Chapter 6
Physics of the Cell Membrane

Ben Ovryn, Terrance T. Bishop, and Diego Krapf

Abstract The plasma membrane is a phospholipid bilayer that links the cell to the outside world. The surface of eukaryotic cells consists of both hydrophilic and lipophilic lipid molecules and essentially an equal amount of protein. Some of these proteins are easily removed, while others are deeply integrated with the membrane. The amphipathic lipid molecules self-assemble in aqueous environments to spontaneously form bilayers with their hydrophilic head toward the water and hydrophobic face hiding toward the interior. In this chapter, we first discuss the variety of membrane proteins and evaluate a few examples. Then, we deal with membrane fusion, a process that is ubiquitous to cellular function. Membrane fusion is a process that requires concerted action of proteins and lipids. We dissect the features that make possible this complex process and its tightly regulated temporal evolution. At last, we examine how the material properties of the membrane give rise to its fluid mosaic elastic behavior and present a rigorous mathematical exploration of the energy associated with membrane bending. We introduce the Helfrich Hamiltonian and apply it to the shape and free energy of a membrane as it de-adheres from a flat substrate.

6.1 The Phospholipid Bilayer

Robert Hooke made the first documented microscopic observations of cells more than 350 years ago [1], but these early visualizations were not able to reveal organelles or nuclei. More than 200 years would elapse between Hooke's coining of the term "cells" and an understanding that cells had independent walls [2]. It would take at least

B. Ovryn
New York Institute of Technology, NewYork, NY, USA
e-mail: bovryn@nyit.edu

T. T. Bishop
Southern Illinois University, Carbondale, IL, USA
e-mail: Terrance.bishop@siu.edu

D. Krapf (✉)
Georgetown University, Washington, DC, USA
e-mail: diego.krapf@colostate.edu

K. B. Blagoev and H. Levine (eds.), *Physics of Molecular and Cellular Processes*,
Graduate Texts in Physics, https://doi.org/10.1007/978-3-030-98606-3_6

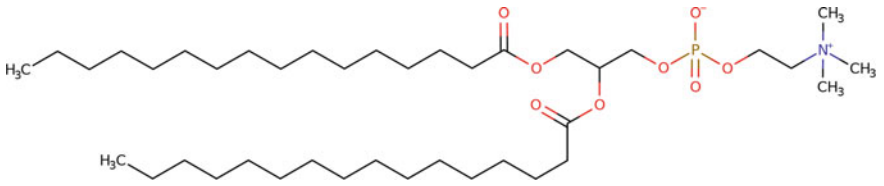

Fig. 6.1 2D structure of a glycerophospholipid. The glycerophospholipid phosphatidylcholine, one of the four main lipids in eukaryotic membranes, has glycerol esterified to two hydrophobic (nonpolar) fatty acid tails and a phosphate group with a hydrophilic choline head. When aligned in a bilayer, the choline head group, phosphate and glycerol have a high probability of being close to the aqueous interface, while the lipid tails, beginning with the C2 carbons, (i.e., the carbonyl group) form the hydrocarbon core. (Image from ChemIDplus TOXNET database (chem.nlm.nih.gov/chemidplus), Dipalmitoylphosphatidylcholine, registry number 2644-64-6)

another 100 years until the fluid mosaic model was articulated to describe the plasma membrane of eukaryotic cells. It is now understood that the phospholipid bilayer of eukaryotic cells consists of both hydrophilic and hydrophobic lipid molecules and essentially an equal amount of protein. These amphipathic lipid molecules self-assemble in aqueous environments to form bilayers with their hydrophilic head toward the water and hydrophobic face hiding toward the interior via an energetically favorable process (in the latter part of the chapter, the more complicated role of membrane fusion is introduced). Although the term lipid bilayer underscores the protein content of the plasma membrane, this name more aptly conveys the significant amount of molecules like glycolipids and cholesterol in the membrane. These molecules combine to provide both the fluid and the mechanical properties of eukaryotic plasma membranes (the contributions of cytoskeletal filaments to the properties of the plasma membrane are not emphasized in this chapter). Often the plasma membrane is depicted as a highly symmetric arrangement of amphipathic lipid molecules, however, there is actually a significant difference in the lipid distribution between the inner and outer monolayers.

The phospholipids in the eukaryotic plasma membrane are predominately glycerophospholipids, consisting of two hydrophobic fatty acid tails (hydrocarbon chains) and a hydrophilic head group. Phosphatidylcholine, a glycerophospholipid that has a terminal choline head group (Fig. 6.1), constitutes nearly a fifth of the lipid content of the red blood cell plasma membrane and nearly 40% of the endoplasmic reticulum membrane and the inner and outer mitochondrial membranes. By contrast, the cell membrane of Escherichia coli, considered the quintessential bacterial membrane [3], essentially consists of a single lipid (phosphatidylethanolamine).

The extended structure of lipid molecules can be approximated as either cylindrical, when the cross-sectional area of the head group is essentially equivalent to the tails, or conical, when there is an asymmetry between these two regions. When cylindrical lipids are arrayed in a bilayer, the relaxed state produces a planar shape. Conversely, if lipid molecules with a large head group are arrayed together, the bilayer has an intrinsic positive (convex) curvature, where each monolayer has a complemen-

tary curvature. In order to form a planar bilayer, bending a positively curved bilayer requires an energy $E = \kappa c_o^2/2$ to reorient the molecules, where κ and c_o represent the bending modulus and the intrinsic or spontaneous curvature, respectively. Using typical values cited in published literature, $\kappa \approx 30\ kT$ [4, 5] and $c_o \approx 0.26\ \text{nm}^{-1}$ [6], yields a bending energy $E \approx 1\ kT\ /\text{nm}^2$ (these concepts will be revisited in greater detail in a subsequent part of the chapter).

6.2 Membrane Proteins

The phospholipids in the bilayer provide both a barrier that separates the cell from the outside world and a solvent for proteins that enable selective transmembrane communication [6]. The fluid mosaic model developed by S.J. Singer and Garth Nicolson, in 1972, encapsulates the idea that hydrophobic proteins can be a stable and integral part of the hydrophobic lipid bilayer [7]. These membrane proteins are often termed "integral" proteins so as to distinguish them from proteins that are peripherally associated with the membrane and easily removed [8]. In this chapter, we use the terms membrane protein and integral membrane protein interchangeably. It is rather remarkable that hydrophobic proteins can reside inside of the hydrophobic core of the lipid bilayer. In fact, because of an intricate and choreographed set of processes, membrane proteins are inserted and folded in the membrane so as to minimize free energy. Membrane proteins and soluble proteins have structures that are governed by thermodynamics with interiors populated by hydrophobic residues and polar residues toward the exterior. Membrane proteins, however, face an additional constraint imposed by the bilayer orientation, defined by the transmembrane axis (perpendicular to the bilayer), such that all spatially extended membrane proteins reflect the preference of some amino acids for the cytoplasmic face (the "positive-inside" rule) [9]. In order for membrane proteins to traverse the thickness of the bilayer, the α-helices and β-strands are generally longer than soluble proteins. This is illustrated in Fig. 6.2 which shows the seven membrane spanning α-helices (shown as coils) and β strands (shown as ribbons) of rhodopsin in the membrane with the top (red) surface representing the exterior of the bilayer (obtained from the RCSB Protein Data Bank [10]).

The hydrophobic match between the membrane protein and the bilayer drives the fluid mosaic so as to minimize the energy penalty caused by having either longer or shorter protein hydrophobic length than the thickness of the relaxed bilayer hydrophobic core. When the hydrocarbons of the bilayer do not adequately cover the hydrophobic regions of the membrane protein, the tails of the lipids can either extend or the protein can be tilted within the bilayer. Because membrane proteins are two to three orders of magnitude, more rigid than the bilayer, the thickness of the bilayer changes so as to reduce the hydrophobic mismatch which results in an energetic cost as the membrane reconfigures. A quantitative analysis of the energy associated with membrane bending will be examined in considerable detail in the subsequent part of the chapter.

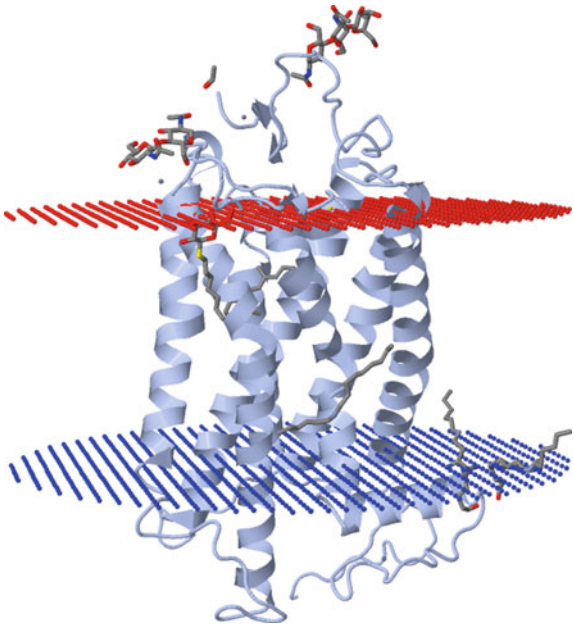

Fig. 6.2 Bovine rhodopsin shown in the plasma membrane. The α-helices and β strands of the seven transmembrane rhodopsin are shown. The exterior portion of the bilayer is on the top (red); the N-terminus of the protein is on the top and the C-terminus is on the bottom right. The hydrodynamic thickness is 3.12 ± 0.15 nm (image from the RCSB Protein Data Bank (www.rcsb.org) of PDB ID 1U19 [10])

6.2.1 Integral Proteins

Integral proteins are also known as transmembrane proteins because they have one or more segments which span the membrane. These proteins have three segments: exoplasmic, cytosolic, and a membrane spanning region. The exoplasmic and cytosolic domains typically have hydrophilic exteriors to interact with an aqueous environment, where the membrane spanning region is typically hydrophobic to interact with the hydrocarbon core of the phospholipid bilayer. Here, the exoplasmic space is regarded as one of three options: the lumen of the ER or Golgi or the cell exterior. In order to span the membrane, integral membranes typically contain one or more α helices or multiple β barrels. Membrane proteins can be further subdivided into classes (I–IV) that depend upon the orientation of the protein with respect to the cytosolic and exoplasmic space (the transmembrane axis), as well as the number of passes. Each of these four types of integral proteins is synthesized at the ER along with certain GPI-linked proteins as well. The synthesis of each of these five ER membrane proteins is fine-tuned to achieve specific protein orientation [11] (Fig. 6.3).

6.2.1.1 Type I Embedded Proteins

Type I proteins have an N-terminal signal sequence that directs them to the ER as well as a membrane spanning α helix. The signal-recognition particle (SRP) binds to its signal sequence on a nascent peptide chain as well as the large ribosomal sub-

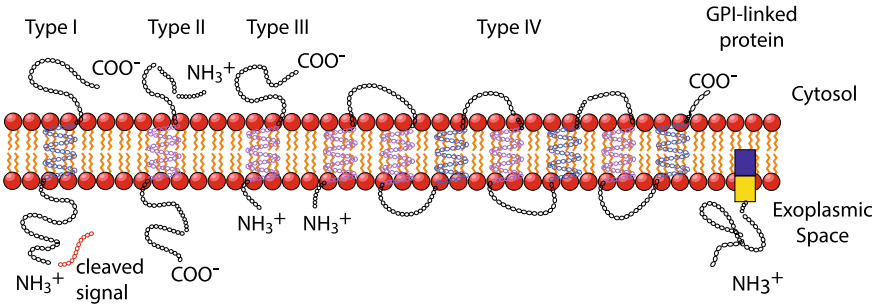

Fig. 6.3 Five potential orientations of integral membrane proteins. The classification of these proteins depends upon their orientation with respect to the membrane, as well as their number of passes through the membrane. The exoplasmic space consists of either the ER or Golgi lumen and the extracellular space in the plasma membrane

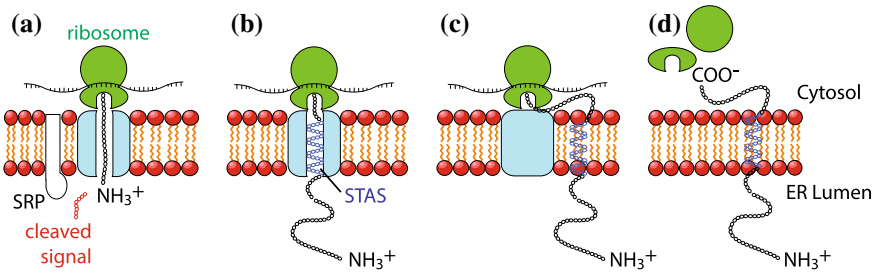

Fig. 6.4 Insertion mechanism for Type I proteins into the ER membrane. a Nascent protein with its N-terminus in the exoplasmic space, i.e., the ER lumen. **b** Protein synthesis pauses at the stop-transfer sequence. **c** The protein is transferred from the translocon to the lipid bilayer. **d** finished Type 1 protein with its C-terminus in the cytoplasm

unit. This nascent protein–ribosome complex is targeted to the SRP receptor on the ER membrane. Once at the membrane, GTP binds the SRP and SRP receptor to strengthen their interaction. The nascent protein is transferred from the SRP/large ribosomal subunit to the translocon, which requires opening the translocation channel. Once in the translocation channel, the protein continues to be synthesized until an internal stop-transfer anchor sequence (STAS) is reached which halts the synthesis. The Stop-transfer sequence is transferred laterally from the translocation channel to the phospholipid bilayer. Upon completion of their synthesis, the ribosomal subunits are released into the cytosol, and the newly synthesized Type I integral proteins are able to freely diffuse in the membrane (Fig. 6.4).

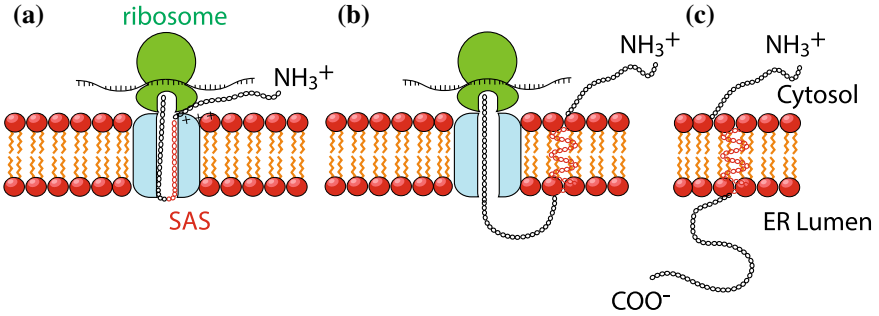

Fig. 6.5 Insertion mechanism for Type II proteins into the ER membrane. a The nascent chain is oriented so that the N-terminus is localized in the cytoplasm. **b** The signal-anchor sequence (SAS) is transferred laterally to the bilayer. **c** A completed Type II protein

6.2.1.2 Type II and Type III Proteins

Type II and Type III proteins do not contain the cleavable signal sequence of a Type I protein, but rather they contain an internal signal-anchor sequence that serves as both a signal sequence to the ER and a membrane anchor sequence. As the Type II and Type III proteins are being translated, their internal signal-anchor sequence (SAS) is synthesized and bound by an SRP which directs the ribosome nascent chain complex to the ER membrane. The orientation of Type III is similar to that of Type I proteins, in that the N-terminus is in the exoplasmic space and the C-terminus is in the cytosol. The orientation of Type II is opposite of Type III orientation. This difference in orientation between Type II and Type III is accomplished by differences in the positioning of the internal signal-anchor sequence. For Type II proteins, the nascent chain is oriented within the translocon with its N-terminal portion toward the cytosol. Next, as the protein chain continues to grow, the Signal-anchor sequence is transferred laterally anchoring the sequence to the phospholipid bilayer. Once protein synthesis is done, the C-terminus of the polypeptide is released into the lumen and the ribosomal subunits are released into the cytosol. Type III synthesis is very similar except the positively charged residues are closer to the C-terminus than the N-terminus resulting in an opposite orientation within the translocon compared to Type II. This results in Type III having its N-terminus in the ER lumen, and its C-terminus in the cytosol (Figs. 6.5 and 6.6).

6.2.1.3 Multiple Topogenic Sequences

Types I, II, and III all utilize a single topogenic sequence. For Type I proteins, there is a single internal stop-transfer anchor sequence, and for Types II and II, there is one internal signal-anchor sequence. Type IV proteins (multipass proteins) have multiple internal topogenic sequences. These membrane spanning α helixes can act to direct the protein to the ER, to anchor the protein, or to stop the transfer of the protein. Type

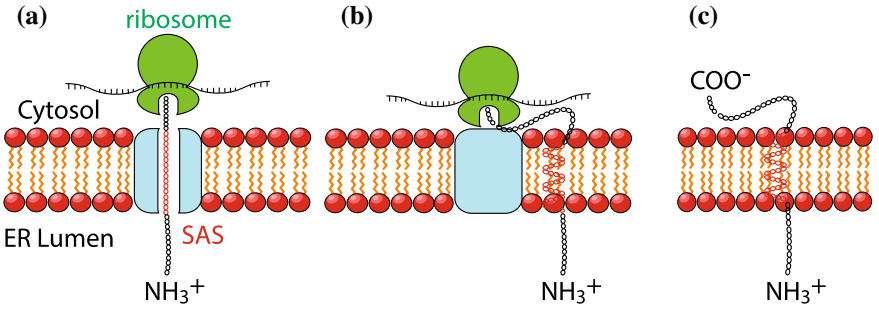

Fig. 6.6 Insertion mechanism for Type III proteins into the ER membrane. Note the presence of a stop-transfer sequence instead of a signal-anchor sequence instead of a stop-transfer anchor sequence. **a** The nascent chain is oriented so that the N-terminus is in the ER lumen. **b** The protein is transferred laterally from the translocon when the SAS is reached. **c** A completed Type III protein

IV proteins are further categorized by the orientation of their N-terminus with respect to the membrane. This N-terminus arrangement is determined by the charge of the sequences that follows the hydrophobic sequence closest to the N-terminus. Type IV proteins are also classified by the number of transmembrane α helices. If there is an even number of membrane spanning regions, the N-terminus and the C-terminus will be facing toward the same side of the membrane, where an odd number of membrane spanning regions would result in opposite orientations of the two termini.

6.2.1.4 Type IV Proteins

Type IV proteins can be classified according to having their N-terminus within the cytoplasmic or extracellular side. Those with N-terminus in the cytosol include proteins like glucose transporters and most ion channel proteins. The hydrophobic section neighboring the N-terminus is responsible for commencing the insertion of the nascent protein sequence into the ER membrane. The first N-terminal α helix and the following odd numbered ones serve as signal-anchor sequences, while the even sequences act as stop-transfer anchor sequences.

Type IV proteins with the N-Terminus in exoplasmic space include the family of G-protein-coupled receptors. The hydrophobic α-helix closest to the N-terminus contains a bunch of positively charged amino acids. These positively charged amino acids act to assist with the insertion of the growing peptide into the translocon with the N-terminus lengthening into the lumen. As the chain continues to grow, it is implanted into the membrane of the ER by alternating type II signal-anchor and stop-transfer sequences.

6.2.1.5 GPI-Linked Proteins

GPI-linked proteins are covalently attached to an amphipathic molecule known as Glycosylphosphatidylinositol (GPI). These GPI-linked proteins are initially synthesized in a similar fashion to other transmembrane proteins, however, a dense sequence of amino acids serves as a signal to be recognized by transamidase. Transamidase simultaneously cleaves the original anchor sequence while transferring the remaining protein to a GPI anchor. One benefit of a GPI anchor is the ability to easily diffuse on the surface of the phospholipid bilayer.

6.2.2 Peripheral Proteins

One method of peripheral protein association is accomplished by disruption of the electrostatic interaction between basic groups of peripheral protein and anionic lipids. The inner leaflet of the plasma membrane of animal cells is composed of 20% anionic lipids that provide negative charges. A peripheral protein can be phosphorylated in the membrane spanning region. This addition of a phosphate is referred to as an electrostatic switch because the phosphate can be cleaved by phosphatases altering the peripheral proteins' membrane binding potential.

Another method of membrane association involves post-translational modification of a peripheral protein to add an acyl chain. When this mechanism is used, a lipid anchor is permanently added in the post-translational modification. This anchor is too weak to associate a peripheral protein to the membrane, so a second acyl chain (or prenyl group) is added. The second chain doubles the strength of the interaction providing sufficient strength to associate the peripheral protein and the membrane. Acyltransferases can be used to modify the affinity of the proteins for the membrane.

The third method of regulation involves the binding of ligands such as nucleotides or Ca^{2+}. Binding of GTP to the G protein ADP- ribosylation factor causes a conformational change in the protein that exposes an amphipathic helix which can insert itself into the bilayer. Ca^{2+} can trigger similar effects. Recoverin is a neuronal calcium binding protein in photoreceptor cells of the eye. When Ca^{2+} binds recoverin, it induces a conformational change where a bound myristoyl group extends out to become a membrane. This "Myristoyl switch" can be seen in other peripheral proteins involved in signal transduction [12]. Finally, peripheral proteins can bind with a specific lipid component of the membrane like DAG or polyphosphorylated inositol.

6.2.3 Receptors

Cell surface receptors are integral membrane proteins that are responsible for communicating external signals to the cell. For surface cell receptors, there are two properties that characterize their interaction with a given ligand, specificity and affinity. Speci-

ficity refers to the ability of a protein to refer one molecule preferentially to another, whereas the affinity refers to the strength of the binding interaction. Molecular complementarity is the idea that the shape and the chemical surface of a binding site must be complementary to its ligand. When a ligand binds to its receptor, a conformational change occurs in that receptor which is communicated across the membrane spanning domain resulting in the subsequent activation of protein in the cytosol. The process of taking an extracellular signal and converting it into an intracellular process is referred to as signal transduction.

6.2.3.1 G Protein-Coupled Receptors—GPCR

GPCRs are the largest family of receptors in cells. They typically contain seven transmembrane helices and have an N-terminus in the extracellular space and their C-terminus in the cytoplasmic space. GPCR activation method involves conformational changes of GPCR, especially in the regions facing the cytosol.

Olfaction is a process which utilizes a G protein-coupled receptor. Olfaction neurons express one type of odor receptor, and in order to detect a wide variety of scents, 400 distinct receptors are used in concert to distinguish 10^{12} different scents. The olfaction signaling cascade begins when an odorant binds the 7-transmembrane odorant receptor. The direct effector is adenylyl cyclase which leads to a cascade that results in an influx of calcium ions and sodium ions. This influx leads to an action potential that is communicated to the brain. For olfactory processes, it is important to have a mechanism for desensitization after prolonged activation. This mechanism is accomplished by GPCR kinase (GRK) that phosphorylates the GPCR on multiple sites, and by arrestin that binds to the phosphorylated GPCR, effectively desensitizing it.

6.2.3.2 Receptor Tyrosine Kinases

Receptor tyrosine kinases are integral membrane proteins with a single transmembrane domain. Their extracellular domain is variable and the cytoplasmic part contains a tyrosine kinase domain. The activation mechanism for receptor tyrosine kinases (RTK) is dependent on kinase domain dimerization and transphosphorylation. RTKs accomplish this activation along with the help of dimeric signal ligands, juxtapose positioning, and release of steric hindrance.

Activation of RTK by epidermal growth factor leads to dimerization and autophosphorylation within the C-terminal tail of the receptor. This phosphorylation recruits in a GTP exchange factor (GEF) and various adaptor containing proteins leading to the activation of Ras and downstream MAP kinase pathways. Activated RTKs can also stimulate phospholipase C activity to generate lipid-derived second messengers.

6.2.3.3 Cytokine Receptors

Cytokine receptors are integral membrane proteins with a single transmembrane domain. Cytokine receptors have no kinase domain, instead they have a kinase tightly associated with their intracellular domains. The associated kinase is Janus kinase (JAK, also known as "just another kinase"), which binds to the cytoplasmic tail. The activation of cytokine receptors is dependent on kinase domain dimerization through interaction with dimeric signal ligands. The binding of a dimeric signal ligand leads to conformational changes in pre-formed dimers.

When a ligand binds, it leads to a conformational change that places the dimers of the cytokine receptor closer together. These dimers JAK region trans-phosphorylates one another. These phosphorylated sites act as binding regions for signal transducer and activator of transcription (STAT). STAT is phosphorylated and released to the cytoplasm where it forms a homodimer. The phosphorylated STAT dimer enters the nucleus and leads to the expression of various genes. One Gene that is activated is a suppressor of cytokine signaling which exhibits negative feedback on the system.

6.2.3.4 Receptor Serine/Threonine Kinases

Serine/threonine kinase receptors are integral membrane proteins with a single transmembrane domain. Serine/threonine kinase receptors feature a Type I receptor and a Type II receptor positioned just next to each other within the membrane. The type I receptor has an inactive kinase domain, whereas the type II receptor has an active kinase domain. The human genome encodes for seven type I receptors and five Type II receptors. Activation of serine/threonine occurs through kinase domain dimerization via trans-activation of type I kinase domain by Type II kinase domain.

Transforming growth factor β (TGFβ) brings together the type I and type II receptors. The type II active kinase phosphorylates the Type I's kinase. R-Smad binds the type I receptor and forms a complex with Smad anchor for receptor activation (SARA). The type I kinase phosphorylates R-Smad which promotes its dissociation from the receptor and SARA. Now in the cytoplasm, R-Smad binds Co-Smad, and the Smad hetero-oligomer enters the nucleus. Smad associates with DNA-binding proteins to activate or inhibit the transcription of specific genes.

6.2.3.5 Integrins and the Glycocalyx

Integrin, an important cell surface receptor, is a glycoprotein with an alpha and beta subunit [13]. As a superfamily, integrins are significantly N-linked glycosylated [14]. The hierarchical processes that lead to integrin binding to protein ligands on the extracellular matrix and the formation of an integrin adhesion requires integrin activation [15–17]. Upon activation, the affinity for a ligand increases as ligand-binding site epitopes in integrin's extracellular domains become exposed. Subsequent to integrin activation, the stalks of the conformationally activated heterodimer have relatively

short projections (tens of nanometers) from the plasma membrane into the extracellular space. These short integrins molecules diffuse among a sea of molecules, called the pericellular matrix or glycocalyx, which also project toward the extracellular matrix. The glycocalyx consists of proteoglycans that have a core protein and many covalently linked glycosaminoglycan (GAG) chains. Heparan sulfate (HS), chondroitin sulfate, and hyaluronan (HA) are prominent GAGs present on most cells (HA is a GAG without a core protein). The longer diffusing molecules of the glycocalyx suggest that the ligand-binding domains of a homogenous population of activated, short integrins are prevented from reaching the extracellular matrix [5]. An analysis of the free energy of these diffusing molecules and the binding energy associated with the plasma membrane, so as to displace integrins toward the extracellular matrix, predicts the magnitude of the energy barrier which prevents the formation or nucleation of an integrin adhesion patch (a cluster of bound integrins). It is conjectured that a stochastic ensemble of a few actin polymerizing filaments is required to nucleate an adhesion [5]. As the adhesion patch grows in size, it becomes a nascent adhesions or a focal complex [18, 19].

6.2.3.6 Surface Cell Receptor Examples

Endocrine Signaling
Endocrine signaling can carry information in the form of hormones at great distances. These hormones are synthesized by endocrine organs and act on various target cells distant from their site of synthesis. The movement of hormones is typically accomplished by blood or other extracellular fluids.

Hormones fall into four general classes: small water-soluble molecules, peptide hormones, lipophilic molecules, which are detected by extracellular receptors, and lipophobic molecules which are detected by intracellular receptors. Hormones have temporal and structural specificity, only being released when needed, and detected only by the cell which they are meant to target. Some examples of the different hormonal classes are the following:

- Small water-soluble molecules: histamine, epinephrine
- Peptide hormones: insulin, luteinizing hormone
- Lipophobic molecules with extracellular receptors: prostaglandins
- Lipophilic molecules with intracellular receptors: cortisol, progesterone

Paracrine Signaling
Paracrine signaling occurs when signaling molecules released by a cell are destined to affect target cells in close proximity to the signal source. The release of neurotransmitters from synaptic vesicles is a prime example of paracrine signaling. The neurotransmitters diffuse across the synaptic cleft where they bind to dendritic receptors

Autocrine
In autocrine signaling, cells respond to signals that they released themselves. Some growth factors act in this fashion. Autocrine stimulation occurs when a cell produces growth factors that act on itself. Many tumors lack balance in the autocrine signaling processes and over-release growth factors which lead to additional tumor proliferation.

6.3 Membrane Fusion

Membrane fusion is a process shared by a variety of cellular processes ranging from fertilization, intracellular trafficking, viral entry, tissue formation, carcinogenesis, and more. And while the goal of membrane fusion is different for each of these cases, they all share similarities within the mechanism they employ to accomplish their function. This section delves into the mechanisms and driving forces for membrane fusion.

6.3.1 Intermediate Structures

The basic structures leading to the fusion of two lipid bilayers are outlined in Fig. 6.7 and summarized below. Mechanistically, it involves bending of the membranes, forming a hemifusion stalk where the proximal leaflets are allowed to mix their components, expanding this stalk into a diaphragm and eventually forming a pore.

Hemifusion is an intermediate stage of membrane fusion that results in lipid mixing between the proximal leaflets of two membranes, but not mixing of the contents of the distal membranes. Hemifusion leads to a distinct membrane structure known as a fusion stalk (Fig. 6.7c). The formation of this fusion stalk depends on a combination of elastic deformations within the membrane leaflets. Necessary deformations include bending, splaying of the hydrocarbon tails of lipids molecules, and tilting of these lipids with respect to the membrane [20]. A more detailed discussion of the physics of membrane deformations is presented later in this section.

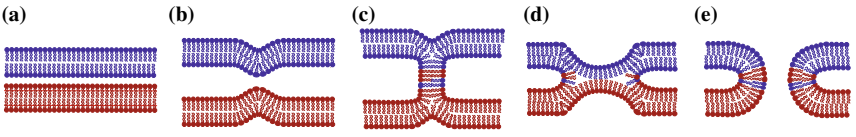

Fig. 6.7 Sketch of membrane fusion leading to the formation of a fusion pore [22]. (**a**) Two membranes prior to fusion initiation. (**b**) A membrane protrusion forms reducing the hydration repulsion energy between the membrane leaflets coming into contact. (**c**) The formation of the hemifusion stalk results in mixing of proximal leaflets but does not involve mixing of the distal leaflets. (**d**) Expansion of the hemifusion stalk results in hemifusion diaphragm. (**e**) A fusion pore forms, either from the hemifusion diaphragm or directly from the fusion stalk

For biological membranes, lipid species can modulate hemifusion formation but generally the physiologically available lipids in the plasma membrane do not provide sufficient driving forces for hemifusion to take place. For example, the hemifusion diaphragm forms spontaneously in fully hydrated membrane systems only when the membrane lipid composition is almost entirely phosphatidylethanolamine (PE) [21]. This condition is not realistic in cell membranes, so it becomes apparent that additional forces via proteins integrated into the membrane are necessary to drive the reaction to completion. Within the cell fusion processes, the hemifusion pore is almost universally formed as an intermediate step to successful fusion events. In these successful events, the hemifusion diaphragm must transition into an expanding fusion pore.

A key difference between the hemifusion diaphragm and the fusion pore is the mixing of lipids from both the proximal and the distal layers of the cell membrane as well as the mixing of the contents initially separated by the apposed membranes. A key similarity between the hemifusion diaphragm and the fusion pore is that in both the cases, their formation depends on specific lipid compositions. Fusion pores in viral fusion and exocytosis have been seen to open on the order of microseconds with a diameter near 2 nm. For the next few milliseconds to seconds, the fusion pore can exhibit an irregular opening and closing pattern known as flickering. However, for cases where there is a successful fusion event, the fusion pore exhibits a gradual expansion that is irreversible. It has been shown that fusion pores are essentially lipidic, that is, they can form in exclusively lipid environment without the help of additional proteins [23]. However, additional fusion proteins can have a strong influence on their properties.

6.3.2 Membrane Tension as a Driving Force

One of the most energetically unfavorable stages of the fusion process is the formation and the expansion of a fusion pore. During this process, the cell drives the expansion of fusion pores by actively modulating local membrane surface tension. Furthermore, membrane tension plays key roles in prefusion and early fusion stages [24].

Due to the fluid nature of lipid membranes, besides the traditional membrane mechanical forces, membrane tension is also affected by lipid flow share rates. These dynamic contributions can be highly anisotropic. On the other hand, static tension can be considered in terms of an analogue to hydrostatic pressure in the 3D liquid case. Therefore, one would expect static tension to be homogeneous across the entire membrane. However, anisotropies can easily develop in the plasma membrane tension. For example, membrane domains, which are widespread in the plasma membrane have different tensions on both sides of the domain boundary. In the same way, that two bulk phases intersect at a surface, which is associated with a surface tension γ_s; two membrane phases intersect at a line, associated with an interfacial line tension γ_l. The tension difference $\Delta\tau$ within a domain of radius R can be computed from the interfacial line tension: $\Delta\tau = \gamma_l / R$. Interfacial line tensions in lipid bilayer domains

have been measured from the thermodynamics of domain nucleation rate, given that in thermodynamic equilibrium, the nucleation rate J follows an Arrhenius law

$$J = J_0 \exp(-\Delta G / k_B T), \tag{6.1}$$

where $\Delta G \propto \gamma_l^2$ is the activation barrier to nucleation and $k_B T$ is thermal energy and J_0 is a constant [25]. Additionally, membrane line tensions are also estimated by shape analysis [26].

Typical tensions of the plasma membrane are in the range of 0.01–0.04 mN/m but, in specialized cells, it can reach much higher values. As mentioned above, the cells modulate membrane tension as a mechanism to drive specific processes such as fusion pore expansion. The forces that modulate membrane tension are related to four general processes:

1. Hydrostatic pressure is controlled by osmotic imbalance between the cytosol and the extracellular medium, i.e., differences in the total ionic concentrations. Due to the nature of hydrostatic pressure, its contributions are homogenous across the cell membrane.
2. The plasma membrane interacts with the cortical cytoskeleton adjacent to the cell membrane. These forces can introduce local heterogeneities and specific local membrane curvatures. Interactions with the cytoskeleton can further promote or inhibit the formation of membrane domains.
3. Adhesion processes provide strong forces that tether the membrane at specific locations.
4. The consumption or abscission of membrane material effectively increases or decreases the area of the membrane, which in turn, alters membrane tension.

6.3.3 Fusion Proteins

Successful fusion events rely on the help of a variety of proteins expressed on the surface of each cell participating in the fusion event. These proteins are necessary to bring the membranes within close contact with one another, catalyze the formation of the initial intermediates, and drive the process to a successful completion. Accomplishing these tasks relies on the coordinated effort of numerous families of proteins [27]. A handful of the key players are highlighted below.

6.3.3.1 Soluble N-Ethylmaleimide-Sensitive-Factor Attachment Protein Receptors - SNAREs

Soluble SNAREs are a superfamily of small proteins that have been termed the workhorses for fusion. There is great variability in sizes and structures of SNAREs, the only truly identifiable feature between them being a SNARE motif. The Snare

motif contains 60–70 amino acids that includes eight heptad repeats for coiled coil domains. Most SNARES contain a C-terminal domain that serves as a membrane anchor. SNARE proteins from opposing membranes associate with core complexes during fusion, and dissociate after, giving rise to a SNARE cycle.

Combining the correct SNARE proteins will lead to spontaneous assemblage into four helical bundles in which the SNARE motifs are arranged in a parallel fashion (Fig. 6.8). When this parallel complex is assembled, SNARE proteins complete their function with an elegant mechanism. SNAREs in opposite membranes intertwine with one another and pull their opposing membranes into closer proximity. There exist four subfamilies of SNARE proteins: Qa-SNAREs, Qb-SNAREs, Qc-SNAREs, and R-SNAREs. All functional SNARE complexes rely on a SNARE from each subfamily. There exists ample evidence that SNARE proteins are responsible for driving membranes into close proximity, but it is still unclear exactly how SNAREs accomplish this mechanism. SNARE proteins remain the best candidate for inducing

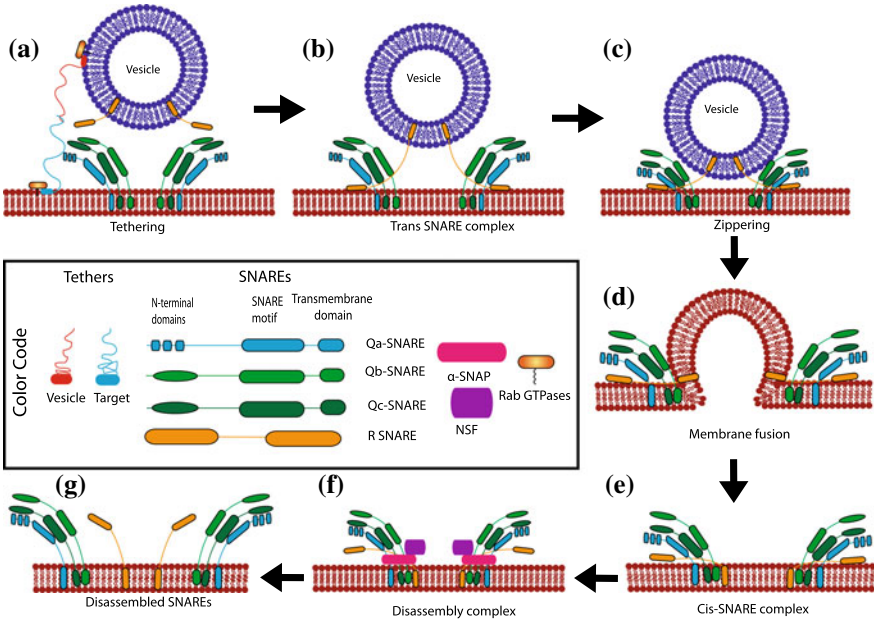

Fig. 6.8 SNARE-mediated vesicle fusion. **a** Tethering process whereby surface molecules on both the membrane target and the vesicle initiate the fusion process at distances up to 50 nm. **b** Formation of the trans-SNARE complex from proteins on the vesicle and the target membrane. **c** Zippering that occurs as the SNAREs on apposed membranes coil together in a spontaneous process. As the SNAREs coil, they form a four-helix complex and the two membranes are brought together. **d** The two membranes merge and their contents are combined. **e** The cis-SNARE complex refers to the energetically favorable complex that is formed over the course of the fusion reaction. **f** Disassembly of cis complex requires recruitment of the cytosolic protein α-SNAP, which, in turn, recruits the ATPase NSF to disassemble the cis-SNARE complex in an energy-dependent manner. **g** Disassembled SNAREs can be used again in additional membrane fusion events

successful fusion by inducing and stabilizing the transition state that leads to the opening of a fusion pore.

6.3.3.2 Sec1/Munc18-Like Proteins—SM

SM proteins are hydrophilic and typically contain 650–700 residues. When SM proteins are deleted, the respective fusion event fails, demonstrating the necessity of SM proteins for fusion. In the human genome, there exists seven varieties of SM proteins which suggests that they are versatile fusion agents that are capable of functioning in multiple fusion reactions. All SM proteins are arch-shaped molecules consisting of three domains and a major v-shaped cleft in the middle. There are four distinct ways that SM proteins interact with SNAREs. The most common type of interaction occurs when the SM binds directly to the N-terminus of a corresponding Qa-SNARE. This binding is sequence specific and has been observed in all fusion reactions within the Golgi complex and the endoplasmic reticulum which suggests its importance in secretory processes. SM proteins like Munc18a, 18b, and 18c can also bind directly to their corresponding Qa-SNARE. This process requires a distinct closed conformation of syntaxins.

SM mediated fusion in yeast cells has been shown not to bind to Qa-SNAREs but rather to fully assembled SNARE complexes. The exact mechanism of this process is yet to be determined. The final mechanism occurs in endosomal/vacuolar fusion via SM proteins interacting with SNAREs indirectly via SNARE binding proteins. Current evidence suggests that SM proteins regulate SNARE assembly in a fashion that is coupled to membrane attachment. It has also been suggested that SM proteins may act as syntaxin chaperones. SM proteins can interact with individual SNAREs, or SNARE complexes to help drive fusion to completion.

6.3.3.3 Rab Proteins

Rab proteins are members of the GTP-binding family. Similar to other proteins in this family, when in a GTP bound state, Rab proteins' initiate downstream effector proteins. The activation and deactivation of Rab proteins is modulated by both GTP exchange factors and GTPase activating proteins. Rab proteins range between 21 and 25 kDa and contain many regions that are highly conserved across the Ras superfamily. Most Rab proteins are tightly associated with the membrane via post-translation addition of two geranylgeranyl (20-carbon polyisoprenoid) groups. The Rab family contains approximately 70 members who act as the master regulators of intracellular vesicle transport. A small fraction of Rab proteins are found within the cytosol, where majority of Rab proteins are localized to membranes of transport vesicles and their target compartments.

6.3.4 Electrostatic Forces

From a physical point of view, the fusion of lipid bilayers in an aqueous environment takes place in two steps [28]. In the first step, the membranes are brought together, so that the process must overcome the repulsive electrostatic forces between the two outer leaflets. The second step involves destabilization of the boundary between the hydrophilic and hydrophobic portion of the bilayer. These transition states rely on forces that minimize the exposure of nonpolar surfaces to water. The successful culmination of these two steps involves the formation of an aqueous fusion pore. The time scale for a successful fusion reaction has been found to depend exponentially on membrane tension, with increased membrane tension leading to more rapid fusion processes.

Example 6.1 Example of Membrane Fusion—Viral Entry Viruses with an envelope require membrane fusion to infect the cell. To achieve this, first and foremost, the virus expresses many fusion proteins locked into a prefusion conformation [29]. Two transitions must occur for the virus cell to become fusogenic. The first process is known as priming and commonly occurs via proteolytic cleavage. The second process is known as triggering and occurs as a result of ligand binding. Examples of ligands that trigger this transition are protons and co-receptor expressed on a cell surface. Viral fusion proteins have been termed "suicide enzymes" because they undergo a irreversible priming step and only function once. This is vastly different when compared to the mechanism of SNARE proteins which exhibit a ATP dependent cycle.

Viral fusion proteins fall into three categories. The first class of fusion proteins are trimers of a single-chain precursor, which requires proteolytic cleavage to become active. The second class achieves priming by cleavage of a heterodimeric partner protein, and the final class of fusion protein has no priming mechanism. When the virus fuses its envelope to its target membrane, its internal contents are released into the cell to wreak havoc.

6.4 Energy Required to Bend a Membrane

6.4.1 Fluid Properties of the Plasma Membrane

Because the membrane has both fluid-like and elastic properties, it is useful to begin with an examination of a curved fluid surface under tension and subsequently look at the effect of elasticity. Consider a transition between two fluids (e.g., water and air). If the transition between the two media occurs over a very thin layer, it can be treated as a distinct boundary which is commonly called a surface. Ultimately, the free energy of the plasma membrane will be analyzed, but it is often more common to analyze the forces and the shape of the surface. Figure 6.9 illustrates a pressure

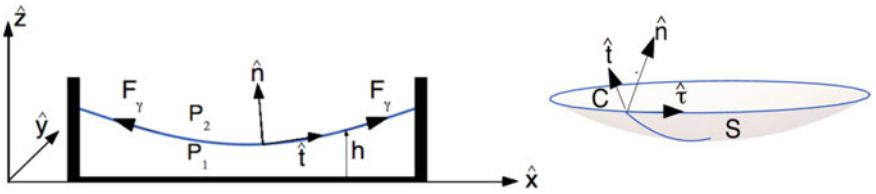

Fig. 6.9 Forces on a fluid surface. Two forces are shown acting on the fluid interface: (1) a pressure difference $\Delta P = P_2 - P_1$, acting in the direction $\hat{\mathbf{n}}$ normal to the surface S, and (2) a surface tension γ, which acts along the boundary C, in the direction $\hat{\mathbf{t}}$ tangent to S and perpendicular to C. The unit vector $\hat{\boldsymbol{\tau}}$ points along the boundary C

difference $\Delta P = P_2 - P_1$, across a surface. Because small creatures can walk on the surface of a fluid, there must be enough force generated to balance their weight. This force arises from the surface tension of the fluid. Surface tension, γ, has units of force/length (i.e. N/m) which is effectively an energy density (J/m^2); water at 25 °C, has a surface tension $\gamma \approx 0.072$ N/m [30].

The shape of the surface S, shown in Fig. 6.9, may be characterized by a height $z = h(x, y)$, above the xy plane; a surface with this parameterization is known as a Monge patch and the Monge parameterization will be invoked throughout this chapter [31, 32]. At any point, $\mathbf{r}(x, y, h(x, y))$, on this smooth surface there is a unit normal vector $\hat{\mathbf{n}}$, which points from medium 1 (with pressure P_1) to medium 2. If the surface is in equilibrium, the pressure difference in the normal direction $\hat{\mathbf{n}}$ must be balanced by the surface tension $\mathbf{F}_\gamma$. The surface tension along the boundary C acts in the direction $\hat{\mathbf{t}}$, where $\hat{\mathbf{t}}$ is a unit vector tangent to the surface. Force balance implies the total force $\mathbf{F}_T$ must cancel [33],

$$\mathbf{F}_T = \iint_S (P_2 - P_1)\, \hat{\mathbf{n}}\, dS + \oint_C \gamma\, \hat{\mathbf{t}}\, d\lambda = 0, \tag{6.2}$$

where the integrals are over the surface and the line perimeter C, respectively. In order to show how these two forces balance, the line integral can be re-stated as a surface integral using Stokes' theorem [34]. From the geometry of Fig. 6.9, Stokes' theorem states $\oint_C \mathbf{f} \cdot \hat{\boldsymbol{\tau}}\, d\lambda = \iint_S (\nabla \times \mathbf{f}) \cdot \hat{\mathbf{n}}\, dS$ where $\mathbf{f}$ is any vector and $\hat{\boldsymbol{\tau}}$ points along C. Letting $\mathbf{f} = \gamma \hat{\mathbf{n}} \times \boldsymbol{\mu}$, where $\boldsymbol{\mu}$ is a constant vector, Stokes' theorem yields

$$\oint_C (\gamma \hat{\mathbf{n}} \times \boldsymbol{\mu}) \cdot \hat{\boldsymbol{\tau}}\, d\lambda = \iint_S \left[\nabla \times (\gamma \hat{\mathbf{n}} \times \boldsymbol{\mu})\right] \cdot \hat{\mathbf{n}}\, dS. \tag{6.3}$$

The line integral in Eq. 6.3 now has a similarity to the second integral in Eq. 6.2, but it is important to note that Eq. 6.3 is an integral along $\hat{\boldsymbol{\tau}}$, whereas the second integral in Eq. 6.2 is an integral along $\hat{\mathbf{t}}$. Using the vector identity $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a})$ yields $(\gamma \hat{\mathbf{n}} \times \boldsymbol{\mu}) \cdot \hat{\boldsymbol{\tau}} = \boldsymbol{\mu} \cdot (\gamma \hat{\boldsymbol{\tau}} \times \hat{\mathbf{n}})$ and from Fig. 6.9, $\hat{\mathbf{t}} = \hat{\boldsymbol{\tau}} \times \hat{\mathbf{n}}$, so the integrand on the left-hand side of Eq. 6.3 becomes $(\gamma \hat{\mathbf{n}} \times \boldsymbol{\mu}) \cdot \hat{\boldsymbol{\tau}} = \boldsymbol{\mu} \cdot \gamma\, \hat{\mathbf{t}}$. The integrand on the right hand side of Eq. 6.3 can be simplified using the vector identity $\nabla \times$

$(\mathbf{A}\times\mathbf{B}) = \mathbf{A}(\nabla\cdot\mathbf{B}) - \mathbf{B}(\nabla\cdot A) + \mathbf{B}\cdot\nabla\mathbf{A} - \mathbf{A}\cdot\nabla\mathbf{B}$, such that $\left[\nabla\times(\gamma\hat{\mathbf{n}}\times\boldsymbol{\mu})\right]\cdot\hat{\mathbf{n}} = \left\{\boldsymbol{\mu}\cdot\nabla(\gamma\hat{\mathbf{n}}) - \boldsymbol{\mu}\left[\nabla\cdot(\gamma\hat{\mathbf{n}})\right]\right\}\cdot\hat{\mathbf{n}} = \boldsymbol{\mu}\cdot\left\{\hat{\mathbf{n}}\cdot\nabla(\gamma\hat{\mathbf{n}}) - \hat{\mathbf{n}}\left[\nabla\cdot(\gamma\hat{\mathbf{n}})\right]\right\}$. Therefore, Eq. 6.3 becomes

$$\oint_C \boldsymbol{\mu}\cdot\gamma\hat{\mathbf{t}}\,d\lambda = \iint_S \boldsymbol{\mu}\cdot\left\{\hat{\mathbf{n}}\cdot\nabla(\gamma\hat{\mathbf{n}}) - \hat{\mathbf{n}}\left[\nabla\cdot(\gamma\hat{\mathbf{n}})\right]\right\}\,dS. \tag{6.4}$$

For a constant surface tension γ, Eq. 6.4 simplifies considerably,

$$\oint_C \gamma\,\hat{\mathbf{t}}\,d\lambda = -\iint_S \gamma\hat{\mathbf{n}}(\nabla\cdot\hat{\mathbf{n}})\,dS. \tag{6.5}$$

Therefore, Eq. 6.2 can be written as $\mathbf{F}_T = \iint_S \left[\Delta P - \gamma\left(\nabla\cdot\hat{\mathbf{n}}\right)\right]\hat{\mathbf{n}}\,dS = 0$, yielding

$$\Delta P = \gamma\left(\nabla\cdot\hat{\mathbf{n}}\right), \tag{6.6}$$

which relates the pressure difference to the curvature of the surface, $\nabla\cdot\hat{\mathbf{n}}$. If the pressure difference is not zero, then the surface must be curved. Conversely, for a flat surface in equilibrium $\Delta P = 0$. Introducing $H = -(\nabla\cdot\hat{\mathbf{n}})/2$ into Eq. 6.6, the Young–Laplace equation (6.7) is obtained [33]

$$\Delta P = P_{inside} - P_{outside} = -2H\gamma. \tag{6.7}$$

With the height of the surface characterized as $z = h(x, y)$, a straightforward approach to obtain the unit normal vector $\hat{\mathbf{n}}$ is to consider the surface defined by the equation $f(x, y, z) = z - h(x, y) = 0$ and use Eq. 6.8 [31],

$$\hat{\mathbf{n}} = \frac{\nabla f}{\sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 + \left(\frac{\partial f}{\partial z}\right)^2}} = \frac{\nabla(z - h(x, y))}{|\nabla(z - h(x, y))|}. \tag{6.8}$$

Therefore

$$\hat{\mathbf{n}} = \frac{\left\{-\frac{\partial h(x,y)}{\partial x}, -\frac{\partial h(x,y)}{\partial y}, 1\right\}}{\sqrt{\left[\frac{\partial h(x,y)}{\partial x}\right]^2 + \left[\frac{\partial h(x,y)}{\partial y}\right]^2 + 1}}. \tag{6.9}$$

In order to simplify the notation, it is easier to let $h = h(x, y)$ and introduce a shorthand for the partial derivatives $h_i = \frac{\partial h}{\partial i}$, $h_{ii} = \frac{\partial^2 h}{\partial i^2}$, and $h_{ij} = \frac{\partial^2 h}{\partial i\partial j}$ with i and j being x or y. With $H = -(\nabla\cdot\hat{\mathbf{n}})/2$ and Eq. 6.9,

$$H = \frac{h_{\mathrm{xx}}\left(h_y^2 + 1\right) - 2h_x h_y h_{\mathrm{xy}} + h_{\mathrm{yy}}\left(h_x^2 + 1\right)}{2\left(h_x^2 + h_y^2 + 1\right)^{3/2}}. \tag{6.10}$$

For a 2D surface with h independent of y, i.e., $h = h(x)$, Eq. 6.10 simplifies to

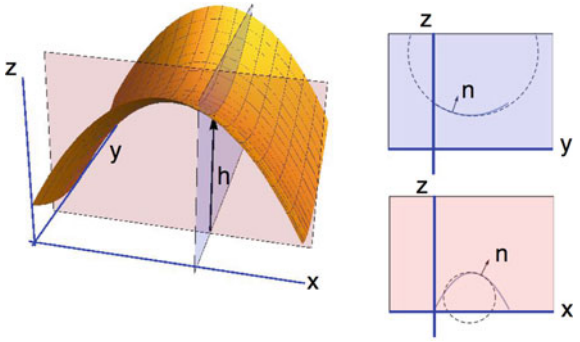

Fig. 6.10 A surface characterized by two radii of curvature. The surface is characterized in the Monge parameterization with $h(x, y)$. Two orthogonal planes show the two different normals and their osculating circles

$$H = \frac{h''(x)}{2\left[1 + h'(x)^2\right]^{3/2}}. \tag{6.11}$$

For a sphere of constant radius R, the height becomes $h(x, y) = \sqrt{R^2 - x^2 - y^2}$ and $H = -1/R$; the minus sign is consistent with a normal vector which points outward. Therefore, for a spherical bubble, $\Delta P = 2\gamma/R$.

Figure 6.10 shows a somewhat more complicated surface and the curvatures (with their osculating circles) in two orthogonal planes chosen along the xz and yz planes. Although there is nothing necessarily special about these two orthogonal planes, there are two orthogonal planes, called the principal planes, for which the curvatures will assume extreme values (a maximum and minimum). Once these two curvatures (k_1 and k_2) are determined, the curvature in other planes may be obtained from their linear combination. Additionally, the curvature H is the mean of these curvatures, $H = \frac{1}{2}(k_1 + k_2)$ [31, 32]. In order to completely characterize the bending of the surface, another curvature, called the Gaussian curvature, is required. The Gaussian curvature is $K = k_1 k_2$. For a surface characterized as $h = h(x, y)$, the Gaussian curvature is given as Eq. 6.12 and the principal curvatures can be determined.

$$K = \frac{h_{xx}h_{yy} - h_{xy}^2}{\left(h_x^2 + h_y^2 + 1\right)^2}. \tag{6.12}$$

$$\begin{aligned} k_1 &= H + \sqrt{H^2 - K}, \\ k_2 &= H - \sqrt{H^2 - K}. \end{aligned} \tag{6.13}$$

The mean curvature takes on a special significance because surfaces characterized by $k_1 = -k_2$ ($H = 0$ and $K < 0$) are called minimal surfaces because they have minimal surface area; a necessary condition for a minimal surface is that it has vanishing mean curvature. Any perturbation in the curvature of a minimal surface will increase the surface area. Clearly, a sphere is not a minimal surface, but it does

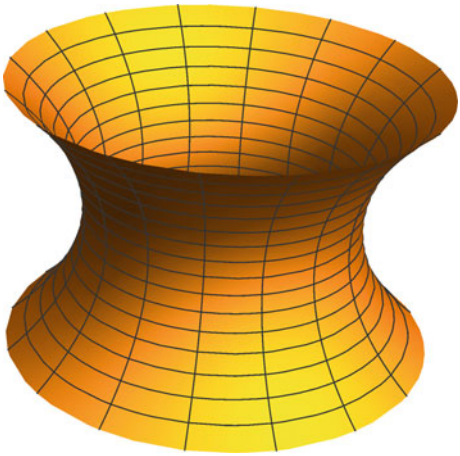

Fig. 6.11 The catenoid is a minimal surface. When $h(x, y) = c\cosh^{-1}\left(\frac{\sqrt{x^2+y^2}}{c}\right)$, we obtain $H = 0$

have the important property of minimizing the ratio of its surface area to volume. Certain surfaces of revolution (e.g., plane, cylinder, catenoid), have been known since the mid 1800s to be minimal surfaces. Although it is straightforward (as shown below) to solve the forward problem and test if a surface is minimal, the inverse problem of searching for minimal surfaces is much more difficult and is of historical importance in the development of the calculus of variations; in the mid-1700s, Lagrange applied variational methods to search for minimal surfaces; in Sect. 6.4.3, an Euler–Lagrange equation will be used to find the surface of minimal energy associated with bending of the plasma membrane.

The problem of finding a surface of smallest area that is bounded by a closed contour is referred to as Plateau's Problem because Joseph Plateau performed many cool experiments with soap films (mixtures of water, glycerin and an emulsifier, sodium oleate) between brass wire frames [35]. Figure 6.11 shows a catenoid which is the shape a thin film adopts when two soapy rings of fixed radius are slowly pulled apart; before the film ruptures, the surface forms catenoid shape so as to minimize its total energy. With $h(x, y) = c\cosh^{-1}\left(\frac{\sqrt{x^2+y^2}}{c}\right)$, it can be readily verified using Eqs. 6.10 and 6.13, that $H = 0$.

6.4.2 *Bending Energies and the Helfrich Hamiltonian*

Unlike a Newtonian fluid, the phospholipid bilayer has elastic properties, but like a fluid, the membrane cannot support in-plane shear. Energy is required, however, for other types of deformations, such as bending. From the theory of elasticity, the energy density associated with bending a thin plate so that it obtains a curvature, C, is $E_\kappa = \frac{1}{2}\int_{\text{membrane}} \kappa\, C^2\, dA$, where the bending modulus $\kappa = YI$ is the product of

the Young's modulus, Y, and the principal moment of inertia [36, 37]. For phospholipid membranes, the bending modulus $\kappa \approx 30\,kT$ [4, 5]. Therefore, the free energy associated with bending the plasma membrane should have a dependence upon the bending modulus and the square of the curvature.

The free energy of membrane stretching and bending is a scalar and can only be a function of curvature terms which are also invariant with respect to a change in coordinate systems. Both the mean curvature, H, and Gaussian curvature, K, are invariant with respect to a change of coordinates and depend only upon the direction of the surface normal, $\hat{\mathbf{n}}$ [32]. For a closed membrane, such as a vesicle, it makes sense to distinguish between an inward and outward pointing normal vector, however, this does not hold for the symmetry associated with a flat membrane. Accordingly, the free energy associated with membrane deformation is expected to have a quadratic dependence upon the mean curvature, but not a linear dependence. By contrast, the Gaussian curvature is an intrinsic property of the surface and does not depend upon the sign of $\hat{\mathbf{n}}$ [32]. Equation 6.14, which shows the dependence of the free energy associated with membrane deformation upon the curvatures, is often referred to as the Helfrich Hamiltonian because it is analogous to Helfrich's 1973 derivation of the energy required to stretch and bend the membrane [31, 38–40]. Helfrich referred to the two elastic moduli (κ and κ_G in Eq. 6.14), as the splay and saddle splay modulus to emphasize that his derivation followed closely from work on the curvature elasticity of liquid crystals [31, 32]. Equation 6.14, is called the Helfrich equation and often the Canham–Helfrich free energy to afford credit to Canham who formulated an energy functional in 1970: [41]

$$E = \iint_S \left(2\kappa\,(H - c_o)^2 + \gamma + \kappa_G K\right) dA + \iiint_V \Delta P\, dV \tag{6.14}$$

Equation 6.14 demonstrates that the energy is proportional to the square of the mean curvature and contains a term, c_o, which represents the spontaneous curvature of the membrane. When $H = c_o$, the free energy is minimal. It is often appropriate to assume that the spontaneous curvature equals zero because the membrane is symmetric. If the molecules have a preferential bending direction or in the presence of proteins which bend the membrane, than the spontaneous curvature will not be zero. In many instances, the topology of the surface does not change and the dependence of E upon the Gaussian curvature K can be ignored because the integral of the Gaussian curvature is a constant. For example, for a sphere, $K = 1/R^2$ and $E(K) = \kappa_G \iint K\,dA = \frac{\pi}{2}\kappa_G$. Therefore, assuming that both c_o and ΔP equal zero and ignoring the Gaussian term, the Helfrich free energy can be written as

$$E = \iint \left(2\kappa H^2 + \gamma\right) dA. \tag{6.15}$$

The free energy required to bend a flat membrane may be obtained from the Helfrich integral (Eq. 6.15) which integrates the energy density, $e_m = 2\kappa H^2 + \gamma$, over the entire membrane. Equation 6.15 is a complicated, non-linear differential equa-

tion because both the mean curvature H and differential area dA depend upon gradients in $h(x, y)$. With $\mathbf{r}(x, y, z) = \{x, y, h(x, y)\}$, the area element dA can be determined from the tangent vectors $\bar{\mathbf{e}}_x = \{1, 0, h_x\}$ and $\bar{\mathbf{e}}_y = \{0, 1, h_y\}$ as $dA = \left|\mathbf{e}_x(x, y) \times \mathbf{e}_y(x, y)\right| dx\, dy$, which yields $dA = dx\, dy\sqrt{1 + h_x^2 + h_y^2}$. Fortunately, it may generally be assumed that the bending is weak such that the gradients in $h(x, y)$ can be approximated. Under this simplification, called the small gradient approximation, terms of $\mathcal{O}\left[(\nabla h)^2\right]$ and higher are ignored. The small gradient approximation applied to Eq. 6.10 yields: $H \approx \frac{1}{2}(h_{xx} + h_{yy})$ and $dA \approx \left[1 + \frac{1}{2}(h_x^2 + h_y^2)\right] dxdy$. Inserting these approximations into Eq. 6.15 (and disregarding constant surface integrals), yields the simplified form of the Helfrich equation

$$E = \frac{1}{2} \iint \left[\kappa(h_{xx}^2 + h_{yy}^2) + \gamma(h_x^2 + h_y^2)\right] dx\, dy. \tag{6.16}$$

6.4.3 *Free Energy and Shape of a Bent Membrane*

As an example, applying the Helfrich Hamiltonian to the geometry of the membrane in Fig. 6.12, the membrane shape is given as $h = h(x)$ and the free energy in Eq. 6.16 becomes

$$E = \frac{1}{2} \iint (\kappa h_{xx}^2 + \gamma h_x^2) dx\, dy. \tag{6.17}$$

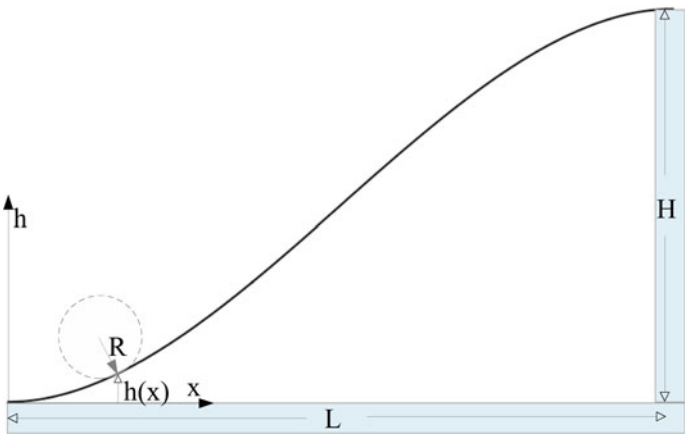

Fig. 6.12 Bending of the plasma membrane to form a projection of height H at a distance L from a point adhesion. Assuming an initially flat membrane which remains adhered to the ECM at the origin, energy bends the membrane to a maximum distance H at length L from the origin

The equilibrium shape, $h(x)$, will minimize the energy subject to the membrane's boundary conditions. It is easy to test if a given shape, $h(x)$, satisfies Eq. 6.17, however, not every shape will minimize the total energy. Therefore, this forward method does not prove that the shape, $h(x)$ minimizes the energy. Fortunately, the inverse problem of predicting the shape that will produce a stationary value of the energy, E, can be solved using the calculus of variations developed by Euler and Lagrange [36, 42]. Their variational approach states that the solution, $h(x)$, satisfies the Euler–Lagrange equation. In 1D, the Euler–Lagrange equation is particularly simple to formulate: given the functional $F = \int_a^b H\left[x, f(x), f'(x)\right] dx$, the Euler–Lagrange equation becomes $\frac{\partial F}{\partial f} - \frac{d}{\mathrm{dx}} \frac{\partial F}{\partial f'} = 0$. The Euler–Lagrange equation with the Helfrich energy (Eq. 6.17) as the functional yields

$$\kappa\, h''''(x) - \gamma h''(x) = 0, \tag{6.18}$$

for the shape of the membrane $h(x)$. This differential equation is sometimes called the shape equation. It can be verified by direct substitution that the solution is

$$h(x) = \frac{\kappa e^{-\frac{\sqrt{\gamma}x}{\sqrt{\kappa}}}\left(c_1 e^{\frac{2\sqrt{\gamma}x}{\sqrt{\kappa}}} + c_2\right)}{\gamma} + c_3 + c_4 x. \tag{6.19}$$

We can solve for the four unknowns in Eq. 6.19 by applying the appropriate boundary conditions. As shown in Fig. 6.12, the membrane is bound to the ECM at the origin and at $x = L$. Therefore, we have $h(0) = 0$; $h'(0) = 0$; $h(L) = H$ and $h'(L) = 0$. The solution to the four equations readily yields four constants. Introducing the variable $\lambda = \sqrt{\kappa/\gamma}$, which has units of length, the constants are given by

$$c_1 = -\frac{\gamma H}{\sqrt{\gamma}\sqrt{\kappa}L\left(e^{\frac{\sqrt{\gamma}L}{\sqrt{\kappa}}} + 1\right) - 2\kappa\left(e^{\frac{\sqrt{\gamma}L}{\sqrt{\kappa}}} - 1\right)}, \tag{6.20}$$

$$c_2 = -\frac{\gamma H e^{\frac{\sqrt{\gamma}L}{\sqrt{\kappa}}}}{\sqrt{\gamma}\sqrt{\kappa}L\left(e^{\frac{\sqrt{\gamma}L}{\sqrt{\kappa}}} + 1\right) - 2\kappa\left(e^{\frac{\sqrt{\gamma}L}{\sqrt{\kappa}}} - 1\right)}, \tag{6.21}$$

$$c_3 = \frac{H\sqrt{\kappa}\left(e^{\frac{\sqrt{\gamma}L}{\sqrt{\kappa}}} - 1\right)}{2\sqrt{\kappa}\left(e^{\frac{\sqrt{\gamma}L}{\sqrt{\kappa}}} - 1\right) - \sqrt{\gamma}L\left(e^{\frac{\sqrt{\gamma}L}{\sqrt{\kappa}}} + 1\right)}, \tag{6.22}$$

$$c_4 = \frac{\sqrt{\gamma}H\left(e^{\frac{\sqrt{\gamma}L}{\sqrt{\kappa}}} + 1\right)}{\sqrt{\gamma}L\left(e^{\frac{\sqrt{\gamma}L}{\sqrt{\kappa}}} + 1\right) - 2\sqrt{\kappa}\left(e^{\frac{\sqrt{\gamma}L}{\sqrt{\kappa}}} - 1\right)}. \tag{6.23}$$

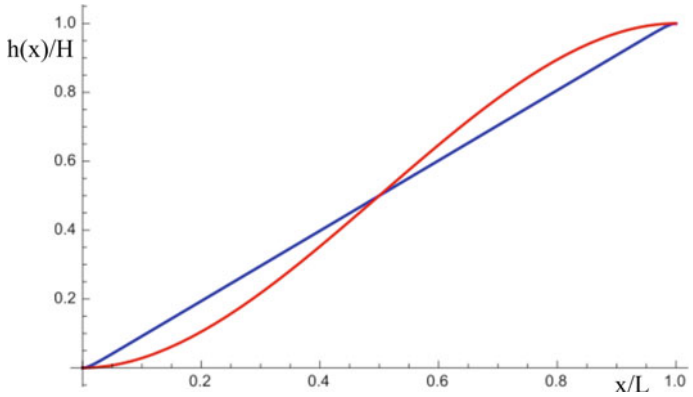

Fig. 6.13 Shape of the membrane $h(x)$ depends upon ratio of the length scale λ. For very small λ, the large tension prevents bending of the membrane ($\lambda/L = 0.01$, blue). By contrast, when the tension is very low, the membrane bends considerably ($\lambda/L = 1$, red)

Inserting these four constants into Eq. 6.20, the shape $h(x)$ is given by

$$h(x) = \frac{H\left(\lambda - \lambda e^{L/\lambda} + x e^{L/\lambda} + \lambda e^{\frac{L-x}{\lambda}} - \lambda e^{x/\lambda} + x\right)}{2\lambda + L e^{L/\lambda} - 2\lambda e^{L/\lambda} + L}. \tag{6.24}$$

Thus, the energy required to bend the membrane into this shape can be obtained by inserting Eq. 6.24 into Eq. 6.17

$$\begin{aligned} E &= \int_0^L \kappa \left(h''(x)^2 + \frac{h'(x)^2}{\lambda^2}\right) dx \\ &= \frac{H^2\kappa\left(e^{L/\lambda} + 1\right)}{\lambda^2\left(2\lambda + L e^{L/\lambda} - 2\lambda e^{L/\lambda} + L\right)} \\ &= \frac{H^2\kappa}{\lambda^2\left(L - 2\lambda \tanh\left(\frac{L}{2\lambda}\right)\right)}. \end{aligned} \tag{6.25}$$

The shape of the membrane, which depends upon the length scale λ, is shown in Fig. 6.13. In the limiting case where surface tension goes to zero, $\gamma \to 0$, the shape, $h(x)$, is given by

$$\lim_{\lambda\to\infty} h(x) = \frac{3Hx^2}{L^2} - \frac{2Hx^3}{L^3}. \tag{6.26}$$

References

1. R.. Hooke, *Micrographia: Or Some Physiological Descriptions of Minute Bodies Made by Magnifying Glasses, with Observations and Inquiries Thereupon* (Dover Publications Reprint Edition, 2003)
2. J.L. Robertson, J. Gen. Physiol. **150**(11), 1472 (2018)
3. C. Sohlenkamp, O. Geiger, FEMS Microbiol. Rev. **40**(1), 133 (2016)
4. E. Evans, W. Rawicz, Phys. Rev. Lett. **64**(17), 2094 (1990)
5. E. Atilgan, B. Ovryn, Biophys. J . **96**(9), 3555 (2009)
6. O.S. Andersen, R.E. Koeppe, Ann. Rev. Biophys. Biomol. Struct. **36**, 107 (2007)
7. S.J. Singer, G.L. Nicolson, Science **175**(4023), 720 (1972)
8. B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, ISBN **1174808063**, 1392 (2007)
9. S.H. White, W.C. Wimley, Ann. Rev. Biophys. Biomol. Struct. **28**(1), 319 (1999)
10. T. Okada, M. Sugihara, A.N. Bondar, M. Elstner, P. Entel, V. Buss, J. Mol. Biol. **342**(2), 571 (2004)
11. H. Lodish, A. Berk, C.A. Kaiser, C. Kaiser, M. Krieger, M.P. Scott, A. Bretscher, H. Ploegh, P. Matsudaira, et al., *Molecular Cell Biology* (Macmillan, 2008)
12. M. Luckey, *Membrane Structural Biology: with Biochemical and Biophysical Foundations* (Cambridge University Press, 2014)
13. R.O. Hynes, Science **300**(5620), 755 (2003)
14. J. Gu, Regulation of integrin functions by N-glycans. Glycoconj. J. **21**, 9 (2004)
15. R. Zaidel-Bar, C. Ballestrem, Z. Kam, B. Geiger, J. Cell Sci. **116**(22), 4605 (2003)
16. M. Cohen, D. Joester, B. Geiger, L. Addadi, ChemBioChem **5**(10), 1393 (2004)
17. C.G. Galbraith, K.M. Yamada, J.A. Galbraith, Science **315**(5814), 992 (2007)
18. M.A. Schwartz, Cold Spring Harb. Perspect. Biol. **2**(12), a005066 (2010)
19. B. Geiger, K.M. Yamada, Cold Spring Harb. Perspect. Biol. **3**(5), a005033 (2011)
20. Y. Kozlovsky, L.V. Chernomordik, M.M. Kozlov, Biophys. J . **83**(5), 2634 (2002)
21. L.V. Chernomordik, M.M. Kozlov, Cell **123**(3), 375 (2005)
22. H.R. Marsden, I. Tomatsu, A. Kros, Chem. Soc. Rev. **40**(3), 1572 (2011)
23. L.V. Chernomordik, M.M. Kozlov, Nat. Struct. Mol. Biol. **15**(7), 675 (2008)
24. M.M. Kozlov, L.V. Chernomordik, Curr. Opin. Struct. Biol. **33**, 61 (2015)
25. C.D. Blanchette, W.C. Lin, C.A. Orme, T.V. Ratto, M.L. Longo, Langmuir **23**(11), 5875 (2007)
26. T. Baumgart, S.T. Hess, W.W. Webb, Nature **425**(6960), 821 (2003)
27. D. Ungar, F.M. Hughson, Ann. Rev. Cell Dev. Biol. **19**(1), 493 (2003)
28. R. Jahn, T. Lang, T.C. Südhof, Cell **112**(4), 519 (2003)
29. S.C. Harrison, Nat. Struct. Mol. Biol. **15**(7), 690 (2008)
30. J.C. Berg, *An Introduction to Interfaces & Colloids: The Bridge to Nanoscience* (World Scientific, 2010)
31. S. Safran, *Statistical Thermodynamics Of Surfaces And Membranes* (CRC Press, Interfaces, 2018)
32. R.D. Kamien, Rev. Mod. Phys. **74**(4), 953 (2002)
33. L. Landau, E. Lifshitz. *Fluid Mechanics: Course of Theoretical Physics*(butterworth, 1987)
34. G.B. Arfken, H.J. Weber, *Mathematical Methods for Physicists* (Academic, 2001)
35. R. Courant, H. Robbins **2**, 901 (1956)
36. A.S. Saada, *Elasticity: Theory and Applications* (Krieger Publishing Company, 1983)
37. L. Landau, E. Lifshitz, *Theory of Elasticity* (Pergamon Press, Oxford, 1959)
38. W. Helfrich, Zeitschrift für Naturforschung c **28**(11–12), 693 (1973)
39. H. Deuling, W. Helfrich, Journal de Physique **37**(11), 1335 (1976)
40. H. Deuling, W. Helfrich, Biophys. J. **16**(8), 861 (1976)
41. P.B. Canham, J. Theor. Biol. **26**(1), 61 (1970)
42. G.A. Maugin, in *Continuum Mechanics Through the Eighteenth and Nineteenth Centuries* (Springer, 2014), pp. 7–32