Charge Distribution of Signal Peptides and Signal Anchor Type II

Masahiro Gomi [1] gomi2@cc.tuat.ac.jp
Takatsugu Hirokawa [12] takatsugu@din.or.jp
Shigeki Mitaku [1] mitaku@cc.tuat.ac.jp

[1] Tokyo University of Agriculture and Technology, Department of Biotechnology, Tokyo 184-8588, Japan
[2] Present address: Ryoka System Inc.

1 Introduction

Hydrophobic signal sequences, that is, signal peptides (SP) and type I and type II signal-anchor sequences (SA-1, SA-2), are responsible for secretion of mature proteins and translocation of C-terminal and N-terminal segments, respectively [1]. Almost all signal sequences show positively charged residues at their N-terminal end [2]. However, it is difficult to discriminate difference between SP and SA from the information of amino acid sequences alone, because those sequences have very similar characteristics [3]. Particularly, any clear sequence motif has not been reported so far in the C-terminal end region at which SP is cut off. Only so-called 1-3 rule has been reported, which postulates that small residues are usually observed at the position of -1 and -3 from the cleavage site of SP [3]. In this work, we have carried out comparative studies of the charge distribution of SP and SA type II and revealed significant helical periodicity of positive charge at the C-terminal end of SA type II.

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 SP, 123 for SA-II of eukaryote and 73 for SA-II 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. Then, we have calculated two parameters characterizing the charge distribution of signal sequences: (1) net charges (the difference between the number of positively charged residues and negatively charged ones) per a supposed helix and (2) total charges (the sum of the number of charged residues) per a helix.

3 Results and Discussion

hen all amino acid sequences were aligned by a positively charged residue at the N-terminal end of signal sequences, charged residues are rarely found in the hydrophobic regions from 1-st to about 15-th position. The lack of strongly polar residues in this region of SP and SA-2 is quite natural, since polar groups have large electric self energy in the nonpolar environment of membrane. Whereas, both positive and negative charges increased at the C-terminal end. Although the distribution of charged residues seems similar between SP and SA, the net charges and the total charges showed significant difference between the two kinds of sequences. The charge distribution in the region between 18-th to 35-th residues exhibited marked helical periodicity of positive net charges for SA type II. However, amino acid sequences of SP in the same region did not show any helical periodicity and the average net charge was rather negative. This difference in the distribution of net charges suggested that the helical structure of hydrophobic segments is extended to the C-terminal region in SA-II, and the putative helices in C-terminal end region appear to interact with the negatively-charged part of machinery for translocation of signal anchors. The distribution of total number of charged residues was similar between SP and SA with an exception of the lack of charged residues at the positions 18-th and 19-th residues in SA-II. This narrow and deep hole in the charge distribution could not be observed for SP of eukaryote as well as prokaryote. Although the physical meaning of this feature is not clear yet, it may be related to the cleavage of SP or the inhibition of the cleavage of SA. The difference in the charge distribution between SP and SA found in this work is statistically significant, and the discrimination between the two types of signal sequences may become possible by analyzing the charge distribution around the C-terminal end.

References

[1] Plath, K., Mothes, W., Wilkinson, B.M., Stirling, C.J., 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.