# A novel approach to recognize peptide functions in microorganisms: Establishing systems biology-based relationship networks to better understand disease causes and prevention

**Erich Gombocz** [1], **James Candlin**[2]

[1] IO Informatics, 2550 Ninth Street, Suite 114, Berkeley, CA 94710, USA
[2] SAGE-N Research, 1525 McCarthy Blvd, Suite 1000, Milpitas, CA 95035, USA
Correspondence: egombocz@io-informatics.com

ABSTRACT (*300 words*)

Characterization, identification and understanding of microorganisms in their impact on diseases has been challenging because although there is a wealth of information on microbial proteins (and even complete organism genomes) available, using it in context of the biological system to identify functional mechanisms involved as disease causes has been fragmented. Although recognized as important, investigating similarities across organisms in their infective behavior is still in an early stage.

This poster describes a novel approach to identify peptides from different microorganism with common mechanism of actions, and to categorize them as potential biomarkers to detect microbial threats prior to onset of disease symptoms.

To balance complexity of systems biology networks against broadly applicable results, the workflows consist of: (1) peptide analysis of microorganism via MS and identification by sequence scoring; (2) annotation of pre-characterized peptides with relevant genes to check identification correctness; (3) use of public resources (PATRIC, ICTV, VIDA, Viral ORFeome, miRBase) to semantically integrate taxon-specific functional genomic and pathway information, cluster by sequence similarity to homologous protein families (HPFs), then annotate them (functional classification, related protein structures, protein length, boundaries of conserved regions, organism-specific genes). Next, incorporate disease-related pathways (BioCyc, KEGG) for further enrichment. The resulting knowledge network consists of functionally annotated peptides in their relationships to diseases. (4) Identify peptides with similar disease-causing functions appearing in multiple pathogens for further generalization and network reduction. Semantic (SPARQL) queries result in key pathway intersections common in the disease, which further decreases the number of target proteins. (5) Establish a set of peptide markers as applicable molecular signatures (Applied Semantic Knowledgebases) and test their validity as decision support in multiplexed assays.

Future applications will apply this technology to automated screening for biological threats, to characterize origin and type of disease and to develop preventive measures (drugs or vaccines) effective for several classes of microorganism.

2 Tables, 3 figures