PROTEIN LOCALIZATION IN CELLS

R. NAGARAJ
Centre for Cellular and Molecular Biology, Hyderabad 500 007, India.

ABSTRACT

An important step in the biosynthesis of cellular proteins is the correct location of the proteins within various compartments in the cell. Many proteins are also secreted out of cells. In this review, the chemical nature of the 'molecular address' or signal sequences, the various components of the cells export machinery and the models that have been proposed to explain how the specific sorting of proteins occurs, are discussed.

INTRODUCTION

AN important step in the biosynthesis of cellular proteins is their correct location within various compartments in the cell. Also many cells are capable of protein secretion. The manner in which cells accurately sort out their proteins to various destinations has been the subject of intense research in recent years[1–4]. To-date, the 'molecular address' which enables proteins to be secreted out of the cell, integrate into organelles like mitochondria, chloroplasts and nucleus have been identified. Several other protein molecules which have a role in the protein 'export' process have also been identified. This review focuses attention on the chemical nature of the 'molecular address' or signal sequences that ensure proteins reach their correct destinations within the cell and the models that have been proposed to explain how the specific sorting of proteins occurs. Finally, some of the important questions that remain to be answered in the intracellular sorting of proteins are discussed.

CHEMICAL NATURE OF SIGNAL SEQUENCES

Proteins that are secreted out of eukaryotic cells and those destined for organelles like mitochondria and chloroplasts as well as proteins localized in compartments like outer membranes and periplasmic space in gram negative bacteria like *E. coli* are synthesized as precursors with amino terminal extensions of 20–40 residues[1–3]. These amino terminal precursor regions which are referred to as signal sequences are cleaved off from precursor proteins by peptidases present in the inner membrane in *E. coli*[5] and in the mitochondrial matrix[6] and endoplasmic reticulum[7] in eukaryotes. Discrete peptide sequences which specifically target proteins into the nucleus have been identified[8]. The primary structures of some representative 'signals' are shown in figure 1. An exhaustive list of such sequences has been compiled recently[9].

Requirements for signal sequences to be functional:

(a) Secreted protein signals: These sequences are generally comprised of 15–25 amino acids (figure 1). Although there is no primary structures homology amongst these sequences, they all have a characteristic distribution of amino acids which is illustrated schematically in figure 2. All these sequences have a positively charged basic amino terminal region which is followed by a long stretch of uncharged, mainly hydrophobic amino acids. The carboxy terminal amino acids have a short side chain and is invariably Gly, Ala, Ser or Cys.

Extensive genetic studies in gram-negative bacteria, particularly *E. coli*[10–12] indicate that the hydrophobic region is essential for signal sequences to be able to initiate export of proteins. Introduction of charged amino acids in this region causes the precursor forms of these proteins, to accumulate in the cytoplasm

Eukaryotic signal sequences (ref 9)

Met-Arg-Ser-Leu-Leu-Ile-Leu-Val-Leu-Cys-Phe-Leu-Pro-Leu-Ala-Ala-Leu-Gly (chicken lysozyme)

Met-Lys-Trp-Val-Thr-Phe-Leu-Leu-Leu-Leu-Phe-Ile-Ser-Gly-Ser-Ala-Phe-Ser (pro albumin)

Prokaryotic signal sequences (ref 9)

Met-Met-Ile-Thr-Leu-Arg-Lys-Leu-Pro-Leu-Ala-Val-Ala-Val-Ala-Ala-Gly-Val-Met-Ser-Ala-Gln-Ala-Met-Ala (*E.coli* λ - receptor)

Met-Lys-Ala-Thr-Lys-Leu-Val-Leu-Gly-Ala-Val-Ile-Leu-Gly-Ser-Thr-Leu-Leu-Ala-Gly (*E.coli* Lipoprotein)

Mitochondrial signal sequence (ref 9)

Met-Leu-Phe-Asn-Leu-Arg-Ile-Leu-Leu-Asn-Asn-Ala-Ala-Phe-Arg-Asn-Gly-His-Asn-Phe-Met-Val-Arg-Asn-Phe-Arg-Cys-Gly-Gln-Pro-Leu-Gln

Nuclear signal sequences

Ala-Ala-Phe-Glu-Asp-Leu-Arg-Val-Leu-Ser (Influenza virus nucleoprotein from Davey, J; Dimmock, N.J. and Colman, A. Cell, 1985, 40, 667)

Pro-Lys-Lys-Lys-Arg-Lys-Val-Glu-Asp-Pro (SV40 large T antigen, ref 8)

Figure 1. Primary structure of signal sequences.

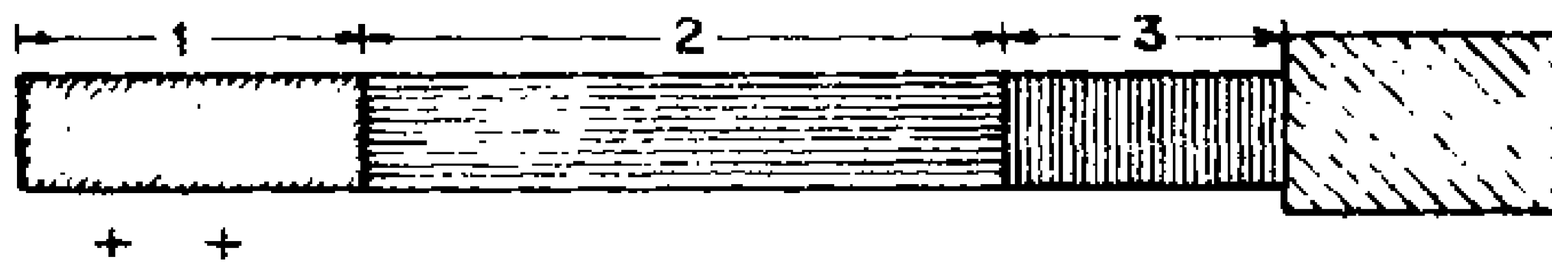

Figure 2. Distribution of amino acids in eukaryotic and prokaryotic signal sequences: *area 1* (////) has one or more basic amino acids like Lys and Arg, *area 2* (==) is comprized of almost exclusively of neutral hydrophobic amino acids like Leu, Ile, Phe, *area 3* (\\\\\) also consists of neutral, apolar amino acids, but of relatively small side chains like Ala, Gly, Ser, Cys.

indicating that for such mutants, export cannot be initiated. Genetic studies[13] have also indicated that the actual length of the hydrophobic region can vary from 12–20 amino acids. The importance of the hydrophobic region for a signal sequence to be able to initiate export has also been demonstrated in an eukaryotic system by *in vitro* reconstitution experiments[14]. Substitution of leu by β-hydroxy-leucine resulted in accumulation of precursor proteins.

(b) Mitochondrial signals: In mitochondria, proteins are localized in the matrix, inner membrane, inter membrane space and outer membrane. Most of these proteins are coded for, by the nucleus. Most of the matrix, inner membrane and inter membrane proteins are made as precursors whereas the outer membrane proteins are not[15]. The length of the precursor region varies between 20 and 40 residues. The number of mitochondrial proteins for which the primary structure is known is relatively few as compared to the signals of secretory proteins. As in the case of secretory protein signals, no primary structure homology is discernible.

In all cases, however, there is a strikingly overall basic composition and most of these sequences are devoid of acidic residues[15]. *In vitro* mutagenesis experiments[16] have indicated

that the central portion of the signal sequence is essential for import of the precursor protein to the mitochondria. The importance of certain Arg residues for function has also been demonstrated. However, such studies have not been extensive and clearly more data is necessary to attempt structure-function correlation.

(c) Nuclear signals: These signals have only been recently identified[8] and the primary structures of only few sequences are available. These sequences unlike those discussed in (a) and (b) are not found at the amino terminus of nucleus bound protein and are also not cleaved off the proteins after entry into the nucleus. The signals identified so far are highly basic (figure 1) and again there is no primary structure homology. With the limited amount of information available on these sequences, structure-function relationship is difficult.

DO REGIONS OTHER THAN THE SIGNAL SEQUENCE HAVE A ROLE IN THE SORTING OF PROTEINS?

One may ask the question whether signal sequences have all the information for sorting proteins to various destinations. Recent studies using recombinant DNA techniques and *in vitro* reconstitution systems suggest that signal sequences (mitochondrial and non-mitochondrial) may indeed have all the information for proper sorting[17-22]. Cytoplasmic proteins like globin have been found to translocate across the ER membranes when a signal sequence is present in a hybrid protein containing a signal sequence followed by α chain of globin[17]. Likewise cytosolic proteins have been transported to mitochondria[20-22], chloroplasts[18] and nucleus[19] with the aid of the appropriate signal sequences. In fact, it has been shown that the signal sequences of proteins destined for mitochondrial matrix and inner membrane can direct cytosolic proteins to these respective destinations when present in 'hybrid' proteins which are comprised of these sequences and a cytosolic protein[23]. However such a specificity is not observed with prokaryotic signal sequences i.e. the information to direct protein to either the outer membranes or periplasmic space does not seem to be present in the signal sequence.

OTHER COMPONENTS OF THE CELLS' EXPORT MACHINERY

While genetic techniques have yielded considerable information about the requirements of signal sequences to be functional in prokaryotes (which probably holds true for eukaryotic signals too), extensive biochemical studies, particularly *in vitro* reconstitution experiments in eukaryotes have revealed how segeration of ribosomes synthesizing secretory proteins at membrane sites i.e. the rough endoplasmic reticulum (ER) occurs. A protein-RNA complex, the signal recognition particle (SRP)[24], a receptor for SRP, fancifully named as the 'docking protein'[25] and ribosome binding proteins, the ribophorins[26], which aid this process have been isolated from microsomes. The manner in which these proteins are supposed to bring the ribosomes synthesizing secretory proteins to the endoplasmic reticulum membrane is illustrated schematically in figure 3. According to this scheme[27], SRP recognizes and binds to the signal sequence of the secretory protein that is being synthesized on ribosomes. This binding results in the arrest of translation of the secretory protein from its

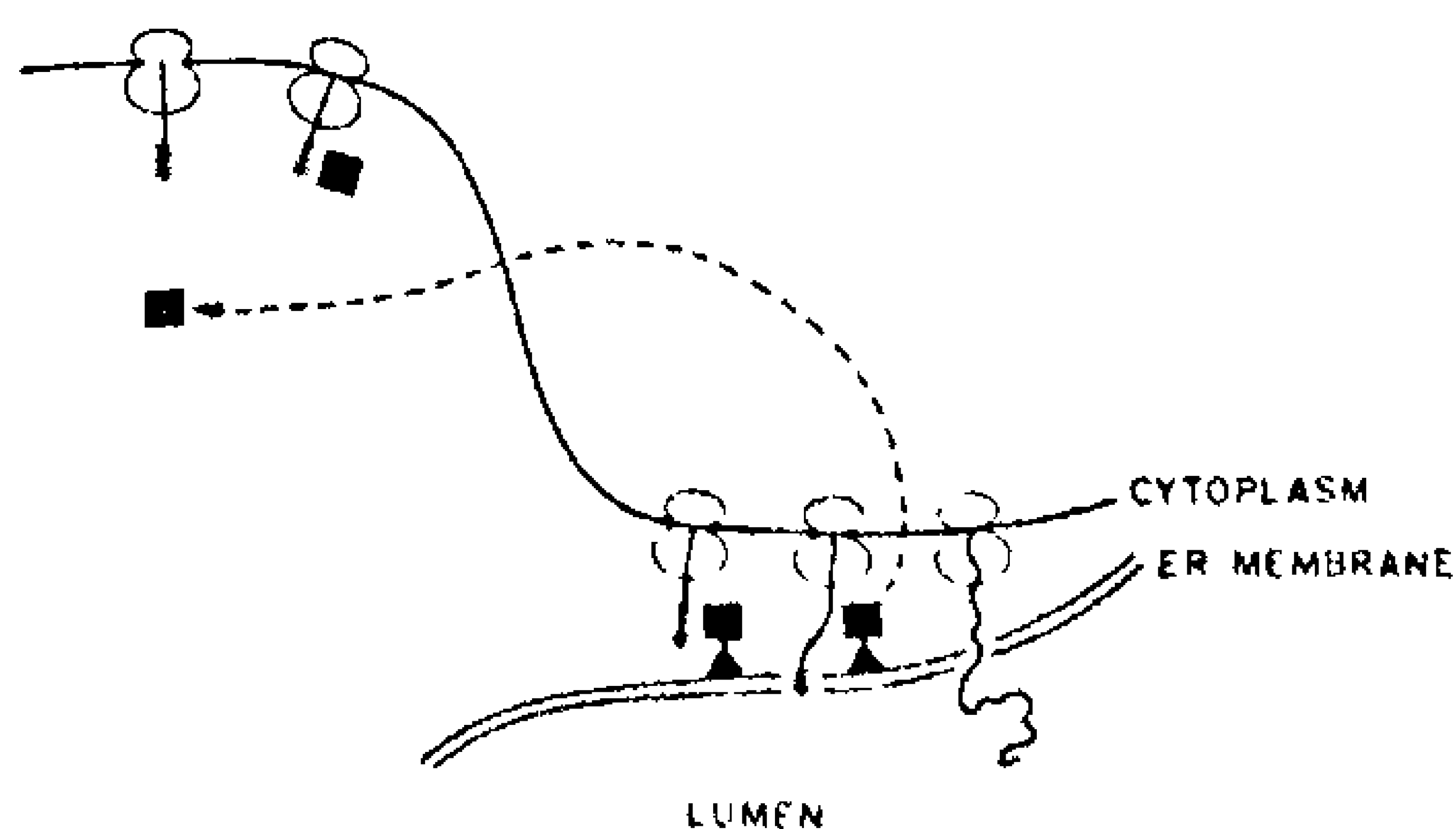

Figure 3. Schematic representation of the manner in which proteins destined for secretion are targeted to the endoplasmic reticulum in eukaryotes. ■, SRP; ▲, docking protein; ≈, Signal sequence.

mRNA. The SRP-nascent secretory protein-ribosome complex then move to the ER where binding of SRP to the docking protein results in the release of the translation arrest. The translocation arrest is observed only when SRP is present as the protein is getting synthesized. SRP does not bind to fully synthesized precursor proteins. Many intrinsic membrane proteins which do not have cleavable sequences, are also targeted to the ER as illustrated in figure 3.

The transfer of secreted proteins across the endoplasmic reticulum is initiated before the synthesis of the protein is complete i.e. the translocation is co-translational. The scheme illustrated in figure 3 is not operational if experiments are done with the fully synthesized precursor protein. Very recent studies indicate that for recognition of precursor proteins by SRP, the association of the protein-mRNA complex to ribosomes is essential[28]. In fact SRP recognizes almost completely synthesized precursor proteins if still bound to polysomes. While it is not clear whether regions other than the signal sequence are recognised by SRP in precursor proteins, the presence of the signal sequence is essential for SRP induced translation arrest. This fact has been demonstrated in the case of insulin by Wiedman *et al*[29]. Removal of the signal sequence of pre-proinsulin by recombinant DNA techniques abolished interaction between the polypeptide and SRP indicating the absolute requirement of signal sequence for recognition by SRP.

In prokaryotes, particularly in *E. coli*, several genetic loci that may specify components of the cellular export machinery have been identified[11]. However, the products of these genes have not been characterized. Unlike in eukaryotes, both co-translational and post-translational (i.e. after the synthesis of precursor protein is complete) modes translocation are observed in prokaryotes.

There is no unequivocal evidence for the participation of proteins like SRP and docking protein in the translocation of protein into mitochondria, chloroplasts and nucleus. Unlike secreted proteins, the mitochondria, chloroplast or nucleus bound proteins are assembled post-translationally.

The peptidases which remove signal sequences from the precursor proteins[5–7] have been identified in *E. coli*, in the endoplasmic reticulum membrane and in mitochondria. Eukaryotic precursor proteins are processed by *E. coli* signal peptidase[30] and vice versa[31]. Mitochondrial signal peptidase is specific only for mitochondrial precursors.

WHAT IS COMMON IN SIGNAL SEQUENCES THAT IS RECOGNIZED BY COMPONENTS OF THE CELL'S EXPORT MACHINERY?

Since primary structure homology is absent in signal sequences, it is evident that some other common structural feature (or features) is recognized by components of the cells export machinery. Analysis of secondary structure by the method of Chou and Fasman[32] also does not indicate the presence of any common secondary structural feature. However, experimental studies on the conformation of synthetic signal sequences having different primary structures indicate a high α-helix content particularly in hydrophobic environment[33–35]. Also genetic studies on the signal sequence of the lam B gene product, the λ-receptor, in *E. coli*[36] and circular dichroism studies[34] on synthetic peptides corresponding to wild type, mutant and pseudorevertant lam B signal sequences indicate the presence of α-helical conformation only in wild type and pseudorevertant sequences which can initiate export of λ-receptor. The mutant signal sequence which cannot initiate export of λ-receptor, does not have an α-helical conformation. Thus, helical conformation may be one of the recognition elements in signal sequences.

The overall hydrophobicity of signal sequences may be another recognition element since genetic studies and *in vitro* reconstitution experiments clearly indicate that reduction of the length of the hydrophobic stretch of signal sequences results in accumulation of the corresponding precursor proteins.

TRANSLOCATION OF PROTEINS ACROSS MEMBRANES

The manner in which ribosomes synthesizing proteins destined for secretion reach the endoplasmic reticulum at least in eukaryotes is reasonably clear (figure 3). Although molecules like SRP and docking protein have not been characterized in prokaryotes, it is likely that a mechanism similar to the one shown in figure 3 operates in prokaryotes also. Once this targeting to the membrane site is achieved, proteins have to cross a membrane barrier to reach their correct destination. Several models have been proposed to explain this important step of protein export.

Several workers have proposed models[39–42] for translocation of proteins across membranes which envisages direct partitioning of the protein into the lipid bilayer. The partitioning is initiated by interaction of the signal sequence with the lipid bilayer. In fact the free energy change on going from aqueous to hydrophobic environment for signal sequences strongly favour partitioning into the lipid bilayer[40]. The interaction with the lipid bilayer is facilitated by binding of the positively charged amino terminus of the signal sequence with the negatively charged membrane surface. Engelman and Steitz[42] have also tried to explain how integral membrane proteins can attain their topography in their helical hairpin hypothesis which is based on the spontaneous penetration of the lipid bilayer by peptide helices which assume a hairpin like configuration. Insertion of a helical hairpin into the bilayer will initiate secretion if the second helix is polar. Secretion of the newly synthesized protein would continue until or unless a hydrophobic segment is encountered.

Blobel and coworkers[27] have proposed a model which envisages translocation of proteins across membranes through aqueous channels formed by proteins present in the membrane rather than by perturbation of the lipid components of membranes.

Existing experimental evidence however does not unequivocally favour any of the proposed models. While it has been shown that signal sequences do interact with lipid vesicles[43], how exactly this interaction facilitates passage of polypeptide chain particularly those which are hydrophilic in nature, through the lipid bilayers remains to be established. Likewise membrane proteins that can form channels and allow the passage of nascent polypetide chains through them have not yet been identified.

ENERGETICS OF TRANSLOCATION

Translocation of proteins across membranes in *E. coli*[44, 45] as well as in mitochondria[15] requires energy. The form of energy required is the electrochemical potential and not ATP[44]. Studies in *E. coli* have indicated that both components of the electrochemical potential i.e. the electrical component and the proton gradient (Δ pH) are required for translocation of proteins across membranes[45]. In the case of mitochondria[15], an energized membrane is required only for the assembly of inner membrane and matrix proteins and not outer membrane proteins. Unlike in *E. coli*, it is not clear if both components of the electrochemical potential are important for import of proteins in mitochondria. It is clear that the transmembrane potential does not serve as a simple electrophoretic gradient in *E. coli* and mitochondria. The energy requirements for translocation of proteins across the endoplasmic reticulum membrane is still not clear.

SUMMARY AND CONCULSIONS

Intense research over the past 15 years or so has revealed the chemical nature of the 'molecular address' that ensure correct sorting of proteins and also other components of the cells that participate in the sorting process. However, some important questions still remain to be answered satisfactorily. They are as follows:

(i) How do proteins actually cross membranes? We have shown[46] that synthetic signal sequ

ences render lipid vesicles permeable to cations and dyes and also cause their fusion. However, it remains to be demonstrated that fairly hydrophilic peptide chain can penetrate the lipid bilayer. Models for protein translocation which favour direct partitioning of protein into the lipid bilayer are somewhat simplistic and may not represent the actual state of affairs. Although alternate models in which protein translocation is proposed to occur through aqueous channels formed by proteins in the ER have been proposed, no such proteins have been identified or characterized.

(ii) How are mitochondria bound proteins targeted correctly?

Mitochondrial proteins coded by the nucleus are assembled post-translationally. The differences in mitochondrial and secretory protein signal sequences presumably helps in preventing the mitochondrial precursors from being targeted to ER. How these proteins selectively reach the mitochondria correctly is still not clear. It is likely that molecules similar to SRP and docking protein may be involved. These molecules, however, are yet to be identified.

(iii) After crossing the membrane barrier, how are proteins selectively sorted out to organelles like lysosomes, peroxisomes or various regions of the plasma membrane?

(vi) How is the energy requirement for proteins translocation across membranes met?

(v) What are the hidden signals in proteins that are not processed but nevertheless cross a membrane barrier like ovalbumin.

Although satisfactory answers to these questions have not been obtained as of now, they should be forthcoming in years to come.

2 August 1986; Revised 17 September 1986

1. Kreil, G., *Annu. Rev. Biochem.*, 1981, **50**, 317.
2. Waksman, A., Hubert, P., Cremel, G., Rendon, A. and Claude, B., *Biochim. Biophys. Acta*, 1980, **604**, 249.
3. Silhavy, T. J., Benson, S. A. and Emr, S. D., *Microbiol. Rev.*, 1983, **47**, 313.
4. Wickner, W. and Lodish, H. F., *Science*, 1986, **230**, 400.
5. Zwizinski, C. and Wickner, W., *J. Biol. Chem.*, 1980, **255**, 7973.
6. Bohni, P., Gasser, S., Leaver, C. and Schatz, G., In: *The organization and expression of the mitochondrial genome*, (eds) A. M. Kroon and C. Saccone, Elsevier, Amsterdam, 1980, p. 423.
7. Evans, E. A., Gilmore, R. and Blobel, G., *Proc. Natl. Acad. Sci. USA*, 1986, **83**, 581.
8. Dingwall, C., *Trends Biochem. Sci.*, 1985, **10**, 64.
9. Watson, M. E. E., *Nucl. Acid Res.*, 1984, **12**, 5145.
10. Hall, M. N. and Silhavy, T., *J. Annu. Rev. Genet.*, 1981, **15**, 91.
11. Michaelis, S. and Beckwith, J., *Annu. Rev. Microbiol.*, 1982, **36**, 435.
12. Bankaitis, V. A., Ryan, P. J., Rasmussen, B. A. and Bassford, P. J. Jr., *Current topics in membranes and transport*, 1985, **24**, 105.
13. Ryan, J. P., Duncan, M. C., Bankaitis, V. A. and Bassford, P. J. Jr., *J. Biol. Chem.*, 1986, **261**, 3389.
14. Hortin, G. and Boime, I., *Proc. Natl. Acad. Sci. USA*, 1980, **77**, 1356.
15. Hay, R., Bohni, P. and Gasser, S., *Biochem. Biophys. Acta*, 1984, **779**, 65.
16. Horwich, A. L., Kalousek, F., Fenton, W. A., Pollock, R. A. and Rosenberg, L. E., *Cell*, 1986, **44**, 451.
17. Lingappa, V. R., Chaidez, J., Yost, C. S. and, Hedgpeth, J., *Proc. Natl. Acad. Sci. USA*, 1984, **81**, 456.
18. Schreier, P. H., Sefton, E., Schell, J. and Bohnert, H. J., *EMBO*, 1985, **4**, 25.
19. Hall, M. N., Hereford, L. and Herskowitz., I., *Cell*, 1984, **36**, 1057.
20. Hurt, E. C., Pesold-Hurt, B., Suda, K., Oppliger, W. and Schatz, G., *EMBO*, 1985, **4**, 2061.
21. Horwich, A. L., Kalousek, F., Mellman, I. and Rosenberg, L., *EMBO*, 1985, **4**, 1129.
22. Emr, S. D., Vassarotti, Garrett, J., Geller, B. L., Takeda, M. and Douglas, M. G., *J. Cell Biol.*, 1986, **102**, 523.
23. Van Loon, A. P. E. M., Brandl, A. W. and Schatz, G., *Cell*, 1986, **44**, 801.
24. Walter, P., Ibrahimi, I. and Blobel, G., *J. Cell Biol.*, 1981, **91**, 545.
25. Meyer, D. I., Krause, E. and Dobberstein, B., *Nature (London)*, 1982, **297**, 647.

26. Kreibich, G., Ulrich, B. L. and Sabatini, D. D., *J. Cell Biol.*, 1978, **77**, 464.
27. Walter, P., Gilmore, R. and Blobel, G., *Cell*, 1984, **38**, 5.
28. Perara, G., Rothman, R. E. and Lingappa, V. R., *Science*, 1986, **232**, 348.
29. Weidmann, M., Huth, A. and Rapoport, T. A., *Biochim. Biophys. Res. Commun.*, 1986, **134**, 790.
30. Talmadge, K., Brosius, J. and Gilbert, W., *Nature* (*London*), 1981, **294**, 176.
31. Muller, M., Ibrahimi, I., Chang, C. N., Walter, P. and Blobel, G., *J. Biol. Chem.*, 1982, **257**, 11860.
32. Austen, B. M., *FEBS Lett.*, 1979, **103**, 308.
33. Rosenblatt, M., Beaudette, N. V. and Fasman, G. D., *Proc. Natl. Acad. Sci. USA*, 1980, **77**, 3983.
34. Briggs, S. M. and Gierasch, L. M., *Biochemistry*, 1984, **23**, 3111.
35. Katakai, R. and Iizuka, Y., *J. Am. Chem. Soc.*, 1984, **106**, 5715.
36. Shinnar, A. E. and Kaiser, E. T., *J. Am. Chem Soc.*, 1984, **106**, 5006.
37. Laxma Reddy, G. and Nagaraj, R., *Biochim. Biophys. Acta*, 1985, **831**, 340.
38. Emr, S. D. and Silhavy, T. J., *Proc. Natl. Acad. Sci. USA*, 1983, **80**, 4579.
39. Inouye, M. and Halegoua, S., *Crit. Rev. Biochem.*, 1979, **7**, 339.
40. Von Heijne, G., *Eur. J. Biochem.*, 1981, **116**, 419.
41. Wickner, W., *Annu. Rev. Biochem.*, 1979, **48**, 23.
42. Engelman, D. M. and Steitz, T. A., *Cell*, 1981, **23**, 411.
43. Nagaraj, R., *FEBS Lett.*, 1984, **165**, 79.
44. Bakker, E. P. and Randall, L. L., *EMBO*, 1948, **3**, 895.
45. Gasser, S. M., Daum, G. and Schatz, G., *J. Biol. Chem.*, 1982, **257**, 13034.
46. Joseph, M., Laxma Reddy, G. and Nagaraj, R., *Biochim. Biophys. Acta*, 1986 (Communicated).