This Week's Citation Classic

(Research Group for Theoretical Biophysics, Department of Theoretical Physics, Royal Institute of Technology, Stockholm, Sweden)

This paper reported a statistical study of the amino-acid sequences of secretory signal peptides. In particular, it described that only small, uncharged residues are allowed in positions -3 and -1 relative to the site of cleavage between the signal peptide and the mature protein. This observation served as a basis for a scheme that predicts the most likely cleavage site when only the primary sequence of the precursor protein is known. (The SCI® indicates that this paper has been cited in over 380 publications.)

Prediction of Cleavage in Secretory Proteins

G. von Heijne
Department of Molecular Biology
Center for Biotechnology
Huddinge University Hospital
Karolinska Institute
S-141 86 Huddinge
Sweden

September 23, 1988

As a graduate student in Clas Blomberg's Research Group for Theoretical Biophysics at the Royal Institute of Technology in Stockholm, I had one scientifically very fruitful idea: to brush up my rusty high-school French. A demanding teacher made me subscribe to La Recherche, a French popular-science magazine. Flipping through its pages one day, I stumbled across a short piece on protein secretion. It described the classic G. Blobel and B. Dobberstein paper1 that presented the first full-blown version of the signal hypothesis. A small figure illustrated the main idea: a signal peptide initiating cotranslational protein translocation across the membrane of the endoplasmic reticulum (ER). The hydrophobic signal peptide was shown as somehow squeezing through the membrane via a hydrophobic membrane, ending up, after cleavage, as a freely soluble peptide in the lumen of the ER.

This didn't make sense to me: a hydrophobic peptide ought to be anchored in the membrane, most likely with its charged amino-terminal end remaining in the cytoplasm. I later found out that this was the essence of the so-called "loop model." Fortunately, I didn't know this at the time, or I would never have been drawn into the field of protein sorting.2 At any rate, this inspired me to write a paper dealing with the energetics of a polypeptide chain passing through a lipid bilayer.3

I then got interested in the primary sequences of secreted proteins, and a study of the then-known signal peptides was a fairly obvious step. Again, I didn't know that this had been done before on smaller collections of sequences,4 and it turned out that my sample was just the right size for discerning what has later become known as the (-3,-1)-rule for the cleavage site between the signal peptide and the mature protein; only small, uncharged residues are allowed in positions -3 and -1. As it happened, an equally well-cited paper5 with essentially the same message was published by D. Perlman and H. Halvorson within a few weeks of my paper.

The main reason for the many citations is that genes and cDNAs for secretory proteins represent a large proportion of current DNA-sequencing efforts. The (-3,-1)-rule allows one to make a reasonable prediction of the site of signal peptide cleavage in such proteins. If a few thousand new protein sequences are deduced from their DNA sequences per year, and if, say, 20 percent of these represent secretory proteins, and if a good number of the papers reporting these sequences cite the (-3,-1)-rule, one is bound to end up with quite a few centimeters of Science Citation Index® column-space. It is thus simple mass-market effects, rather than profound insight or theoretical sophistication, that marks the success of this Citation Classic. As for a moral, I guess that the story underlines the well-documented importance of ignorance and French in all scientific work.

References: