Following the leader: bacterial protein export through the Sec pathway

Anastassios Economou

Significant strides have been made during the past 20 years in our understanding of protein secretion across the bacterial inner membrane. Specialized chaperones select secretory polypeptide chains and usher them to a membrane-embedded preprotein translocase. This unique molecular machine envelops the polymeric substrate and migrates along its length in defined, energy-dependent steps. Consequently, preproteins are gradually pumped into the periplasm where they acquire their native, folded conformation.

A. Economou is in the Institute of Molecular Biology and Biotechnology-FORTH, and also in the Dept of Biology, University of Crete, PO Box 1527, Iraklio, Crete, GR-71110, Greece.
tel: +30 81 391167,
fax: +30 81 391166,
e-mail: aeconomo@imbb.forth.gr

Information transfer in the living cell does not end with the biosynthesis of a polypeptide chain. The familiar central dogma of molecular biology comes with essential small print: a polypeptide chain will only fulfil its designated function if it is correctly folded. Additionally, for a significant subset of cellular proteins (more than 30% of the total proteome), folding must take place in the correct subcellular, membrane-enclosed compartment. The modern cell has developed elaborate mechanisms to deal with both the membrane translocation and the secretion of polypeptides, as well as with their subsequent folding: enter the protein translocases (also called translocons) and chaperones. This review focuses on protein secretion, which is, in effect, nothing more than a means of topologically guiding and temporally controlling protein folding.

Protein-secretion research was inspired by a conundrum: how can a long heteropolymer that contains hydrophilic and hydrophobic regions and that acquires a tightly folded native conformation cross a hydrophobic lipid membrane? Evolution has solved this problem not once but several times, using dissimilar molecular machines in each case. In Bacteria, there are at least five non-homologous protein export systems, allowing them not only to construct their membranes and cell envelope but also to secrete an army of toxins, adhesins and hydrolytic enzymes. Of these systems, only one, the Sec (for secretion) system, is essential for cell viability and this system has also been found in the Archaea and Eukarya[1].

Currently, all the essential components of the Sec pathway are known and an impression of the secretion mechanism and its energetics has emerged[1–4]. In contrast to the situation in the mammalian endoplasmic reticulum, protein secretion across the bacterial plasma membrane mainly occurs post-translationally. This distinguishing feature has allowed full reconstitution of the bacterial reaction using purified components[2,4].

The Sec pathway can be divided into three distinct but sequential and interdependent stages: targeting, translocation and release (Fig. 1). In stage I, preprotein substrates are guided to exit sites in the membrane. In stage II, the exiting chain crosses the the lipid bilayer, probably through the translocase. Finally, in stage III, the translocated chain is released and allowed either to acquire its native folded state in the periplasm or to proceed to the outer membrane for integration (reviewed in Ref. 3).

Catalysts and energetic requirements

At least 11 proteins and one RNA species are known to be directly involved in the secretion pathway in *Escherichia coli* (Fig. 1b). Stage I of the reaction requires two secretion-specific chaperones: SecB (Refs 5–7) and a ribonucleotide complex, the signal-recognition particle (SRP), which is composed of the Ffh protein and a 4.5S-RNA species[4,8–11]. SecB and the SRP both recognize subsets of secreted polypeptides, with the SRP being particularly important for the targeting of polytopic (multiple-membrane-spanning) membrane proteins[8–11]. In certain cases, the additional involvement of housekeeping chaperones, such as GroEL (Ref. 12), DnaK (Ref. 13) and trigger factor[8], has also been documented.

Stage II of the reaction is catalysed by preprotein translocase, a biochemically fascinating and complex molecular machine (Fig. 1b). The translocase has a membrane-spanning domain containing heterotrimers of the polytopic membrane proteins SecY, SecE and SecG. Using electron microscopy, it has been estimated that up to three SecYE dimers can assemble in a quasipentagonal, ring-shaped structure built around a putative pore[14]. In contrast to SecY and SecE, SecG is not essential for viability in most *E. coli* strains tested[15], is not conserved outside the Bacteria and is apparently not required for the formation of the SecYE ring[14].

In addition to these mainly structural subunits, the translocase has a mechanical motor device, the SecA ATPase. SecA is essential and unique to Bacteria[1]; the eukaryotic translocase uses a different ATPase. SecA is a large, elongated, dimeric molecule[16] comprising two primary domains[2,4]: the ATPase (amino-terminal)

0966-842X/99/$ - see front matter © 1999 Elsevier Science Ltd. All rights reserved. PII: S0966-842X(99)01555-3

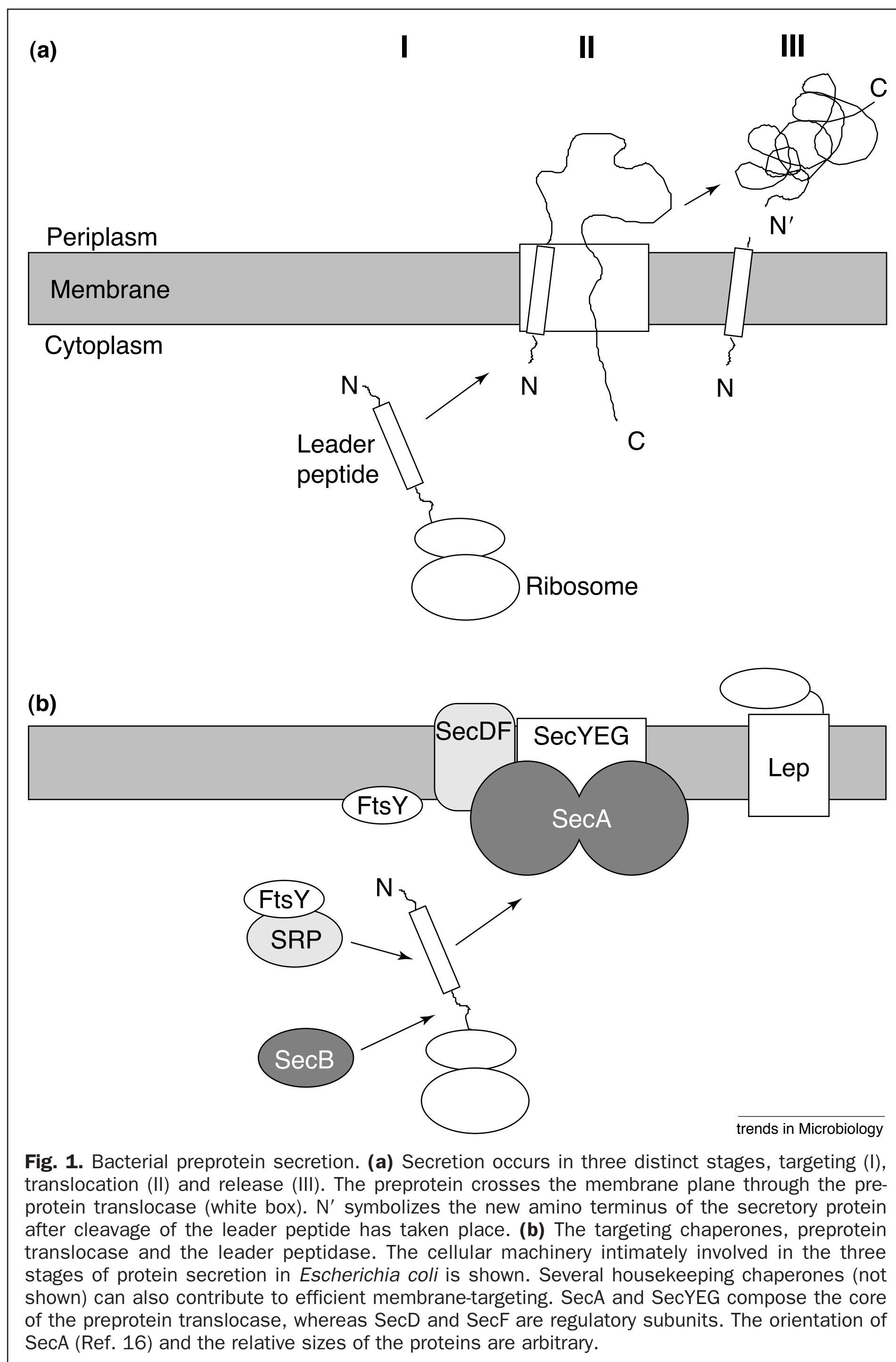

Fig. 1. Bacterial preprotein secretion. **(a)** Secretion occurs in three distinct stages, targeting (I), translocation (II) and release (III). The preprotein crosses the membrane plane through the preprotein translocase (white box). N′ symbolizes the new amino terminus of the secretory protein after cleavage of the leader peptide has taken place. **(b)** The targeting chaperones, preprotein translocase and the leader peptidase. The cellular machinery intimately involved in the three stages of protein secretion in *Escherichia coli* is shown. Several housekeeping chaperones (not shown) can also contribute to efficient membrane-targeting. SecA and SecYEG compose the core of the preprotein translocase, whereas SecD and SecF are regulatory subunits. The orientation of SecA (Ref. 16) and the relative sizes of the proteins are arbitrary.

domain and the dimerization (carboxy-terminal) domain (Fig. 2). The carboxy-terminal domain allows SecA to bind to SecYEG, leading to the assembly of the functional translocase core[2,17–23]. SecA, unequivocally, is central to bacterial secretion, as it not only provides a physical link between most of the reaction components (the functional regions and ligands are shown in Fig. 2) but also provides a mechanism for the conversion of energy into intersubunit communication and translocase work[4,22–24]. The auxiliary, nonessential[1–3] translocase subunits SecD and SecF optimize the secretion reaction[2,25]. Acidic phospholipids are also an important component of the active preprotein translocase and are required for activation of SecA, correct leader-peptide–membrane interaction and SecYEG stability[26]. The sources of energy for the translocase machinery are both chemical (ATP) and electrochemical [proton motive force (PMF)]. Whereas ATP is essential, the PMF enhances translocation rates[2,4].

Stage III involves the leader peptidase, Lep. Lep is a serine peptidase that removes leader peptides after transfer of the mature domain of the exiting chain to the *cis* side of the membrane[27]. A homologous protein is responsible for cleaving the leader peptides of secreted lipoproteins.

The information discovered to date is the result of pioneering work using *E. coli* as a model system. Bacterial genomics has vindicated this choice by establishing that *E. coli* can indeed serve as a true paradigm for bacterial protein secretion. All the essential *sec* and SRP-related genes are present in every bacterium whose genome sequence has become available. Nevertheless, interesting variations on a theme have also been revealed and more are expected. For example, in *Mycobacterium tuberculosis*, two distinct *secA* orthologues exist in the same cell. These two SecAs might have different substrate specificities. Some Gram-positive bacteria, such as *Bacillus subtilis*, carry at least five different Lep isozymes[28] and their SecD and SecF proteins are fused in a single polypeptide[29].

The substrate

The secretion pathway is used by hundreds of substrates (including many heterologous ones used in the biotechnology industry) that do not share any sequence similarity. How does the cell discriminate between cytoplasmic-resident proteins and secreted proteins, and maintain correct subcellular sorting? The information on the final cellular destination of all secretory polypeptides is contained in their primary sequence in the form of a 20–30-residue amino-terminal extension known as the leader, or signal, peptide. Although leader peptides do not show strict conservation of their primary sequence, they do share common chemical properties,

including a central hydrophobic core[30]. Leader peptides are not physical constituents of the native protein and are removed from the preprotein by Lep in a precise endoproteolytic event.

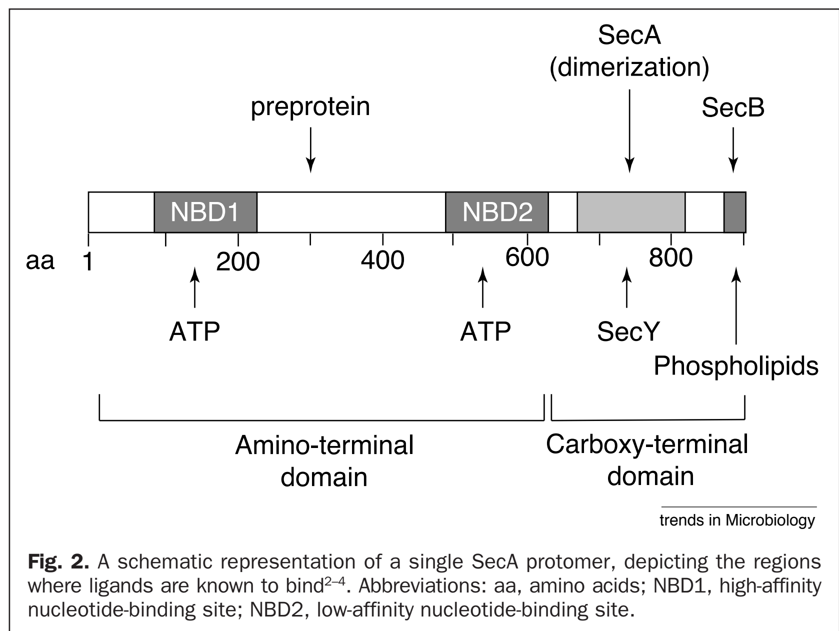

Fig. 2. A schematic representation of a single SecA protomer, depicting the regions where ligands are known to bind[2–4]. Abbreviations: aa, amino acids; NBD1, high-affinity nucleotide-binding site; NBD2, low-affinity nucleotide-binding site.

Targeting the preprotein to the membrane

By virtue of their leader peptides, secretory proteins get trapped in the export pathway by association with chaperones. Cytoplasmic proteins, which do not carry leader sequences, generally escape the attention of chaperones and are allowed to fold in the cytosol. Nascent leader peptides delay the folding of the whole polypeptide[2,6,7], thus allowing chaperones such as GroEL (Ref. 12) or SecB (Refs 5–7) to bind. In contrast to SecB, which binds to the mature part of the preprotein[7], the SRP specifically recognizes the leader peptide and, together with the FtsY GTPase[8,10], mainly targets hydrophobic, polytopic proteins[9–11]. The final outcome of these different binding reactions is the arrival of preproteins at the inner surface of the cell membrane.

In a remarkable demonstration of biochemical efficiency and fidelity, all the different ternary complexes of preproteins and chaperones bind to the same membrane receptor: SecA, the peripheral subunit of the translocase. SecA binding is facilitated by the leader peptide, which specifically recognizes SecA even in the absence of any other auxiliary factors[31]. In addition, the mature domain of the polypeptide chain, as well as the chaperones, has an affinity for SecA (Refs 2,5,12). A specific binding site for SecB has been identified at the extreme carboxyl terminus of SecA (Ref. 5). Despite the fact that up to 50% of cellular SecA is cytosolic, this fraction is unlikely to contribute to preprotein targeting, as it has a low affinity for SecB (Ref. 32) and for the preprotein substrate[2,31].

On initiation of the translocation reaction, the mature part of the preprotein is dislodged from the SecB chaperone and transferred onto SecA that is already bound to SecYEG at the membrane[33]. Following this transfer, SecB is no longer required and is expelled from the membrane[5] and recycled.

Crossing the sea of lipid

ATP-driven SecA cycling

Stage II of the secretion reaction takes place wholly within the membrane plane. As a result, and despite remarkable progress, many aspects of this reaction stage are still poorly understood (reviewed in Ref. 4). A central mechanistic element of translocase activity is that SecA uses the energy from ATP to fuel cycles of membrane insertion and de-insertion[23,24]. In the inserted state, significant parts of SecA are buried in the membrane and become inaccessible to exogenously added proteases[23]. The binding energy of ATP at the high-affinity nucleotide-binding site NBD1 (Fig. 2) promotes membrane insertion and hydrolysis of ATP drives de-insertion. Initially, only the carboxy-terminal third of SecA was shown to become protease protected in a complete *in vitro* translocation reaction[23] and to be exposed to the periplasmic phase in non-translocating membranes[34,35]. Subsequently, using different labelling techniques, it was demonstrated that some amino-terminal regions could also be embedded in the membrane[21,36,37]. A more-detailed examination of membrane-inserted SecA is required as the probes used to date give only a coarse topological picture. Several residues of the presumed membrane-inserted SecA are probably still exposed to solvent[37,38]. In some cases, the extreme carboxyl terminus is found facing the cytosol[23]; in others, it faces the periplasm[35]. An inherent problem with some of these early experimental techniques was that they could not discriminate between SecA bound to SecYEG and SecA bound to lipid. A significant recent advance is the isolation of chemically pure SecA–SecYEG in a micellar solution[39]. This system will permit the study of translocase oligomerization, structure and conformation during catalysis.

The membrane insertion of SecA occurs at SecYEG (Refs 18,20,23,40). Phospholipids have not been detected in the immediate vicinity of membrane-inserted SecA (Refs 41,42), suggesting that ATP-driven SecA membrane penetration could take place largely within a proteinaceous environment. Such an environment would be provided by the SecYE ring structure[14]. The exciting possibility remains that SecYE, together with SecG (not included in the study of Meyer *et al.*[14]), forms a rather flexible complex that can accommodate SecA insertion and cyclical movement.

Once inserted, SecA can be de-inserted from the membrane following ATP hydrolysis[18,22,23,43]. SecA de-insertion occurs only during translocation[22,23,25] and allows recycling of the enzyme for further translocation reactions. This reaction cycle makes use of the ATP interactions at NBD1. The low-affinity

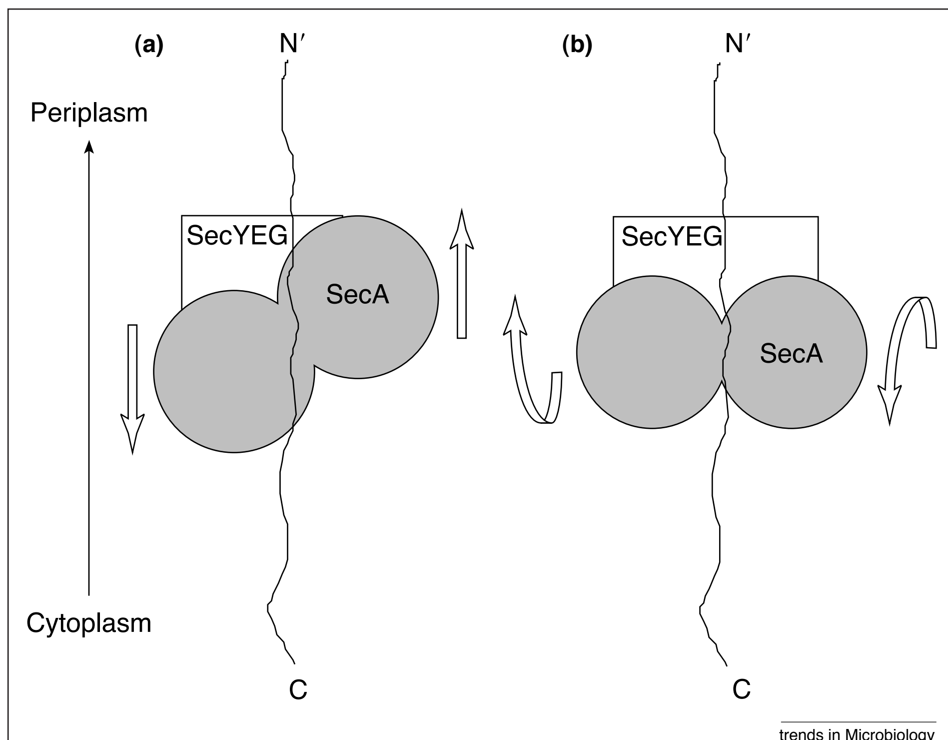

Fig. 3. Two models of translocase processivity. The translocase 'walks'[4] along its polymeric substrate in defined steps[48] as a result of ATP-driven SecA membrane cycling[23,24]. The size of each step is probably dictated by the degree of movement of, and the distance covered by, the SecA domains during conformational changes, whether in the form of deep membrane-penetration **(a)** or relative rotational movement of the two protomers **(b).** The two SecA protomers might participate in different binding subreactions, thus contributing to SecA asymmetry, which is expected to be important for processive forward movement[4]. In this schematic representation, the preprotein translocase is shown in the absence of the essential phospholipid component[26] for simplicity and to emphasize its similarity with other processive cellular machines known to occur in the cytoplasm, such as nucleic acid polymerases and helicases. N′ represents the new amino-terminus of the secretory protein after cleavage of the leader peptide has taken place.

NBD2 (Fig. 2) is essential for protein translocation but not for SecA cycling[24] and its precise role remains undetermined.

Regulation of SecA cycling

Recently, several previously identified disparate translocase subreactions have been shown to be linked to the central reaction of SecA cycling. SecG undergoes an inversion of its topology that is coupled to insertion of SecA into the membrane[44]; this change has been proposed to facilitate SecA cycling[25,39,42,44,45]. The contribution of SecG can be more clearly observed when translocation is compromised: at low temperatures[44], in the absence of SecDF (Ref. 25) or at low PMF levels[22]. Conversely, in the absence of SecG, two other regulatory translocase subunits, SecD and SecF, become important[43]. SecDF has been shown to stabilize membrane-inserted SecA[24,46]. Finally, the well-documented stimulation of ATP-driven translocation by the PMF (Ref. 2) is exerted via SecA de-insertion from the membrane[22]. The PMF might affect SecY conformation directly[22,47]. For the first time, these exciting observations place the elusive molecular mechanism of PMF-driven translocation within experimental reach.

SecA membrane cycling as a mechanical device

SecA membrane cycling offers a mechanistic path for the conversion of energy into work (i.e. the forward movement of the preprotein through the export channel). Under the same conditions required for SecA insertion, and in the same time frame, short stretches of the substrate also enter the membrane plane and become protected from proteases. The SecA-mediated forward preprotein movement occurs in two distinct steps[48]. Initially, the mere binding of SecA to the exiting polypeptide chain causes limited translocation of ~20 amino acid residues. Subsequently, ATP binding to SecA promotes the forward translocation of another 15–20 amino acid residues. At this stage, preproteins are threaded through the translocase and are close to SecA and SecY (Refs 2,4). SecDF-mediated stabilization of inserted SecA prevents the reverse movement of translocation intermediates[46]. When the ATP bound to SecA is hydrolysed, the polymeric preprotein (or rather parts of it) is released from SecA into the SecYE 'pore'. Multiple rounds of SecA insertion and de-insertion are proposed to drive the translocation of a whole polypeptide chain in segmental fashion, without the substrate ever leaving the translocase. When the substrate is halfway through the translocase, the PMF alone can complete translocation if SecA has been removed[2,46].

The combination of the two SecA mechanical strokes described above leads to a model of enzymatic processivity that provides a simple basic framework within which bacterial translocation can be viewed[4]. These mechanistic events allow the translocase to 'walk' along the length of the polymeric substrate (Fig. 3). One exciting possibility is that processive steps are generated by each of the two SecA protomers during non-simultaneous insertion (Fig. 3a) or rotational movement (Fig. 3b) of the two protomers in the membrane. To test this model, it is essential that SecA membrane cycling is examined further using quantitative biophysical and biochemical methods.

Going native

Once a portion of the exiting chain has been exposed to the periplasm, the Lep cleavage site is available for proteolytic attack. Lep is a membrane protein with a large periplasmically exposed hydrophilic domain that carries the catalytic centre[27]. This domain could be positioned over the translocase exit in proximity to

the leader peptide cleavage sites. Once the leader peptide is cleaved, the remaining part of the mature chain is no longer attached to the membrane and can begin folding as soon as sufficient chain segments exit the translocase. Based on the observation that antibodies against the large periplasmic domain of SecD prevent preprotein release from the membrane, a role for SecD in this stage of the reaction has been proposed[49]. However, the same role has not been observed for the *B. subtilis* SecDF protein[29]. Given the involvement of SecDF in stabilizing membrane-inserted SecA[44,46], it is possible that SecDF influences periplasmic release indirectly, via one of the essential Sec proteins. As mutations in practically any Sec protein lead to the accumulation of membrane-threaded and unreleased intermediates[2,3], it follows that stage III reactions are intimately connected to productive ongoing translocation in the preceding stage.

Several periplasmic-folding catalysts have recently been identified. These proteins accelerate specific subreactions, such as disulphide-bond formation or proline isomerization, and, in concert with other factors, play an important role in the folding of soluble periplasmic proteins as well as outer-membrane proteins[3,50]. A connection between these folding reactions and a secreted polypeptide exiting the translocase remains to be established.

Conclusions

Protein secretion is a complex, multistep reaction. With a complete inventory of the biochemical components that relay the substrate from the cytoplasm to the periplasm, the field has now entered a mature phase. The emphasis has shifted to determining the atomic structures of the components and the biophysical parameters and kinetics of intermediate reaction stages. The first low- and high-resolution structures of components of the Sec pathway are already available. The combination of structure determination with traditional biochemistry and molecular biology has allowed us to take full advantage of the data provided by bacterial genomics. Understanding the molecular mechanics of this reaction promises to teach us a lot about basic features of other nanomachines that handle polymeric substrates in the modern cell, such as polymerases, proteases, chaperones and helicases. Moreover, we can now manipulate the secretion pathway to allow bacterial production of biopharmaceuticals[51] and industrial enzymes as secreted proteins.

Acknowledgements

This paper is dedicated to Prof. W. Wickner, who started it all. Research in my lab is supported by grants from the European Union Directorate of Science and Technology (TMR-ERBFMRXCT960035, Biotech2-BIO4-CT97-2244 and Biotech 2-BIO4-CT98-0051), the Greek Secretariat of Research and Technology (GRI-088-97 and EPETII No: 97 EKBAN 2-17) and by the University of Crete Research Fund (ELKE; KA 1194).

Questions for future research

- What is the structure of the translocase at atomic resolution?
- What is the molecular mechanism of translocase enzyme processivity?
- What are the quantitative bioenergetics of the ATP- and PMF-driven translocation components?
- What is the molecular mechanism of preprotein recognition by the translocase?
- How is translocation and lipid-bilayer insertion of the interacting polytopic membrane proteins achieved by SRP and SecA interaction?
- How does protein secretion integrate with cell physiology?

References

1 Pohlschroder, M. *et al.* (1997) *Cell* 91, 563–566
2 Duong, F. *et al.* (1997) *Cell* 91, 567–573
3 Danese, P.N. and Silhavy, T.J. (1998) *Annu. Rev. Genet.* 32, 59–94
4 Economou, A. (1998) *Mol. Microbiol.* 27, 511–518
5 Fekkes, P., van der Does, C. and Driessen, A.J.M. (1997) *EMBO J.* 16, 6105–6113
6 Randall, L.L. *et al.* (1998) *Methods Enzymol.* 290, 444–459
7 Randall, L.L. *et al.* (1998) *Protein Sci.* 7, 1195–2000
8 Valent, Q.A. *et al.* (1995) *EMBO J.* 14, 5494–5505
9 Ulbrandt, N.D., Newitt. J.A. and Bernstein, H.D. (1997) *Cell* 88, 187–196
10 Valent, Q.A. *et al.* (1998) *EMBO J.* 17, 2504–2517
11 de Gier, J.W. *et al.* (1998) *Proc. Natl. Acad. Sci. U. S. A.* 95, 646–651
12 Bochkareva, E.S., Solovieva, M.E. and Girshovich, A.S. (1998) *Proc. Natl. Acad. Sci. U. S. A.* 95, 478–483
13 Wild, J. *et al.* (1996) *J. Bacteriol.* 178, 3608–3613
14 Meyer, T.H. *et al.* (1999) *J. Mol. Biol.* 285, 1789–1800
15 Bost, S. and Belin, D. (1997) *J. Biol. Chem.* 272, 4087–4093
16 Shilton, B. *et al.* (1998) *FEBS Lett.* 436, 277–282
17 Snyders, S., Ramamurthy, V. and Oliver, D. (1997) *J. Biol. Chem.* 272, 11302–11306
18 Matsumoto, G., Yoshihisa, T. and Ito, K. (1997) *EMBO J.* 16, 6384–6393
19 Manting, E.H., van der Does, C. and Driessen, A.J.M. (1997) *J. Bacteriol.* 179, 5699–5704
20 van der Wolk, J.P. *et al.* (1998) *EMBO J.* 17, 3631–3639
21 Chen, X., Brown, T. and Tai, P.C. (1998) *J. Bacteriol.* 180, 527–537
22 Nishiyama, K-I. *et al.* (1999) *EMBO J.* 18,1049–1058
23 Economou, A. and Wickner, W. (1994) *Cell* 78, 835–843
24 Economou, A. *et al.* (1995) *Cell* 83, 1171–1181
25 Duong, F. and Wickner, W. (1997) *EMBO J.* 16, 2756–2768
26 van Klompenburg, W. and deKruijff, B. (1998) *J. Membr. Biol.* 162, 1–7
27 Paetzel, M., Dalbey, R.E. and Strynadka, N.C. (1998) *Nature* 396, 186–190
28 Bron, S. *et al.* (1998) *J. Biotechnol.* 64, 3–13
29 Bolhuis, A. *et al.* (1998) *J. Biol. Chem.* 273, 21217–21224
30 Izard, J.W. and Kendall, D.A. (1994) *Mol. Microbiol.* 13, 765–773
31 Miller, A., Wang, L. and Kendall, D.A. (1998) *J. Biol. Chem.* 273, 11409–11412
32 den Blaauwen, T. *et al.* (1997) *Biochemistry* 416, 35–38
33 Fekkes, P. *et al.* (1998) *Mol. Microbiol.* 29, 1179–1190
34 Kim,Y.J., Rajapandi, T. and Oliver, D.B. (1994) *Cell* 78, 845–853
35 van der Does, C. *et al.* (1996) *Mol. Microbiol.* 22, 619–629
36 Eichler, J. and Wickner, W. (1997) *Proc. Natl. Acad. Sci. U. S. A.* 94, 5574–5581
37 Ramamurthy, V. and Oliver, D.B. (1997) *J. Biol. Chem.* 272, 23239–23246
38 den Blaauwen, T. *et al.* (1997) *Biochemistry* 36, 9159–9168
39 van der Does, C. *et al.* (1998) *Biochemistry* 37, 201–210

40 Matsumoto, G., Mori, H. and Ito, K. (1998) *Proc. Natl. Acad. Sci. U. S. A.* 95, 13567–13572
41 Eichler, J., Brunner, J. and Wickner, W. (1997) *EMBO J.* 16, 2188–2196
42 van Voorst, F. *et al.* (1998) *Biochemistry* 37, 12261–12268
43 Suzuki, H., Nishiyama, K. and Tokuda, H. (1998) *Mol. Microbiol.* 29, 331–341
44 Nishiyama, K-I., Suzuki, T. and Tokuda, H. (1996) *Cell* 85, 71–81
45 Ramamurthy, V., Dapic, V. and Oliver, D. (1998) *J. Bacteriol.* 180, 6419–6423
46 Duong, F. and Wickner, W. (1997) *EMBO J.* 16, 4871–4879
47 Nouwen, N., de Kruijff, B. and Tommassen, J. (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 5953–5957
48 van der Wolk, J.P., de Wit, J.G. and Driessen, A.J.M. (1997) *EMBO J.* 16, 7297–7304
49 Matsuyama, S., Fujita, Y. and Mizushima, S. (1993) *EMBO J.* 12, 265–270
50 Missiakas, D. and Raina, S. (1997) *J. Bacteriol.* 179, 2465–2471
51 Binnie, C., Cossar, J.D. and Stewart, D.I. (1997) *Trends Biotechnol.* 15, 315–320

Mapping regulatory networks in microbial cells

Ruth A. VanBogelen, Kenneth D. Greis, Robert M. Blumenthal, Travis H. Tani and Rowena G. Matthews

Genome sequences are the blueprints of diverse life forms but they reveal little information about how cells make coherent responses to environmental changes. The combined use of gene fusions, gene chips, 2-D polyacrylamide gel electrophoresis, mass spectrometry and 'old-fashioned' microbial physiology will provide the means to reveal a cell's regulatory networks and how those networks are integrated.

R.A. VanBogelen and K.D. Greis are in the Parke-Davis Pharmaceutical Research Division, Warner-Lambert, Ann Arbor, MI 48105, USA; K.D. Greis is currently in the Structural Biology Section, Procter & Gamble, Health Care Research Center-Discovery, 8700 Mason-Montgomery, Mason, OH 45040-9462, USA; R.M. Blumenthal is in the Dept of Microbiology and Immunology, Medical College of Ohio, Toledo, OH 43614-5806, USA; T.H. Tani and R.G. Matthews are in the Biophysics Research Division and Dept of Biological Chemistry, University of Michigan, Ann Arbor, MI 48109-1055, USA.*
**tel: +1 734 764 9459,*
fax: +1 734 764 3323,
e-mail: rmatthew@umich.edu

Living cells are the ultimate in complex systems, having tens of thousands of components that have co-evolved over billions of years in order to function well together in dynamically changing environments. The stakes for understanding these particular complex systems are high, with implications for human health, agriculture and the ecological health of the biosphere. The nucleotide sequences for entire genomes are now available for a rapidly growing number of microorganisms, and we can identify, with some precision, the full complement of gene products these organisms can make. However, this situation is similar to being given a list of electronic components but no wiring diagram: although some components might suggest a function, understanding the connections is essential.

Consider a unicellular prokaryote. Its ability to respond quickly to complex environmental changes is essential for its survival. Although such organisms could, in theory, have specific regulators tailored to each of their several-thousand genes, this would represent a tremendous genetic burden. Instead, regulators usually control numerous genes and sometimes even control other regulators. These networks of regulators provide important integration of the cell's responses to various environmental shifts[1]. Identifying regulatory networks is essential if the goal is to predict or control a cell's behavior.

There has recently been an explosion of technical advances for studying cells, making it possible to analyse regulatory networks that could not have been examined even a decade ago. This review summarizes the available strategies for determining the wiring diagram (regulatory networks) for organisms that have a genomic 'parts list'. We will not focus on the specific mechanisms by which regulatory proteins act, because these mechanisms are often highly species-dependent. We also warn against being seduced by the alluring charms of the new methods; they are powerful and essential, but can give a seriously skewed view of the cell if used in isolation. A combination of several of the new methods must be coupled with more established methods to elucidate global regulatory networks.

Studying global responses and regulatory networks

Adaptation by an organism to an environmental stimulus is not simply a group of independent responses by individual genes, but rather a coordinated series of linked events. Perhaps the most difficult tasks in studying global regulation are determining the appropriate experiments to elucidate the components

0966-842X/99/$ - see front matter © 1999 Elsevier Science Ltd. All rights reserved. PII: S0966-842X(99)01540-1