Analysis of cleaved N-terminal sequences coming from MS/MS proteomics for E.coli and S.cerevisiae

Dmitry Ivankov*
Technical University of Munich

Stefano Bonissoni
University of California at San Diego

Pavel Pevzner
University of California at San Diego

Dmitrij Frishman
Technical University of Munich

Abstract

Sub-cellular localization is an important aspect of protein function. It determines the molecular environment in which proteins operate, in particular the availability of interaction partners. In bacteria, the majority of proteins execute their function in the cytoplasm, but a sizeable fraction of gene products are directed to other cellular locations — the cytoplasmic membrane, cell wall and extracellular space in Gram-positive bacteria and the cytoplasmic membrane, the periplasm, the outer membrane and the extracellular space in Gram-negative bacteria [1]. In eukaryotes more than 20 different localizations can be distinguished [2] with the major ones being cytoplasm, nucleus, mitochondria, extracellular space, various kinds of membranes, as well as chloroplasts in plants.

Post-translational transport of proteins to cellular compartments involves translocation across at least one membrane. In many cases proteins are targeted to their cellular destinations by means of short sequence motifs that are involved in molecular interactions with membrane receptors. In particular, proteins directed to the secretory pathway, to mitochondria, and to chloroplasts typically contain signal peptides, mitochondrial targeting peptides, and chloroplast transit peptides, respectively, located at their N-terminus. These three types of targeting motifs that are cleaved off after translocation by signal peptidases constitute the focus of our research project.

The two major pathways for exporting unfolded and folded proteins are the essential and universal Sec (general secretion) system and the Tat (twin arginine translocation) system [3]. The Sec translocase is found in the cytoplasmic membrane of all bacteria, archaea, in the thylakoid membrane of plant chloroplasts, and in the endoplasmic reticulum of eukaryotic cells. In eukaryotes proteins secreted through the Sec-pathway are further directed to other compartments via the vesicle sorting route. The Tat pathway exists in many bacteria and archaea, in the thylakoid membranes of plant plastids, and, presumably, in plant mitochondria [4].

*ivankov13@gmail.com

Here we examine targeting signals of E.coli and S.cerevisiae derived by correlating mass spectral data with the genome structure in the framework of a proteogenomics approach. The analysis focuses exclusively on peptides with a non-tryptic N-terminus with no upstream coverage [5]. Such peptides indicate that the upstream N-terminal protein fragments, not directly observed in MS experiment, were cleaved in vivo by some peptidase. Analysis of the E.coli and S.cerevisiae genomes as well as of the previously studied S.oneidensis [5] genome shows that in most cases it is the signal peptidase I and mitochondrial peptidases (in case of yeast).

Analysis of E.coli shows that out of 42 cleaved N-terminal peptides detected by MS proteomics about 60% are signal peptides, about 30% are not signal peptides, and the remaining 10% can not be easily classified. Out of signal peptides 80% comes from periplasm proteins, while 12% comes from inner/outer membrane proteins. 8% are exported via the Tat-signal pathway.

In S.cerevisiae the number of cleaved N-terminal peptides detected in MS proteomics is an order higher than in E.coli, namely, 413 peptides. Of them more than one third are confident signal/transit peptides and about 40% are probably not signal or transit peptides. About 70% of the confident signaling sequences are mitochondrial transit peptides, while the remaining sequences are apparently cleaved by signal peptidase I. Comparison of the found mitochondrial target peptides with those annotated in UniProt reveals different scenarios of action of mitochondrial peptidases: 1) the so called *R-2* motif is indicative of the cleavage by the mitochondrial processing peptidase (MPP), 2) the *R-3* motif corresponds to processing by MPP followed by intermediate cleavage peptidase (Icp55), 3) the *R-10* motif mediates the action of the MPP followed by octapeptidase (Oct1), and finally 4) the *R-11* may be the evidence of the experimentally yet unknown combined action of MPP, Oct1, and Icp55.

To summarize, proteogenomics data for E.coli and S.cerevisiae shows that most of the in vivo proteolytic cleavages in the cell detected by MS are cleavages of signal and transit peptides. Even for such well-studied model organisms we observed many signal/transit peptides not yet confirmed by any experiment. Extrapolation of the results to the poorly studied genomes implies that the proteogenomics approach has a great potential for expanding our knowledge about signal sequences.

References

[1] J.L. Gardy, F.S.L. Brinkman: Methods for predicting bacterial protein subcellular localization. *Nat Rev Micro*, 4:741-751, 2006.

[2] A. Pierleoni, P.L. Martelli, P. Fariselli, R. Casadio: eSLDB: eukaryotic subcellular localization database. *Nucleic Acids Res*, 35:D208-212, 2007.

[3] P. Natale, T. Bruser, A.J. Driessen Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane–distinct translocases and mechanisms. *Biochim Biophys Acta*, 1778:1735-1756, 2008.

[4] C. Robinson, A. Bolhuis: Tat-dependent protein targeting in prokaryotes and chloroplasts. *Biochim Biophys Acta*, 1694:135-147, 2004.

[5] N. Gupta, S. Tanner, N. Jaitly, J.N. Adkins, M. Lipton, et al.: Whole proteome analysis of post-translational modifications: applications of mass-spectrometry for proteogenomic annotation. *Genome Res*, 17:1362-1377, 2007.