

University of Groningen

## Mastering data pre-processing for accurate quantitative molecular profiling with liquid chromatography coupled to mass spectrometry

Mitra, Vikram

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*  
2017

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Mitra, V. (2017). Mastering data pre-processing for accurate quantitative molecular profiling with liquid chromatography coupled to mass spectrometry [Groningen]: University of Groningen

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



university of  
groningen

# **Mastering data pre-processing for accurate quantitative molecular profiling with liquid chromatography coupled to mass spectrometry**

## **PhD thesis**

to obtain the degree of PhD at the  
University of Groningen  
on the authority of the  
Rector Magnificus Prof. E. Sterken  
and in accordance with  
the decision by the College of Deans.

This thesis will be defended in public on

Monday 3 July 2017 at 09.00 hours

by

**Vikram Mitra**

born on 3 July 1985  
in Kolkata, India

**Supervisors**

Prof. P.L. Horvatovich

Prof. R.P.H. Bischoff

Prof. A.K. Smilde

**Assessment Committee**

Prof. M. Biehl

Prof. L. Martens

Prof. G.L. Corthals

# Table of contents

<b>Table of contents</b> .....	<b>3</b>
<b>Chapter 1. General Introduction</b> .....	<b>5</b>
1.1 DATA PRE-PROCESSING STEPS INVOLVED IN ANALYSIS OF LABEL-FREE LC-MS/MS PROTEOMICS DATASETS....	9
1.2 SCOPE OF THE THESIS.....	10
REFERENCES .....	12
<b>Chapter 2. Monotonic shifts and orthogonality in single-stage LC-MS(/MS) data</b> .....	<b>16</b>
2.1 ABSTRACT.....	16
2.2 INTRODUCTION.....	17
2.3 ACCURATE ALIGNMENT OF SINGLE-STAGE LC-MS(/MS) DATA.....	22
2.4 CAUSES FOR SHIFTS AND ORTHOGONALITY IN SINGLE-STAGE LC-MS DATA.....	28
2.5 CONCLUSION.....	39
REFERENCES .....	41
<b>Chapter 3. Inversion of peak elution order prevents uniform time alignment of complex liquid-chromatography coupled to mass spectrometry datasets</b> .....	<b>48</b>
3.1 ABSTRACT.....	48
3.2 INTRODUCTION.....	49
3.3 MATERIALS AND METHODS.....	52
3.4 THEORY.....	52
3.5 RESULTS AND DISCUSSIONS .....	58
3.6 CONCLUSION AND OUTLOOK.....	72
REFERENCES .....	74
<b>Chapter 4. Assessment of identification transfer in data dependent LC-MS/MS experiments</b> .....	<b>78</b>
4.1 ABSTRACT.....	78
4.2 INTRODUCTION .....	80
4.3 MATERIALS AND METHOD.....	82
4.4 RESULTS.....	86
4.5 CONCLUSION.....	95
REFERENCES .....	98
<b>Chapter 5. Identification of analytical factors affecting complex proteomics profiles acquired in a factorial design study with ANOVA – simultaneous component analysis.</b> .....	<b>102</b>
5.1 ABSTRACT.....	102
5.2 INTRODUCTION.....	104
5.3 METHODS.....	106
5.4 SAMPLE PREPARATION AND DESCRIPTION OF EXPERIMENTAL FACTORS .....	108
5.5 LC-MS ANALYSIS.....	110
5.6 ASCA ANALYSIS.....	111
5.7 RESULTS & DISCUSSIONS .....	112
5.8 CONCLUSION.....	120
REFERENCES .....	122
<b>Chapter 6. Summary and future outlook</b> .....	<b>124</b>
REFERENCES .....	127
<b>Dutch Summary (Samenvatting)</b> .....	<b>130</b>
<b>Supporting information for chapter 3</b> .....	<b>135</b>
RAT CSF SAMPLING AND LC-MS ACQUISITION .....	135
MOUSE EXPERIMENTAL DESIGN DATASET.....	138
MONTE-CARLO SIMULATED DATASET .....	141
DETAILED DESCRIPTION OF THE TIME ALIGNMENT ALGORITHM.....	142

PRE-PROCESSING OF SINGLE STAGE LC-MS DATA .....	142
INTENSITY-RANK-BASED PEAK MATCHING OF LC-MS DATA (LEFT PART OF STEP 1. IN FIGURE 1).....	142
OPTIMIZING PEAK MATCHING PARAMETERS OF SINGLE STAGE LC-MS PEAK LISTS (PARAMETER OPTIMISATION IN STEP 1. IN FIGURE 1) .....	143
SELECTION OF ACCURATELY MATCHED PEAK PAIRS (STEP 2. IN FIGURE 1) .....	146
MONOTONIC NON-LINEAR ALIGNMENT FUNCTION (STEP 3. IN FIGURE 1).....	146
PROBABILITY OF PEAK ELUTION ORDER SIMILARITY BETWEEN TWO CHROMATOGRAMS (STEP 3. IN FIGURE 1) ...	147
RETENTION TIME CORRECTION (STEP 4. IN FIGURE 1).....	148
HARDWARE AND SOFTWARE ENVIRONMENT .....	149
PEPTIDE IDENTIFICATION PARAMETERS .....	149
PARAMETERS USED FOR OPTIMISATION OF INTENSITY-BASED PEAK MATCHING .....	150
LABELS LEGEND OF G-SCORE VERSUS 2D-KS PLOT IN FIGURE 2.....	150
SUPPLEMENTARY FIGURES FOR CHAPTER 3.....	151
<b>Supporting information for chapter 4.....</b>	<b>164</b>
$X_{\text{REA}}$ SCORE CALCULATION .....	164
SUPPLEMENTARY FIGURES FOR CHAPTER 4.....	164
<b>Supporting information for chapter 5.....</b>	<b>170</b>
MATERIALS AND METHODS .....	170
LC-MS/MS (QTOF) .....	170
DATA PRE-PROCESSING AND QUANTIFICATION.....	170
ANNOTATION OF THE ALIGNED PEAK MATRIX .....	171
PARAMETERS OF THE SIMULATED DATASET .....	172
MAIN STEPS REQUIRING BIOINFORMATICS INTERVENTION.....	173
SUPPLEMENTARY FIGURES FOR CHAPTER 5.....	173
<b>Acknowledgements .....</b>	<b>184</b>
<b>List of publications.....</b>	<b>186</b>
PART OF THIS THESIS.....	186
OTHER PUBLICATIONS .....	186