Protein targeting in bacterial cells should be a rather simple process, since few sub-cellular compartments exist as compared to eukaryotic cells where targeting to the nucleus, mitochondria, peroxysomes, endoplasmic reticulum or chloroplasts is required. In most cases of bacterial protein targeting, the idea is that a tag is contained within the protein, and this label gives the cell information about where the protein should be delivered. This is reminiscent of the tagging of your luggage at the airport to have it delivered where appropriate.

However, over the last few years, and especially in Gram negative bacteria, it has emerged that the protein targeting process involves much more diverse tags than expected and that the nature of some of them has still not been clearly elucidated. In this issue of *Journal of Bacteriology*, findings by the group of Cecile Wandersman highlight very well the complexity of this process (18). In essence, the targeting signal may sometimes be simple and universal, such as the classical N terminal signal peptide, which is used to
deliver proteins to the Sec machinery and to allow translocation of proteins across the
cytoplasmic membrane (24). Such a mechanism is very similar to the one used in
eukaryotic cells for targeting proteins into the lumen of the endoplasmic reticulum. A
slightly different signal was later discovered, which contains a twin arginine motif. This
signal is required for the translocation of proteins already folded in the cytoplasm, and is
known as the Tat (twin arginine translocation) signal peptide (4). The twin-arginine
translocation (Tat) system also operates in the chloroplast thylakoid (20).

In Gram negative bacteria the complexity of the targeting mechanism comes from the fact
that these organisms have evolved a large number of molecular machines which are
involved in protein translocation across the outer membrane. These translocated proteins
can then remain anchored to the cell surface or be secreted into the extracellular medium.
In all cases, although the end result is identical, the mechanisms involved are drastically
different. Therefore the cell needs to select the suitable substrates for each of these
machineries, using a specific recognition process, and avoiding all invalid pairing, which
may jeopardize the function of the system.

These secretion machineries, called type I to type VI secretion systems (T1SS to T6SS)
may allow either translocation across all compartments of the cell envelope, or solely
across the outer membrane (5, 8, 10, 12, 13, 16). The targeting signal can sometimes be
easily identifiable as it is for the Sec-dependent signal peptide. An obvious signal for
targeting was also believed to exist in the T1SS, whose substrates carry a non-cleavable C
terminal signal able to bind the ABC component of the T1SS. This signal does not have a
conserved amino acid sequence, but a helical secondary structure spanning about 30-60
residues that in principle cannot be removed without loss of targeting/secretion (27).
However, studies in the laboratory of Cecile Wandersman have previously shown that the T1SS-dependent HasA substrate from *Serratia marcescens* can be efficiently targeted to the secretion machinery despite lacking 14 C-terminal amino acids (6). Although targeting is efficient, secretion is impaired and the machinery is jammed. In this issue of *Journal of Bacteriology*, the same group has further characterized the additional sites that are important to bind the secreted substrate to the secretion machinery (18), thereby demonstrating that targeting does not solely rely on the C-terminal residues.

In brief, a large number of studies based on haemolysin HlyA, an *Escherichia coli* T1SS substrate, suggested that the secretion signal lies within the last 53 amino acids. This region includes a potential 18-amino-acid amphiphilic alpha-helix, a cluster of charged residues, and a weakly hydrophobic terminal sequence (17). Furthermore, the nucleotide-binding domain of HlyB and a C-terminal 23 kDa fragment of HlyA have been shown to interact with each other in a specific manner (3). This specificity is confirmed by the fate of a chimera made from the *Pseudomonas aeruginosa* T1SS substrate AprA fused to the C terminus of HlyA. This protein was targeted to the Hly transporter and not to the AprA transporter (11). In this context the general view has been that the targeting might rather be post-translational, since the C terminus of the secreted protein has to emerge from the ribosome to be accessible for the secretion machinery.

The observation by Cecile Wandersman’s group that C-terminal truncation of HasA does not prevent targeting is a real step forward in understanding the dynamic of the system. The fact that the truncation blocks the secretion, indeed suggests that the role of the C terminus might not be crucial in the early stages of the process but rather at later stages and particularly for the release of the secreted protein from the T1SS machine. By
performing a systematic mutagenesis resulting in the insertion of pentapeptides within the HasA sequence, Cecile Wandersman and collaborators point out additional regions, located outside of the C terminus that are required for efficient targeting. This is an elegant strategy, since the selected mutations prevent targeting and thus relieve the blocking imposed by the truncated HasA. The fact that the presence of multiple insertions has an additive effect in decreasing the targeting efficiency, might suggest that these motifs are linear and act sequentially rather than forming a conformational motif, whose disruption will result in a yes/no targeting phenotype. This is also supported by the fact that unfolded HasA is still efficiently targeted to the HasD ABC transporter of the T1SS machinery. Finally, since the HasA structure is available (2), it was observed that the different regions identified as potential additional secretion signals do not fall into a unique patch within the HasA 3D structure.

The idea of a conformational signal for recognition by the secretion machinery has been carried over in other cases. With the type II secretion system (T2SS), the proteins are first translocated into the periplasm in a Sec- or Tat-dependent manner before being targeted to the T2SS (25). Once in the periplasm, the protein folds and is transported through a large secretin channel across the outer membrane. It has been shown that folding is a prerequisite, which suggests that the signal is a conformational patch (7, 26). Several studies in the laboratory of Anthony Pugsley identified distinct regions within the *Klebsiella oxytoca* pullulanase (PulA) amino acid sequence, which are important for targeting to the T2SS Pul machinery. These two 80-amino-acid stretches, designated A and B, can promote secretion of β-lactamase and endoglucanase CelZ (14). Since the 3D structure of PulA is not known it is difficult to predict whether these two regions could
form a conformational signal, or whether as in the case of HasA, these regions are additional sequences, which could facilitate progressive recognition of the substrate by the secretion machinery.

This case of a T2SS-dependent substrate is the only one for which folding has clearly been shown to be a prerequisite to transport. The fact that in other cases unfolding is the likely state of the protein for translocation does not make the task easier for identifying secretion signals. In *Helicobacter pylori* the translocation of the T4SS-dependent substrate CagA, depends on the presence of its 20 C-terminal amino acids, in which a number of positively charged residues can be found (15). The *Bartonella* T4SS effector proteins BepA-G also display a short positively charged tail sequence required for targeting (22). Moreover, replacing the C-terminal region of the T4SS-dependent CagA with that of other type IV-secreted proteins reconstitutes CagA translocation competence. However, this does not confer a change in specificity, as it is the case with the T1SS, where replacement of the AprA C-terminal sequence by the HlyA C-terminal sequence not only restores secretion competence but also switches the machinery to which the substrate is targeted (11). Finally, the C-terminal signal of the T4SS substrate does not seem sufficient for targeting since removing N-terminal part of the CagA protein also renders the protein translocation-incompetent as well (15). This indicates that, as for HasA described in the study by the Wandersman laboratory, additional sequences may contribute to optimal targeting and further translocation.

The nature and role of all of the secretion signals may be many-fold. They could be used to directly interact with a component of the machinery, as it is clear for T1SS substrates that interact with the ABC component of the T1SS machinery (3). Alternatively, they can
provide the binding site for a chaperone, which will target the complex towards the machinery. For example, it has been shown that CagF is a chaperone that interacts with a 100 aa-long region adjacent to the C-terminal secretion signal of CagA. Furthermore, CagF binding precedes recognition of the C-terminal CagA translocation signal, and both steps are required to recruit CagA to the type IV translocation channel (19). In the case of the type III secretion machinery, the secretion signal was found at the N terminus of the effector (23). Strikingly, from other T3SS studies the nature of the signal has been proposed to be located within the secondary structure of the mRNA rather than being directly dependent on the amino acid sequence (1). Despite a lot of studies ongoing in this area, no one has yet succeeded in reconciling all of these different concepts.

The idea presented in the paper by the Wandersman laboratory, suggesting that one part of the secretion signal is required for targeting and that other regions, such as the C terminus of HasA, are required for dissociation and release of the substrate from the machinery, is a very attractive concept. In the case of the type V secretion system, also called the two-partner secretion pathway, TpsA is the secreted substrate, whereas TpsB is the transporter localized in the outer membrane (16). It has often been suggested that a motif within the secreted protein TpsA, known as TPS motif, is responsible for targeting by binding to a domain within the TpsB transporter, which is also known as the POTRA domain (9). However, it was recently shown that a Pseudomonas aeruginosa TpsA-like protein, CupB5, could jam a transporter, CupB3, which lacks its POTRA domain (21). This indicates that in the absence of POTRA, targeting of a TpsA to the translocation machinery is possible, however completion of the secretion mechanism is affected. That is exactly what is observed with HasA lacking its C terminus, which is targeted but jams
the secretion machinery. A generalization of such a kinetically ordered role of various domains in the secreted protein at various stages of the secretion process could be the path to follow in future research on secretion signals. The protein is first recognized and then moves along the machinery and this requires a series of recognition events, which in combination maintain the efficiency and specificity of the secretion process. Identification of such series of sequential recognition events is reminiscent of studies that have shown, using an original method called TrIP, that DNA transported across the T4SS channel can be trapped at different stages during the transport process (8). This technique allowed the identification of the sequential protein partners during the translocation process.

In conclusion, we have a long way to go to decipher the full complexity of protein targeting in bacteria. This is an intriguing puzzle that so far remains unanswered. However, the elegant combination of biochemical and genetic approaches used in the study on HasA by the Wandersman laboratory is one example of the way to follow to solve this issue.

References


