NWO/RGC JOINT RESEARCH SCHEME
COMPLETION REPORT

Project Reference Number

D-HK006/11T-II

Project Title

Development of complex, high-throughput data processing workflows for LC-MS based proteomics

Particulars

	Hong Kong team	Dutch team
Name of Principal Investigator (with title)	English: Henry H. N. Lam Chinese: 林熙寧	Péter Horvatovich
Name of Co-Investigator (if any)	English: Chinese:	
Institution or Institutional affiliation	[ ] CityU [ ] HKU [ ] CUHK [✓] HKUST [ ] HKBU [ ] LU [ ] HKIEd [ ] PolyU	University of Groningen, Faculty of Mathematics and Natural Sciences, Department of Pharmacy, Analytical Biochemistry
Other project team members (if any)		

Funding Period

	1st year	2nd year (if applicable)
Start Date	1 July 2011	1 July 2012
Completion Date	30 June 2012	30 June 2013

Objective(s) as per original application

1. Development of generic, complex, accurate LC-MS(/MS) proteomics data processing workflow by combination of single stage mass spectrometry and data dependent MS/MS processing tools.
2. Development of peptide and protein identification workflow by combination of MS/MS spectra obtained by electron transfer dissociation and collision induced dissociation to improve protein identification quality.
3. Connecting LC-MS(/MS) based discovery to MRM method development and data analysis.
4. Providing easy-to-use, high-throughput data processing services for the developed data processing workflows using modern parallel computational facility with large computational power.

ease attach relevant document(s)]

i) Outline of proposed research and results obtained

We proposed to develop a proteomics data processing workflow by integrating the data analysis expertise in the Lam group (primarily MS/MS-based) and the Horvatovich group (primarily LC/MS-based) (Objective 1 and 3). To this end, we have cross-trained our group members through invited seminars and informal interactions during our visits. To allow the Dutch team to readily use the MS/MS data processing workflow in the Lam group, the Trans-Proteomic Pipeline (TPP) suite has been implemented in the Galaxy framework. Through sharing his knowledge and his msCompare software, Prof. Horvatovich has also helped the Lam group develop a new LC-MS alignment platform, called LWBMatch (manuscript in review in *Bioinformatics*).

Two other collaborative projects have been initiated through the mutual visits. First, we recognized that workflows using multi-dissociation (CID/ETD) tandem mass spectrometry methods (Objective 2) is most needed for the analysis of intact glycopeptides, one of the last frontiers of computational proteomics. Therefore, we started a project, together with a leading glycobiologist in the Netherlands, Prof. Manfred Wuhrer of Leiden University Medical Center, to combine LC-MS and multi-dissociation MS/MS data analysis methods for intact glycopeptides identification. Preliminarily we have obtained some promising results in the identification of intact glycopeptides through building and searching a predicted spectral library of glycopeptides CID spectra.

Second, through this connection, the Lam group has joined the Dutch team in the Chromosome-based Human Proteome Project (C-HPP), an internationally-coordinated project to map the human proteome in the next 5 years. This project has just been started and is expected to deliver MS-based and antibody-based assays for all human proteins at its completion. The Dutch team is responsible for mapping all proteins encoded by genes on Chromosome 5.

ii) Significance of research results

The integration of TPP in Galaxy is expected to benefit both the Horvatovich group and the research community in the Netherlands at large, since it is through this platform that the bioinformatics groups in the Netherlands share their software tools. The glycopeptides identification project is the first to attempt high-throughput, system-wide profiling of intact glycopeptides. If successful, this project will significantly impact the field of glycobiology, as it will vastly speed up the identification of glycoproteins, with site-specific glycan structure information. The C-HPP is expected to deliver accurate and high-throughput protein assays for all human proteins to the hands of the biologists.

iii) Research output

One publication in review from the Lam group benefited significantly from this collaboration. The manuscript submitted to *Bioinformatics* is attached.

iv) Potential for or impact on further research collaboration

As described in i) above, two potentially high-impact collaborative projects were initiated through the mutual visits and will continue for at least the next 3-5 years. We have also agreed to look into possibility of student exchanges (or joint supervision of new PhD students) to facilitate more interactions in the future.