The Computational Molecular Biomedicine group is pursuing N major research topics as well as collaborative projects in diverse biomedical research areas. Most projects combine broad knowledge of molecular biology with state-of-the-art mathematical modeling and machine learning techniques.

1. Proteomics data analysis method and software development

The goal of this project is to develop high-performance, high-throughput analysis methods and software tools for proteomics datasets. In particular, we are interested in applying machine learning techniques on both CPU and GPU architectures to improve peptide identification and quantification from mass spectrometry data. Example application areas include identification of (i) crosslinked peptides whose mass spectra are generally 3-4 times more complex than those found in typical proteomics studies and (ii) novel peptide sequences and modifications that currently do not exist in reference databases.

As an illustration of a possible impact of this project, the use of zero-length chemical crosslinking – an experimental technique for probing structural information of protein complexes, which has previously been restricted to studies of simple proteins with less than 100 kDa sequence complexity, became practical on a red cell membrane system with much larger than 1 M Da sequence complexity with the aid of a software tool, ZXMiner, in ~2010-2013. Importantly, ZXMiner also identifies crosslinked peptides with high confidence – a critical concern in any study of protein structure where few misidentifications would lead to incorrect structure model. With improved data throughout and quality from recent generations of high resolution mass spectrometer, application of machine learning techniques to crosslinked peptide identification is expected to greatly enhance the confidence of crosslinked peptide identification and enable routine analyses of even more complex biological systems.

This project will be performed in collaboration with the Systems Biology Center at Chulalongkorn University (CUSB) and Dr. David W. Speicher’s proteomics research laboratory at the Wistar Institute in Philadelphia, PA USA.

1. Integrative analysis of high-throughput biomedical datasets

[Public datasets + internal datasets]

[Computational analyses + biological interpretation]

[What are primary biological topics]

1. Mathematical modeling and simulation of chromatin folding process

[]