Proteogenomic analysis of bacterial signal peptides

Dmitry N. Ivankov[1], Samuel H. Payne[2], Stefano Bonissone[3], Si Wu[4], Roslyn N. Brown[2], Ljiljana Paša-Tolić[4], Matthias Birschenk[5], Iris Antes[5], Richard D. Smith[2], Michael Y. Galperin[6], Pavel A. Pevzner[3,7], Dmitrij Frishman[1]

[1]*Department of Genome-Oriented Bioinformatics, Technische Universität München, Freising, Germany;*

[2]*Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA;*

[3]*Bioinformatics Program, University of California San Diego, La Jolla, CA, USA;*

[4]*Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland WA, USA;*

[5]*Theoretical Chemical Biology and Protein Modelling Group, Technische Universität München, Freising, Germany;*

[6]*National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA;*

[7]*Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA, USA;*

Over the last five years proteogenomics (using mass spectroscopy to identify proteins predicted from genomic sequences) has emerged as a promising approach to the high-throughput identification of protein N-termini, which remains a problem in genome annotation. Comparison of the experimentally determined N-termini with those predicted by sequence analysis tools allows identification of the signal peptides and therefore making? conclusions on the cytoplasmic or extracytoplasmic (periplasmic or extracellular) localization of the respective proteins.

Our proteogenomic study of the signal peptides in *Escherichia coli* K-12 recovered more than a third of all signal peptides that had been experimentally determined during the past three decades and confirmed at least 31 additional signal peptides, mostly in the known exported proteins, which had been previously predicted but not validated. The filtering of putative signal peptides for the peptide length and the presence of an eight-residue hydrophobic patch and a typical signal peptidase (SPase) cleavage site proved sufficient to eliminate the false-positive hits. Surprisingly, the results of this proteogenomics study, as well as a re-analysis of the *E. coli*

genome with the latest version of SignalP program, show that the fraction of proteins containing signal peptides is only about 10%, or half of previous estimates.

Currently experimentally known signal peptides catalogued in common databases like UniProt come from only a narrow taxonomic sampling. As a result, the dominant view is that signal peptides cleaved in the general secretion route (Sec-pathway) are defined by a canonical AxA motif. Our experimental proteomics data from 32 bacterial and archaeal organisms from eight phyla demonstrate that this AxA dogma is incorrect. Discoveries include fundamentally distinct signal peptide motifs (VxA, SfS, SfA, and LfA) from Alphaproteobacteria, Spirochaetes, and Euryarchaeota. Surprisingly, divergent motifs correlate with a proteome-wide reduction in alanine. This suggests a unique evolutionary co-adaptation of the SPaseI enzyme and its cleavage site in response to reduced alanine content. Computational analyses reveal numerous major evolutionary clades that have replaced the canonical signal peptide sequence with novel motifs.

To investigate the structural basis of the observed phenomenon we have also performed molecular docking simulations of six organism-specific peptides into the corresponding native peptidases. We observed that the canonical peptide from *E. coli* binds consistently stronger to all peptidases than the corresponding organism-specific peptides. Thus, based on the structural analysis, we conclude that the shift in pattern is not caused by differences in the SPaseI structures.

The research was supported by DFG International Training and Research Group RECESS (Regulation and Evolution of Cellular Systems), NSF award EF-0949047, NIH Intramural Research Program at the National Library of Medicine, NIAID IAA-Y1-A1-8401, PNNL is operated for the DOE by Battelle under Contract DEAC05-76RLO 1830.

1. S.H.Payne et al. (2012) Unexpected diversity of signal peptides in prokaryotes. *MBio*, **3:**e00339-12.

2. D.N.Ivankov et al. (2013) How many signal peptides are there in bacteria? Environ. Microbiol., **15:**983-990.