4), we propose a quadripartite model in which the S-component binds the substrate and the EcfAA'T-S-component complex transports it across the membrane (see figure 2.6). The translocation process is expected to be coupled to ATP hydrolysis in the two ATPase domains: EcfA₂ or EcfAA'. The subunit stoichiometry was confirmed experimentally for the shared EcfAA'T-S-component complexes of *Lactococcus lactis* (see chapter 3), but different stoichiometries have been reported for other (dedicated) ECF-complexes (Finkenwirth et al., 2010).

### 2.5 Discussion

The broad distribution, functional versatility, and modular assembly of the ECF module dependent transport systems are summarized in figure 2.1 and 2.6. The ECF transporters form a novel class of membrane transporters that can be classified into dedicated ECF transporters, this group includes transporters from diverse microbial lineages (170 species out of the 365 genomes that were analyzed in 2008) that have a dedicated EcfAT or EcfAA'T module encoded in the same gene cluster as an S-component, and shared ECF transporters (a total of 459 transporters in 91 species from the 365 analyzed genomes) that employ a universal EcfAT or EcfAA'T-module that is encoded by a separate gene cassette and shared by many
Figure 2.5. Genomic organization of the dedicated ECF transporters containing S-components from the TrpP, RibU, PanT, PdxU2, ThiW, QueT, and CblT families. Genes encoding substrate capture S-components and EcfA, (EcfA') and EcfT components of the dedicated energy-coupling modules are shown by black, red, and blue arrows, respectively.

different unlinked S-components. Shared ECF transporters are ubiquitous in the phyla Firmicutes and Thermotogales and also occur in some members of the Archaea. It must be emphasized that the distinction between dedicated and shared ECF transporters has been based solely on genomic organization. It cannot be excluded that S-components from dedicated systems interact with the ECF module from a shared system (if present in the same organism), and vice versa.

The S-components identified could be classified into at least 20 protein families that correspond to different substrate specificities (table 2.1). Most of them are integral membrane proteins of comparable sizes (155 to 230 residues) that have six predicted transmembrane domains. The NikM and CbiM proteins, which form a single family in the Pfam database, are larger (210 to 250 residues) and are predicted to have seven transmembrane domains. Sequence comparisons of representative S-components from 18 families revealed very little overall pair-wise identity between
the proteins from different families. However, recent X-ray crystallography studies showed that S-components share the same fold (see chapter 1 and 4) (Zhang et al., 2010; Erkens et al., 2011).

The ECF transporters identified in this study are mechanistically unique. Their substrate specificity is mediated by integral membrane proteins (S-components), which form active transporters in the presence of the ECF module. How do ECF transporters relate to other transporters of the ABC transporter superfamily? All ABC transporters couple ATP hydrolysis to substrate uptake or efflux (Davidson & Chen, 2004; Davidson et al., 2008). All ABC importers and exporters share a four-component architecture comprised of two transmembrane and two ATP-hydrolyzing domains. In addition, extra domains/subunits are often found. Prokaryotic importers, other than the ECF transporters, have additional extracytoplasmic soluble proteins (SBPs, for substrate-binding proteins) that mediate substrate binding and delivery to the respective transmembrane components. ECF transporters are characterized by (i) the absence of extracytoplasmic substrate binding proteins and their replacement by integral membrane proteins and (ii) the shared use of energy-coupling EcfaAT modules by many highly diverse S-components. Such sharing is occasionally seen in SBP-dependent ABC transporters, but it always involves very similar substrates and substrate binding proteins (see chapter 1 and (Higgins & Ames, 1981)). A less fundamental but nonetheless marked characteristic of ECF transporters is a predilection for vitamins.

Finally, as noted at the outset, numerous human pathogens such as Mycoplasma, Ureaplasma, and Streptococcus strains rely totally upon transporters to obtain vitamins and other essential metabolites due to the absence of the corresponding de novo biosynthetic pathways. Many of these microorganisms use ECF transporters, and indeed, certain ecf genes have been found to be essential for the growth and survival of Streptococcus pneumoniae and Mycoplasma genitalium (Glass et al., 2006; Thanassi et al., 2002). All components of ECF transporters, especially the unique S-components and EcfaT proteins, are thus potential targets for antibiotic development. In fact, the centrality of the EcfaT to the uptake of multiple compounds makes it a classic Achilles’ heel.
Figure 2.6. **Modular architecture of energy-coupling factor (ECF) transporters.** ECF transporters consist of a substrate-capture protein (S-component: ‘S’) and a dedicated (left-hand part) or a shared (right-hand part) energy-coupling module composed of two ABC ATPases (EcfA: ‘A’) and a conserved transmembrane protein (EcfT: ‘T’). Prokaryotes contain up to 22 different ECF systems. The shared use of a transporter module among many different systems was previously unprecedented and is predominantly found in Gram-positive bacteria including many human pathogens.