

Name: **David duVerle**

Title: CaMPDB: a resource for calpain and modulatory proteolysis (9<sup>th</sup> International Workshop on Bioinformatics and Systems Biology)

Workshop report :

## 1. Work Presentation

The work presented was CaMPDB, a database resource on modulatory proteolysis, focusing on calpain, a well-studied intracellular protease which regulates substrate functions by proteolytic processing. CaMPDB contains sequences of calpains, substrates and inhibitors as well as substrate cleavage sites, collected from the literature.

The presentation focused on two main aspects of CaMPDB project:

- 1) database content and features
- 2) calpain cleavage prediction.

For the first part, I outlined the work that went into gathering the data from heterogeneous source and providing a clear interface to use it, as well as different experimental validations and statistics that were ran on the data content.

The second part described in more details the effort to use the manually curated dataset of calpain substrates in order to build a cleavage prediction tool that can be used from the CaMPDB website, but also provide insights into the mechanism of modulatory proteolysis by calpain. Results of different machine-learning approaches to this problem were presented, along with analysis and future direction of the project.

## 2. Interactions

The presentation seemed to elicit interest and yielded many questions from audience members.

Particular points discussed with other researchers following the presentation where:

- Specifics of calpain's inhibition by calpastatin and the possibility to use calpastatin data gathered for CaMPDB in order to improve the understanding of the cleavage mechanisms of calpain.
- The method used to expand the set of curated substrates (104 sequences) to a set of computationally expanded substrates (1914 sequences) and whether such data could possibly be used to improve results obtained through supervised-learning methods.
- Goals and limitations of the different machine learning methods used (PSSM and SVM) compared to those considered in future direction of the project (HMM). Ways to further analyze the results in order to extract information on the underlying biological model.

In addition, many notes and suggestions were exchanged with other projects sharing similar machine-learning approaches (SVM/HMM) and the different issues that had to be dealt with in order to achieve proper results.

若手研究者インターナショナル・トレーニング・プログラム(ITP)  
バイオインフォマティクスとシステムズバイオロジーの国際連携教育研究プログラム  
ワークショップ参加レポート

若手研究者インターナショナル・トレーニング・プログラム(ITP)  
バイオインフォマティクスとシステムズバイオロジーの国際連携教育研究プログラム  
ワークショップ参加レポート