# Analysis of cell envelope proteins

Mickaël Desvaux, Michel Hébraud

# 12 Analysis of Cell Envelope Proteins

*Mickaël Desvaux and Michel Hébraud*

## CONTENTS



## 12.1 INTRODUCTION

From a morphological point of view, the most fundamental dichotomy within prokaryotes (the term "prokaryotes" is used here in its primary etymological sense—that is, single-celled organisms without nuclei as opposed to eukaryotes, without any further phylogenetic considerations[1]) is between those bound by a single biological membrane (monoderm prokaryotes)—that is, the cytoplasmic membrane, and those bound by two concentric but topologically different membranes (diderm prokaryotes)—that is, the inner membrane (cytoplasmic membrane) and the asymmetric outer membrane.[2] In accordance with holistic and teleonomic concepts, organisms are far more than mere collections of genes,[3,4] and such difference in membrane organization, and thus cell compartmentation, is not trivial but has profound phylogenetic, structural, metabolic, and physiological implications. Based on the most recent advances in biological evolution and megaclassification of organisms,[5–7] monoderm prokaryotes are regrouped under the term Monodermata (also called Unibacteria), which essentially includes Archaea together with Posibacteria (formerly called Gram-positive bacteria).

It is worth stressing that the term "Gram-positive bacteria" is terminologically ambiguous, especially for researchers interested in aspects related to bacterial cell envelope (e.g., protein secretion or surface proteins).[8] From its origin, a positive or negative result given by Gram staining method

indicates whether or not bacteria retain the stain respectively. Later on, the difference in staining was related to profound divergence in structural organization of the cell envelope, briefly: (1) a cytoplasmic membrane surrounded by a thick cell wall in Gram-positive bacteria, and (2) a cytoplasmic membrane surrounded by a thin cell wall beneath the outer membrane in Gram-negative bacteria. Molecular analyses further revealed that, contrary to Gram-negative bacteria, Gram-positive bacteria correspond to a phylogenetically coherent grouping of prokaryotes within the domain Bacteria with phylum BXIII Firmicutes (low G+C mole percent) and phylum BXIV Actinobacteria (high G+C mole percent).[9,10] However, from Gram staining to cell envelope organization to taxonomic grouping, each step represents some approximations, which often result in misleading or incoherent statements in the literature. For example, some members of Firmicutes and Actinobacteria phyla do not retain Gram stain because of (1) the absence of a cell wall (e.g., bacteria from the genus *Mycoplasma*), (2) a too thin cell wall (e.g., some members of the genus *Clostridium*), or (3) the presence of a waxy outer sheath preventing penetration of the stain (e.g., species from the genus *Mycobacterium*).

Inversely, some bacteria not taxonomically related to Gram-positive bacteria retain the Gram stain (e.g., some members of the phylum BIV Deinococcus-Thermus). More confusingly, some bacteria clearly possessing a Gram-negative-like cell envelope architecture are in fact phylogenetically related to the taxonomic group of Gram-positive bacteria (e.g., *Thermotoga maritima* currently classified in phylum BII Thermotogae,[11] or *Fusobacterium nucleatum* belonging to phylum BXXI Fusobacteria).[12,13] Some other phyla regroup bacteria exhibiting both cell envelope structures (Gram-negative-like or Gram-positive-like cell envelope)—for example, BVI Chloroflexi or BVII Thermomicrobia.[14] Even in some deep branches of the phylum Firmicutes, some bacteria clearly exhibit Gram-negative cell envelope ultrastructure (e.g., in genus *Desulfotomaculum*, *Selenomonas*, *Syntrophomonas,* or *Coprothermobacter*).[2] Therefore, it appears in numerous cases that the term "Gram-positive bacteria" cannot describe at once a particular Gram staining result, cell envelope organization, and taxonomic group; thus, when employing this term it is extremely important to specify what it refers to. Because of fewer terminological ambiguities, the terms "Monodermata" or "monoderm bacteria" will be preferred to describe prokaryotic cells surrounded by a single biological membrane but without any further phylogenetic considerations. For the purpose of the present review, the term "Gram-positive bacteria" will be used to describe bacteria with a cell envelope composed of (1) a cytoplasmic membrane, and (2) a cell-wall composed at least of peptidoglycan.

*Listeria* species are monoderm bacteria possessing a thick cell wall retaining Gram stain and belonging to phylum Firmicutes, class Bacilli, order Bacillales, and family Listeriaceae,[9] and as such are Gram-positive bacteria in all meaning of the term. *L. monocytogenes* is undoubtedly the species that has attracted most attention, considering its frequent occurrence in food coupled with a high mortality rate.[15] Still, the genus *Listeria* comprises six species: (1) two pathogenic ones (*L. monocytogenes,* a human pathogen, and *L. ivanovii,* a ruminant pathogen), and (2) four nonpathogenic relatives (*L. innocua*, *L. seeligeri, L. welshimeri,* and *L. grayi*.)[16,17] Only two completed *L. monocytogenes* genome sequences are currently available—*L. monocytogenes* 1/2a EGD-e and 4b F2365[18,19]—but several other strains are being unassembled[18] or sequenced (http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi). Among other species, *L. innocua* CLIP11262[19] and *L. welshimeri* SLCC5334[20] are the only genomes available, but *L. ivanovii* PAM55, *L. seeligeri* SLCC3954, and *L. grayi* CLIP12515 are currently being sequenced.[17] Since the genomes of *L. monocytogenes* 1/2a F6854 and 4b H7858 are unfinished, some genes cannot be properly identified; also, final assembly of these genomic sequences may reveal homologues at a later date. Because no clear conclusion can be drawn from genomic analysis of unfinished genomes,[21] this review will only focus on completed genome sequences of *L. monocytogenes* strains.

## 12.2 PROTEIN SECRETION SYSTEMS

Within the cell envelope, *Listeria* species can exhibit a large variety of proteins; some of them can even interact with the cell surroundings and thus constitute the surfaceome (i.e., the subset of pro-

tein exposed on the bacterial cell surface). It is worth reminding that, on one hand, cell wall is not an impermeable barrier and cell envelope proteins can interact with the environment without ever having a domain that leaves the confine of the cell wall[8] and that the extracellular milieu can penetrate the cell wall, so proteins do not necessarily need to poke out into the environment.[22] On the other hand, protein localization into the cell envelope is no guarantee that it is cell surface exposed *stricto sensu* as proteins can be masked by overlying components such as capsule polymer, for example.[8] Nevertheless, for the purpose of the present review, cell surface proteins will refer to gene products that are attached to the cell wall and/or cytoplasmic membrane and interacting with the external side, whereas cell envelope proteins will refer to all gene products present within the cell wall and/or the cytoplasmic membrane.

While cell surface proteins are systematically cell envelope proteins, the opposite is not necessarily true (e.g., proteins attached to the cytoplasmic membrane but interacting only with the cytoplasm). Still, all cell surface proteins (and most cell envelope proteins) must be first translocated to the cytoplasmic membrane via a protein secretion system before attaching to membrane or cell wall components and thus remaining in contact with the external side. Concerning the functions of cell envelope proteins, they are extremely diverse, ranging from transporters and enzymes involved in various metabolic pathways (such as carbohydrates, proteins, nucleotides, or lipids), signal transductions, adhesion and colonization determinants, to virulence factors. It is worth stressing that among cell surface proteins, some so-called moonlighting proteins can be present.[23] Such proteins are multifunctional in the sense that they conduct enzymatic and/or nonenzymatic activities, sometimes taking part in widely divergent pathways, especially when present at different subcellular locations. For example, enolase, a cytoplasmic protein normally involved in glycolytic pathways, was found on the listerial cell surface, which can bind to human plasminogen.[24]

In Didermata (corresponding to Gram-negative bacteria, also called Negibacteria),[5,25] six major protein secretion systems (numbered from Type I to Type VI, i.e., T1SS to T6SS) are currently recognized and are restricted to these microorganisms.[26–29] In fact, protein secretion systems are categorized primarily by translocation mechanisms across the outermost lipid bilayer, which corresponds to the outer membrane in diderm bacteria but to the cytoplasmic membrane in monoderm prokaryotes. To date in monoderm bacteria, six systems are described as allowing protein secretion[30–33]—that is, protein transport from inside to outside cell cytoplasm—namely, (1) the Sec pathway (secretion, TC #3.A.5; TC#: transport classification number),[34] (2) the Tat pathway (twin-arginine translocation, TC #2.A.64), (3) the FEA (flagella export apparatus, TC #3.A.6.1), (4) the FPE (fimbrilin-protein exporter, TC #3.A.14), (5) the holins (hole-formers, TC#1.E.), and (6) the Wss (WXG100 secretion system, proteins with WXG motif of ~100 residues). To be complete, the MscL family (large conductance mechanosensitive ion channel, TC #1.A.22) and the putative Tad (tight adherence) apparatus could also be added to the list,[35,36] even though experimental evidence is not currently available in monoderm bacteria. Once translocated by one of these systems, a protein can remain associated to the cell envelope, be released into the extracellular milieu, or be translocated into a host cell.

As depicted in Figure 12.1, identification of protein secretion systems in *Listeria* involved screening of genome coding sequences (CDS) against various databases as well as bibliographic analyses. From there, Sec, Tat, FPE, FEA, holins, and Wss were identified in *L. monocytogenes*[37] (Figure 12.2). While some components of these secretion systems have been experimentally investigated, in *Listeria,* protein translocation per se has never been ascertained in any of them yet.

### 12.2.1 Sec System

The presence, remarkable conservation, and essential nature of the Sec translocon in all living cells have given rise to the notion of a general secretory pathway (GSP) but also led to confusing statements in the literature.[38] As illustrated in Figure 12.3, all components of Sec translocon are encoded in *L. monocytogenes*. In addition to the SecYEG protein conducting channel, the signal recognition

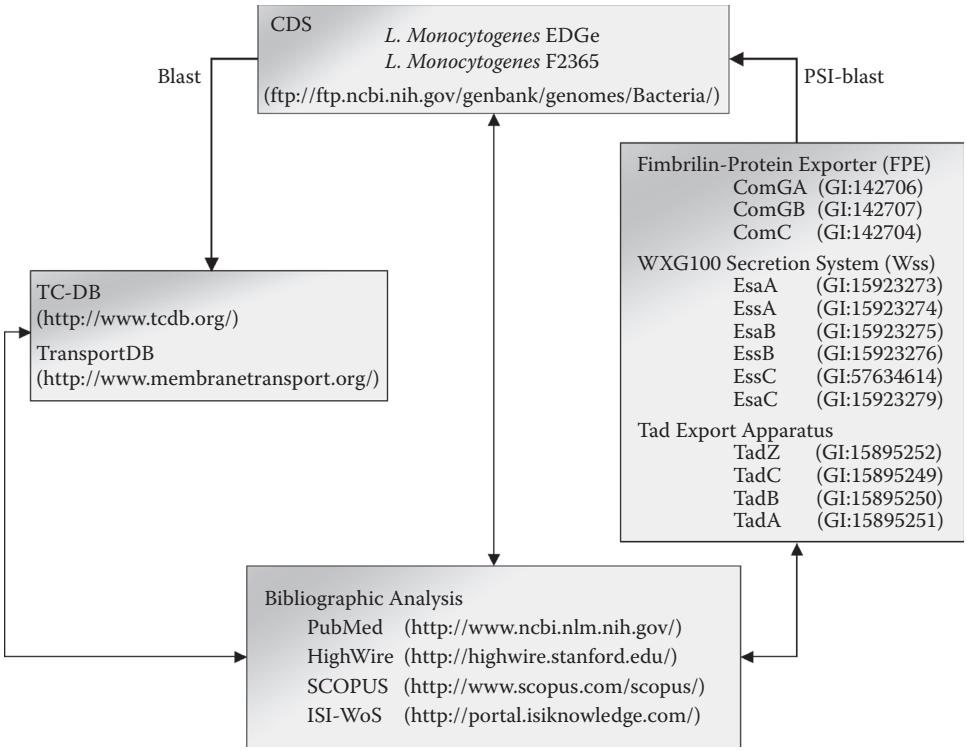


**FIGURE 12.1** Genomic identification of protein secretion systems in *Listeria* species.[37] Prior to bioinformatic analysis, complete genome, coding sequences (CDS), and original annotation data sets were downloaded from GenBank. Each CDS was screened for the capacity to encode a component of a protein secretion system following BLAST against TCDB[146] and TransportDB.[248] These analyses revealed the presence of Sec components and partners as well as FEA subunits, Tat components and holins. MscL and ABC transporter truly implicated in protein secretion could not be identified. The identification of FPE was based on PSI-BLAST searches using GenBank amino acid sequences of ComGA, ComGB, and ComC from *B. subtilis* as queries. Similarly, Wss was identified using EsaA, EssA, EsaB, EssB, EssC, and EsaC from *S. aureus* as amino-acid sequence queries. Using protein sequences of *Clostridium acetobutylicum* as queries,[33] Tad system components could not be identified. Overall, bibliographic analyses were also performed from various databases.

particle (SRP) and the SRP receptor are ubiquitous and essential in all domains of life.[39] In *E. coli*, SRP interacts with nascent signal peptide for cotranslational translocation and specific integration of inner membrane proteins, whereas the targeting factor and chaperone SecB interacts with the mature part of the protein and allows post-translational translocation via Sec.[40] As in all Gram-positive bacteria,[41] SecB and CsaA (analogous to SecB in *B. subtilis*[30]) are absent from *L. monocytogenes*. In *E. coli*, three auxiliary proteins (SecD, SecF, and YajC) form a transmembrane complex loosely associated with SecYEG and increase the overall efficiency of protein translocation through the cytoplasmic membrane.[42]

Contrary to SecDF-YajC, the cytosolic ATPase SecA is essential to Sec-dependent translocation in bacteria as it provides the driving force for stepwise export of the protein.[43] A SecA paralogue (i.e., SecA2) has been identified in several Gram-positive bacteria including *L. monocytogenes*.[44] Contrary to *Streptococcus gordonii,* for example,[45] presence of SecA2 in *L. monocytogenes* is not accompanied by duplication of SecY. While SecA2 is not essential and its relationship with SRP/Sec is unknown, it clearly allows the secretion of a subset of proteins in *L. monocytogenes* (e.g., Iap,[44] NamA,[46] and FbpA[47]). Interestingly, the membrane protein FbpA lacks a putative N-terminal signal peptide. As in *B. subtilis,*[30] two paralogues of YidC could be identified in *L. monocytogenes*: SpoIIIJ and YqjG.[37] In *E coli*, the polytopic membrane protein YidC is necessary for cotranslational

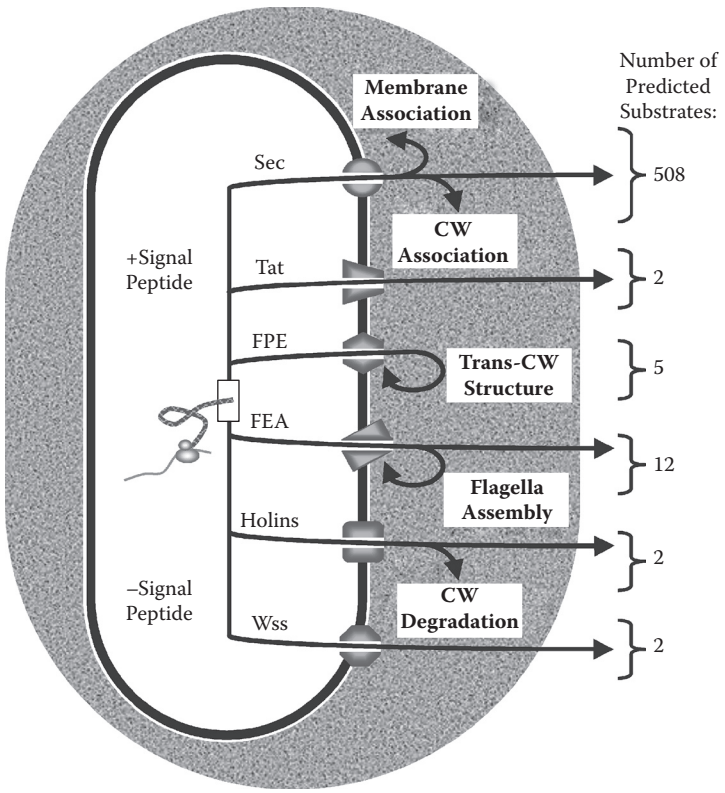


**FIGURE 12.2** Schematic overview of protein secretion pathways in L. monocytogenes EGD-e.37 Proteins to be translocated can exhibit (+) or not (–) an N-terminal signal peptide (with the exception of Sec pathway, which can translocate proteins with or without signal peptide by alternative mechanisms). Proteins translocated via the Sec pathway remain membrane associated or cell wall associated, are released into the extracellular milieu, or would even be injected into an eukaryotic host cell. Proteins exported via Tat would most certainly be cell surfaced or released into the extracellular milieu. FPE would be involved in the formation of transcell-wall structures. FEA is involved in flagella assembly. Proteins exported by holins seem secreted into the extracellular milieu or involved in cell wall degradation. WXG100 proteins would be secreted into the extracellular milieu. The number of translocated proteins by each pathway is given from most recent estimations. CW, cell wall; Sec, secretion; FPE, fimbrilin-protein exporter; Tat, twin-arginine translocation; FEA, flagella export apparatus; Wss, WXG100 (proteins with WXG motif of ~100 amino acyl residues) secretion system.

insertion of all integral membrane proteins (IMPs).[48] YidC is a versatile pathway since it can be Sec-, SecA-, and/or SecB independent. In *B. subtilis*, studies have showed that SpoIIIJ and YqjG play a role in the folding of several secreted proteins and can work independently to insert integral membrane proteins.[49]

Signal peptide of translocated preprotein is cleaved off by a membrane-bound signal peptidase (SPase). Different classes of N-terminal signal peptide are recognized and are cleaved by different types of SPases. Signal peptides of proteins targeted to Sec are of two classes: class 1 and class 2. Class 2 signal peptides are present in lipoproteins and are cleaved off by SPase II (for further details, see section 12.3.1.2). As depicted in Figure 12.3, precursor proteins exhibiting a class 1 signal peptide meet different fates; that is, they can (1) insert in cytoplasmic membrane and thus become integral membrane proteins (for further details, see section 12.3.1.1), (2) remain attached covalently or noncovalently to cell wall components (for further details, see section 12.3.2), (3) be released into the extracellular milieu, or (4) be injected into a eukaryotic host cell via pore formed by Sec-secreted listeriolysin O in a process called cytolysin-mediated translocation (CMT).[50,51] It is worth noting that CMT has never been as yet reported in *Listeria*.

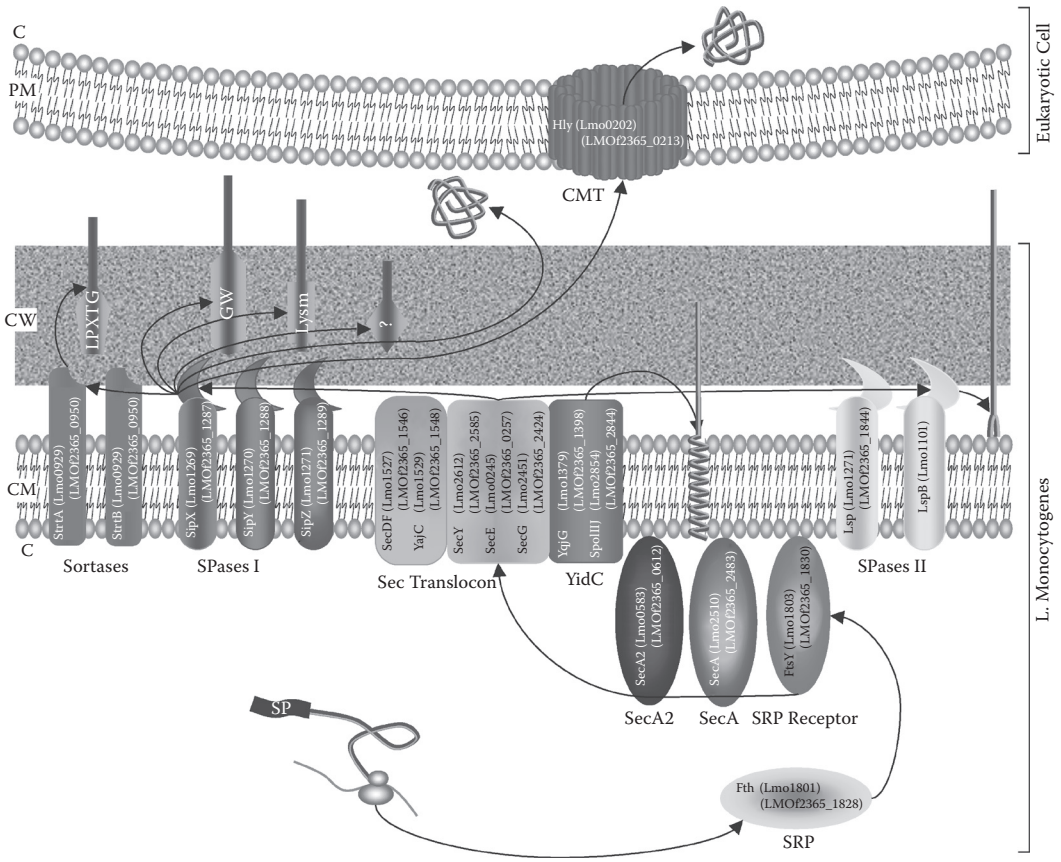


**FIGURE 12.3** Schematic representation of the Sec pathway in *L. monocytogenes*.[37] N-terminal signal peptide is recognized by SRP before cotranslational translocation of the protein through the Sec translocon in a SecA-dependent manner. Some proteins with or without a signal peptide can also be translocated in a SecA2-dependent manner. Integral membrane proteins integrate into the CM via YidC homologues in Sec-dependent or -independent manner; such proteins bear stop-transfer sequence and can exhibit signal peptide or not, which can be cleaved or not. Lipoproteins, which bear signal peptide of class 2 cleavable by SPases II, are covalently attached to long-chain fatty acids of the CM. Proteins bearing class 1 signal peptide cleavable by SPases I are (1) secreted into the extracellular milieu or could even be injected into an eukaryotic host cell following CMT thanks to pores formed by oligomerization of listeriolysin O; (2) bound to CW components via cell binding motifs (i.e., GW, LysM, or uncharacterized motifs); or (3) covalently attached to CW by sortases because of the presence of C-terminal LPXTG motif. C, cytosol; PM, plasmic membrane; CW, cell wall; EM, extracellular milieu; CM, cytoplasmic membrane; SP, signal peptide; SPase, signal peptidase; SRP, signal recognition particle; CMT, cytolysin mediated translocation.

In Gram-positive bacteria, some Sec-dependent signal peptides exhibit a YSIRK motif (PF04650) present at the beginning of the H-domain. This motif is required for efficient protein secretion and is systematically associated with an LPXTG motif, even though the opposite is not true. Class 1 signal peptides are not always cleaved as the H-domain can serve of transmembrane anchor domain as observed in SPases I. Three SPases I have been uncovered and characterized in *L. monocytogenes*: SipX, SipY, and SipZ.[52,53] Deletion of *sipY* genes had no detectable effect, whereas SipX and SipZ had overlapping substrate specificity.[52] *lsp* was demonstrated as encoding a genuine SPase II[54] and a second SPase II—LspB (Lmo1101)—was recently uncovered by genomic analysis but only in *L. monocytogenes* EGD-e.[37]

While some proteins cleaved by SPases I can remain noncovalently bound by various cell wall binding domains (for further details, see section 12.3.2.2), covalent attachment of proteins to

cell wall requires sortases. Proteins emerging from the Sec apparatus and exhibiting an LPXTG-like motif C-terminally located (for further details, see section 12.3.2.1) are recognized by membrane-associated sortase.[55] Transpeptidase sortase attacks the TG bond of the LPXTG-like motif, capturing cleaved polypeptide as a thioester-linked acyl enzyme at its active site cystein residue.[56] Subsequently, this complex is resolved by the nucleophilic attack of the amino group of the cross-bridge within lipid II precursor. Based on phylogenetic analyses, sortases are now classified into four classes, designated A, B, C, and D.[57] In *L. monocytogenes*, two sortases are present (SrtA and SrtB; Figure 12.3).

As observed in other Gram-positive bacteria, sortase of class A (also called SrtA subfamily) in *L. monocytogenes* is encoded only once in the genome, resembles a Type II membrane protein, and is necessary for the anchoring of the majority of LPXTG-containing proteins.[58] Sortase of class B (SrtB subfamily) recognizes a particular type of sorting signal (i.e., an NXZTN motif), which suggests a lower stringency of the recognition motif of SrtB compared to SrtA.[59] Captivatingly, from investigations in *Streptococcus pyogenes* and *Staphylococcus aureus*, glycosylated LPXTGase, an enzyme that cleaves the C-terminal LPXTG motif, is the first enzyme found that is produced by nonribosomal peptide (NRP) synthesis.[60,61] It is known that NRP synthesis (and similarly related polyketide synthesis) occurs in Bacilli class, where NRPs are assembled in the cytoplasm by large megaproteins called NRP synthetases consisting of a series of active modules carrying out catalysis and modification of the tethered growing peptide chain.[62] However, investigations in *S. aureus* suggest that enzymes responsible for cell wall assembly may also be involved in the construction of LPXTGase.[61] Finally, it cannot be excluded that such a nonribosomally synthesized enzyme be also present and involved in LPXTG-like protein anchoring in *L. monocytogenes*.[37]

Substrates of the Sec system are generally considered as exhibiting an N-terminal signal peptide composed of three domains: (1) The N-domain contains positively charged amino terminus, (2) the H-domain is a hydrophobic core region, and (3) the C-domain contains the cleavage site.[63] It must be emphasized, however, that it is not the case for all proteins (e.g., some SecA2-dependent and/or YidC-dependent proteins). Still, the presence of an N-terminal signal peptide indicates a protein is targeted to membrane. Despite lack of amino acid sequence similarity, signal peptides can be detected with good accuracy by various documented and publicly available applications (Table 12.1). The first methods developed were SigCleave and SPScan, which were implementations of a simple weight matrix approach.[64] While SigCleave is part of the EMBOSS suite and also available by an interface on the World Wide Web, SPScan is only available as part of the GCG suite and thus requires ability to work under Unix-like environment. Comparing the two programs, SPScan has clearly better predictive performance in terms of secretory protein and cleavage site recognition, especially for prokaryotic proteins.[65] Nearly a decade later, SignalP, a promising method based on a

**TABLE 12.1**
**Bioinformatic Resources for Prediction of Bacterial N-Terminal Signal Peptides**

| Application | Method | Webserver | Ref. |
|---|---|---|---|
| SigCleave | Position weight matrix | http://bioweb.pasteur.fr/seqanal/interfaces/sigcleave.html | 64 |
| SPScan | Position weight matrix | none | 64 |
| SignalP | Neural network<br>Hidden Markov model | http://www.cbs.dtu.dk/services/SignalP/ | 67 |
| PrediSi | Position weight matrix | http://www.predisi.de/ | 69 |
| SOSUIsignal | Global physicochemical analysis | http://bp.nuap.nagoya-u.ac.jp/sosui/sosuisignal/ | 70 |
| Phobius | Hidden Markov model | http://phobius.binf.ku.dk/ | 71 |
| PSORTb | Support vector machine<br>Hidden Markov model | http://www.psort.org/psortb/ | 74 |
| SPdb | BLAST | http://proline.bic.nus.edu.sg/spdb/ | 76 |

neural network, was released[66] and has undoubtedly become the most popular method for predicting N-terminal signal peptide. Since the first available version 1.1, SignalP has been substantially improved up to the latest version 3.0.[67]

While version 1.1 is definitively out of date, both versions 2.0 and 3.0 use either a neural network (NN) or HMM. When comparing SignalP v2.0-NN, -HMM, and SPScan, it appears that (1) SPScan predicts correctly more proteins as secreted than SignalP v2.0-NN or -HMM; (2) SignalP v2.0-NN and -HMM are superior in predicting the correct cleavage site; (3) SignalP v2.0-NN lags behind SPScan and SignalP v2.0-HMM in classifying correctly the proteins, the latter providing the best prediction; and (4) SignalP v2.0-NN is the best for predicting of the correct cleavage site.[65] In other words, these methods are complementary in predicting an N-terminal signal peptide. The main improvement in SignalP v3.0 is increased accuracy in prediction of signal peptidase cleavage sites.[67] In comparative analyses, SignalP3.0 performs significantly better than other machine learning and HMM methods. Despite performance improvement in the latest SignalP v3.0, however, it appears that SignalP v2.0-NN remains the best signal prediction program.[68]

A position weight matrix approach was improved by a frequency correction, which takes into consideration the amino acid bias (i.e., PrediSi).[69] SOSUIsignal is a global structure analysis based on physicochemical features of the three signal peptide domains—N-, H-, and C-domains—and discriminates between cleavable and anchoring signal sequences.[70] Since a signal peptide contains a hydrophobic H-domain, there is a risk of erroneously identifying a transmembrane α-helix as a signal peptide or, conversely, classifying a protein with a signal peptide H-domain region as an IMP. In order to discriminate between the two, a combined TM topology and signal peptide predictor has been developed: Phobius.[71] Phobius significantly reduces false classifications of signal peptides compared to SignalP. Another machine learning approach used for prediction of signal peptides is support vector machine (SVM), which can predict signal peptides with great accuracy.[72] Such an implementation of an SVM combined with an HMM is part of PSORTb,[73] now applicable to both Gram-positive and Gram-negative bacteria.[74] Finally, SPdb, a repository of experimentally determined and computationally predicted signal peptides, is also accessible via BLAST (basic local alignment search tool) search.[75,76]

It can be stressed again that these analyses only predict the presence of signal peptide, meaning that the protein is targeted to the cytoplasmic membrane. However, it does not necessarily mean the protein is translocated across the cytoplasmic membrane via Sec or released into the extracellular milieu. Indeed, proteins translocated via Tat or FPE also possess N-terminal signal peptides with additional features, which are not identified by the previous tools (Table 12.1). Thus, final prediction of a protein translocated via Sec requires additional inspections (see sections 12.2.2 and 12.2.3). Concerning proteins translocated by the Sec system and possessing a signal peptide, they can (1) be released into the extracellular medium or injected into a host cell, (2) remain associated to the cell wall by covalent or noncovalent interactions, or (3) remain associated to the cytoplasmic membrane by transmembrane domains (including H-domain of uncleaved signal peptide) or be lipoproteins (see section 12.3). Thus, final localization prediction of Sec substrates requires a combination of tools for prediction of function, motifs, and TMDs. It is also recommended to combine these results with those from tools dedicated to prediction of protein subcellular localization in Gram-positive bacteria (Table 12.2).

NNPSL was the first tool developed for such prediction and is based on an NN.[77] SubLoc,[78] PSORTb,[73] CELLO[79] (recently extended to prediction in Gram-positive bacteria[80]), and LOCtree[81] are basically SVM. These tools have their own advantages and weaknesses,[82] and some of them, like PSORTb, combine a variety of individual predictors. Proteome Analyst is a novel type of machine-learning classifier that involves several steps in the prediction process, such as BLAST search against Swiss-Prot database and naïve Bayesian classifiers.[83] From the most recent studies on performance of prediction tools, PSORTb and Proteome Analyst achieve the highest overall precision.[84] Gpos-PLoc, another type of ensemble classifier, was recently developed where several basic

**TABLE 12.2**
**Bioinformatic Resources for Prediction of Subcellular Localization of Proteins in Gram-Positive Bacteria**

| Application | Method | Webserver | Ref. |
|---|---|---|---|
| NNPSL | Neural network | http://www.doe-mbi.ucla.edu/~astrid/astrid.html | 77 |
| SubLoc | Support vector machine | http://www.bioinfo.tsinghua.edu.cn/SubLoc/ | 78 |
| PSORTb | Support vector machine<br>Ensemble classifier | http://www.psort.org/psortb/ | 73 |
| CELLO | Support vector machine | http://cello.life.nctu.edu.tw/ | 80 |
| LOCtree | Support vector machine | http://cubic.bioc.columbia.edu/services/loctree/ | 81 |
| Proteome Analyst | Ensemble classifier | http://pa.cs.ualberta.ca:8080/pa/ | 83 |
| Gpos-PLoc | Ensemble classifier | http://202.120.37.186/bioinf/Gpos/ | 85 |
| DBSubLoc | BLAST | http://www.bioinfo.tsinghua.edu.cn/~guotao/intro.html | 86 |
| PSORTdb | BLAST | http://db.psort.org/ | 87 |
| PA-GOSUB | BLAST | http://www.cs.ualberta.ca/~bioinfo/PA/GOSUB/ | 88 |
| Augur | Ensemble classifier | http://bioinfo.mikrobio.med.uni-giessen.de/augur/ | 80 |

classifiers were fused and optimized for predicting subcellular localization of Gram-positive bacterial proteins.[85] Finally, several databases (derived from previously described prediction tools) are available following BLAST search (DBSubLoc,[86] PSORTdb,[87] and PA-GOSUB[88]). Augur is another database especially dedicated to protein localization on cell surface of Gram-positive bacteria.[89] Once again, final prediction of secreted proteins (and localization) should combine results from these various bioinformatic tools.[84]

Using SignalP v2.0 to predict signal peptide region and TopPred v2.0 to exclude other transmembrane domains, 86 proteins were predicted as secreted into the extracellular medium from genomic analysis of *L. monocytogenes* EGD-e.[19] In *L. monocytogenes* F2365, 420 proteins were predicted with a putative N-terminal signal peptide, including 2 with a YSIRK motif.[18] Performing extensive genomic analyses, which combined results from SignalP v2.0, SigCleave, SOSUI, PSORT, and TMPinGS, the number of proteins bearing an N-terminal signal peptide was estimated at 525 in *L. monocytogenes* EGD-e, including 255 IMPs and 270 exported proteins where 121 would be released into the extracellular milieu.[90] All 14 virulence factors characterized so far in *L. monocytogenes* are most likely translocated via the Sec translocon.[37] Among the 121 proteins originally predicted as secreted via Sec and released into the extracellular milieu, a closer look revealed that four prepilins—that is, ComGC (Lmo1345), ComGD (Lmo1344), ComGE (Lmo1343), and ComGG (Lmo1341)—should be removed from the output since they would form trans-cell-wall structure following translocation via FPE.[37] Proteomic analysis of supernatant from liquid culture of *L. monocytogenes* EGD-e allowed the identification of 54 out of 117 proteins predicted as extracellular, including virulence factors Hly, PlcA, and PlcB.[90]

### 12.2.2 Tat Pathway

The term twin-arginine translocation (Tat) was coined from the systematic presence of RR motif in signal peptide of proteins translocated via this secretion system.[91] The [ST]RRXFLK motif straddles the N-domain and H-domain of N-terminal signal peptide.[92] Contrary to the Sec translocon, the main feature of this pathway is its ability to translocate proteins in a folded state. General knowledge on the precise succession and mechanistic events leading to protein secretion via this pathway remains rudimentary.[93] The generally accepted translocation model was first proposed by Mori and Cline,[94] where Tat translocation follows a cycle in which TatBC functions in the specific recognition of the substrate and TatA functions as the pore-forming unit. An alternative model proposes

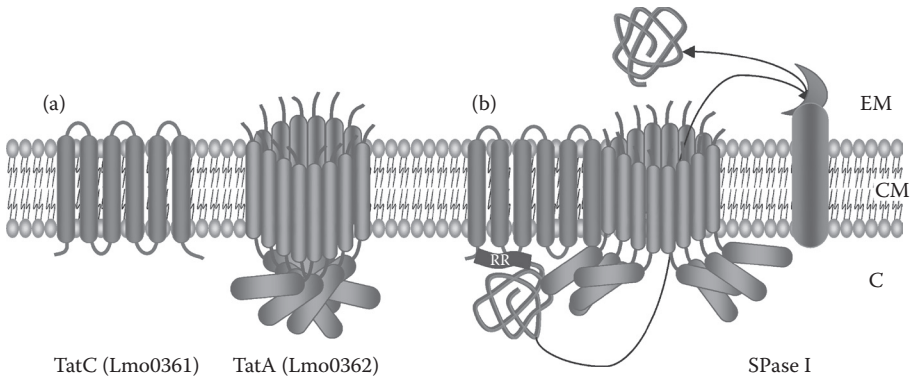


**FIGURE 12.4** Tat translocon in *L. monocytogenes*.[37] During Tat secretion, the general model proposes a cyclical assembly of components. (A) In resting state, Tat machinery components are separately present in the cytoplasmic membrane (i.e., TatB and TatA). (B) Once Tat substrate protein precursor binds to the TatC in an energy-independent step, this complex associates with TatA in a step driven by transmembrane proton electrochemical gradient. This association would persist until completion of protein transport across the membrane driven by proton motive force. Tat signal peptide is subsequently cleaved by SPase I and Tat machinery components disassembled, as depicted in (A). RR, twin-arginine motif in Tat signal peptide; EM, extracellular milieu; CM, cytoplasmic membrane; C, cytoplasm.

that membrane integration could precede Tat-dependent translocation and the membrane targeting process may require ATP-dependent N-terminal unfolding-steps energy.[95]

Still, components of Tat translocon differ in number between Gram-negative and Gram-positive bacteria.[30] The most baffling difference is the absence of TatB from all Gram-positive bacteria sequenced so far, although it is an essential components of Tat translocon in *E. coli*, which is used as a paradigm.[96] As in most Gram-positive bacteria, Tat translocon in *L. monocytogenes* is encoded in one locus and is composed of only two proteins, TatA and TatC[37] (Figure 12.4). TatC is a large IMP generally considered as the primary site for signal-peptide recognition.[97] TatA is a membrane protein that oligomerizes to form a protein-conducting channel where the number of subunits would adjust in function of the Tat substrate size.[98] In TatA, a cytoplasmic lid region acts as a gate and would open following association of TatC–substrate complex with TatA, then inducing conformation change and protein translocation driven by proton motive force. Translocated protein is finally released after cleavage by SPAse I.[96] A Tat translocon does not seem to be systematically present in *L. monocytogenes* as no component could be identified in *L. monocytogenes* F2365. The Tat system has never been experimentally investigated in *Listeria*; thus, its expression, functionality, involvement of one or three SPases I, or proteins secreted via this pathway remain unknown.[37]

Three tools are currently available to discern Tat substrates (Table 12.3) and TATFIND was the first program especially devoted to such identification.[99] In its original available version, TATFIND v1.2, prediction was based on two criteria: (1) presence of conserved Tat motif ZRRZZZ within the first 35 amino acid residues, where Z represents a defined set of permitted residues; and (2) presence

**TABLE 12.3**
**Bioinformatic Resources for Prediction of Tat Signal Peptides**

| Application | Method | Webserver | Ref. |
|---|---|---|---|
| TATFIND | Physicochemical analysis and regular expression | http://signalfind.org/tatfind.html | 100 |
| TatP | Neural network and regular expression | http://www.cbs.dtu.dk/services/TatP/ | 101 |
| TATPred | Naïve Bayesian network | http://www.jenner.ac.uk/logP/JennerlogPcalc.htm | 102 |

of an uncharged stretch of at least 13 amino acids downstream of the twin arginine. In the latest version, TATFIND v1.4, search for a single charged residue in positions +2 and +5 relative to the RR was included.[100] TatP v1.0 incorporates signal peptide and cleavage site prediction based on a combination of two artificial neural networks followed by a postfiltering of the output based on regular expression RR[FGAVML][LITMVF].[101] Compared to TATFIND v1.2, TatP generates far fewer false positive but slightly more false negative predictions. TATPred is the latest algorithm based on naïve Bayesian network developed for prediction of Tat substrates.[102] Compared to TatP, TATPred appears as the most robust and reliable predictor with higher sensitivity of prediction.

According to TATFIND search, only two Tat substrates could be identified in *L. monocytogenes* EGD-e.[99] One of these putative Tat substrates, however, is also present in *L. monocytogenes* F2365, where the Tat system is not encoded.[37] These substrates have never been reported as present in the extracellular milieu of *L. monocytogenes*. While it has been long assumed that the RR motif was highly specific and conserved in Tat substrates, it must be stressed that substitutions of one arginine, or in some cases both arginines, by lysine[103] or that natural proteins harboring very distantly related RR motifs[104] could still permit targeting and translocation via the Tat pathway.[105] This indicates that Tat system specificity is more flexible than originally thought and thus presence of Tat substrate cannot be systematically identified by bioinformatic analysis.

### 12.2.3 FPE

Components of fimbrilin-protein exporter (FPE) of Gram-positive bacteria are homologous to proteins required for secretion of substrate proteins in Gram-negative bacteria, namely, some ATPase and IMP components of the Type II protein secretion system (T2SS), Type 4 piliation sytem (Tfp), and Type IV protein secretion system (T4SS), as well as archaeal flagella.[106] As in all Gram-positive bacteria where it has been reported so far,[33,107,108] components of FPE in *L. monocytogenes* are encoded in a *comG* operon, except for *ComC* located elsewhere on the chromosome. Protein exporters of the FPE family consist of two constituents—ComGA and ComGB—that would function together in an ATP-hydrolysis-dependent export of proteins across the cytoplasmic membrane[109,110] (Figure 12.5). ComGA is an ATPase localized to the cytoplasmic side of the membrane that could participate in modeling of pilus-like structure.[109] As a homologue to PilC of Tfp and PulF of T2SS,[109] ComGB is an IMP having three putative TMDs that could play the role of a protein-conducting channel.[111] ComC is a Type 4 prepilin peptidase involved in cleavage of N-terminal signal peptide of class 3;[112] this signal peptidase belongs to the aspartic acid protease family.[113]

While ComC is required for maturation, translocation, and assembly of prepilins, an initial translocation event across the cytoplasmic membrane has not been clearly elucidated. As prepilin signal peptide is cleaved at the cytoplasmic side between the N- and H-domains, prepilins are certainly not translocated by the Sec or Tat pathways and the hypothesis of ComGAB involvement is favored. However, YidC contribution cannot be excluded[30]; prepilins were originally thought to insert spontaneously in the membrane bilayer but with the current knowledge of membrane protein insertion this hypothesis should not be privileged (see section 12.3). Four Type 4 prepilins are encoded in *comG* locus by the *comGC*, *comGD*, *comGE,* and *comGG* genes[110]; ComGF is presumably an IMP. Once maturated and translocated, pilins form a trans-cell-wall macromolecular complex where monomers are covalently linked by disulphide bonds.[114] Since this structure is involved in bacterial competence and does not form a proper Type 4 pilus, it was named competence pseudopilus.

In *B. subtilis*, Type 4 prepilins exhibit N-terminal signal peptides with a conserved motif [KR]G▼F[TSI][LTY][VLIP][EA] located between the N- and H-domains where ▼ indicates the predicted cleavage site.[110] In *Listeria*, the motif is slightly different—that is, [NPRS][GA]▼F[TS]L[VLP][EF]—and is found in five putative prepilins (i.e., ComGC, ComGD, ComGE, ComGF, and ComGG).[37] In *B. subtilis*, the highly conserved phenylalanine at position +1 is aminomethylated by ComC, which thus appears bifunctional as it is also involved in prepilin processing.[30] Using ScanProsite syntax,[115] search for consensus motif [GA]F[TS]LX[EF] located between the N- and

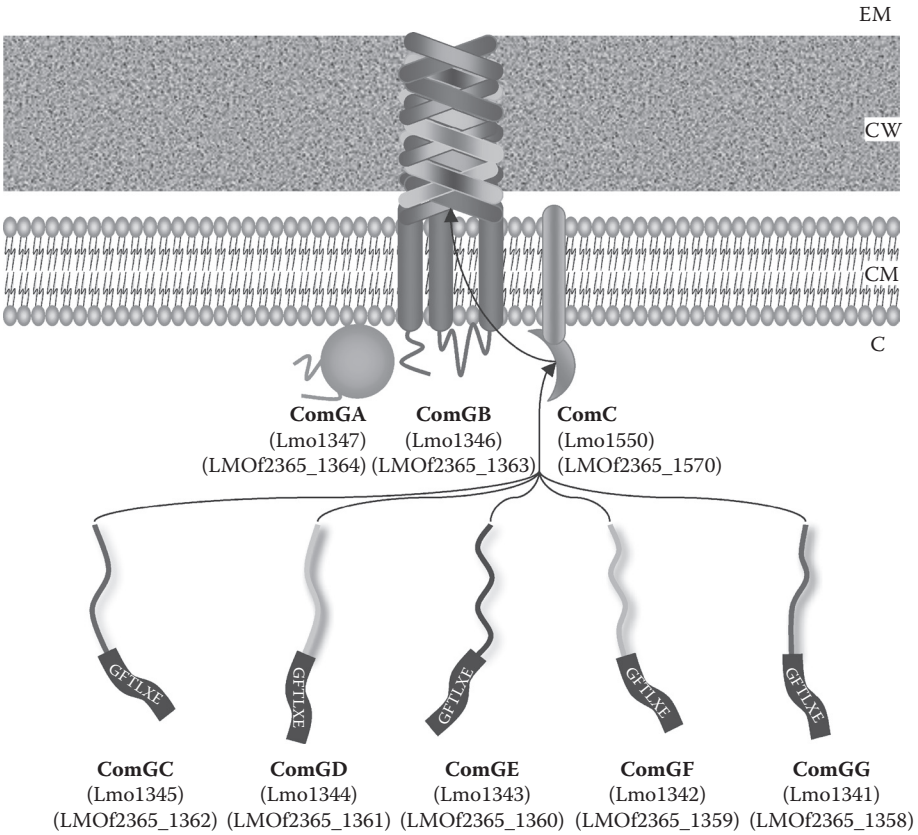


**FIGURE 12.5** FPE in *L. monocytogenes*.[37] The prepilins initially float in the CM; initial insertion into the membrane is certainly Sec or Tat independent but the involvement of ComGAB remains to be ascertained. After processing by ComC signal peptidase, ComGA and ComGB would be involved in assembly of pseudopilins to form a trans-cell-wall pilus-like structure. GFTLXE, conserved [GA]F[TS]LX[EF] motif in Type 4 prepilin signal peptide from *Listeria*; CW, cell wall; EM, extracellular milieu; CM, cytoplasmic membrane; C, cytoplasm.

H-domains of predicted signal peptide can thus be performed in order to identify putative FPE substrates in *Listeria*. The FPE system has never been experimentally investigated in *Listeria*; thus, its expression, functionality, and involvement in bacterial competence remain to be established.[108]

### 12.2.4 FEA

*L. monocytogenes* produces up to six peritricheous flagella, which are down-regulated at 37°C, although variation from one strain to another was reported.[116,117] Regulation of listerial flagella is not entirely understood and appears rather complex since at least five regulators involved in its expression have been identified so far: FlaR,[118] PrfA,[119] DegU,[120] MogR,[121] and GmaR (Lmo0688 also called WcaA).[122] Interestingly, the antirepressor GmaR is bifunctional since it also functions as a glycosyltransferase for flagellin FlaA[122] and glycosylation with β-O-linked N-acetylglucosamine was indeed established for FlaA.[123] This investigation constituted the first description of β-O-GlcNac post-translational modification on a prokaryotic protein, though flagella glycosylation is not essential for motility in *L. monocytogenes*.[124] As motility mediators, flagella are important in colonization of abiotic surfaces and host cell invasion but do not function as adhesins.[124,125]

Interestingly, FlaA was also demonstrated as exhibiting a peptidoglycan-hydrolyzing activity that might play a role during flagella assembly.[126,127] Indeed, some flagellar components are assembled

on the bacterial cell surface where local digestion of cell wall sacculus might be required—namely, for (1) the rod proteins (i.e., FlgB, FlgC, FliE, and FlgG), (2) the hook/junction proteins (i.e., FliK, FlgD, FlgE, FlgK, and FlgL), and (3) the filaments proteins FlaA and FliD. As in Gram-negative bacteria,[128–132] these proteins lack a cleavable N-terminal signal peptide and are presumably translocated by the flagella export apparatus (FEA) composed of FlhA, FlhB, FliH, FliL, FliP, FliO, and FliH. In *Listeria*, all flagella components are encoded in a single flagellar–motility–chemotaxis cluster of 41 genes,[37] where FEA and its potential substrates could be identified by homology search. In Didermata, T3SS refers to a secretion system where translocation apparatus is homologous to injectisomes (T3aSS) and flagella (T3bSS),[51,133] both of which are involved in secretion of extracellular proteins.[134,135] As already stressed, however, this terminology is restricted to Gram-negative bacteria. In monoderm bacteria, involvement of FEA in secretion of extracellular protein has only been suggested in *Bacillus thuringensis*.[136]

### 12.2.5 Holins

Holins (hole-formers) are small membrane proteins of phage origin that essentially control endolysin function in a process leading to bacterial apoptosis.[137–139] A current model for the holin-endolysin system proposes that holins accumulate in the cytoplasmic membrane, whereas endolysins accumulate in the cytoplasm[140,141] (Figure 12.6). At a programmed time, holins oligomerized to form pores in the cytoplasmic membrane, allowing release of endolysins into the extracytoplasmic space leading to cell lysis following cell wall degradation and membrane disruption. Homo-oligomeric pore complexes formed by holins would provide a passive but specific translocation system.[142] Generally, holin and its specific endolysin are genetically encoded in tandem. Some holin genes possess a dual start motif, which results in the expression of two distinct proteins with dramatically opposed function since one would promote autolysis (holin) and the other would inhibit it (antiholin).[143] Such regulation can also occur between proteins encoded at different loci (e.g., *lrgAB*/*cidAB* operons in *Staphylococcus aureus*).[144,145] Holins are an extremely diverse group of proteins with 23 distinct families recognized in TC-DB (transport classification database),[146] although they can be grouped into three classes based on membrane topologies.[147] Class 1 holins exhibit three helical TMDs, whereas class 2 holins have two TMDs. Besides classes 1 and 2, which cover most holins, a third class was identified on the basis of T protein of phage T4 where only a single TMD is present.[148]

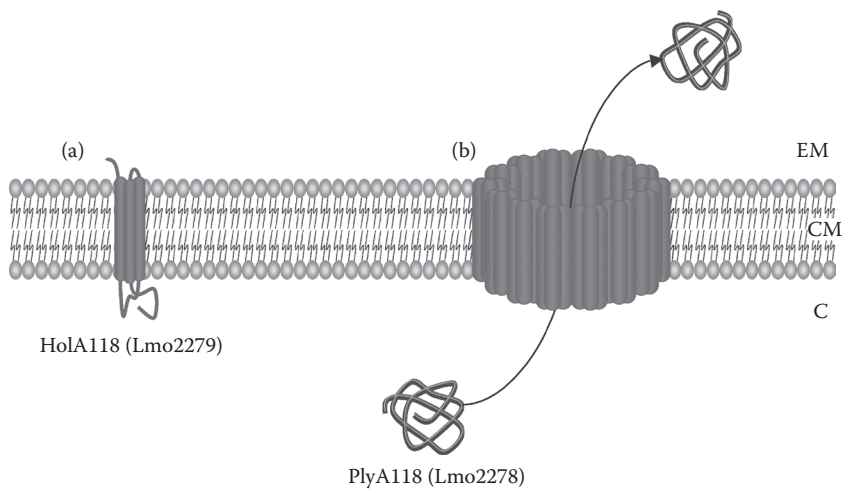


**FIGURE 12.6** ϕA118 holin-endolysin system in *L. monocytogenes* EGD-e.[37] (A) HolA118 is a class 1 holin (i.e., with 3 TMDs). (B) HolA118 oligomerizes in the CM to form a pore allowing translocation and activation of endolysin Ply118. EM, extracellular milieu; CM, cytoplasmic membrane; C, cytoplasm.

The number of holins encoded in *Listeria* varies between strains and species[37]; only one holin could be identified in *L. monocytogenes* F2365, whereas five holins were found encoded in nonpathogenic strain *L. innocua* CLIP11262. However, only three distinct families of holins were identified in *L. monocytogenes* (i.e., as belonging to bacteriophage 118, TcdE, and bacteriophage 11 families). Holins of ϕA118 family (HolA118) were first identified and investigated in *L. monocytogenes* although a homologue is at least also encoded in *L. innocua*.[37,149] HolA118 is not encoded by all *L. monocytogenes* species as it is absent from *L. monocytogenes* F2365. Native holin HolA118 is a 93-amino-acid-long protein belonging to class 1, but its encoding gene is subjected to dual translational initiation, which leads to a second 83-amino-acid-long protein called HolA118(83) acting as an antiholin.[150] Gene encoding phage lysin of ϕA118 (PlyA118) systematically clusters with gene encoding HolA118.[18,151]The endolysin PlyA118 is an L-alanoyl-D-glutamate peptidase hydrolyzing the cross-linking bridges of cell wall peptidoglycan and thus responsible for bacterial lysis in a programmed cell death.

Holins belonging to TcdE family are encoded in all sequenced *Listeria* but as ϕ11 holins they have never been experimentally investigated. TcdE holin was investigated in *Clostridium difficile,* where toxigenic strains produce two large toxins, TcdA and TcdB, of major importance in bacterial virulence, which would be translocated across the cytoplasmic membrane by TcdE.[152,153] While in *C. difficile* all these genes are encoded within a pathogenicity locus, no genes coding for toxins or virulence factors could be identified in *Listeria*.[37] However, a putative autolysin lacking a signal peptide was systematically present (i.e., genes encoding Lmo0129 and LMOf2365_0147 in *L. monocytogenes* EGD-e and F2365, respectively). Although this particular holin family has never been investigated per se in *Listeria*, proteomic analysis in *L. monocytogenes* EGD-e disclosed the presence of Lmo0129 in supernatant of bacterial cultures, suggesting this secretion pathway is active in this species.[90] In *Listeria*, ϕ11 holins were only identified in unassembled genome of *L. monocytogenes* F6854 and nonpathogenic *L. innocua* CLIP11262, where they systematically clustered with genes encoding amidases presumably involved in cell wall degradation.[37]

### 12.2.6 Wss

Wss stands for proteins with WXG motif of ~100 residues (WXG100) secretion system.[154] WXG100 is a new superfamily of proteins around 100 amino acids long, possessing a coil–coil domain and bearing a conserved WXG motif located in the middle region. First identified members of this superfamily were paralogues ESAT-6 (early secreted antigen target of 6 kDa) and CFP-10 (culture filtrate protein 10) from *Mycobacterium tuberculosis*. While ESAT-6 and CFP-10 are specific and experimentally investigated proteins, WXG100 (PF06013) is an established and generic terminology more appropriate to describe protein members of this family, especially those that have not been experimentally investigated yet. No generic terminology for the different components of Wss apparatus has been established yet. Presence of a novel protein secretion system was clearly suggested by bioinformatic analysis.[154] In *B. subtilis*, genes encoding WXG100 proteins appeared to cluster systematically with *yukab*, which are predicted to encode membrane bound ATPases with FtsK/SpoIIIE domains. Similar genetic organization was observed in some *Corynebacterium, Mycobacterium, Streptomyces, Bacillus, Clostridium, Listeria,* and *Staphylococcus* species.[31,154] YukAB homologues appear encoded as single or two CDS. To date, Wss seems phylogenetically restricted to Gram-positive bacteria and has been experimentally investigated only in *M. tuberculosis*, *M. smegmatis,* and *S. aureus*.

In *Mycobacterium*, two WXG100 proteins are secreted: ESAT-6 and CFP-10.[155] Recently, a C-terminal signal sequence required for secretion via Wss was unraveled in CFP-10.[156] *Mycobacterium* Wss apparatus was named Snm (secretion in mycobacteria) and is composed at least of[155,157]: (1) Snm1, Snm2, and Snm6 containing NTP-binding motifs (where Snm1 and Snm2 are homologous to YukAB); (2) Snm4, which is an IMP; (3) Snm5 and Snm7 with uncharacterized functions,

and Snm8 (i.e., a membrane anchored serine protease). Snm permits translocation of ESAT-6 and CFP-10 as well as their heterodimerization.[158,159] In *S. aureus*, the Wss was named Ess (ESAT-6 secretion system) and is encoded in a locus composed of eight CDS including two WXG100 paralogues—EsxA (Ess extracellular protein A) and EsxB—as well as:[160] (1) EssC (Ess protein C) homologous to YukAB; (2) EssA, EssB, and EsaA (ESAT-6 secretion accessory protein A), which are IMPs; and (3) EsaB and EsaC, which predict cytoplasmic chaperones. Compared to *B. subtilis,* where a putative Wss was primarily uncovered, EssB and EsaB appear homologous to YukC and YukD, respectively. In *S. aureus*, no homologue to Snm4, Snm5, Snm6, Snm7, or Snm8 was found, whereas in mycobacteria, no homologue to EssA, YukC, EsaA, and YukD could be identified. In both *M. tuberculosis* and *S. aureus,*[160,161] Wss is important and critical for bacterial pathogenicity, though the function of WXG100 proteins in virulence remains obscure.[162]

Synteny is highly conserved between Wss encoding loci of *S. aureus* and *L. monocytogenes.*[160] However, compared to *Mycobacterium* species or *S. aureus*, only a single copy of Wss locus is present in each sequenced *Listeria* genome.[37] Following homology with *S. aureus* and *Mycobacterium*, Wss in *L. monocytogenes* is represented in Figure 12.7. From one report,[163] it seems that WXG100 protein is not required for virulence of *L. monocytogenes*. Still, protein expression, system functionality, and involvement in bacterial virulence of Wss remain to be established.

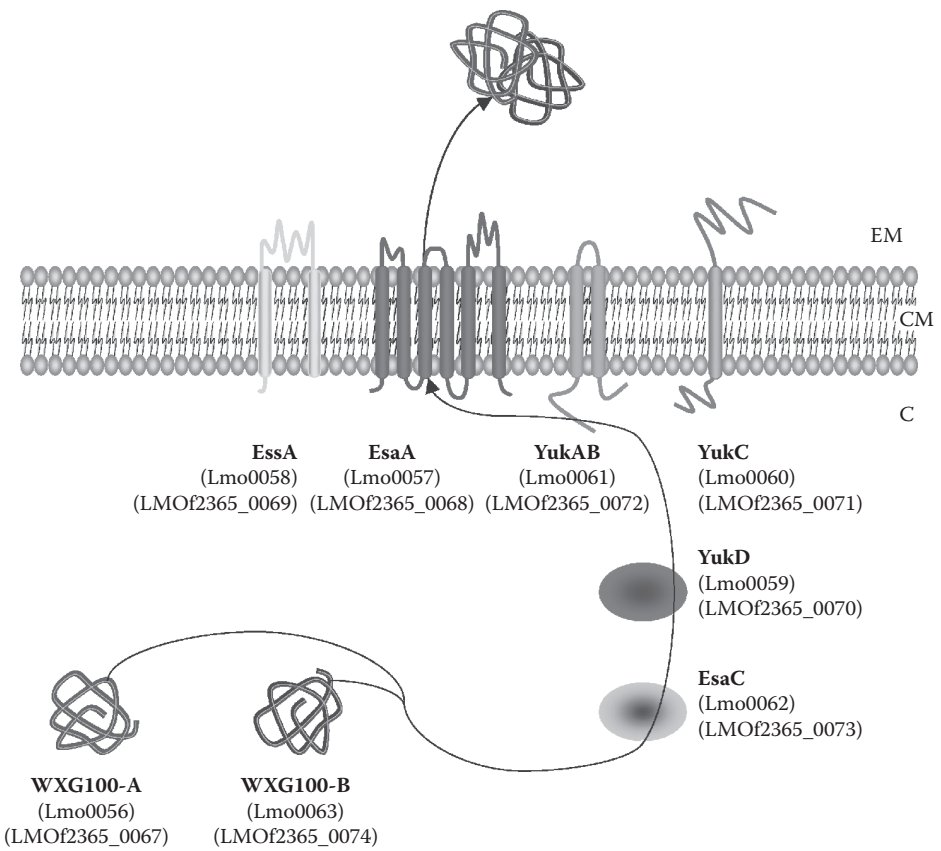


**FIGURE 12.7** Wss in *L. monocytogenes*.[37] WXG100 proteins would interact with putative cytoplasmic chaperones YukD and EsaC before being translocated by the Wss apparatus constituted of EssA, EsaA, YukC, and YukAB; in the course of translocation the two WXG100 proteins would finally form a heterodimer. EM, extracellular milieu; CM, cytoplasmic membrane; C, cytoplasm.

## 12.3 CELL ENVELOPE-ASSOCIATED PROTEINS

Cell envelope of Gram-positive bacteria is primarily composed of a single biological membrane (i.e., the cytoplasmic membrane) and a cell wall made up of peptidoglycan (which in turn consists of linear polysaccharide chains cross-linked by short peptides).[164] Besides peptidoglycan, the rigid cell-wall of Gram-positive bacteria contains large amounts of wall-associated polymers, also called "secondary" cell wall polymers (SCWPs), which can be classified into three distinct groups: (1) teichoic acids (i.e., polyol phosphate polymers, including lipoteichoic acids), (2) teichuronic acids, and (3) other neutral or acidic polysaccharides that cannot be assigned to the two former groups (e.g., lipoglycans).[55,165,166] The SCWPs, present in various proportions, are either covalently linked to the peptidoglycan backbone (i.e., teichoic acids) or tethered to a lipid anchor moiety. Except for teichoic and teichuronic acids, the structure and biosynthesis of other SCWPs are largely unknown. It must be stressed that in almost all phylogenetic branches of Archaea and Bacteria, the cell envelope is also constituted of a proteinaceous S-layer (regular crystalline surface layer), which forms the outermost cell-wall layer.[167] The S-layer entirely coats the bacterial cell surface and is composed of (glyco)proteins, which bind by noncovalent interactions to cell wall components and are arrayed in a two-dimensional lattice.[167] S-layer, however, is not present in all Gram-positive bacteria as it is absent from all members of *Listeria* genus. Within the cell envelope, proteins can associate with cytoplasmic membrane or cell wall components.[8]

### 12.3.1 Membrane-Associated Proteins

Membrane-associated proteins include membrane integrated proteins as well as peripheral membrane proteins. Being different from membrane integrated proteins, peripheral membrane proteins do not possess membrane spanning domains. Membrane integrated proteins are anchored within the lipid bilayer and thus systematically exhibit hydrophobic transmembrane domains (TMDs), which are normally α-helices for proteins found in the cytoplasmic membrane. Peripheral membrane proteins include (1) lipoproteins, (2) subunits of membrane-associated complexes, and (3) proteins interacting with membrane components due to electrostatic and/or hydrophobic/steric properties.[168] Following recommendations of the Gene Ontology (GO) Consortium for describing location of cellular components (one of the three organizing principles of GO with biological process and molecular function),[169] two classes of membrane-related location are distinguished (Figure 12.8). First, intrinsic to plasma membrane (GO:0031226) refers to proteins with covalently attached moiety embedded in the cytoplasmic membrane, which splits into (1) integral to plasma membrane (GO:0005887) corresponding to membrane integrated proteins, where some part of the peptide sequence spans all or part of the cytoplasmic membrane; and (2) anchored to plasma membrane (GO:0046658) corresponding to proteins tethered to the cytoplasmic membrane by a nonpolypeptidic covalently attached anchor: lipoproteins. Second, extrinsic to plasma membrane (GO:0019897) refers to proteins neither anchored by covalent bonds to any moiety nor directly embedded in the cytoplasmic membrane; some of these proteins can be (1) primarily present in the cytoplasm (GO:0005737) but interact with membrane components, or (2) subcomponents localized within protein complex (GO:0043234).

#### 12.3.1.1 Integral Membrane Proteins

As already mentioned, all bacterial IMPs are presumably inserted into the cytoplasmic membrane via YidC homologues[48,170,171] (i.e., SpoIIIJ and YqjG in Gram-positive bacteria[30]; see section 12.2.1). The Sec-independent function of YidC homologues is conserved and essential for bacterial cell growth as it works like a membrane protein insertase.[172] YidC plays a major role in the folding step of transmembrane-spanning domains but the exact mechanism of functioning is not fully understood.[173] YidC would facilitate the insertion of membrane proteins by providing a special amphiphilic surface, which would overcome the repulsion of the hydrophobic protein segments by polar head groups. In addition, polar residues seem to be protected against the hydrocarbon core of

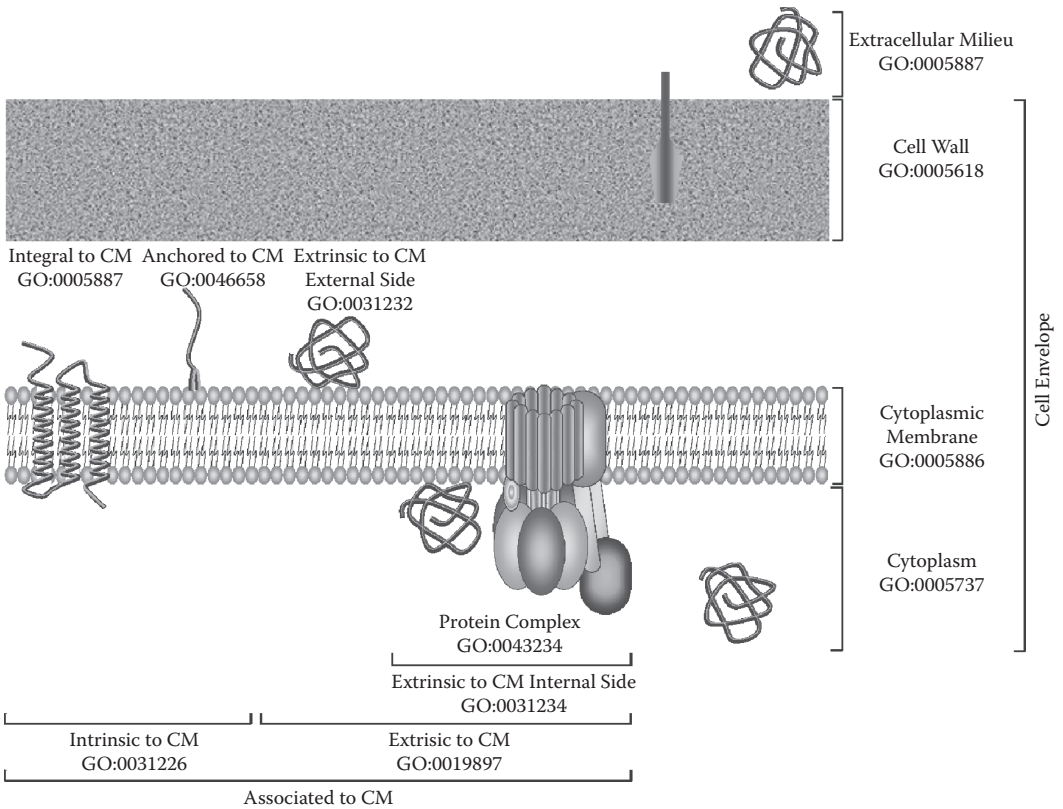


**FIGURE 12.8** Description of protein localization following GO in Gram-positive bacteria.[179] In monoderm bacteria four subcellular compartments can be distinguished: (1) cytoplasm, (2) cytoplasmic membrane, (3) cell wall, and (4) extracellular milieu. A membrane-associated protein can be intrinsic or extrinsic. Proteins intrinsic to CM are either integral to membrane (i.e., integral membrane proteins) or anchored to CM, essentially lipoproteins with the restriction of lipoproteins having TMDs. Proteins extrinsic to CM can be located on the external or internal side of the CM (i.e., in exoplasmic or cytoplasmic compartment, respectively). They can interact more or less temporarily with membrane components or be part of membrane protein complex (e.g., $F_1F_0$ATP synthetase δ subunit) as indicated on the schema. A protein can also have multiple final localization. Importantly, because cell wall of Gram-positive bacteria is permeable, extracellular milieu penetrates it. CM, cytoplasmic membrane; GO, gene ontology.

the membrane by YidC. In the case of Sec-dependent translocation, the protein would be stabilized and then folded by contact with YidC after leaving the Sec YEG channel. It is quite possible that the transmembrane segments could fold and interact with each other even within SecYEG-YidC machinery. It has been suggest that YidC functions as an assembly site for hydrophobic domains, so it may be necessary for its attaching to the individual subunits of multisubunit membrane complex.[174] It is worth noting that, in *E. coli*, targeting, translocation, and insertion of IMPs are considered cotranslational and thus SRP dependent.[175]

Translocation of polypeptide chain is promoted by signal peptides and interrupted by another type of topogenic element called stop-transfer sequence[176,177]; both types of topogenic sequences act as α-helical transmembrane domains. Multiple uncleaved signal peptides can be found all along the amino acid sequences; when located N-terminally, they can be cleaved or not. These types of topogenic elements have a $C_{out}$-$N_{in}$ topology and when uncleaved are also called signal-anchors or Type II signals.[178] Similarly, one or more stop-transfer sequences with an $N_{out}$-$C_{in}$ topology can be present in polypeptide chain and are also called Type II signals. Single-spanning membrane proteins are discriminated on the basis of Type I or Type II signal (Figure 12.9). Polytopic membrane

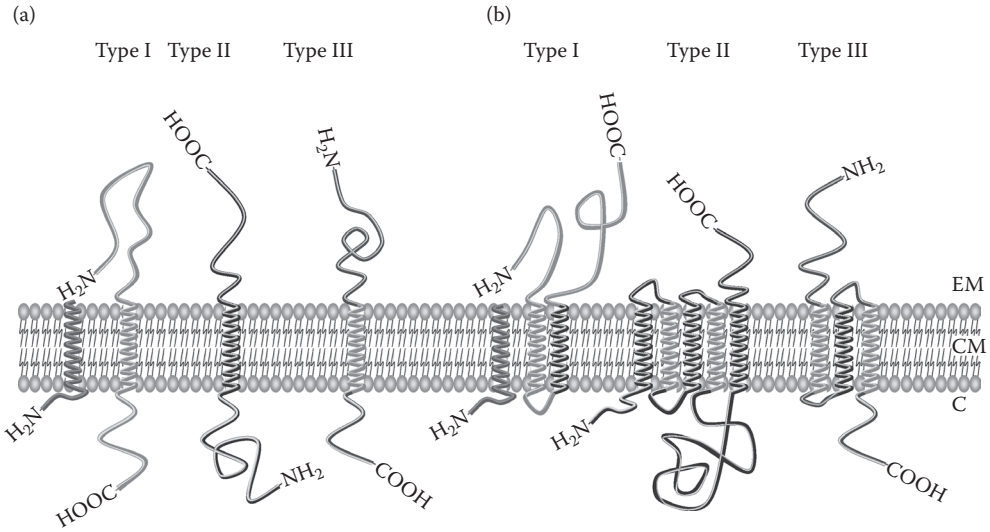


**FIGURE 12.9** Classification and topology of IMP in cytoplasmic membrane.[179] (A) Three types of single-spanning membrane proteins can be discriminated: (1) Type I proteins possess a cleavable N-terminal signal peptide and thus have a Type I signal or stop-transfer sequence with $N_{out}$-$C_{in}$ topology; (2) Type II proteins have a Type II signal or signal-anchor sequence with a $C_{out}$-$N_{in}$ topology, which can correspond to an uncleavable N-terminal signal peptide; and (3) Type III proteins have reverse signal-anchor sequence (i.e., with a $N_{out}$-$C_{in}$ topology) and are sometimes described as Type I proteins without a cleavable signal peptide since the reverse signal-anchor sequence is a Type I signal. (B) Three types of multispanning-membrane proteins (i.e., with a number of TMDs higher than 1) can be distinguished based on whether the most N-terminal TMD is either (1) cleaved by a SPase (i.e., Type I); (2) spans the membrane with an $N_{out}$-$C_{in}$ orientation (i.e., Type II); or (3) have a $C_{out}$-$N_{in}$ orientation. Various numbers of TMDs are present in multispanning-membrane proteins where alternates Type I and II signals. EM, extracellular milieu; CM, cytoplasmic membrane; C, cytoplasm.

proteins are built up of a series of Types I and II modules that initiate and halt the translocation of the polypeptide chain. Such IMPs are classified on the basis of the orientation of the most N-terminal TMD spanning the lipid bilayer[179] (Figure 12.9).

Numerous tools are available to predict IMPs and their topology. Table 12.4 is an attempt to review all of them. These tools are based on various approaches, such as (1) statistical analyses (e.g., TMpred or TMSTAT; (2) hydrophobicity analyses (e.g., SOSUI or TopPred); (3) NNs (e.g., PREDTMR or PHDhtm); (4) SVMs (e.g., SVMtm); or (5) HMMs (e.g., HMMTOP or THUMBUP). Some of them—for example, ConPred or TUPS—combine results of several models. Readers are invited to study related publications listed in Table 12.4 in order to get further insight into the methods used. The total number of IMPs encoded in *L. monocytogenes* genomes is estimated to be 1204 and 733 in *L. monocytogenes* EGD-e and F2365, respectively.[18] Virulence factors ActA and SvpA are cell surface exposed IMPs of *L. monocytogenes* exhibiting a hydrophobic tail (i.e., a carboxyl terminal region containing a hydrophobic domain followed by positively charged residues).[180,181] Following genomic analysis, a total of 11 surface proteins with hydrophobic tails have been predicted in *L. monocytogenes* EGD-e.[19] Contrary to what is sometimes assumed, cell surface IMPs should not be restricted only to proteins with a hydrophobic tail[182]; indeed, depending on the number and organization of Type I and Type II signals in IMPs, final protein topology can result in the cell surface exposure of functional domains located not only N- or C-terminally but also in loops. With FbpA as an example,[47] it can be noticed that from experimental investigations, some proteins appear located within the cytoplasmic membrane despite the absence of predicted signal peptide and TMD.

**TABLE 12.4**
**Bioinformatic Resources for Prediction of TMDs**

| Application | Method | Webserver | Ref. |
|---|---|---|---|
| TMpred | Statistical analysis | http://www.ch.embnet.org/software/TMPRED_form.html | 249 |
| TopPred | Hydrophobicity analysis | http://bioweb.pasteur.fr/seqanal/interfaces/toppred.html | 250 |
| PHDhtm | Neural network | http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_htm.html | 251 |
| DAS | Hydrophobicity analysis | http://www.sbc.su.se/~miklos/DAS/ | 252 |
| TMAP | Statistical analysis | http://bioweb.pasteur.fr/seqanal/interfaces/tmap.html | 253 |
| TSEG | Hydrophobicity analysis | http://www.genome.jp/SIT/tsegdir/tseg_exe.html | 254 |
| TMHMM | Hidden Markov model | http://www.cbs.dtu.dk/services/TMHMM-2.0/ | 255 |
| SOSUI | Hydrophobicity analysis | http://bp.nuap.nagoya-u.ac.jp/sosui/ | 256 |
| PREDTMR | Neural network | http://athina.biol.uoa.gr/PRED-TMR2/ | 257 |
| kPROT | Statistical analysis | http://bioinfo.weizmann.ac.il/kPROT/ | 258 |
| TMSTAT | Statistical analysis | http://bioinfo.mbb.yale.edu/tmstat/ | 259 |
| HMMTOP | Hidden Markov model | http://www.enzim.hu/hmmtop/ | 260 |
| TMFinder | Hydrophobicity analysis | http://www.bioinformatics-canada.org/TM/ | 261 |
| DAS-TMfilter | Hydrophobicity analysis | http://www.enzim.hu/DAS/DAS.html | 262 |
| SPLIT | Statistical analysis | http://split.pmfst.hr/split/ | 263 |
| THUMBUP | Hidden Markov model | http://sparks.informatics.iupui.edu/Softwares-Services_files/thumbup.htm | 264 |
| UMDHMMTMHP | Hidden Markov model | http://sparks.informatics.iupui.edu/Softwares-Services_files/umdhmm.htm | 264 |
| TUPS | Combination | http://sparks.informatics.iupui.edu/Softwares-Services_files/tups.htm | 264 |
| BPROMPT | Bayesian belief network | http://www.jenner.ac.uk/BPROMPT/ | 265 |
| SVMtm | Support vector machine | http://ccb.imb.uq.edu.au/svmtm/svmtm_predictor.shtml | 266 |
| ConPred | Combination | http://bioinfo.si.hirosaki-u.ac.jp/~ConPred2/ | 267 |
| HMMTM | Hidden Markov model | http://biophysics.biol.uoa.gr/HMM-TM/ | 268 |
| MINNOU | Hydrophobicity analysis | http://minnou.cchmc.org/ | 269 |
| MEMSAT | Statistical analysis | http://bioinf.cs.ucl.ac.uk/psipred/ | 270 |

### 12.3.1.2 Lipoproteins

In monoderm bacteria, lipoproteins are attached to the outer surface of the cytoplasmic membrane via a covalently bound lipid anchor.[32] Systematically, these proteins are first translocated in a Sec-dependent manner and thus possess N-terminal signal sequences. Such signal peptides, however, belong to class 2 as it exhibits a conserved lipobox motif in the C-domain.[30] It can be noticed, however, that in *E. coli* YidC plays an important role in targeting and translocation of some lipoproteins.[183] Lipobox includes invariably a cysteine residue located just after the cleavage site of signal peptide. After translocation of the prolipoprotein, a common post-translational modification involves a prolipoprotein diacylglyceryl transferase (Lgt), which adds an N-acyl diglyceride group from a glycerophospholipid to the SH-group of the lipobox cysteine.[184] This thioether linkage allows protein anchoring to the membrane thanks to the insertion of the diacylglyceryl group into the lipid bilayer. Subsequently, SPase II (also called Lsp for lipoprotein signal peptidase) cleaves off the signal peptide and the cysteine becomes the N-terminal residue.[185] In contrast to *E. coli*, however, lipidation by Lgt (Lmo2482) in *Listeria* is neither essential for bacterial growth nor a prerequisite for activity of Lsp.[186] Once signal peptide is cleaved off, the amino-terminal cysteine residue is usually acylated at the free amino group by a phospholipid/apolipoprotein transacylase (Lnt), resulting

**TABLE 12.5**
**Bioinformatic Resources for Prediction of Lipoproteins**

| Application | Method | Webserver | Ref. |
|---|---|---|---|
| ScanProsite | Profile (PS51257) search<br>Pattern (G+LPP) search | http://www.expasy.org/tools/scanprosite/ | 189, 190 |
| DOLOP | Pattern search | http://www.mrc-lmb.cam.ac.uk/genomes/dolop/ | 191 |
| LipoP | Hidden Markov model | http://www.cbs.dtu.dk/services/LipoP/ | 192 |
| SPEPlip | Neural network | http://gpcr.biocomp.unibo.it/predictors/ | 193 |
| LipPred | Naïve Bayesian network | http://www.jenner.ac.uk/LipPred/ | 194 |

in protein anchoring to membrane long chain fatty acid.[187] This additional post-translational modification step, however, does not seem to be conserved in all bacteria,[184] as an *lnt* gene is apparently lacking from all sequenced members of Firmicutes phylum, and *Listeria* species is no exception.[188] It can be further noticed that some lipoproteins are IMPs integrated to the cytoplasmic membrane by TMDs in a YidC-dependent manner.[173]

As summarized in Table 12.5, several resources can be applied for genomic identification of lipoproteins. In PROSITE,[189] lipobox motif was previously referred to as PS00013 and defined by the regular expression {DERK}(6)-[LIVMFWSTAG](2)-[LIVMFYSTAGCQ]-[AGS]-C, where {DERK}(6) means that none of the four amino acids are allowed in the first six positions relative to the cleavage site. The pattern had two additional rules: (1) The cysteine must be between positions 15 and 35, and (2) at least one positively charged residue (K or R) must be within one of the first seven N-terminal residues. This pattern (i.e., a qualitative motif description based on a regular expression-like syntax) is now replaced by a profile (i.e., a quantitative motif description based on the generalized profile syntax), referred to as PS51257 and defined as prokaryotic membrane lipoprotein lipid attachment site profile. Originally, lipobox search using PS00013 was known to generate a significant proportion of false-positives, which prompted the need to improve the syntax of this regular expression. Thus, a refined pattern named G+LPP (for Gram-positive lipoprotein) and using PROSITE syntax—that is, [MV]-X(0,13)-[RK]-{DERKQ}(6,20)-[LIVMFESTAG]-[LVIAM] [IVMSTAFG][AG]-C—was developed.[190] This pattern appears more specific for the identification of Gram-positive bacterial lipobox and allows greater discrimination against false-positives compared with PS00013. Thus, lipobox can be predicted by scanning polypeptidic sequences for the presence of PS51257 profile or G+LPP pattern using ScanProsite.[115]

However, correct sequence assignment as putative lipoprotein also requests that the lipobox is localized within an N-terminal Sec-dependent signal peptide where it covers H- and C-domains; signal peptide can be predicted following analysis with previously described tools (Table 12.1). DOLOP compiles similar criteria by scanning query sequences for presence of (1) a lipobox within the first 40 residues from the N-terminus with the consensus as [LVI][ASTVI][ASG][C], (2) positively charged amino acid in n-domain of signal peptide, and (3) at least 7–22 residues between the predicted lipobox and the charged residue.[191] LipoP is based on an HMM and discriminates among lipoprotein signal peptides (cleaved by SPase II), other signal peptides (cleaved by SPase I), n-terminal membrane helices, and cytoplasmic protein following attribution of scores.[192] Despite having been primarily developed for Gram-negative bacteria, LipoP can also efficiently identify lipoproteins in Gram-positive bacteria. The only feeble point may be that when handling lipoproteins with transmembrane regions, the HMM misses, in some cases, the lipoprotein signal peptide. SPEPlip is an NN-based method for prediction of signal peptide and integrating a regular expression search based on PROSITE pattern.[193] LipPred is based on test against a naïve Bayesian network allowing the identification of lipoprotein with a calculated index for prediction confidence.[194] When compared to other available methods, LipPred can be considered as the most accurate for detection of lipoprotein signal sequence and SPase II cleavage site. Finally, lipoprotein-associated domains

can be searched using HMMs from Pfam or Tigrfam: namely, in Firmicutes, PF00938, PF01347, PF01540, PF02030, PF03180, PF03202, PF03260, PF03304, PF03305, PF03330, PF03640, PF04200, PF05481, TIGR00363, TIGR00413, TIGR01742, TIGR01533, and TIGR02898. Proteins identified following this approach should, however, be considered with great care and ideally confirmed by methods listed previously for lipoprotein modification/processing motif.

Following search for presence of PS00013 pattern and signal peptide using SignalP v2.0, 68 lipoproteins were originally identified in the genome of *L. monocytogenes* EGD-e.[19] Pfam and Tigrfam searches allowed the identification of 70 lipoproteins in *L. monocytogenes* F2365, whereas the number of lipoproteins was estimated at 63 in *L. monocytogenes* EGD-e.[18] Using the computational pipeline Augur,[89] where lipoproteins are identified on the basis of G+LPP pattern match, 65 lipoproteins were identified in *L. monocytogenes* EGD-e.[18] Generating a new HMM from 26 verified lipoproteins by proteomic analysis in *L. monocytogenes* EGD-e, the number of lipoproteins was reestimated down to 62 in *L. monocytogenes* EGD-e and 56 in *L. monocytogenes* F2365.[186] Despite discrepancy in predicted number, lipoproteins constitute the largest family of putative surface proteins in *Listeria* (68 out of 133 originally predicted in *L. monocytogenes* EGD-e).[19] These lipoproteins are putatively involved in various metabolic pathways (e.g., as substrate-binding components of ABC transport systems); remarkably, no biological function could be assigned for a large proportion. In spite of their predominance as surface proteins of Gram-positive bacteria, very few lipoproteins have been biochemically characterized.[195] In *Listeria*, only five have been more specifically investigated and thus confirmed, at least partially, in term of biological function: (1) TcsA (Lmo1388), a $CD4^+$ T cell-stimulating antigen presented by major histocompatibility complex class II molecules[196]; (2) GbuC (Lmo1016), a glycine betaine binding-protein part of an ABC transport system[197]; (3) the substrate binding protein OpuCC (Lmo1426) part of an ABC L-carnitine transporte[198]; (4) OppA (Lmo2196), another ABC substrate binding protein mediating the transport of oligopeptides[199]; and (5) the virulence factor LpeA (Lmo1847) involved in bacterial entry into eukaryotic infected cells.[200]

#### 12.3.1.3 Extrinsic Membrane Proteins

No bioinfomartic tool is currently available to identify such proteins, which are most of the time primarily predicted as extracellular or cytoplasmic depending on their presence on the external or internal side of the cytoplasmic membrane (Figure 12.8). Thus, their identification requires a deep understanding of bacterial physiology and excellent general literature survey. Some of these proteins, which are not intrinsic to the cytoplasmic membrane, can be subunits of membrane protein complexes such as $F_1F_0$ATP synthetase (GO:0045260), fumarate reductase complex (GO:0045284), or ABC (ATP binding cassette) transporters (GO: 0043190). Some other extrinsic membrane proteins can interact more or less temporarily with other membrane components, including other membrane-associated proteins. For example, in the SRP-dependent pathway, ribosomal proteins interact with Sec translocon in the course of cotranslational translocation,[201] or in two-component systems, response regulators interact with membrane bound sensors.[202] It should also be noticed that some cytoplasmic proteins can associate with the lipid bilayer by weak interactions and by no means be functionally associated with membrane components. To date, the number of extrinsic membrane proteins has never been estimated in *L. monocytogenes*.

### 12.3.2 Cell Wall-Associated Proteins

Cell wall-associated proteins are either covalently linked to peptidoglycan when possessing a C-terminal LPXTG motif or noncovalently linked to cell wall components by a cell wall-binding domain (CBD).[8] In Gram-positive bacteria, six CWBDs are currently characterized: CWBD of Type 1 (CWBD1), CWBD of Type 2 (CWBD2), Lysin motif domain (LysM), GW modules, S-layer homology domain (SLHD), and WXL domain (WXL).[8,203] In *Listeria*, however, only proteins with

LPXTG, LysM, GW, and WXL motifs have been identified so far.[182,203,204] In *L. monocytogenes*, proteins exhibiting such motifs systematically possess Sec-dependent N-terminal signal peptide. These motifs can be found using RPS-BLAST (reverse position-specific BLAST)[205] or HMM[206] from different databases—namely, InterPRO,[207] Pfam,[208] SMART,[209] TIGRfam,[210] and SuperFamily.[211]

#### 12.3.2.1 LPXTG Motif

LPXTG motif (IPR001899, PF00746, TIGR01167) is found in proteins covalently attached to cell wall by sortases[55] (see section 12.2.1). This motif consists of a pattern varying around LPXTG, a hydrophobic domain, and a positively charged tail.[212] Cross-bridging of the protein to cell wall by sortase would occur in four steps.[213,214] Following translocation across the Sec apparatus, membrane-associated sortase recognizes the LPXTG motif and cleaves it before linking proteins to cell wall precursor lipid II.[215] The proteins thus linked to lipid II are further incorporated into the cell wall by transglycosylation and transpeptidation reactions that generate peptidoglycan. Forty-one proteins with LPXTG motif substrates of StrA have been identified in *L. monocytogenes* EGD-e,[19] whereas two proteins with an NXZTN motif are recognized by StrB.[181,216] In *L. monocytogenes* F2365, a total of 44 LPXTG-like proteins have been identified.[18] StrA is required for bacterial virulence as among its protein substrates several virulence factors have been identified, such as InlA.[59,217] Compared to StrA, StrB plays a minor role both in terms of number of proteins anchored to cell wall and involvement in bacterial virulence, although virulence factor SvpA is substrate to StrB.[216]

#### 12.3.2.2 Noncovalently Attached Cell Wall Proteins

Even though most cell-associated proteins contain an LPXTG motif in *Listeria*, several proteins bear domains involved in noncovalent attachment to the components of cell wall. Three motifs are clearly established as involved in noncovalent cell wall attachment in *Listeria*: LysM, GW, and WXL.[182,203,204] Besides these known attachment domains, other proteins found in the cell wall are retained by putative CWBDs, alternative or unknown mechanisms.

##### *12.3.2.2.1 LysM*

LysM (IPR002482, PF01476, SM00257, SSF54106) is a motif about 40 residues long and composed of three α-helices with a general peptidoglycan binding function.[218] It is found in a variety of enzymes mostly involved in bacterial cell wall degradation. When present, this motif is often repeated several times in the protein sequence. Interestingly, proteins bearing this motif can attach to surface of Gram-positive bacteria other than the ones that synthesized it.[218] In *L. monocytogenes*, several proteins bear LysM domains; among others, P60 (also called Iap or CwhA) can be cited as it is also considered as a virulence factor.[219] It is worth stressing here that contrary to previous assumption,[19,182] NlpC/P60 domain should not be considered as the motif involved in cell wall binding *stricto sensu* (see section 12.3.2.2.4). In *L. monocytogenes* EGD-e and F2365 genomes, six LysM proteins, including P60 orthologue, were identified following bioinformatic analysis (Desvaux and Hébraud, unpublished data).

##### *12.3.2.2.2 GW*

GW (SSF82057) module was originally identified in *L. monocytogenes* within internalin InlB.[220] This module is about 80 amino acids, contains a highly conserved dipeptide Gly-Trp, and interacts with lipoteichoic acids allowing cell surface attachment.[182] GW modules are found in multicopy, as in InlB, where three copies are present in the C-terminal region, or in Ami, where eight modules are present. It also appears that the higher the number of GW modules is, the stronger is the attachment to the bacterial cell wall[221,222]; proteins exhibiting only one GW module would not be surface attached at all. It is interesting to note that GW modules are related to Src homology-3 (SH3) clan (CL0010) and more specifically prokaryotic SH3 of Type 3, or SH3b (IPR013247, PF08239), but are unlikely to act as functional mimics of SH3 domains since proline-binding sites are blocked or

destroyed in GW domains.[223] In *L. monocytogenes* EGD-e, nine GW proteins were identified, most of which (including Ami) exhibit an amidase domain.[19,204] This indicates that this class of protein would be mainly involved in cell wall degradation in *L. monocytogenes,* although Ami is also considered as a virulence factor involved in bacterial adhesion to infected cells.[182,224]

#### *12.3.2.2.3 WXL*

Following genomic analysis of *Enterococcus faecalis*, a novel C-terminal cell wall binding motif was uncovered and named WXL domain.[203] This conserved domain is characterized by a first WXL motif and an additional YXXX[LIV]TWXLXXXP motif found further downstream; the two WXL domains are separated by between 66 and 247 residues. WXL domain was found in CDS of several genomes of low G+C Gram-positive bacteria, including *L. monocytogenes* where four proteins bearing such domain were identified.[203] In *E. faecalis*, it was demonstrated that WXL domain is a determinant of bacterial subcellular protein. Indeed, its presence conferred cell surface display of the protein, whereas specific deletions into the domains prevented its display. Moreover, neither proteins nor carbohydrates were necessary for cell wall attachment but peptidoglycan was a binding ligand for WXL domain. As LysM, WXL can attach to cell wall of a variety of Gram-positive bacteria. From genome-wide analysis of Gram-positive bacteria,[225] it appeared that genes encoding WXL proteins seem to cluster and that some N-terminal regions of these proteins are involved in utilization of plant complex polysaccharides. It was also suggested that some WXL proteins might mediate interactions between different bacteria species.[203] In *Listeria*, physiological function of such proteins awaits to be established and their presence on bacterial cell surface remains to be demonstrated.

#### *12.3.2.2.4 Other Noncovalently Attached Cell Wall Proteins*

ChW motif (IPR006637, SM00728, PF07538) stands for clostridial hydrophobic domain with a conserved W residue and was first uncovered from bioinformatic analysis of *Clostridium acetobutylicum* genome.[226] As GW, ChW contains a highly conserved Gly-Trp dipeptide motif and was suggested to be involved in cell surface attachment.[33] This repetitive domain can be found several times along the protein sequence (up to nine copies). ChW proteins were speculated to be part of a molecular complex on bacterial cell surface dedicated to degradation of polymer and surface adhesion.[226] One putative serine protease bearing three copies of ChW motif is encoded in genome of *L. monocytogenes* F2365 (Desvaux and Hébraud, unpublished data); its expression, secretion, cell surface display, and function remain to be established. Similarly, function of SH3b is as yet unknown, but *Staphylococcus simulans* lysostaphin contains such a domain in its C-terminal region.[227] Since this region mediates protein binding to bacterial cell wall, SH3b may have this function.

It is important to note that despite the absence of cell wall binding motifs in some enzymes involved in cell wall degradation, such proteins have affinity for cell wall components via their enzymatic active site. Thus, secreted proteins with cell wall degradation domains, such as NlpC/P600 (IPR000064, PF00877)[228] or N-acetylmuramoyl-L-alanine (IPR002508, PF01520), can be localized in cell wall. However, such domains should not be considered as cell wall binding motifs per se since primary function of these enzymes is to cleave cell wall components following when they find a new cleavage site or are released into the extracellular milieu. In *L. monocytogenes*, these enzymes are involved in numerous cellular processes.[126]

It is now well known that many metabolic enzymes can be surface localized in Gram-positive bacteria.[229] Such proteins lack N-terminal signal peptide and are supposedly secreted by pathways alternative to the known ones. SecretomeP is a bioinformatic tool dedicated to the prediction of such proteins.[230] However, instead of secretion through the cytoplasmic membrane, these proteins could be released from the bacterial cell following autolysis and then attached to cell surface of nonlysed cells. In *Streptococcus pneumoniae*, release of cytoplasmic proteins is triggered by competent cells and originates from lysis of noncompetent cells.[231] This tightly controlled phenomenon was named allolysis and involves several cell wall hydrolases. Once released and associated with bacterial cell surface, such proteins generally moonlight.[23] For example, in *S. pneumoniae*, the glycolytic enzyme

glyceraldehyde 3-phosphate dehydrogenase genolase exhibits plasmin(ogen)-binding activity on the bacterial cell surface and thus significantly enhances bacterial virulence.[232] In *L. monocytogenes*, several proteins primarily predicted as cytoplasmic were also identified in the cell wall fraction, including enolase, which was demonstrated to bind human plasminogen.[24] Functions of other cytoplasmic proteins found at this subcellular location (e.g., chaperone DnaK, elongation factor TU, or glyceraldehyde-3-phosphate dehydrogenase) remain to be elucidated as well as protein motifs involved in cell wall attachment.

## 12.4 CONCLUSIONS AND PERSPECTIVES

The proteomic technologies are certainly the most powerful and appropriate to provide global and accurate information on the expression, structure, and function of proteins. Since the first description of protein extraction and separation using two-dimensional gel electrophoresis (2-DE) in 1975,[233–235] many advances (use of new detergents, immobilized pH gradient, new apparatus for IsoElectroFocalization [first dimension], and SDS-PAGE [second dimension]) have been brought to improve protein solubilization and resolution as well as reproducibility and implementation of the techniques. Over the past two decades, 2-DE progressively became the classical method of choice to separate and compare complex mixtures of proteins and was mainly applied for soluble intracellular proteins. In the field of bacteriology, an increasing number of investigations using comparative proteomic approaches was devoted to the characterization of adaptive responses—namely, to various physicochemical stresses or to the effect of a gene mutation. At this time, however, protein identification was difficult and time consuming as it essentially involved Edman degradation and sequence alignment from short and on limited numbers of amino acid sequences. Consequently, these early studies generally remained quite descriptive and phenomenological.

However, two occurrences gave a considerable impetus to proteomic analyses: (1) the availability of ever growing amounts of genomic sequence data, and (2) important advances in mass spectrometry technology for ionization and detection of large molecules such as peptides and proteins. Indeed, data obtained with mass spectrometry analysis—namely, peptide mass fingerprinting or fragmentation—could from then be matched against databases of all known gene products and thus greatly facilitated protein identification. The use of 2-DE and MS tools to separate and identify proteins is now widespread in all domains of life science. One of the consequences of this remarkable progress was the possibility to establish 2-DE databases containing several hundreds of identified protein spots available on proteome reference maps. Thus, the first bacterial 2-DE database was established on *E. coli* cytosoluble proteins separated into different pH gradients and regularly brought up to date.[236,237] Similar but generally more limited 2-DE databases are available for other bacterial species, including *L. monocytogenes* EGD-e[238] (http://www.clermont.inra.fr/proteome).

Another consequence was the possibility and need to investigate further the different cell compartments and thus following cell fractionation to explore thoroughly the different subcellular proteomes that, in Gram-positive bacteria, include (1) cytoplasmic proteins, (2) membrane associated proteins, (3) cell-wall associated proteins, and (4) proteins secreted in the extracellular milieu. While extracting and separating cytoplasmic and extracellular proteins could be achieved rather easily and efficiently, classical 2-DE procedures failed to give a good overview of proteins present in the cell envelope—that is, proteins associated to cytoplasmic membrane or cell wall. Beyond the well-known limitations of 2-DE gel-based technology (the inability to separate or to reveal low-abundance proteins, high molecular mass and extreme p*I* [isoelectric point] proteins), other limitations appeared much more problematic for the cell surface subproteomes due to the intrinsic properties of cell envelope associated proteins. Indeed, multitransmembrane proteins are generally highly hydrophobic and are either almost impossible to solubilize during the extraction procedure or not recovered in the second dimension of 2-DE due to self-aggregation and irreversible precipitation in IEF. On the other hand, proteins noncovalently or *a fortiori* covalently attached to the cell wall peptidoglycan are very difficult to extract and require specific and laborious treatments not directly

compatible with classical 2-DE separation. Altogether, this considerably hampers proteomic analysis in classical 2-DE gel-based technology.

Different strategies are now developed to tackle the difficulties to analyze these cell envelope subproteomes. These strategies can associate different protocols of protein extraction with different techniques of separation and mass spectrometry.[239] Several studies have attempted to extract membrane associated proteins of Gram-positive bacteria by combining protocols described for Gram-negative or eukaryotes organisms.[240] Thus, the extraction procedures could include enzymatic treatment, fractionation of broken cell by centrifugation, use of chemical agents such as zwitterionic detergents for solubilization of hydrophobic proteins,[241] solvents for delipidation,[242] or protein extraction and separation.[243] For example, such a combinational approach has been used to efficiently characterize by 2-DE the cell-wall and membrane-associated subproteomes of the Gram-positive bacterium *Staphylococcus xylosus*.[244]

A different protocol originally developed for *Bacillus cereus*[245] has been applied for the global extraction of *L. monocytogenes* cell-surface proteins combining the protein solubilization by SDS with a classical SDS-PAGE separation.[246] The 1-DE and 2-DE separations were both used to characterize the cell wall subproteome of *L. monocytogenes*.[24] In this case, proteins were extracted by the sequential action of two salts at high concentration and their identification was performed by N-terminal sequencing and peptide mass fingerprint obtained with matrix assisted laser desorption ionization time of flight mass spectrometer (MALDI-TOF MS). It is interesting to note that among the 55 identified proteins, only 27 possessed a peptide signal, including 4 proteins with cell wall-binding motifs (2 GW proteins and 2 LysM proteins), 20 lipoproteins, and 3 proteins with no predictable surface association motif. The 28 remaining proteins without peptide signal were primarily predicted with cytoplasmic functions and nothing could explain how they managed to cross the cytoplasmic membrane or how they associated with bacterial cell wall. Such unusual localization of cytoplasmic proteins leads to the suggestion that they could moonlight on the bacterial cell surface, although no experimental evidence could back up such a hypothesis.

More recent strategies and technologies consist in analyzing peptide mixture obtained by tryptic digestion of cell envelope protein samples issued from stringent protocols extraction (e.g., combining cell mechanic broken) fractionation by centrifugation, and treatment with high concentration of SDS at 100°C, then at 80°C.[58,247] An alternative approach consists in "shaving" the bacterial surface with a specific protease (such as trypsin) to cleave surface-exposed proteins.[239] After lyophilization to remove SDS, the peptide hydrolysate is then separated by two-dimensional liquid chromatography coupled to tandem mass spectrometer (2-D LC MS/MS). This separation technique, termed "shotgun proteomic" or multidimensional protein identification technology (MudPIT), uses a two-dimensional liquid chromatography to separate a tryptic peptide mixture where a strong cation exchange is applied in the first dimension and a reverse phase is applied in the second dimension. The separated peptides are subjected online to analysis by fragmentation (MS/MS) in an electrospray ionization MS. Peptide fragmentation spectrum is further used to identify the original protein via query against databases.

Besides a significant gain of time, the use of 2-D LC MS/MS overcomes all limitations of gel-based 2-DE previously cited. Even hydrophobic proteins can be identified thanks to amino acid cleavage sites accessible to tryptic digestion in exposed regions of the protein. These new approaches allowed extraction of proteins following treatments that are not always compatible with classical 2-DE. Consequently, the set of proteins identified with high number of transmembrane spanning regions or LPXTG motif (i.e., covalently anchored to cell wall) has been significantly enlarged. The other development of 2-D LC MS/MS concerns the possibility to perform quantitative proteomics for comparative analysis of samples pretreated with amino acid tags or labels such as the ICAT™ (isotope-coded affinity tags), iTRAQ™, or SILAC technologies.[239]

The years to come will undoubtedly see the development and improvement of these new exploring methods of subproteomes. Everyone will have the possibility to map protein expression and to compare several biological samples with high throughput, sensibility, and resolution. In spite of

this very attractive progress and considering at least its complementarity, classical 2-DE technique remains irreplaceable. Indeed and contrary to LC MS/MS approaches, gel-based 2-DE allows one to separate simultaneously several hundreds of proteins at once and to visualize shifts due to post-translational modifications. The implementation of two or more complementary proteomic strategies would be one of the keys to generate valuable information on the role of cell envelope proteins in pathogenic processes, bacterial communication, sensing of and exchange with its environment, motility, and adhesion on and colonization of biotic or abiotic surfaces.

## REFERENCES


1. Woese, C.R., A new biology for a new century, *Microbiol. Mol. Biol. Rev.*, 68, 173, 2004.
2. Shatalkin, A.I., Highest level of division in classification of organisms. 3. Monodermata and didermata, *Zh. Obshch. Biol.*, 65, 195, 2004.
3. Danchin, A., *The Delphic boat: What genomes tell us,* Harvard University Press, Cambridge, MA, 2003.
4. Monod, J., *Le hasard et la nécessité: Essai sur la philosophie naturelle de la biologie moderne,* Éditions du Seuil, Paris, 1970.
5. Cavalier-Smith, T., Rooting the tree of life by transition analyses, *Biol. Direct.*, 1, 19, 2006.
6. Cavalier-Smith, T., The neomuran origin of archaebacteria, the negibacterial root of the universal tree and bacterial megaclassification, *Int. J. Syst. Evol. Microbiol.*, 52, 7, 2002.
7. Gupta, R.S., The natural evolutionary relationships among prokaryotes, *Crit. Rev. Microbiol.*, 26, 111, 2000.
8. Desvaux, M. et al., Protein cell surface display in Gram-positive bacteria: From single protein to macromolecular protein structure, *FEMS Microbiol. Lett.*, 256, 1, 2006.
9. Garrity, G. M., *Bergey's manual of systematic bacteriology*, 2nd ed., Springer, Berlin, 2001.
10. Woese, C.R., Bacterial evolution, *Microbiol. Rev.*, 51, 221, 1987.
11. Snel, B., Huynen, M.A., and Dutilh, B.E., Genome trees and the nature of genome evolution, *Annu. Rev. Microbiol.*, 59, 191, 2005.
12. Gupta, R.S. and Griffiths, E., Critical issues in bacterial phylogeny, *Theor. Popul. Biol.*, 61, 423, 2002.
13. Mira, A. et al., Evolutionary relationships of *Fusobacterium nucleatum* based on phylogenetic analysis and comparative genomics, *BMC Evol. Biol.*, 4, 50, 2004.
14. Botero, L.M. et al., *Thermobaculum terrenum* gen. nov., sp. nov.: A nonphototrophic Gram-positive thermophile representing an environmental clone group related to the *Chloroflexi* (green nonsulfur bacteria) and *Thermomicrobia*, *Arch. Microbiol.*, 181, 269, 2004.
15. Roberts, A.J. and Wiedmann, M., Pathogen, host and environmental factors contributing to the pathogenesis of listeriosis, *Cell Mol. Life Sci.*, 60, 904, 2003.
16. Vaneechoutte, M. et al., Comparison of PCR-based DNA fingerprinting techniques for the identification of *Listeria* species and their use for atypical *Listeria* isolates, *Int. J. Syst. Bacteriol.*, 48, 127, 1998.
17. Hain, T., Steinweg, C., and Chakraborty, T., Comparative and functional genomics of *Listeria* spp., *J. Biotechnol.*, 126, 37, 2006.
18. Nelson, K.E. et al., Whole genome comparisons of serotype 4b and 1/2a strains of the food-borne pathogen *Listeria monocytogenes* reveal new insights into the core genome components of this species, *Nucl. Acids Res.*, 32, 2386, 2004.
19. Glaser, P. et al., Comparative genomics of *Listeria* species, *Science*, 294, 849, 2001.
20. Hain, T. et al., Whole-genome sequence of *Listeria welshimeri* reveals common steps in genome reduction with *Listeria innocua* as compared to *Listeria monocytogenes*, *J. Bacteriol.*, 188, 7405, 2006.
21. Fraser, C.M. et al., The value of complete microbial genome sequencing (you get what you pay for), *J. Bacteriol.*, 184, 6403, 2002.
22. Demchick, P. and Koch, A.L., The permeability of the wall fabric of *Escherichia coli* and *Bacillus subtilis*, *J. Bacteriol.*, 178, 768, 1996.
23. Jeffery, C.J., Moonlighting proteins, *Trends Biochem. Sci.*, 24, 8, 1999.
24. Schaumburg, J. et al., The cell wall subproteome of *Listeria monocytogenes*, *Proteomics*, 4, 2991, 2004.
25. Gupta, R.S., Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes, *Microbiol. Mol. Biol. Rev.*, 62, 1435, 1998.

26. Henderson, I.R. et al., Type V protein secretion pathway: The autotransporter story, *Microbiol. Mol. Biol. Rev.*, 68, 692, 2004.
27. Salmond, G.P. and Reeves, P.J., Membrane traffic wardens and protein secretion in Gram-negative bacteria, *Trends Biochem. Sci.*, 18, 7, 1993.
28. Henderson, I.R. et al., Renaming protein secretion in the Gram-negative bacteria, *Trends Microbiol.*, 8, 352, 2000.
29. Economou, A. et al., Secretion by numbers: Protein traffic in prokaryotes, *Mol. Microbiol.*, 62, 308, 2006.
30. Tjalsma, H. et al., Signal peptide-dependent protein transport in *Bacillus subtilis*: A genome-based survey of the secretome, *Microbiol. Mol. Biol. Rev.*, 64, 515, 2000.
31. Pallen, M.J., Chaudhuri, R.R., and Henderson, I.R., Genomic analysis of secretion systems, *Curr. Opin. Microbiol.*, 6, 519, 2003.
32. Tjalsma, H. et al., Proteomics of protein secretion by *Bacillus subtilis*: Separating the "secrets" of the secretome, *Microbiol. Mol. Biol. Rev.*, 68, 207, 2004.
33. Desvaux, M. et al., Genomic analysis of the protein secretion systems in *Clostridium acetobutylicum* ATCC824, *Biochim. Biophys. Acta-Mol. Cell Res.*, 1745, 223, 2005.
34. Busch, W. and Saier, M.H., Jr., The transporter classification (TC) system, *Crit. Rev. Biochem. Mol. Biol.*, 37, 287, 2002.
35. Ajouz, B. et al., Release of thioredoxin via the mechanosensitive channel MscL during osmotic downshock of *Escherichia coli* cells, *J. Biol. Chem.*, 273, 26670, 1998.
36. Kachlany, S.C. et al., Nonspecific adherence by *Actinobacillus actinomycetemcomitans* requires genes widespread in *Bacteria* and *Archaea*, *J. Bacteriol.*, 182, 6169, 2000.
37. Desvaux, M. and Hébraud, M., The protein secretion systems in *Listeria*: Inside-out bacterial virulence, *FEMS Microbiol. Rev.*, 30, 774, 2006.
38. Desvaux, M. et al., The general secretory pathway: A general misnomer? *Trends Microbiol.*, 12, 306, 2004.
39. Cao, T.B. and Saier, M.H., Jr., The general protein secretory pathway: Phylogenetic analyses leading to evolutionary conclusions, *Biochim. Biophys. Acta-Biomembr.*, 1609, 115, 2003.
40. Valent, Q.A. et al., The *Escherichia coli* SRP and SecB targeting pathways converge at the translocon, *EMBO J.*, 17, 2504, 1998.
41. Van Wely, K.H.M. et al., Translocation of proteins across the cell envelope of Gram-positive bacteria, *FEMS Microbiol. Rev.*, 25, 437, 2001.
42. de Keyzer, J., van der Does, C., and Driessen, A.J., The bacterial translocase: A dynamic protein channel complex, *Cell Mol. Life Sci.*, 60, 2034, 2003.
43. Economou, A. et al., SecA membrane cycling at SecYEG is driven by distinct ATP binding and hydrolysis events and is regulated by SecD and SecF, *Cell*, 83, 1171, 1995.
44. Lenz, L.L. and Portnoy, D.A., Identification of a second *Listeria secA* gene associated with protein secretion and the rough phenotype, *Mol. Microbiol.*, 45, 1043, 2002.
45. Bensing, B.A. and Sullam, P.M., An accessory *sec* locus of *Streptococcus gordonii* is required for export of the surface protein GspB and for normal levels of binding to human platelets, *Mol. Microbiol.*, 44, 1081, 2002.
46. Lenz, L.L. et al., SecA2-dependent secretion of autolytic enzymes promotes *Listeria monocytogenes* pathogenesis, *Proc. Natl. Acad. Sci. USA*, 100, 12432, 2003.
47. Dramsi, S. et al., FbpA, a novel multifunctional *Listeria monocytogenes* virulence factor, *Mol. Microbiol.*, 53, 639, 2004.
48. Froderberg, L. et al., Versatility of inner membrane protein biogenesis in *Escherichia coli*, *Mol. Microbiol.*, 47, 1015, 2003.
49. Tjalsma, H., Bron, S., and van Dijl, J.M., Complementary impact of paralogous Oxa1-like proteins of *Bacillus subtilis* on post-translocational stages in protein secretion, *J. Biol. Chem.*, 278, 15622, 2003.
50. Madden, J.C., Ruiz, N., and Caparon, M., Cytolysin-mediated translocation (CMT): A functional equivalent of Type III secretion in Gram-positive bacteria, *Cell*, 104, 143, 2001.
51. Desvaux, M. et al., Type III secretion: what's in a name? *Trends Microbiol.*, 14, 157, 2006.
52. Bonnemain, C. et al., Differential roles of multiple signal peptidases in the virulence of *Listeria monocytogenes*, *Mol. Microbiol.*, 51, 1251, 2004.
53. Raynaud, C. and Charbit, A., Regulation of expression of type I signal peptidases in *Listeria monocytogenes*, *Microbiology*, 151, 3769, 2005.

54. Reglier-Poupet, H. et al., Maturation of lipoproteins by type II signal peptidase is required for phagosomal escape of *Listeria monocytogenes*, *J. Biol. Chem*., 278, 49469, 2003.
55. Navarre, W.W. and Schneewind, O., Surface proteins of Gram-positive bacteria and mechanisms of their targeting to the cell wall envelope, *Microbiol. Mol. Biol. Rev.*, 63, 174, 1999.
56. Mazmanian, S.K., Ton-That, H., and Schneewind, O., Sortase-catalyzed anchoring of surface proteins to the cell wall of *Staphylococcus aureus*, *Mol. Microbiol*., 40, 1049, 2001.
57. Dramsi, S., Trieu-Cuot, P., and Bierne, H., Sorting sortases: A nomenclature proposal for the various sortases of Gram-positive bacteria, *Res. Microbiol*., 156, 289, 2005.
58. Pucciarelli, M.G. et al., Identification of substrates of the *Listeria monocytogenes* sortases A and B by a nongel proteomic analysis, *Proteomics*, 5, 4808, 2005.
59. Bierne, H. et al., Inactivation of the *srtA* gene in *Listeria monocytogenes* inhibits anchoring of surface proteins and affects virulence, *Mol. Microbiol*., 43, 869, 2002.
60. Lee, S.G., Pancholi, V., and Fischetti, V.A. , Characterization of unique glycosylated anchor endopeptidase that cleaves the LPXTG sequence motif of cell surface proteins of Gram-positive bacteria, *J. Biol. Chem*., 277, 46912, 2002.
61. Lee, S.G. and Fischetti, V.A., Purification and characterization of LPXTGase from *Staphylococcus aureus*: The amino acid composition mirrors that found in the peptidoglycan, *J. Bacteriol*., 188, 389, 2006.
62. Kleinkauf, H. and von Dohren, H., The nonribosomal peptide biosynthetic system—On the origins of structural diversity of peptides, cyclopeptides and related compounds, *Antonie Van Leeuwenhoek*, 67, 229, 1995.
63. Fekkes, P. and Driessen, A.J., Protein targeting to the bacterial cytoplasmic membrane, *Microbiol. Mol. Biol. Rev.*, 63, 161, 1999.
64. von Heijne, G., A new method for predicting signal sequence cleavage sites, *Nucleic Acids Res*., 14, 4683, 1986.
65. Menne, K.M., Hermjakob, H., and Apweiler, R., A comparison of signal sequence prediction methods using a test set of signal peptides, *Bioinformatics*, 16, 741, 2000.
66. Nielsen, H. et al., Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites, *Protein Eng.*, 10, 1, 1997.
67. Bendtsen, J.D. et al., Improved prediction of signal peptides: SignalP 3.0, *J. Mol. Biol*., 340, 783, 2004.
68. Zhang, Z. and Henzel, W.J., Signal peptide prediction based on analysis of experimentally verified cleavage sites, *Protein Sci*., 13, 2819, 2004.
69. Hiller, K. et al., PrediSi: Prediction of signal peptides and their cleavage positions, *Nucleic Acids Res*., 32, W375, 2004.
70. Gomi, M., Sonoyama, M., and Mitaku, S., High performance system for signal peptide prediction: SOSUIsignal, *Chem-Bio Info. J*., 4, 142, 2004.
71. Käll, L., Krogh, A., and Sonnhammer, E.L., A combined transmembrane topology and signal peptide prediction method, *J. Mol. Biol*., 338, 1027, 2004.
72. Cai, Y.D., Lin, S.L., and Chou, K.C., Support vector machines for prediction of protein signal sequences and their cleavage sites, *Peptides*, 24, 159, 2003.
73. Gardy, J.L. et al., PSORT-B: Improving protein subcellular localization prediction for Gram-negative bacteria, *Nucleic Acids Res*., 31, 3613, 2003.
74. Gardy, J.L. et al., PSORTb v.2.0: Expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis, *Bioinformatics*, 21, 617, 2005.
75. Altschul, S.F. et al., Basic local alignment search tool, *J. Mol. Biol*., 215, 403, 1990.
76. Choo, K.H., Tan, T.W., and Ranganathan, S., SPdb—A signal peptide database, *BMC Bioinformatics*, 6, 249, 2005.
77. Reinhardt, A. and Hubbard, T., Using neural networks for prediction of the subcellular location of proteins, *Nucleic Acids Res*., 26, 2230, 1998.
78. Hua, S. and Sun, Z., Support vector machine approach for protein subcellular localization prediction, *Bioinformatics*, 17, 721, 2001.
79. Yu, C.S., Lin, C.J., and Hwang, J.K., Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions, *Protein Sci*., 13, 1402, 2004.
80. Yu, C.S. et al., Prediction of protein subcellular localization, *Proteins*, 64, 643, 2006.
81. Nair, R. and Rost, B., Mimicking cellular sorting improves prediction of subcellular localization, *J. Mol. Biol*., 348, 85, 2005.

82. Dönnes, P. and Höglund, A., Predicting protein subcellular localization: Past, present, and future, *Geno. Prot. Bioinfo.*, 2, 209, 2004.
83. Lu, Z. et al., Predicting subcellular localization of proteins using machine-learned classifiers, *Bioinformatics*, 20, 547, 2004.
84. Gardy, J.L. and Brinkman, F.S., Methods for predicting bacterial protein subcellular localization, *Nat. Rev. Microbiol.*, 4, 741, 2006.
85. Shen, H.B. and Chou, K.C., Gpos-PLoc: An ensemble classifier for predicting subcellular localization of Gram-positive bacterial proteins, *Protein Eng. Des. Sel.*, 20, 39, 2007.
86. Guo, T. et al., DBSubLoc: Database of protein subcellular localization, *Nucleic Acids Res.*, 32, D122, 2004.
87. Rey, S. et al., PSORTdb: A protein subcellular localization database for bacteria, *Nucleic Acids Res.*, 33, D164, 2005.
88. Lu, P. et al., PA-GOSUB: A searchable database of model organism protein sequences with their predicted gene ontology molecular function and subcellular localization, *Nucleic Acids Res.*, 33, D147, 2005.
89. Billion, A. et al., Augur—A computational pipeline for whole genome microbial surface protein prediction and classification, *Bioinformatics*, 22, 2819, 2006.
90. Trost, M. et al., Comparative proteome analysis of secretory proteins from pathogenic and nonpathogenic *Listeria* species, *Proteomics*, 5, 1544, 2005.
91. Berks, B.C., Palmer, T., and Sargent, F., Protein targeting by the bacterial twin-arginine translocation (Tat) pathway, *Curr. Opin. Microbiol.*, 8, 174, 2005.
92. Müller, M., Twin-arginine-specific protein export in *Escherichia coli*, *Res. Microbiol.*, 156, 131, 2005.
93. Lee, P.A., Tullman-Ercek, D., and Georgiou, G., The bacterial twin-arginine translocation pathway, *Annu. Rev. Microbiol.*, 60, 373, 2006.
94. Mori, H. and Cline, K., Post-translational protein translocation into thylakoids by the Sec and ΔpH-dependent pathways, *Biochim. Biophys. Acta-Mol. Cell Res.*, 1541, 80, 2001.
95. Bruser, T. and Sanders, C., An alternative model of the twin arginine translocation system, *Microbiol. Res.*, 158, 7, 2003.
96. Sargent, F., Berks, B.C., and Palmer, T., Pathfinders and trailblazers: A prokaryotic targeting system for transport of folded proteins, *FEMS Microbiol. Lett.*, 254, 198, 2006.
97. Alami, M. et al., Differential interactions between a twin-arginine signal peptide and its translocase in *Escherichia coli*, *Mol. Cell*, 12, 937, 2003.
98. Gohlke, U. et al., The TatA component of the twin-arginine protein transport system forms channel complexes of variable diameter, *Proc. Natl. Acad. Sci. USA*, 102, 10482, 2005.
99. Dilks, K. et al., Prokaryotic utilization of the twin-arginine translocation pathway: A genomic survey, *J. Bacteriol.*, 185, 1478, 2003.
100. Rose, R.W. et al., Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway, *Mol. Microbiol.*, 45, 943, 2002.
101. Bendtsen, J.D. et al., Prediction of twin-arginine signal peptides, *BMC Bioinformatics*, 6, 167, 2005.
102. Yen, M.R. et al., Sequence and phylogenetic analyses of the twin-arginine targeting (Tat) protein export system, *Arch. Microbiol.*, 177, 441, 2002.
103. Ize, B., Gerard, F., and Wu, L.F., *In vivo* assessment of the Tat signal peptide specificity in *Escherichia coli*, *Arch. Microbiol.*, 178, 548, 2002.
104. Ignatova, Z. et al., Unusual signal peptide directs penicillin amidase from *Escherichia coli* to the Tat translocation machinery, *Biochem. Biophys. Res. Commun.*, 291, 146, 2002.
105. Robinson, C. and Bolhuis, A., Tat-dependent protein targeting in prokaryotes and chloroplasts, *Biochim. Biophys. Acta-Mol. Cell Res.*, 1694, 135, 2004.
106. Peabody, C.R. et al., Type II protein secretion and its relationship to bacterial Type 4 pili and archaeal flagella, *Microbiology*, 149, 3051, 2003.
107. Chung, Y.S. and Dubnau, D., All seven *comG* open reading frames are required for DNA binding during transformation of competent *Bacillus subtilis*, *J. Bacteriol.*, 180, 41, 1998.
108. Claverys, J.P. and Martin, B., Bacterial "competence" genes: Signatures of active transformation, or only remnants? *Trends Microbiol.*, 11, 161, 2003.
109. Chen, I. and Dubnau, D., DNA uptake during bacterial transformation, *Nat. Rev. Microbiol.*, 2, 241, 2004.
110. Dubnau, D., Binding and transport of transforming DNA by *Bacillus subtilis*: The role of Type-4 pilin-like proteins—A review, *Gene*, 192, 191, 1997.

111. Filloux, A., The underlying mechanisms of Type II protein secretion, *Biochim. Biophys. Acta*, 1694, 163, 2004.
112. Chung, Y.S. and Dubnau, D., ComC is required for the processing and translocation of comGC, a pilin-like competence protein of *Bacillus subtilis*, *Mol. Microbiol.*, 15, 543, 1995.
113. LaPointe, C.F. and Taylor, R.K., The type 4 prepilin peptidases comprise a novel family of aspartic acid proteases, *J. Biol. Chem.*, 275, 1502, 2000.
114. Chen, I., Provvedi, R., and Dubnau, D., A macromolecular complex formed by a pilin-like protein in competent *Bacillus subtilis*, *J. Biol. Chem.*, 281, 21720, 2006.
115. de Castro, E. et al., ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins, *Nucleic Acids Res.*, 34, W362, 2006.
116. Peel, M., Donachie, W., and Shaw, A., Temperature-dependent expression of flagella of *Listeria monocytogenes* studied by electron microscopy, SDS-PAGE and western blotting, *J. Gen. Microbiol.*, 134, 2171, 1988.
117. Way, S.S. et al., Characterization of flagellin expression and its role in *Listeria monocytogenes* infection and immunity, *Cell Microbiol.*, 6, 235, 2004.
118. Sanchez-Campillo, M. et al., Modulation of DNA topology by *flaR*, a new gene from *Listeria monocytogenes*, *Mol. Microbiol.*, 18, 801, 1995 .
119. Michel, E. et al., Characterization of a large motility gene cluster containing the *cheR*, *motAB* genes of *Listeria monocytogenes* and evidence that PrfA down-regulates motility genes, *FEMS Microbiol. Lett.*, 169, 341, 1998.
120. Knudsen, G.M., Olsen, J.E., and Dons, L., Characterization of DegU, a response regulator in *Listeria monocytogenes*, involved in regulation of motility and contributes to virulence, *FEMS Microbiol. Lett.*, 240, 171, 2004.
121. Grundling, A. et al., *Listeria monocytogenes* regulates flagellar motility gene expression through MogR, a transcriptional repressor required for virulence, *Proc. Natl. Acad. Sci. USA*, 101, 12318, 2004.
122. Shen, A. et al., A bifunctional O-GlcNAc transferase governs flagellar motility through antirepression, *Genes Dev.*, 20, 3283, 2006.
123. Schirm, M. et al., Flagellin from *Listeria monocytogenes* is glycosylated with β-O-linked N-acetylglucosamine, *J. Bacteriol.*, 186, 6721, 2004.
124. Lemon, K.P., Higgins, D.E., and Kolter, R., Flagellar motility is critical for *Listeria monocytogenes* biofilm formation, *J. Bacteriol.*, 189, 4418, 2007.
125. O'Neil, H.S. and Marquis, H., *Listeria monocytogenes* flagella are used for motility, not as adhesins, to increase host cell invasion, *Infect. Immun.,* 74, 6675, 2006.
126. Popowska, M., Analysis of the peptidoglycan hydrolases of *Listeria monocytogenes*: Multiple enzymes with multiple functions, *Pol. J. Microbiol.*, 53, 29, 2004.
127. Popowska, M. and Markiewicz, Z., Murein-hydrolyzing activity of flagellin FlaA of *Listeria monocytogenes*, *Pol. J. Microbiol.*, 53, 237, 2004.
128. Macnab, R.M., How bacteria assemble flagella, *Annu. Rev. Microbiol.*, 57, 77, 2003.
129. Aizawa, S.I., Bacterial flagella and Type III secretion systems, *FEMS Microbiol. Lett.*, 202, 157, 2001.
130. Journet, L., Hughes, K.T., and Cornelis, G.R., Type III secretion: A secretory pathway serving both motility and virulence, *Mol. Membr. Biol.* 22, 41, 2005.
131. Macnab, R.M., Type III flagellar protein export and flagellar assembly, *Biochim. Biophys. Acta-Mol. Cell Res.*, 1694, 207, 2004.
132. Minamino, T. and Namba, K., Self-assembly and Type III protein export of the bacterial flagellum, *J. Mol. Microbiol. Biotechnol.*, 7, 5, 2004.
133. Troisfontaines, P. and Cornelis, G.R., Type III secretion: More systems than you think, *Physiology (Bethesda)*, 20, 326, 2005.
134. Young, G.M., Schmiel, D.H., and Miller, V.L., A new pathway for the secretion of virulence factors by bacteria, the flagellar export apparatus functions as a protein-secretion system, *Proc. Natl. Acad. Sci. USA*, 96, 6456, 1999.
135. Cornelis, G.R., The Type III secretion injectisome, *Nat. Rev. Microbiol.*, 4, 811, 2006.
136. Ghelardi, E. et al., Requirement of *flhA* for swarming differentiation, flagellin export, and secretion of virulence-associated proteins in *Bacillus thuringensis*, *J. Bacteriol.*, 184, 6424, 2002.
137. Wang, I.N., Smith, D.L., and Young, R., Holins: The protein clocks of bacteriophage infections, *Annu. Rev. Microbiol.*, 54, 799, 2000.
138. Bayles, K.W., Are the molecular strategies that control apoptosis conserved in bacteria? *Trends Microbiol.*, 11, 306, 2003.

139. Young, R. and Blasi, U., Holins: Form and function in bacteriophage lysis, *FEMS Microbiol. Rev.*, 17, 191, 1995.
140. Bernhardt, T.G. et al., Breaking free: "Protein antibiotics" and phage lysis, *Res. Microbiol.*, 153, 493, 2002.
141. Loessner, M.J., Bacteriophage endolysins—Current state of research and applications, *Curr. Opin. Microbiol.*, 8, 480, 2005.
142. Ziedaite, G. et al., The holin protein of bacteriophage PRD1 forms a pore for small-molecule and endolysin translocation, *J. Bacteriol.*, 187, 5397, 2005.
143. Bläsi, U. and Young, R., Two beginnings for a single purpose: The dual-start holins in the regulation of phage lysis, *Mol. Microbiol.*, 21, 675, 1996.
144. Groicher, K.H. et al., The *Staphylococcus aureus* lrgAB operon modulates murein hydrolase activity and penicillin tolerance, *J. Bacteriol.*, 182, 1794, 2000.
145. Rice, K.C. et al., The *Staphylococcus aureus cidAB* operon: Evaluation of its role in regulation of murein hydrolase activity and penicillin tolerance, *J. Bacteriol.*, 185, 2635, 2003.
146. Saier, M.H., Jr., Tran, C.V., and Barabote, R.D., TC-DB: The transporter classification database for membrane transport protein analyses and information, *Nucleic Acids Res.*, 34, D181–D186, 2006.
147. Young, R., Bacteriophage holins: Deadly diversity, *J. Mol. Microbiol. Biotechnol.*, 4, 21, 2002.
148. Ramanculov, E. and Young, R., Functional analysis of the phage T4 holin in a lambda context, *Mol. Genet. Genomics*, 265, 345, 2001.
149. Loessner, M.J. et al., Complete nucleotide sequence, molecular analysis and genome structure of bacteriophage A118 of *Listeria monocytogenes*: Implications for phage evolution, *Mol. Microbiol.*, 35, 324, 2000.
150. Vukov, N. et al., Functional regulation of the *Listeria monocytogenes* bacteriophage A118 holin by an intragenic inhibitor lacking the first transmembrane domain, *Mol. Microbiol.*, 48, 173, 2003.
151. Loessner, M.J., Wendlinger, G., and Scherer, S., Heterogeneous endolysins in *Listeria monocytogenes* bacteriophages: A new class of enzymes and evidence for conserved holin genes within the siphoviral lysis cassettes, *Mol. Microbiol.*, 16, 1231, 1995.
152. Tan, K.S., Wee, B.Y., and Song, K.P., Evidence for holin function of *tcdE* gene in the pathogenicity of *Clostridium difficile*, *J. Med. Microbiol.*, 50, 613, 2001.
153. Mukherjee, K. et al., Proteins released during high toxin production in *Clostridium difficile*, *Microbiology*, 148, 2245, 2002.
154. Pallen, M.J., The ESAT-6/WXG100 superfamily—and a new Gram-positive secretion system? *Trends Microbiol.*, 10, 209, 2002.
155. Converse, S.E. and Cox, J.S., A protein secretion pathway critical for *Mycobacterium tuberculosis* virulence is conserved and functional in *Mycobacterium smegmatis*, *J. Bacteriol.*, 187, 1238, 2005.
156. Champion, P.A. et al., C-terminal signal sequence promotes virulence factor secretion in *Mycobacterium tuberculosis*, *Science*, 313, 1632, 2006.
157. Stanley, S.A. et al., Acute infection and macrophage subversion by *Mycobacterium tuberculosis* require a specialized secretion system, *Proc. Natl. Acad. Sci. USA*, 100, 13001, 2003.
158. Okkels, L.M. and Andersen, P., Protein–protein interactions of proteins from the ESAT-6 family of *Mycobacterium tuberculosis*, *J. Bacteriol.*, 186, 2487, 2004.
159. Renshaw, P.S. et al., Conclusive evidence that the major T-cell antigens of the *Mycobacterium tuberculosis* complex ESAT-6 and CFP-10 form a tight, 1:1 complex and characterization of the structural properties of ESAT-6, CFP-10, and the ESAT-6*CFP-10 complex. Implications for pathogenesis and virulence, *J. Biol. Chem.*, 277, 21598, 2002.
160. Burts, M.L. et al., EsxA and EsxB are secreted by an ESAT-6-like system that is required for the pathogenesis of *Staphylococcus aureus* infections, *Proc. Natl. Acad. Sci. USA*, 102, 1169, 2005.
161. Guinn, K.M. et al., Individual RD1-region genes are required for export of ESAT-6/CFP-10 and for virulence of *Mycobacterium tuberculosis*, *Mol. Microbiol.*, 51, 359, 2004.
162. Brodin, P. et al., ESAT-6 proteins: Protective antigens and virulence factors? *Trends Microbiol.*, 12, 500, 2004.
163. Way, S.S. and Wilson, C.B., The *Mycobacterium tuberculosis* ESAT-6 homologue in *Listeria monocytogenes* is dispensable for growth *in vitro* and *in vivo*, *Infect. Immun.*, 73, 6151, 2005.
164. Shockman, G.D. and Barrett, J.F., Structure, function, and assembly of cell walls of Gram-positive bacteria, *Annu. Rev. Microbiol.*, 37, 501, 1983.
165. Archibald, A.R., Structure and assembly of the cell wall in *Bacillus subtilis*, *Biochem. Soc. Trans.*, 13, 990, 1985.

166. Schäffer, C. and Messner, P., The structure of secondary cell wall polymers: How Gram-positive bacteria stick their cell walls together, *Microbiology*, 151, 643, 2005.
167. Sára, M. and Sleytr, U.B., S-Layer proteins, *J. Bacteriol.*, 182, 859, 2000.
168. Klein, C. et al., The membrane proteome of *Halobacterium salinarum*, *Proteomics*, 5, 180, 2005.
169. Harris, M.A. et al., The Gene Ontology (GO) database and informatics resource, *Nucleic Acids Res.*, 32, D258, 2004.
170. Samuelson, J.C. et al., YidC mediates membrane protein insertion in bacteria, *Nature*, 406, 637, 2000.
171. Luirink, J. et al., Biogenesis of inner membrane proteins in *Escherichia coli*, *Annu. Rev. Microbiol.*, 59, 329, 2005.
172. Dalbey, R.E. and Kuhn, A., YidC family members are involved in the membrane insertion, lateral integration, folding, and assembly of membrane proteins, *J. Cell Biol.*, 166, 769, 2004.
173. Van Bloois, E. et al., Distinct requirements for translocation of the N-tail and C-tail of the *Escherichia coli* inner membrane protein CyoA, *J. Biol. Chem.*, 281, 10002, 2006.
174. Van der Laan, M., Nouwen, N.P., and Driessen, A.J., YidC—An evolutionary conserved device for the assembly of energy-transducing membrane protein complexes, *Curr. Opin. Microbiol.*, 8, 182, 2005.
175. de Gier, J.W. and Luirink, J., The ribosome and YidC. New insights into the biogenesis of *Escherichia coli* inner membrane proteins, *EMBO Rep.*, 4, 939, 2003.
176. Dalbey, R.E. and Chen, M., Sec-translocase mediated membrane protein biogenesis, *Biochim. Biophys. Acta-Mol. Cell Res.*, 1694, 37, 2004.
177. von Heijne, G., Membrane-protein topology, *Nat. Rev. Mol. Cell Biol.*, 7, 909, 2006.
178. White, S.H. and von Heijne, G., The machinery of membrane protein assembly, *Curr. Opin. Struct. Biol.*, 14, 397, 2004.
179. Goder, V. and Spiess, M., Topogenesis of membrane proteins: determinants and dynamics, *FEBS Lett*, 504, 87, 2001.
180. Kocks, C. et al., *L. monocytogenes*-induced actin assembly requires the *actA* gene product, a surface protein, *Cell*, 68, 521, 1992.
181. Borezee, E. et al., SvpA, a novel surface virulence-associated protein required for intracellular survival of *Listeria monocytogenes*, *Microbiology*, 147, 2913, 2001.
182. Cabanes, D. et al., Surface proteins and the pathogenic potential of *Listeria monocytogenes*, *Trends Microbiol.*, 10, 238, 2002.
183. Froderberg, L. et al., Targeting and translocation of two lipoproteins in *Escherichia coli* via the SRP/Sec/YidC pathway, *J. Biol. Chem.*, 279, 31026, 2004.
184. Tjalsma, H. et al., The role of lipoprotein processing by signal peptidase II in the Gram-positive eubacterium *Bacillus subtilis*. Signal peptidase II is required for the efficient secretion of α-amylase, a nonlipoprotein, *J. Biol. Chem.*, 274, 1698, 1999.
185. Hayashi, S. and Wu, H.C., Lipoproteins in bacteria, *J. Bioenerg. Biomembr.*, 22, 451, 1990.
186. Baumgärtner, M. et al., Inactivation of Lgt allows systematic characterization of lipoproteins from *Listeria monocytogenes*, *J. Bacteriol.*, 189, 313, 2007.
187. Sankaran, K. and Wu, H.C., Lipid modification of bacterial prolipoprotein. Transfer of diacylglyceryl moiety from phosphatidylglycerol, *J. Biol. Chem.*, 269, 19701, 1994.
188. Garcia-Del Portillo, F. and Cossart, P., An important step in *Listeria* lipoprotein research, *J. Bacteriol.*, 189, 294, 2007.
189. Hulo, N. et al., The PROSITE database, *Nucleic Acids Res.*, 34, D227, 2006.
190. Sutcliffe, I. and Harrington, D.J., Pattern searches for the identification of putative lipoprotein genes in Gram-positive bacterial genomes, *Microbiology*, 148, 2065, 2002.
191. Babu, M.M. and Sankaran, K., DOLOP-database of bacterial lipoproteins, *Bioinformatics*, 18, 641, 2002.
192. Juncker, A.S. et al., Prediction of lipoprotein signal peptides in Gram-negative bacteria, *Protein Sci.*, 12, 1652, 2003.
193. Fariselli, P., Finocchiaro, G., and Casadio, R., SPEPlip: The detection of signal peptide and lipoprotein cleavage sites, *Bioinformatics*, 19, 2498, 2003.
194. Taylor, P.D. et al., LipPred: A Web server for accurate prediction of lipoprotein signal sequences and cleavage sites, *Bioinformation*, 1, 335, 2006.
195. Sutcliffe, I.C. and Russell, R.R., Lipoproteins of Gram-positive bacteria, *J. Bacteriol.*, 177, 1123, 1995.
196. Sanderson, S., Campbell, D.J., and Shastri, N., Identification of a CD4+ T cell-stimulating antigen of pathogenic bacteria by expression cloning, *J. Exp. Med.*, 182, 1751, 1995.

197. Ko, R. and Smith, L.T., Identification of an ATP-driven, osmoregulated glycine betaine transport system in *Listeria monocytogenes*, *Appl. Environ. Microbiol.*, 65, 4040, 1199.
198. Fraser, K.R. et al., Identification and characterization of an ATP binding cassette L-carnitine transporter in *Listeria monocytogenes*, *Appl. Environ. Microbiol.*, 66, 4696, 2000.
199. Borezee, E., Pellegrini, E., and Berche, P., OppA of *Listeria monocytogenes*, an oligopeptide-binding protein required for bacterial growth at low temperature and involved in intracellular survival, *Infect. Immun.*, 68, 7069, 2000.
200. Reglier-Poupet, H. et al., Identification of LpeA, a PsaA-like membrane protein that promotes cell entry by *Listeria monocytogenes*, *Infect. Immun.*, 71, 474, 2003.
201. Gal, L. et al., CelG from *Clostridium cellulolyticum*: A multidomain endoglucanase acting efficiently on crystalline cellulose, *J. Bacteriol.*, 179, 6595, 1997.
202. Gaudin, C. et al., CelE, a multidomain cellulase from *Clostridium cellulolyticum*: A key enzyme in the cellulosome? *J. Bacteriol.*, 182, 1910, 2000.
203. Brinster, S., Furlan, S., and Serror, P., C-terminal WXL domain mediates cell wall binding in *Enterococcus faecalis* and other Gram-positive bacteria, *J. Bacteriol.*, 189, 1244, 2007.
204. Popowska, M. and Markiewicz, Z., Classes and functions of *Listeria monocytogenes* surface proteins, *Pol. J. Microbiol.*, 53, 75, 2004.
205. Altschul, S.F. et al., Gapped BLAST and PSI-BLAST: A new generation of protein database search programs, *Nucleic Acids Res.*, 25, 3389, 1997.
206. Eddy, S.R., Hidden Markov models, *Curr. Opin. Struct. Biol.*, 6, 361, 1996.
207. Mulder, N.J. et al., New developments in the InterPro database, *Nucleic Acids Res.*, 35, D224, 2007.
208. Bateman, A. et al., The Pfam protein families database, *Nucleic Acids Res.*, 32, D138, 2004.
209. Schultz, J. et al., SMART, a simple modular architecture research tool: Identification of signaling domains, *Proc. Natl. Acad. Sci. USA*, 95, 5857, 1998.
210. Selengut, J.D. et al., TIGRFAMs and genome properties: Tools for the assignment of molecular function and biological process in prokaryotic genomes, *Nucleic Acids Res.*, 35, D260, 2007.
211. Wilson, D. et al., The SUPERFAMILY database in 2007: Families and functions, *Nucleic Acids Res.*, 35, D308, 2007.
212. Pallen, M.J. et al., An embarrassment of sortase—A richness of substrates? *Trends Microbiol.*, 9, 97, 2001.
213. Marraffini, L.A., Dedent, A.C., and Schneewind, O., Sortases and the art of anchoring proteins to the envelopes of Gram-positive bacteria, *Microbiol. Mol. Biol. Rev.*, 70, 192, 2006.
214. Paterson, G.K. and Mitchell, T.J., The biology of Gram-positive sortase enzymes, *Trends Microbiol.*, 12, 89, 2004.
215. Ton-That, H., Marraffini, L.A., and Schneewind, O., Protein sorting to the cell wall envelope of Gram-positive bacteria, *Biochim. Biophys. Acta-Mol. Cell Res.*, 1694, 269, 2004.
216. Bierne, H. et al., Sortase B, a new class of sortase in *Listeria monocytogenes*, *J. Bacteriol.*, 186, 1972, 2004.
217. Garandeau, C. et al., The sortase SrtA of *Listeria monocytogenes* is involved in processing of internalin and in virulence, *Infect. Immun.*, 70, 1382, 2002.
218. Steen, A. et al., Cell wall attachment of a widely distributed peptidoglycan binding domain is hindered by cell wall constituents, *J. Biol. Chem.*, 278, 23874, 2003.
219. Dussurget, O., Pizarro-Cerda, J., and Cossart, P., Molecular determinants of *Listeria monocytogenes* virulence, *Annu. Rev. Microbiol.*, 58, 587, 2004.
220. Gaillard, J.L. et al., Entry of *L. monocytogenes* into cells is mediated by internalin, a repeat protein reminiscent of surface antigens from Gram-positive cocci, *Cell*, 65, 1127, 1991.
221. Braun, L. et al., InlB: An invasion protein of *Listeria monocytogenes* with a novel type of surface association, *Mol. Microbiol.*, 25, 285, 1997.
222. Jonquieres, R. et al., Interaction between the protein InlB of *Listeria monocytogenes* and lipoteichoic acid: A novel mechanism of protein association at the surface of Gram-positive bacteria, *Mol. Microbiol.*, 34, 902, 1999.
223. Marino, M. et al., GW domains of the *Listeria monocytogenes* invasion protein InlB are SH3-like and mediate binding to host ligands, *EMBO J.*, 21, 5623, 2002.
224. Milohanic, E. et al., Transcriptome analysis of *Listeria monocytogenes* identifies three groups of genes differently regulated by PrfA, *Mol. Microbiol.*, 47, 1613, 2003.

225. Siezen, R. et al., *Lactobacillus plantarum* gene clusters encoding putative cell-surface protein complexes for carbohydrate utilization are conserved in specific Gram-positive bacteria, *BMC Genomics*, 7, 126, 2006.
226. Nölling, J. et al., Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*, *J. Bacteriol.*, 183, 4823, 2001.
227. Baba, T. and Schneewind, O., Target cell specificity of a bacteriocin molecule: A C-terminal signal directs lysostaphin to the cell wall of *Staphylococcus aureus*, *EMBO J.*, 15, 4789, 1996.
228. Anantharaman, V. and Aravind, L., Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes, *Genome Biol.*, 4, R11, 2003.
229. Scott, J.R. and Barnett, T.C., Surface proteins of Gram-positive bacteria and how they get there, *Annu. Rev. Microbiol.*, 60, 397, 2006.
230. Bendtsen, J.D. et al., Nonclassical protein secretion in bacteria, *BMC Microbiol.*, 5, 58, 2005.
231. Guiral, S. et al., Competence-programmed predation of noncompetent cells in the human pathogen *Streptococcus pneumoniae*: Genetic requirements, *Proc. Natl. Acad. Sci. USA*, 102, 8710, 2005.
232. Bergmann, S., Rohde, M., and Hammerschmidt, S., Glyceraldehyde-3-phosphate dehydrogenase of *Streptococcus pneumoniae* is a surface-displayed plasminogen-binding protein, *Infect. Immun.*, 72, 2416, 2004.
233. Klose, J., Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals, *Humangenetik*, 26, 231, 1975.
234. O'Farrell, P.H., High resolution two-dimensional electrophoresis of proteins, *J. Biol. Chem.*, 250, 4007, 1975.
235. Scheele, G.A., Two-dimensional gel analysis of soluble proteins. Characterization of guinea pig exocrine pancreatic proteins, *J. Biol. Chem.*, 250, 5375, 1975.
236. VanBogelen, R.A. et al., The gene-protein database of *Escherichia coli*: Edition 5, *Electrophoresis*, 13, 1014, 1992.
237. Pasquali, C. et al., Two-dimensional gel electrophoresis of *Escherichia coli* homogenates: The *Escherichia coli* SWISS-2DPAGE database, *Electrophoresis*, 17, 547, 1996.
238. Folio, P. et al., Two-dimensional electrophoresis database of *Listeria monocytogenes* EGDe proteome and proteomic analysis of mid-log and stationary growth phase cells, *Proteomics*, 4, 3187, 2004.
239. Cordwell, S.J., Technologies for bacterial surface proteomics, *Curr. Opin. Microbiol.*, 9, 320, 2006.
240. Santoni, V., Molloy, M., and Rabilloud, T., Membrane proteins and proteomics: Un amour impossible? *Electrophoresis*, 21, 1054, 2000.
241. Santoni, V. et al., Membrane proteomics: use of additive main effects with multiplicative interaction model to classify plasma membrane proteins according to their solubility and electrophoretic properties, *Electrophoresis*, 21, 3329, 2000.
242. Mastro, R. and Hall, M., Protein delipidation and precipitation by tri-n-butylphosphate, acetone, and methanol treatment for isoelectric focusing and two-dimensional gel electrophoresis, *Anal. Biochem.*, 273, 313, 1999.
243. Deshusses, J.M. et al., Exploitation of specific properties of trifluoroethanol for extraction and separation of membrane proteins, *Proteomics*, 3, 1418, 2003.
244. Planchon, S. et al., Proteomic analysis of cell envelope from *Staphylococcus xylosus* C2a, a coagulase-negative *Staphylococcus*, *J. Proteome Res.*, in press, 2007.
245. Kotiranta, A. et al., Surface structure, hydrophobicity, phagocytosis, and adherence to matrix proteins of *Bacillus cereus* cells with and without the crystalline surface protein layer, *Infect. Immun.*, 66, 4895, 1998.
246. Tresse, O. et al., Comparative evaluation of adhesion, surface properties, and surface protein composition of *Listeria monocytogenes* strains after cultivation at constant pH of 5 and 7, *J. Appl. Microbiol.*, 101, 53, 2006.
247. Calvo, E. et al., Analysis of the *Listeria* cell wall proteome by two-dimensional nanoliquid chromatography coupled to mass spectrometry, *Proteomics*, 5, 433, 2005.
248. Ren, Q., Kang, K.H., and Paulsen, I.T., TransportDB: A relational database of cellular membrane transport systems, *Nucl. Acids Res.*, 32, D284, 2004.
249. Hofmann, K., and Stoffel, W., PROFILEGRAPH: An interactive graphical tool for protein sequence analysis, *Comput. Appl. Biosci.,* 8, 331, 1992.ne
250. Claros, M.G., and von Heijne, G., TopPred II: An improved software for membrane protein structure predictions, *Comput. Appl. Biosci.,* 10, 685, 1994.
251. Rost, B. et al., Transmembrane helices predicted at 95% accuracy, *Protein Sci.,* 4, 521, 1995.

252. Csero, M. et al., Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: The dense alignment surface method, *Protein Eng.,* 10, 673, 1997.
253. Persson, B., and Argos, P., Prediction of membrane protein topology utilizing multiple sequence alignments, *J. Protein Chem.,* 16, 453, 1997.
254. Kihara, D., Shimizu, T., and Kanehisa, M., Prediction of membrane proteins based on classification of transmembrane segments, *Protein Eng.,* 11, 961, 1998.
255. Sonnhammer, E.L., von Heijne, G., and Krogh, A., A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.,* 6, 175, 1998.
256. Hirokawa, T., Boon-chieng, S., Mitaku, S., SOSUI: Classification and secondary structure prediction system for membrane proteins, *Bioinformatics,* 14, 378, 1998.
257. Pasquier, C., and Hamodrakas, S.J., An hierarchical artificial neural network system for the classification of transmembrane proteins, *Protein Eng.,* 12, 631, 1999.
258. Pilpel, Y., Ben-Tal, N., and Lancet, D., kPROT: A knowledge-based scale for the propensity of residue orientation in transmembrane segments. Application to membrane protein structure prediction, *J. Mol. Biol.,* 294, 921, 1999.
259. Senes, A., Gerstein, M., and Engelman, D.M., Statistical analysis of amino acid patterns in transmembrane helices: The GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions, *J. Mol. Biol.,* 296, 921, 2000.
260. Tusnady, G.E., and Simon, I., The HMMTOP transmembrane topology prediction server. *Bioinformatics,* 17, 849, 2001.
261. Deber, C.M. et al., TM Finder: A prediction program for transmembrane protein segments using a combination of hydrophobicity and nonpolar phase helicity scales, *Protein Sci.,* 10, 212, 2001.
262. Cserzo, M. et al., On filtering false positive transmembrane protein predictions, *Protein Eng.,* 15, 745, 2002.
263. Juretic, D., Zoranic, L., and Zucic, D., Basic charge clusters and predictions of membrane protein topology, *J. Chem. Inf. Comput. Sci.,* 42, 620, 2002.
264. Zhou, H. et al., Web-based toolkits for topology prediction of transmembrane helical proteins, fold recognition, structure and binding scoring, folding-kinetics analysis and comparative analysis of domain combinations, *Nucleic Acids Res.,* 33, W193, 2005.
265. Taylor, P.D., Attwood, T.K., and Flower, D.R., BPROMPT: A consensus server for membrane protein prediction, *Nucleic Acids Res.,* 31, 3698, 2003.
266. Yuan, Z., Mattick, J.S., and Teasdale, R.D., SVMtm: Support vector machines to predict transmembrane segments, *J. Comput. Chem.,* 25, 632, 2004.
267. Arai, M. et al., ConPred II: A consensus prediction method for obtaining transmembrane topology models with high reliability, *Nucleic Acids Res.,* 32, W390, 2004.
268. Bagos, P.G., Liakopoulos, T.D., and Hamodrakas, S.J., Algorithms for incorporating prior topological information in HMMs: Application to transmembrane proteins, *BMC Bioinform.,* 7, 189, 2006.
269. Cao, B. et al., Enhanced recognition of protein transmembrane domains with prediction-based structural profiles, *Bioinformatics*, 22, 303, 2006.
270. Jones, T.D., Improving the accuracy of transmembrane protein topology prediction using evolutionary information, *Bioinformatics,* 23, 538, 2007.