# Sequence Analysis for Discrimination of Signal Peptides and Signal Anchor Sequences

**Masahiro Gomi** gomi2@proteome.bio.tuat.ac.jp
**Shigeki Mitaku** mitaku@cc.tuat.ac.jp
Tokyo University of Agriculture and Technology, Department of Biotechnology, Koganei, Tokyo 184-8588, Japan

## 1 Introduction

Signal peptides (SP) are responsible for the secretion of mature proteins, and signal-anchor sequences of type I and type II (SA-1, SA-2) translocate C-terminal and N-terminal segments, respectively [1].

However, it is difficult to discriminate difference between SP and SA-2 from the information of amino acid sequences alone, because the difference between those sequences does not have significant motif [3].

Particularly, any clear sequence motif has not been reported so far in the C-terminal end region at which SP is cut off, although typical signal sequences show positively charged residues at their N-terminal end [2].

There are some reports of discrimination between SP and SA-2, but their accuracies are not satisfactory for practical application.

Previously, we compared the distribution of electric charges charges at the C-terminal end of SA type II and found that the helical periodicity of charges is statistically large.

In this work, we developed a new index for the discrimination between SP and SA-II, which enables the discrimination rate of about 80%.

## 2 Material and Methods

We have prepared three kinds of dataset from SWISS-PROT: (1) signal peptides from eukaryote, (2) the same type of sequences from prokaryote and (3) signal anchors type II. The datasets were selected whose pair-wise sequence homology with any other sequences is lower than 25% and the number of data was 71 for SA-II, 123 for SP of eukaryote and 73 for SP of prokaryote. We have aligned all sequences by the positively charged residues, K and R, ubiquitously found at the N-terminal end of signal sequences. Statistical analysis has been practiced at downstream region from this aligned point consists of 36 amino acid residues. We have defined simple index (SS index:Ala:-0.5 Cys:0 Asp:-1 Glu:-1 Phe:1 Gly:-0.5 His:0 Ile:1 Lys:0 Leu:1 Met:0 Asn:-1 Pro:-2 Glu:-1 Arg:0 Ser:-0.5 Thr:-0.5 Val:1 Trp:0 Tyr:0) for discrimination of signal peptides based on the result of frequency analysis which reveals differences between SP and SA-II in tendency of the appearance of amino acid residue. Then, we calculated mean SS Index in downstream of aligned point.

## 3 Results and Discussion

All amino acid sequences were aligned by a positively charged residue at N-terminal end of signal sequences. In the region between 10-th to 27-th residues from the positive charge at the N-terminal end exhibited significant difference in mean SS index distribution. The peak of the average SS-index distribution was significantly larger in SA-II than SP. When SP and SA-II discriminated by the

threshold value of 0.083, the selectivity was 83%. This result is the best selectivity in the problem of the discrimination between SP and SA-II.

It seems very natural that there is significant difference between SP and SA-II in the C-terminal region of signal sequences, because the cleavage site is located at the C-terminal end of SP. Although cleavage site hardly has distinct sequence motif, SS index has to show an essential factor of the discrimination between SP and SA-II.

We have defined SS Index so that it reflects the hydrophobicity (positive) and negative charge (negative) of amino acid residues, and dare to remove the influence of positive charged residue from this index, because in C-terminal end of signal sequences, the frequency of positive charged residue seems very similar. Therefore, the influence of positive charged residue in discrimination SP and SA-II by SS Index was considered as noise. The result of mean SS index analysis exhibited tendency clearly that SP (SA-II) was distributed to negative (positive) regions. This result suggests that negative charge and hydrophobicity of C-terminal end residues of signal sequence is important for characterization of SP and SA-II.

## References


[1] Plath, K., Mothes, W., Wilkinson, B.M., Stirling, C.J., and Rapoport, T.A., Signal sequence recognition in posttranslational protein transport across the yeast ER membrane, *Cell*, 94(6):795–807, 1998.

[2] Von Heijne, G., Computer-assisted identification of protein sorting signals and prediction of membrane proteins topology and structure, *Advances in Computational Biol.*, 2:1–14, 1996.

[3] Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G., Identification of prokaryotic and eukaryotic signal prediction of their cleavage sites, *Protein Eng.*, 10(1):1–6 , 1997.