Property-controlling Enzymes at the Membrane Interface

Changrong Ge
Cover picture: Proposed membrane orientation of monoglucosyldiacylglycerol synthase from *Acholeplasma laidlawii*. 
To my family

献给我的家人
Abstract

Monotopic proteins represent a specialized group of membrane proteins in that they are engaged in biochemical events taking place at the membrane interface. In particular, the monotopic lipid-synthesizing enzymes are able to synthesize amphiphilic lipid products by catalyzing two biochemically distinct molecules (substrates) at the membrane interface. Thus, from an evolutionary point of view, anchoring into the membrane interface enables monotopic enzymes to confer sensitivity to a changing environment by regulating their activities in the lipid biosynthetic pathways in order to maintain a certain membrane homeostasis. We are focused on a plant lipid-synthesizing enzyme DGD2 involved in phosphate shortage stress, and analyzed the potentially important lipid anchoring segments of it, by a set of biochemical and biophysical approaches. A mechanism was proposed to explain how DGD2 adjusts its activity to maintain a proper membrane. In addition, a multivariate-based bioinformatics approach was used to predict the lipid-binding segments for GT-B fold monotopic enzymes. In contrast, a soluble protein Myr1 from yeast, implicated in vesicular traffic, was also proposed to be a membrane stress sensor as it is able to exert different binding properties to stressed membranes, which is probably due to the presence of strongly plus-charged clusters in the protein. Moreover, a bacterial monotopic enzyme MGS was found to be able to induce massive amounts of intracellular vesicles in *Escherichia coli* cells. The mechanisms involve several steps: binding, bilayer lateral expansion, stimulation of lipid synthesis, and membrane bending. Proteolytic and mutant studies indicate that plus-charged residues and the scaffold-like structure of MGS are crucial for the vesiculation process. Hence, a number of features are involved governing the behaviour of monotopic membrane proteins at the lipid bilayer interface.
List of Publications

This thesis is based on the following papers, which will be referred to in the text by their Roman numerals:


III. A Georgiev*, **C Ge***, Å Wieslander. (2011) Basic Clusters and Amphipathic Helices Contribute to Interactions of Myr1/Syh1 with Membrane Phospholipids. (Manuscript)


* These authors have contributed equally to the paper.

Reprints were made with permission from the publishers.

**Additional publications:**

### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CL</td>
<td>Cardiolipin</td>
</tr>
<tr>
<td>DGD</td>
<td>Digalactosyldiacylglycerol synthase</td>
</tr>
<tr>
<td>DGD1</td>
<td>Digalactosyldiacylglycerol synthase 1 from <em>Arabidopsis thaliana</em></td>
</tr>
<tr>
<td>DGD2</td>
<td>Digalactosyldiacylglycerol synthase 2 from <em>Arabidopsis thaliana</em></td>
</tr>
<tr>
<td>DGS</td>
<td>Diglucosyldiacylglycerol synthase from <em>Achleplasma laidlawii</em></td>
</tr>
<tr>
<td>GalDAG</td>
<td>1,2-diacyl-3-O-(β-d-galactopyranosyl)-sn-glycerol</td>
</tr>
<tr>
<td>GalGalDAG</td>
<td>1,2-diacyl-3-O-[α-d-galactopyranosyl-(1→6)-O-β-d-galactopyranosyl]-sn-glycerol</td>
</tr>
<tr>
<td>GlcDAG</td>
<td>1,2-diacyl-3-O-(α-d-glucopyranosyl)-sn-glycerol</td>
</tr>
<tr>
<td>GlcGlcDAG</td>
<td>1,2-diacyl-3-O-[α-d-glucopyranosyl-(1→2)-O-α-d-glucopyranosyl]-sn-glycerol</td>
</tr>
<tr>
<td>MGS</td>
<td>Monoglucosyldiacylglycerol synthase from <em>Achleplasma laidlawii</em></td>
</tr>
<tr>
<td>MGD1 (2, 3)</td>
<td>Monogalactosyldiacylglycerol synthase 1, 2, and 3, respectively, from <em>Arabidopsis thaliana</em></td>
</tr>
<tr>
<td>Myr</td>
<td>Homolog of suppressor of myo2 mutant in Yeast</td>
</tr>
<tr>
<td>PA</td>
<td>Phosphatidic acid</td>
</tr>
<tr>
<td>PC</td>
<td>Phosphatidylcholine</td>
</tr>
<tr>
<td>PE</td>
<td>Phosphatidylethanolamine</td>
</tr>
<tr>
<td>PG</td>
<td>Phosphatidylglycerol</td>
</tr>
<tr>
<td>PI</td>
<td>Phosphatidylinositol</td>
</tr>
<tr>
<td>PIs</td>
<td>Phosphoinositides, phosphorylated derivatives of PI</td>
</tr>
<tr>
<td>PS</td>
<td>Phosphatidylserine</td>
</tr>
<tr>
<td>SQDG</td>
<td>Sulfoquinovosyl diacylglycerols</td>
</tr>
</tbody>
</table>
1. Introduction

Biological membranes are not only the physical permeability barriers separating cells from the environment, but also the sites where numerous biochemical events take place. A great variety of molecules require the licenses granted by the membrane to travel across this 5~8 nm thick and greasy “fence”. The *structural base* of a biological membrane consists primarily of a continuous elastic lipid bilayer with a diverse set of proteins associated with or embedded into the lipid bilayer. Usually, there are hundreds of physicochemically distinct lipid species in a typical lipid bilayer, but their distribution varies both in time and in space. This permits the characteristic diversity of membrane shapes of intracellular organelles and cells, resulting in a myriad of life forms. In particular, a growing body of evidence demonstrates that membrane proteins play a leading role in shaping the lipid bilayer, hence the life forms.

However, organisms are constantly subject to internal and/or external stimuli and have to respond by adjusting metabolic pathways in order to maintain proper functioning. Since the membrane is the first physical line opposing the external stimuli and also a mechanical fence for protecting the cell, its homeostatic condition is of high importance. It is known that the membrane properties, including the lipid composition, the acyl chain packing, the lipid phase transitions, surface charge, and the stored elastic curvature stress, can be regulated by the cell to meet different environmental conditions. However, the molecular mechanisms underlying these biological regulations remain poorly understood. One of the intriguing questions is how the cell is able to sense and respond to a changing environment. Unlike proteins, there is no genetic code for the lipids, therefore the other major component in the membrane - membrane proteins (protein/lipid ratio in *Escherichia coli* plasma membrane ≈ 3/1, w/w), especially those enzymes implicated in lipid biosynthesis pathways, are believed to play important roles in controlling lipid properties in order to maintain membrane homeostasis. Membrane proteins are a specialized group of proteins in that they carry out a diverse set of vital cellular functions in the membrane rather than in the cytosol. This implies that the interaction between membrane proteins and surrounding lipids might play important roles in regulating protein functions, which would eventually affect the lipid profiles as well as membrane properties.

Hence, the *objectives* of this thesis are to understand how membrane proteins, especially certain interface enzymes involved in lipid biosynthesis by sitting at the membrane interface, are able to sense and respond to changes in the lipid environment. The first part of the thesis is devoted to the background by generally discussing the membrane lipids, membrane proteins, and the interactions between them. The second
part is the summary of the Papers on which the thesis is mainly based. Paper I and Paper II are focused on a plant glycolipid-synthesizing enzyme by elucidating the structural features that govern its membrane binding properties under condition of fluctuating phosphate supply. In comparison, an analogous bacterial enzyme has been analyzed in Paper IV and Paper V, in which the mechanisms underpinning an unexpected vesiculation process were investigated to understand how the enzyme deforms the plasma membrane into variously sized vesicles in the bacterial cytoplasm. In Paper III, a soluble yeast protein was found to be able to sense “membrane stress” by its transient association with vesicular traffic components.
2. Biological membranes

If there were no cell membranes, there would be no life on earth. Life was probably established when a membrane had emerged, enclosing “something” in a defined compartment and separating it from the surroundings. Cellular membranes are first proposed in late 19th century to be a “lipid-impregnated boundary layer” (1), since then the model has evolved for decades to fit new experimental observations, until 1972, when the modern view of membrane structure, known as the fluid mosaic model (2), was presented. It reflects some basic features of membrane structure - a fluid structure with many proteins embedded in, or attached to the lipid bilayer where all lipid and protein molecules diffuse more or less easily. However, advances in membrane structural and biological studies over time unfolded a more complex picture - the membrane is more like a mosaic two-dimensional fluid with heterogeneous lipid and protein regions varying in composition and thickness (3). This biological boundary is characterized by a 5 to 8 nm thick membrane where numerous molecular processes take place. The biological membrane is selectively permeable as it harbors a variety of channels and transporters that are involved in exchanging numerous substances between the cell (or organelle) and the environment. In addition, the plasma membrane contains different micro-domains, different lipid composition, and different protein-lipid ratio, etc. Moreover, the flexibility endows membranes with the ability to generate extraordinarily diverse shapes of cells and organelles.

Phospholipids and glycolipids constitute the major lipid classes in biological membranes, and offer a continuous and amorphous matrix. In contrast to the dominant lipid bilayer matrix, which mainly provides structural support to the cell or the cellular organelle, membrane proteins play important cellular functions, such as signal transduction, energy transduction, intracellular communication, lipid synthesis, and protein translocation. Membrane proteins tend to associate with each other in the planar membrane space as oligomeric or heteromeric aggregates. These membrane protein complexes are normally resistant to the disruption of lipid mimicking molecules - detergents which are widely used to solubilize membrane proteins from native membrane. Likewise, certain lipids are also prone to be segregated as patchy domains. The membrane offers a meeting point for lipids and proteins where they are “playing with” each other in certain cellular processes through specific interaction (4). Therefore, cells need to sense and respond to environmental situations not only by regulating protein biosynthesis, but also by adjusting membrane lipid properties, such as composition, acyl chain length, unsaturation level, and spontaneous curvature.
2.1 Membrane lipids

Non-covalent interactions, such as van der Waal forces, electrostatic interactions, and hydrogen bonds, are all contributing to the maintenance of this continuous lipid bilayer in which the hydrophobic effect is the major driving force. In eukaryote cells, a high number of genes encode enzymes synthesizing a variety of lipids, which implies that lipid is not only the key element for the physical membrane barrier, but also a critical component in many cellular functions (5). The major biological functions of lipids in the cell include membrane barrier, energy storage and cellular signaling.

Membrane lipids can be divided into three main classes: glycerolipids, sphingolipids and sterols (6). For simplicity, I will only discuss the glycerolipids, which are the main lipid species I have been focusing on in my PhD study. Glycerolipids can be further classified into two major subgroups based on their head-group properties - glycerophospholipids and glyceroglycolipids. Glycerol lipids are composed of a glycerol backbone in which the two hydroxyl group positions sn-1 and sn-2 are substituted with fatty acids through ester (or ether) bonds to generate two hydrophobic chains, and the third sn-3 hydroxyl group can be substituted with a variety of moieties such as phosphate, alcohols, amino acids, or sugars (e.g. glucose/galactose) to constitute the polar head-group. The variation of chains and head-groups bound to the glycerol backbone, allows thousands of glycerolipids with different physical and chemical properties to exist in an eukaryotic cell.

Glycerophospholipids

Glycerophospholipids, often referred to as phospholipids, carrying a polar phosphate head group is the predominant group in animal, yeast and many Gram-negative bacteria. Based on the polar phosphate head-group properties, the major structural and functional phospholipids in biological membranes are phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylserine (PS), phosphatidylinositol (PI), phosphatidylglycerol (PG), cardiolipin (CL) and phosphatidic acid (PA) (7). Their structures are shown in Fig. 1. However, their existence and abundance are indeed dependent on the cell types and organelles. For instance, the phospholipids in Escherichia coli inner membrane consist of 70-80% PE, 20-25% PG, and 5-10% CL (8), but with trace amount of metabolic intermediates of other lipid species. PC accounts for about 50% of the phospholipids in most mammalian endoplasmic reticulum membranes, Golgi membranes and plasma membranes (9). Mitochondria contain much more CL than other organelles, which may reflect its bacterial origin.
The minor lipid species - PI, is an important second messenger participating in essential metabolic processes in plants, fungi and animals, as it can be phosphorylated by a series of kinases on hydroxyl groups of the inositol ring to generate seven different PI derivative lipids that are also involved in various vesicular traffic events (11-13). Although there is a wealth of data concerning the specific biological roles of these phospholipids, the mechanisms behind such uneven distribution still remain unclear. Moreover, these phospholipids are also asymmetrically distributed across the lipid bilayer, vertically as well as laterally. This non-uniform distribution is regulated by the cell to fulfill certain biological roles but is also determined by the individual lipid species mainly implicated in specific cellular process. In the eukaryote cell, the main lipid synthesis occurs in the endoplasmic reticulum (ER) (14), which produces most of the lipids, although Golgi is also a site for lipid synthesis and sorting (15). Conversely, in the bacterium E. coli, the plasma membrane is the site for phospholipid synthesis (16). Phospholipid synthesis is found to proceed in the inner membrane from where mature lipids can be translocated to the outer membrane. It is noted that most of the enzymes involved in phospholipid biosynthesis in E. coli are membrane-bound. These membrane-bound lipid-synthesizing enzymes can be affected by the biophysical properties of the membrane, such as lipid composition, saturation of acyl chains and surface charge density. In turn, they are also able to regulate lipid synthesis to maintain a membrane homeostasis (17).

Figure 1. Schematic representations of common phospholipids. Structures of phospholipid polar headgroups, R1, R2, R1’ and R2’ refer to fatty acyl chains.
Glyceroglycolipids

Glyceroglycolipids which contain glucose or galactose, in some cases other sugars with different glycosidic linkages in their head group, are the major abundant class in algae, higher plant, and Gram-positive and many photosynthetic bacteria. Monogalactosyl-diacylglycerol (GalDAG) and digalactosyl-diacylglycerol (GalGalDAG), also referred to as galactolipids, are the most predominant lipids in thylakoid membrane of chloroplasts (and on Earth) in which they together constitute about 70% of total lipids. Sulfoquinovosyl-diacylglycerol (SQDG) and PG are the other two major structural components and are also found at a significant level in plant leaves (18, 19). The structures of these three glycolipids are illustrated in Fig. 2A. The two glycolipids GalDAG and SQDG are exclusively located in plastid membranes, while GalGalDAG and PG are also found in extraplastidic membranes (18, 20). Moreover, all these four lipids are also found as integral lipids in the protein crystal structures of the photosystem II core complex and located in the interfaces among protein subunits (21-23). Besides, analyses of crystal structures of other photosynthetic systems such as the photosystem I complex (24, 25), the Cytochrome b$_6$f complex (26-28), and the light harvesting complex (29), also reveal the presence of galactolipids. Taken together, this may indicate that, galactolipids are not only the bulk structural components of thylakoid membranes but also key players in photosynthesis, and this is well in agreement with the results from the study of galactolipid-deficient A. thaliana mutants (30, 31). In addition to participating in photosynthesis, galactolipids, especially GalGalDAG are also found to play important physiological roles in maintaining cellular membrane homeostasis under certain stress conditions (32-34). For instance, during phosphate shortage, GalGalDAG was accumulating in plastidial and extraplastidial membranes to substitute the deprived phospholipids (35-37).

It has been hypothesized that there are two parallel pathways for galactolipid biosynthesis (GalDAG and GalGalDAG) in plants - the eukaryotic and prokaryotic pathways, involving lipid-synthesizing enzymes residing both in ER and plastid membranes (38). In the eukaryotic pathway, potential galactolipid precursors such as PA (39), PC (40, 41), DAG (42) and lysoPC (43) are thought to be assembled in the ER and then transported to the plastid for galactolipids synthesis. So far, it is still unclear how these precursors are transported through this pathway, though several possible mechanisms have been proposed (38, 44), such as vesicular traffic (45-47), physical association between ER and plastid membrane (48, 49), and spontaneous partition of lyso-PC across the cytosol (50). Conversely, in the prokaryotic pathway, the precursor lipid PA is assembled entirely in plastids followed by formation of GalDAG and GalGalDAG (38).
Interestingly, in non-photosynthetic cell wall-less bacterium *Acholeplasma laidlawii*, almost 50% of total lipids in plasma membrane are made up by two different glycolipids - monoglucosyl-diacylglycerol (GlcDAG) and diglucosyl-diacylglycerol (GlcGlcDAG) (Fig. 2B) in which the headgroup region contains glucose moieties rather than galactose. These two glycolipids are also found in related Gram-positive bacteria. It was also shown, that the molar ratio between these two glycolipids is crucial for maintaining proper bilayer packing properties in the *A. laidlawii* plasma membrane (51, 52).

Several enzymes implicated in the glycolipid biosynthetic pathways of bacteria and higher plants have been identified and described recently. Most of them are believed to consist of integral monotopic membrane proteins. In higher plants such as *A. thaliana*, three monogalactosyl-diacylglycerol synthases (namely MGD1, MGD2 and MGD3) and two digalactosyl-diacylglycerol synthases (namely DGD1 and DGD2), have been shown to participate in the glycolipid synthesis. All these enzymes are...
UDP-Gal dependent glycosyltransferases (GTs) and localized in the chloroplast envelope membranes (53, 54). In contrast, in most Gram-positive bacteria only one homolog has been found with monoglycosyl-diacylglycerol or diglycosyl-diacylglycerol synthase activity, such as in A. laidlawii (named aMGS and aDGS), which are both UDP-Glc dependent glycosyltransferases (55-57).

2.2 Membrane curvature

The plasticity of biological membranes enables itself to be reconstructed by the cell into various membrane structures with a great diversity of shapes, exemplified by the intracellular transport vesicles varying in a broad range of sizes in eukaryotic cells. During endocytosis, a small fraction of the plasma membrane bulges inward and is pinched off from the plasma membrane, then reshaped into vesicles, which are transported to different compartments for processing. Besides, the complex structures of ER membranes, and the Golgi apparatus is also interconnected to a network, comprised of tubles, cylinders and disc-shaped membranes (58, 59). Membrane shape can be geometrically regarded as membrane curvature, which is the consequence of interplay between lipid and protein “packing shapes” in the membrane (60). An increasing body of evidence suggests that membrane proteins provide the leading force for shaping the membrane by either direct or indirect mechanisms, which can change the elasticity of membrane (59, 61-65).

Spontaneous curvature

The intuitive driving force to generate membrane curvature or shape a planar membrane is related to lipids, which have distinguishable physical-chemical properties from each other. Based on physical packing shapes, lipids can be classified into three main groups - cylindrical, conical and inverted conical shapes (66) (Fig. 3A). The different shapes, caused by either the saturation/unsaturation of acyl chains, or the different relationship between polar head group size and the acyl chain lateral areas, are related to the spontaneous curvature of the lipids (67, 68). Spontaneous curvature is an intrinsic property that can be determined by physical properties of a given lipid molecule. Cylindrical shaped lipids, such as PC, GlcGlcDAG, and GalGalDAG with similar lateral size of polar head group and hydrophobic acyl chain, exert zero force to form spontaneous curvature and therefore form planar lipid bilayers; Conical lipids like PE, GlcDAG, and GalDAG, having smaller head group cross-sectional area than...
that of their acyl chains, can form negatively curved lipid bilayer; In contrast, positive spontaneous curvature is exerted by inverted conical lipids like lysophospholipids with smaller chain area than that of their polar head groups \((69, 70)\). Cylindrical lipids are also referred to as bilayer-prone lipids, while the conical and inverted conical lipids are therefore non-bilayer prone. Individual lipids in bilayers, interacting with surrounding lipids alongside the membrane normal by attraction and repulsion, give rise to the so-called \textit{lipid lateral stress profiles} describing the molecular forces present at different depths of the cross-sectional lipid bilayer \((71-73)\).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{lipid_packing_shapes.png}
\caption{Lipid packing shapes and membrane curvature. A. Lipids with different spontaneous curvature; B. Generation of a planar membrane by combating the opposite bending tendency due to the presence of non-bilayer prone lipids.}
\end{figure}

\textbf{Stored curvature elastic stress}

A biological membrane is not a homogeneous lipid matrix, but consists of two lipid monolayers with a number of lipid species. Different lipid species exhibit different spontaneous curvatures either positive or negative. Therefore, there is a need to insert different lipid species into a planar membrane in order to generate a functional lipid bilayer entity by overcoming the spontaneous bending tendency due to the presence of non-bilayer prone lipids \((74)\) (Fig. 3B). Even though a planar membrane is formed, the individual non-bilayer prone lipid is forced to pack in a cylindrical compartment shape rather than its preferable conical or inverted conical shape. The energy cost in packing together different lipid species is referred to as \textit{stored curvature elastic stress} which
can be either enhanced or relieved by integral or peripheral membrane proteins. A couple of membrane proteins (*e.g.* lipid-synthesizing enzymes) are shown to be able to sense and respond to the stored curvature stress under certain stress conditions in order to maintain proper cellular functions (52, 75-77). This can be achieved by a so-called feedback mechanism in which the membrane-bound proteins (enzymes) can be modulated to synthesize the appropriate amount of desirable lipid products, which in return influence the activity of corresponding enzymes.

**Membrane bending**

Since membrane lipids in the cell do exhibit spontaneous curvature and there is intrinsic stored elastic stress energy in the planar membrane, could these forces permit the membrane itself to assume a broad range of shapes exhibited by cells or cellular organelles? It was shown (58, 78), that the energy required to bend the membrane is much higher than the energy provided by the thermal fluctuation of membrane lipids alone. Thus, the membrane bending can only be achieved by joint efforts contributed by complex interactions between lipids and proteins (78). A number of proteins, permanently or transiently associating with the membrane, can directly or indirectly shape the membrane into diverse dynamic structures as observed in the cell. It has also been shown, that not only lipids in the membrane can be modified by proteins to change either the spontaneous curvature or the asymmetric distribution across the bilayer, but also proteins can directly impose specific physical constraints on the membrane surface. The former refers to both enzymes like flippases and scramblases, which can translocate lipids between the two leaflets of the bilayer to generate lipid asymmetry across the membrane (79-83). The phospholipase family enzymes (phospholipase A, B, C and D) (84) are able to change spontaneous curvature feature of lipids by hydrolyzing specific bonds in phospholipids. However, in this section, emphasis will be placed on the membrane proteins that can physically impose mechanical forces on the membrane surface. Fig. 4 illustrates the two major strategies by which proteins can physically bend the membrane - hydrophobic insertions (85) and scaffolding mechanisms (59, 86), which are not mutually exclusive but inter-related to a certain extent.

As for the hydrophobic bilayer insertion, some proteins may cause lateral expansion of only one monolayer with respect to its counterpart within the same bilayer, by inserting amphipathic helices or small hydrophobic segments shallowly into the lipid bilayer, which can eventually lead to “squeezing” of the membrane. Epsins (87, 88) were the first proteins shown to deform membrane into tubules by the amphipathic
alpha helix, located in the ENTH domain at its N-terminus and interacting with phosphatidylinositol-4,5-biphosphate. The insertion of the amphipathic alpha helix can force the neighboring lipids in the same leaflet to expand laterally with respect to the other leaflet. This in turn can lead to the space of one monolayer occupied by lipids is greater than that of the other monolayer which eventually results in bending the membrane. Arf1 (89) is a small G protein involved in vesicular trafficking, and can also be anchored to the membrane interface via its amphipathic alpha helices that are embedded into lipid bilayer to deform the membrane.

With regard to scaffolding mechanisms, some proteins either in single or polymeric forms may work as a scaffold to shape the underlying membrane, or stabilize an already deformed membrane due to their intrinsic “banana-like” shape. This can be exemplified by the key participants in vesicular traffic events, such as the clathrin complex (90-92), dynamin proteins (93-95), COPI/II proteins (96), and BAR superfamily proteins (97-99). Clathrin is implicated in the exocytosis process, and can be recruited to the membrane surface, and then polymerize to form a rigid structure locally framing the membrane. N-BAR domain-containing proteins insert an N-terminal amphipathic helix into the lipid bilayer and fit their intrinsic “banana-like” shape to curve membrane surface, in order to exert its curvature-inducing role.

**Figure 4. Two mechanisms for bending membranes.** A, Proteins insert hydrophobic segments into a membrane monolayer, causing the curvature stress. B, Scaffolding proteins have a rigid curved shape (intrinsic or formed by several interacting molecules) interacting with the membrane, forcing the bilayer to adopt the same curvature.
Bacterial membrane vesicles

Membrane vesicles are ubiquitous in eukaryotic cells, and vesicular trafficking plays indispensable roles in various cellular processes such as endocytosis, exocytosis, and protein sorting. These vesicles, budded off from either plasma membrane or organelle membranes, bridge the physical distance gaps by transporting various materials among the different cellular compartments. However, except for some photosynthetic bacteria (100-102), membrane vesicles, especially intracellular vesicles, are rarely observed in prokaryotic cells. This can be explained by both the absence of subcellular compartments and the small size of prokaryotic cells. As usually is the case in prokaryotic cells, the cellular space is small, so the biological molecules are easily accessible to all of the cytoplasmic space by simple diffusion. On the other hand, outer membrane vesicles released from the outer membrane of Gram-negative bacteria into the surroundings have been known for decades (103, 104), and these outer membrane vesicles play important roles in various processes, such as delivering toxins (105), virulence factors (105) and DNA (106), mediating cell-cell communication (107), and presenting antigens for initiating the immune system (108).

Interestingly, it was noted that overexpression of certain endogenous membrane proteins in the Gram-negative bacterium E. coli can usually enlarge cell size, which is probably due to the incorporation of extra foreign proteins into the plasma membrane and the inhibition of cell division, leading to lipid lateral expansion (109). But more than that, in some cases such as overexpressing ATP synthase or its β-subunit (110), fumarate reductase (111), sn-glycerol-3-phosphate acyltransferase PIsB (112), LamB-LacZ hybrid proteins (113), sp6.6 or the chemotaxis receptor Tsr (114), can cause formation of stacked or tubular membrane structures in the cytoplasm. All of them are trans-membrane proteins, and they may form polymeric forms through specific protein-protein interactions between the extended cytoplasmic regions. Therefore, this can probably result in extra membrane formation in the cytoplasmic space in order to accommodate the overexpressed transmembrane proteins.

However, some peripheral membrane proteins were also found to be able to induce intracellular vesicles from the inner membrane when overexpressed in E. coli cells. MurG (115), one of the key enzymes implicated in peptidoglycan precursor biosynthesis, was the first monotopic membrane protein found to generate vesicles under overexpression conditions. MurG is believed to interact with plasma membrane via a hydrophobic patch surrounded by some basic amino acid residues, which is also a characteristic “anchor” feature for several peripheral membrane proteins. CL content was substantially higher in MurG overexpressing cell membranes as well as in vesicles, than in non-overexpressing cells. Therefore, the anionic phospholipid CL,
interacting with MurG, was thought to play a special role in the regulation of peptidoglycan synthesis, probably through the CL synthase. LpxB (116) is involved in synthesizing the bacterial outer membrane constituent lipid A, and also a drug target in developing new antibiotics. Overexpression of LpxB also generated uniform tubules accumulating along the cytosolic side of inner membrane.

Membrane vesicles induced by MGS

So far, to the best of my knowledge, the most striking example for vesicle formation in *E. coli* is by the glycolipid-synthesizing enzyme monoglucosyl-diacylglycerol synthase (MGS) (Paper IV). This enzyme, synthesizing one of the major glycolipids – GlcDAG in *A. laidlawii* by transferring a glucose moiety from UDP-glucose to the head region of diacylglycerol, was able to induce massive formation of intracellular vesicles under certain conditions of overexpression. Most of the vesicles pinched off from the inner membrane vary in size from 50 to 100 nm. Approximately 60% of these vesicles weight are lipids, which is substantially higher than that of the inner membrane (40%). The lipid composition in these vesicles was ~40% PE, ~10% PG, ~10% CL and ~40% of the foreign GlcDAG, whilst wild type *E. coli* cell inner membrane contains 70-80% PE, 20-25% PG, and 5-10% CL. GlcDAG is a nonbilayer-prone lipid, therefore there seems to be no significant difference in terms of the ratio between nonbilayer-prone and bilayer-prone lipids.

As seen from the SDS-PAGE profiles of purified native vesicles, more than 90% of proteins in the vesicles are MGS molecules (Paper IV), which are most likely located on the outer surface of vesicles. 17 unique proteins from the purified vesicles were identified by mass spectroscopy and are listed in Table 1. These distinct functional proteins, normally sorted to inner membrane, outer membrane, or cytoplasm, were simultaneously found in the vesicles, and this indicates that the vesiculation or pinching-off vesicles by MGS is most likely a non-specific process. This may lead to trapping some proteins that are still in the biosynthesis or assembly/folding process.

Interestingly, components of the Sec protein translocon apparatus such as SecA and SecD were identified, and this raised the possibility that some membrane proteins might be translocated to the vesicles through the Sec translocon, which is probably also located in the vesicles. Therefore, one of the potential applications of these vesicles can be to facilitate membrane protein overexpression, which is usually a bottleneck in membrane biology research.
The lipid phase transition profile on these vesicles was also analyzed by DPH fluorescence polarization (unpublished data), and it turns out that there is no obvious influence even in the prevailing presence of MGS molecules. In addition, the distance of the vesicular membrane (lipid bilayer + interfacial water phase) was determined to be approximately 4.30 nm by small angle X-ray diffraction (Ge et al. unpublished data). Normally the thickness of *E. coli* cell native inner membranes is about 3.75 nm (117). Since the size of MGS is roughly 4 × 5 nm, this suggests the insertion of MGS should be substantially deep, as also indicated by MD simulation of other several monotopic membrane proteins (118).

**Table 1. List of vesicle proteins identified by MALDI-MS from SDS-PAGE gel**

<table>
<thead>
<tr>
<th>Swiss Prot Code</th>
<th>Protein Name</th>
<th>Location</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>P21513</td>
<td>Ribonuclease E</td>
<td>Cytoplasm</td>
<td>Hydrolase</td>
</tr>
<tr>
<td>B1LC9</td>
<td>Magnesium-translocating P-type ATPase</td>
<td>IM</td>
<td>Hydrolase</td>
</tr>
<tr>
<td>Q03660</td>
<td>1,2-diacylglycerol 3-glucosyltransferase</td>
<td>IM/Cytoplasm</td>
<td>Glycosyltransferase</td>
</tr>
<tr>
<td>P10408</td>
<td>Protein translocase subunit secA</td>
<td>IM/Cytoplasm</td>
<td>Protein transport</td>
</tr>
<tr>
<td>P00722</td>
<td>Beta-galactosidase</td>
<td>Cytoplasm</td>
<td>Hydrolase</td>
</tr>
<tr>
<td>P0AG91</td>
<td>Protein-export membrane protein secD</td>
<td>IM</td>
<td>Protein transport</td>
</tr>
<tr>
<td>P0ABB2</td>
<td>ATP synthase subunit alpha</td>
<td>IM</td>
<td>Hydrolase</td>
</tr>
<tr>
<td>P0ABB6</td>
<td>ATP synthase subunit beta</td>
<td>IM</td>
<td>Hydrolase</td>
</tr>
<tr>
<td>P0A910</td>
<td>Outer membrane protein A</td>
<td>OM</td>
<td>Porin activity</td>
</tr>
<tr>
<td>P02931</td>
<td>Outer membrane protein F</td>
<td>OM</td>
<td>Porin activity</td>
</tr>
<tr>
<td>P0ABA8</td>
<td>ATP synthase gamma chain</td>
<td>IM</td>
<td>Proton transport</td>
</tr>
<tr>
<td>D3JYW3</td>
<td>TEM-1b beta-lactamase</td>
<td>Cytoplasm</td>
<td>beta-lactamase</td>
</tr>
<tr>
<td>P0A7R5</td>
<td>30S ribosomal protein S10</td>
<td>Cytoplasm</td>
<td>rRNA binding</td>
</tr>
<tr>
<td>Q8X9A8</td>
<td>Succinate dehydrogenase, iron sulfur protein</td>
<td>IM</td>
<td>TCA cycle</td>
</tr>
<tr>
<td>P0A13</td>
<td>ATP-dependent zinc metalloprotease FtsH</td>
<td>IM</td>
<td>Hydrolase</td>
</tr>
<tr>
<td>P0AC43</td>
<td>Succinate dehydrogenase flavoprotein subunit</td>
<td>IM</td>
<td>TCA cycle</td>
</tr>
</tbody>
</table>

Even though the vesiculation mediated by the aforementioned peripheral membrane proteins in *E. coli* cell seems to be an artificial, interfering process, in which the resulting vesicles are physiologically irrelevant, the mechanisms for how these proteins induce vesicles are still unclear. Therefore, investigating the vesiculation mechanisms can provide knowledge not only about how these proteins interact with the membrane, but also how bacterial cells regulate their cellular processes under stress conditions.

MGS was able to produce ~40% GlcDAG of the total membrane lipids when overexpressed in *E. coli* (Paper IV), and could the vesicles we observed be attributed to the presence of the extra GlcDAG in the membrane? Since GlcDAG is a nonbilayer-prone lipid, it may facilitate membrane curvature, which in turn leads to membrane bending. To prove this, several inactive MGS variants were constructed by substituting key residues in the conserved EX7E motif (Paper IV). All MGS variants were found still to be able to generate massive amounts of vesicles in the absence of
GlcDAG. Hence, the vesiculation is most likely attributed to the overall structural features of the MGS protein rather than its nonbilayer-prone lipid product GlcDAG. This conclusion is further supported by the other structurally similar enzyme DGS (Paper IV), synthesizing GlcGlcDAG from the lipid product GlcDAG of the MGS enzyme. DGS can also induce intracellular vesicles in the absence of its substrate GlcDAG in *E. coli*, though in less amount than MGS under similar conditions.

Figure 5. Proposed vesiculation mechanisms for MGS. Schematic illustration of the steps of how the monotopic protein MGS inserts into the membrane interface, expands the inner membrane, preferentially interacts with anionic lipids and causes bending of the lipid bilayer, eventually leading to vesiculation (Paper IV).

However, several possible steps in the mechanisms have been proposed to explain why MGS is able to deform membranes into vesicles (Paper IV and V). Some of them are sketched in Fig. 5. Firstly, MGS is a monotopic protein, and it interacts with membranes by penetrating into the lipid bilayer without reaching out of the other side (monolayer) of the membrane. The hydrophobic patches at the lipid contact surface of MGS may be intercalated into the membrane like a wedge, and this may cause one lipid monolayer to expand laterally due to the lipid area difference generated at both ends of the “wedge”. Under strong overexpression conditions, more and more “wedge” shaped MGS will be inserted into the membrane and hence cause continuous
lipid lateral expansion. This would generate membrane curvature and finally “squash” the membrane to bend inward.

Secondly, given the overall structure of MGS, it was noted, that MGS is a bent shaped protein with a cleft constituting the catalytic region between the two domains. Besides, MGS was also shown to interact with membranes through both its two structurally similar domains, but with different binding strengths (Paper V). This indicates that MGS may bend the membrane by either imposing its own intrinsic curvature forces to the membrane surface due to its bent shaped features, or scaffolding the membrane by polymerized forms. Vesiculation can be affected or even abolished by MGS variants in which the overall bent shape of MGS was disrupted by genetically splitting MGS into two single domains (Paper V). Moreover, truncating the C terminus regions (blue region in Fig. 6), which structurally encompasses the N domain for rigidifying the bent shape, can totally inhibit its vesiculation (Paper V).

![Figure 6. Structural model of A. laidlawii glycolipid glycosyltransferase MGS.](image)

The truncated C-terminal region (residue 371 to 398, packing back to N domain) in Paper V is marked in blue. The unmodeled very C-terminal region (residue 388 to 398) is also indicated by blue dotted line.

Finally, there are clustered positively charged residues on the N domain surface of MGS, especially one segment with paired KR and RK residues that may interact strongly with anionic phospholipids like PG and CL in the plasma membrane (Fig. 7). It has been proposed that phospholipid synthesis in *E. coli* is governed by the surface (anionic) charge density of the membrane, which is mainly contributed by acidic
phospholipids PG and CL (17). PG is also a pace-keeper in the biosynthetic pathway of phospholipid synthesis in *E. coli*, by regulating its synthesis to maintain membrane homeostasis through a feedback mechanism (120). Therefore, PG synthesis is interregulated by the other major zwitterionic phospholipid PE. In other words, the more PG synthesized, the more PE synthesized. Theoretically, MGS could neutralize the negatively charged surface by interacting with PG and CL through its positively charged residues, and this could influence activities of membrane bound lipid-synthesizing enzymes to increase the synthesis rate of anionic lipids PG and CL. In order to keep a membrane lipid homeostasis, the synthesis of zwitterionic lipid PE must also be up-regulated to respond to the increased amounts of PG and/or CL. So the increased amount of phospholipids, especially nonbilayer-prone ones, such as PE and GlcDAG, can further facilitate the vesiculation process since it can meet the demand for more lipids to generate vesicles.

Figure 7. Membrane lipid biosynthesis in *E. coli* cell. The three major phospholipids are indicated in red and major lipid-synthesizing enzymes are indicated in green. PlsB, glycerol-3-phosphate O-acyltransferase; PlsC, 1-acyl-sn-glycerol-3-phosphate acyltransferase; CdsA, CDP-diacylglycerol synthase; PssA, phosphatidylserine synthase; Psd, phosphatidylserine decarboxylase; PgsA, phosphatidylglycerophosphate synthase; PgpABC, phosphatidylglycerophosphate phosphatase; Cls, CL synthase; GP, *sn*-glycerol-1-phosphate transferase; Dgk, diacylglycerol kinase.
2.3 Membrane traffic

Membrane traffic is an essential cellular process, involving numerous steps and components, by which biological materials are transported among an array of functionally distinct membrane-bound compartments. There are two major membrane trafficking pathways in eukaryotic cells. One is called the secretory pathway by which proteins and other macromolecules are transported to the various destinations inside or even outside of the cell, while the other is the endocytic pathway by which extracellular substances are internalized into the cell. All cellular compartments, including the endoplasmic reticulum, Golgi apparatus, endosomes, lysosomes, mitochondria, chloroplasts, and the plasma membrane are involved in membrane trafficking. Membrane traffic is generally mediated by small vesicles. These vesicles are generated from one “donor” membrane and are then transported to the “target” membrane where materials are exchanged by fusion and fission processes.

The secretory pathway consists of a set of sequentially inter-connected compartments. This pathway starts from ER, and then continues through the intermediate compartment, cis-Golgi network, and to the trans-Golgi network, where vesicles are ready for sorting. Secretory proteins, as well as lipids and carbohydrates, are modified in Golgi and then transported to the trans-Golgi network where vesicles carrying the cargo molecules are sorted to various destinations by the cellular sorting machineries. Through this secretory pathway, different cellular compartments are functionally linked to provide a series of posttranslational modifications of proteins. Conversely, the endocytic pathway is initiated at the plasma membrane, then the resulting vesicles formed are transported through different stages of endosomes, including early endosome and late endosome, to the Golgi stacks for sorting or to the lysosome recycling (121). There are four major routes by which solutes can be transported through the endocytic pathway: clathrin-mediated endocytosis, caveola-dependent entry route, macropinocytosis, and phagocytosis (122). The basic functions of endocytosis include nutrients uptake, receptor down-regulation, receptor signaling, neurotransmission, and pathogen entry (123).

A number of stage-specific proteins have been identified to be implicated in intracellular membrane trafficking, and they possess various structurally distinct domains that are able to mediate protein-protein interactions and/or protein-lipid interactions. According to their specific roles in cellular transporting machineries, all these proteins can be classified into several subgroups, including Small GTPases accessory proteins, coat proteins, coat adaptor proteins, sorting proteins, fission proteins, fusion proteins, and motor proteins (124-126). A wealth of studies have been
devoted in the past decades to characterize the functions of individual proteins and their specific roles in the overall membrane trafficking.

As the key components of cellular organelles and the transporting vesicles, lipids were shown to play important regulatory roles in membrane trafficking. As mentioned previously, a cell contains hundreds of different lipid species but they are distributed heterogeneously among their subcellular compartments. Lipid species and/or their concentrations are not uniformly distributed among all the trafficking vesicles, but vary at different stages of membrane traffic. Besides, there is also lipid asymmetry across the bilayers of trafficking vesicles (127). In addition to this heterogeneous lipid distribution regulated by the cell, the roles of individual lipid species have also been appreciated in the past years. For instance, Phosphoinositides (PIs), though the minor phosphorylated lipid species in the cell, have been found to be crucial for membrane trafficking as they flag identity of the different membrane compartments (128), such as the PI(4,5)P2 and PI(5)P at the plasma membrane, PI(3)P in early endosome, and PI(3,5)P2 in late endosome or lysosome (13). It has to be pointed out that the distribution of PIs is not universal but vary at various steps of trafficking. In fact, the regulatory roles of PIs in membrane trafficking are tightly linked to their metabolism, which is mainly regulated by PI kinases and PI phosphatases (129, 130). Lipid rafts in the plasma membrane, enriched in sphingolipids and sterols, can favor the fission and fusion processes, as they are able to interact with SNARE proteins in synaptic vesicles (131). Phospholipids like PA, PC, and PS are not only the building blocks for those subcellular compartments, but also found to be important in regulating the membrane trafficking by interacting with different protein components, though their exact cellular functions remain largely unknown. Among various key players involved in membrane trafficking, Phosphoinositides (PIs) are of special interest due to their versatile nature, which is determined by the fast and efficient inter-conversion in phosphatidylinositol and its phosphorylated derivatives. PIs are minor lipid species in the cell, but are distributed in almost all cellular compartments. The local PI species and their concentrations are tightly regulated by the cell to meet the requirements by different cellular compartments in trafficking pathways.

PIs are regarded as organisers during membrane trafficking by recruiting and assembling proteins and/or protein complexes (13). These proteins and/or protein complexes are mainly peripheral or soluble but possessing PI-binding domains, such as the PH, PX, FYVE, ENTH and GLUE domains, which can directly interact with the inositol head-group (132). These PI-binding domains differ in sizes, amino acid composition, secondary structures, affinities and/or specificities for PI species, etc (132). In particular, these PI-binding proteins are not acting alone in associating with trafficking machineries, but cooperate with partners from other proteins to enhance binding affinities.
Given the omnipresence of membrane trafficking in many cellular functions and also the key regulatory roles of PIs, it is conceivable that new PI-binding domains still remain to be characterized. Myr1 (Paper III), a newly proposed membrane stress sensor can bind to membrane lipids but with relatively low affinities, which is also the characteristic feature for most lipid-binding proteins implicated in membrane trafficking. Two selected domains (coiled-coil domain and C-terminal domain shown in Fig. 8) derived from Myr1 were found to bind preferentially to monophosphorylated PIs, but they are not predicted to be or belong to any known PI-binding domain. These two domains are mainly composed of positive charged residues. It was thought, that electrostatic attractions between these residues and negatively charged lipid head-groups, play major roles in mediating their membrane association. However, the observed preferential PIs binding is probably due to its secondary structural features. It has also been noted that, synergistic cooperation between weak lipid binding sites within one protein can promote membrane association by giving enough specificity and strength (132, 162). For instance, two single separate C2 domains in synaptotagmin did not show lipid binding capacity to granules or lipids extracted from granules, both together did act synergistically to bind to PS/PC vesicles (133).

**Figure 8. Schematic illustration of Myr1 with predicted conserved domains.** GYF domain, coiled-coil domain and C-terminal domain are mapped to the primary sequence, and the positively charged residues for selected peptides are marked in blue (Paper III).
2.4 Lipid remodeling in membranes

Actually, all living organisms are persistently challenged by a wide spectrum of biotic and/or abiotic environmental stress factors, therefore various strategies have evolved to maintain cellular homeostasis. Numerous molecular receptors localized in the plasma membrane work as antennas for perceiving and transducing external signals to the cell, and then a number of signaling response networks can be established to cope with those external stimuli. Tons of information in the context of global responses has been collected over the past decades to understand how cells respond to environmental stresses by transcriptome, proteome, and metabolome analyses. For simplicity, focus in this section will be directed to the responses exerted by plants under various environmental stresses including nutrients shortage, drought stress, elevated temperature, freezing, etc (32, 134-142). In particular, lipid remodeling in membranes by a set of lipid-synthesizing enzymes will be discussed. Understanding how membrane proteins (lipid-synthesizing enzymes) sense and respond to the environmental changes by adjustments in expression levels and/or metabolic activities will give us more information about how plants adapt to environments.

Glycolipid biosynthesis

As mentioned briefly, the glycolipids GalDAG and GalGalDAG are the predominant lipid constituents in chloroplast membranes and in cyanobacteria. They are also found as integral lipids in several crystal structures of photosynthetic complexes, suggesting specific structural roles in photosynthesis.

In plants, GalDAG is synthesized by GalDAG synthases, which utilize UDP-Gal and sn-1,2-diacylglycerol (DAG) as substrates, and transfer the galactose from UDP-Gal to the sn-3 position of DAG. So far, three GalDAG synthases have been identified in *A. thaliana*, namely MGD1, MGD2, and MGD3 (53). These three isoforms differ in substrate specificity and subcellular localization. MGD1, referred to as a type A enzyme, utilizes DAG originating from the plastid as substrate, and is considered to be located in the inner envelope membrane of the chloroplast. In contrast, MGD2 and MGD3, referred to as type B enzymes, use DAG imported from ER as substrate, and are located in the outer envelope membrane of chloroplast. (143-145).

One *A. thaliana* weak allele, *mgd1-1* (146), carrying an insertion in the promoter region of the *mgd1* gene, contains only ~40% of GalDAG wild type amounts, which
suggests that the MGD1 enzyme contributes to the bulk GalDAG synthesis. Besides, the reduced amounts of chlorophyll synthesized and abnormal structure of thylakoid membrane in this mutant, further indicate the importance of GalDAG in photosynthesis. In another mgd1 null mutant, mgd1-2 (31), there is complete lack of galactolipids in the cell, leading to severe defects in chloroplast biogenesis, photosynthesis, and embryogenesis. In contrast, MGD2 and MGD3 enzymes are thought to be less important for GalDAG normal bulk biosynthesis, and found to be mainly accumulated (expressed) in non-photosynthetic tissues, such as roots (36, 144).

It was also noted, that the biosynthesis of β-GalDAG in cyanobacteria is different from plants (147-149), even though cyanobacteria are thought to be the endosymbiotic ancestors of chloroplasts. Instead of UDP-Gal, UDP-Glc is utilized by cyanobacteria to synthesize the intermediate precursor β-GlcDAG. Then the glucose head in GlcDAG is modified to galactose by an epimerase to give rise to GalDAG. One protein denoted as sll1377 from *Synechocystis* sp. PCC6803 was reported to synthesize the intermediate GlcDAG (150).

Two enzymes have been identified to catalyze the synthesis of GalGalDAG in *A. thaliana*, namely DGD1 and DGD2. Similar to the aforementioned GalDAG synthases, they also use UDP-Gal as the soluble substrate to form GalGalDAG by transferring the galactose from UDP-Gal to the head group of GalDAG. DGD1 and DGD2 are localized in the outer envelope membrane of chloroplasts (53, 144). In comparison to the primary sequence of DGD2, mature DGD1 possesses a large N terminal extension region that is required for chloroplast outer envelope insertion, or intermembrane contact. Under normal growth conditions, DGD1 is the major enzyme contributing to the biosynthesis of GalGalDAG. It was shown in the *A. thaliana dgd1-1* mutant, that the amount of GalGalDAG was reduced to ~10% of wild-type (151). This mutant also displayed a severe dwarf growth and defects in photosynthesis. DGD2 is not the key player in synthesizing GalGalDAG under normal conditions, but was found to play a crucial role in synthesizing extra GalGalDAG to surrogate the reduced phospholipids under phosphate shortage conditions (36, 37). In addition, an UDP-Gal-independent enzyme GGGT was also found to synthesize GalGalDAG in a double dgd1-1, dgd2-1 null mutant (37). GGGT can utilize GalDAG as the donor of galactose instead of UDP-Gal to form GalGalDAG by condensing two GalDAG molecules. The resulting GalGalDAG derived from the GGGT pathway, can be further glycosylated by GGGT to give rise to GalGalGalDAG. Note that all galactolipids produced by the GGGT enzyme differ in glycosidic bond configuration from those synthesized by DGD1 or DGD2 (54, 152).
Glycolipid remodeling upon stress

Under abiotic or biotic stress conditions, plants adjust membrane composition in order to save phosphate and maintain proper cellular functions. In this section, the case of phosphate deprivation stress will be selected for illustrating the stress influences on plants in terms of membrane lipid profiles. Phosphate is an essential nutrient for plants because it constitutes one of the key building units for numerous cellular molecules including many metabolites, nucleic acids, and phospholipids. It was also estimated that around 30% of global cropland areas suffer from phosphorus deficiency (153). Hence, this agronomic phosphorus imbalance across the globe can significantly affect crop growths and reduce crop yields.

Figure 9. Changes of chloroplast lipid composition during phosphate deprivation (18). GalDAG (MGDG) and GalGalDAG (DGDG) usually represent about 60 mol % of total lipids in green leaves of A. thaliana. Phospholipids including PC, PE, and PG contribute to the remaining 40 mol % of the membrane lipids. Under phosphate shortage conditions in a Pho1 mutant, the expression level of glycolipids was up-regulated while phospholipids amounts decreased. In particular, GalGalDAG increased from 14% up to 24%, but the other major glycolipid GalDAG remained constant.

The A. thaliana mutant Pho1 was unable to transport phosphate from root to shoot (154), therefore it provided a good model system for analyzing how plants respond to phosphate shortage in terms of membrane lipid profiles. As shown in Fig. 9 (18), GalDAG and GalGalDAG are the most abundant lipid species and usually represent about 60 mole % of total lipids in green leaves of A. thaliana. The remaining about 40
mole % lipids are contributed by the other group of polar lipids - phospholipids including PC, PE, and PG. In comparison to wild type, the expression levels of glycolipids in the Pho1 mutant were up-regulated while phospholipid amounts decreased. In particular, SQDG and to a lesser extent GalGalDAG increased dramatically but the other major glycolipid GalDAG remained constant. Thus, synthesis of the non-phosphorous lipids GalGalDAG and SQDG was increased to provide surrogate lipids for the reduced level of phospholipids, as they also belong to the bilayer-forming lipid classes as do most of the major phospholipids.

Interestingly, the replacement of phospholipids with GalGalDAG is not a specific in situ “one to one” mode. Normally, GalGalDAG is restricted to plastid membranes and barely detected in extra-plastidial membranes. However, up to 70% of phospholipids in the plasma membrane of oats can be replaced with GalGalDAG under phosphate starvation (138, 156-158), which suggests the occurrence of inter-membrane replacement of lipids between organelles. Moreover, GalGalDAG was found to be exclusively accumulated in the cytosolic leaflet of the oat root plasma membrane, surrogating the partially degraded phospholipids, while the apoplastic leaflet was occupied by acylated sterol glycosides, which were suggested to maintain plasma membrane integrity by increasing lipid acyl chain ordering (158).

Gene expression analyses in A. thaliana reveals that, except for DGD1, all genes encoding the glycolipid-synthesizing enzymes were up-regulated, including, MGD2, MGD3, DGD1, DGD2, SQD1 and SQD2 (36, 159, 160). Extensive efforts have been made in the past decades to determine the extent to which each of them contributes in lipid remodeling. The mRNA level of SQD1 gene, which encodes one enzyme involved in sulfolipids synthesis, as well as its protein expression level are both up-regulated. This probably leads to the increased amount of SQDG under phosphate-limiting conditions. GalDAG, the precursor of GalGalDAG, remains constant in response to phosphate shortage, and it does not accumulate in extra-plastidial membranes. GalGalDAG is usually restricted to plastid membranes, but was found to accumulate in extra-plastidic membranes including the plasma membrane, the tonoplast membrane, and the mitochondrial membrane under phosphate deficiency (161). But the mechanism for the transportation of GalGalDAG from plastids to extra-plastid membranes remains unknown. It was found that the transcription of the DGD1 and DGD2 genes is also induced during phosphate deprivation (36). Analyzing the fatty acyl chain profiles of accumulating GalGalDAG species revealed (146, 151), that DGD1 is mainly contributing to the GalGalDAG accumulated in chloroplast membrane, while DGD2 is responsible for synthesizing the GalGalDAG accumulated in extra-plastidial membranes. Since there are no chloroplasts in roots, the large amount of GalGalDAG accumulated in oat roots under phosphate limitation is presumably due to DGD2. GalGalDAG synthesis is also triggered in nitrogen-fixing
nodules of soybean and Lotus, where it serves as a bilayer component of the peribacteroid membrane (PBM) (163). Notably, the transcription of MGD2 and MGD3, paralogs of MGD1, was also up-regulated during phosphate shortage (36, 145). A dramatic reduction of GalGalDAG accumulation in roots was observed in mgd3 mutant, indicating that MGD3 is crucial for GalGalDAG biosynthesis in non-photosynthetic tissues under phosphate starvation. Further decrease of GalGalDAG was observed in a mgd2/mgd3 double mutant, which has almost no extra-plastidial accumulation of GalGalDAG (164). Taken together, DGD2 together with MGD2 and MGD3 are able to form a DGD1-independent pathway for synthesizing GalGalDAG in non-photosynthetic tissues.

Fates of phospholipids

The reduction of phosphate by breaking down phospholipids in response to phosphate shortage is presumably linked to its essential roles in cellular functions. The liberated phosphate from breakdown of phospholipids can be used for either incorporation into macromolecules such as DNA and RNA, or participating in signal transduction networks through phosphorylation and/or de-phosphorylation. So far, there are two pathways proposed to mediate the phospholipid breakdown during phosphate deficiency (54, 139). One involves two enzymatic steps, in which a phospholipase D (PLD) catalyzes the first step to give rise to PA, then PA is hydrolyzed by PA phosphatase (PAP) in the second step to release DAG and phosphate (54). The other pathway is catalyzed by phospholipase C (PLC) to remove the headgroup in phospholipids (54, 156). The expression of PLDZ2, one of twelve PLD-encoding genes in A. thaliana, was found to be up-regulated under phosphate shortage (35), and its corresponding mutant caused moderate defects in GalGalDAG accumulation in roots (35). Two PAP enzymes in A. thaliana, namely PAP1 and PAP2, were also proposed to mediate phospholipid degradation under phosphate shortage, because there was a defect in accumulation of GalGalDAG in extra-plastid membranes in a double PAP1 PAP2 knock-out mutant (165). There have so far been six non-specific PLCs (NPC1-NPC6) identified in A. thaliana. Only the transcriptional level of NPC4 and NPC5 can be stimulated in response to phosphate deficiency, suggesting their potential roles in remodeling phospholipids profiles, which were further supported by studies on their knockdown mutants (156).
2.5 Membrane-bound proteins

Membrane-bound proteins are classified into peripheral and integral proteins based on their lipid-protein interactions. Peripheral membrane proteins are loosely associated with the membrane surface mainly by means of ionic forces, electrostatic interactions, and/or hydrogen bond interactions (166, 167). Besides, post-translational lipidation is also another way to anchor proteins to the membrane (168). Peripheral membrane proteins can sometimes be recruited to the membrane surface upon signaling and activation. The interactions are reversible, therefore extraction of peripheral membrane proteins can be done without detergent addition (169). High ionic strength or alkaline buffers can strip the peripheral membrane protein off from the membrane by eliminating their relatively weak associations.

In contrast, integral membrane proteins are tightly integrated into the membrane. They cannot be extracted without usage of lipid mimic molecules such as detergents (169, 170). There are two major structurally distinct integral membrane protein types in the biological membrane: α-helical and β-barrel membrane proteins (171). α-helical membrane proteins span the membrane with one or more helices and are present in the inner membrane of bacteria or the plasma and internal membrane of eukaryotes. These helical protein structures have been found dominant as various membrane receptors and channels (171). β-barrels are found exclusively in the outer membranes of Gram-negative bacteria or the outer membrane of chloroplast and mitochondria. The functions of β-barrels include ion channel, nutrient uptake, and so on (171, 172).

Based on the modes of insertion into the lipid bilayer, integral membrane proteins can be further classified into monotopic, bitopic and polytopic proteins (173). Monotopic proteins are permanently bound to only one side of membrane without transmembrane segments. Bitopic protein contains only one transmembrane segment while polytopic protein transverses the membrane with more than one transmembrane segment.

There have been about 300 crystal structures solved for integral membrane proteins until October 2011, but less than 10% of them are monotopic proteins (according to “Membrane proteins of known 3D structure” at http://blanco.biomol.uci.edu/mpstruc). Both experimental and bioinformatics data have been scarce for the interaction of monotopic proteins with membranes. Notably, monotopic proteins are comprised of a large group of interface enzymes that can utilize both hydrophobic (usually lipids or fatty acids) and hydrophilic substrates (174). Some monotopic proteins are involved in lipid metabolism by synthesizing new lipid species or modifying existing lipids,
which can in turn influence membrane properties. With advances in structural bioinformatics, new insights into the molecular interaction between monotypic proteins and lipid bilayer have been obtained by atomistic and coarse-grained molecular dynamics simulations (175-177). The extent of penetrating into hydrocarbon chain region seems to vary substantially for different monotypic proteins. Some proteins penetrate shallowly into the membrane and are situated at the water-lipid interface regions, while other proteins can insert more deeply into the hydrocarbon core, and some even distort the lipid bilayer due to the strong lipid-protein interactions, resulting changes in membrane thickness and bending (176). Basic residues (Arg, Lys, and His) and hydrophobic side chains (Phe, Leu, Ile, and Val) were found to play key roles for the membrane insertion. In particular, the number of basic residues around the interface region seems to correlate with the depth of insertion, and also contribute to the perturbation of the lipid bilayer (176).

**Glycosyltransferases**

*Glycosyltransferases* (GTs) constitute a large family of enzymes that are implicated in the biosynthesis of complex carbohydrates and glycoconjugates by catalyzing the transfer of mono-saccharide or oligo-saccharide moieties from a donor to an acceptor (178-182). The donors are mainly nucleotide-activated sugars including UDP, GDP, and CMP activated monosaccharides. The acceptors can be a variety of molecules such as oligosaccharides, proteins, peptides, lipids, and inorganic phosphate. There are about 1% of coding regions in most genomes estimated to encode GTs, which reflects the fundamental role of glycosylation in biological systems. So far, nearly 50,000 amino acid sequences have been identified with proven (minority) or putative GT activities, and these are grouped into 93 families in the Carbohydrate Active Enzyme Database (http://www.cazy.org) on the basis of sequence similarity (183). These families of GTs differ in structure and function. GT2 and GT4, accounting for about half of the members, are the two largest families. Moreover, almost all archael GTs are classified into these two families, hence GT2 and GT4 are believed to be the ancestor families of GTs (184).

Though the sequence similarities between GTs are low, there are only two main structural subclasses: *GT-A* and *GT-B* (Fig. 10) (178, 180, 181, 185). The GT-A fold enzymes have two dissimilar domains: the N terminal domain, consisting of a Rossmann-like motif, is responsible for recognizing the nucleotide activated sugar donor; the C terminal domain is mainly composed of mixed β-sheets, and is involved in acceptor binding. As for the GT-B fold enzymes, they contain two similar
Rossmann fold domains. The N-terminal domain provides binding sites for sugar donor, while the C-terminal domain is responsible for recognizing the acceptor. The two domains in both GT-A and GT-B fold enzymes are connected by a linker region, and the catalytic site is localized in the cleft region between the two domains. In addition, a third glycosyltransferase fold (GT-C) has been proposed based on iterative BLAST searches followed by structural comparisons. This GT-C fold is that of a large hydrophobic integral membrane protein having between 8 and 13 transmembrane helices and an active site located in a loop region (186).

**Figure 10. Two main structural subclasses for glycosyltransferases.** The GT-A fold is represented by the nucleotide-diphospho-sugar transferase SpsA (PDB code: c) from *Bacillus subtilis*, and the GT-B fold is shown by the mannosyltransferase PimA (PDB code: 2GEJ) from *Mycobacterium smegmatis*.

The new glycosidic bond formed by GTs can occur with either inversion or retention of configuration at the anomeric center. Thus GTs can be categorized into two main subgroups: reverting GTs and retaining GTs (178, 180, 181). For the reverting GTs, one simple displacement mechanism was proposed, where the reactive hydroxyl group of the acceptor sugar is deprotonated, resulting in a nucleophilic attack on the C1 carbon of the donor. Usually, a general base (asp or glu) is involved in the deprotonation of the reactive hydroxyl of the acceptor. In contrast, a two-step mechanism was proposed for retaining GTs. A donor-enzyme intermediate is formed at the first step, but the second step is still less known. Divalent metal ions such as Mn$^{2+}$ and Mg$^{2+}$ are required for many glycosyltransferases as they can bind to the active sites, resulting in conformational changes of enzymes to assist the mobilization of substrates and release of products, respectively (178, 180, 181, 187).
GT-B monotopic proteins

The occurrence of numerous molecules eliciting amphipathic properties, leads to the existence of a special group of enzymes that can manipulate two strikingly different substrates for releasing an amphipathic molecule. These enzymes can be either integral or water-soluble (174). The membrane is the major source for providing hydrophobic substances while hydrophilic molecules are ubiquitously present in the cytoplasm. Therefore, the monotopic enzymes seem to be the easiest way to bring together these two different substrates (174). The reactive site of a monotopic enzyme is usually located at the water-lipid interface for an easy access to hydrophobic substrates from the membrane and hydrophilic substrates from the cytoplasm. As mentioned above, GT-B fold enzymes possess two Rossmann fold like domains, and this feature indeed could facilitate the reaction involving both hydrophobic and hydrophilic substrates. It has been shown that in some GT-B monotopic enzymes (36, 57, 115, 116, 188-191), one domain contains the binding site for the hydrophobic substrate (usually lipids), while the other domain is responsible for binding the hydrophilic substrate (usually nucleotide diphosphate activated sugar). The catalytic reaction is performed in the cleft region between the two domains, which allows both substrates meeting each other without leaving from their original native environments.

Recently, several crystal structures for GT-B monotopic proteins have been solved, including MurG (GT-28) (192), PimA (GT-4) (193), and GumK (GT-70) (194). This allows us to get more insights into the function and dynamics of enzymes with regard to membranes. A beautiful example is from PimA (188, 189), an essential mannosyltransferase from *M. smegmatis* and *M. tuberculosis*. PimA utilizes GDP-Man as sugar donor to catalyze the first step in the biosynthesis of phosphatidyl-[myo]-inositol mannosides (PIMs) by transferring the mannose moiety to the 2-position of the inositol group of the acceptor substrate phosphatidyl-[myo]-inositol, followed by further modification to give rise to PIMs (195). An amphipathic α-helix from the N-terminal domain of PimA was found to be essential for membrane association, as substitution of four conserved positively charged residues significantly affected its membrane interaction (188). Additionally, this amphipathic α-helix is crucial for activity in vivo, and thus is essential for the growth of mycobacteria. Such amphipathic α-helices represent a common structural feature for monotopic proteins and was also found in other GT-B monotopic proteins such as MurG (115) and MGS (196). As for MGS, an amphipathic helix derived from the N-terminal domain was shown to interact with the lipid bilayer, by a combination of structural, biochemical and biophysical approaches (196). Moreover, both substrates were shown to cause
significant conformational changes of PimA but exerted opposite effects \(\text{(189)}\). The sugar substrate GDP-Man is able to induce an open to closed transition by rigidifying the enzyme through interdomain rearrangements, while binding the lipid acceptor PI caused a more relaxed state of PimA, resulting in the release of products. This open-close-open transition confers the ability of PimA to manipulate two substrates \textit{in situ} to give rise to the final products. It was also noted for several GT-B monotopic enzymes, that the activities can be substantially increased by the addition of non-substrate anionic lipids. This suggests that membrane environments \textit{(i.e.,} membrane thickness, surface charge density, lipid phases, curvature stress, etc.) exert substantial influences on the functions of the enzymes, probably by causing conformational changes or mechanical movements with regard to membrane. Notably, a large fraction of GTs are anchored in the ER and Golgi membrane by one or more transmembrane segments, suggesting the potential interplay between membrane lipids and the GTs during the glycosylation processes.

The general structural features of GT-B monotopic proteins were discussed in Paper I. Basically, the hydrophobic segments, penetrating deeply into the hydrophobic core, are mostly surrounded by positively charged clusters, contributing to electrostatic association at the membrane interface. In addition, the N-terminal domains have commonly higher calculated pI values than the C-terminal domains, reflecting the presence of positively charged surfaces. A model for lipid interaction of a GT-B monotopic protein is sketched in Fig. 11.

![Figure 11. Membrane association of GT-B monotopic proteins.](image)

The N-domain is the most likely lipid-anchored region, while the C-domain may associate weakly with the membrane, which facilitates conformational changes upon substrate binding. The active site is located at the membrane interface and is easily accessible to both hydrophobic and hydrophilic substrates.
2.6 Lipid - protein interactions

Though the lipid compositions of defined sub-cellular compartments do not seem to vary dramatically, the local area of lipid species do vary substantially according to different stages of cellular processes. Additionally, more and more evidence suggest, that there are different separated lipid domains both in eukaryotic and prokaryotic cells (197), which accommodate special proteins and/or protein complexes implicated in certain cellular functions such as membrane traffic and signal transduction. Lipid rafts (198-200) are one of the well-studied domains present in eukaryote plasma membrane, which are enriched in cholesterol and sphingolipids, and to which GPI-conjugated proteins are stably anchored. Lipid rafts provide a specialized microenvironment where different protein molecules interact to initiate a diverse set of signal transduction processes. Cardiolipin-enriched domains were also confirmed in bacterial cells, and they are located predominantly at cell poles and the division septum (201). Thus, the lateral inhomogeneities make the membrane a more complex system than that of the fluid mosaic model, which was considered as a fluid homogeneous entity. Integral membrane proteins, permanently interacting with different lipids, perform their functions properly in the membrane. Even for peripheral membrane proteins, which interact transiently with membranes, the interplay with the membrane through physical interactions is still the prerequisite for exerting their cellular functions. Therefore, it is not surprising, that the structure and function of a membrane protein can be influenced by the surrounding lipids.

Based on the extent of protein interactions, lipids in the membrane can be classified into three major groups: annular lipids, non-annular lipids, and bulk lipids (61, 203). Annular lipids are the lipids situated on the hydrophobic and/or hydrophilic surfaces of membrane proteins, and these annular lipids tend to form a lipid shell on the protein surface due to slower exchange rates (higher affinities), resulting in a protein-lipid complex. Annular lipids in these protein-lipid complexes are in dynamic exchange processes with the neighboring bulk lipids, but at a relatively slow rate due to their direct binding to the proteins. Non-annular lipids are frequently found at the interfaces of multi-subunits of membrane proteins or membrane protein complexes. Non-annular lipids, fit snugly into the cavities or clefts between protein subunits, and are crucial for the assembly, folding, and function of some membrane proteins (203, 204). They are exchanging with neighboring lipids at an even slower rate compared to annular lipids. Bulk lipids are not in direct physical contact with membrane proteins and diffuse randomly in the membrane plane, but they provide the bulky matrix to support the proteins (203).
There are several mechanisms proposed to explain the interactions between lipids and membrane proteins, and these can be classified into two main groups: non-specific and specific interactions (64, 205), which are not mutually exclusive. Non-specific interactions are related to the hydrophobic mismatch between membrane protein transmembrane segments and the lipid bilayer. Membrane proteins reside in the membrane by matching its hydrophobic segments with the hydrophobic regions of lipid bilayer to achieve optimal thermodynamic stability. Differences of thickness between membrane and the protein can shift the matching patterns, thus affect the structure and function of the protein. Fluidity, surface charge density, and elastic stored curvature stress are all properties of lipid bilayers, which can affect the functions of membrane proteins without any specific binding between lipids and protein residues. In contrast, the specific interactions involve specific binding between head-groups and/or fatty acyl chains of lipids and residues in the proteins (205, 206).

**Non-specific interactions**

The non-specific interaction between membrane proteins and lipids is the consequence of hydrophobic matching (206). The hydrophobic core regions of a lipid bilayer fit well with the hydrophobic length of membrane proteins in order to minimize the energetic costs of exposing their hydrophobic region to the polar water phase (61). Adaptability of this hydrophobic matching is important for the structure and function of membrane proteins. This can be illustrated by the Ca$^{2+}$-ATPase from skeletal muscle sarcoplasmic reticulum (208). Its activity was assayed by *in vitro* in artificial liposomes composed of PC with various acyl chain lengths. The highest activity was observed with a fatty acyl chain length of C$_{18}$, while decreased activities were found in both shorter and longer acyl chain length. Therefore, the C$_{18}$ length of the PC lipid provide the optimal hydrophobic matching condition for its activity, and hydrophobic mismatching resulted from changes in the acyl chain length caused activity reduction probably by affecting its conformational states (209, 210).

Similarly, activities of integral membrane proteins can also be affected by other membrane properties, such as membrane surface charge density, fluidity, and stored membrane curvature stress. LacY (lactose permease) from the inner membrane of *E. coli* is a symporter, utilizing the free energy released from downhill translocation of H$^{+}$ to drive the energy dependent uphill accumulation of substrate against a concentration gradient. It has been shown both *in vitro* and *in vivo*, that PE is required for its uphill accumulation of substrate. Interestingly, other neutral and nonbilayer-prone lipids such as PC and GlcDAG could also support its uphill transport function.
But its activity was abolished in the bilayer-prone lipid - GlcGlcDAG. This indicates, that both neutral head-group and nonbilayer-prone properties of lipids are important for the activity of LacY (8, 211, 212).

Besides polytopic proteins, the influence of membrane properties on monotopic and peripheral membrane proteins was also observed. For example, the activities of DGD2 (Paper I), PimA (188), DGS (56, 213), and MGS (56) were significantly stimulated in anionic lipid-containing systems, suggesting that the surface charge property of the membrane contributed by anionic lipids might induce the conformational changes and/or local rearrangements of these monotopic proteins. Membrane stored curvature stress due to the presence of different ratios of bilayer and nonbilayer-prone lipids were also able to modulate the activity. For instance, in comparison to PC, PE stimulated the activity of CTP:phosphocholine cytidylyltransferase, a rate-limiting enzyme involved in PC biosynthesis (214). Similarly, PE was also able to enhance the activity of protein kinase C (215) and DGS (75, 213). In contrast, nonbilayer lipids such as PE, and GlcDAG were found to decrease the activity of DGD2 (Paper I), probably a result of the modulation of protein orientation with respect to the membrane plane.

Lipids can also work as “molecular chaperones” to assist membrane protein insertion, folding, and assembly through non-specific interactions (203). It is known that membrane proteins are integrated laterally into the lipid bilayer subsequently after translocation through the Sec translocon. The orientation of a membrane protein with respect to the membrane plane is thought to mainly be determined by its primary sequence and the translocon insertion machinery (216-218). However, a growing body of evidence shows that the properties of the surrounding bulk lipids can also influence the membrane protein topology by non-specific interaction. The beautiful examples are from LacY, PheP, and GabP which have been extensively studied in the past decade and their membrane topologies varied in different membrane lipid environments (211, 212).

The crystal structure of LacY exhibits two six-transmembrane α-helical domains, separated by a large hydrophilic cavity. However, the misfolding of this protein was observed in PE-lacking E. coli cells, in which the N-terminal six transmembrane helix-bundle of LacY adopted the inversed orientation with respect to the very C-terminal five transmembrane helix bundles, with the seventh transmembrane domain located outside the membrane. Interestingly, the wild type orientation could be restored in PC or GlcDAG in an E. coli strain lacking PE. It was proposed that the neutral lipids PC or GlcDAG can neutralize the highly negative surface charge of PE-lacking plasma membrane that is mainly given by the anionic lipids CL and PG. GalDAG, bearing no net charge and having hydrogen-bonding properties, is more like
the neutral native phospholipid PE. Therefore, they could create an optimal membrane surface charge density to support the wild type orientation of LacY by diluting the negative surface charge density (of PG and CL) as PE does (211, 212).

Moreover, the misfolding of LacY can also be prevented in PE-lacking strains, by either eliminating the negatively charged residues or introducing positively charged residues in the cytoplasmic side of the N-terminal helix bundle (219). This further indicates that the lipid-protein charge interaction determines the final membrane protein topology. Similar findings were also reported in the literature (211, 212).

**Specific interactions**

The specific polar interactions between the head-group region (phosphodiester and/or identity head moiety) of lipids and polar residues from membrane proteins mainly contribute to the stabilization of lipid binding as observed in several crystal structures of membrane proteins (203, 220, 221). These lipids, often referred to as non-annular lipids, are considered essential components for the structure and function of the protein (17, 203, 205, 222). As mentioned before, some lipids, such as PG, SQDG, GalDAG, and/or GalGalDAG were frequently found localized in the grooves or clefts of crystal structures of some protein complex (220, 221, 223), such as the photosynthetic machineries, and the yeast cytochrome bc1 complex. These lipids, being both mediator molecules for assembling subunits and effector molecules for triggering conformational changes, are crucial for both integrity and functions of these protein complexes. Table 2 lists the lipids identified in some membrane protein complexes implicated in oxygenic photosynthesis.

There have been no conserved specific lipid binding sites detected in integral membrane proteins, but some residues were found frequently in association with specific lipids (205, 207). For instance, three residues like KKY, RKY and HRN are binding tightly to the phosphodiester groups of the acidic phospholipids CL (205). It should be stated here that, the binding site is not always linear in its primary sequence, but in most cases is composed of polar residues from different protein subunits, as observed in the yeast cytochrome bc1 complex (220). Notably, aromatic residues are very often involved in lipid binding (224). Tyrosine residues interact with phosphodiester groups by either ion pairing or hydrogen bonding and are preferentially localized at the membrane interface region. Likewise, tryptophan is also frequently found at the membrane interface region, hydrogen bonding with phosphodiester group through its indole nitrogen atom. Additionally, lipid binding is
further stabilized due to the hydrophobic effect between fatty acyl chains and the transmembrane segments.

Table 2. Identification of lipid species in photosynthetic machineries, adapted from (223)

<table>
<thead>
<tr>
<th>Protein complex</th>
<th>Organism</th>
<th>No. of lipids/monomer</th>
<th>PDB code</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS I</td>
<td><em>Thermosynechococcus elongatus</em></td>
<td>3 PG, 1 GalDAG</td>
<td>1jb0 (24)</td>
</tr>
<tr>
<td>PS II</td>
<td><em>Thermosynechococcus elongatus</em></td>
<td>11 GalDAG, 7 GalGalDAG, 5 SQDG, 2 PG</td>
<td>3bz1, 3bz2 (22)</td>
</tr>
<tr>
<td>PS II</td>
<td><em>Thermosynechococcus elongatus</em></td>
<td>6 GalDAG, 4 GalGalDAG, 3 SQDG, 1 PG</td>
<td>2axt (23)</td>
</tr>
<tr>
<td>cyt b6f</td>
<td><em>Mastigocladus laminosus</em></td>
<td>2 DOPC</td>
<td>1vf5 (26)</td>
</tr>
<tr>
<td>cyt b6f</td>
<td><em>Mastigocladus laminosus</em></td>
<td>2 DOPC, 1 SQDG</td>
<td>2e74, 2e75, 2e76 (28)</td>
</tr>
<tr>
<td>cyt b6f</td>
<td><em>Chlamydomonas reinhardii</em></td>
<td>2 GalDAG, 1 SQDG</td>
<td>1q90 (27)</td>
</tr>
<tr>
<td>LHC II</td>
<td><em>Pisum sativum</em></td>
<td>1 PG, 1 GalGalDAG</td>
<td>2bhw (225)</td>
</tr>
<tr>
<td>LHC II</td>
<td><em>Spinacia oleracea</em></td>
<td>1 PG, 1 GalGalDAG</td>
<td>1rwt (29)</td>
</tr>
</tbody>
</table>

However, there are at least 10 identified structural domains from peripheral membrane proteins, exhibiting lipid binding capacities (132). As mentioned earlier, these lipid-binding domains differ substantially in many aspects from each other such as size, secondary structure, target preference, and the cofactor dependence. Some domains show specific binding to specific lipids. For example, C1 domain, a zinc-finger motif of about 50 amino acids found in protein kinase C isoforms and DAG kinases, binds preferentially to DAG (132). Pleckstrin Homology (PH) domain, present in numerous proteins with membrane-association functions, is a β-sandwich motif consisting of 100 amino acids. It contains a stereospecific binding site for different PIs species (132). Likewise, the FYVE (Fab1, YOTB, Vac1, EEA1) domains, a mixture of α/β motif containing about 70 amino acids, bind specifically to PI(3)P
which is present in different stages of endosomes (132). Some domains such as the C1 domain and the annexin domain require the presence of cytosolic calcium ions that confer positive charge at the membrane interface to attract negatively charged headgroup of PS in the membrane (132). In addition, many peripheral membrane proteins were also found to associate with membranes via undefined basic and/or hydrophobic clusters (226, 227).

2.7. Approaches to study lipid-protein interactions

Given the central biological roles of lipid-protein interactions in a number of cellular processes and the presence of numerous proteins exerting lipid-binding capacities, a variety of techniques based on biochemical, biophysical, and/or bioinformatics principles, have been developed for analyzing the protein-lipid interactions. For the virtue of simplicity, details will not be mentioned for all of them here, which is outside the scope of the thesis. The most commonly used biophysical techniques for probing the lipid-protein (peptide) interaction, as deduced from literature, are x-ray crystallography, electron diffraction, NMR spectroscopy, atomic force microscopy (228), infrared spectroscopy (119, 155), mass spectroscopy, electron paramagnetic resonance (229), differential scanning calorimetry, isothermal titration calorimetry (230), flow cytometry (231), fluorescence quenching (232), surface plasma resonance (233), and circular dichroism (Paper II). Most of them have been reviewed (234, 235).

Blotting is also a fast way to analyze the lipid binding features of a given protein (236). The purified target protein is incubated with a nitrocellulose membrane (filter) onto which a spectrum of different lipids are immobilized, followed by a suitable antibody against the protein. Then the binding features can be obtained by analyzing the immunoblotting signal. As was shown by the Lipid Strip™ assay in Paper I, several GFP-fused DGD2 short peptides were analyzed by this approach to detect their potential lipid binding capacities. However, this approach comes with the disadvantage that the lipids immobilized onto the membrane are multilayered aggregates, and different from the natural lipid bilayer environment.

Another wildly used approach for probing the lipid protein interaction is to separate lipid-bound proteins from free (“soluble”) proteins by different ways including ultracentrifugation, ultrafiltration, and dialysis (237). In this case, the separated protein and/or lipid concentrations can be measured to quantify the binding properties, such as binding strength and binding kinetics. Figure 12 depicts the sucrose-loaded vesicle-binding assay used in Paper I, II, and V. The sucrose-loaded liposomes containing the desired lipid species are mixed with proteins of interest, then
ultracentrifugation can be used to separate the lipid-bound proteins (in pellets) from free proteins (in supernatant). In our case, GFP was genetically fused to the peptides as a fluorescence reporter for quantifying the peptide amounts both in pellet (lipid-bound peptides) and in supernatant (free unbound peptides), and a small portion of Bodipy-labeled (“fluorescent”) lipid was also incorporated to the liposome in order to monitor the integrity of liposome during the process of binding assay.

Figure 12. Liposome binding assay based on ultracentrifugation. Purified GFP-fusion proteins are incubated with sucrose-loaded liposomes followed by ultracentrifugation. The lipid bound GFP-fusions are obtained in the pellets while unbound ones are in the supernatant.

Moreover, the bioinformatics approaches have been shown to be powerful on probing the molecular interaction between protein and lipid. Though the tightly bound lipid molecules have been observed in many crystal structures of membrane proteins, the molecular mechanisms underpinning the specific lipid interactions are still poorly understood. Their structures define the functions of membrane proteins, and especially the conformational changes triggered by surrounding lipids can dramatically affect the functions of membrane proteins. Therefore, molecular modeling combined with other computational studies can bypass the experimental difficulties and gain us more insights into these interactions, which correlate the structural features with functions of membrane proteins. Atomistic simulation and “coarse grain” simulation are two widely used approaches for analyzing lipid protein interactions (139, 157, 238-241). The latter can probe more complex membrane protein-lipid systems due to its simplified representations of all atoms (176, 240, 242), while atomistic simulation can
provide more accurate information but also requires more computational resources (240). An integral membrane protein-lipid insertion database, based on coarse grain simulations, has been established (CGDB, http://sbcb.bioch.ox.ac.uk/cgdb/) (243). It contains orientation information of more than 100 unique membrane proteins that are simulated in a simplified DPPC bilayer. From this database, some important general features can be obtained, for example, tyrosine and tryptophan residues are mostly found localized at the membrane interface, while glycine and proline residues tend to prefer the hydrocarbon core region. Recently, a comprehensive database for membrane-annotated protein structures (MAPS, http://bic.boseinst.ernet.in/gautam/maps) (244), based on simple hydrophobic lipid-protein mismatch energy and a flexible lipid/water boundary, has been established. It contains almost all solved membrane protein structures deposited in the PDB database with annotation on the membrane contact regions. Other algorithms based computational programs were also reported for identifying the lipid contact regions of membrane proteins (245, 246).

Notably, almost all molecular modeling requires highly reliable membrane protein structures, which unfortunately are only available for a few membrane proteins. As is well known, the primary sequence of a protein is the major determinant for its final structure, therefore also defines the function, though other factors such as protein insertion machineries and lipid bilayer properties are also involved. Examining primary sequence properties can also deduce membrane contact regions such as transmembrane segments. The “Membrane Protein Explorer” prediction facility (MPEx, http://blanco.biomol.uci.edu/mpex/), based on thermodynamic and biological principles, was designed for this purpose to understand the membrane protein topology by examining their primary sequences (247). It is based on the hydropathy scale of amino acids, which was obtained by measuring the transfer free energy of amino acids between water and membrane. The interface scale within this program can be used to predict the membrane association regions for monotopic proteins by calculating the hydrophobicity. However, more and more evidence suggest that the positively charged residues play crucial roles in anchoring peripheral membrane proteins or monotopic proteins into the lipid bilayer through electrostatic attraction (166). Therefore, the complete physical and chemical properties for amino acids must be considered when predicting the lipid contacting regions of a membrane protein by computational strategies that are dependent on primary sequences. As a comparison to MPEx, a novel computational approach based on multivariate comparison of polypeptide sequence properties was introduced in Paper III for predicting the lipid contacting regions. This new approach is derived from statistical calculations of primary sequence features of GT-B monotopic proteins and was shown to perform better predictions than MPEx in the case of the monotopic protein DGD2.
Multivariate data analyses for protein sequences

The primary sequence of a polypeptide is a combination of 20 amino acids and is the determinant both for its structure and function. Very distant amino acids in the primary sequence may come close enough to constitute motifs and/or domains, which are usually the key workplaces in proteins. Therefore, the traditional way of aligning primary sequences to find the conserved structural and functional features sometimes may not be successful for some proteins. However, multivariate analysis has been shown to be able to provide a powerful way for extracting important sequence features from primary sequence of proteins and then correlate these features with protein properties including locations, activities, and structural folding patterns (248-251). The assumption for applying multivariate analysis on protein sequences is that the primary sequence of a protein contains the determinant information for the properties of interest.

Amino acids that contain a number of physical and chemical features can be described in a number of measured and computed parameters such as hydration potential, isoelectric point, molecular mass, etc. However, not all the features are related to the question of interest, and to reduce the extreme complexity of description for each amino acid, so-called z scales were used (252). These z scales are derived from a multi-property matrix comprising of 29 physical-chemical experimental parameters for the 20 coded amino acids with principal component analysis (PCA) (252). For simplicity, three principle z scales were used in our studies, and they can be tentatively interpreted as z1 for “hydrophobicity”, z2 for “bulk of side-chain”, and z3 for “electronic properties”. As shown in Fig.1 in Paper II, each amino acid in a polypeptide was described by three z-scales and then the periodic physical properties in the polypeptides were calculated by autocross-covariance (ACC), which combines auto covariances between the same z-scale in each position, and cross covariances between two different z-scales in each position (253, 254). The ACC terms were calculated for each polypeptide and a new uniform data matrix X was created, which is detailed in Paper II.

Partial least-squares projections to latent structures discriminant analysis (PLS-DA) is a frequently used classification approach (255), which was developed to explore the relationships between a matrix X and a matrix Y (Paper II). Usually, matrix Y was composed of dummy variables, hence a value of 0 was given to one class and 1 for the other class. In Paper II, PLS-DA is based on the assumption that membrane-associated GT-B protein sequences have common features that are different from that of soluble GT-B proteins, and these features can be identified by PLS-DA for predicting the lipid contacting regions by systematic examination of the
sequences of interest. Thus, PLS-DA was first applied between these two matrices to build a model that could separate these sequence membership based on their periodic physical properties-ACC terms. Then the discriminant ACC terms, defining the separation of membrane associated GT-B from soluble GT-B protein, can be obtained for predicting the lipid contact regions for a given sequence. The detailed procedures concerning the model establishment, model evaluation and model applications are all described in Paper II.
3. Summary of Papers

The main aim of my PhD project is to understand how interface enzymes are able to sense and respond to a changing lipid environment, and to find the important structural features governing this adaptability. To this end, Paper I and Paper II are focused on a plant lipid-synthesizing enzyme. The potential important lipid anchoring segments were analyzed by a set of biochemical and biophysical approaches. A regulatory mechanism was proposed to explain how DGD2 mediates the stressed membrane properties by synthesizing more GalGalDAG to surrogate the deprived phospholipids, which were used for other critical cellular processes. This modulatory mechanism may also be important for other similar monotopic enzymes. Besides, a multivariate-based bioinformatics tool was established to predict the interface binding segments for GT-B monotopic enzymes. In contrast, a peripheral soluble protein Myr1 in Paper III, implicated in vesicular traffic, was found to associate with lipids, especially PIs. Additionally, it also exerted different binding responses to stressed membranes, containing different lipids. Hence, Myr1 was proposed to be a membrane stress sensor by exposing its strongly positively charged clusters to the lipid species present in the intracellular vesicles. In Paper IV, we found a bacterial monotopic enzyme MGS was able to generate large amounts of intracellular vesicles, which are most likely due to its own structural features. In Paper V, analysis of the structural features controlling its lipid anchoring was conducted to MGS, in order to understand how it induces the high amounts of vesicles in the cell. It turns out, that the scaffold-like shape of its structure is most likely one of the key elements crucial for the vesiculation.

Paper I - Tryptophan residues promote membrane association for a plant lipid glycosyltransferase involved in phosphate stress

In this Paper, we are aiming to understand the structural features of a lipid-synthesizing GT enzyme that governs the properties of photosynthetic membranes constituted mainly by two intrinsically distinct glycolipids - the nonbilayer-prone GalDAG and the bilayer-prone GalGalDAG. As is well known, the relative amount between these two glycolipids affect the membrane curvature as well as the lateral stress profiles of the membrane, and hence most likely influence the functions of membrane embedded proteins or protein complexes. In contrast, the proteins that are anchored to the membrane may also adjust their activities by sensing the given lipid environments, in order to keep the physiological balance between the different lipid species present in the membrane. atDGD2, a key enzyme contributing to the synthesis
of GalGalDAG in extraplastidial membranes under phosphate deficiency, was found in this study to be able to "sense" the surrounding lipid environments, and was also able to adjust its activity with regard to the given lipid species. The critical residues for activity were identified and the lipid anchoring regions were also mapped.

It has been shown, that DGD2, together with type B GalDAG synthases, is involved in DGD1-independent biosynthetic pathway for GalGalDAG synthesis under phosphate shortage. DGD2 is not present in all plants and algae in which, however, DGD1 is the absolutely required enzyme for synthesizing GalDAG. DGD1 possesses an N-terminal extension region that is absent in DGD2. Except for this difference, the remaining C-DGD1 region is very similar to (the full) DGD2 and they share very similar traits in the GT-B core region. They possess same structural feature - double Rossmann folds, and also have similar hydropathy profiles and interface-binding segments. Moreover, they contain similar numbers and pairs of positively charged residues, which are thought to be important for the membrane association.

So far, there has been no structure solved for plant GalGalDAG synthases, therefore fold recognition and modeling were used to gain more insights into potential regulatory features of them. In comparison, DGS, a bacterial analog of DGD2 from *A. laidlawii*, was also included in this study, since it was shown to be involved in mediating properties of the plasma membrane by adjusting its activity. All model structures of C-DGD1, DGD2, and DGS share very similar features: Two domains, each of which contains a Rossmann fold structural motif, are separated by a catalytic cleft region, and the N-domain has higher number of positively charged residues than the C-domain, most likely leading to its lipid anchoring.

To challenge the quality of model structures, docking of soluble substrates UDP-Gal and UDP-Glc into the proposed catalytic sites were performed for DGD2 and DGS, respectively. A selected set of amino acids that appeared close in space to the substrates were substituted as alanine, and activity assays were conducted in a detergent mixed-micelle system. Almost all of the tested variants were found to have lower activity, indicating the successful attempts to establish model structures.

Except for its upregulated transcription level during different stress conditions, DGD2 was believed to be able to adjust its activity to control membrane features by controlling the biosynthesis of its product - GalGalDAG. Activity assays performed in bilayer-like micelles showed that three different nonbilayer-prone lipids GlcDAG, DOG and DOPE all decreased activity of DGD2, implying that increasing curvature stress by these additives has a negative effect on GalGalDAG synthesis. Moreover, several potential "activator" lipids, varying in charge and spontaneous curvature stress, were tested in detergent-mixed micelles. It turns out, that the anionic lipids PG, PA and PS have strong stimulatory effects on DGD2 activity, but DGD2 was not
responsive towards CL even though it is structurally related to PG. This may suggest the presence of PG-specific binding site(s) in DGD2. Hence, the species and relative amounts of anionic lipids present in the membranes may indeed play a potential role as selective “controllers” of DGD2 activities in situ, affecting GalGalDAG synthesis.

The kinetic behavior of DGD2 was also investigated to gain more insights into its interaction with the membrane (i.e., how the enzyme may respond to or even be dependent on its lipid environment). It was shown in Figure 3, that for the soluble substrate UDP-Gal, increasing PG concentrations correlated with higher \( V_{\text{max}} \) values and also higher affinities, while for the lipid substrate GalDAG, increasing PG concentrations also stimulated the synthesis rate, but the affinities remain unchanged. This indicates local changes in anionic lipid concentration during stress conditions may have a substantial effect on DGD2 activity.

According to MPEx predictions, C-DGD1, DGD2 and DGS enzymes all lack transmembrane segments but possess a number of segments that are probably anchored to the membrane, and therefore they are classified as monotopic membrane proteins. As for DGD2, five out of nine potential interface binding segments derived from MPEx analyses were selected and mapped onto the model structure of DGD2. Interestingly, all these five segments are localized on the protein surface, and all of them contain at least one Trp residue. It has been also known that, tryptophan is frequently found situated in the re-entrant regions of membrane proteins due to its special physical and chemical properties. Additionally, all the Trp residues are exposed towards the surface. All these features indicate the potential important functional roles of Trp residues. Therefore, the first attempt was to analyze the activity influence on these Trp residues by substituting them to either Phe or Ala. As expected, all variants were affected in activity, and some even lost activity totally. There is a roughly linear correlation between the free energy of transfer \( \Delta G \) value of the segments containing a Trp residue and the measured activities of the respective Trp variants as shown in Figure 4. In DGS, the bacterial homolog of DGD2, substitution of Trp residues to either Phe or Ala were conducted on three predicted sites, and activities were also tested against the respective short segments possessing the Trp residues. Similar results were also observed in the case of DGS. This strongly highlights the functional importance of Trp residues in this kind of monotopic enzymes.

Could these Trp residues also be involved in membrane association? Or could these Trp residues work as “sensor sites” or contributing to a sensor region for monitoring the changing lipid environments? To answer these questions, the potential lipid binding properties of the five selected DGD2 segments were tested both in vivo and in vitro. These short segments, or the respective variants in which Trp was changed to
Ala, were genetically fused to a fluorescence reporter GFP, and were then expressed in *E. coli* cell to check their subcellular localizations in order to evaluate the lipid binding features *in vivo*. Moreover, the liposome binding assay plus lipid strip binding assay were also performed in order to verify the lipid binding properties quantitatively and/or qualitatively. All these binding studies indicate that the three segments—Segment III, IV and V are potentially membrane anchors for DGD2.

In conclusion, the increased synthesis of GalGalDAG under phosphate starvation or other stress conditions may be regulated by DGD2 at the *protein activity level*. The anionic lipids present in the chloroplast membrane may exert a modulator role by affecting the activity of DGD2. Three Trp-containing segments of DGD2, localized both in the N- and C-domains may interact with stimulatory anionic phospholipids, which determine the orientation of DGD2 at the membrane interface. Hence, a regulatory model of DGD2 anchoring at the membrane interface was proposed in this study as shown in Fig. 13. In this model, segments IV and V are permanent anchor points for the N domain in the interface. Segment V of C domain interacts with the membrane interface in a more flexible way—either in an up-down mode or in a rolling mode. The positive charges situated at the lipid contact surface may play a determining role in regulating membrane interaction of the whole enzyme, leading to the different efficiencies in product formation.

![Figure 13. The proposed model for interface regulation of atDGD2 activity (Paper I)](image-url)
Paper II – Lipid-interacting regions in phosphate stress glycosyltransferase 
atDGD2 from *A. thaliana*

In Paper II, those five short segments analyzed by biochemical approaches in Paper I, were further analyzed by a combination of biophysical methods in order to gain more and new insights into the interaction of DGD2 with lipid bilayers. Fluorescence and NMR spectroscopy were used to study the extent of membrane interaction of the selected short segments. We also performed a structural analysis by CD spectroscopy, followed by a NMR structure determination of one of the segments. Moreover, a novel multivariate-based program was established for predicting the interface binding segments for monotopic GT-B proteins.

In comparison with other similar membrane interacting GTs of similar structures, there are no conserved sites, either in amino acid sequences or localization in the proteins. This indicates that the binding sites are specialized for a given membrane interacting GT protein. Therefore, we developed a multivariate-based method to predict binding segments. We first retrieved the sequence features from a training dataset consisting of two groups of GT-B enzymes, which are referred to as either membrane-associated GT-B proteins, or soluble GT-B proteins. Then these features were used to construct a decent multivariate model in which these two groups of enzymes are separated. For a given protein sequence, it will be first processed into short segments, followed by prediction against the established model to gain a ranking score from the model, and this score indicates its propensity to be membrane-associated. Finally, the scores for all the segments derived from the given protein were plotted against its primary sequence in order to evaluate their propensity to be membrane-associated. DGD2 was tested in this study by this approach, and three segments were predicted to be strongly membrane-associated, which overlapped with MPEx predictions. But the other two segments (segment I and II) predicted by MPEx were not likely to be membrane-associated according to our multivariate-based approach. This prediction is in perfect agreement with *in vitro* liposome binding studies both in Paper I and Paper II. Hence, segment IV, V and VI are the lipid anchor sites while segment I and II are probably implicated in the accommodation of hydrophobic acceptor substrate GalDAG.

Secondary structural features of four short segments were analyzed by CD spectroscopy in the presence of different lipid environments. As shown in Figure 5 (Paper II), no structure induction, and only limited lipid interaction were observed for segment I and segment II, which is consistent with multivariate-based prediction as mentioned above. Segment VI is able to form some structure when the anionic lipid concentration in the vesicles was raised up to 30%. In contrast, segment VI was observed to show secondary structures even in the presence of lower amounts of
anionic lipids. However, zwitterionic lipids were not able to induce any obvious secondary structure changes of all short segments tested, implying that charge property of anionic lipids is a prerequisite for anchoring the whole enzyme to the lipid bilayer. Diffusion NMR and fluorescence quenching experiments further support the findings that segment IV and VI interact strongly with anionic lipids due to the electrostatic interactions. Moreover, the structure of segment VI in DPC micelles was successfully solved as an amphipathic helix by a solution NMR technique. A similar amphipathic helix was found in another monotopic lipid GT from mycobacteria - PimA, and this helix, located close to the reaction site of PimA seems to influence enzyme activity in a lipid environment, as suggested from substitution studies. Hence, the conformational change in this segment of DGD2 is more likely to be important for both its activity and lipid anchoring.

In conclusion, we solved the structure of segment VI to be an amphipathic helix in the presence of lipids, and this segment has strong interactions with anionic lipids. The interaction of this segment with lipids is probably mediated by the presence of positively charged residues and Trp residues. The other three segments were predicted by a multivariate-based approach to have different lipid binding affinities, which were then evaluated by a combination of biochemical and biophysical approaches. The work in Paper II gives further strong support for the regulatory model of DGD2 under phosphate stress conditions proposed in Paper I.

Paper III - Basic clusters and amphipathic helices contribute to interactions of Myr1/Syh1 with membrane phospholipids

In this study, we focused on a soluble protein Myr1 from yeast, which was previously shown to be able to rescue temperature sensitive growth of ypt6Δ and ric1Δ null strains, implying its association with the mRNA decay machinery under physiologically relevant conditions. In the present work, we further demonstrate that Myr1 is also able to rescue the vacuole fragmentation phenotype of the ypt6Δ and ric1Δ mutant strains. Though it was originally classified as a soluble protein (without transmembrane segments), it is still believed that, the molecular mechanism behind this interaction is most likely related to its membrane association capacity. Therefore, we characterized the structural regions in Myr1 that are relevant to the lipid binding, and also analyzed its responses to the stressed membrane environments.

We observed partial restoration of vacuolar morphology defects in ypt6Δ and ric1Δ mutants in a Myr1 over-expressed cell strain, implying a functional interaction of Myr1 with RIC1 and YPT6. But this interaction seems not to be direct, since Myr1 is not localized in late Golgi as demonstrated by in vivo fluorescence tracking. It was
also found that the mRNA decay process was interconnected with membrane stress, since both events can be simultaneously induced by the membrane surface-localizing drug chlorpromazine. Therefore, we hypothesize, that Myr1 is able to sense the membrane stress by directly transiently interacting with intracellular traffic. To this end, we further focused on elucidating the lipid anchoring regions of Myr1 and its ability to sense the stressed lipid environments.

Myr1 is predicted to have three conserved domains: a GYF domain, a coiled-coil domain and a C-terminal domain. The GYF domain is a well-known protein motif implicated in protein-protein interactions, therefore only the coiled-coil domain and the C-terminal domain were selected for downstream studies. It was shown by a liposome binding assay that two domains are both able to bind to the yeast lipids, but with relative low binding affinities ($K_m$ within the range of $\mu$M), which is consistent with its role as membrane sensor - transiently interacting with vesicles. In addition, the other two domains denoted as C and D, possessing a coiled-coil domain and C-terminal domains, respectively, were found to show different binding profiles for the stressed membrane, which is composed of total lipid extracts from stressed cells. This further supports its capacity to sense membrane stress (i.e. lipid changes). To narrow down the binding regions of Myr1, three short segments enriched in positively charged residues (but few hydrophobic ones) were selected and genetically fused with GFP for lipid binding assessment. It turns out that all these three peptides are able to bind anionic lipids PG, PS and PI, and that the interaction is probably due to the presence of highly positively charged residues in the peptides, leading to electrostatic attraction with the negative surface of anionic lipid bilayers. The binding to PI lipids led us to test more phosphorylated PI species against these three peptides by the lipid strip assay, since PIs are extremely important in eukaryotic intracellular membrane traffic events. Only one peptide was found to bind preferentially to mono-phosphorylated PI(3)P, PI(4)P and PI(5)P. Moreover, the other two segments C and D all exert binding affinity to mono-phosphorylated PI lipid, indicating that binding of Myr1 to the lipids is probably mediated by various regions and that cooperative binding is the most likely way to regulate its transient lipid interaction.

In conclusion, the soluble Myr1 can sense “membrane stress” by its lipid binding regions, which are preferentially binding to PI species. The electrostatic interactions may determine its lipid binding profiles due to the presence of strongly positively charged residues in certain Myr1 regions. It was also suggested that the various lipid binding regions of Myr1 may use its lipid binding properties in a synergistic way in order to bind specifically to the lipid species with increased affinities.
Paper IV - Massive formation of intracellular membrane vesicles in *Escherichia coli* by a monotopic membrane-bound lipid glycosyltransferase

In Paper IV, the optimization of the overexpression level of one bacterial monotopic lipid-synthesizing GT-MGS led us to discover that the protein can generate massive formation of intracellular vesicles in *E. coli* cells, to an extent rarely seen for other membrane proteins. Lipid regions stained by FM4-64 were observed in the intact cell cytoplasm by fluorescence microscopy, and cryo-transmission electron microscopy (cryo-EM) revealed that the cell was filled with variously sized vesicles. Most of the isolated vesicles (sucrose density centrifugation) had spherical shapes between 50 and 100 nm in diameter. Site-directed mutation studies suggested that the capacity of vesiculation by MGS seems to be inherent to the protein itself. Based on these observations and other analogous reports in literature, several mechanisms were proposed in the Paper to explain the *vesiculation mechanisms*.

MGS is a lipid-synthesizing monotopic GT-B family protein, and expression in *E. coli* can reached up to 330 mg per liter of growth medium after purification procedures, which is substantially higher than those reported for other membrane proteins in the literature. The observed banded appearance through the elongated cell (Fig.1 in Paper IV) suggests the presence of lipid clusters in the cytoplasm, which are rarely seen in bacteria. To further confirm these lipid clusters, sucrose gradient centrifugation was applied to separate these lipid clusters, since they are bound to be of high lipid/protein ratio. They migrate to the top sucrose fraction (Figure 2). As expected, light lipid clusters were observed around the 30% sucrose fraction. This strongly suggests that these lipid clusters are most likely the newly formed vesicles. In order to confirm the morphology of these vesicles, cryo-EM was used to analyze the intact cell. As seen from Figure 3, the outer cell membrane was continuous, whereas the vesicles appeared to be pinched off from the inner membrane. Isolated vesicles by sucrose gradient centrifugation were also analyzed by cryo-EM. In comparison with vesicles formed by outer membranes (256), the MGS-induced vesicles are slightly smaller with diameter mainly between 50 and 100 nm.

SDS-PAGE profiles for the vesicles revealed similar patterns to those of outer membranes or inner membranes (Figure 5). MGS is the predominant protein fraction in the vesicles (>90%) and other protein bands were also analyzed by MS. These proteins are basically from all the locations of the cell and this strongly indicates the pinching-off by MGS is a rather random process. We also analyzed the lipid compositions of the vesicles (Figure 6) and they are basically the same as those of inner membrane, outer membrane or even the whole cell membrane.

The presence of large fractions of non-bilayer prone lipid GlcDAG in the cell membrane led us to hypothesize, that GlcDAG may be implicated in the vesiculation
process by changing the membrane stored curvature stress or other lipid bilayer features. Site-directed mutagenesis was performed in a conserved catalytic region called E(D)X7E motif, which is present in all GT-4 family enzymes. Interestingly, these mutants were still able to give rise to massive vesicles, though the synthesis of GlcDAG is inhibited. This strongly indicates the contribution of protein structural features rather than the presence of nonbilayer-prone lipid GlcDAG, is more important in the induction of the vesicles.

Several mechanisms were proposed in this Paper to explain how these vesicles can be generated by MGS. (i) The packing shape of the protein is important, and especially the overall structure is potentially a physical force driving membrane bending, which was exemplified in the context of “wedge-shaped” proteins. (ii) The shallow penetration of the segments into one leaflet of lipid bilayer may cause the lateral expansion and also induce membrane curvature, resulting in membrane bending. (iii) The positively charged residues at the binding interface may determine the penetration pattern of the protein. The distribution and combination of positively charged residues in the binding region may affect the binding and the bending. (iv) The stimulation of lipid synthesis is another indirect way to aid the vesiculation. The interaction between positively charged residues and anionic lipids can result in the neutralization of surface charge density. Because the synthesis of non-bilayer prone PE in E. coli is mainly governed by the surface charge density given by PG and CL, the high production of PG, which is required to maintain certain surface charge density, can stimulate the synthesis of more PE. Therefore, the requirement of more lipids for vesiculation can be completed by the coupled synthesis of two major phospholipids PG and PE.

In summary, MGS is able to induce massive amounts of vesicles in the cytoplasm of E. coli, and the vesiculation capacity of MGS is most likely due to its own structural features. The shape and insertion profiles of it seem to be important for membrane bending. Additionally, the potential neutralization of surface charge density may also stimulate the coupled synthesis of phospholipids, leading to more lipids required for vesiculation.

**Paper V - Modulation of Escherichia coli cell membrane by a monotopic lipid glycosyltransferase - an exploration of potential mechanisms**

In Paper V, we investigated how the structural features of MGS correlate with its vesiculation capacity. The binding of MGS is a prerequisite for bending the membrane, and therefore we identified the lipid binding regions of it by a combination of limited proteolysis and Mass Spectrometry. Systematic screening of potential binding
segments was performed to narrow down the binding sites, and this was done in a similar way as for DGD2 in Paper I and Paper II. Moreover, C-terminal truncated variants of MGS, compromising the rigid scaffold-like structure, were unable to induce vesicles, implying the importance of its overall structure on generating vesicles.

Segment fragments from both N and C domains were identified by limited proteolysis, followed by MS. It was noted that, the lipid-associated fragments remaining after trypsin digestion were still large. This may be due to the rigid folding pattern of each domain, which is resistant to the digestion, and/or the absence of solvent exposed lysine and arginine residues (the targets for proteolysis). Based on the model structure, a set of short segments were selected from both domains, and all of them are most likely situated in the lipid contact regions. The lipid binding properties of these GFP-fused segments were investigated by the liposome binding assay that was also used in Paper I and Paper II. Except for one segment, which was previously shown to be lipid-binding but was not successfully over-expressed in this study, other segments displayed different binding patterns. One segment from the C-domain binds preferentially to anionic lipids, while the other segment within the C-domain did not show any binding to the tested lipids. Two segments from the N domain were all able to bind to lipids, but with different affinities. However, the binding affinities for all segments tested are lower than observed for those of DGD2 in Paper I and Paper II. Given the strong deformation capacity of MGS on membranes in vivo, the potential cooperation between these weak-binding segments may strengthen the overall binding of MGS. This was further corroborated by subcellular location studies in living cells by fluorescence microscopy.

In addition to the identification of lipid anchoring regions of MGS, the overall scaffold-like structure of it was also challenged by truncating different lengths of residues from the C-terminal tail. The purpose of doing this is to assess the contribution of overall structure feature of MGS to the vesiculation. Three C-terminal truncated MGS variants were constructed, and the vesiculation capacity as investigated by flow cytometry. As shown in Figure 4, except for the shortest variant, which exerts similar forward- and side-scattering profiles with wild type MGS, the other two forms were unable to induce the vesicles. The linking region between the two domains was totally impaired in the latter two variants, and this leads to a more flexible MGS, rather than its native rigid scaffold-like features. The activities of three variants were also measured in a detergent mixed-micelles system that was also used in Paper I. It turns out that the activity of three variants was strongly affected as compared to that of wild type MGS, and the activity for the other two variants, was even totally abolished. This indicates that truncating the C-terminal residues, most of which contributes to the linking region, impairs the conformation states of MGS, therefore affects the catalytic region.
In conclusion, the two domains of MGS both mediate its lipid anchoring, but with different binding strengths. The linking region (helix) between two domains rigidifies the overall structure of MGS, and the scaffold-like structure is important for both its catalytic function and its vesiculation properties.
4. Perspectives

Monotopic proteins constitute a special group of membrane proteins in that they are involved in a variety of biochemical events taking place at the membrane interface. In particular, the monotopic enzymes are able to offer a reaction center at the membrane interface for bringing together two biochemically distinct molecules, one from the polar cytoplasm and one from the hydrophobic membrane. Then the resulting hydrophobic or amphiphilic product is released into the membrane. Thus, the intuitive feature of anchoring at the membrane interface enables monotopic enzyme to also sense the constantly changing lipid environment and then to respond by regulating the biosynthesis pathway of lipids in order to maintain a certain membrane homeostasis.

In Paper I and Paper II, the lipid binding regions were mapped out for one of these interface enzymes (DGD2), and also the regulatory binding models were proposed to explain how the physical movements or the conformational changes, resulting from interaction with its neighboring lipids, affect its catalytic functions. So far, it is still largely unknown how exactly an organism is able to maintain proper cellular functions by regulating these lipid-synthesizing interface enzymes to control the lipid biosynthesis and transport. However, a macroscopic view of the cell and/or organism by “omics” studies would provide invaluable information about responses under stress conditions. Additionally, structural biology combined with bioinformatics research would gain us more and new insights into the dynamic regulation exerted by monotopic proteins and lipids at the molecular level.

Likewise, structural information is also required for another interface enzyme described in Paper IV and V (MGS), in order to understand how it re-shapes a membrane into vesicles. Then, potential engineering can be done on this enzyme to control its vesiculation efficiency, which can be used to aid overexpression of other membrane proteins by providing more lipid bilayer in the cell.

In Paper III, the potential membrane stress sensor protein Myr1 was found to bind specifically to PI derivatives, suggesting an important role in the intracellular traffic events in yeast cell. This interaction is most likely mediated by the short basic clusters distant in primary sequence. Hence, elucidating the structural features of those basic clusters can help us understanding more about its participation in the stressed membrane traffic events, which can also be related to its other potential roles in RNA degradation.

Moreover, an important area of future investigation in clarifying the interactions between monotopic proteins and membrane lipids is to characterize the tightly bound
lipid species. These lipids, including annular and non-annular lipid, do exert structural and functional influences on the monotopic proteins. In particular, the anionic lipids have frequently been found to be key regulators for both activity and conformation of monotopic proteins. In addition, the elucidation of the characteristic residues and/or regions on the surface of monotopic proteins showing specific lipid binding affinity, can also gain us a better understanding the importance of these monotopic lipid-synthesizing enzymes in regulating the native biological membrane properties.
5. Acknowledgement

Over the past 4 years, I have been very grateful to everybody who has been involved in making my PhD period such a wonderful experience. Therefore, I would like to take the opportunity to acknowledge the following persons who have made the completion of this thesis possible.

Åke, thanks for accepting me into your group. You are the best supervisor I have ever met. Your guidance is so helpful both in science and in daily life. Your patience and trust from the first moment encouraged me to overcome most of the scientific challenges I encountered during my PhD study. Your “crazy” ideas and scientific “smelling” are always inspiring us to explore further. And also your enthusiasm for science is always a strong motivator to us. Your updated reference alerts (about 400 papers in total concerning about my project) are always something I had been expecting every day in the past 4 years.

Jean-marie, my co-supervisor, thanks for your dedication to my project, and has always been encouraging me to go further. Vincent, thanks for technical support and scientific discussions.

Stefan, for being so nice in the Big Exam, and I did learn a lot from the discussions, for spending much time on an earlier draft of this thesis with very useful suggestions, and also for taking care of all PhD students in DBB.

I am especially grateful to previous and current group members who have made a direct contribution to the ideas and experimental procedures of my work: Amélie, for introducing me into the plant lipid-synthesizing enzymes, and also for the great collaboration. Cannot remember how many emails we had been writing to each other during past years. Alex, for guiding me in the lab from very beginning, and also teaching me the statistics, which turned out to be very helpful. Your daily “funny” stories were always spreading laugh in the office. Hanna, for the very helpful discussions in the group meetings, the inspiring advices on my project and the great collaboration in the lab. Malin and Maria, for your kindness and the time we spent together. Veronica, for your technical help in the lab work. Eleonora and Enzo, for your time spent in my project and your invaluable input to the thesis. Ljubica, for having nice daily talks in and out of science.

Candan, for such a wonderful friend and colleague both in and out of science, for the great time we had been sharing in and out of lab, for interesting daily talks on almost everything, and also I have been learning a great deal from you.
I would like to give my special thanks to Pedro, Tiago, Helen, Catarina, Pilar, Beata for being so nice to take me into the “gang” and the great time we spent during lunch and other occasions, though I was absent sometimes. Particular thanks to Salomé for your bright smile which dilutes out all the stress I have whenever you are around, and your time and the very helpful advices on my Big Exam preparation. And thanks are also due to Mousumi for being so kind around and your time on preparing me for the Big Exam.

Lena, thanks for the great collaboration and those inspiring scientific discussions. Scarlett, thanks for the scientific talks and collaboration.

Birgitta, for introducing me into the biogenesis of cyanobacterial membranes, and the good collaboration we got. Tatiana, for those interesting discussions we had.

Andreas, thanks for the scientific discussions on infrared spectroscopy which inspired me a lot.

Elzbieta, for introducing me into the field of dual targeting peptides, and the interesting conversations we had, and also inviting me for the dinner. Erika, for nice talks and collaboration, I am sure we will soon get our findings out of door.

Peter, thanks for the interesting discussions in the Big Exam and in my pre-dissertation seminar.

Anki, Bogos, Håkan, Ann, Lotta, Haidi, Maria, Peter, Tornbjörn, and Eddie for all the technical support in the past years.

Charlotta, for the nice collaboration and I think we will soon obtain something interesting from your protein.

I am also especially grateful to all people in EU-Biocontrol network. Regine, for such a good coordinator and your time and efforts spent in the program. Nicola and Liane, for all the practical assistance. Goran, Malgorzata, Jerzy, Corina, Agnieszka, Yana, Julia, Anna, Barbara, Yasemin, Nina, and Catarina for all the time we spent in the workshops and conferences, and those numerous scientific discussions did really help me to be more biophysics-prone.

I owe a special acknowledgement to Erik, Fabrice, Véronique, Guy, Mouna, Michel, Vinciane, Cédric, Paméla, Matthieu, Chantal, Rabia, and Anastassia for creating such a good environment in SFMB, ULB. I was really happy to have one year training there which benefited me a lot both in science and in life.
Thanks are also due to all the people in DBB, Stockholm and SFMB, Brussels. Thanks for all the help through scientific daily life.

I am also indebted to a group of Chinese friends. And I am so fortunate to have so many of you around and really enjoyed the time we spent in Stockholm. Youqian Deng, Yanling Wang, Wei Xiao, Yingbo Lin, and Hao Pang, just to name a few, for always being there whenever I ask for help. Though most of you are back to China, I present my sincere wishes and kind thoughts to all of you. A special thanks to Zhi Wang who always helped me in practical issues from the day I arrived in Sweden and also the time on helping me to analyze the cell by fluorescence microscopy.

At last, but not least, my undying gratitude goes to my family, my parents and my sister, who have been always there to encourage me in life, and showered me with endless love and bless. Also the greatest thanks go to my wife Lei Cheng, for always being there with me through many ups and downs in past years, and I owe you a lot.
## 6. References


4. Sammanfattning på svenska