European Journal of Cell Biology 97 (2018) 422–441

Contents lists available at ScienceDirect

European Journal of Cell Biology

journal homepage: www.elsevier.com/locate/ejcb

Review

A comprehensive review of signal peptides: Structure, roles, and applications

Hajar Owji[a,b], Navid Nezafat[a,b], Manica Negahdaripour[a,b], Ali Hajiebrahimi[a,b], Younes Ghasemi[a,b,c,d,*]

[a] Pharmaceutical Sciences Research Center, Shiraz University of Medical Sciences, Shiraz, Iran
[b] Department of Pharmaceutical Biotechnology, School of Pharmacy, Shiraz University of Medical Sciences, Shiraz, Iran
[c] Department of Medical Biotechnology, School of Advanced Medical Sciences and Technologies, Shiraz University of Medical Sciences, Shiraz, Iran
[d] Biotechnology Research Center, Shiraz University of Medical Sciences, Shiraz, Iran

ARTICLE INFO

Keywords:
Signal peptide
Structure
Functions
Secretory pathways

ABSTRACT

Signal peptides (SP) are short peptides located in the N-terminal of proteins, carrying information for protein secretion. They are ubiquitous to all prokaryotes and eukaryotes. SPs have been of special interest in several scientific and industrial fields, including recombinant protein production, disease diagnosis, immunization, and laboratory techniques. Recently, the role of SPs in recombinant protein production has gained too much attention. Herein, several studies have been reviewed to elucidate the precise structure and function of SPs, particularly the optimized ones for recombinant protein production. However, some features of SPs still have remained obscure. In this review, some approaches concerning elucidation and optimization of SPs are discussed, and pragmatic conclusions and suggestions for future studies are also proposed. Moreover, a summary of secretory pathways, evolutionary changes, functions, applications, and different types of SPs is mentioned. At last, current limitations and prospects are discussed.

1. Introduction

A great majority of secretory proteins in all domains of life carry a short peptide at their N-terminals, called signal peptide (SP). SPs act as zip codes marking the protein secretion pathway as well as the protein target location. In addition to protein targeting, a number of critical functions with or without regard to the passenger proteins have been attributed to SPs. They have come handy in diverse fields from recombinant protein production to disease diagnosis and vaccination. Of special importance, they have been shown to be a promising tool in biotherapeutic production. Different computational and experimental studies have been carried out to elucidate SP features. Moreover, a number of methods and tools have been devised for exploring different features of SPs. This review delves into the literature of SP and provides the knowledge from the structural and functional point of view as well as signal peptide applications.

2. Discovery of signal peptides

In 1971, for the first time, Gunter Blobel and David Sabatini hypothesized the presence of intrinsic information for translocation of protein into the membrane (Blobel and Sabatini, 1971). In 1972, Milstein and his co-workers reported that the light chain of IgG becomes mature in the endoplasmic reticulum vesicles in myeloma cell (Milstein et al., 1972). This process takes place by cleaving the amino-terminal of nascent IgG. In 1975, Blobel postulated that proteins transferring to different parts of the cell bear varied targeting sequence. Afterwards, in 1983, the architecture of these targeting sequences, naming as N-terminal signal peptides (also leader peptide or leader sequence), was clarified by Heijne (von Heijne, 1983). Finally, in 1991 the Nobel prize

Abbreviations: SP, signal peptide; SPase, signal peptidase; ER, endoplasmic reticulum; LHCGR, human luteinizing hormone receptor; CADA, cyclotriazadisulfonamide; hCD4, human CD4; aa, amino acid; PMF, proton motive force; VSV-G, vesicular stomatitis virus G-protein; MBP, maltose-binding protein; ALP, alkaline phosphatase; Cel-CD, cellulase catalytic domain; RBP, ribose-binding peptide; GM-CSF, granulocyte-colony-stimulating factors; BEVS, baculovirus expression vector system; LCMV, lymphocytic choriomeningitis virus; TorA, TAMO reductase; OPH, organophosphorus hydrolase; Csn, chitosanase; pGH, porcine growth hormone; IFNα2, interferon alpha 2; PTHrP, parathyroid hormone related protein; TIR, translation initiation region; TOM, translocase of outer membrane; TIM, translocase of inner membrane; MPP, mitochondrial processing peptidase; MIP, mitochondrial intermediate peptidase; SPP, stroma processing peptidase; TPP, thylakoid processing peptides; SPC, signal peptidase complex; NK, natural killer cell; JEV-E, pseudotyped Japanese encephalitis virus envelope; PSM, phenol soluble modulin; CaM, calmodulin; ECP, eosinophil cationic protein; TGF-α, transforming growth factor-α; HAS, human serum albumin; DNP, dynamic nuclear polarization; MAS, magic angle spinning; NMR, nuclear magnetic resonance

* Corresponding author at: Department of Pharmaceutical Biotechnology, School of Pharmacy, Shiraz University of Medical Sciences, P.O. Box 71345-1583, Shiraz, Iran.
E-mail address: ghasemiy@sums.ac.ir (Y. Ghasemi).

https://doi.org/10.1016/j.ejcb.2018.06.003
Received 15 January 2018; Received in revised form 18 June 2018; Accepted 18 June 2018

Table 1
Roles and application of SPs.

Application	Type of SP	Evidence and Implication	Ref.
Recombinant protein production		• Increasing recombinant protein production to a commercially significant level • Preventing inclusion body formation and protein degradation by protease	(Ohmuro-Matsuyama and Yamaji, 2017) (Mergulhao et al., 2005)
Selection marker	SacB	Responsible for sugar intolerance of SacB as a selection marker in *E. coli*	(Gao et al., 2017)
Responsible for human diseases	Preparathyroid hormone	Mutation causes autosomal familial isolated hypoparathyroidism	(Datta et al., 2007)
	Preproinsulin	Mutations are associated with the onset of diabetes	(Bonfanti et al., 2009)
	LHCGR	Variations in SP affect receptor biogenesis causing leydig cell hypoplasia	(Vezzoli et al., 2015)
		Attributing 26 human diseases in 21 human proteins to mutations in SPs	(Jarjanazi et al., 2008)
Drug targets	hCD4	Targets of CADA, reducing expression of CD4 on thymocyte by disturbing SP structure, and protecting from HIV and SIV infections	(Vermeire et al., 2014)
Signal-exon trap		Method used to distinguish secretory/membrane proteins by SP	(Péterfy et al., 2000)
Gene therapy	IL-2	Increasing the serum level of therapeutic molecule by optimized SP	(Zhang et al., 2005)
Diagnostic biomarkers	Epidermal growth factor domain-containing protein-1	Levels of SP complement herald pulmonary embolism	(Dirican et al., 2016)
Acting as an intramolecular chaperone	Murine C4b	Retarding the folding of binding protein	(Ogata et al., 1993)
Retarding protein folding by having non-optimal codons	MBP and β-lactamase	Replacement of non-optimal codons resulted in a significant decrease of secretion	(Zalucki et al., 2009; Kang et al., 2014)
Escaping from the immune system	HIV gp120	Postponing the cleavage of proteins, preventing the exposure of premature protein to the immune system	(Li et al., 1996)
Controlling cytokine secretion	Human interleukin 15	Cleavage in two steps: one rapid and one slow step; providing additional control on cytokine secretion	(Kurys et al., 2000)
Controlling translation and transcription	*E. coli* interferon-12b	Changing the SP codons alters the mRNA secondary structure and translation initiation	(Ramanan et al., 2010)
N-terminal glycosylation of protein	Human interleukin-15 isoform	Containing a long SP, essential for protein glycosylation	(Kurys et al., 2000)
Membrane integration of nascent peptide	Prion proteins	Controlling membrane integration	(Ott and Lingappa, 2004)
Nuclear localization signal	Mouse mammary tumor virus envelope protein and its splice variant	Having a long SP containing nuclear localization signal	(Dultz et al., 2008)
Functioning as an antigen	Human HLA-C	Forming a complex with HLA-E on the surface of cells, causing apoptosis by NK cells	(Ulbrecht et al., 2000)
	Human cytomegalovirus UL40	Possessing a mimotope of the HLA-C SP: Infection of human cells with this virus downregulates HLA-C and maintains the level of HLA-E → non recognition of immune cells by NK cells and cytotoxic T-cells	
Functioning as a trans-membrane domain	*Lassa virus* glycoprotein-C	Remaining membrane anchored and processing glycoprotein-C into GP-1 and GP-2	(Schrempf et al., 2007; Eichler et al., 2003; Duda et al., 2004)
	Junin virus glycoprotein-C	Adopting a specific topology in the membrane, ensuring the full assembly of glycoprotein-C	(Agnihothram et al., 2007)
	Foamy virus envelope protein	Integrating membrane, incorporating into viral particles, and interacting with Gag proteins	(Duda et al., 2004; Wilk et al., 2001)
	Cyclophilin SP from the cattle parasite	Remaining membrane anchored	(Rusch and Kendall, 2007; Ebel et al., 2004)
Infectivity role	*Arena virus* glycoprotein-C	• Pseudoviral infection and cell surface expression of glycoprotein-C • Fusion to the cell membrane of the host by myristoylation site of the SP	(Schrempf et al., 2007; York and Nunberg, 2006; Bederka et al., 2014)
	Foamy virus envelope protein	Controlling the balance between viral and sub-viral particles released from infected cells by ubiquitination	(Stanke et al., 2005)
	Alpha virus membrane proteins (P62)	Playing a role in spike heterodimerization	(Lobigs et al., 1990)
	Alpha virus 6 K peptide	Being responsible for insertion of the protein E3 (the viral membrane protein), assembly of the virus and budding processes	(Liljeström and Garoff, 1991; Yao et al., 1996; Loewy et al., 1995)
	S. aureus PVL^+	Mediating adhesion of bacteria to the heparan Sulfates in extracellular matrix, causing necrotizing pneumonia	(Tristan et al., 2009)
	S. aureus Agr quorum-sensing protein	Being similar to the small peptide toxins of *S. aureus* (PSM) leading to cytolysis and inflammation	(Schwartz et al., 2014)
	JEV-E	Enhancing packaging efficiency and increasing the number of released viral particles	(Liu et al., 2017)
Regulatory function	SP fragments of pre-prl and p-gp160 protein (related to the human immunodeficiency virus-1)	Binding to the CaM, inhibiting Ca-dependent signaling processes	(Martoglio et al., 1997; Martoglio and Dobberstein, 1998)
	ECP	Upregulation of TGF-α in causing asthma and allergic inflammatory disease	(Chang et al., 2007)

(*continued on next page*)

Table 1 (*continued*)

Application	Type of SP	Evidence and Implication	Ref.
Vaccine candidates	HLA system Bacterial and tumor peptide	• Having high densities of T-cell and B-cell epitopes, useful in designing multiepitope vaccines • Activating innate immunity and the cell-mediated responses • Acting in a TAP-independent pathway • Targeted delivery of antigen to the ER, eliminating the overexpression of xenogeneic proteins and diminishing misbalance in homeostasis	(Aladin et al., 2007; Kovjazin et al., 2011; Kovjazin and Carmon, 2014; Perez-Trujillo et al., 2017)

in physiology or medicine was awarded to Blobel for the discovery of SPs (Blobel, 2000). Signal peptide is a general term referred to a wide variety of signaling sequences. Nevertheless, N-terminal signal peptides (SPs) are the most commonly studied signaling sequence, thus the generic name of SP has been primarily applied to them.

3. Applications and importance of signal peptides

Day by day, the demand for biotherapeutics and recombinant proteins is increasing. Hence, several prokaryotic and eukaryotic hosts have been widely accepted for cytoplasmic expression of recombinant protein. However, there are several obstacles in the large-scale production of recombinant proteins, among which inclusion body formation and protein degradation via proteases are the important factors. On the other hand, endogenous proteins may interfere with the folding of a recombinant secretory protein. The above-mentioned factors, as well as the complicated downstream purification process of protein production, will result in loss of protein yield (Jonet et al., 2012). Moreover, the yield of recombinant protein is not only related to the expression level but also to translocation efficiency. Two major components, including secretory machinery and SPs, determine the capability and efficiency of translocation. The translocation efficiency could be increased by using alternative SPs from heterologous species (Kober et al., 2013; Zamani et al., 2015; Negahdaripour et al., 2017; Mousavi et al., 2017). It was shown that the use of SPs has increased protein production to commercially significant levels (Ohmuro-Matsuyama and Yamaji, 2017; Low et al., 2013; Mergulhao et al., 2005).

SPs are of special importance from other aspects than production of recombinant proteins. About 50% of proteins that have been identified in the extracellular proteins of *Bacillus subtilis* contain typical SPs (Humphery-Smith and Hecker, 2006). More than 90% of the secretory proteins in *Escherichia coli* are SP-dependent (Rusch and Kendall, 2007). A significant amount of eukaryotic secretory proteins, which is more than 20% of mouse and human proteome, possesses SPs as well (Kanapin et al., 2003). The considerable proportion of SPs in various taxa signifies their study. Additionally, a large number of human diseases is caused by mutations in the SPs. A comprehensive literature survey was carried out to find SP- associated diseases. It was revealed that 26 diseases were attributed to the SP mutations and impairment in 21 human proteins (Jarjanazi et al., 2008). For example, a single mutation in the hydrophobic region of pre parathyroid hormone SP impairs the hormone secretion, causing autosomal familial isolated hypoparathyroidism (Datta et al., 2007). Another example is mutation in the SP of preproinsulin that is associated with the onset of diabetes (Bonfanti et al., 2009). A new variant SP of the human luteinizing hormone receptor (LHCGR) affects receptor biogenesis via triggering Leydig cell hypoplasia (Vezzoli et al., 2015). Moreover, SPs are intriguing targets of drugs. For instance, cyclotriazadisulfonamide (CADA) down-modulates human CD4 (hCD4) in a highly selective manner through interfering with the related SP. In fact, CADA disturbs the completion of SP inversion to a hairpin-looped structure, thus, inhibits SP cleavage; and subsequently the CD4 polypeptide is less expressed on the surface of thymocytes. As a result, cells will be protected from HIV and SIV infection, whereas the immune functions attributed to $CD4^+$ T-cell will be maintained (Vermeire et al., 2014). The success of gene therapies highly depends on the high transfection efficiency to the target organ, which could be achieved by increasing the secretion rate and serum levels of the therapeutic molecule via selecting optimized SPs (Zhang et al., 2005). It is possible to take advantage of SPs as diagnostic biomarkers in a number of diseases. For instance, levels of SP complement for epidermal growth factor domain-containing protein-1 herald pulmonary embolism (Dirican et al., 2016).

SPs are implemented in lab applications together with their fundamental capacities in industry and therapeutic fields. In this context, there is a technique called “signal-exon trap”, which is used to distinguish secretory proteins or membrane proteins on the genomic scale using SPs (Péterfy et al., 2000). Interestingly, SPs are crucial for the selective activity of a selection marker in labs. In this view, sugar intolerance for the SacB, the gene used as a selection marker in *E. coli*, only appeared when translocation was SP dependent (Gao et al., 2017).

Taken together, SPs are ubiquitous in diverse fields. In addition to the abovementioned roles, SPs have other applications, including controlling the rate of protein secretion, determining protein folding-state, affecting downstream trans-membrane behavior and N-terminal glycosylation, nuclear localization signal, playing role in viral/bacterial infectivity, and applying as potential vaccine candidates. Roles and applications of SPs are summarized in Table 1.

4. Secretory systems in prokaryotes and eukaryotes

Several different secretory pathways have evolved among organisms: 1) to come up with the secretory needs of rapidly growing organisms such as bacteria and yeasts, 2) to offset the low rate of protein synthesis versus the high rate of secretion, and 3) to be able to secrete proteins with different characteristics. Additionally, the type of secretory pathway, which is determined by the SP’s features, affects protein localization in the cell as well as post/co-translational modification of protein (Colombo et al., 2005; Anandatheerthavarada et al., 1999; Rutkowski et al., 2003; Hegde and Bernstein, 2006). Therefore, having a general view of different secretory pathways enables us to optimize protein secretion by choosing the most appropriate SPs or manipulating them with respect to the attributes of the secretory protein. Different types of secretory pathways depending on the N-terminal signal, the related organisms, and their specific features are summarized in Table 2. In this review, only the general secretory pathways are mentioned.

4.1. Sec pathway

4.1.1. Introduction of the Sec pathway

Sec pathway or post-translational translocation is a pathway in which proteins remain unfolded to be sufficiently recognized and translocated via the secretory machinery. In other words, proteins that undergo post-translational modifications are secreted via the SecB pathway. The Sec pathway, which seems to be conserved in all classes of lives, is classified as SecB or SecA2 dependent pathway. The SecB pathway is considered as the prevailing pathway in all classes of bacteria and SecA2 is regarded Gram-positive specific (Green and Mecsas,

Table 2
Secretory pathways relying on N-terminal signal sequence (SPs).

Secretory pathway	Species	Specific features	Ref.
Sec-dependent pathway	Present in all taxa Prevail in Gram-negative bacteria	• Conducting post-translational translocation • Having higher capacity for recombinant protein production • Preventing unwanted degradation and folding of proteins • ATP-dependent	(Rusch and Kendall, 2007; Green and Mecsas, 2016; Yuan et al., 2010; Plath et al., 1998; Mordkovich et al., 2015; Papanikou et al., 2007; Hardy and Randall, 1991)
SRP-dependent pathway	Present in all taxa Prevail in eukaryotes	• Conducting co-translational translocation • Appropriate for rapid folding proteins • ATP-dependent	(Yuan et al., 2010; Rapoport et al., 1996) (Luirink and Dobberstein, 1994; Bornemann et al., 2008; Lange et al., 2016)
Tat-dependent pathway	Present in all taxa Prevail in Gram-positive bacteria and archaea	• Carry mature and rapid-folding proteins • Mostly carry SPs with twin-Arg motif • Useful pathway for recombinant protein production in Bacillus • Appropriate for translocation of multimeric and co-factor dependent- enzymes • Dependent on PMF	(Yuan et al., 2010; Palmer and Berks, 2012; van Dijl et al., 2002; Berks et al., 2000; Müller, 2005; Ulfig et al., 2017; Nielsen et al., 1997)
ABC transporter-associated pathway	Present in all taxa Prevail in Gram-positive bacteria	• Translocate proteins not compatible with the previous pathways including lantibiotics and pheromones • Transporting SPs called leader peptides • ATP-dependent	(Fath and Kolter, 1993) (Quentin et al., 1999)
YidC pathway	Present in all taxa	• Responsible for membrane insertion of proteins • Can act in parallel with other pathways	(Yuan et al., 2010; Samuelson et al., 2000; Scotti et al., 2000; Yi and Dalbey, 2005)
SecA2 pathway	Gram-positive specific	• Used in stress condition and virulence of bacteria • Translocate protein through SecYEG secretion system	(Green and Mecsas, 2016; Rigel and Braunstein, 2008; Bensing et al., 2014)
SecA2-SecY2 (aSec) pathway	Gram-positive specific	• Responsible for translocation of large, highly-glycosylated anchor proteins with serine-rich repeats (such as virulent protein) • Translocate proteins through SecY2 channel	(Bensing et al., 2014; Siboo et al., 2008; Mistou et al., 2009)
Injectosome pathway	Gram-positive specific Observed in *Streptococcus pyogens*	• Translocation of Sec-transported toxins between bacterium and the host cell • Functional analog of T3SS	(Madden et al., 2001; Ghosh et al., 2010)
Type II secretion system (T2SS)	Gram-negative specific	• Secrete enzymes and virulence factor out of cell wall • Transport proteins translocated by Sec or Tat pathway • Transport folded proteins	(Korotkov et al., 2012)
Type III secretion system (T3SS) (injectosome) (needle and syringe)	Gram-negative specific	• Transport unfolded proteins, esp virulence factor, in one-step process • Carry uncleavable SPs	(Büttner, 2012)
Type V secretion system (T5SS)	Gram-negative specific	• Transport SecB-translocated virulence factors • Self-secretion of protein harboring a β-barrel domain • Including three pathways, autotransporter, two-partner, and chaperone-usher	(Green and Mecsas, 2016; Leyton et al., 2012)

2016). More than 90% of *E. coli* proteins (Rusch and Kendall, 2007) follow the Sec pathway; however, this pathway is absent or of trivial importance in most of the Gram-positive bacteria (Yuan et al., 2010). In eukaryotes, small secretory proteins, including yeast proteins and organellar proteins, opt for the Sec pathway (Plath et al., 1998). Notably, the Sec pathway is preferable in the biotechnological processes, due to three reasons: 1) it has a higher capacity of protein production than other secretory pathways such as Tat, 2) the contact between the pro-protein and intracellular protease is minimal in this pathway, as they are substantially coupled to the secretory components, 3) SecB-dependent proteins are directed to the periplasm or outside of the cell (Mordkovich et al., 2015; Papanikou et al., 2007).

4.1.2. The SEC pathway components and functionality

Sec machinery consists of SecB (or DnaK), SecA, SecY, SecE, SecG, and SecDF proteins (Rusch and Kendall, 2007). Pre-proteins bind to SecB for maintaining transport-competent state. In fact, SecB acts as a chaperone, which prevents the protein folding before secretion (Hardy and Randall, 1991), and subsequently, the pre-protein is guided to SecA. As SecA was bound to the pre-protein and ATP, it was inserted into the cell membrane. In the cell membrane, the SecA/pre-protein complex forms a translocation complex with SecY/SecE/SecG (Economou and Wickner, 1994). SecY and SecE form a channel conducting protein out of the cell, and SecG activates translocation. The protein is conducted by the energy released from ATP and proton gradient (Papanikou et al., 2007). Finally, the mature protein is released by SecD (or SecDF complex) as the signal peptidase (SPase) cleaves the SP (S-i et al., 1993) (Fig. 1). Cleavage by SPase is an event that takes place in the most aforesaid pathways., SecA and SecY are so critical in the Sec system so that their advantageous mutations can compensate defects in the SP (Rusch and Kendall, 2007).

4.2. SRP pathway

4.2.1. Introduction of the SRP pathway

SRP pathway or co-translational translocation is a pathway occurring co-translationally, which prevents any form of cytoplasmic modification. This pathway seems to be present in all domains of life (Yuan et al., 2010). Proteins translocated across the endoplasmic reticulum (ER) membrane utilize the SRP pathway, particularly the ones with more than 100 amino acids (aas) (Rapoport et al., 1996). It was hypothesized that the SRP-dependent mechanism dominates in eukaryotes. However, accumulating data suggests that some eukaryotic proteins are translocated via SRP-independent mechanisms (Nilsson

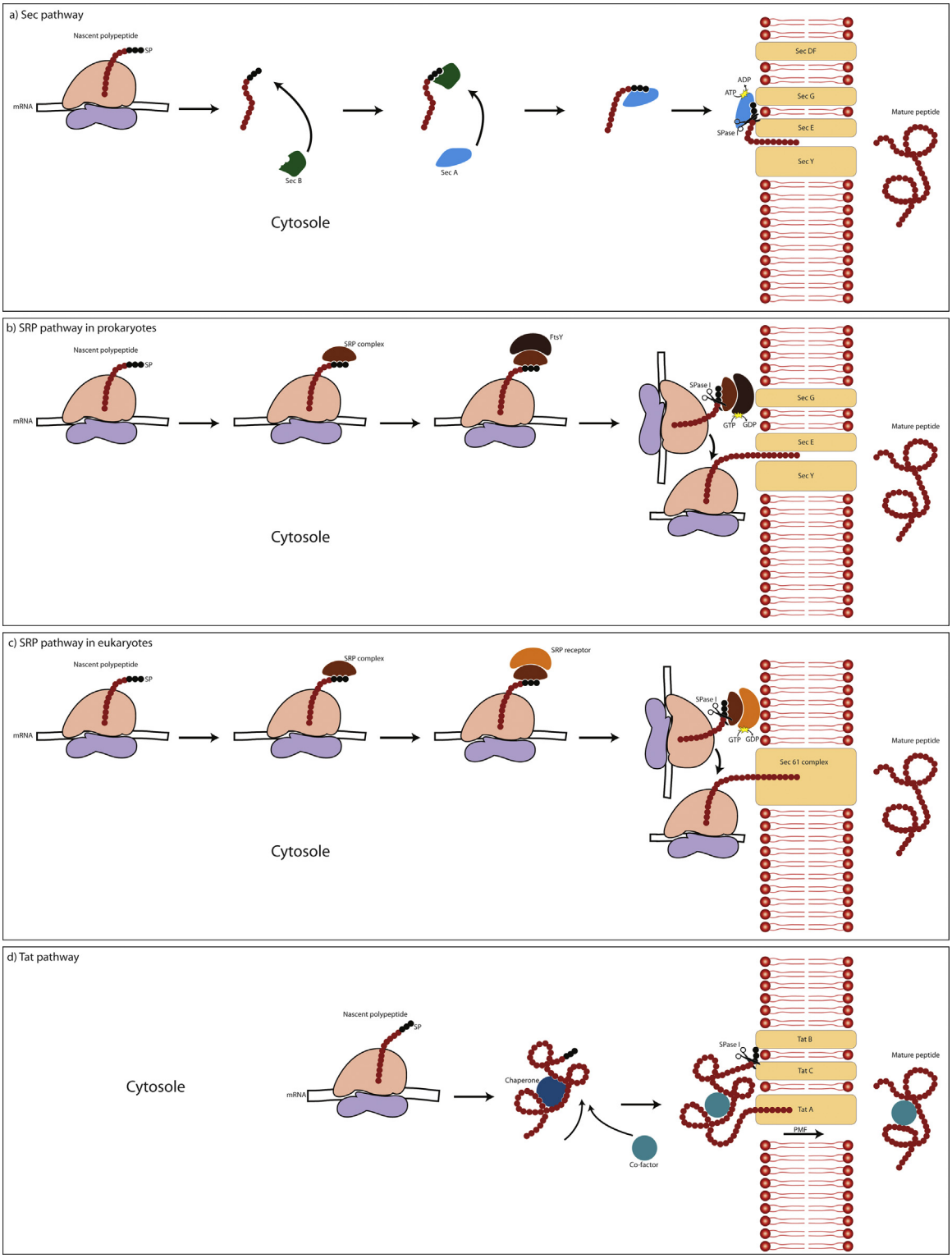

Fig. 1. General secretory pathways in eukaryotes, prokaryotes, and archaea. a) Sec pathway. Sec machinery consists of SecB, SecA, SecY, SecE, SecG, and SecDF proteins. Nascent polypeptide binds to the SecB, which acts as a chaperone and prevents folding of the protein before secretion. After that, the protein is guided to the SecA. As SecA binds to the pre-protein inserts, into the cell membrane, forming a translocation complex with SecY/SecE/SecG. SecY, and SecE form a channel conducting protein out of the cell, and SecG activates translocation. The protein is conducted by the energy released from ATP. Finally, the mature protein is released as the SPase I cleaves the SP. b) SRP pathway in prokaryotes. The SRP pathway consists of the SRP complex and the FtsY receptor. Recognition of SP by SRP is initiated in the tunnel of ribosomes or directly after the emergence of a nascent polypeptide chain on the ribosome. The SP-SRP complex is targeted to the same translocon in the Sec pathway using an FtsY/SRP receptor. c) SRP pathway in eukaryotes. SRP pathway consists of the SRP complex, SRP receptor, and translocon 61 complex. The mechanism of this pathway is completely similar to the prokaryote one. d) Tat pathway. Tat secretory system generally consists of three membrane proteins, called TatA, TatB, and TatC. Folding occurs soon after synthesis of the protein by chaperones. The mature protein-SP complex is recognized by the TatBC complex. It recruits TatA that translocates the protein through a conducting channel. The mature protein is translocated by PMF.

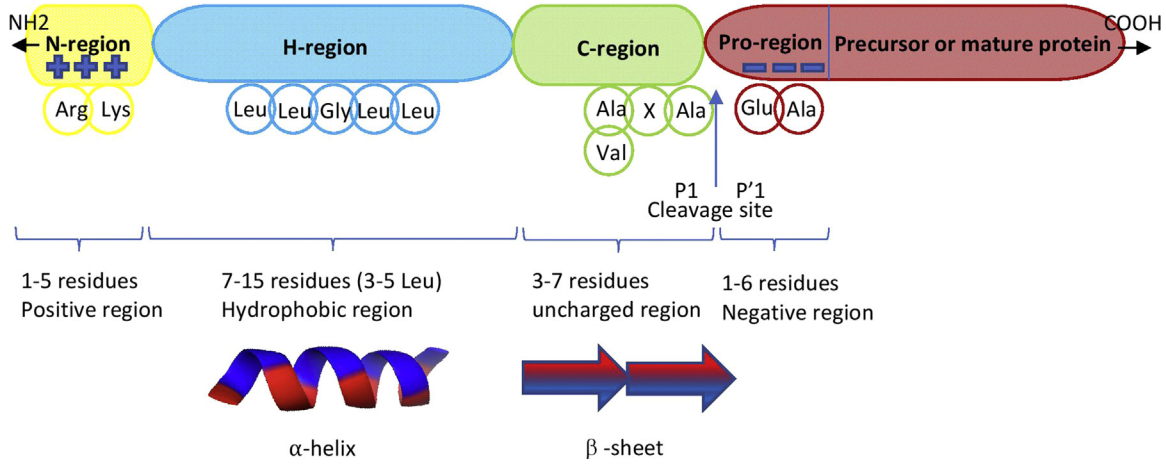

Fig. 2. The general structure of an SP. It is composed of three main parts: 1) N-region- the positive-charged domain 2) H-region- the hydrophobic core, forming α-helix 3) C-region- the cleavage site, forming β-sheet. The initial part of protein, important to protein secretion, is called Pro-region. Residues before the cleavage site are called as P1 and after the cleavage site are called as P'1. Cleavage occurs at AXA or VXA motif.

et al., 2015).

4.2.2. The SRP pathway components and functionality

The SRP pathway consists of the SRP complex, the FtsY receptor in bacteria or SRP receptor in eukaryotes, and the translocon complex. SRP complex is a ribonucleoprotein, consisting of a 7S RNA and six protein subunits (SRP9, SRP14, SRP19, SRP54, SRP68, and SRP72) in eukaryotes, and 4.5S RNA and the Ffh protein in prokaryotes (Luirink and Dobberstein, 1994). The archaeal SRP pathway is an intermediate between bacterial and eukaryotic SRP system. Recognition of SP by SRP begins in the tunnel of ribosomes or directly after the emergence of a nascent polypeptide chain on the ribosome. In this way, protein folding starts in a very initial stage (Bornemann et al., 2008; Lange et al., 2016). The SP-SRP complex is targeted to the same translocon in Sec pathway using FtsY/SRP receptor. In the eukaryotes, the ER membrane-spanning translocon Sec61, composed of Sec61, Sec62, and Sec63, directs proteins out of the ER (Deshaies and Sanders, 1991). In addition to Sec61 complex, for some of the proteins, two additional membrane proteins called translocating chain associated membrane protein (TRAM) or translocon-associated protein (TRAP) are essential for protein translocation (Görlich et al., 1992; Fons et al., 2003). These accessory translocons are crucial in mammalian cells, whereas in the yeast, Sec61 is sufficient for protein translocation (Fig. 1).

4.3. Tat pathway

4.3.1. Introduction of Tat pathway

The twin-Arg transport system, shortly called Tat pathway, is a Sec-independent pathway that translocates proteins in a mature state. Tat pathway has gotten its name from the SPs carrying the twin-Arg motif (Palmer and Berks, 2012). It is found in bacteria, archaea, plant chloroplast, and a few plant mitochondria with a higher prevalence in archaea and Gram-positive bacteria, respectively (Yuan et al., 2010; van Dijl et al., 2002). Notably, low secretion efficiency of heterologous proteins in Bacillus species is attributed to the specific properties of Sec pathway, which is used in the production of commercially important recombinant proteins. As a consequence, it has been suggested to take advantage of Sec-independent pathways, including the Tat pathway in Bacillus species (van Dijl et al., 2002). Whereas proteins with slow folding rates are translocated via the Sec pathway, rapidly-folding proteins favor the Tat pathway. Additionally, multimeric complexes such as enzyme complex or periplasmic proteins in *E. coli*, which are bound to cofactors (e.g. flavin and iron-sulfur clusters), are secreted via the Tat pathway (Berks et al., 2000).

4.3.2. Tat pathway components and functionality

Tat secretory system generally consists of three membrane proteins called TatA, TatB, and TatC. Nevertheless, the types and numbers of translocase components depend on the species (Gram-positive, Gram-negative, or archaea). TatA and TatC are the minimal prerequisites in Tat pathway (Yuan et al., 2010). The mature protein-SP complex is recognized by the TatBC complex, recruiting TatA, which translocates protein through a conducting channel or destabilizing cell membrane (Müller, 2005; Ulfig et al., 2017). In fact, the proton motive force (PMF) is sufficient to initiate the translocation of protein (Robinson et al., 1993). Herein, Tat pathway is evolutionary advantageous, because the only energy needed for protein translocation is the PMF (Garg and Gould, 2016). Accordingly, the prevalent secretion pathway in *Corynebacterium glutamicum* is the Tat pathway, although the majority of SPs have been identified as the Sec-types (Watanabe et al., 2009). Moreover, oligomerization of Tat translocon partly depends on the PMF (Nielsen et al., 1997). It was revealed that TatB is responsible for recognition of the twin-Arg motif of an SP, while TatC has close contact with the hydrophobic region of the SP (Fig. 1).

5. Signal peptide structure

A typical SP has 25–30 residues (von Heijne, 1990a). Longer SPs (up to ~140 residues) are usually found in eukaryotes; however, they have been also observed in viral proteins and bacterial autotransporters (Hiss and Schneider, 2009; Szabady et al., 2005). Long eukaryotic SPs are mainly organelle-targeting sequences. Longer SPs mostly remain stable after protein maturation and add extra functions to protein targeting (Hiss and Schneider, 2009). Tat SPs are longer than the Sec or SRP SPs; they have an average length of 36 aa (Tjalsma et al., 2004). However, SPs could be as small as 16 aa such as the SP of albumin and vesicular stomatitis virus G-protein (VSV-G) (Kapp et al., 2013). The general structure of an SP is composed of three main parts: 1) N-region: the positive-charged domain, 2) H-region: the hydrophobic core, and 3) C-region: the cleavage site (von Heijne, 1990a) (Fig. 2). Although all SPs have this general structure, they have remarkable differences in their aa compositions. High variation in SP composition is responsible for their high capacity for protein translocation. It was assumed that SPs are tolerant of a wide range of mutations and capable of secretion in evolutionary distant heterologous hosts (Von Heijne, 1985; Zheng and Gierasch, 1996). However, this belief was an issue of debate thereafter (Duffy et al., 2010). SP variations alter translocation efficiency, cleavage sites, and even the post-cleavage events. In order to address the different positions in SP structure, any position before cleavage site is called as P1 (position -1), P2 (position -2), and so on; and any position

Table 3a
Studies concerning the SP structure, N-region.

Types of study	Consequences	Ref.
Lacking positive residues at +2 /+3 of twin-Arg in *B. subtilis*	Not directing the SP toward the Tat pathway in spite of having twin-Arg	(van Dijl et al., 2002)
Substitution of twin-Arg with a pair of Lys	Impairing the interaction between TatC and the SP	(Ulfig et al., 2017)
Positive charge balance	Determining SP orientation	(Nilsson et al., 2015)
Net positive charge over the C-region in eukaryotes	Easy access of SPase to the cleavage site	(Amaya et al., 2015) (Hartmann et al., 1989)
Net positive charge over the C-region in prokaryotes	Locating cleavage site at the preplasmic part	(Paetzel et al., 2002)
Analysis of 475 bacterial and 1877 eukaryotic SPs	Less positive charge in the eukaryotic N-region	(Choo and Ranganathan, 2008)
Less positive charge of ER SPs compared with mitochondrial SPs	Avoiding crosstalk between mitochondrial and ER proteins in eukaryotes	(Berks, 2015)
Presence of a positive charge in spite of having a net negative charge	Effective secretion in Levansucrase in *Bacillus*	(Lammertyn and Anné, 1998)
Net charge of +1 in the MBP and ALP	Preferable for secretion but not required	(Puziss et al., 1992) (Nesmeyanova et al., 1997)
Increasing the net positive charge of SP from +2 to +5	Higher intracellular solubility and periplasmic export for recombinant glucanotransferase in *E. coli*	(Ismail et al., 2011)
Systematic screening of *B. subtilis* SP	Positive charge of +2 is favorable	(Degering et al., 2010) (Brockmeier et al., 2006)
Analysis of pro-entrotoxin SP	R4D substitution reduced the net charge from +2 to +1, not affecting pro-entrotoxin translocation efficiency Continuous mutations of N-region not affecting secretion efficiency Not having Arg at the second position of mutant and wild type SP	(Mordkovich et al., 2015)
E1V substitution in the SP of Cel-CD	Increasing the net positive charge of N-region from -1 to 0, inhibiting the inner membrane translocation of protein	(Gao et al., 2016)
Increasing the positive charge	Decreasing inner membrane secretion	(Jonet et al., 2012)
L to D/E substitution in the SP of antibodies	Significant increase of antibody secretion	(Ohmuro-Matsuyama and Yamaji, 2017) (Haryadi et al., 2015)
Rise in the negative charge	Preventing protein aggregation and enhance soluble expression	(Joshi et al., 2012) (Kvam et al., 2010)
Substitution of AAA with the other codons	Secretion abolishment	(Ahn et al., 2006)
Having Lys or Arg in the second residues of MBP and RBP	High export rates	(Puziss et al., 1992)
Lys, Tyr, or His at the second position of ALP's SP	Increasing secretion	(Nesmeyanova et al., 1997)
Substitution of AAA with AAT in MBP and ALP	Not affecting protein export	(Zalucki et al., 2007)
Substitution of AAA with AAG	Deleterious effects	(Zalucki et al., 2007)
Fusion of SPs (PelB, OmpA, and CSP) harboring Lys as the second residue	Efficient secretion of difficult-to-express proteins like GM-CSF	(Sletta et al., 2007)
Met-Arg-Phe conversion to Met-Phe-Arg or Met-Phe-Lys in prepro-α-factor SP	45-75% reduction in translocation	(Green et al., 1989)
Increasing basicity of the IL2 SP	Promoting secretion efficiency of ALP and endostatin both *in vitro* and *in vivo*	(Zhang et al., 2005)
Photocrosslinking investigation of implementing mutations reducing positive charge	No effect on SP-SRP association	(Nilsson et al., 2015)
Introduction of Arg in the BEVS SP	Not resulting in an increase in recombinant protein secretion	(Futatsumori-Sugai and Tsumoto, 2010)

after the cleavage site is called as P'1 (position +1), and P'2 (position +2). There are several studies concerning elucidation of SP composition as well as optimization of their structure. In the following sections, the major aspects of SP regions are mentioned and summarized in Tables 3a–d.

5.1. N-region: the positive-charged domain

5.1.1. Introduction of the N-region

The N-terminal part of SP is called N-region, generally characterized by the presence of basic residues (Table 3a). This region highly varies in length but usually, it has an average length of five residues (von Heijne, 1990a). The N-region is responsible for interactions with 1) negatively charged phosphate group of the lipid bilayer, essential for efficient translocation of proteins (Van Voorst and De Kruijff, 2000), 2) the phosphate backbone of SRP complex or SecB chaperone, which directs SPs through SecA/YEG translocase (Batey et al., 2000), and 3) SecA ATPase. The positive charge offsets for the uncharged formylated N-terminal Met in bacterial proteins, particularly in the Gram-positive bacteria. Conversely, the initial Met in eukaryotes is unformylated and charged (von Heijne, 1984). Thus, without considering the long SPs, N-region tends to be the shortest in eukaryotes and the longest in Gram-positive bacteria (von Heijne, 1990a; Choo and Ranganathan, 2008). Based on secretory systems, the length of N-region of Tat-dependent SPs in *B. subtilis* is about twofold more than the Sec-dependent SPs, having 13–14 residues (van Dijl et al., 2002).

5.1.2. The importance of positive charge and its difference among various SPs

The positive charge in N-region, as well as positive charge in the downstream of SP, determines SP orientation. In other words, the balance of positive charges across the SP confers an orientation favoring or preventing translocation (Nilsson et al., 2015). In both eukaryotic and prokaryotic SPs, the charge difference between N-region and C-region is mainly positive in favor of the N-region. As a result, the C-region would be easily available for the active site of SPase, while the N-region remains in the cytosol (Amaya et al., 2015; Hartmann et al., 1989; Paetzel et al., 2002). However, talking about the net positive charge of the N-region itself, the story would be different within eukaryotes and prokaryotes. Extensive analysis of bacterial and eukaryotic SPs depicted a greater inclination for a net positive charge within prokaryotic N-regions compared to the eukaryotic ones. The N-regions of Gram-positives are also more positive than their Gram-negative counterparts (Low et al., 2013; Choo and Ranganathan, 2008; Degering et al., 2010; Brockmeier et al., 2006). Furthermore, ER SPs are less positive than other organelle-targeting SPs. Interestingly, the charge difference in ER SPs avoids crosstalk between mitochondrial and ER proteins (Berks, 2015). SPs of different secretory pathways harbor different levels of positive charge. For instance, the N-regions of Tat SPs harboring the twin-Arg motif are more positively charged. Notably, carrying the twin-Arg motif is not adequate in the *B. subtilis* SP, and positive residues at the position of +2 and +3 to Arg is necessary to direct the SP toward Tat pathway (van Dijl et al., 2002).

Table 3b
Studies concerning the SP structure: H-region.

Types of study	Consequences	Ref.
Changing the H-region hydrophobicity	Affect the ATPase activity of SecA	(Mori et al., 1997)
Increasing the hydrophobicity	Increasing secretion and periplasm accumulation of the heavy chain of monoclonal antibody in *E. coli*	(Zhou et al., 2016)
Increasing hydrophobicity of the IL2 SP	Promoting secretion efficiency of ALP and endostatin both *in vitro* and *in vivo*	(Zhang et al., 2005)
IFNα2 had a long run of hydrophobic amino acids and only two hydrophilic residues	Highly efficient	(Román et al., 2016)
The average hydrophobicity of 56% and 61% with respect to the percentage of the amino acids G, A, V, L, I, M, F, W, and P in *Bacillus*	Highly efficient	(Degering et al., 2010; Brockmeier et al., 2006)
Mutations in the *E. coli* TorA SP translocated via Tat pathway	Reducing hydrophobicity, impairing translocation efficiency	(Ulfig et al., 2017)
Being the SP of cutinase, a helix-beta protein	Helical propensity become less important	(Zhang et al., 2016)
Missense mutation (Val9Ala) in the pGH SP	Changed the quantity of the mRNA at steady state by forming different mRNA secondary structures, resulting in an expression decrease	(Cheng et al., 2016)
Presence of at least five Leu residues	Efficient translocation of ALP	(Rusch et al., 1994)
Having up to ten Leu residues	More efficient in translocation than the wild type ones	(Chen et al., 1996)
Reducing the hydrophobicity of *S. aureus* SP	Complete abolishment of the secretion of mature protein	(Mordkovich et al., 2015)
Hydrophobicity increases in the OmpF SP	Positive charge of the N-region would not take into account in translocation efficiency	(Hikita and Mizushima, 1992a; Rusch et al., 2002)
Photocrosslinking investigation of implementing mutations reducing hydrophobicity	Drastic effects on SP-SRP association	(Nilsson et al., 2015)
E.coli lipoprotein SP containing nine or more Leu residues in the H-region of OmpF	Translocated efficiently, regardless of the N-region positive charge	(Hikita and Mizushima, 1992b)
OmpF with eight Leu residues	Depended on the positive charge of the N-region	(Hikita and Mizushima, 1992b)
Insertion of charged residues	Defective SP *in vivo*	(Rusch and Kendall, 2007)
Mutations in eukaryotic SPs, reducing hydrophobicity	Abolish Sec61 gating activity	(Oliver et al., 1995)
DNP MAS NMR data: presence of hydrophobic domains	Stabilization of SP-SRP interactions	(Lange et al., 2016)
LamB SP harboring four more hydrophobic residues than its parent peptide	Directing protein into SRP-dependent pathway	(Bowers et al., 2003)
Substitution of Gly with Cys and Leu in the PhoE SP	Shifting protein secretion from SecB to SRP-dependent pathway	(Adams et al., 2002)
Difference in the hydrophobicity	Distinguish the Tat and Sec pathway in *E. coli* but not *B. subtilis*	(van Dijl et al., 2002)
Moderate hydrophobicity, in the SP of OPH from Flavobacterium	Being able to use both Tat and Sec	(Kang et al., 2016)
Finding H-motifs with respect to species	Not being random hydrophobic residues	(Duffy et al., 2010)
Altering the composition and the order of residues	Resulting in different secretion efficiency	(Han et al., 2017)
Introduction of aromatic amino acids into the SP	Preventing heterologous secretion of Cel-CD from *E. coli* inner membrane	(Gao et al., 2016)
Three Leu deletion in the mammalian H-region	Having 80% of wild-type SP crosslinking efficiency to SRP and 0.5-2.5% translocation efficiency of wild type	(Nilsson et al., 2015)
Five Leu deletion in the mammalian H-region	Reducing crosslinking efficiency to 35% and translocation efficiency to the background level	(Nilsson et al., 2015)
Substitution of the conserved region “QQWS” in the PTHrP with four tandem Leus	Impeding cleavage significantly	(Amaya et al., 2015)
Having conserved region of “QQWS” in the PTHrP	Being less hydrophobic and able of dual localization	(Amaya et al., 2015)
Helix-breakers at the edge of H-domain	Facilitating the cleavage by SPase	(Dalbey et al., 1997)
Insertion of Gly in the middle of recombinant glucanotransferase SP	Doubled protein secretion compared with the wild type	(Jonet et al., 2012)
Insertion of helix breaker in the LHCGR SP	Abolishing receptor expression causing leydig cell hypoplasia	(Vezzoli et al., 2015)

5.1.3. Is a rise in the positive charge always favorable?

On one hand, a rise in the net positive charge enhances translocation rate and efficiency. The net positive charge of +1 was favorable for maltose-binding protein (MBP) and alkaline phosphatase (ALP) secretion even though it was not a necessity (Puziss et al., 1992; Nesmeyanova et al., 1997). An increase in the net positive charge of the L-asparaginase SP from +2 to +5 in the mutant resulted in higher intracellular solubility and periplasmic export for recombinant glucanotransferase in *E. coli* (Ismail et al., 2011). On the other hand, an increase in the net positive charge is neither always favorable nor necessary for protein secretion. A frameshift mutation in the gene of pro-enterotoxin SP, resulting in the R4D substitution, reduced the net charge from +2 to +1. Herein, the mutation changed the primary structure of SP, whereas it did not affect pro-enterotoxin translocation efficiency (Mordkovich et al., 2015). Surprisingly, the mutation (E1V) in the cellulase catalytic domain (Cel-CD) SP, inhibited the inner membrane translocation of protein by increasing the net positive charge of the N-region from -1 to 0 (Gao et al., 2016). Consequently, augmenting the net positive charge causes protein aggregation, jamming of secretion translocase, and activation of stress-related genes (Low et al., 2013; Joshi et al., 2012; Kvam et al., 2010). Sometimes only the presence of a positively charged residue appears to be sufficient in spite of having a negative net charge. For example, in the *Bacillus* levansucrase, the presence of positive charge in the N-region, resulted in enzyme secretion, despite the net negative charge of this region (Lammertyn and Anné, 1998).

5.1.4. Amino acid composition of the N-region

What types of residues are responsible for the positive charge of the N-region? Lys or Arg as the second residues resulted in an efficient secretion of MBP, Ribose-binding peptide (RBP), difficult-to-express proteins such as granulocyte-colony-stimulating factors (GM-CSF), and proteins in baculovirus expression vector system (BEVS) (Puziss et al., 1992; Sletta et al., 2007; Futatsumori-Sugai and Tsumoto, 2010). In addition to Lys, Tyr, or His at the same position increased ALP's secretion (Nesmeyanova et al., 1997). In eukaryotes, the precise order of residues in the N-region is critically important rather than the mere presence of positive residues. For example, positive residue insertion in the N-region of yeast invertase did not affect secretion; however, conversion of the wild pre-pro-α-factor SP harboring Met-Arg-Phe to Met-Phe-Arg or Met-Phe-Lys resulted in 45–75% reduction in translocation (Green et al., 1989). Moreover, the occurrence of Arg and Lys in the eukaryotic N-regions, particularly at the second position, is somewhat weak compared with prokaryotic ones, and they are more frequent in

the Gram-positive's N-region than Gram-negative's (Mordkovich et al., 2015; Nielsen et al., 1997; Edman et al., 1999). Tat and Sec substrates could be distinguished based on their N-region. The N-region of Tat SPs harbors the consensus motif of (S/T)-R-R-x-F-L-K (Berks et al., 2000). Several studies have shown that the substitution of twin-Arg with a pair of Lys has impaired the interaction between the TatC component of translocase and the SP (Ulfig et al., 2017).

5.1.5. The codon bias in the N-region

The initial codons of N-region (codons 2–5, particularly the second codon) influence maximal secretion. In about 40% of SPs in *E. coli,* the second codon is AAA, which codes for Lys. Substitution of AAA with other codons coding for Lys such as AAG abolishes secretion (Ahn et al., 2006; Zalucki et al., 2007). After AAA, the best codon is AAT, coding for Asp. Substitution of AAA with AAT in the SP of MBP and ALP did not affect protein export (Zalucki et al., 2007). Consequently, through the natural selection, high-translation-initiation-rate codons were preferred to the ones coding positively charged aas (Zalucki et al., 2009).

Altogether, the N-region is not as conserved and well-studied as the other regions of an SP, thus not playing a major role in SP optimization. However, three factors are considered in the N-region for the optimization of SPs: 1) carrying a higher positive charge than the C-region to endow an appropriate structure and not to interfere with the secretion capacity of the host, 2) containing basic residues such as Lys and Arg for an optimal interaction with the secretory components, and 3) harboring high-translation-initiation-rate codons such as AAA and AAT, particularly in the second codon.

5.2. H-region: the hydrophobic core

5.2.1. Introduction of the H-region

The milestone of SP is its central hydrophobic part, called H-region (Table 3b) (Hegde and Bernstein, 2006). The H-region consists of ~7-15 residues (von Heijne, 1990a). However, long H-regions were also reported; for example, the one reported in the SP of the tyrosine kinase receptor, Xmrk (*Xipophorus* melanoma receptor kinase) (Schartl et al., 1998). Generally, the H-region in Gram-positive bacteria is considered to be longer than Gram-negative and eukaryotic ones (Choo and Ranganathan, 2008). In addition, some of the SPs are quite different, since they have two hydrophobic regions rather than one such as the SP of glycoprotein-C in lymphocytic choriomeningitis virus (LCMV) (Schrempf et al., 2007). It is of special notice that a very short H-region, irrespective of its hydrophobicity, would be non-functional (Hikita and Mizushima, 1992a).

5.2.2. The importance of hydrophobicity of the H-region

Hydrophobicity of the H-region determines 1) the conformation of SP and its orientation toward cell membrane, 2) SP cleavage, affecting the rate and efficiency of protein translocation, 3) the secretory pathway of protein, and 4) protein processing such as N-linked glycosylation. As the hydrophobicity decreases, protein processing and translocation will be slower or completely quenched. Mutations in the H-region of eukaryotic SPs and pro-enterotoxin B SP from *Staphylococcus aureus*, which led to a reduced hydrophobicity, abolished the Sec61 gating activity and secretion of the mature protein, respectively (Mordkovich et al., 2015; Oliver et al., 1995). Highly hydrophobic H-regions increased secretion of the monoclonal antibody heavy chain in *E. coli*, ALP, endostatin, and interferon alpha 2 (IFNα2) (Zhang et al., 2005; Zhou et al., 2016; Román et al., 2016). A hydrophobic H-region can eliminate the need for a positive charge in the N-region. As reported in an experiment, when the hydrophobicity of OmpF was increased by raising the number of Leu to nine or more, the positive charge did not count in translocation efficiency (Hikita and Mizushima, 1992a; Rusch et al., 2002). Similarly, in one study, mutations in the hydrophobic region drastically affected SP-SRP association; nevertheless, the presence of positive charges in the N-region had mild effects on SRP-SP affinity (Nilsson et al., 2015). The consequences of interfering with the H-region hydrophobicity are not always manifested *in vitro*. Insertion of charged residues in the H-region did not affect the processing of SP *in vitro*; nonetheless, the SP became defective *in vivo* (Rusch and Kendall, 2007). Even in the Tat SPs, for an optimal interaction between Tat precursor proteins and the TatBC receptor complex, Tat motif is not sufficient, and a minimum hydrophobicity of the H-region is obligatory as observed in the TAMO reductase (TorA) SP translocated of *E. coli* (Ulfig et al., 2017). Interestingly, hydrophobicity appears to be important also at the mRNA level by maintaining an optimum codon in the H-region. The missense mutation (Val9Ala) in the H-region of porcine growth hormone (pGH) SP resulted in an expression decrease. Herein, the mutation changed the quantity of the mRNA in a steady state by forming different mRNA secondary structures, thereby affecting the accessibility of secretory machinery to mRNA (Cheng et al., 2016).

A rise in the hydrophobic residues is not always desired. If the hydrophobicity increases too much, a great number of immature proteins with non-native conformations would jam behind the translocase, which threatens cell viability besides decreasing the protein secretion. The hydrophobic residues in this region confer the propensity of adopting α-helical conformation, thus have a membrane-spanning role. However, the helical propensity of the SP becomes less important for proteins rich in helices. For example, the helical property was important for the efficient function of the xylanase SP, a β–rich enzyme, unlike cutinase SP, a helix-beta protein (Zhang et al., 2016). Furthermore, increasing the H-region hydrophobicity augments the inner membrane secretion; however, the outer membrane secretion decreases. Notably, heterologous proteins, secreted from *E. coli,* possessed hydrophilic and more polar SPs (Gao et al., 2016; Li et al., 2014; Teng et al., 2011).

5.2.3. Effect of the H-region hydrophobicity on selecting the secretory pathway

The hydrophobicity extent of the H-region determines what kind of secretory pathway (Sec, SRP or Tat) should be picked up (Ng et al., 1996; De Gier et al., 1998). Highly hydrophobic SPs direct proteins toward SRP pathway rather than Sec pathway. Dynamic nuclear polarization (DNP) magic angle spinning (MAS), nuclear magnetic resonance (NMR) data, as well as photocrosslinking investigations, suggested that hydrophobic domains stabilize SP-SRP interactions, while less hydrophobic SPs disfavor interaction with SRP (Nilsson et al., 2015; Lange et al., 2016). For example, adding or substitution with more hydrophobic residues in LamB and PhoE shifted protein secretion from SecB to the SRP pathway (Bowers et al., 2003; Adams et al., 2002). Berks et al. (2000) stated that the H-region of Tat SPs is less hydrophobic and with a higher occurrence of Gly and Thr than Leu. The hydrophobicity threshold for a specific pathway differs among diverse species. Proteins under a specific threshold of hydrophobicity are not recognized by the yeast and bacterial SRPs unlike mammalian SRPs (Hegde and Bernstein, 2006). In *E. coli*, Tat SPs contain less hydrophobic H-region than Sec SPs. By contrast, there is no significant difference in the length and hydrophobicity of the H-region of Tat SPs and Sec SPs in *B. subtilis* (van Dijl et al., 2002). The SPs with a moderate hydrophobicity will be able to use both Tat and Sec machinery, as seen in the SP of organophosphorus hydrolase (OPH) from Flavobacterium (Kang et al., 2016). Overall, the need for more hydrophobic substrates declines in the SRP, Sec, and Tat pathway, respectively. Hydrophobicity seems to be more important for prokaryotes than eukaryotes as well as Gram-negatives than Gram-positives.

5.2.4. Amino acid composition of the H-region

There is a great preference for specific types, order, and number of hydrophobic residues in the H-region. Leu predominates in prokaryotes and eukaryotes. Having at least five Leu residues and increasing to ten boosted the translocation in *E. coli* (Rusch et al., 1994; Chen et al.,

1996). Although H-regions are different throughout species, they are unique within a specific species. H-region in a specific species contains particular sets of conserved H-motifs rather than random hydrophobic residues. These H-motifs are predictable up to ~75–90% within a particular species (Duffy et al., 2010). Literally, SPs with the same hydrophobicity are not functionally interchangeable (Hegde and Bernstein, 2006). For instance, SP of *Trypanosoma brucei* was not applicable to canine ER microsomal system, while trypanosome SP and canine SP both had the same hydrophobicity (Ahmed et al., 1998). In eukaryotes, a tri-Leu structure is universal for the H-motif. However, there is a greater bias for Leu in mammalian SPs. For example, Leu abounds in the human H-motifs, while Ser emerges frequently instead of Leu in *Saccharomyces cerevisiae*. The *E. coli* H-motifs are rich in Ala besides Leu (Nilsson et al., 2015; Duffy et al., 2010; Palazzo et al., 2007). Moreover, Phe is unique to human H-motif, and polar residues are not found in them (Duffy et al., 2010). Consistently, the introduction of aromatic aas into the SP, including L13 F mutation, prevented heterologous secretion of Cel-CD from *E. coli* inner membrane (Gao et al., 2016). Correspondingly, altering the composition and the order of residues in the H-region of *S. aureus* α-toxin resulted in a different secretion efficiency in *E. coli*. Highly efficient SPs possessed hydrophobic residues in the middle of the H-domain or near to the C-region; by contrast, placing hydrophobic residues near the N-region reduced secretion. It is suggested that there is an interaction between the hydrophobic residues of the H-region and the two other regions (Han et al., 2017). In contrast to what stated up to now, there are a number of SPs in human with less hydrophobic H-region, which translocate via SRP-independent mechanisms, including parathyroid hormone-related protein (PTHrP). PTHrP harbors a QQWS region in its H-region and a hydroxyl group in the boundary of the H-region and the C-region. This structure is unfavorable for ER translocation, yet it is fundamental for dual localization of PTHrP (nuclear localization as well as ER translocation). Surprisingly, substitution of the conserved region “QQWS” with four tandem Leu impeded cleavage significantly (Amaya et al., 2015). Taken together, three to five Leu residues conquered the H-region. Meanwhile, an optimized H-region for the SP of a target protein should be devised on the basis of the previously authenticated H-region of similar sorts of proteins.

5.2.5. Helix-breaker in the H-region

Insertion of helix breaker residues, Gly, Pro, and Ser in the middle of H-region gives a hairpin-like structure to the SP, easing insertion of the SP into the membrane and cleavage by SPase (Driessen and Van der Does, 2002; Dalbey et al., 1997). Presence of a helix breaker also directs protein toward a specific secretory pathway. Helix breaker could be found at any other position rather than at the edge of H-region. Changing the position of helix breaker would change the protein export level (Low et al., 2013). The insertion of Gly in the middle of recombinant glucanotransferase SP doubled protein secretion compared with the wild type, which did not contain the helix breaker (Jonet et al., 2012). The presence of helix-breaker is not always favorable as in the *B. subtilis* Tat SP and the LHCGR (Vezzoli et al., 2015; van Dijl et al., 2002). Moreover, only about 60% of the studied *B. subtilis* SPs had a helix-breaker, which was mostly Gly (Tjalsma et al., 2004).

5.3. C-region: the cleavage site

5.3.1. Introduction of the C-region

Cleavage could be the potential destiny of the SP after translocating a protein to the periplasmic space or out of the cell, where the C-region comes up to be important (Table 3c). The C-terminal part of an SP is called C-region consisting of ~3–7 neutral or polar aas (von Heijne, 1990a). The five-residue length was considered optimum for a high cleavage efficiency. Increasing the length to more than nine would impair or completely abolish secretion (Suominen et al., 1995). As the C-region gets longer, the cleavage site would be far from the membrane surface. Cleavage efficiency determines the protein secretion level. Hence, cleavage is the rate-limiting step of protein secretion (Geukens et al., 2004). Intriguingly, translation initiation region (TIR) overlaps with the nucleotide sequence of C-region, consequently optimization of C-region will affect the translational strength (Zhou et al., 2016).

5.3.2. Amino acid composition of the C-region

SPs of diverse origins differ slightly in length, charge, and hydrophobicity; nevertheless, they seem to be more conserved in the cleavage site (von Heijne and Abrahmsèn, 1989). C-region is composed of residues conferring an extended β-conformation, thereby provides the binding site for SPase. The critical points of the C-region are position 1 and 3 prior to the cleavage site, well-recognized as the (−3, −1) rule or AXA motif (von Heijne, 1984). This cleavage site is universal in prokaryotic, ER, and organellar SPs; whereas it differs from the cleavage sites of other types of SPs, including the ones in lipoprotein or prepilin (Paetzel et al., 2002). Ala aa is preferred at positions -3 and -1. Substitution of Ala at these positions with other residues, particularly at position -1, changes the cleavage site (von Heijne, 1990a). Studies of the SP of MBP (Fikes et al., 1990), the M13 procoat (Shen et al., 1991), and ALP (Karamyshev et al., 1998) depicted that Gly, Ser, and Cys are also tolerated at these two positions. Rarely, Thr is seen at position -1. Ala at position -3 could be substituted by other neutral aas having a larger side chain, including Val, Leu, and Ile, since SPase provides a larger room at this position by having a bigger pocket size than position -1 (Choo et al., 2005). Still, big polar, aromatic, and charged aas are not generally preferable at position -3. For instance, proteomic analysis of *B. subtilis* revealed that 71% of SPs in this organism contain AXA motif. Nevertheless, 18% of the identified extracellular proteins contain VXA motif, denoting that cleavage at VXA motif is prevalent in the *B. subtilis* SPs (van Roosmalen et al., 2004). Similarly, the best-performing *Lactobacillus* SP have the VXA motif (Mathiesen et al., 2008). In eukaryotic SPs, different aas instead of Ala and Val, such as Pro, are likely to be much easily tolerated at P1 and P3 compared with the prokaryotes (Nagahora et al., 1988). The occurrence of AXA motif through analysis of 1877 eukaryotic, 168 Gram-positive, and 307 Gram-negative SPs was 14.5%, 47.0%, and 58.9%, respectively (Choo and Ranganathan, 2008). The cleavage site might change. For example, in the pre MBP bearing other residues than Ala, Ser, Gly, Cys, and Thr at P1, cleavage happens at two residues upstream (Fikes et al., 1990). In summary, the C-region near to the cleavage site has been strictly conserved in prokaryotes, unlike eukaryotes. Furthermore, presence of Ala at position -1 is crucial, irrespective of the species.

Other positions of the C-region rather than positions -3 to -1 are substantially variable and remained to be obscure. The canonical motif of AXA provides substantial interaction of the SP and SPase, while other residues have minimal interaction (Ting et al., 2016). However, there are some observations for these aas. Large aas, or even negatively charged ones are allowed at P2 (Karamyshev et al., 1998; Mathiesen et al., 2008). β-Turn forming residues such as Gly and Pro are preferred outside the AXA motif, particularly at P5 (Cheng et al., 2016; Ting et al., 2016).

C-region features vary depending on the secretory pathway or species the same as N- and H-region. For instance, the C-region of Tat SPs tends to be positively charged in order to be discriminated from the Sec SPs (Yuan et al., 2010). In contrast to other typical SPs, the SP C-region in silkworm proteins and recombinant proteins secreted by BEVS has completely different characteristics. This C-region tends to be polar. Moreover, the introduction of Asp into the C-region of recombinant SPs increased secretion in BEVS (Futatsumori-Sugai and Tsumoto, 2010).

5.4. Pro-region

5.4.1. The importance of Pro-region

Residues located at downstream of the cleavage site and before the mature protein is called pro-region or export initiation domain

Table 3c
Studies concerning the SP structure: C-region.

Types of study	Consequences	Ref.
Substitution of Ala at the positions -1 and -3 with other residues, particularly at the position -1	Changing the cleavage site	(von Heijne, 1990a)
Studying MBP, M13 procoat, and ALP	Gly, Ser, and Cys could be tolerated at -1 and -3. Rarely Thr could be seen at the position -1	(Fikes et al., 1990; Shen et al., 1991; Karamyshev et al., 1998)
Having a bigger pocket size at the position -3 than the position -1 of the SPase	Ala could be substituted by other neutral amino acids having larger side chain, including Val, Leu, and Ile	(Choo et al., 2005)
Proteomic analysis of *B. subtilis*	Containing VXA motif by 18% of the identified extracellular proteins	(van Roosmalen et al., 2004)
Analysis of *Lactobacillus* SP	The best performing had VXA motif and negatively charged residue at -2	(Mathiesen et al., 2008)
Analysis of eukaryotic SPs	Pro can be tolerated at the position -1	(Nagahora et al., 1988)
Analysis of 1877 eukaryotic, 168 Gram-positive, and 307 Gram-negative SPs	Occurrence of AXA motif was 14.5%, 47.0%, and 58.9%, respectively	(Choo and Ranganathan, 2008)
Absence of Pro at the position -2 of *E. coli* alkaline phosphatasen	Impairing or inhibiting SPase I function	(Karamyshev et al., 1998)
Analysis of position -4 and -5 of *E. coli* alkaline phosphatasen	Gly is seen to some extent	(Karamyshev et al., 1998)
Analysis of Gram-positive and Gram-negative SPase	Residues at P2 are buried into a pocket in Gram-negative SPase in contrast to Gram-positive	(Ting et al., 2016)
Presence of Pro in the boundary of the H-region and the C-region	Inducing conformational changes from α-helix to β-strand	(von Heijne, 1990a)
Analysis of bacterial SPs	Necessary to have Pro at P5 in the SP of *S. aureus* and at P5-P6 in other bacterial SPs	(Choo and Ranganathan, 2008; Schallenberger et al., 2012)
Introduction of Pro to the positions -6 to -4 in the eukaryotic SP	Promoting protein processing	(Yamamoto et al., 1989)
Mutations intervening with or reducing the probability of the β-turn formation near to the cleavage site	Decreasing the chance of protein processing	(Cheng et al., 2016)
Deletion of residues at the position -7 to -5 of pro-enterotoxin B SP of *S. aureus*	Complete loss of SP function	(Mordkovich et al., 2015)
Increasing the length to more than nine	Impairing or completely abolishing secretion	(Suominen et al., 1995)

(Table 3d). This region could extend up to 30 aa. The statistical analysis of bacterial SPs revealed that the residues at P1' to P6' are of great importance, since they might be involved in binding to SPase I (Low et al., 2013; Auclair et al., 2012). It is thought that pro-region is responsible for high protein exports, particularly in Gram-negative and Gram-positive bacteria. Deletion of this region results in the accumulation of proteins in the cytoplasm and formation of inclusion bodies (Kim et al., 2007). The recombinant protein fused to position +2 was exported more efficiently than the one fused to position +1 (Geukens et al., 2004). In addition to controlling protein export, proper folding is affected by the Pro-region. Deletion of 16 aa in the downstream of the cleavage site in the Streptomyces transglutaminase impaired protein secretion as well as protein solubility (Liu et al., 2011). Truncation of this region in *Pseudomonas aeruginosa* lipoxygenase resulted in expression of an insoluble protein in *E. coli* (Lu et al., 2016).

5.4.2. Amino acid composition of the pro-region

Neutral or acidic aas are preferable in the Pro-region. Substitution of acidic residues with the basic ones will slow down the protein export and change the cleavage site (von Heijne, 1990a). A negatively charged pro-region is responsible for loop formation with the positively charged N-region; therefore, it compensates the electrochemical potential of the bacterial inner membrane. However, positive charge is tolerated easily in the cleavage site downstream of eukaryotic SPs, since they lack an electrochemical potential in the ER membrane (Low et al., 2013; von Heijne, 1990a). Residues in this region, particularly at P1', should be small and flexible; otherwise, they will cause steric hindrance for the interaction of SPase and SP. For instance, Pro at P1' precluded cleavage

Table 3d
Studies concerning the SP structure: Pro-region.

Types of study	Consequences	Ref.
Deletion of this region	Accumulation of proteins in the cytoplasm and formation of inclusion bodies	(Kim et al., 2007)
Deletion of 16 aa in the pro-region *Streptomyces* transgultaminase	Impairing protein secretion as well as protein solubility	(Liu et al., 2011)
Pro-region truncation in *Pseudomonas aeruginosa* lipoxygenase	Expression of insoluble protein in *E. coli*	(Lu et al., 2016)
Substitution of acidic residues with the basic ones	Slowing down the protein export and change the cleavage site	(von Heijne, 1990a)
Introduction of three Ala and three Gly	Increasing the cleavage specificity of the *Streptomyces* transglutaminase up to 1.5 fold compared with the wild type	(Chen et al., 2013)
Analysis of pro-region in different species	A striking bias for acidic residues at P2' and a complete absence of aromatic residues at the same position for *E. coli* and *B. subtilis*, no bias for the mentioned amino acids in yeast and archeae	(Zalucki and Jennings, 2017)
Analysis of α-amylase in *C. glutamicum*	A Glu at P1'; necessary for secretion	(Watanabe et al., 2009)
Analysis of pro-region in phosphoglycerate kinase	Having the optimum charge of -2 and -3	(Tian and Bernstein, 2009)
Analysis of pro-region in glyceraldehyde-3-phosphate dehydrogenase and enolase	Having the optimum charge of -6	(Tian and Bernstein, 2009)
Charged residue at P1'	Interfering with protein translocation	(Von Heijne, 1994)
Pro at P1'	Causing steric hindrance and precluding cleavage	(Barkocy-Gallagher and Bassford, 1992)
Computational modeling of SP-SPase complex	P1', P3', and P4' are buried in the small cleft of SPase I	(Choo and Ranganathan, 2008)
Studying SRP and SP interaction in eukaryotes through quantitative analysis of photo-adducts	Charged residues in the pro-region have minor effects on SRP-SP association	(Nilsson et al., 2015)

(Barkocy-Gallagher and Bassford, 1992; Von Heijne, 1994). In a recent study, a striking bias for acidic residues at P2' and a complete absence of aromatic residues at the same position for *E. coli* and *B. subtilis* were observed. Nevertheless, no bias for the mentioned aas was observed in yeast and archaea (Zalucki and Jennings, 2017). The optimum charge and the magnitude of negative charge differ from -2 to -6 and depends on the passenger protein (Tian and Bernstein, 2009).

5.5. Long SPs

As mentioned earlier in this manuscript, there are some sorts of SPs that are substantially longer than the typical ones, known as long SPs. The structure of long SPs is composed of different parts compared with short SPs. Long SPs are made up of three main domains: N-domain, C-domain, and transition part (Hiss et al., 2008). They are not as recognized as their short counterparts. Nonetheless, we briefly mention them here.

1) N-domainN-domain is the N-terminal part and responsible for directing a protein to mitochondria. The N-domain consists of positively charged residues such as Arg, which confer the native conformation of SP in the plasma membrane. N-domain locates the C-domain in the inside of membrane and allows protein-protein interactions. 2) C-domainC-domain is the C-terminal part and has the general structure of a canonical SP (N-region, H-region, and C-region). C-domain is sufficient to translocate a protein, but is not effective enough compared with the complete SP. It is responsible for ER-translocation.3) Transition partTransition part is the area that connects the N- and C-domains and consists of three overlapping β-turns. β-turns are not exclusive to long SPs, but are also present in short SPs. Short SPs have a fewer number of β-turns (Hiss and Schneider, 2009). The transition area is responsible for ER translocation. Translocation efficiency was decreased by the omission of this area (Hiss et al., 2008).

6. Types of signal peptides

The majority of SPs that are transported via Sec/Tat/SRP pathways in all life domains have a tripartite structure sharing general similarities as described in the previous sections. Although SPs could be expressed in heterologous hosts (Kober et al., 2013), they are not always recognized by the translocons from other species (Kim et al., 2002). On the other hand, the secretory machinery of a heterologous host might recognize the foreign SP, although secretion might not be as efficient as the native SP. For instance, in the *E. coli* expression system, the OmpA SP was more efficient than the native *Bacillus* SP for both expression and secretion of *B. subtilis* chitosanase (Csn); however, the SP cleavage was not precise (Pechsrichuang et al., 2016). As a consequence, SPs could be classified from different aspects, including secretory pathways, target location, and structure. In Table 4 and Fig. 3, organellar-targeting peptides and structural classifications of bacterial and eukaryotic SPs are compared in details. In Fig. 4, different fates of a peptide synthesized by the ribosome based on its SP are demonstrated.

6.1. Nonclassical signal sequences

As stated in the beginning, SP term generally refers to the amino terminal signal sequences. Hence, other signal sequences located at other locations than N-terminal are considered as nonclassical signal sequences. For instance, peroxisomal targeting signal and nuclear localization signal, located either at the internal or the C-terminal parts of proteins, are outliers. The general structure of these signaling sequences is completely far from the common SPs (Von Heijne, 1990b). Some proteins being targeted to the cell wall or out of the cell also bear nonclassical SPs. Examples are the human UDP-glucuronosyltransferase with internal signals, the virulence factor of *Mycobacterium tuberculosis*, and *E. coli* haemolysin harboring C-terminal secretion signals (Hiss et al., 2008). Cell-wall-binding repeats, found in some bacteria, are the C-terminal signal sequences that direct enzymes to specific sites functioning in the assembly and turnover of the cell wall (Baba and Schneewind, 1996; Tjalsma et al., 2000). Some of the proteins anchoring cell wall have a C-terminal sorting signal in addition to the N-terminal SP (Schneewind et al., 1992). It is of special notice that all of the secretory proteins do not contain signal sequences. These kinds of proteins might be secreted via nonclassical secretory pathways (Tjalsma et al., 2004; Wang et al., 2015). Herein, fibroblast growth factor 1 and 2, interleukin 1 (IL-1), and precursor of IL-16 could be mentioned (Bendtsen et al., 2005).

7. Signal peptidases

7.1. SPase role and classification

Protein export is accompanied by SP cleavage, although it is not always necessary for protein translocation. Human cytomegalovirus SP (US2) and cyclophilin SP from the cattle parasite are two examples bearing a non-cleavable SP in spite of the -3, -1 motif presence (Rusch and Kendall, 2007; Gewurz et al., 2002; Ebel et al., 2004). Interestingly, SPases are not only responsible for the cleavage of SPs, but they indeed catalyze the final steps in protein maturation (Rusch and Kendall, 2007). In sum, SPases are classified into three types, designated as type I, II, and, IV based on their protein targets. Most of the secreted proteins, translocated by any type of the secretory pathways, are cleaved by SPases I, while lipoproteins and prepilin are exclusively cleaved by SPase II and SPase IV, respectively (Paetzel et al., 2002; Auclair et al., 2012). Talking about SPase other than the SPase type I is out of the scope of this review.

7.2. SPase I classification

SPase I is a membrane-bound serine protease, responsible for the cleavage of SP and releasing the mature protein. It could be subdivided into two major subfamilies: 1) prokaryotic (p)-type: found in bacteria, mitochondria (mitochondrial SPase Imp1/Imp2), and chloroplast (TPP) 2) eukaryotic/ ER type: found in eukaryotes, archaea, and some species of bacteria. Processing of the intramitochondrial sorting SP and the thylakoid sorting SP is mediated by the mitochondrial SPase Imp1/Imp2 and TPP, respectively. The ER SPase I, also called signal peptidase complex (SPC), is responsible for the cleavage of protein after exporting into the ER lumen. SPC is much more complex than its prokaryotic counterparts. SPase I vary in length, complexity, and localization in the membrane throughout different species (Paetzel et al., 2002; Ng et al., 2007). They could even be different within a specific species. The species of *E. coli*, yeast, and archaea have only one copy of SPase I, while seven and two copies of SPase I are found in *B. subtilis* and other eukaryotes, respectively (Auclair et al., 2012; van Wely et al., 2001). Generally, the number of SPase I copies in Gram-positive bacteria exceeds other species (Paetzel et al., 2002).

7.3. SPase I structure and catalytic site

In spite of the substantial difference of SPase I in different species, there are five evolutionary conserved residues in all of them, designated as Box A–E. Box A is the anchoring domain, while the others constitute catalytic domains (Ng et al., 2007). The catalytic domain is composed of two pockets, named S1 and S3, which accommodate side chains of residues at P1 and P3. These two pockets and the residues bridging them are highly conserved, conferring a high fidelity to the enzymes (Ting et al., 2016). Despite the high fidelity of the enzyme, it is functional in heterologous hosts, which means ER SPase could be replaced by the P-type or vice versa. In fact, only pockets are in contact with the SP, while trans-membrane domain might act freely in the lipid bilayer. This is advocated by the idea that a truncated SPase lacking the membrane-spanning domain is capable of SP processing (Carlos et al.,

Table 4
Types of Signal peptides.

Classification of eukaryotic signal peptides based on the target organ

Type	Structure		Role		Cleavage	Specific features	Translocon	Ref.
ER signal sequence	Canonical tripartite structure		Targeting to the ER		SPase I	Higher Leu content versus the bacteria	Sec61/62/63 complex	(Nielsen et al., 1997)
Mitochondrial targeting peptide	Matrix signal	Bipartite structure	Targeting peptide into the mitochondria	Directing proteins into inter-membrane space of mitochondria	Mitochondrial processing peptidase (MPP) and mitochondrial intermediate peptidase (MIP)	1) 15-50 aa 2) Rich in Ala, Leu, Lys, and in particular Arg, conferring an amphipathic α-helix 3) Lacking negatively charged acidic residues	Translocase of outer membrane (TOM) and Translocase of inner membranes (TIM)	(Economou and Wickner, 1994; Garg and Gould, 2016; Paetzel et al., 2002; Von Heijne et al., 1989; Martin et al., 1991)
	Intra-mitochondrial sorting signal	Canonical tripartite structure		Sorting protein into the matrix	SPase I			
Chloroplast transit peptide	Stroma targeting peptide	Tripartite structure of 1) Uncharged N-terminal domain 2) Positively charged central core varying in length 3) C-terminal forming amphiphilic β-strand	Directing proteins into the stroma		Cleaving stroma targeting peptide	Having bulky aromatic amino acids	Translocase of outer membrane (TOC): having phe recognizing domain in cytosol translocase of inner membranes (TIC)	(Paetzel et al., 2002; von Heijne and Abrahmsèn, 1989; Von Heijne, 1990b; Patron and Waller, 2007; Sommer et al., 2011; Köhler et al., 2015)
	Thylakoid sorting signal	Tripartite structure	Essential for sorting protein into the thylakoids		Cleaving thylakoid sorting signal			

Structural classification of bacterial signal peptides

Type	Structure	Role	Cleavage	Specific features	Translocon	Ref
Archetypical SPs	Canonical tripartite structure	Common role of bacterial SP	SPase I	–	Sec YEG complex	(Blobel and Sabatini, 1971)
Lipoprotein SPs	Canonical tripartite structure and lipobox in the C-region	Protein targeting, possessing information about lipid modification	SPase II	–	Sec YEG complex	(Hayashi and Wu, 1990)
Prepilin SPs	Have a short basic region without any hydrophobic domain	Targeting the type IV pilin	SPase IV	–	Sec YEG complex	(Strom and Lory, 1993)
Pheromone and bacteriocin SPs = leader peptide	Canonical tripartite structure lacking the H-domain	Post-translational modification of lantibiotic; preventing premature antimicrobial activity	SPase I	–	Sec YEG complex	(van Kraaij et al., 1999)

Structural classification of eukaryotic signal peptides

Type	Structure	Role	Cleavage	Specific features	Translocon	Ref.
Tandem SP	Consisting of two individual SPs	Dual targeting (e.g., targeting to chloroplast and mitochondria)	Two SPase I	Possessions of an intermediate of two features; positive charge and bulky aromatic amino acids	Sec61 or Sec YEG complex	(van Kraaij et al., 1999)
Ambiguous SP	Not composed of distinct SPs, a mixture of SPs	Dual targeting	One SPase I	Having intermediate features of two SPs	Sec61 or Sec YEG complex	(Hansson et al., 2004)
Alternative targeting SPs	Consisting of two individual SPs	Primary targeting	One SPase I	Capable of leading to dual targeting, but having one cleavage site	Sec61 or Sec YEG complex	(Hiss and Schneider, 2009)
SPs with cryptic signals	Contain two or more SPs, one of which is a cryptic signal	Primary targeting	SPase I	Having only primary targeting capacity	Sec61 or Sec YEG complex	(Hiss and Schneider, 2009)

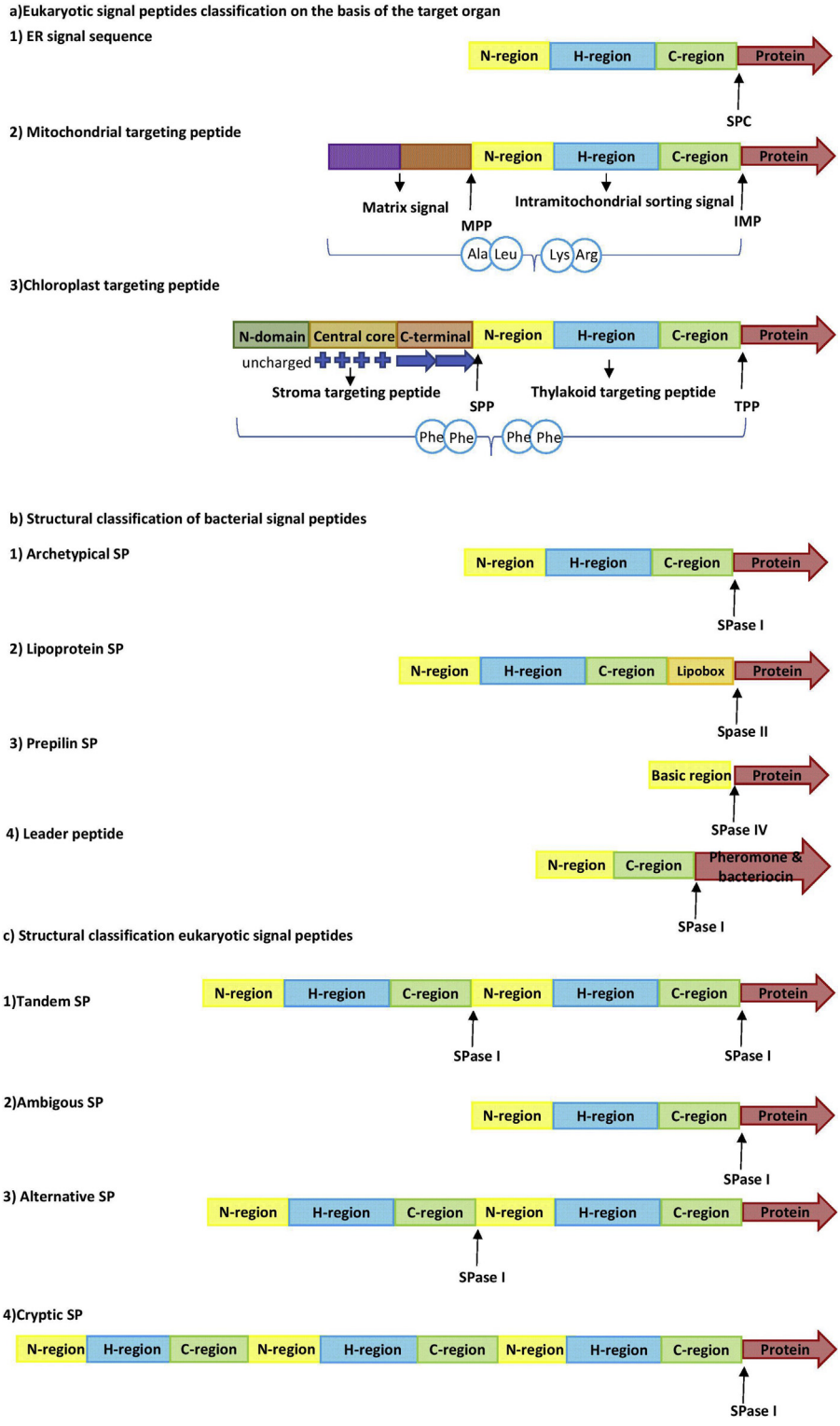

Fig. 3. SP classification. a) Eukaryotic signal peptides classification on the basis of the target organ. ER signal sequence has a common tripartite structure. Mitochondrial targeting peptide is composed of matrix signal and intramitochondrial sorting signal. Chloroplast transit peptide is composed of stroma targeting peptide and thylakoid sorting signal. Intramitochondrial sorting signal and thylakoid sorting signal have a common tripartite structure. Mitochondrial targeting peptide has positively charged residues, while chloroplast transit peptide has bulky amino acids such as Phe. Upward arrows represent the cleavage site. Horizontal arrows represent β-sheet. b) Structural classification of bacterial signal peptides. Archetypical SP has a common tripartite structure. Lipoprotein SPs are distinguished by the presence of lipobox in the C-region. Prepilin SPs only have a basic region. Leader SPs lack the hydrophobic region. Cleavage in different bacterial SPs occurs by different SPase. c) Structural classification of eukaryotic signal peptides. Tandem SPs, ambiguous SPs, and alternative SPs can lead to dual targeting. Tandem SPs and alternative SPs consist of two distinct SPs, while ambiguous SP is a mixture of SPs. Tandem SP is cleaved by two SPase, while alternative SP has one cleavage site. SPs with cryptic signals are the SPs that contain two or more SPs, one of which is a cryptic signal. They have a single cleavage site, thus having primary targeting capacity.

2000). The major catalytic site of SPase I in prokaryotes is Ser-Lys, which is different from the catalytic site of Ser-His-Asp in eukaryotes. SPase I of *B. subtilis* in contrast to other bacterial SPase I, SipW, utilizes Ser-His paired residues (Paetzel et al., 2002).

7.4. Evolutionary analysis of signal peptides

SPs have undergone evolutionary changes similar to other proteins. Although the structure of SPs could be variable, their function and gating efficiency have been highly conserved throughout evolution, particularly between Gram-negative bacteria and eukaryotes (Kim et al., 2002; Pohlschröder et al., 1997; Belin et al., 2004). Mutational studies on the H5N1 hemagglutinin SP, retrieved from the GenBank, categorized SPs into two different structural subclasses: 1) uncharged, with a compact conformation 2) carrying +2 electric charges, with an extended conformation. This study assumed that SPs undergo mutations more rapidly than the proteins in order to assume the appropriate conformational state (Weltman et al., 2007). Another study also showed

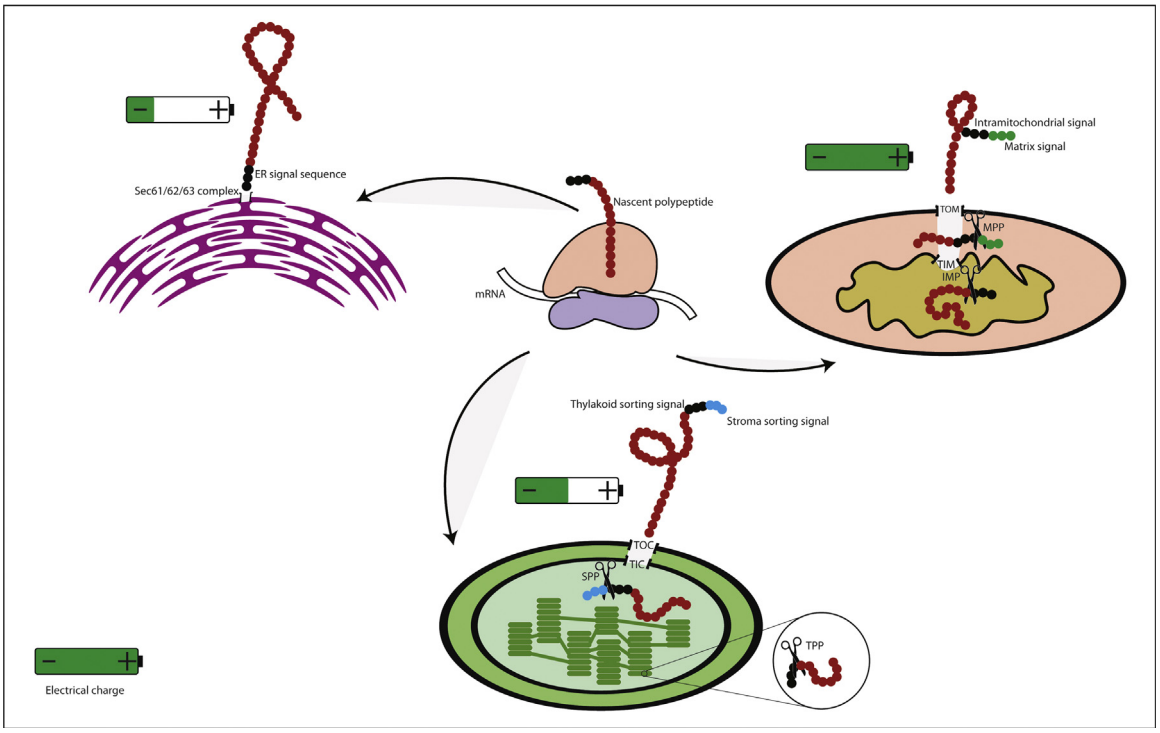

Fig. 4. Signal peptides of different organelles. ER signal sequence, mitochondrial targeting peptide, and chloroplast transit peptide are distinguished on the basis of structural difference and electrical charge. ER signal sequence translocates protein through the Sec61 complex. Mitochondrial targeting peptide and chloroplast transit peptide translocate protein through translocons of the outer and inner membranes, called TOM and TIM (in the mitochondria), TOC and TIC in the chloroplast.

that there is a constant selective pressure on the charge of SPs. A minimal change in charge of a SP determines its targeting location. Presence of the electrochemical gradient across the organelle's membrane favors the positive charge of the SP. The positive charge of SP for mitochondrial proteins is selected as a result of the presence of an electrochemical gradient across the inner membrane of mitochondria. On the other hand, the less positively charged SPs harboring bulky aromatic aas target the protein to the plastid. That difference could be attributed to these organelles' bacterial ancestors as well as the impact of selective pressure. Moreover, the eukaryotic SPs are less positively charged than their prokaryotic counterparts (Garg and Gould, 2016). Hence, through the evolution, the charge and composition of SPs have changed with respect to their targets.

8. Current limitations and prospects

8.1. Limitation of signal peptide application

SPs represent a promising role in recombinant protein production, whereas one should tackle several hindrances to reach the best performance of SPs. Talking about *E. coli* as one of the most encouraging biological production systems, a major obstacle shows up. Recombinant proteins harboring SP mostly pass through the inner membrane of *E. coli*; however, passage through both membranes is somewhat impossible (Mergulhao et al., 2005). There are few studies denoting complete excretion of recombinant proteins from *E. coli.* Even if such a study exists, the culture medium was optimized, or the SP remained attached to the protein (Gao et al., 2016; Lee et al., 2001). The aforementioned SPs were notably different from a typical SP and the secretion mechanisms were unknown. Since *E. coli* is one of the most common hosts for production of recombinant proteins, SPs appropriate for complete secretion requires additional investigation. Several factors other than SP work hand in hand to lead to an efficient secretion. For example, the fusion of SP to the protein might affect the mRNA stability; therefore, different levels of mRNA and protein precursors are produced. Undeniably, the effective interaction between the SP and the secretory apparatus along with the availability of chaperones and proteases significantly determine protein translocation. For heterologous protein secretion, an optimal balance between targeting efficiency and folding capacity should be maintained. If the protein targeting is increased inappropriately to the folding capacity, the improperly folded proteins will accumulate at the translocation site, e.g., the cell wall. Thus, stress signal related proteases are activated to degrade the improperly folded proteins. Unfortunately, the properly folded proteins get degraded in addition to the unfolded ones resulting in an overall reduction of secretion efficiency (Hyyryläinen et al., 2001; Darmon et al., 2002). Therefore, applying the best SP for one protein to another protein would not always be pragmatic. A further downside is that although it is generally believed that SP utilization can reduce inclusion body formation and protein aggregation, some exceptions are seen. An instance of SP inapplicability to the protein is the *E. coli* thioredoxin, in which the protein without SP had higher secretion than the one with the SP fusion. Fusion of malE with the SP led to thermodynamic destabilization and irreversible aggregation of the unfolded protein (Singh et al., 2013). Consequently, SPs do not always promise an efficient secretion.

8.2. Limitations of studying signal peptides

Some earlier studies concluded that SPs are such similar and flexible that could be easily interchanged and applied to heterologous proteins (Rapoport et al., 1996). However, this concept could be the subject of debate from several aspects. First of all, one might wonder how the components of a secretory system face with the substantial degeneracy of SPs as well as maintaining high specificity and fidelity. Additionally, SPs have undergone changes through the evolution, which question the sequence diversity and flexibility of SPs. In other words, specific evolutionary conservation has been observed among SPs, which refuses the substantial possibility of SP interchanging. Moreover, considerable functions attributed to SPs other than their role in the protein

biogenesis imply that there should be a level of structural specificity among them. Several site-directed mutagenesis studies or exchanging SPs corroborate the idea that there should be an optimal consistency between the SP and the passenger protein (Weltman et al., 2007; Pfeiffer et al., 2006). Different combination of SPs and passenger protein affect protein translocation efficiency by controlling the translocation machine in a substrate-specific manner (Kim et al., 2002). These substrate-specific properties of signals were evolutionarily conserved and fundamental for the proper protein biogenesis. On the other hand, there are few studies regarding the high translocation efficiency of SPs for heterologous proteins or great similarity of SPs of various origins. For example, the SPs of cone snail (*Conus*) peptides, as a rich pharmaceutical treasure, were all highly conserved among all members of the same superfamily (Wang and Chi, 2004). However, based on few conflicting studies, over-generalization should not be done. Although the examples of similar SPs for various proteins are intriguing to be recruited in recombinant proteins production, in the future studies, SPs should be precisely scrutinized regarding their cargo and host. Designing SPs should be substrate-specific rather than based on some general rules.

To date, computational SP prediction methods are based on the information gained from a limited number of experimentally validated SPs. Undoubtedly, there is a need for high-throughput methods of identification of SPs. Genome-scale identification, e.g. proteogenomics could be implemented, at least for model organisms such as *E. coli*. In one study, only half of the previously estimated SPs were approved by proteogenomics analysis (Ivankov et al., 2013). Surprisingly, through high throughput sequencing of SP libraries in *B. subtilis* and *B. licheniformis*, highly efficient SPs did not share any significant similarities in hydrophobicity, N-domain positive charge, and/or cleavage site. Moreover, there was no relationship between the calculated D-score (the score showing the probability of being an SP), and the secretion efficiency (Degering et al., 2010; Brockmeier et al., 2006). Therefore, a detailed analysis of the best-performing SPs with respect to the protein and the host as well as more powerful prediction tools are needed.

A further minus side of studying SP concerns long SPs. Limited data is known regarding long SPs. There is a lack of appropriate software tools to predict long SPs. Protein translocation insufficiency in short SPs might be related to the lack of protein-protein interactions, which is conferred by long SPs. Moreover, considering the vast diversity of functions other than protein translocation that are attributed to the long SPs, it is crucial to study and take advantage of them in the future. Another flaw in studying SPs is that all proteins harboring SP are not secreted out of the cell. It was assumed that 5–20% of proteins with SP might not translocate across the ER membrane and would remain in the cytosol (Hegde and Bernstein, 2006). This flaw leads to researchers' misconduct at ease. Perhaps it is better to rely on SPs that were approved to be secreted out of the cell.

Conflict of interest

Authors declare no conflict of interests.

Acknowledgment

This study was supported by Grant no. 13435 from the Research Council of Shiraz University of Medical Sciences, Shiraz University of Medical Sciences, Shiraz, Iran.

References

Adams, H., Scotti, P.A., de Cock, H., et al., 2002. The presence of a helix breaker in the hydrophobic core of signal sequences of secretory proteins prevents recognition by the signal-recognition particle in *Escherichia coli*. FEBS J. 269 (22), 5564–5571. http://dx.doi.org/10.1046/j.1432-1033.2002.03262.x.

Agnihothram, S.S., York, J., Trahey, M., et al., 2007. Bitopic membrane topology of the stable signal peptide in the tripartite Junín virus GP-C envelope glycoprotein complex. J Virol. 81 (8), 4331–4337. http://dx.doi.org/10.1128/JVI.02779-06.

Ahmed, A.-Q., Teilhet, M., MENSA-WILMOT, K., 1998. Species-specificity in endoplasmic reticulum signal peptide utilization revealed by proteins from *Trypanosoma brucei* and *Leishmania*. Biochem J. 331 (2), 521–529. http://dx.doi.org/10.1042/bj3310521.

Ahn, J.-H., Hwang, M.-Y., Lee, K.-H., et al., 2006. Use of signal sequences as an in situ removable sequence element to stimulate protein synthesis in cell-free extracts. Nucleic Acids Res. 35 (4), e21. http://dx.doi.org/10.1093/nar/gkl917.

Aladin, F., Lautscham, G., Humphries, E., et al., 2007. Targeting tumour cells with defects in the MHC Class I antigen processing pathway with CD8+ T cells specific for hydrophobic TAP-and Tapasin-independent peptides: the requirement for directed access into the ER. Cancer Immunol. Immunother. 56 (8), 1143–1152. http://dx.doi.org/10.1007/s00262-006-0263-2.

Amaya, Y., Nakai, T., Miura, S., 2015. Evolutionary well-conserved region in the signal peptide of parathyroid hormone-related protein is critical for its dual localization through the regulation of ER translocation. J. Biochem. 159 (4), 393–406. http://dx.doi.org/10.1093/jb/mvv111.

Anandatheerthavarada, H.K., Biswas, G., Mullick, J., et al., 1999. Dual targeting of cytochrome P4502B1 to endoplasmic reticulum and mitochondria involves a novel signal activation by cyclic AMP-dependent phosphorylation at ser128. EMBO J. 18 (20), 5494–5504. http://dx.doi.org/10.1093/emboj/18.20.5494.

Auclair, S.M., Bhanu, M.K., Kendall, D.A., 2012. Signal peptidase I: cleaving the way to mature proteins. Protein Sci. 21 (1), 13–25. http://dx.doi.org/10.1002/pro.757.

von Heijne, G., 1990a. The signal peptide. J. Membr. Biol. 115 (3), 195–201. http://dx.doi.org/10.1007/BF01868635.

Von Heijne, G., 1990b. Protein targeting signals. Curr. Opin. Cell Biol. 2 (4), 604–608. http://dx.doi.org/10.1016/0955-0674(90)90100-S.

Baba, T., Schneewind, O., 1996. Target cell specificity of a bacteriocin molecule: a C-terminal signal directs lysostaphin to the cell wall of *Staphylococcus aureus*. EMBO J. 15 (18), 4789.

Barkocy-Gallagher, G., Bassford, P., 1992. Synthesis of precursor maltose-binding protein with proline in the+ 1 position of the cleavage site interferes with the activity of *Escherichia coli* signal peptidase I in vivo. J. Biol. Chem. 267 (2), 1231–1238.

Batey, R.T., Rambo, R.P., Lucast, L., et al., 2000. Crystal structure of the ribonucleoprotein core of the signal recognition particle. Science 287 (5456), 1232–1239. http://dx.doi.org/10.1126/science.287.5456.1232.

Bederka, L.H., Bonhomme, C.J., Ling, E.L., et al., 2014. Arenavirus stable signal peptide is the keystone subunit for glycoprotein complex organization. MBio 5 (6), e02063–14. http://dx.doi.org/10.1128/mBio.02063-14.

Belin, D., Guzman, L.-M., Bost, S., et al., 2004. Functional activity of eukaryotic signal sequences in *Escherichia coli:* the ovalbumin family of serine protease inhibitors. J. Mol. Biol. 335 (2), 437–453. http://dx.doi.org/10.1016/j.jmb.2003.10.076.

Bendtsen, J.D., Kiemer, L., Fausbøll, A., et al., 2005. Non-classical protein secretion in bacteria. BMC Microbiol. 5 (1), 58. http://dx.doi.org/10.1186/1471-2180-5-58.

Bensing, B.A., Seepersaud, R., Yen, Y.T., et al., 2014. Selective transport by SecA2: an expanding family of customized motor proteins. Biochem. Biophys. Acta 1843 (8), 1674–1686. http://dx.doi.org/10.1016/j.bbamcr.2013.10.019.

Berks, B.C., 2015. The twin-arginine protein translocation pathway. Annu. Rev. Biochem. 84, 843–864. http://dx.doi.org/10.1146/annurev-biochem-060614-034251.

Berks, B.C., Sargent, F., Palmer, T., 2000. The Tat protein export pathway. Mol. Microbiol. 35 (2), 260–274. http://dx.doi.org/10.1046/j.1365-2958.2000.01719.x.

Blobel, G., 2000. Protein targeting (Nobel lecture). Chembiochem 1 (2), 86–102. http://dx.doi.org/10.1002/1439-7633(20000818)1:2%3C86::AID-CBIC86%3E3.0.CO;2-A.

Blobel, G., Sabatini, D.D., 1971. In: Manson, L.A. (Ed.), Ribosome-Membrane Interaction in Eukaryotic Cells. Biomembranes. Plenum Press, Springer, New York p. 193–5.

Bonfanti, R., Colombo, C., Nocerino, V., et al., 2009. Insulin gene mutations as cause of diabetes in children negative for five type 1 diabetes autoantibodies. Diabetes Care 32 (1), 123–125. http://dx.doi.org/10.2337/dc08-0783.

Bornemann, T., Jöckel, J., Rodnina, M.V., et al., 2008. Signal sequence–independent membrane targeting of ribosomes containing short nascent peptides within the exit tunnel. Nat. Struct. Mol. Biol. 15 (5), 494–499. http://dx.doi.org/10.1038/nsmb.1402.

Bowers, C.W., Lau, F., Silhavy, T.J., 2003. Secretion of LamB-LacZ by the signal recognition particle pathway of *Escherichia coli*. J. Bacteriol. 185 (19), 5697–5705. http://dx.doi.org/10.1128/JB.185.19.5697-5705.2003.

Brockmeier, U., Caspers, M., Freudl, R., et al., 2006. Systematic screening of all signal peptides from *Bacillus subtilis:* a powerful strategy in optimizing heterologous protein secretion in Gram-positive bacteria. J. Mol. Biol. 362 (3), 393–402. http://dx.doi.org/10.1016/j.jmb.2006.07.034.

Büttner, D., 2012. Protein export according to schedule: architecture, assembly, and regulation of type III secretion systems from plant-and animal-pathogenic bacteria. Microbiol. Mol. Biol. Rev. 76 (2), 262–310. http://dx.doi.org/10.1128/MMBR.05017-11.

Carlos, J.L., Paetzel, M., Brubaker, G., et al., 2000. The role of the membrane-spanning domain of type I signal peptidases in substrate cleavage site selection. J. Biol. Chem. 275 (49), 38813–38822. http://dx.doi.org/10.1074/jbc.M007093200.

Chang, H.T., Kao, Y.L., Wu, C.M., et al., 2007. Signal peptide of eosinophil cationic protein upregulates transforming growth factor-alpha expression in human cells. J. Cell. Biochem. 100 (5), 1266–1275. http://dx.doi.org/10.1002/jcb.21120.

Chen, H., Kim, J., Kendall, D.A., 1996. Competition between functional signal peptides demonstrates variation in affinity for the secretion pathway. J. Bacteriol. 178 (23), 6658–6664.

Chen, K., Liu, S., Wang, G., et al., 2013. Enhancement of *Streptomyces* transglutaminase activity and pro-peptide cleavage efficiency by introducing linker peptide in the C-terminus of the pro-peptide. J. Ind. Microbiol. Biotechnol. 40 (3-4), 317–325. http://dx.doi.org/10.1007/s10295-012-1221-y.

Cheng, Y., Liu, S., Lu, C., et al., 2016. Missense mutations in the signal peptide of the

porcine GH. Pituitary 19 (4), 362–369. http://dx.doi.org/10.1007/s11102-016-0713-6.
Choo, K.H., Ranganathan, S., 2008. Flanking signal and mature peptide residues influence signal peptide cleavage. BMC Bioinform. 9 (12), S15. http://dx.doi.org/10.1186/1471-2105-9-S12-S15.
Choo, K.H., Tan, T.W., Ranganathan, S., 2005. SPdb–a signal peptide database. BMC Bioinform. 6 (1), 249. http://dx.doi.org/10.1186/1471-2105-6-249.
Colombo, S., Longhi, R., Alcaro, S., et al., 2005. N-myristoylation determines dual targeting of mammalian NADH-cytochrome b (5) reductase to ER and mitochondrial outer membranes by a mechanism of kinetic partitioning. J. Cell Biol. 168 (5), 735–745. http://dx.doi.org/10.1083/jcb.200407082.
Dalbey, R.E., Lively, M.O., Bron, S., et al., 1997. The chemistry and enzymology of the type I signal peptidases. Prot. Sci. 6 (6), 1129–1138. http://dx.doi.org/10.1002/pro.5560060601.
Darmon, E., Noone, D., Masson, A., et al., 2002. A novel class of heat and secretion stress-responsive genes is controlled by the autoregulated CssRS two-component system of Bacillus subtilis. J. Bacteriol. 184 (20), 5661–5671. http://dx.doi.org/10.1128/JB.184.20.5661-5671.2002.
Datta, R., Waheed, A., Shah, G.N., et al., 2007. Signal sequence mutation in autosomal dominant form of hypoparathyroidism induces apoptosis that is corrected by a chemical chaperone. Proc. Natl. Acad. Sci. 104 (50), 19989–19994. http://dx.doi.org/10.1073/pnas.0708725104.
De Gier, J.-W.L., Scotti, P.A., Sääf, A., et al., 1998. Differential use of the signal recognition particle translocase targeting pathway for inner membrane protein assembly in *Escherichia coli*. Proc. Natl. Acad. Sci. 95 (25), 14646–14651.
Degering, C., Eggert, T., Puls, M., et al., 2010. Optimization of protease secretion in *Bacillus subtilis* and *Bacillus licheniformis* by screening of homologous and heterologous signal peptides. Appl. Environ. Microbiol. 76 (19), 6370–6376. http://dx.doi.org/10.1128/AEM.01146-10.
Deshaies, R.J., Sanders, S.L., 1991. Assembly of yeast Sec proteins involved in translocation into the endoplasmic reticulum into a membrane-bound multisubunit complex. Nature 349 (6312), 806. http://dx.doi.org/10.1038/349806a0.
Dirican, N., Duman, A., Sağlam, G., et al., 2016. The diagnostic significance of signal peptide-complement C1r/C1s, Uegf, and Bmp1-epidermal growth factor domain-containing protein-1 levels in pulmonary embolism. Ann. Thorac. Med. 11 (4), 277. http://dx.doi.org/10.4103/1817-1737.191876.
Driessen, A.J., Van der Does, C., 2002. Protein export in bacteria. In: Ross Dalbey, GvH (Ed.), Protein Targeting, Transport and Translocation Academic, 1st edn ed. G. Academic Press, San Diego p. 47-73.
Duda, A., Stange, A., Lüftenegger, D., et al., 2004. Prototype foamy virus envelope glycoprotein leader peptide processing is mediated by a furin-like cellular protease, but cleavage is not essential for viral infectivity. J. Virol. 78 (24), 13865–13870. http://dx.doi.org/10.1128/JVI.78.24.13865-13870.2004.
Duffy, J., Patham, B., Mensa-Wilmot, K., 2010. Discovery of functional motifs in h-regions of trypanosome signal sequences. Biochem. J. 426 (2), 135–145. http://dx.doi.org/10.1042/BJ20091277.
Dultz, E., Hildenbeutel, M., Martoglio, B., et al., 2008. The signal peptide of the mouse mammary tumor virus Rem protein is released from the endoplasmic reticulum membrane and accumulates in nucleoli. J. Biol. Chem. 283 (15), 9966–9976. http://dx.doi.org/10.1074/jbc.M705712200.
Ebel, T., Pellé, R., Janoo, R., et al., 2004. A membrane-anchored *Theileria parva* cyclophilin with a non-cleaved amino-terminal signal peptide for entry into the endoplasmic reticulum. Vet. Parasitol. 121 (1), 65–77. http://dx.doi.org/10.1016/j.vetpar.2004.02.007.
Economou, A., Wickner, W., 1994. SecA promotes preprotein translocation by undergoing ATP-driven cycles of membrane insertion and deinsertion. Cell 78 (5), 835–843. http://dx.doi.org/10.1016/S0092-8674(94)90582-7.
Edman, M., Jarhede, T., Sjöström, M., et al., 1999. Different sequence patterns in signal peptides from mycoplasmas, other gram-positive bacteria, and *Escherichia coli*: a multivariate data analysis. Proteins: Struct. Funct. Bioinf. 35 (2), 195–205. http://dx.doi.org/10.1002/(SICI)1097-0134(19990501)35:2%3C195::AID-PROT6%3E3.0.CO;2-P.
Eichler, R., Lenz, O., Strecker, T., et al., 2003. Identification of *Lassa virus* glycoprotein signal peptide as a trans-acting maturation factor. EMBO Rep. 4 (11), 1084–1088. http://dx.doi.org/10.1038/sj.embor.7400002.
Fath, M.J., Kolter, R., 1993. ABC transporters: bacterial exporters. Microbiol. Rev. 57 (4), 995–1017.
Fikes, J., Barkocy-Gallagher, G., Klapper, D., et al., 1990. Maturation of *Escherichia coli* maltose-binding protein by signal peptidase I in vivo. Sequence requirements for efficient processing and demonstration of an alternate cleavage site. J. Biol. Chem. 265 (6), 3417–3423.
Fons, R.D., Bogert, B.A., Hegde, R.S., 2003. Substrate-specific function of the translocon-associated protein complex during translocation across the ER membrane. J. Cell Biol. 160 (4), 529–539. http://dx.doi.org/10.1083/jcb.200210095.
Futatsumori-Sugai, M., Tsumoto, K., 2010. Signal peptide design for improving recombinant protein secretion in the baculovirus expression vector system. Biochem. Biophys. Res. Commun. 391 (1), 931–935. http://dx.doi.org/10.1016/j.bbrc.2009.11.167.
Gao, D., Luan, Y., Liang, Q., et al., 2016. Exploring the N-terminal role of a heterologous protein in secreting out of *Escherichia coli*. Biotechnol Bioeng. 113 (12), 2561–2567. http://dx.doi.org/10.1002/bit.26028.
Gao, S., Yao, S., Hart, D.J., et al., 2017. Signal peptide-dependent protein translocation pathway is crucial for the sucrose sensitivity of SacB-expressing *Escherichia coli*. Biochem. Eng. J. 122, 71–74. http://dx.doi.org/10.1016/j.bej.2017.03.002.
Garg, S.G., Gould, S.B., 2016. The role of charge in protein targeting evolution. Trends Cell Biol. 26 (12), 894–905. http://dx.doi.org/10.1016/j.tcb.2016.07.001.
Geukens, N., Frederix, F., Reekmans, G., et al., 2004. Analysis of type I signal peptidase affinity and specificity for preprotein substrates. Biochem. Biophys. Res. Commun. 314 (2), 459–467. http://dx.doi.org/10.1016/j.bbrc.2003.12.122.
Gewurz, B.E., Ploegh, H.L., Tortorella, D., 2002. US2, a human cytomegalovirus-encoded type I membrane protein, contains a non-cleavable amino-terminal signal peptide. J. Biol. Chem. 277 (13), 11306–11313. http://dx.doi.org/10.1074/jbc.M107904200.
Ghosh, J., Anderson, P.J., Chandrasekaran, S., et al., 2010. Characterization of *Streptococcus pyogenes* β-NAD+ Glycohydrolase re-evaluation of enzymatic properties associated with pathogenesis. J. Biol. Chem. 285 (8), 5683–5694. http://dx.doi.org/10.1074/jbc.M109.070300.
Görlich, D., Hartmann, E., Prehn, S., et al., 1992. A protein of the endoplasmic reticulum involved early in polypeptide translocation. Nature 357 (6373), 47–52. http://dx.doi.org/10.1038/357047a0.
Green, E.R., Mecsas, J., 2016. Bacterial secretion systems–an overview. Microbiol. Spectr. 4 (1). http://dx.doi.org/10.1128/microbiolspec.VMBF-0012-2015.
Green, R., Kramer, R., Shields, D., 1989. Misplacement of the amino-terminal positive charge in the prepro-alpha-factor signal peptide disrupts membrane translocation in vivo. J. Biol. Chem. 264 (5), 2963–2968.
Han, S., Machhi, S., Berge, M., et al., 2017. Novel signal peptides improve the secretion of recombinant *Staphylococcus aureus* Alpha toxin H35L in *Escherichia coli*. AMB Exp. 7 (1), 93. http://dx.doi.org/10.1186/s13568-017-0394-1.
Hansson, C.A., Frykman, S., Farmery, M.R., et al., 2004. Nicastrin, presenilin, APH-1, and PEN-2 form active γ-secretase complexes in mitochondria. J. Biol. Chem. 279 (49), 51654–51660. http://dx.doi.org/10.1074/jbc.M404500200.
Hardy, S.J., Randall, L.L., 1991. A kinetic partitioning model of selective binding of nonnative proteins by the bacterial chaperone SecB. Science 251 (4992), 439.
Hartmann, E., Rapoport, T.A., Lodish, H.F., 1989. Predicting the orientation of eukaryotic membrane-spanning proteins. Proc. Natl. Acad. Sci. 86 (15), 5786–5790.
Haryadi, R., Ho, S., Kok, Y.J., et al., 2015. Optimization of heavy chain and light chain signal peptides for high level expression of therapeutic antibodies in CHO cells. PloS One 10 (2), e0116878. http://dx.doi.org/10.1371/journal.pone.0116878.
Hayashi, S., Wu, H.C., 1990. Lipoproteins in bacteria. J. Bioenergy Biomembr. 22 (3), 451–471.
Hegde, R.S., Bernstein, H.D., 2006. The surprising complexity of signal sequences. Trends Biochem. Sci. 31 (10), 563–571. http://dx.doi.org/10.1016/j.tibs.2006.08.004.
Hikita, C., Mizushima, S., 1992a. The requirement of a positive charge at the amino terminus can be compensated for by a longer central hydrophobic stretch in the functioning of signal peptides. J. Biol. Chem. 267 (17), 12375–12379.
Hikita, C., Mizushima, S., 1992b. Effects of total hydrophobicity and length of the hydrophobic domain of a signal peptide on in vitro translocation efficiency. J. Biol. Chem. 267 (7), 4882–4888.
Hiss, J.A., Schneider, G., 2009. Architecture, function and prediction of long signal peptides. Brief. Bioinform. 10 (5), 569–578. http://dx.doi.org/10.1093/bib/bbp030.
Hiss, J.A., Resch, E., Schreiner, A., et al., 2008. Domain organization of long signal peptides of single-pass integral membrane proteins reveals multiple functional capacity. PLoS One 3 (7), e2767. http://dx.doi.org/10.1371/journal.pone.0002767.
Humphery-Smith, I., Hecker, M., 2006. Microbial Proteomics: Functional Biology of Whole Organisms. John Wiley & Sons.
Hyyryläinen, H.L., Bolhuis, A., Darmon, E., et al., 2001. A novel two-component regulatory system in *Bacillus subtilis* for the survival of severe secretion stress. Mol. Microbiol. 41 (5), 1159–1172. http://dx.doi.org/10.1046/j.1365-2958.2001.02576.x.
Ismail, N.F., Hamdan, S., Mahadi, N.M., et al., 2011. A mutant L-asparaginase II signal peptide improves the secretion of recombinant cyclodextrin glucanotransferase and the viability of *Escherichia coli*. Biotechnol. Lett. 33 (5), 999–1005. http://dx.doi.org/10.1007/s10529-011-0517-8.
Ivankov, D.N., Payne, S.H., Galperin, M.Y., et al., 2013. How many signal peptides are there in bacteria? Environ. Microbiol. 15 (4), 983–990. http://dx.doi.org/10.1111/1462-2920.12105.
Jarjanazi, H., Savas, S., Pabalan, N., et al., 2008. Biological implications of SNPs in signal peptide domains of human proteins. Proteins 70 (2), 394–403. http://dx.doi.org/10.1002/prot.21548.
Jonet, M.A., Mahadi, N.M., Murad, A.M.A., et al., 2012. Optimization of a heterologous signal peptide by site-directed mutagenesis for improved secretion of recombinant proteins in *Escherichia coli*. J. Mol. Microbiol. Biotechnol. 22 (1), 48–58. http://dx.doi.org/10.1159/000336524.
Joshi, S.N., Butler, D.C., Messer, A. (Eds.), 2012. Fusion to a Highly Charged Proteasomal Retargeting Sequence Increases Soluble Cytoplasmic Expression and Efficacy of Diverse Anti-Synuclein Intrabodies. MAbs. Taylor & Francis. http://dx.doi.org/10.4161/mabs.21696.
Kanapin, A., Batalov, S., Davis, M.J., et al., 2003. Mouse proteome analysis. Genome Res. 13 (6b), 1335–1344. http://dx.doi.org/10.1101/gr.978703.
Kang, Z., Yang, S., Du, G., et al., 2014. Molecular engineering of secretory machinery components for high-level secretion of proteins in *Bacillus* species. J. Ind. Microbiol. Biotechnol. 41 (11), 1599–1607. http://dx.doi.org/10.1007/s10295-014-1506-4.
Kang, D.G., Seo, J.H., Jo, B.H., et al., 2016. Versatile signal peptide of Flavobacterium-originated organophosphorus hydrolase for efficient periplasmic translocation of heterologous proteins in *Escherichia coli*. Biotechnol. Prog. 32 (4), 848–854. http://dx.doi.org/10.1002/btpr.2274.
Kapp, K., Schrempf, S., Lemberg, M.K., et al., 2013. Post-targeting functions of signal peptides. Madame Curie Bioscience Database [Internet]. Landes Bioscience, Austin (TX).
Karamyshev, A.L., Karamysheva, Z.N., Kajava, A.V., et al., 1998. Processing of *Escherichia coli* alkaline phosphatase: role of the primary structure of the signal peptide cleavage region. J. Mol. Biol. 277 (4), 859–870. http://dx.doi.org/10.1006/jmbi.1997.1617.
Kim, S.J., Mitra, D., Salerno, J.R., et al., 2002. Signal sequences control gating of the

protein translocation channel in a substrate-specific manner. Dev. Cell 2 (2), 207–217. http://dx.doi.org/10.1016/S1534-5807(01)00120-4.
Kim, C.-K., Lee, S.-Y., Kwon, O.-J., et al., 2007. Secretory expression of active clostripain in *Escherichia coli*. J. Biotechnol. 131 (3), 346–352. http://dx.doi.org/10.1016/j.jbiotec.2007.07.936.
Kober, L., Zehe, C., Bode, J., 2013. Optimized signal peptides for the development of high expressing CHO cell lines. Biotechnol. Bioeng. 110 (4), 1164–1173 0.1002/bit.24776.
Köhler, D., Dobritzsch, D., Hoehenwarter, W., et al., 2015. Identification of protein N-termini in *Cyanophora paradoxa* cyanelles: transit peptide composition and sequence determinants for precursor maturation. Front. Plant Sci. 6. http://dx.doi.org/10.3389/fpls.2015.00559.
Korotkov, K.V., Sandkvist, M., Hol, W.G., 2012. The type II secretion system: biogenesis, molecular architecture and mechanism. Nat. Rev. Microbiol. 10 (5), 336. http://dx.doi.org/10.1038/nrmicro2762.
Kovjazin, R., Carmon, L., 2014. The use of signal peptide domains as vaccine candidates. Hum. Vaccin Immunother. 10 (9), 2733–2740. http://dx.doi.org/10.4161/21645515.2014.970916.
Kovjazin, R., Volovitz, I., Daon, Y., et al., 2011. Signal peptides and trans-membrane regions are broadly immunogenic and have high CD8+ T cell epitope densities: implications for vaccine development. Mol. Immunol. 48 (8), 1009–1018. http://dx.doi.org/10.1016/j.molimm.2011.01.006.
Kurys, G., Tagaya, Y., Bamford, R., et al., 2000. The long signal peptide isoform and its alternative processing direct the intracellular trafficking of interleukin-15. J. Biol. Chem. 275 (39), 30653–30659. http://dx.doi.org/10.1074/jbc.M002373200.
Kvam, E., Sierks, M.R., Shoemaker, C.B., et al., 2010. Physico-chemical determinants of soluble intrabody expression in mammalian cell cytoplasm. Prot. Eng. Des. Sel. 23 (6), 489–498. http://dx.doi.org/10.1093/protein/gzq022.
Lammertyn, E., Anné, J., 1998. Modifications of Streptomyces signal peptides and their effects on protein production and secretion. FEMS Microbiol. Lett. 160 (1), 1–10. http://dx.doi.org/10.1111/j.1574-6968.1998.tb12882.x.
Lange, S., Franks, W.T., Rajagopalan, N., et al., 2016. Structural analysis of a signal peptide inside the ribosome tunnel by DNP MAS NMR. Sci. Adv. 2 (8), e1600379. http://dx.doi.org/10.1126/sciadv.1600379.
Lee, J., Saraswat, V., Koh, I., et al., 2001. Secretory production of Arthrobacter levan fructotransferase from recombinant *Escherichia coli*. FEMS Microbiol. Lett. 195 (2), 127–132. http://dx.doi.org/10.1111/j.1574-6968.2001.tb10509.x.
Leyton, D.L., Rossiter, A.E., Henderson, I.R., 2012. From self sufficiency to dependence: mechanisms and factors important for autotransporter biogenesis. Nat. Rev. Microbiol. 10 (3), 213. http://dx.doi.org/10.1038/nrmicro2733.
Li, Y., Bergeron, J., Luo, L., et al., 1996. Effects of inefficient cleavage of the signal sequence of HIV-1 gp 120 on its association with calnexin, folding, and intracellular transport. Proc. Natl. Acad. Sci. 93 (18), 9606–9611.
Li, X., Wang, H., Zhou, C., et al., 2014. Cloning, expression and characterization of a pectate lyase from *Paenibacillus* sp. 0602 in recombinant *Escherichia coli*. BMC Biotechnol. 14 (1), 18. http://dx.doi.org/10.1186/1472-6750-14-18.
Liljeström, P., Garoff, H., 1991. Internally located cleavable signal sequences direct the formation of Semliki Forest virus membrane proteins from a polyprotein precursor. J. Virol. 65 (1), 147–154.
Liu, S., Zhang, D., Wang, M., et al., 2011. The pro-region of *Streptomyces hygroscopicus* transglutaminase affects its secretion by *Escherichia coli*. FEMS Microbiol. Lett. 324 (2), 98–105. http://dx.doi.org/10.1111/j.1574-6968.2011.02387.x.
Liu, H., Wu, R., Yuan, L., et al., 2017. Introducing a cleavable signal peptide enhances the packaging efficiency of lentiviral vectors pseudotyped with Japanese encephalitis virus envelope proteins. Virus Res. 229, 9–16. http://dx.doi.org/10.1016/j.virusres.2016.12.007.
Lobigs, M., Zhao, H., Garoff, H., 1990. Function of Semliki Forest virus E3 peptide in virus assembly: replacement of E3 with an artificial signal peptide abolishes spike heterodimerization and surface expression of E1. J. Virol. 64 (9), 4346–4355.
Loewy, A., Smyth, J., Von Bonsdorff, C., et al., 1995. The 6-kilodalton membrane protein of Semliki Forest virus is involved in the budding process. J. Virol. 69 (1), 469–475.
Low, K.O., Mahadi, N.M., Illias, R.M., 2013. Optimisation of signal peptide for recombinant protein secretion in bacterial hosts. Appl. Microbiol. Biotechnol. 97 (9), 3811. http://dx.doi.org/10.1007/s00253-013-4831-z.
Lu, X., Wang, G., Feng, Y., et al., 2016. The N-terminal α-helix domain of *Pseudomonas aeruginosa* Lipoxygenase Is required for its soluble expression in *Escherichia coli* but not for catalysis. J. Microbiol. Biotechnol. 26 (10), 1701–1707. http://dx.doi.org/10.4014/jmb.1602.02027.
Luirink, J., Dobberstein, B., 1994. Mammalian and *Escherichia coli* signal recognition particles. Mol. Microbiol. 11 (1), 9–13. http://dx.doi.org/10.1111/j.1365-2958.1994.tb00284.x.
Madden, J.C., Ruiz, N., Caparon, M., 2001. Cytolysin-mediated translocation (CMT): a functional equivalent of type III secretion in gram-positive bacteria. Cell 104 (1), 143–152. http://dx.doi.org/10.1016/S0092-8674(01)00198-2.
Martin, J., Mahlke, K., Pfanner, N., 1991. Role of an energized inner membrane in mitochondrial protein import. Delta psi drives the movement of presequences. J. Biol. Chem. 266 (27), 18051–18057.
Martoglio, B., Dobberstein, B., 1998. Signal sequences: more than just greasy peptides. Trends Cell Biol. 8 (10), 410–415. http://dx.doi.org/10.1016/S0962-8924(98)01360-9.
Martoglio, B., Graf, R., Dobberstein, B., 1997. Signal peptide fragments of preprolactin and HIV-1 p-gp160 interact with calmodulin. EMBO J. 16 (22), 6636–6645. http://dx.doi.org/10.1093/emboj/16.22.6636.
Mathiesen, G., Sveen, A., Piard, J.C., et al., 2008. Heterologous protein secretion by *Lactobacillus plantarum* using homologous signal peptides. J. Appl. Microbiol. 105 (1), 215–226. http://dx.doi.org/10.1111/j.1365-2672.2008.03734.x.
Mergulhao, F., Summers, D.K., Monteiro, G.A., 2005. Recombinant protein secretion in *Escherichia coli*. Biotechnol. Adv. 23 (3), 177–202. http://dx.doi.org/10.1016/j.biotechadv.2004.11.003.
Milstein, C., Brownlee, G., Harrison, T.M., et al., 1972. A possible precursor of immunoglobulin light chains. Nat. New Biol. 239 (91), 117–120. http://dx.doi.org/10.1038/newbio239117a0.
Mistou, M.-Y., Dramsi, S., Brega, S., et al., 2009. Molecular dissection of the secA2 locus of group B *Streptococcus* reveals that glycosylation of the Srr1 LPXTG protein is required for full virulence. J. Bacteriol. 191 (13), 4195–4206. http://dx.doi.org/10.1128/JB.01673-08.
Mordkovich, N., Okorokova, N., Veiko, V., 2015. Structural and functional organization of the signal peptide of pro-enterotoxin B from *Staphylococcus aureus*. Appl. Biochem. Microbiol. 51 (6), 641. http://dx.doi.org/10.1134/S0003683815060101.
Mori, H., Araki, M., Hikita, C., et al., 1997. The hydrophobic region of signal peptides is involved in the interaction with membrane-bound SecA. Biochim. Biophys. Acta 1326 (1), 23–36. http://dx.doi.org/10.1016/S0005-2736(97)00004-7.
Mousavi, P., Mostafavi-Pour, Z., Morowvat, M.H., et al., 2017. In silico analysis of several signal peptides for the excretory production of reteplase in *Escherichia coli*. Curr. Proteom. 14 (4), 326–335. http://dx.doi.org/10.2174/1570164614666170809144446.
Müller, M., 2005. Twin-arginine-specific protein export in *Escherichia coli*. Res. Microbiol. 156 (2), 131–136. http://dx.doi.org/10.1016/j.resmic.2004.09.016.
Nagahora, H., Fujisawa, H., Jigami, Y., 1988. Alterations in the cleavage site of the signal sequence for the secretion of human lysozyme by *Saccharomyces cerevisiae*. FEBS Lett. 238 (2), 329–332. http://dx.doi.org/10.1016/0014-5793(88)80506-4.
Negahdaripour, M., Nezafat, N., Hajighahramani, N., et al., 2017. In silico study of different signal peptides for secretory production of interleukin-11 in *Escherichia coli*. Curr. Prot. 14 (2), 112–121. http://dx.doi.org/10.2174/1570164614666170106110848.
Nesmeyanova, M.A., Karamyshev, A.L., Karamysheva, Z.N., et al., 1997. Positively charged lysine at the N-terminus of the signal peptide of the *Escherichia coli* alkaline phosphatase provides the secretion efficiency and is involved in the interaction with anionic phospholipids. FEBS Lett. 403 (2), 203–207. http://dx.doi.org/10.1016/S0014-5793(97)00052-5.
Ng, D.T., Brown, J.D., Walter, P., 1996. Signal sequences specify the targeting route to the endoplasmic reticulum membrane. J. Cell Biol. 134 (2), 269–278. http://dx.doi.org/10.1083/jcb.134.2.269.
Ng, S.Y., Chaban, B., VanDyke, D.J., et al., 2007. Archaeal signal peptidases. Microbiology 153 (2), 305–314. http://dx.doi.org/10.1099/mic.0.2006/003087-0.
Nielsen, H., Engelbrecht, J., Brunak, S., et al., 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. 10 (1), 1–6. http://dx.doi.org/10.1093/protein/10.1.1.
Nilsson, I., Lara, P., Hessa, T., et al., 2015. The code for directing proteins for translocation across ER membrane: SRP cotranslationally recognizes specific features of a signal sequence. J. Mol. Biol. 427 (6), 1191–1201. http://dx.doi.org/10.1016/j.jmb.2014.06.014.
Ogata, R.T., Mathias, P., Bradt, B.M., et al., 1993. Murine C4b-binding protein. Mapping of the ligand binding site and the N-terminus of the pre-protein. J. Immunol. 150 (6), 2273–2280.
Ohmuro-Matsuyama, Y., Yamaji, H., 2017. Modifications of a signal sequence for antibody secretion from insect cells. Cytotechnology 1–8. http://dx.doi.org/10.1007/s10616-017-0109-0.
Oliver, J., Jungnickel, B., Görlich, D., et al., 1995. The Sec61 complex is essential for the insertion of proteins into the membrane of the endoplasmic reticulum. FEBS Lett. 362 (2), 126–130. http://dx.doi.org/10.1016/0014-5793(95)00223-V.
Ott, C.M., Lingappa, V.R., 2004. Signal sequences influence membrane integration of the prion protein. Biochemistry 43 (38), 11973–11982. http://dx.doi.org/10.1021/bi049156s.
Paetzel, M., Karla, A., Strynadka, N.C., et al., 2002. Signal peptidases. Chem. Rev. 102 (12), 4549–4580. http://dx.doi.org/10.1021/cr010166y.
Palazzo, A.F., Springer, M., Shibata, Y., et al., 2007. The signal sequence coding region promotes nuclear export of mRNA. PLoS Biol. 5 (12), e322. http://dx.doi.org/10.1371/journal.pbio.0050322.
Palmer, T., Berks, B.C., 2012. The twin-arginine translocation (Tat) protein export pathway. Nat. Rev. Microbiol. 10 (7), 483. http://dx.doi.org/10.1038/nrmicro2814.
Papanikou, E., Karamanou, S., Economou, A., 2007. Bacterial protein secretion through the translocase nanomachine. Nat. Rev. Microbiol. 5 (11), 839. http://dx.doi.org/10.1038/nrmicro1771.
Patron, N.J., Waller, R.F., 2007. Transit peptide diversity and divergence: a global analysis of plastid targeting signals. Bioessays 29 (10), 1048–1058. http://dx.doi.org/10.1002/bies.20638.
Pechsrichuang, P., Songsiriritthigul, C., Haltrich, D., et al., 2016. OmpA signal peptide leads to heterogenous secretion of *B. subtilis* chitosanase enzyme from *E. coli* expression system. Springerplus 5 (1), 1200. http://dx.doi.org/10.1186/s40064-016-2893-y.
Perez-Trujillo, J.J., Garza-Morales, R., Barron-Cantu, J.A., et al., 2017. DNA vaccine encoding human papillomavirus antigens flanked by a signal peptide and a KDEL sequence induces a potent therapeutic antitumor effect. Oncol. Lett. 13 (3), 1569–1574. http://dx.doi.org/10.3892/ol.2017.5635.
Péterfy, M., Gyuris, T., Takács, L., 2000. Signal-exon trap: a novel method for the identification of signal sequences from genomic DNA. Nucleic Acids Res. 28 (7). http://dx.doi.org/10.1093/nar/28.7.e26. e26-e.
Pfeiffer, T., Pisch, T., Devitt, G., et al., 2006. Effects of signal peptide exchange on HIV-1 glycoprotein expression and viral infectivity in mammalian cells. FEBS Lett. 580 (15), 3775–3778. http://dx.doi.org/10.1016/j.febslet.2006.05.070.
Plath, K., Mothes, W., Wilkinson, B.M., et al., 1998. Signal sequence recognition in posttranslational protein transport across the yeast ER membrane. Cell 94 (6),

795–807. http://dx.doi.org/10.1016/S0092-8674(00)81738-9.
Pohlschröder, M., Prinz, W.A., Hartmann, E., et al., 1997. Protein translocation in the three domains of life: variations on a theme. Cell 91 (5), 563–566. http://dx.doi.org/10.1016/S0092-8674(00)80443-2.
Puziss, J.W., Harvey, R., Bassford, P., 1992. Alterations in the hydrophilic segment of the maltose-binding protein (MBP) signal peptide that affect either export or translation of MBP. J. Bacteriol. 174 (20), 6488–6497. http://dx.doi.org/10.1128/jb.174.20.6488-6497.1992.
Quentin, Y., Fichant, G., Denizot, F., 1999. Inventory, assembly and analysis of *Bacillus subtilis* ABC transport systems. J. Mol. Biol. 287 (3), 467–484. http://dx.doi.org/10.1006/jmbi.1999.2624.
Ramanan, R.N., Tik, W.B., Memari, H.R., et al., 2010. Effect of promoter strength and signal sequence on the periplasmic expression of human interferon-12b in *Escherichia coli*. Afr. J. Biotechnol. 9 (3).
Rapoport, T.A., Jungnickel, B., Kutay, U., 1996. Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes. Annu. Rev. Biochem. 65 (1), 271–303. http://dx.doi.org/10.1146/annurev.bi.65.070196.001415.
Rigel, N.W., Braunstein, M., 2008. A new twist on an old pathway–accessory Sec systems. Mol. Microbiol. 70 (1), 271. http://dx.doi.org/10.1111/j.1365-2958.2008.06433.x.
Robinson, C., Klösgen, R., Herrmann, R., et al., 1993. Protein translocation across the thylakoid membrane-a tale of two mechanisms (FEBS 125020). FEBS Lett. 325 (1-2), 67–69. http://dx.doi.org/10.1016/0014-5793(93)81415-V.
Román, R., Miret, J., Scalia, F., et al., 2016. Enhancing heterologous protein expression and secretion in HEK293 cells by means of combination of CMV promoter and IFNα2 signal peptide. J. Biotechnol. 239, 57–60. http://dx.doi.org/10.1016/j.jbiotec.2016.10.005.
Rusch, S.L., Kendall, D.A., 2007. Interactions that drive Sec-dependent bacterial protein transport. Biochemistry 46 (34), 9665–9673. http://dx.doi.org/10.1021/bi7010064.
Rusch, S.L., Chen, H., Izard, J.W., et al., 1994. Signal peptide hydrophobicity is finely tailored for function. J. Cell Biochem. 55 (2), 209–217. http://dx.doi.org/10.1002/jcb.240550208.
Rusch, S.L., Mascolo, C.L., Kebir, M.O., et al., 2002. Juxtaposition of signal-peptide charge and core region hydrophobicity is critical for functional signal peptides. Arch. Microbiol. 178 (4), 306–310. http://dx.doi.org/10.1007/s00203-002-0453-z.
Rutkowski, D.T., Ott, C.M., Polansky, J.R., et al., 2003. Signal sequences initiate the pathway of maturation in the endoplasmic reticulum lumen. J. Biol. Chem. 278 (32), 30365–30372. http://dx.doi.org/10.1074/jbc.M302117200.
Samuelson, J.C., Chen, M., Jiang, F., et al., 2000. YidC mediates membrane protein insertion in bacteria. Nature 406 (6796), 637. http://dx.doi.org/10.1038/35020586.
Schallenberger, M.A., Niessen, S., Shao, C., et al., 2012. Type I signal peptidase and protein secretion in *Staphylococcus aureus*. J. Bacteriol. 194 (10), 2677–2686. http://dx.doi.org/10.1128/JB.00064-12.
Schartl, M., Wilde, B., Hornung, U., 1998. Triplet repeat variability in the signal peptide sequence of the *Xmrk* receptor tyrosine kinase gene in *Xiphophorus* fish. Gene 224 (1), 17–21. http://dx.doi.org/10.1016/S0378-1119(98)00520-4.
Schneewind, O., Model, P., Fischetti, V.A., 1992. Sorting of protein A to the staphylococcal cell wall. Cell 70 (2), 267–281. http://dx.doi.org/10.1016/0092-8674(92)90101-H.
Schrempf, S., Froeschke, M., Giroglou, T., et al., 2007. Signal peptide requirements for lymphocytic choriomeningitis virus glycoprotein C maturation and virus infectivity. J. Virol. 81 (22), 12515–12524. http://dx.doi.org/10.1128/JVI.01481-07.
Schwartz, K., Sekedat, M.D., Syed, A.K., et al., 2014. The AgrD N-terminal leader peptide of *Staphylococcus aureus* has cytolytic and amyloidogenic properties. Infect. Immun. 82 (9), 3837–3844. http://dx.doi.org/10.1128/IAI.02111-14.
Scotti, P.A., Urbanus, M.L., Brunner, J., et al., 2000. YidC, the *Escherichia coli* homologue of mitochondrial Oxa1p, is a component of the Sec translocase. EMBO J. 19 (4), 542–549. http://dx.doi.org/10.1093/emboj/19.4.542.
Shen, L.M., Lee, J.I., Cheng, S., et al., 1991. Use of site-directed mutagenesis to define the limits of sequence variation tolerated for processing of the M13 procoat protein by the *Escherichia coli* leader peptidase. Biochemistry 30 (51), 11775–11781. http://dx.doi.org/10.1021/bi00115a006.
S-i, Matsuyama, Fujita, Y., Mizushima, S., 1993. SecD is involved in the release of translocated secretory proteins from the cytoplasmic membrane of *Escherichia coli*. EMBO J. 12 (1), 265.
Siboo, I.R., Chaffin, D.O., Rubens, C.E., et al., 2008. Characterization of the accessory Sec system of *Staphylococcus aureus*. J. Bacteriol. 190 (18), 6188–6196. http://dx.doi.org/10.1128/JB.00300-08.
Singh, P., Sharma, L., Kulothungan, S.R., et al., 2013. Effect of signal peptide on stability and folding of *Escherichia coli* thioredoxin. PloS One 8 (5), e63442. http://dx.doi.org/10.1371/journal.pone.0063442.
Sletta, H., Tøndervik, A., Hakvåg, S., et al., 2007. The presence of N-terminal secretion signal sequences leads to strong stimulation of the total expression levels of three tested medically important proteins during high-cell-density cultivations of *Escherichia coli*. Appl. Environ. Microbiol. 73 (3), 906–912. http://dx.doi.org/10.1128/AEM.01804-06.
Sommer, M.S., Daum, B., Gross, L.E., et al., 2011. Chloroplast Omp85 proteins change orientation during evolution. Proc. Natl. Acad. Sci. 108 (33), 13841–13846. http://dx.doi.org/10.1073/pnas.1108626108.
Stanke, N., Stange, A., Lüftenegger, D., et al., 2005. Ubiquitination of the prototype foamy virus envelope glycoprotein leader peptide regulates subviral particle release. J. Virol. 79 (24), 15074–15083. http://dx.doi.org/10.1128/JVI.79.24.15074-15083.2005.
Strom, M.S., Lory, S., 1993. Structure-function and biogenesis of the type IV pili. Ann. Rev. Microb. 47 (1), 565–596. http://dx.doi.org/10.1146/annurev.mi.47.100193.003025.
Suominen, I., Meyer, P., Tilgmann, C., et al., 1995. Effects of signal peptide mutations on processing of *Bacillus stearothermophilus* α-amylase in *Escherichia coli*. Microbiology 141 (3), 649–654. http://dx.doi.org/10.1099/13500872-141-3-649.
Szabady, R.L., Peterson, J.H., Skillman, K.M., et al., 2005. An unusual signal peptide facilitates late steps in the biogenesis of a bacterial autotransporter. Proc. Natl. Acad. Sci. U. S. A. 102 (1), 221–226. http://dx.doi.org/10.1073/pnas.0406055102.
Teng, C., Jia, H., Yan, Q., et al., 2011. High-level expression of extracellular secretion of a β-xylosidase gene from Paecilomycesthermophila in *Escherichia coli*. Bioresour. Technol. 102 (2), 1822–1830. http://dx.doi.org/10.1016/j.biortech.2010.09.055.
Tian, P., Bernstein, H.D., 2009. Identification of a post-targeting step required for efficient cotranslational translocation of proteins across the *Escherichia coli* inner membrane. J. Biol. Chem. 284 (17), 11396–11404. http://dx.doi.org/10.1074/jbc.M900375200.
Ting, Y.T., Harris, P.W., Batot, G., et al., 2016. Peptide binding to a bacterial signal peptidase visualized by peptide tethering and carrier-driven crystallization. IUCrJ 3 (1), 10–19. http://dx.doi.org/10.1107/S2052252515019971.
Tjalsma, H., Bolhuis, A., Jongbloed, J.D., et al., 2000. Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. Microbiol. Mol. Biol. Rev. 64 (3), 515–547. http://dx.doi.org/10.1128/MMBR.64.3.515-547.2000.
Tjalsma, H., Antelmann, H., Jongbloed, J.D., et al., 2004. Proteomics of protein secretion by *Bacillus subtilis*: separating the "secrets" of the secretome. Microbiol. Mol. Biol. Rev. 68 (2), 207–233. http://dx.doi.org/10.1128/MMBR.68.2.207-233.2004.
Tristan, A., Benito, Y., Montserret, R., et al., 2009. The signal peptide of *Staphylococcus aureus* panton valentine leukocidin LukS component mediates increased adhesion to heparan sulfates. PLoS One 4 (4), e5042. http://dx.doi.org/10.1371/journal.pone.0005042.
Ulbrecht, M., Martinozzi, S., Grzeschik, M., et al., 2000. Cutting edge: the human cytomegalovirus UL40 gene product contains a ligand for HLA-E and prevents NK cell-mediated lysis. J. Immunol. 164 (10), 5019–5022. http://dx.doi.org/10.4049/jimmunol.164.10.5019.
Ulfig, A., Fröbel, J., Lausberg, F., et al., 2017. The h-region of twin arginine signal peptides supports productive binding of bacterial Tat precursor proteins to the TatBC receptor complex. J. Biol. Chem. http://dx.doi.org/10.1074/jbc.M117.788950. jbc.M117.788950.
van Dijl, J.M., Braun, P.G., Robinson, C., et al., 2002. Functional genomic analysis of the *Bacillus subtilis* Tat pathway for protein secretion. J. Biotechnol. 98 (2), 243–254. http://dx.doi.org/10.1016/S0168-1656(02)00135-9.
van Kraaij, C., de Vos, W.M., Siezen, R.J., et al., 1999. Lantibiotics: biosynthesis, mode of action and applications. Nat. Prod. Rep. 16 (5), 575–587. http://dx.doi.org/10.1039/A804531C.
van Roosmalen, M.L., Geukens, N., Jongbloed, J.D., et al., 2004. Type I signal peptidases of Gram-positive bacteria. Biochim. Biophys. Acta 1694 (1), 279–297. http://dx.doi.org/10.1016/j.bbamcr.2004.05.006.
Van Voorst, F., De Kruijff, B., 2000. Role of lipids in the translocation of proteins across membranes. Biochem. J. 347 (3), 601–612. http://dx.doi.org/10.1042/bj3470601.
van Wely, K.H., Swaving, J., Freudl, R., et al., 2001. Translocation of proteins across the cell envelope of Gram-positive bacteria. FEMS Microbiol. Rev. 25 (4), 437–454. http://dx.doi.org/10.1111/j.1574-6976.2001.tb00586.x.
Vermeire, K., Bell, T.W., Van Puyenbroeck, V., et al., 2014. Signal peptide-binding drug as a selective inhibitor of co-translational protein translocation. PLoS Biol. 12 (12), e1002011. http://dx.doi.org/10.1371/journal.pbio.1002011.
Vezzoli, V., Duminuco, P., Vottero, A., et al., 2015. A new variant in signal peptide of the human luteinizing hormone receptor (LHCGR) affects receptor biogenesis causing leydig cell hypoplasia. Hum. Mol. Gen. 24 (21), 6003–6012. http://dx.doi.org/10.1093/hmg/ddv313.
von Heijne, G., 1983. Patterns of amino acids near signal-sequence cleavage sites. FEBS J. 133 (1), 17–21. http://dx.doi.org/10.1111/j.1432-1033.1983.tb07424.x.
von Heijne, G., 1984. Analysis of the distribution of charged residues in the N-terminal region of signal sequences: implications for protein export in prokaryotic and eukaryotic cells. EMBO J. 3 (10), 2315.
von Heijne, G., Abrahmsèn, L., 1989. Species-specific variation in signal peptide design implications for protein secretion in foreign hosts. FEBS Lett. 244 (2), 439–446. http://dx.doi.org/10.1016/0014-5793(89)80579-4.
Von Heijne, G., 1985. Signal sequences: the limits of variation. J. Mol. Biol. 184 (1), 99–105. http://dx.doi.org/10.1016/0022-2836(85)90046-4.
Von Heijne, G., 1994. Design of Protein Targeting Signals and Membrane Protein Engineering. Walter de Gruyter, Berlin, New York, pp. 263–279.
Von Heijne, G., Steppuhn, J., Herrmann, R.G., 1989. Domain structure of mitochondrial and chloroplast targeting peptides. FEBS J. 180 (3), 535–545. http://dx.doi.org/10.1111/j.1432-1033.1989.tb14679.x.
Wang, C.-Z., Chi, C.-W., 2004. Conus peptides—a rich pharmaceutical treasure. Acta Biochim. Biophys. Sin. 36 (11), 713–723. http://dx.doi.org/10.1093/abbs/36.11.713.
Wang, G., Xia, Y., Gu, Z., et al., 2015. A new potential secretion pathway for recombinant proteins in Bacillus subtilis. Microb. Cell Fact. 14 (1), 179. http://dx.doi.org/10.1186/s12934-015-0374-6.
Watanabe, K., Tsuchida, Y., Okibe, N., et al., 2009. Scanning the *Corynebacterium glutamicum* R genome for high-efficiency secretion signal sequences. Microbiology 155 (3), 741–750. http://dx.doi.org/10.1099/mic.0.024075-0.
Weltman, J.K., Skowron, G., Loriot, G.B., 2007. Influenza A H5N1 hemagglutinin cleavable signal sequence substitutions. Biochem. Biophys. Res. Commun. 352 (1), 177–180. http://dx.doi.org/10.1016/j.bbrc.2006.10.184.
Wilk, T., Geiselhart, V., Frech, M., et al., 2001. Specific interaction of a novel foamy virus Env leader protein with the N-terminal Gag domain. J. Virol. 75 (17), 7995–8007. http://dx.doi.org/10.1128/JVI.75.17.7995-8007.2001.
Yamamoto, Y., Taniyama, Y., Kikuchi, M., 1989. Important role of the proline residue in the signal sequence that directs the secretion of human lysozyme in *Saccharomyces cerevisiae*. Biochemistry 28 (6), 2728–2732. http://dx.doi.org/10.1021/

bi00432a054.
Yao, J.S., Strauss, E.G., Strauss, J.H., 1996. Interactions between PE2, E1, and 6K required for assembly of alphaviruses studied with chimeric viruses. J. Virol. 70 (11), 7910–7920.
Yi, L., Dalbey, R.E., 2005. Oxal/Alb3/YidC system for insertion of membrane proteins in mitochondria, chloroplasts and bacteria. Mol. Membr. Biol. 22 (1-2), 101–111. http://dx.doi.org/10.1080/09687860500041718.
York, J., Nunberg, J.H., 2006. Role of the stable signal peptide of Junin arenavirus envelope glycoprotein in pH-dependent membrane fusion. J. Virol. 80 (15), 7775–7780. http://dx.doi.org/10.1128/JVI.00642-06.
Yuan, J., Zweers, J.C., Van Dijl, J.M., et al., 2010. Protein transport across and into cell membranes in bacteria and archaea. Cell. Mol. Life Sci. 67 (2), 179–199. http://dx.doi.org/10.1007/s00018-009-0160-x.
Zalucki, Y.M., Jennings, M.P., 2017. Signal peptidase I processed secretory signal sequences: selection for and against specific amino acids at the second position of mature protein. Biochem. Biophys. Res. Commun. 483 (3), 972–977. http://dx.doi.org/10.1016/j.bbrc.2017.01.044.
Zalucki, Y.M., Power, P.M., Jennings, M.P., 2007. Selection for efficient translation initiation biases codon usage at second amino acid position in secretory proteins. Nucleic Acids Res. 35 (17), 5748–5754. http://dx.doi.org/10.1093/nar/gkm577.
Zalucki, Y.M., Beacham, I.R., Jennings, M.P., 2009. Biased codon usage in signal peptides: a role in protein export. Trends Microbiol. 17 (4), 146–150. http://dx.doi.org/10.1016/j.tim.2009.01.005.
Zamani, M., Nezafat, N., Negahdaripour, M., et al., 2015. In silico evaluation of different signal peptides for the secretory production of human growth hormone in *E. coli*. Int. J. Pept. Res. Ther. 21 (3), 261–268. http://dx.doi.org/10.1007/s10989-015-9454-z.
Zhang, L., Leng, Q., Mixson, A.J., 2005. Alteration in the IL-2 signal peptide affects secretion of proteins in vitro and in vivo. J. Gene Med. 7 (3), 354–365. http://dx.doi.org/10.1002/jgm.677.
Zhang, W., Yang, M., Yang, Y., et al., 2016. Optimal secretion of alkali-tolerant xylanase in *Bacillus subtilis* by signal peptide screening. Appl. Microbiol. Biotechnol. 100 (20), 8745–8756. http://dx.doi.org/10.1007/s00253-016-7615-4.
Zheng, N., Gierasch, L.M., 1996. Signal sequences: the same yet different. Cell 86 (6), 849–852. http://dx.doi.org/10.1016/S0092-8674(00)80159-2.
Zhou, Y., Liu, P., Gan, Y., et al., 2016. Enhancing full-length antibody production by signal peptide engineering. Microb. Cell. Fact. 15 (1), 47. http://dx.doi.org/10.1186/s12934-016-0445-3.