BIOCHEMISTRY AND MOLECULAR BIOLOGY EDUCATION
Vol. 33, No. 3, pp. 188–193, 2005

Articles

Concepts of Protein Sorting or Targeting Signals and Membrane Topology in Undergraduate Teaching*

Received for publication, June 25, 2004, and in revised form, December 3, 2004

Bor Luen Tang‡ and Felicia Yu Hsuan Teng

From the Department of Biochemistry, National University of Singapore, Singapore 117597, Republic of Singapore

The process of protein biogenesis culminates in its correct targeting to specific subcellular locations where it serves a function. Contemporary molecular and cell biology investigations often involve the exogenous expression of epitope- or fluorescent protein-tagged recombinant molecules as well as subsequent analysis of protein-protein interactions *in vitro* and *in vivo*. Fundamental knowledge of targeting signals that direct a polypeptide to various organelles or membrane domains is essential for the proper design of such recombinant molecules. A fundamental concept of membrane compartmentalization is also often useful for the interpretation of the preliminary results of interaction screens. Knowledge in targeting signals and post-translational dynamics of proteins should therefore be given sufficient emphasis in an undergraduate biochemistry or molecular biology curriculum. Such knowledge is essential, particularly for undergraduates or fresh graduates embarking on research projects in a cell and molecular biology laboratory.

Keywords: Epitope-tagging, organelles, secretory pathway, targeting signals.

Making expression constructs for the exogenous or ectopic expression of genes in mammalian cells is a commonplace activity in any contemporary molecular or cell biology laboratory [1]. Very often, such expression constructs may entail the appendage of markers or tags to a polypeptide to aid detection where primary antibodies are not available, or they may function to differentiate the exogenously expressed protein from the endogenous pool. Exogenous expressions of tagged gene products of choice in cells are useful in many ways. One common use is to demonstrate or confirm possible protein-protein interactions (either direct interaction or indirect interact within multicomponent complexes) via co-immunoprecipitation analysis.

Tasks of construct design and making may be assigned to an undergraduate research student or a fresh graduate student as an exercise, serving as a prelude to more complicated genetic manipulations or simply as a starting point of an independent project. The basic DNA manipulation techniques involved (such as the polymerase chain reaction, DNA restriction, ligation, transformation) are often well covered in laboratory courses. The availability of commercial reagent kits has also greatly simplified the actual bench work involved. However, rational design of expression constructs requires more than an ability to adhere to basic DNA manipulation protocols. Many polypeptides carry within their primary sequence signals that direct their post-translational targeting to a particular organelle or compartment of the cell. Some of the most common signals for mammalian cells are summarized in Fig. 1. Any disruption of these targeting signals during genetic manipulation may result in mis-targeting of the resultant recombinant polypeptide with consequential loss of its original functions or worse, the gain of spurious, undesirable activity that would complicate analysis.

FRONT OR BACK, IN OR OUT?

Epitope-tagging procedures nowadays enjoy a great variety of lavishly designed (in terms of multiple cloning sites) and commercially available expression vectors with epitope tags such as myc, hemagglutinin, FLAG, and others. Users are also spoiled for choices when it comes to the many different fluorescent protein tags now available in the market [2]. How does one decide where to put the tag? The most common position for tagging is to append it at either the N terminus or the C terminus of one's gene of interest, taking care that the reading frame is maintained at the joining site. The obvious reason for introducing a tag at the termini is the relative ease of DNA manipulations involved (compared with intrasequence tagging) and the inherent (but sometimes incorrect) assumption that the introduction of foreign sequences at the termini would probably be the least disruptive to the normal folding of the polypeptide.

The consequence of irrational N-terminal tagging or appendage of a fluorescent protein marker is the risk of possible disruption of signal sequences found at the N terminus of polypeptides. For mammalian cells, the most

‡ To whom correspondence should be addressed: Dept. of Biochemistry, National University of Singapore, 8 Medical Dr., Singapore 117597, Republic of Singapore. Tel.: 65-6874-1040; Fax: 65-6779-1453; E-mail: bchtbl@nus.edu.sg.

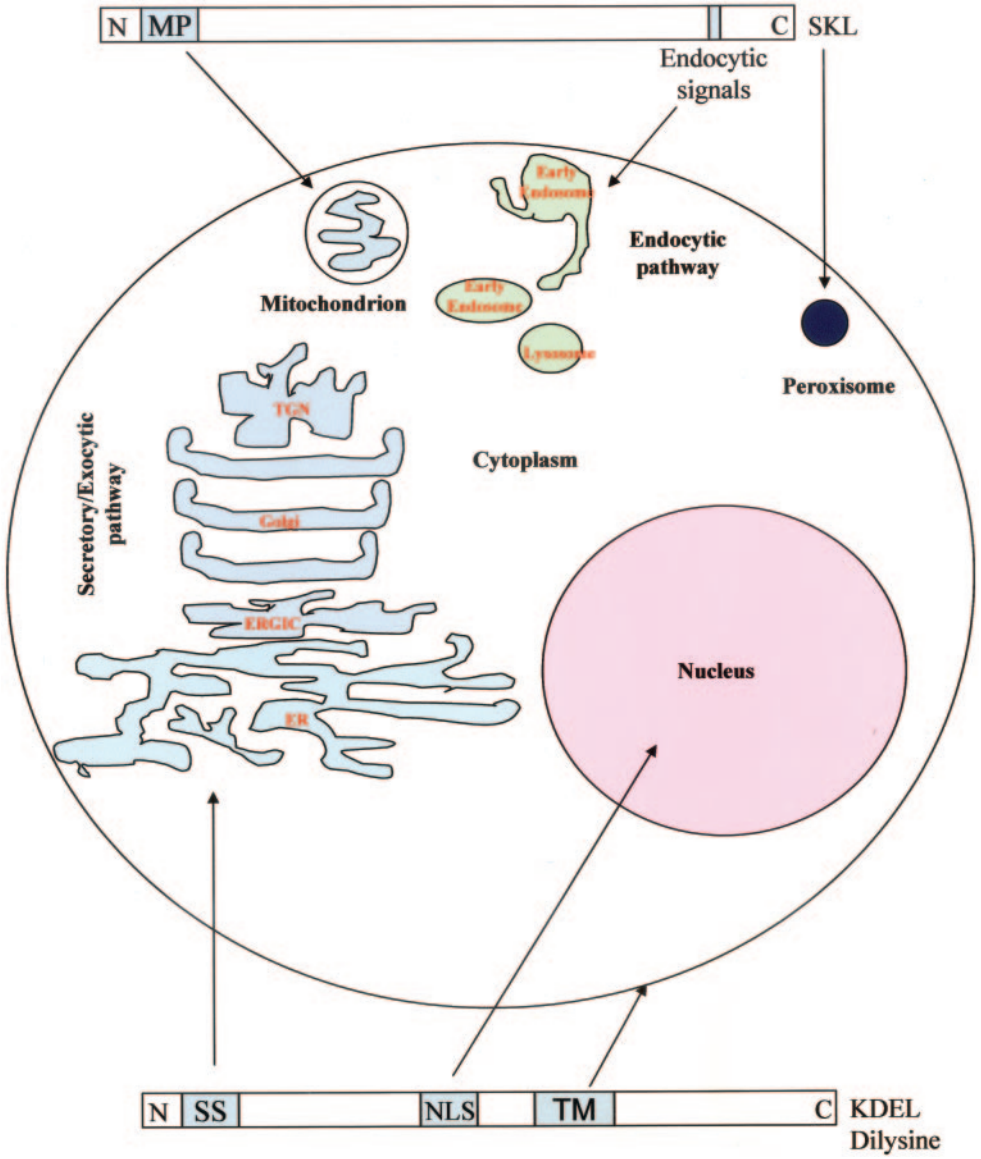

FIG. 1. **Common targeting signals to different subcellular locations in a mammalian cell.** Depicted are major subcellular compartments where a polypeptide, upon emerging from a ribosome in the cytoplasm, can potentially be targeted to. Entry into the secretory (exocytic) pathway usually requires the presence of a functional ER-targeting signal peptide (*SS*) located at the N terminus (*N*) of soluble or membrane-spanning proteins. Soluble proteins would eventually be secreted, and those with transmembrane domains (*TM*) will be anchored on the plasma membrane. In some cases, hydrophobic internal topogenic sequences can serve the function of the signal sequence in ER targeting and translocation as well as membrane anchorage. A technical discussion of the ER translocation mechanism is beyond the scope of this article. Its research, however, involves classical biochemical and molecular biology approaches that are pedagogically useful for undergraduate teaching (see for example Ref. 11). ER resident proteins may contain a KDEL or dilysine motif at the C terminus. Polypeptides not targeted to the secretory pathway remain in the cytoplasm unless they are imported into the nucleus via NLS or targeted to the peroxisome. A large majority of that occurs through the type I peroxisomal targeting signal (*SKL*) found at the C terminus (*C*). Mitochondrial proteins can be targeted into the mitochondria via a mitochondrial-targeting presequence (*MP*) or via internal sequences not depicted here. Some membrane spanning proteins, particular receptors of various growth factors, contain cytoplasmic endocytic signals that facilitate the process of receptor-mediated endocytosis triggered by ligand binding.

common ones are the endoplasmic reticulum (ER)[1] targeting and translocation signals and mitochondrial targeting signals. The former takes the form of a stretch of hydrophobic amino acids flanked by charged residues, whereas the latter is an amphipathic helix with alternating hydrophobic and charge or polar residues. Other less frequently encountered signals include the diarginine signal for ER retention or retrieval. Most mammalian polypeptides destined for the secretory (or exocytic) pathway are cotranslationally targeted and translocated across the ER membrane, a process mediated by the recognition of an ER-targeting signal by the signal recognition particle [3]. Disruption of ER targeting and translocation simply means that these secreted proteins (extracellular matrix components, hormones, growth factors, etc.) and cell surface

[1] The abbreviations used are: ER, endoplasmic reticulum; SKL, serine-lysine-leucine; KDEL, lysine-aspartate-glutamate-leucine; GPI, glycosylphosphatidylinositol; NLS, nuclear localization signal; EGFP, enhanced green fluorescent protein; AP, alkaline phosphatase; DPPIV, dipeptidyl peptidase IV.

proteins spanning the plasma membrane (receptors, adhesion molecules, etc.) would not be directed into the secretory (or exocytic) pathway. As such, these polypeptide are likely to remain (if translation is complete) in the cytosol, and chances are these would be nonfunctional. A worse scenario is that the accumulation of these polypeptides in the cytosol is somewhat toxic to certain key metabolic processes, resulting in impaired cell health or viability. Any observations made under the circumstances could be particularly misleading if one happens to be doing cell death-related studies.

Attaching a C-terminal tag, on the other hand, can also potentially disrupt several types of protein sorting or targeting signals. The type 1 peroxisomal targeting signal is the tripeptide sequence serine-lysine-leucine (SKL) located at the C terminus [4]. Two common ER retrieval signals, the dilysine motif and the tetrapeptide lysine-aspartate-glutamate-leucine (KDEL) motif, are located at the C terminus of many ER resident proteins. Another way whereby a C-terminal tag or marker could disrupt proper protein targeting is the disruption of signals for the incorporation of lipid anchors. For example, many members of the Ras superfamily carry sequences that signal the attachment of lipid anchors at their C termini [5]. A class of plasma membrane proteins, some of which are splice isoforms of cell adhesion molecules or receptors, have a glycosylphosphatidylinositol (GPI) linker [6]. The molecular signals engaging the lipid modification enzyme complexes reside at the C terminus of the above proteins and would almost certainly be disrupted by the addition of any sequence with a substantial length that could serve as an epitope tag.

IS AN *IN VITRO* INTERACTION MEANINGFUL IN VIEW OF MEMBRANE COMPARTMENTALIZATION *IN VIVO*?

A type of preliminary research data that could lead to meaningless pursuits should a fundamental grasp of the concepts of protein targeting or sorting be lacking is the results of protein-protein interaction screens. One type of project often assigned to junior research students is a massive, brute force interaction screen for proteins that may interact with a particular protein of interest (the “bait”). Classical screening methods, such as affinity purification with matrix-coupled recombinant proteins or antibodies, the yeast two-hybrid screen, and the more advanced protein-based chips, could potentially yield a large number of identities, including false-positives. At the end of a screen, one is often left bewildered by which ones to pursue further. Some of the false-positives may seem superficially interesting, especially when these are molecules that are also being intensely studied by others (the “hot” proteins). The need for further investigation with some of these, however, could be safely ruled out based on the logic that some of these proteins are unlikely to ever “see” the bait in an intact cell. For example, the immunoglobulin-like domains of some cell adhesion molecules may appear to interact *in vitro* (or in the case of the yeast two-hybrid screen, in the yeast nuclei) with cytoplasmic proteins such as members of the Ras superfamily GTPases or cytoplasmic signaling adaptors. If one considers the fact that these immunoglobulin-like domains are essentially lumenal all the time until it becomes oriented extracellularly, it is clear that they could never come into contact with cytoplasmic proteins (or for that matter, soluble nuclear proteins) in an intact cell.

DECIDING WHERE OR HOW TO TAG

Unless the polypeptide one wishes to modify has been thoroughly studied by others, and all targeting signals are known, the first thing one should do is to put the sequence through some targeting or sorting signal analysis software. A particularly useful and comprehensive analysis tool is the PSORTII program hosted at the University of Tokyo site (psort.ims.u-tokyo.ac.jp). One should watch out for potential signals at the N or C terminus and avoid putting a tag at these sites.

What if there are targeting or sorting signals at both the N and C terminus and the disruption of either would result in mis-targeting? A case in question is GPI-linked proteins. These are usually plasma membrane proteins carrying a cleavable ER-targeting signal sequence at the N terminus and a GPI-modification signal at the C terminus. Loss of the ER-targeting signal would exclude it from the secretory pathway and prevent its appearance on the plasma membrane. Loss of the GPI modification is likely to result in its secretion and not anchorage. In such a case, more complicated manipulation in order to introduce an intrasequence tag is unavoidable. One possibility is to engineer a tag immediately after the signal sequence. The tag would essentially emerge as a neo-N terminus after the signal sequence is cleaved.

There are no hard and fast rules to intrasequence tagging. One should however pay attention to two aspects. The first is to avoid disrupting targeting signals that are located away from the N or C termini. These include nuclear localization signals (either the classical single-site basic signal or the bipartite signal), signals for endocytosis (dileucine motifs, tyrosine-based endocytic signal, etc.), and membrane-spanning domains. The second point to note is to avoid the interruption of potential functional domains or regulatory sites (such as potential phosphorylation sites) or the disruption of proper folding. A PROSITE search with, for example, the ExPASy server (tw.expasy.org/prosite/) would reveal commonly known domains and post-translational modification sites within the primary sequence of one’s protein of interest.

SELECTIVE TARGETING TO SPECIFIC SUBCELLULAR LOCATIONS

Knowledge of protein sorting and targeting would also help students to design ways to deliberately and selectively enhanced (mis-)targeting of a protein to (or away from) a compartment of interest in order to investigate the physiological consequence of such a manipulation. One such kind of investigations is to enhance the targeting of some proteins to or away from the mitochondria to assess their pro- or anti-apoptotic potential. Other examples include the selective and constitutive targeting of signaling proteins to the plasma membrane or the endosome to assess their role in signaling, the selective exclusion of transcription factors from the nucleus to understand their function in modulating gene expression patterns, or the

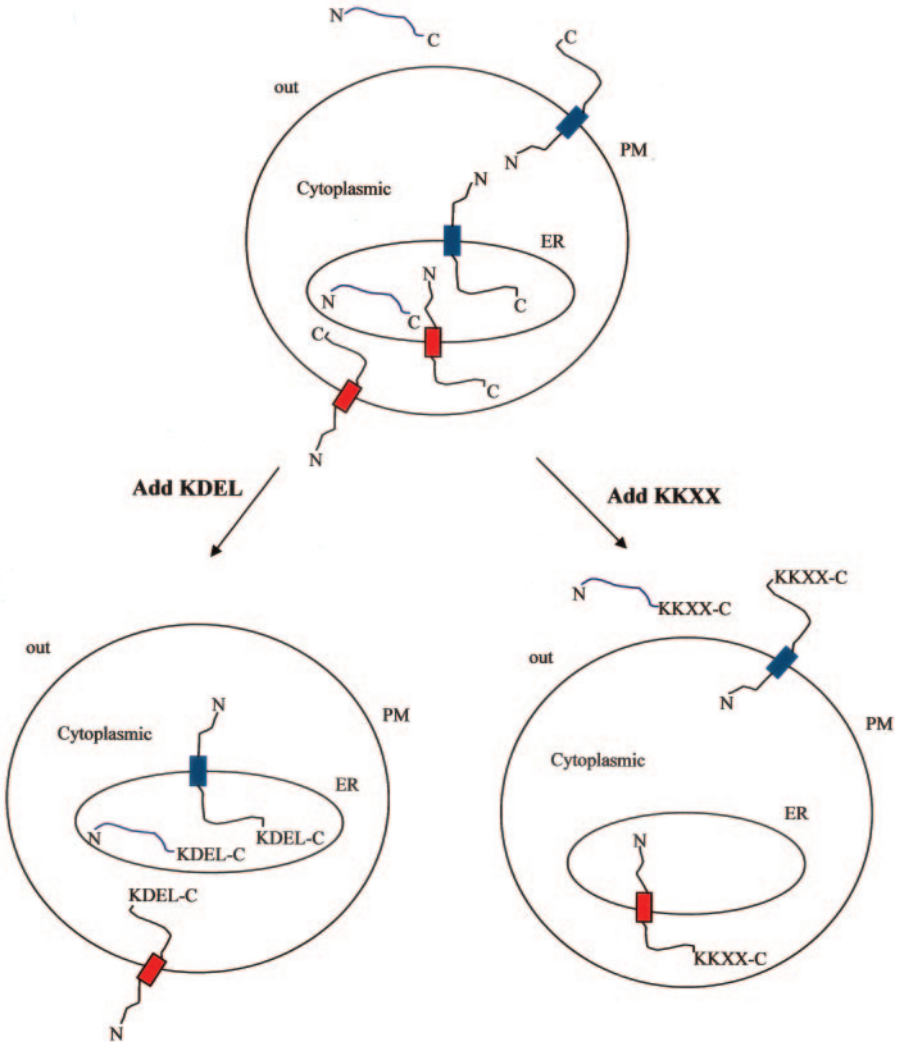

FIG. 2. **A thought experiment illustrating the concepts of membrane topology.** A diagramatic illustration depicting three polypeptides: a lumenal/secretory protein (*blue line*), a type I membrane protein (*black line*, *red* transmembrane domain), and a type II membrane protein (*black line*, *green* transmembrane domain) is shown. The *upper panel* depicts all the possible topological orientations of the three polypeptides, whether they reside inside the cell (*cytoplasmic*) or outside (*out*) or are anchored at the ER membrane or the plasma membrane (*PM*). Students are asked to predict the consequence of artificially appending either the KDEL or the dilysine (KKXX, K-lysine, X-any amino acid) signal to the C terminus of these polypeptides, with the assumption that all three of these would be targeted to and traverse the secretory pathway, ending up at the plasma membrane or being secreted without any intervention. The correct outcome is depicted in the *lower panel* (see text for explanation).

creation of soluble or secreted forms of cell surface molecules to assess their possible autocrine or paracrine functions.

TEACHING THE CONCEPTS AT THE UNDERGRADUATE LEVEL

The concept of protein sorting and targeting as well as membrane compartmentalization will become practically useful and important the moment an undergraduate engages in laboratory projects. It should therefore receive adequate emphasis in the junior undergraduate syllabus, perhaps as part of a cell biology or cellular biochemistry module. Below is a brief description of the author's own experience in teaching such concepts.

Protein targeting signals are introduced in a second-year undergraduate life sciences module entitled "Cell Biology." The module itself is rather extensive, as it covers topics ranging from cell cycle, cell signaling, and cell death/senescence to organelle structures and biogenesis. It is in the latter topic where protein targeting to various organelles is described. Students are first introduced to the various subcellular compartments and various modes of post-translational trafficking dynamics. There is much jargon and many facts in cell biology that, if presented wholesale, could deter even the keenest of students. Most pedagogical emphasis in the teaching of cell biology at a junior or sophomore level is therefore based on the grasp of concepts rather than plain facts. To avoid the overloading of hard facts in our case and to enhance the learning of such concepts, real-life examples of proteins with their respective targeting signals are kept to a minimum. Instead, thought experiments and problem-based learning using hypothetical examples are engaged.

Fig. 2 illustrates one example of a thought experiment alluded to above. The concept emphasized here is the concept of membrane topology and that targeting signals function in specific subcellular and topological contexts. The signals in question here are the dilysine motif and the KDEL signal. Both are ER retrieval signals functioning at the C terminus of a polypeptide, but they are found on different types of molecules. The dilysine motif is typically found in the cytoplasmic part of type I membrane proteins of the ER, whereas the KDEL signals are found in the lumenal proteins of the ER. The former is recognized by the Golgi-ER retrograde transport machinery involving COPI [7], while the latter is recognized by the KDEL receptor [8, 9]. The point to note here is that the dilysine motif would be nonfunctional if it is present (or artificially appended) at the lumenal portion of a molecule. Conversely, the KDEL signal would be nonfunctional if it is cytoplasmically oriented. A simple diagramatic illustration depicting three polypeptide topologies that are most often en-

countered (a lumenal/secretory protein, a type I membrane protein (N terminus cytoplasmic, C terminus lumenal/extracellular), and a type II membrane protein (N terminus lumenal/extracellular, C terminus cytoplasmic)) is shown. Students are asked to predict the consequence of artificially appending either the KDEL or the dilysine (KKXX, K-lysine, X-any amino acid) signal to the C terminus of these polypeptides with the assumption that all three of these would be targeted to and traverse the secretory pathway, ending up at the plasma membrane or secreted without any intervention. With logical reasoning, students would correctly predict that the dilysine KKXX signal would be effective in conferring ER retention for only the type I membrane protein, the C terminus of which is cytoplasmically oriented. They would remember that they were told earlier that the KDEL signal is the ER retrieval signal for ER lumenal proteins. They may surmise that the signal is also likely to confer ER retention for the type II membrane protein (because its C terminus is lumenally oriented) but not the type I protein. At this juncture, they will be told that the experiment was actually done (by one of the authors sometime ago), with a reference to the primary literature [10].

The following is a typical problem-based question given for tutorials.

You are given five polypeptides A, B, C, D, and E and the following information:

A is generally hydrophilic, contains a cleavable signal peptide at the N terminus, and a dilysine motif-like KKXX sequence at the C terminus;

B is generally hydrophilic and contains a classical monopartite nuclear localization signal (NLS);

C is a lysosomal hydrolase;

D is the type II surface membrane protein dipeptidyl peptidase IV;

E contains a tripeptide sequence of SKL at the C terminus.

Question: What is the most probable steady-state subcellular localization of polypeptides A to E in interphase cells before and after a tetrapeptide KDEL sequence is appended to each at the C terminus?

Answer: A, secreted/ER lumen; B, nuclear/nuclear; C, lysosome/ER lumen; D, plasma membrane/ER membrane; E, peroxisome/cytoplasmic.

For A, the student should surmise that the dilysine signal (KKXX) would be nonfunctional (A, being generally hydrophilic, is likely to be lumenal) and would be superseded by a KDEL signal subsequently appended to its C terminus. Correct answers for B, C, and E require the student to be clear about the concept that the polypeptides must first enter the secretory pathway before the KDEL sequence would have any functional significance. Both B (likely to be nuclear) and E (likely to be peroxisomal) are without an ER-targeting signal sequence and therefore could not enter the secretory pathway in the first place. These would not reside in the ER lumen regardless of the other signals they carry. In fact, the basis for their nuclear and peroxisomal targeting is that they remained in the cytosol upon completion of synthesis to allow the nuclear localization signal (NLS) (for B) or the SKL (for E) to be recognized by the respective nuclear and peroxisomal import receptors, a feat not possible if they have a signal sequence and are cotranslationally translocated into the ER. The brighter students could probably also surmise from the ensuing discussion that a polypeptide with both an ER-targeting signal peptide and a NLS is likely to end up in the secretory pathway and not the nucleus. Incidentally the C-terminal context of SKL will be disrupted by the appendage of KDEL, together with E's peroxisomal targeting.

Student feedback pertaining to this topic is typically mixed. While a majority of students in the class is able to comprehend the concepts after working through the thought experiments and the problems, there remained a few who are either unable to follow the logic or find the topic rather confounding. It is not yet clear if they are inherently weak in logical reasoning or have simply chosen to ignore the topic, as they may not see its relevance to other aspects of their training in the life sciences. It is hoped that those who eventually embark on laboratory research projects would find the knowledge useful.

EXTENSION OF CONCEPT TEACHING TO A CELL BIOLOGY LABORATORY EXERCISE

The concepts described above can be readily incorporated into a concurrent laboratory-based exercise or one at a more advanced level, as described below. The necessary starting materials include marker cDNAs whose products are either soluble or membrane proteins. Examples of useful markers include enhanced green fluorescent protein (EGFP: soluble, cytoplasmic), alkaline phosphatase (AP: soluble, secreted), and a plasma membrane type II membrane spanning protein such as dipeptidyl peptidase IV (DPPIV). Various mammalian expression constructs for EGFP tagging are commercially available from companies such as Invitrogen and Clontech (Palo Alto, CA). The *pShooter* series from Invitrogen could be useful as they come with vector-borne mitochondrial presequence and NLS, although these could be generated fairly easily by the polymerase chain reaction. A cell line (and therefore cell culture facilities) that spreads out nicely when cultured on treated glass coverslips and is easily transfected with standard commercially available lipid-based transfection reagents (for example, Lipofectamine or Effectene), such as Chinese hamster ovary cells, would be needed as the host to illustrate subcellular protein localization. A further requirement would be a fluorescent microscope, preferably equipped with some form of image-capturing device (either a film-based camera or a charge-coupled device digital camera).

Depending on the available time and resources, various experiments could be planned. At the very minimum, students can be led through the relatively simple processes of mammalian cell transfection and sample processing (fixation, antibody staining, mounting) for immunofluorescence microscopy. Students could and should familiarize themselves with the staining patterns representing the different compartments of the cell using the basic constructs. These constructs can be made in advance either by teaching or laboratory assistants or by the students themselves. From the above-listed materials, these may include EGFP (largely cytoplasm), EGFP-NLS (nuclear), EGFP-mitochondrial presequence (mitochondrial), EGFP-SKL (peroxiso-

mal), and DPPIV probed with its antibody (plasma membrane). EGFP-KDEL should show the staining of ER. Individual and double labeling with endogenous markers of the various compartments could be done to verify the staining of the exogenously expressed proteins.

More elaborate modifications and manipulations can include tagging KDEL to the C terminus of DPPIV and observing the change of its steady-state localization (from plasma membrane to ER). Students can see for themselves that a dilysine motif similarly attached will not confer ER localization to DPPIV. Another interesting experiment would be to ask what the steady-state localization of EGFP may be if it is simultaneous appended with a ER-targeting signal sequence and a NLS. As mentioned above, because of the nature of co-translational insertion associated with ER targeting, the NLS is unlikely to be functional.

A nonimaging-based demonstration of the function of targeting signals could center on the marker AP, whose enzyme activity is easily assay by standard ELISA-based protocols. Students could compare the effect of, for example, the disruption of the signal sequence or the appendage of KDEL, dilysine motif, or NLS to AP by assaying for the amount of AP secreted into the culture medium.

REFERENCES

[1] J. W. Jarvik, C. A. Telmer (1998) Epitope tagging, *Annu. Rev. Genet.* **32**, 601–618.

[2] M. V. Matz, K. A. Lukyanov, S. A. Lukyanov (2002) Family of the green fluorescent protein: Journey to the end of the rainbow, *Bioessays* **24**, 953–959.

[3] E. Swanton, N. J. Bulleid (2003) Protein folding and translocation across the endoplasmic reticulum membrane, *Mol. Membr. Biol.* **20**, 99–104.

[4] P. E. Purdue, P. B. Lazarow (2001) Peroxisome biogenesis, *Annu. Rev. Cell Dev. Biol.* **17**, 701–752.

[5] F. L. Zhang, P. J. Casey (1996) Protein prenylation: Molecular mechanisms and functional consequences, *Annu. Rev. Biochem.* **65**, 241–269.

[6] S. Mayor, H. Riezman (2004) Sorting GPI-anchored proteins, *Nat. Rev. Mol. Cell. Biol.* **5**, 110–120.

[7] C. Barlowe (2000) Traffic COPs of the early secretory pathway, *Traffic* **1**, 371–377.

[8] A. Murshid, J. F. Presley (2004) ER-to-Golgi transport and cytoskeletal interactions in animal cells, *Cell Mol. Life Sci.* **61**, 133–145.

[9] K. Yamamoto, R. Fujii, Y. Toyofuku, T. Saito, H. Koseki, V. W. Hsu, T. Aoe (2001) The KDEL receptor mediates a retrieval mechanism that contributes to quality control at the endoplasmic reticulum, *EMBO J.* **20**, 3082–3091.

[10] B. L. Tang, S. H. Wong, S. H. Low, W. Hong (1992) Retention of a type II surface membrane protein in the endoplasmic reticulum by the Lys-Asp-Glu-Leu sequence, *J. Biol. Chem.* **267**, 7072–7076.

[11] J. Szeberenyi (2004) Protein transport into the rough endoplasmic reticulum, *Biochem. Educ.* **32**, 121–122.