Review TRENDS in Biochemical Sciences Vol.xxx No.x



# The surprising complexity of signal sequences

**Ramanujan S. Hegde[1] and Harris D. Bernstein[2]**

[1] Cell Biology and Metabolism Branch, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA
[2] Genetics and Biochemistry Branch, National Institutes of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA

**Most secreted and many membrane proteins contain cleavable N-terminal signal sequences that mediate their targeting to and translocation across the endoplasmic reticulum or bacterial cytoplasmic membrane. Recent studies have identified many exceptions to the widely held view that signal sequences are simple, degenerate and interchangeable. Growing evidence indicates that signal sequences contain information that specifies the choice of targeting pathway, the efficiency of translocation, the timing of cleavage and even postcleavage functions. As a consequence, signal sequences can have important roles in modulating protein biogenesis. Based on a synthesis of studies in numerous experimental systems, we propose that substrate-specific sequence elements embedded in a conserved domain structure impart unique and physiologically important functionality to signal sequences.**

## Introduction

The discovery that secreted proteins are encoded with short, removable N-terminal 'signal sequences' that earmark them for export [1] was one of the great breakthroughs in cell biology in the 1970s. Much work in the subsequent 30 years has shown that signal sequences are recognized by dedicated factors that catalyze the transport of secretory precursors (preproteins) across the endoplasmic reticulum (ER) or bacterial cytoplasmic membrane. The key molecules in this pathway are found in all living cells and include a cytosolic targeting factor called the signal recognition particle (SRP), the membrane-bound receptor for SRP, an integral membrane protein conducting channel called the Sec61p complex (in eukaryotes) or the SecY complex (in prokaryotes), and a membrane-bound peptidase that removes signal sequences from preproteins on the lumenal (or periplasmic) face of the membrane.

The signal hypothesis originally predicted that all signal sequences would share a distinctive sequence motif; however, the absence of any 'consensus' quickly became apparent when several preproteins were sequenced. Nevertheless, the concept of 'signal sequence equivalence' prevailed. An early comparative sequence analysis [2] showed that signal sequences have a typical size (~20–30 residues) and a recognizable three-domain structure (a basic 'N domain', a ~7–13 residue hydrophobic 'H domain' and a slightly polar 'C domain'), but otherwise lack any significant homology. By the early 1980s, it had become clear that signal sequences are often readily interchangeable, tolerant of a wide range of mutations [3] and even capable of directing secretion in evolutionarily distant organisms [4,5]. A remarkable study published in 1987 showed that ~20% of random sequences can promote the secretion of invertase in yeast [6], strengthening the conclusion that the only essential feature of signal sequences is a hydrophobic core that is uninterrupted by charged residues. These studies implied that the signal recognition machinery has a high degree of tolerance, but they also created the impression that signal sequences have only a very circumscribed role in protein biogenesis. Thus, it has become common 'wisdom' that signal sequences are simple, interchangeable domains with a low information content.

Although the discovery of species-specific features of signal sequences [7] and a distinct motif in bacterial lipoprotein signal sequences [8] suggested >20 years ago that not all signal sequences are equivalent, hints that signal sequence diversity reflects an underlying functional complexity have emerged only in more recent studies. These studies, often from seemingly unrelated fields and disparate experimental systems, share one important feature: they have each examined the biosynthesis of a preprotein or membrane protein that differs from the 'model' substrates that were originally used to define the basic principles of signal sequence function. Individually each study might seem to describe an 'exception' to the well-established paradigm, but viewed together they suggest that differences among signal sequences could ultimately prove to be as physiologically important as their similarities.

Recent experiments have demonstrated that sequence variation among signal sequences can affect protein targeting and translocation, signal sequence cleavage and even postcleavage events. Perhaps more importantly, these studies have shown that the modulation of specific steps in the recognition and processing of signal sequences can have an essential role in protein biogenesis. Here, we review evidence showing that signal sequences encode far more information than is commonly thought, and we synthesize these studies into a cohesive framework to explain signal sequence diversity.

*Corresponding authors:* Hegde, R.S. (hegder@mail.nih.gov); Bernstein, H.D. (harris_bernstein@nih.gov).
Available online xxxxxx.



Please cite this article as: Ramanujan S. Hegde, Harris D. Bernstein, The surprising complexity of signal sequences, TRENDS in Biochemical Sciences (2006), doi:10.1016/j.tibs.2006.08.004.

## Signal sequences influence the selection of protein targeting pathways

Polypeptides with a signal sequence are first recognized in the cytosol. Early work on mammalian cells identified SRP as the only factor both necessary and sufficient for signal sequence recognition [9]. SRP is a ribonucleoprotein complex that binds to signal sequences as they emerge from the ribosome, slows translation, and targets the ribosome–nascent chain complex to the ER via SRP receptor. By ensuring that preproteins reach the ER at an early stage of translation, SRP is thought to prevent substantial protein folding that would preclude translocation through the Sec61p complex.

Contemporaneous studies on bacteria and yeast identified a different solution to the problem of protein targeting. In these organisms, the first model preproteins that were examined were shown to be fully synthesized, released from ribosomes, and then targeted to the Sec61p or SecY complex posttranslationally [10,11]. Posttranslational targeting requires that preproteins be held in a largely unfolded conformation in the cytosol by chaperones. The discovery that the cotranslational SRP pathway is highly conserved from bacteria to humans [12,13], however, led to the realization that microorganisms use two parallel targeting pathways.

Subsequent studies have shown that targeting pathway selection in yeast and bacteria is dictated largely by features of the signal sequence that influence recognition by SRP (Figure 1). Signal sequence binding is mediated primarily by a deep, flexible hydrophobic groove in the 54-kDa subunit of SRP [14]. The flexibility of residues lining this groove presumably facilitates interaction with a wide range of unrelated hydrophobic sequences. Both X-ray crystallography and *in vivo* secretion studies have suggested that an electrostatic interaction between the phosphate backbone of SRP RNA and basic amino acids in the N domain of signal sequences might also contribute to binding by SRP [15,16]. Differences in the regions of SRP that mediate signal sequence recognition presumably account for the marked variation in substrate specificity that has been observed in divergent organisms. Unlike mammalian SRPs, which recognize all signal sequences, yeast and bacterial SRPs bind predominantly to very hydrophobic signal sequences [17–19]. Proteins whose signal sequences fall below a threshold hydrophobicity are bypassed by SRP and are targeted posttranslationally by default. In *Escherichia coli*, posttranslational translocation is promoted by a ribosome-associated chaperone called trigger factor, which binds to the mature region of preproteins that are bypassed by SRP and is likely to prevent translating ribosomes from docking onto the SecY complex [20,21]. Although completely unrelated to trigger factor by sequence, the nascent-chain-associated complex might have an analogous function in yeast [22].

Why might multiple targeting pathways have evolved? We can envisage two plausible explanations. First, posttranslational translocation might be favored in rapidly growing organisms such as yeast and bacteria because it is compatible with a higher rate of secretion than cotranslational translocation. The higher rate of secretion can be achieved because protein synthesis is almost certainly much slower than protein translocation. By uncoupling these two processes, cells can make maximal use of a limited number of Sec complexes (which are an order of magnitude less abundant than ribosomes). Thus, in yeast and bacteria, cotranslational targeting by the SRP pathway might be used only by proteins that rapidly lose translocation competence in the cytoplasm. Consistent with this hypothesis, a recent study has suggested that only a few secreted proteins are targeted by SRP in *E. coli* [23]. By contrast, most multispanning membrane proteins in *E. coli* are targeted by SRP (which recognizes a transmembrane domain as an internal signal sequence), presumably because these highly hydrophobic proteins would otherwise aggregate in the cytosol. Second, multiple targeting pathways might have evolved to facilitate the delivery of substrates to protein translocation complexes ('translocons') of different composition and functionality (Box 1). Thus, in addition to determining the mode of translocation (cotranslational versus posttranslational), signal sequences might also influence the site of translocation.

Although targeting pathway selection based on signal sequence discrimination by SRP has been observed only in microorganisms, a conceptually related process has recently emerged in mammalian cells. Myristoylation of the N domain of the signal sequence of the NADH cytochrome $b_5$ reductase inhibits its recognition by SRP [24]. This competition between SRP binding and myristoylation targets the nascent polypeptide to either the ER membrane or the mitochondrial outer membrane. Thus, the cell can direct a single enzyme to two distinct compartments by regulating the efficiency of the interactions between SRP and the signal sequence (Figure 1). Other cotranslational modifications, such as phosphorylation of the nascent chain, could similarly influence the interaction of SRP with the signal sequence to regulate protein targeting [25].

## Signal sequence variation affects interactions with the translocon

Once a substrate is targeted to the ER or cytoplasmic membrane, the signal sequence must interact with and engage a translocon. Even though the composition of translocons can differ widely among different organisms and even within a single cell (Box 1), most signal sequence recognition is thought to involve interaction with the Sec61p or SecY complex [26,27]. Because this gating event is a prerequisite for initiating protein translocation, one might imagine that all substrates interact with the core components of the translocon in essentially the same manner. However, several studies in the mammalian system have begun to question this assumption and have suggested that differences in the interaction of diverse signal sequences with the translocon can substantially affect protein biogenesis.

Perhaps the clearest evidence that signal sequences interact differentially with the translocon has emerged from crosslinking and chemical inhibitor studies. The first systematic effort to map the signal sequence-binding site in the Sec61p complex analyzed the interaction between a translocation-arrested substrate (yeast α-factor) and the yeast posttranslational translocon. By selectively



Please cite this article as: Ramanujan S. Hegde, Harris D. Bernstein, The surprising complexity of signal sequences, TRENDS in Biochemical Sciences (2006), doi:10.1016/j.tibs.2006.08.004.

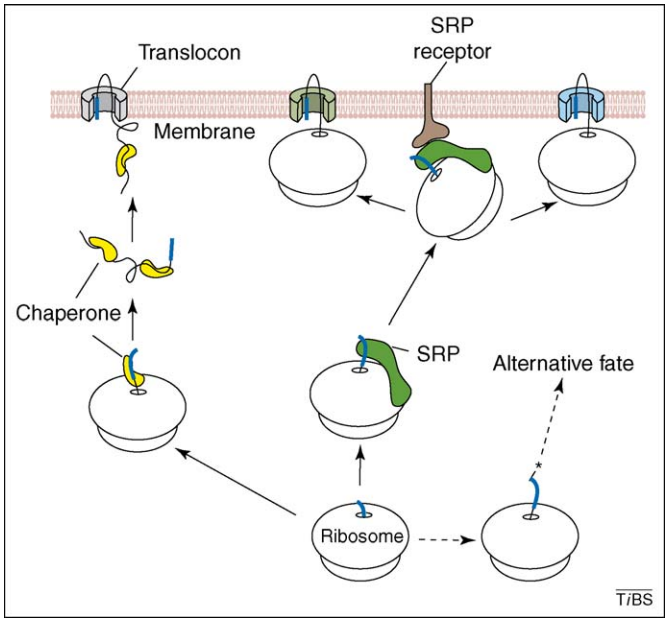


**Figure 1**. Targeting pathway selection is specified by signal sequences. As a polypeptide (black ribbon) containing a signal sequence (blue) emerges from the ribosome, it can interact with SRP (green); chaperones (yellow), including trigger factor in bacteria or nascent-chain-associated complex in eukaryotes; or modifying enzymes (not shown). The choice of these factors is dictated largely by the properties of the signal sequence. Recognition by SRP is usually the decisive event: proteins whose signal sequences are not recognized by SRP are maintained in a loosely folded conformation by chaperones and targeted to the ER or cytoplasmic membrane posttranslationally or are directed to other intracellular destinations (in cases where SRP recognition is blocked by modification of the signal sequence). In principle, ineffective SRP binding or competition between SRP binding and signal sequence modification can result in a polypeptide being routed into more than one biosynthetic pathway. After proteins arrive at the ER or cytoplasmic membrane, signal sequences can also influence their delivery to various translocons (shown in different colors; see Box 1). This illustration represents a composite view derived from several experimental systems; the intention is to convey the concept of targeting pathway diversity rather than to depict the specific pathways in any one organism.

incorporating a modified lysine containing a photoactivatable crosslinking group into the signal sequence, the binding site was mapped to a position surrounded by transmembrane domains 2 and 7 of Sec61p, portions of Sec62p and Sec71p, and lipids [27]. Changing the amino acid used in the crosslinking analysis to a modified phenylalanine, however, noticeably changed the position of the signal sequence in relation to Sec61p [27]. These results imply that the signal sequence binding site is not fixed, but instead varies in a substrate-dependent manner. In addition, a small molecule ('cotransin') has been recently demonstrated to inhibit the interaction between mammalian Sec61p and some, but not other, signal sequences [28,29]. This striking observation provides further evidence that the translocon does not have a simple binding pocket that accommodates all substrates in a uniform manner. Indeed, the crystal structure of an archaeal SecY complex in its inactive state suggests that the putative site of signal sequence binding (between transmembrane domains 2 and 7) is particularly flexible and can accommodate substrates of differing structure and composition [30].

Recognition of some signal sequences by the translocon requires the contribution of accessory factors, which also suggests that differences among signal sequences influence their interaction with the Sec61p complex. In the mammalian system, many signal sequences that promote successful targeting to the Sec61p complex cannot initiate translocation without TRAM [31] and/or the TRAP complex [32], two membrane protein factors that are present at the site of translocation. Likewise, although signal sequences can clearly interact directly with Sec61p in yeast, peripheral components (such as Sec62p) seem to be required for (but are not necessarily directly involved in) the interaction between Sec61p and some signal sequences [27,33,34]. The mechanism by which components such as TRAM, TRAP and Sec62p influence signal sequence interactions with the Sec61p channel, however, remains to be determined.

Additional evidence that signal sequences interact differentially with the translocon has been recently provided by experiments designed to measure the relative efficiencies with which signal sequences cotranslationally 'gate' the translocon to initiate translocation. By using an assay in which the topology of a membrane protein was determined by the timing of translocon gating relative to the emergence of a transmembrane segment during translation, the ability of a signal sequence to initiate

**Box 1. The diversity of translocon composition**

Both eukaryotic and prokaryotic cells seem to contain multiple translocons of differing (but often overlapping) subunit composition that simultaneously function in parallel pathways of translocation. In many cases, substrate specificity is dictated by the accessory factors and maturation factors that associate with a particular translocon.

In *Saccharomyces cerevisiae*, biochemical analyses have identified three distinct translocons. The first is a heptamer [68] comprising the trimeric Sec61p complex (containing Sec61p, Sbh1p and Sss1p) in association with a tetramer (containing Sec62p, Sec63p, Sec71p and Sec72p). This translocon functions in post-translational transport and is not associated with membrane-bound ribosomes or the SRP receptor. The second translocon, which contains only the Sec61p complex, has been implicated in cotranslational translocation by virtue of its tight association with the ribosome and its interaction with the SRP receptor [68]. The third translocon, consisting of a trimeric complex of Ssh1p (a Sec61p homolog), Sbh2p (a homolog of Sbh1p) and Sss1p, has been implicated in the cotranslational translocation of a distinct set of proteins [69,70]. Each of these translocons presumably associates with signal peptidase and oligosaccharyl transferase, but the subunit composition of oligosaccharyl transferase seem to vary slightly depending on the translocon with which it is associated [71].

In the mammalian system, signal peptidase, two distinct oligosaccharyl transferase isoforms, TRAM, the TRAP complex, Sec62, Sec63 and other proteins seem to interact with the Sec61p complex [31,32,72–74]. It is unlikely that these components are all part of a single enormous translocon: instead, they might represent the building blocks of many distinct translocons that operate in parallel with different functional roles. In *Escherichia coli*, the SecY complex (containing SecY, SecE and SecG) coexists in the membrane with another heterotrimer (containing SecD, SecF and YajC) [75]. Because the SecDFYajC complex is much less abundant than the SecY complex, different subsets of these components might be assembled into functionally distinct translocons. Intriguingly, a protein called YidC, which is involved only in membrane protein insertion, seems to function in conjunction with the SecY complex (possibly through an association with SecDFYajC) for some substrates, but in isolation for others [76,77]. In addition, some Gram-positive bacteria encode two distinct SecY homologs. In *Streptococcus gordonii*, one SecY homolog is used specifically to export a large glycoprotein called GspB. Interestingly, GspB is targeted to the accessory SecY channel by an unusually long signal sequence of 90 amino acids [78].



Please cite this article as: Ramanujan S. Hegde, Harris D. Bernstein, The surprising complexity of signal sequences, TRENDS in Biochemical Sciences (2006), doi:10.1016/j.tibs.2006.08.004.

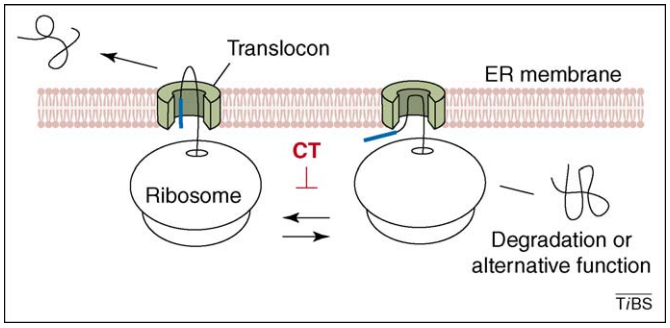


**Figure 2**. Signal sequences interact differentially with the translocon. Sequence diversity influences the efficiency with which signal sequences engage the Sec61p-containing translocon complex. Stable interaction of a signal sequence with the translocon causes a preprotein to adopt a 'looped' orientation (left) and results in gating of the translocon, access of the nascent chain to the noncytosolic side of the membrane, and initiation of translocation. A weaker interaction between the signal sequence and the translocon (right) can result in the nascent polypeptide slipping into the cytosol, sometimes even after the signal peptide has been removed [37–39]. As a result, a fraction of some proteins (e.g. hepatitis B virus precore and calreticulin) is retained in the cytoplasm or is degraded instead of being translocated into the ER. The observation that small molecules such as cotransin (CT) inhibit the initial interaction between the signal sequence and the translocon for some, but not other, signal sequences to block protein translocation selectively [28,29] confirms that signal sequences bind to the translocon in a heterogeneous manner.

translocation within a defined time frame was quantified [35]. This analysis showed that signal sequences have a surprisingly wide range of gating efficiencies that, in many instances, are evolutionarily conserved. The results raise the possibility that a variable fraction (5–20%) of many proteins with signal sequences might not be translocated across the ER membrane but instead might remain in the cytosol (Figure 2). Although this cytosolic population is sometimes relatively minor and often short-lived, its physiological importance has been demonstrated in experiments in which translocation inefficiency has been eliminated by modifying the signal sequence.

One protein whose signal sequence has been designed to generate translocated and nontranslocated populations, both of which are beneficial to the cell, is the abundant ER chaperone calreticulin. Numerous studies over the past ten years have implicated calreticulin in various activities in the cytosol and nucleus [36], although the mechanism by which an ER protein might access these compartments has remained unclear. Recently, the dual localization of calreticulin has been explained by the observation that its signal sequence is slightly inefficient, resulting in a small but detectable cytosolic population (~5%) both *in vitro* and *in vivo* [37]. Eliminating the cytosolic population of calreticulin by using a more efficient signal sequence (derived from preprolactin) inhibits at least one of its proposed cytosolic functions.

The precore protein of hepatitis B virus represents another example in which signal sequence inefficiency seems to have physiological relevance. Although the precore protein clearly has a signal sequence, its translocation is largely aborted at a step after targeting to the translocon (and, curiously, after signal sequence cleavage) [38,39]. Remarkably, both the translocated and nontranslocated forms of the protein might have crucial, but distinct, functional roles in the viral life cycle [40]. In view of the results obtained with calreticulin, it seems likely that the signal sequence of the precore protein has evolved to optimize viral replication by precisely controlling the levels of translocated and nontranslocated populations.

Signal sequence inefficiency might also have pathological consequences under some conditions. The prion protein (PrP), which is a cell-surface glycoprotein whose altered metabolism is implicated in causing neurodegeneration, provides one example of this phenomenon. After PrP is targeted to the ER, a small proportion of the protein (~5–15%) is not translocated completely and instead is made as either a transmembrane ($^{Ctm}$PrP) or a cytosolic (cyPrP) form. When overrepresented, both cyPrP [41] and $^{Ctm}$PrP [42] can be pathogenic. Biochemical studies suggest that the generation of both cyPrP and $^{Ctm}$PrP is caused by a slight but detectable inefficiency of the PrP signal sequence, and that the production of these forms is precluded by increasing the efficiency of the signal sequence [43,44]. By replacing the native signal sequence with a more efficient signal sequence, the survival of cultured cells that express PrP can be enhanced under some experimental conditions [44].

## Signal sequence variation influences events that occur after translocon gating

Two signal sequences that are equally efficient in their ability to initiate translocation can nonetheless differ in other, still poorly understood aspects of their interaction with either the translocon or the signal sequence cleavage machinery. This conclusion has emerged from various studies in which a specific signal sequence has been shown not to influence translocation *per se*, but rather to affect a cotranslational or posttranslational maturation event. The autotransporter EspP, a virulence factor expressed by *E. coli* O157:H7, provides a particularly notable example of a protein whose signal sequence that affects biosynthetic events after the onset of translocation [45].

Autotransporters are large outer membrane proteins that are translocated across the cytoplasmic membrane by means of the SecY complex. After crossing the periplasm, the C-terminal 'β domain' inserts into the outer membrane and promotes transport of a large N-terminal 'passenger domain' into the extracellular space by an unknown mechanism. Similar to ~10% of the known autotransporters, EspP contains a signal sequence with an unusually long N domain (~35 amino acids). Surprisingly, replacement of the native EspP signal sequence with a signal sequence that lacks the long N domain does not affect translocation of the protein across the cytoplasmic membrane, but does inhibit translocation of the passenger domain across the outer membrane [45]. This finding was rationalized by the observation that the unusual N domain facilitates a protracted interaction with the translocon that, in turn, prevents the passenger domain from adopting a transport-incompetent conformation in the periplasm (Figure 3a).

The composition of a signal sequence can similarly influence the behavior of a downstream transmembrane domain or usage of downstream glycosylation sites [46,47]. Conversely, the presence of downstream sequence elements, such as transmembrane domains and glycosylation sites, can influence the functionality of a signal sequence at the N terminus [37,43,48]. The mechanistic basis of these



Please cite this article as: Ramanujan S. Hegde, Harris D. Bernstein, The surprising complexity of signal sequences, TRENDS in Biochemical Sciences (2006), doi:10.1016/j.tibs.2006.08.004.

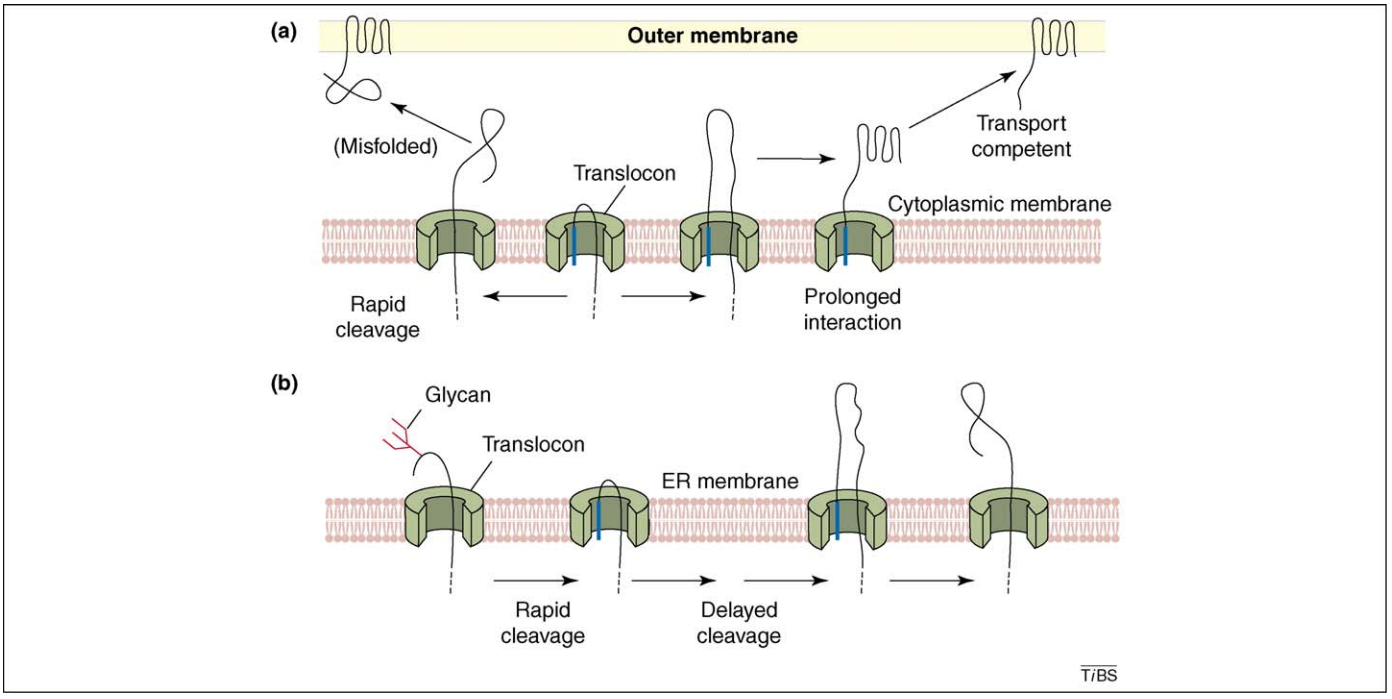


**Figure 3**. The interaction between the signal sequence and the translocon can influence protein biosynthesis. The duration of the interaction between the signal sequence and the translocon before the signal sequence is removed by a signal or leader peptidase can affect downstream events in protein biosynthesis such as folding, glycosylation site usage, association with chaperones, and transmembrane domain integration [45–47,50,79]. **(a)** Available evidence suggests that the signal sequence of the bacterial autotransporter EspP is cleaved slowly to prevent misfolding of the passenger domain into a transport-incompetent conformation in the periplasm [45]. **(b)** A hypothetical eukaryotic protein containing a potential glycosylation site near its N terminus. The glycosylation site is inaccessible to oligosaccharyl transferase (not shown) as long as the signal peptide remains uncleaved. If cleavage occurs rapidly (left), oligosaccharyl transferase can add a glycan to the relatively short, still unfolded, nascent chain. If cleavage occurs after a substantial portion of the nascent chain has been translocated (right), partial folding of the polypeptide masks the glycosylation site.

events is largely unknown, but it might be related to the fact that space within and immediately adjacent to the translocation channel is limited. For example, transmembrane domains and signal sequences are thought to interact with and laterally exit the Sec61p or SecY channel at the same site. Thus, the speed at which this site is vacated by a signal sequence could influence the behavior of an incoming transmembrane domain. Similarly, access to the nascent chain in the translocon by two very large enzymes (such as signal peptidase and oligosaccharyl transferase) is likely to be mutually exclusive for steric reasons [46,49]. Thus, the timing of signal sequence cleavage, which seems to vary considerably among substrates, might influence the use of glycosylation sites near the N terminus (Figure 3b).

Signal sequences seem to regulate the timing of cleavage as a means of controlling not only downstream protein folding and glycosylation events, but also the exit of proteins from the ER. For example, the exceptionally slow cleavage of the native HIV gp120 signal sequence (which requires an hour) causes prolonged retention of the protein in the ER [50]. Replacing the signal sequence markedly accelerates the appearance of gp120 on the cell surface. These observations suggest that the signal sequence prevents premature exposure of gp120 to adjacent cells or to the immune system. An unusually long signal sequence (48 amino acids) associated with interleukin 15 resembles a duplicated signal sequence and is cleaved in two steps [51]. Interestingly, the first half of the signal sequence is removed rapidly, whereas the second half is cleaved extremely slowly. It has been proposed that the slow rate of the second cleavage reaction provides an additional level of control over the secretion of a highly inflammatory cytokine, and that the reaction might be accelerated by an unidentified factor under some conditions. In bacteria, a conserved motif that is associated with many signal sequences in *Staphylococcus aureus* and *Streptococcus pneumoniae*, YSIRK(G/S), also slows the rate of signal sequence cleavage [52], although the physiological importance of this event is unknown. It is not clear in any of these examples whether the slow cleavage of the signal sequence is due to its slow release from the translocon or to its inefficient interaction with the signal peptidase.

Conserved sequence motifs in some bacterial signal sequences direct preproteins to alternative leader peptidases. For instance, a motif that straddles the cleavage site in most *E. coli* lipoprotein signal sequences, LA(G/A)|C, is recognized by an enzyme called signal peptidase II [53]. Signal sequences associated with the major subunit of type IV pili in pathogenic bacteria are thought to promote protein translocation via the SecY complex, but they contain a unique sequence motif, GFTLIE, near the N terminus that delivers the protein to a so-called 'prepilin peptidase' [54]. This enzyme cleaves the signal sequence at a unique location between the N and H domains. Considerable evidence indicates that cleavage of the signal sequence at this site is essential to promote the polymerization of individual pilin subunits into functional pili.

## Postcleavage roles of signal sequences

Given the tremendous functionality of peptides in general, it is intriguing to consider the possibility that signal peptides (or fragments thereof) could have important roles in



Please cite this article as: Ramanujan S. Hegde, Harris D. Bernstein, The surprising complexity of signal sequences, TRENDS in Biochemical Sciences (2006), doi:10.1016/j.tibs.2006.08.004.

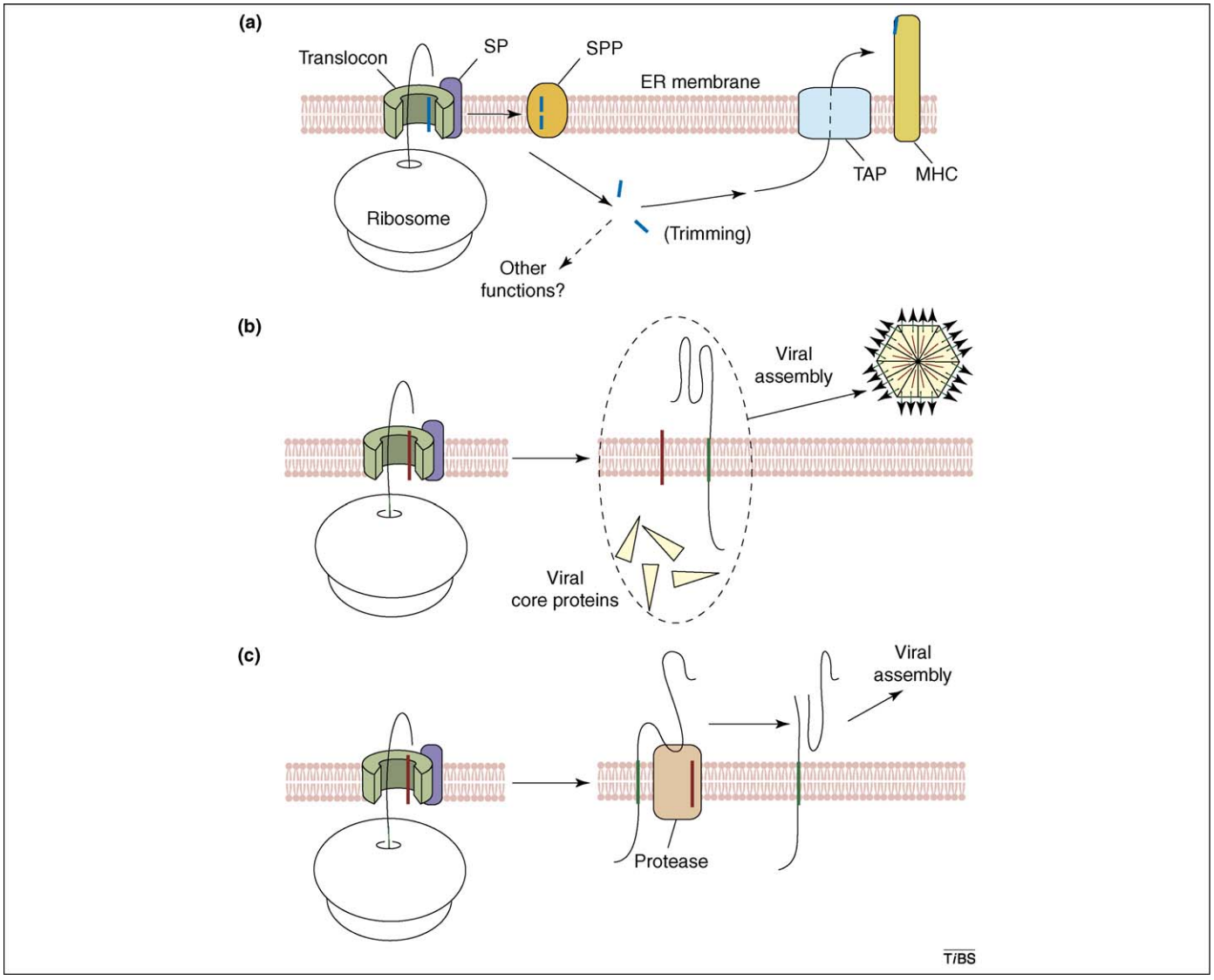


**Figure 4**. Postcleavage functions of signal peptides. **(a)** Fragments of some signal peptides are presented on the cell surface as antigens bound to MHC class I molecules [57,58]. In this pathway, signal peptides that have been removed from preproteins by signal peptidase (SP) are processed further by the intramembrane protease signal peptide peptidase (SPP) [55] and trimmed further in the cytosol. One or more signal peptide fragments are then translocated by the TAP transporter into the ER lumen, where they interact with MHC molecules. Some cleaved signal peptides might also have other roles in cell physiology. **(b,c)** The cleaved signal peptides of some viral glycoproteins mediate functions that are essential for the viral life cycle. For example, the cleaved signal peptides of foamy virus and lymphocytic choriomeningitis virus glycoproteins are incorporated into the viral capsid and might influence capsid assembly and viral infectivity [63,64] (b). In addition, the cleaved signal peptide of the Lassa virus glycoprotein acts as a cofactor in the proteolytic processing of the protein [65] (c).

the cell after they are removed from translocating polypeptides. Although it has long been assumed that cleaved signal peptides are degraded rapidly, the pathways and machinery involved in postcleavage processing are only now being explored. Studies following the fate of the cleaved preprolactin signal peptide have demonstrated further proteolytic processing in the membrane by a signal peptide peptidase and release of the signal peptide fragments into the cytosol [55], where they are degraded by unidentified proteases. Although this pathway clearly represents one means of metabolizing signal peptides, not all signal peptides are processed by signal peptide peptidase and not all products cleaved by signal peptide peptidase are necessarily degraded [56]. Thus, the fate of signal peptides and the fragments derived from them seems to be far more complex than was initially presumed.

Perhaps the first clear evidence that signal peptides have a postcleavage function came from the observation that some self-antigens presented on immune cells are derived from signal peptides (Figure 4a). In a well-studied example, HLA-E presents a nine-residue peptide on the cell surface that corresponds to a conserved sequence found in the signal sequences of several major histocompatibility complex (MHC) class I proteins [57,58]. The presence of peptide-loaded HLA-E on the surface is an indicator of active MHC class I expression and serves as an inhibitory signal to natural killer cells [58,59]. Downregulation of MHC class I expression, which commonly occurs during viral infection, results in both a decrease in HLA-E presentation and destruction by natural killer cells. Thus, generation of a cleaved signal peptide is used to report the intracellular status of a cell at the plasma membrane. Remarkably, human cytomegalovirus has managed to



Please cite this article as: Ramanujan S. Hegde, Harris D. Bernstein, The surprising complexity of signal sequences, TRENDS in Biochemical Sciences (2006), doi:10.1016/j.tibs.2006.08.004.

circumvent this elegant reporting mechanism by evolving a signal sequence on its glycoprotein that contains the same nine-residue sequence found in MHC class I signal sequences [60,61]. Synthesis of the glycoprotein enables the virus to downregulate MHC class I expression but nonetheless to elude detection by natural killer cells by presenting a virus-encoded substitute.

Beyond their role in antigen presentation, cleaved signal peptides have been recently shown or suggested to mediate a surprising array of biological functions. One of the two fragments of the preprolactin signal peptide has been shown to interact with the signaling molecule calmodulin, at least in a cell-free system [62]. The exceptionally long signal peptides (>50 amino acids) of foamy virus and lymphocytic choriomeningitis virus glycoproteins are actually incorporated into the virus particle after liberation by signal peptidase [63,64] (Figure 4b). Although the function of the signal peptide in the virus particle remains unclear, current data suggest that it could be important for maximizing infectivity. Perhaps even more intriguingly, the cleaved signal peptide of the Lassa virus glycoprotein (GP-C), which is also extremely long (58 amino acids), has been shown to act as an essential cofactor in the proteolytic maturation of GP-C [65] (Figure 4c). Replacing the native GP-C signal sequence with a conventional signal sequence does not affect GP-C translocation into the ER, but it abolishes cleavage of the glycoprotein into two fragments by a cellular enzyme. Remarkably, this defect can be rescued by providing the GP-C signal peptide *in trans* on a second preprotein. All of these observations indicate that signal sequences are not necessarily disposable elements with a function that ends at the time of their removal from the mature protein. Indeed, many more functional roles of cleaved signal peptides might remain to be discovered.

## A unifying framework for signal sequence diversity

The principal concept that has emerged here is that the highly degenerate sequence requirements for signal sequence recognition and processing by the core translocation machinery provides an incredibly elastic opportunity to embed additional information. This flexibility in primary sequence is likely to have been exploited during the course of evolution to maintain and to fine-tune any beneficial substrate-specific features of signal sequences that might have arisen by random chance [66]. In a few exceptional cases, unusually long N-domain extensions markedly modify signal sequence function or encode completely novel functions. But most substrate-specific functionalities are likely to be encoded in the relatively subtle differences between signal sequences. Based on the evidence that is currently available, we can envisage three qualitatively different purposes for which signal sequence degeneracy has been used.

First, the specific properties of a signal sequence can be precisely tailored to assure a particular biosynthetic outcome for the substrate to which it is attached. Thus, the use of a specific targeting pathway, the exact timing of signal sequence cleavage, or the length of the association of a signal sequence with the translocon might all be optimized to enable the substrate to fold in a specific manner or to acquire specific modifications (such as glycosylation) necessary for its proper biosynthesis, trafficking and function. Second, signal sequences can be designed to generate multiple functional products from a single mRNA. In some cases, the products might be simply alternatively localized forms of a given protein; in others, the cleaved signal peptide itself might have a second function distinct from its role in directing protein secretion. Third, signal sequences can be engineered to control protein quantity. Here, the properties of the signal sequence are used to determine the amount of the final mature protein and the timing of its production. For example, signal sequences that are targeted, translocated or cleaved inefficiently might effectively limit the amount of particularly potent secretory or membrane proteins that is available. It is intriguing to consider the possibility that this functional inefficiency might be relieved under some intracellular conditions; indeed, evidence already indicates that signal peptide efficiency can vary considerably in different cell types [67].

The challenge in the future will be not only to continue uncovering the diversity in signal sequence functionality, but also to determine whether and how this diversity has been exploited to impart greater control on protein biogenesis to suit constantly changing cellular needs.

## Acknowledgements

We thank Julie Donaldson, Erik Snapp, Cheryl Woolhead and Yihong Ye for critically reading the manuscript and providing helpful suggestions. Work in the authors' laboratories is supported by the intramural research programs of the National Institute of Child Health and Human Development and the National Institutes of Diabetes and Digestive and Kidney Diseases.

## References

1 Blobel, G. and Dobberstein, B. (1975) Transfer of proteins across membranes. I. Presence of proteolytically processed and non-processed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *J. Cell Biol.* 67, 835–851
2 von Heijne, G. (1985) Signal sequences: the limits of variation. *J. Mol. Biol.* 184, 99–105
3 Gierasch, L.M. (1989) Signal sequences. *Biochemistry* 28, 923–930
4 Talmadge, K. *et al.* (1980) Eukaryotic signal sequence transports insulin antigen in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 77, 3369–3373
5 Müller, M. *et al.* (1982) A bacterial secretory protein requires signal recognition particle for translocation across mammalian endoplasmic reticulum. *J. Biol. Chem.* 257, 11860–11863
6 Kaiser, C.A. *et al.* (1987) Many random sequences functionally replace the secretion signal sequence of yeast invertase. *Science* 235, 312–317
7 Abrahmsén, L. *et al.* (1985) Analysis of signals for secretion in the staphylococcal protein A gene. *EMBO J.* 4, 3901–3906
8 von Heijne, G. (1989) The structure of signal peptides from bacterial lipoproteins. *Protein Eng.* 2, 531–534
9 Walter, P. and Blobel, G. (1980) Purification of a membrane-associated protein complex required for protein translocation across the endoplasmic reticulum. *Proc. Natl. Acad. Sci. U. S. A.* 77, 7112–7116
10 Waters, M.G. and Blobel, G. (1986) Secretory protein translocation in a yeast cell-free system can occur posttranslationally and requires ATP hydrolysis. *J. Cell Biol.* 102, 1543–1550
11 Randall, L.L. (1983) Translocation of domains of nascent periplasmic proteins across the cytoplasmic membrane is independent of elongation. *Cell* 33, 231–240
12 Poritz, M.A. *et al.* (1990) An *E. coli* ribonucleoprotein containing 4.5S RNA resembles mammalian signal recognition particle. *Science* 250, 1111–1117



Please cite this article as: Ramanujan S. Hegde, Harris D. Bernstein, The surprising complexity of signal sequences, TRENDS in Biochemical Sciences (2006), doi:10.1016/j.tibs.2006.08.004.

13 Ulbrandt, N.D. *et al.* (1997) The *E. coli* signal recognition particle is required for the insertion of a subset of inner membrane proteins. *Cell* 88, 187–196
14 Keenan, R.J. *et al.* (1998) Crystal structure of the signal sequence binding subunit of the signal recognition particle. *Cell* 94, 181–191
15 Batey, R.T. *et al.* (2000) Crystal structure of the ribonucleoprotein core of the signal recognition particle. *Science* 287, 1232–1239
16 Peterson, J.H. *et al.* (2003) Basic amino acids in a distinct subset of signal peptides promote interaction with the signal recognition particle. *J. Biol. Chem.* 278, 46155–46162
17 Ng, D.T. *et al.* (1996) Signal sequences specify the targeting route to the endoplasmic reticulum membrane. *J. Cell Biol.* 134, 269–278
18 de Gier, J-W. *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. U. S. A.* 95, 14646–14651
19 Lee, H.C. and Bernstein, H.D. (2001) The targeting pathway of *Escherichia coli* presecretory and integral membrane proteins is specified by the hydrophobicity of the targeting signal. *Proc. Natl. Acad. Sci. U. S. A.* 98, 3471–3476
20 Lee, H.C. and Bernstein, H.D. (2002) Trigger factor retards protein export in *Escherichia coli*. *J. Biol. Chem.* 277, 43527–43535
21 de Gier, J-W. and Luirink, J. (2003) The ribosome and YidC. New insights into the biogenesis of *Escherichia coli* inner membrane proteins. *EMBO Rep.* 4, 939–943
22 Grallath, S. *et al.* (2006) L25 functions as a conserved ribosomal docking site shared by nascent chain-associated complex and signal-recognition particle. *EMBO Rep.* 7, 78–84
23 Huber, D. *et al.* (2005) Use of thioredoxin as a reporter to identify a subset of *Escherichia coli* signal sequences that promote signal recognition particle-dependent translocation. *J. Bacteriol.* 187, 2983–2991
24 Colombo, S. *et al.* (2005) *N*-Myristoylation determines dual targeting of mammalian NADH-cytochrome $b_5$ reductase to ER and mitochondria outer membranes by a mechanism of kinetic partitioning. *J. Cell Biol.* 168, 735–745
25 Anandatheerthavarada, H.K. *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, 5494–5504
26 Jungnickel, B. and Rapoport, T.A. (1995) A posttargeting signal sequence recognition event in the endoplasmic reticulum membrane. *Cell* 82, 261–270
27 Plath, K. *et al.* (1998) Signal sequence recognition in posttranslational protein transport across the yeast ER membrane. *Cell* 94, 795–807
28 Garrison, J.L. *et al.* (2005) A substrate-specific inhibitor of protein translocation into the endoplasmic reticulum. *Nature* 436, 285–289
29 Besemer, J. *et al.* (2005) Selective inhibition of cotranslational translocation of vascular cell adhesion molecule 1. *Nature* 436, 290–293
30 Van den Berg, B. *et al.* (2004) X-ray structure of a protein-conducting channel. *Nature* 427, 36–44
31 Voigt, S. *et al.* (1996) Signal sequence-dependent function of the TRAM protein during early phases of protein transport across the endoplasmic reticulum membrane. *J. Cell Biol.* 134, 25–35
32 Fons, R.D. *et al.* (2003) Substrate-specific function of the translocon-associated protein complex during translocation across the ER membrane. *J. Cell Biol.* 160, 529–539
33 Lyman, S.K. and Schekman, R. (1997) Binding of secretory precursor polypeptides to a translocon subcomplex is regulated by BiP. *Cell* 88, 85–96
34 Dunnwald, M. *et al.* (1999) Detection of transient *in vivo* interactions between substrate and transporter during protein translocation into the endoplasmic reticulum. *Mol. Biol. Cell* 10, 329–344
35 Kim, S.J. *et al.* (2002) Signal sequences control gating of the protein translocation channel in a substrate-specific manner. *Dev. Cell* 2, 207–217
36 Coppolino, M.G. and Dedhar, S. (1998) Calreticulin. *Int. J. Biochem. Cell Biol.* 30, 553–558
37 Shaffer, K.L. *et al.* (2005) Regulation of protein compartmentalization expands the diversity of protein function. *Dev. Cell* 9, 545–554
38 Garcia, P.D. *et al.* (1988) Targeting of the hepatitis B virus precore protein to the endoplasmic reticulum membrane: after signal peptide cleavage translocation can be aborted and the product released into the cytoplasm. *J. Cell Biol.* 106, 1093–1104
39 Ou, J.H. *et al.* (1989) Transport of hepatitis B virus precore protein into the nucleus after cleavage of its signal peptide. *J. Virol.* 63, 5238–5243
40 Scaglioni, P.P. *et al.* (1997) Posttranscriptional regulation of hepatitis B virus replication by the precore protein. *J. Virol.* 71, 345–353
41 Ma, J. *et al.* (2002) Neurotoxicity and neurodegeneration when PrP accumulates in the cytosol. *Science* 298, 1781–1785
42 Hegde, R.S. *et al.* (1998) A transmembrane form of the prion protein in neurodegenerative disease. *Science* 279, 827–834
43 Kim, S.J. and Hegde, R.S. (2002) Cotranslational partitioning of nascent prion protein into multiple populations at the translocation channel. *Mol. Biol. Cell* 13, 3775–3786
44 Rane, N.S. *et al.* (2004) Protection from cytosolic prion protein toxicity by modulation of protein translocation. *EMBO J.* 23, 4550–4559
45 Szabady, R.L. *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, 221–226
46 Rutkowski, D.T. *et al.* (2003) Signal sequences initiate the pathway of maturation in the endoplasmic reticulum lumen. *J. Biol. Chem.* 278, 30365–30372
47 Ott, C.M. and Lingappa, V.R. (2004) Signal sequences influence membrane integration of the prion protein. *Biochemistry* 43, 11973–11982
48 Rehm, A. *et al.* (2001) Signal peptide cleavage of a type I membrane protein, HCMV US11, is dependent on its membrane anchor. *EMBO J.* 20, 1573–1582
49 Chen, X. *et al.* (2001) Signal peptidase and oligosaccharyltransferase interact in a sequential and dependent manner within the endoplasmic reticulum. *J. Biol. Chem.* 276, 2411–2416
50 Li, Y. *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. U. S. A.* 93, 9606–9611
51 Kurys, G. *et al.* (2000) The long signal peptide isoform and its alternative processing direct the intracellular trafficking of interleukin-15. *J. Biol. Chem.* 275, 30653–30659
52 Bae, T. and Schneewind, O. (2003) The YSIRK-G/S motif of staphylococcal protein A and its role in efficiency of signal peptide processing. *J. Bacteriol.* 185, 2910–2919
53 Tokunaga, M. *et al.* (1982) Prolipoprotein signal peptidase in *Escherichia coli* is distinct from the M13 procoat protein signal peptidase. *J. Biol. Chem.* 257, 9922–9925
54 Nunn, D.N. and Lory, S. (1991) Product of the *Pseudomonas aeruginosa* gene *pilD* is a prepilin leader peptidase. *Proc. Natl. Acad. Sci. U. S. A.* 88, 3281–3285
55 Weihofen, A. *et al.* (2002) Identification of signal peptide peptidase, a presenilin-type aspartic protease. *Science* 296, 2215–2218
56 Lemberg, M.K. and Martoglio, B. (2002) Requirements for signal peptide peptidase-catalyzed intramembrane proteolysis. *Mol. Cell* 10, 735–744
57 Braud, V. *et al.* (1997) The human major histocompatibility complex class Ib molecule HLA-E binds signal sequence-derived peptides with primary anchor residues at positions 2 and 9. *Eur. J. Immunol.* 27, 1164–1169
58 Borrego, F. *et al.* (1998) Recognition of human histocompatibility leukocyte antigen (HLA)-E complexed with HLA class I signal sequence-derived peptides by CD94/NKG2 confers protection from natural killer cell-mediated lysis. *J. Exp. Med.* 187, 813–818
59 Braud, V.M. *et al.* (1998) HLA-E binds to natural killer cell receptors CD94/NKG2A, B and C. *Nature* 391, 795–799
60 Tomasec, P. *et al.* (2000) Surface expression of HLA-E, an inhibitor of natural killer cells, enhanced by human cytomegalovirus gpUL40. *Science* 287, 1031
61 Ulbrecht, 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, 5019–5022
62 Martoglio, B. *et al.* (1997) Signal peptide fragments of preprolactin and HIV-1 p-gp160 interact with calmodulin. *EMBO J.* 16, 6636–6645
63 Lindemann, D. *et al.* (2001) A particle-associated glycoprotein signal peptide essential for virus maturation and infectivity. *J. Virol.* 75, 5762–5771



Please cite this article as: Ramanujan S. Hegde, Harris D. Bernstein, The surprising complexity of signal sequences, TRENDS in Biochemical Sciences (2006), doi:10.1016/j.tibs.2006.08.004.

64 Froeschke, M. *et al.* (2003) Long-lived signal peptide of lymphocytic choriomeningitis virus glycoprotein pGP-C. *J. Biol. Chem.* 278, 41914–41920
65 Eichler, R. *et al.* (2003) Identification of Lassa virus glycoprotein signal peptide as a trans-acting maturation factor. *EMBO Rep.* 4, 1084–1088
66 Hegde, R.S. (2002) Targeting and beyond: new roles for old signal sequences. *Mol. Cell* 10, 697–698
67 Levine, C.G. *et al.* (2005) The efficiency of protein compartmentalization into the secretory pathway. *Mol. Biol. Cell* 16, 279–291
68 Panzner, S. *et al.* (1995) Posttranslational protein transport in yeast reconstituted with a purified complex of Sec proteins and Kar2p. *Cell* 81, 561–570
69 Finke, K. *et al.* (1996) A second trimeric complex containing homologs of the Sec61p complex functions in protein transport across the ER membrane of *S. cerevisiae*. *EMBO J.* 15, 1482–1494
70 Wittke, S. *et al.* (2002) Recognition of a subset of signal sequences by Ssh1p, a Sec61p-related protein in the membrane of endoplasmic reticulum of yeast *Saccharomyces cerevisiae*. *Mol. Biol. Cell* 13, 2223–2232
71 Yan, A. and Lennarz, W.J. (2005) Two oligosaccharyl transferase complexes exist in yeast and associate with two different translocons. *Glycobiology* 15, 1407–1415
72 Johnson, A.E. and van Waes, M.A. (1999) The translocon: a dynamic gateway at the ER membrane. *Annu. Rev. Cell Dev. Biol.* 15, 799–842
73 Tyedmers, J. *et al.* (2000) Homologs of the yeast Sec complex subunits Sec62p and Sec63p are abundant proteins in dog pancreas microsomes. *Proc. Natl. Acad. Sci. U. S. A.* 97, 7214–7219
74 Meyer, H.A. *et al.* (2000) Mammalian Sec61 is associated with Sec62 and Sec63. *J. Biol. Chem.* 275, 14550–14557
75 Duong, F. and Wickner, W. (1997) Distinct catalytic roles of the SecYE, SecG and SecDFyajC subunits of preprotein translocase holoenzyme. *EMBO J.* 16, 2756–2768
76 Fröderberg, L. *et al.* (2003) Versatility of inner membrane protein biogenesis in *Escherichia coli*. *Mol. Microbiol.* 47, 1015–1027
77 Nouwen, N. and Driessen, A.J. (2002) SecDFyajC forms a heterotetrameric complex with YidC. *Mol. Microbiol.* 44, 1397–1405
78 Bensing, B.A. *et al.* (2005) Determinants of the streptococcal surface glycoprotein GspB that facilitate export by the accessory Sec system. *Mol. Microbiol.* 58, 1468–1481
79 Holden, P. *et al.* (2005) Secretion of cartilage oligomeric matrix protein is affected by the signal peptide. *J. Biol. Chem.* 280, 17172–17179



Please cite this article as: Ramanujan S. Hegde, Harris D. Bernstein, The surprising complexity of signal sequences, TRENDS in Biochemical Sciences (2006), doi:10.1016/j.tibs.2006.08.004.