Genetic approaches to the study of protein targeting.

Anthony P. Pugsley
Unité de Génétique Moléculaire
Institut Pasteur
25, rue du Dr Roux
Paris 75724 Cedex 15
France
Tel: (1) 45 68 84 94; FAX: (1) 45 68 85 21; EARN/BITNET:MAXaPASTEUR

Genetics and protein targeting. No single methodological approach can lead to a complete understanding of any complex biological system. The emphasis placed upon different approaches depends to a great extent on the ease with which they can be adapted to a particular problem. One area of research which has benefitted enormously from the use of classical and in particular molecular genetics is that of protein targeting. This is not to say, however, that a genetic approach can be used in isolation; instead it provides basis for further studies using biochemical and other approaches. There is, for example, no doubt that the great advances in our understanding of the mechanisms of protein targeting in bacteria which arose from the pioneering work on the genetics of the system by Jon Beckwith, Tom Silhavy, Maurice Hofnung and Maxime Schwartz in the late 1970s only came to fruition when Bill Wickner and others developed the first in vitro assays with which to study the process biochemically.

Geneticists either study genome structure and the ways in which genes are regulated and expressed, or use the tools of genetics to study other processes. These tools are numerous and varied; the purpose of this article is to briefly describe some of them and to discuss the results obtained through their use to study protein targeting.

Gene expression systems. The simplest genetic tools used in studies on protein targeting are gene expression systems. Virus-infected cells were often used in early studies on protein targeting in mammalian cells. Major coat proteins of many viruses such as Vesicular Stomatitis virus, Influenza virus or Semliki Forest virus are assembled into viruses at the plasma membrane which they reach via the same secretory pathway (endoplasmic reticulum-transition element-Golgi apparatus-trans Golgi network-secretory vesicles-plasma membrane) as endogenous secretory proteins. Viral genome-encoded proteins are the major products of virus-infected cells, and are therefore readily detectable in simple pulse-chase experiments. Even the pioneering work on protein export in the bacterium Escherichia coli by Bill Wickner and his colleagues was based on studies of procoat (procoat), the major component of the
bacteriophage M13 which assembles in the cytoplasmic membrane of infected bacteria.

Although viruses continue to be extensively used in studies on protein targeting, they are gradually being replaced by custom-made expression systems in which genes for endogenous proteins are placed under the control of a strong foreign promoter. The more sophisticated, "second generation" systems have regulated promoters which allow the amount of protein being produced to be carefully controlled. The entire package of promoter and cloned gene is usually inserted into high copy number plasmids or defective viruses which can be introduced into novel host cells. Plasmids usually carry selectable markers (e.g., antibiotic resistance genes) which allow them to be maintained in the new host. In other cases, the plasmid or defective virus is allowed to integrate into the genome of the host cells, which increases the stability of the recombinant cell lines but reduces the copy number of the cloned gene (and hence the amount of encoded protein which can be produced). These expression systems have opened the way to studying targeted proteins in specially adapted cell lines such as certain epithelial cells or other cells which can be readily grown in culture.

**Gene fusions and hybrid proteins.** The premise that targeted proteins contain one or more signals which direct them into particular targeting pathways and ensure that they are correctly sorted is the basis of much of the present work on protein targeting. In the most complex situation, that of plant cells, a minimum of five distinct routing signals are required to distinguish cytoplasmic proteins from proteins which are routed into the secretory pathway (cell surface, secretory organelles and vacuole), mitochondria, chloroplasts or other plasmids, the nucleus or the peroxisome. In fact, different proteins routed to the same target may have different routing signals and use different pathways, so that the actual number of routing signals which have to be "decoded" by the cell is probably somewhat larger than five. Additional sorting signals are required for the correct localization of proteins once they have entered a particular route. Most if not all routing signals are linear sequences of amino acids. While some sorting signals are also linear amino acid sequences, many may be composed of amino acids which are brought together only in the folded polypeptide (patch signals) and others are known to include prosthetic groups (e.g., the mannose-6-phosphate residues which are the sorting signals on lysosomal hydrolases).

Gene fusions have proven particularly powerful tools for detecting and identifying routing signals and sorting signals composed of short polypeptide segments. The basic principle of the gene fusion approach is that a passenger or reporter protein devoid of any routing and/or sorting signals can be targeted to any location in the cell by fusing it to