General Introduction

1. Membrane transport proteins

Cells of all organisms are surrounded by a lipid membrane that divides space in two compartments: the interior and exterior. The lipids constituting this membrane form a hydrophobic barrier that prevents the free exchange of hydrophilic solutes between both compartments and allows control of the internal environment. The lipid matrix is filled with proteins, that can account for more than half of the dry weight of the membrane as shown for the cytoplasmic membrane of *Escherichia coli* (20). The relevance of these proteins is reflected by the high percentage of all open reading frames (ORFs) (20-30%) predicted to code for membrane-embedded proteins (77,127). A large fraction of these proteins, predicted to be coded by approximately 5-10% of all ORFs, are transport proteins involved in the translocation of solutes from one compartment to another (78).

Transport proteins can be subdivided in four classes in terms of force driving the translocation. The primary transporters (**i**) utilize the energy stored in light (e.g., bacteriorhodopsin, the light-driven proton pump from *Halobacterium salinarium*), redox energy (e.g., respiratory chain components like NADH dehydrogenase I from *E. coli*), or energy-rich molecules such as ATP (e.g., the vitamin B12 transporter BtuCD from *E. coli*). The ATP binding cassette (ABC) transporters, to which BtuCD belongs, form the largest superfamily within the class of primary transporters.

The secondary transporters (**ii**) use the electrochemical energy stored in ion- or solute-gradients to drive transport. Secondary transport can proceed via uniport (one solute is transported down its gradient, so-called facilitated diffusion, e.g the glucose transporter GLUT1 from erythrocytes), symport (multiple different solutes are transported in the same direction, e.g., the lactose:H\(^+\) co-transporter LacY from *E. coli*), or antiport (multiple different solutes are transported in opposite directions, e.g., the Cl⁻:HCO\(_3\)⁻ exchanger AE1 from erythrocytes). The major facilitator superfamily (MFS), of which GLUT1 and LacY are a member, constitutes the largest superfamily of secondary transporters. The ABC and MF superfamilies form the two largest transporter families found in nature, and nearly half of all transporters belong to either one of these families (95).

The group-translocators (**iii**) couple the transport of a solute to its covalent modification. This type of transport is represented by the phosphoenolpyruvate phosphotransferase systems (PEP-PTS) which transfer the high energy phosphate group of PEP via intermediate phosphoproteins to the substrate while it is translocated across the membrane (e.g., the enzyme II-glucose from *E. coli*). Group translocators are found in most prokaryotes, but not in eukaryotes or archaea (76).

In contrast to the above-mentioned transport systems, the channels or pores (**iv**) do not undergo large conformational changes that are stoichiometrically coupled to the translocation of solutes. Instead, the channels open and close in response to an external signal (e.g., membrane tension, ligand or voltage) and allow the flow of small solutes down their concentration gradient (e.g., the mechanosensitive channel of large conductance MscL from *E. coli*).

In Table 1 [adapted from TransportDB (94,95)], the contribution of each class to the total number of transport systems on a genome is shown for representatives of all three kingdoms.
of life. In prokaryotes and archaea, the most abundant classes of transporters are the primary and secondary transport systems. In higher eukaryotes, due to the large number of the channel proteins (76), the secondary transporters and channels constitute the largest classes. Although transport proteins of all classes have received a lot of attention, today, our knowledge on membrane transport is still limited due to the difficulty to obtain high resolution structural information of the proteins and allows at best a posteriori explanations rather than a priori predictions of mechanistic aspects. This thesis focusses on mechanistic details of a subset of membrane transporters from the secondary transporter class.

2. The Major Facilitator Superfamily of secondary transporters

The largest family of transporters within the class of secondary transporters is the Major Facilitator Superfamily. Within the MFS, nearly 30 families can be discriminated. MFS transporters can be found in archaea, bacteria and eukarya (plasma- and organellar membranes), and they transport a variety of small substrates among which are sugars, peptides, charged solutes and hydrophobic compounds. All three mechanisms of secondary transport, that are, uniport, symport and antiport, are represented within the MFS (101).

The majority of MFS transporters are composed of a single polypeptide of 400-600 amino acids with 12 or 14 \( \alpha \)-helical transmembrane segments (TMs). Based on a significant degree of sequence similarity between the N- and C-terminal halves, MFS proteins are thought to originate from an internal gene duplication of a 6 TMs transporter (75). Additional evidence for this came from the high-resolution structures of LacY (3), the glycerol-3-phosphate:phosphate antiporter GlpT (54), from \( E. \) coli (Fig. 1), and the oxalate:formate antiporter OxlT from \( Oxalobacter formigenes \) (51). For these MFS proteins, the N- and C-terminal 6 TMs are related by a pseudo two-fold symmetry as the two halves could be superimposed on each other with only small deviations (2).
In addition, the structures of each of these carrier proteins comply with the “alternating access” model for translocation by non-channel transporters (classes i-iii). This model dictates that the substrate binding site is exposed to either the internal or the external side of the membrane, but never to both. A conformational change is involved in obscuring one site, while exposing the other (57). On the cytoplasmic side, the structures of both GlpT and LacY show a central hydrophilic cavity, which is sealed off on the periplasmic side (Fig. 1). This cavity reaches up to the middle of the membrane, where a substrate molecule is bound in the LacY structure. Although GlpT and LacY belong to different MFS families, the similarities in their structures suggest that secondary and tertiary structural elements are conserved within the MFS (1).

**Figure 1. Helix packing of GlpT and LacY from *E. coli*.** Transmembrane regions of GlpT (A) and LacY (B) are represented as ribbon diagrams, viewed along the membrane normal from the cytoplasmic side. The loop regions have been left out for clarity. The same color scheme is used for A and B. C, schematic depiction of the average helix packing of GlpT and LacY, viewed from the same side as A and B. TMs are labeled with roman numbers. The structures of GlpT and LacY have the PDB accession numbers 1PW4 and 1PV7, respectively.

### 3. The Galactoside-Pentoside-Hexuronide family of the MFS

Within the MFS, the galactoside-pentoside-hexuronide (GPH) family of transporters can be discerned (101). Members of this family catalyze transport of sugars and sugar derivatives (galactosides, glucosides, pentosides and hexuronides) in symport with a cation (mostly Na+ or H+). Members of the GPH-family can be found in both pro- and eukaryotes (73,84,93). Although the similarity of the complete sequences is sufficient to establish homology within the pro- or eukaryotic groups (Fig. 2 [adapted from (93)]), the sequence identity between members from both kingdoms of life is rather low. However, a comparison of the N-terminal domains, a region which has been reported to show maximal sequence similarity for distantly related transport proteins (68,102), did show a sufficient degree of similarity to establish homology between both groups (73). Nevertheless, some residues in the N-terminal domain, that are highly conserved within the prokaryotic group and are suggested to perform catalytically important roles in the translocation process (e.g., charged residues in TM II) (84), are not conserved in the eukaryotic group. Most data on structural and functional aspects of GPH-proteins is derived from studies on two bacterial representatives of the GPH-family, the melibiose permease, MelB from *E.coli*, and the lactose permease, LacS, from *S. thermophilus*. LacS and MelB are part of different subfamilies within the bacterial branch of the GPH-family (Fig. 2).
3.1. Secondary and tertiary structure

GPH-proteins comprise approximately 450 amino acids, and hydropathy profiles predict 12 $\alpha$-helical TMs traversing the membrane in a zigzag fashion. The LacS subfamily deviates from the standard topology by the presence of a C-terminal cytoplasmic domain of approximately 160 amino acids. This protrusion is designated IIA-domain based on its homology with IIAGlc of the *Escherichia coli* glucose PTS, (34% identical residues, depicted in Fig. 6) (86). The topology model, shown for LacS (Fig. 3), has been verified for MelB, using alkaline phosphatase fusions (90). The low resolution projection map of 2D-crystals of MelB (45) shows several discrete high-density regions around a central region of lower density. While the latter could indicate a cavity, the four density peaks most likely represent TMs perpendicular to the membrane normal. The absence of more than four of these regions suggests that the majority of the TMs are tilted with respect to the membrane, which is in agreement with the high-resolution structures of other MFS proteins (3,49,54).

In the absence of high-resolution structures of GPH-proteins, a helix packing model has been constructed (122), based on low resolution structural information derived from genetic screens (second-site suppressors), bioinformatical, biochemical and biophysical studies (summarized in (122)). The model (Fig. 4, adapted from (122)) represents the cytoplasmic face of the...
protein, since most distance constraints are available for this side of the protein. This model consists of an inner ring of six TMs, ordered around a central cavity, and an outer ring of six TMs surrounding the first ring. The inner ring contains the three amphipathic TMs II, IV and XI, that might line a hydrophilic cavity, plus TMs I, VII and X. Most of these TMs harbour residues involved in the binding of the sugar and/or the cation, or the coupling of the substrate and cation flux; other amino acids in these TMs have been proposed to interact with catalytically important residues. Additionally, loop 10-11, located on the cytoplasmic side of the protein, has been positioned within the central cavity as it contains catalytically important regions (82,85) and has been shown to be less flexible than expected for a loop region (110). The outer ring is composed of TMs III, V, VI, VIII, IX and XII. Data on catalytic regions or interactions between amino acids are scarce for the outer ring TMs, making it less likely that these TMs constitute part of the translocation path.

3.2. Substrate binding and transport

GPH-proteins exclusively catalyze transport of substrates (sugars and sugar derivatives) in symport with cations, with a substrate:cation stoichiometry of approximately 1:1 (6,30). The coupling ions reported are Na⁺, Li⁺ and H⁺ for MelB (132), and H⁺ for LacS (30). Acidic residues in TMs I, II and IV have been proposed to contribute to the cation binding site (84). These residues are highly conserved within the prokaryotic GPH-members, however, conservation is low or absent in the eukaryotic GPH-proteins. For these proteins, the significance of other regions implied in cation binding, e.g., TM VII (62), might be larger.

Cation and substrate transport are not independent processes and tight coupling of the translocation of both ligands prevents carrier-mediated leakage of substrate or cation. The use of certain cations seems to be restricted to certain substrates, e.g., MelB can use Na⁺ and H⁺ to transport melibiose, whereas lactose could only be symported with protons (132). Additionally, mutations affecting substrate specificity were found to alter cation recognition as well (10). The isolation of substrate specificity mutants has indicated that multiple regions of the protein contribute to shaping the sugar binding site. Residues in TMs I, IV, VII, X and XI, that are located in the inner ring of our helix packing model (Fig. 4), were found to contain

Figure 3. Topology model of the carrier domain of LacS. This model is based on the topology model of MelB (90). For simplicity the remaining 160 C-terminal amino acids, constituting the IIA-domain are not shown; the 3D structure of the *E. coli* IIA^{Glc}, which is homologous to the LacS-IIA domain, is presented in Fig. 6. The grey horizontal lines indicate the membrane interfaces.
residues implicated in sugar selection. However, the majority of the sugar specificity mutants reside in loop 10-11 (10,21).

Substrate transport catalyzed by GPH-proteins can proceed in electrogenic and electroneutral modes (Fig. 5), that is, when a complete translocation cycle is considered. Electroneutral substrate transport (Fig. 5A), designated counterflow transport, requires the presence of a substrate on both sides of the membrane. Reorientation of the ternary Enzyme:Substrate:Cation complex is followed by release of the substrate and cation on the trans side. Subsequently, a substrate and cation are bound on the trans side and the ternary complex reorients to the cis side of the membrane, where the substrate and cation are released. No net cation or substrate transport occurs, but by using different sugars or sugar derivatives as substrate and countersubstrate (e.g., lactose and galactose) this mode of transport can lead to net sugar import.

Electrogenic substrate transport (Fig. 5B) requires the presence of either a substrate gradient, or a cation or charge gradient (i.e., proton motive force (\(\Delta p\))-driven lactose transport) which drives the accumulation of the co-ligand. In this mode of transport, the reorientation of the ternary Enzyme:Substrate:Cation complex from cis to the trans side is followed by the release of the substrate and cation, but, unlike counterflow transport, no ligands are bound on the trans side of the membrane. Instead, the unliganded carrier reorients to the cis side, leading to a net accumulation of the co-ligand (lactose in case of LacS) at the expense of the electrochemical proton gradient.

Symport requires that the fully liganded and the completely unliganded state of the enzyme can reorient. Mobility of the partial liganded complexes (Enzyme:Substrate or Enzyme:Cation) will lead to dissipation of the gradient(s) driving the transport and disables accumulation of the co-ligand. Examples of such uncoupled proteins are LacS(E67C) and LacS(E379Q), mutants that have lost the ability to accumulate lactose in the presence of a proton motive force, while lactose counterflow proceeds at wildtype rate (82,122).
3.3. Regulation of transport activity

For members of both the prokaryotic MelB and LacS GPH-subfamilies, the specific activity is adjusted to the energy status of the cell by regulatory mechanisms. In *E. coli*, the presence of a PTS substrate leads to dephosphorylation of IIA\textsubscript{Glc}, either via direct transfer of the phosphate group to glucose via IIBC \textsubscript{Glc}, or via rerouting of the phosphoryltransfer to other EII PTSs. The dephosphorylated species of IIA \textsubscript{Glc} interacts with several permeases, among which MelB and LacY, thereby inhibiting their activity (88). This mechanism is termed inducer-exclusion. Based on the structure of the IIA\textsubscript{Glc}-glycerol kinase complex (55) and regions in IIA\textsubscript{Glc} involved in LacY binding (105), most probably, the surface of IIA\textsubscript{Glc} that interacts with HPr and holds the site of phosphorylation (Fig. 6) (128), interacts with MelB as well. Mutants of MelB resistant to IIA\textsubscript{Glc} inhibition contain mutations in the C-terminus (67) and loop 10-11 (89). Additionally, motifs proposed to be involved in the interaction with IIA\textsubscript{Glc} are located in loop 2-3 (106) and loop 6-7 (19) of MelB. It is not known whether inhibition by IIA\textsubscript{Glc} requires a substrate-dependent conformation as suggested for LacY (107).

As stated above, members of the LacS subfamily differ from the other GPH proteins by the presence of a C-terminal IIA domain, homologous to IIA\textsubscript{Glc}. This domain is not required for
transport (83), but is involved in the regulation of the carrier domain. LacS-IIA can be phosphorylated by HPr(His–P) on a histidine residue (85). The rate of phosphoryl transfer between HPr and LacS-IIA is approximately four orders of magnitude slower than between HPr and IIA\textsubscript{Glc} from \textit{E. coli}, which is in line with a role of LacS-IIA in regulation rather than energization of transport (43). Phosphorylation of LacS-IIA accelerates the rate of lactose counterflow by increasing the maximal uptake rate, while only marginally affecting lactose accumulation driven by the $\Delta$\textit{p} (41). In \textit{S. thermophilus}, phosphorylation of LacS is part of the response of the cell to a limiting lactose concentration and allows high lactose transport rates for prolonged periods of time (42).

3.4. Quaternary structure

Although the projection map of 2D-crystals of MelB (45) does not comply with a dimeric state, both LacS and the xyloside transporter, XylP, a member of the GusB-subfamily, from \textit{Lactobacillus pentosus}, have been shown to form dimers. In the detergent-solubilized state, LacS and XylP undergo reversible self-association with a monomer to dimer mode of association, whereas membrane-reconstituted LacS and XylP seem to form dimers only (32,48,109). Since XylP lacks a IIA domain, the subunit interface must be located in the membrane-embedded carrier domain. Within the LacS dimer, functional interactions between the subunits take place. During lactose counterflow transport both subunits function independently, suggesting that each subunit holds a full translocation path and ligand binding sites. However, cooperative behaviour was observed during $\Delta$\textit{p}-driven lactose accumulation. Since $\Delta$\textit{p}-driven transport requires the reorientation of the unliganded carrier (Fig. 5B), unique for this mode of transport, it was proposed that during this step both subunits have intimate functional interactions (124).

![Figure 6. Ribbon diagram of IIA\textsubscript{Glc} from \textit{E. coli}. Residues of IIA\textsubscript{Glc} shown to be involved in the interaction with glycerol kinase (GK) and/or HPr from \textit{E. coli} (summarized in (128)) are color coded dark grey. His-90, that is phosphorylated by HPr, is colored black and shown in ball and sticks representation. A, B and C represent top and 90°-rotated sideviews, respectively. The structure of IIA\textsubscript{Glc} has the PDB accession number 1F3G.]

4. LacS, the lactose transporter from \textit{Streptococcus thermophilus}

\textit{Streptococcus thermophilus} is a Gram-positive lactic acid bacterium that is highly adapted to living in milk and metabolizing the main milk sugar lactose. Analysis of its genome revealed only a limited number of functional genes coding for sugar transporters, carbohydrate degradation and fermentation, compared to other streptococci. Additionally, it contains the
dedicated lactose transporter LacS, absent in most other streptococci (8). Although the published annotation of the genome suggests that LacS is the only secondary transport protein present in *S. thermophilus*, BLAST searches suggest the presence of at least one other secondary transporter, a homologue of the riboflavin transporter RibU from *Lactococcus lactis* (C. Burgess, personal communication). Nevertheless, such a low number of secondary transporters is highly unusual.

LacS catalyzes the symport of galactosides and protons. Internalized lactose is hydrolyzed to galactose and glucose by β-galactosidase. In most *S. thermophilus* strains, only the glucose moiety can enter metabolism, whereas the galactose moiety is excreted via LacS (70). Several characteristics indicate that LacS functions as an efficient exchanger of extracellular lactose for intracellular galactose, and thereby functions as a net glucose importer: i) substrate binding is specific for the (non-metabolizable) galactose moiety of lactose; ii) the affinity of LacS for galactosides is higher on the cytoplasmic side than on the outside-facing side; iii) for both binding sites the affinity is higher for galactose than lactose (125).

Although favourable when lactose is abundantly present, when most of the lactose is metabolized and converted to galactose, resulting in a low lactose and high galactose concentration in the external medium, the specificity and high affinity for galactose decrease the lactose import rate. *S. thermophilus* responds rapidly to the decreased lactose availability by increasing the specific activity of the transporter by phosphorylating the regulatory IIA domain. Furthermore, as a slower response, the amount of transporters in the membrane is increased by releasing CcpA/HPr(Ser-P)-complex mediated inhibition of lacS transcription (42). Taken together, the physiology of *S. thermophilus* seems highly adapted to life in milk. LacS forms one of the key-enzymes allowing the bacterium to inhabit this peculiar ecosystem.

5. **Scope of this thesis**

Detailed structural and functional studies on proteins will ultimately lead to complete understanding of the mechanisms behind enzyme functioning. This will allow the rational engineering of proteins and substrates to enable catalysis of artificial reactions or to modify enzyme functioning in a predetermined manner. Although this comes gradually within reach for soluble proteins, the knowledge on membrane proteins lags far behind. Among other things, this is caused by practical challenges involved in studying membrane proteins, requiring the development of dedicated methodologies. Here, new methodological insights in the functional amplification of membrane proteins are presented. Moreover, this thesis contributes insights in the transport mechanism of secondary transporters from the Major Facilitator Superfamily by describing functional and structural aspects of domain and subunit interfaces.

Amplification of membrane proteins forms one of the major bottlenecks in the path to structural information. Increased rates of membrane protein production often result in the formation of aggregated material (inclusion bodies), from which the extraction of protein is complicated. Owing to the preservation of structure and the ease of protein isolation from the membrane-inserted state, expression in the membrane is generally preferred over isolation from inclusion bodies. In **chapter 2**, the functional overexpression of LacS, using the AraC/P_{BAD} system, is described. Recommendations are presented to optimize overexpression of membrane proteins in general.

**Chapter 3** describes an examination of the role of the LacS-IIA domain in regulating transport activity. Previous studies have shown that LacS-IIA can be phosphorylated by HPr(His~P), and that phosphorylation results in an increase in the rate of lactose counterflow transport.
However, it remained to be answered whether transport by the LacS carrier domain was inhibited by dephosphorylated LacS-IIA, or stimulated by the phosphorylated species. By comparing wildtype and a truncation mutant lacking the IIA domain, we show that phosphorylation of IIA stimulates the translocation activity. A new model on the mode of regulation by LacS-IIA is proposed.

In the membrane, LacS forms a dimer and functional interactions have been shown to occur between the subunits. In order to unravel the mechanistic aspects of the communication between the subunits, one preferentially would like to have high-resolution structural data of the complex. Due to the failures to crystallize LacS thusfar, cysteine cross-linking was used to obtain low-resolution structural information on the contact region of the subunits. The identification of residues near the subunit interface in the LacS dimer is described in chapter 4.

In chapter 5, the current view of the extent and nature of the communication between the subunits is expanded. By studying the interactions between subunits in a covalently linked dimer, using the wildtype protein linked to a mutant completely inactive in transport, additional functional interactions within the LacS dimer were revealed.

In chapter 6, physical and functional interactions between LacS and β-galactosidase (LacZ) are explored. These enzymes catalyze the first two steps in the lactose metabolism of S. thermophilus, that are, uptake and hydrolysis of lactose, and it was unknown whether a close association of the enzymes contributed to the efficient utilization of lactose. The effect of LacZ on the rate of lactose transport by LacS is described.