Mathematical Method Validation Tools for Application to a Proteomics Approach of Postmortem Metabolic Capacity Estimation

Brigitte Desharnais

A Thesis
In the Department
of
Chemistry and Biochemistry

Presented in partial fulfillment of the requirements
For the Degree of
Doctor of Philosophy (Chemistry) at
Concordia University
Montréal, Québec, Canada

July 2019

CONCORDIA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Brigitte Desharnais

Entitled: Mathematical method validation tools for application to a proteomics approach of postmortem metabolic capacity estimation

and submitted in partial fulfillment of the requirements for the degree of

Doctor Of Philosophy (Chemistry)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____Chair
Dr. William Zerges

_____ External Examiner
Dr. Aurelien Thomas

_____ External to Program
Dr. Eric Pedersen

_____ Examiner
Dr. Yves Gelinas

_____ Examiner
Dr. Dajana Vuckovic

_____Thesis Co-Supervisor
Dr. Cameron Skinner

_____Thesis Co-Supervisor
Dr. Pascal Mireault

Approved by _____
Dr. Yves Gelinas, Graduate Program Director

August 29, 2019 _____
Dr. Andre Roy, Dean
Faculty of Arts and Science

Abstract

Mathematical Method Validation Tools for Application to a Proteomics Approach of Postmortem Metabolic Capacity Estimation

Brigitte Desharnais, Ph.D.

Concordia University, 2019

In postmortem cases, forensic toxicologists perform analyses for legal and illegal drugs, volatile substances, poisons and biochemical parameters in order to determine the causes and circumstances of death.

Evaluation of the metabolic capacity of an individual might help to achieve this goal. Knowledge that the deceased had a poor metabolic capacity might help differentiate between medical error and accidental overdose, for example.

Traditionally, DNA analysis of genes encoding for metabolizing enzymes has been used for this purpose. However, the genotype can be quite a poor predictor of phenotype; intervening factors such as sex, age, presence of inducers or inhibitors act as confouding factors.

A proof-of-concept methodology estimating the postmortem metabolic capacity through characterization and quantification of cytochrome P450 (CYP) enzymes in liver tissue is presented here. Combining quantitative proteomics with detection of the peptides bearing mutation sites allowed for a more accurate estimation of the metabolic capacity than genotyping alone.

The current regulatory environment, and best practices, requires forensics bioanalytical methods to be validated. Anticipating the validation of this method, several methodological issues were foreseen.

In order to properly validate the quantitative part of the CYP analysis method, a simple, analyst-independent, and systematic procedure to choose and validate a calibration model (order, weighting) based on statistical analysis was developed. Additionally, the omnipresence of the target analyte(s) in authentic matrix (human liver) calls for a methodology allowing to deal with endogenous concentration(s) of analytes in matrices used to prepare calibration standards and quality control samples. An automated tool was developed to correct for the endogenous analytes' concentration.

Finally, characterization of the CYP enzymes, via the monitoring of peptides bearing a mutation site, requires validation via a qualitative decision point method. Current guidelines about this type of analysis are ill adapted to deal with the binary nature of the results. A more suitable set of guidelines was developed and tested.

These mathematical method validation tools, in combination with the CYP analysis method, provide the necessary framework for metabolic capacity estimation in postmortem cases.

Acknowledgments

Je dédie cette thèse à mes grands-parents. À mon grand-père Vézina qui, me dit-on, aurait été tellement heureux de voir sa petite fille faire des études de troisième cycle universitaire. À ma grand-mère Desharnais, féministe avant l'heure à Guyenne. À ma grand-mère Vézina, qui a dû arrêter l'école en onzième année parce que l'argent commençait à manquer. Deux générations plus tard, une femme de la famille termine un doctorat. Comme quoi les choses progressent... lentement.

First, a few words of acknowledgement to my supervisors, Pascal Mireault and Cameron D. Skinner. I have learned and grown so much under your supervision. Thank you for sharing the breadth of your knowledge with me, and giving me the autonomy to challenge it and build my own.

Félix C., quelle chance et quel bonheur ça a été de travailler avec toi. Merci pour ton support, ton caractère posé, ta droiture, ton éthique de travail et ton merveilleux sens de l'humour. Je ne peux qu'espérer qu'on continue pour les 30 prochaines années.

À l'extraordinaire équipe de toxicologie du LSJML: vous êtes drôles, créatifs, travaillants, empathiques, compétents, que demander de plus? Ce fut une chance immense d'avoir pu compter sur toute votre aide et support.
Lucie, qu'est-ce que j'aurais fait sans toi, vraiment. Entre travail expérimental et support moral, entre demoiselle d'honneur et co-auteure, merci de tout ce travail que tu as fait.
Audrey, voisine de bureau des débuts, merci de ton enthousiasme sans faille, de ta bonne humeur constante et de tes encouragements soutenus pour le sprint final.
Cynthia (C., puisqu'il commence à y en avoir beaucoup!), merci d'avoir mis à ma disposition tes vastes connaissances en chromatographie, validation de méthode et spectromètres

de masse Sciex.
Édith, merci pour ton énergie, ton dévouement, ta bonne humeur et tes encouragements.
Gabrielle, experte des logiciels Sciex et d'Excel, merci pour toutes les collectes et analyses de données pour les projets de calibration et méthodes qualitatives. Merci aussi pour le support dans les moments difficiles.
Julie L., tu as la foi de pouvoir déplacer des montagnes, et le réalisme de savoir qu'il faudra toute une équipe pour y arriver. Merci pour ton support et ta recherche constante de la vérité expérimentale.
Marie-Jo, merci pour tout le travail ingrat fait sans broncher, dans la bonne humeur en plus!
Marc-André, merci pour ta force tranquille et ton travail discret mais imposant sur les méthodes et la révision des articles.
Stéphanie, merci pour ton implication et tes idées en validation de méthode, ton calme et ton efficacité.

Félix, merci de m'avoir accompagnée toutes ces années. Tu es vraiment l'homme le plus classe du monde (malgré tes chaussures en suédine bleue). Avec toi je n'ai jamais à être autre chose que moi-même, et ce fut un havre de paix au milieu de l'ouragan du doctorat. Quand plus rien n'allait, il y avait toujours tes bras pour me réconforter. Je suis plus que chanceuse de t'avoir rencontré... et marié.

Finalement, merci à ma famille, qui m'a supporté les huit (!) dernières années. De ma grand-mère qui a multiplié les prières, à mes parents qui ont écouté mes interminables tirades et découragements, en passant par mon frère et ma soeur qui se sont fâchés en solidarité avec moi. Votre soutien compte plus que je ne saurais le dire, et je suis chanceuse de vous avoir eus.

Musical inspirations: music from the motion picture "She's the Man", Mika (in particular the album "The Origin of Love"), wonderful Alexandra Stréliski and Jean-Michel Blais.

Contribution of Authors

This thesis is a compilation of six manuscripts: three of them are already published [1–3], two have been accepted for publication with modifications [4, 5] and the final one is under redaction [6].

In the publications presented in Chapter 2:

Procedure for the Selection and Validation of a Calibration Model I —Description and Application
Brigitte Desharnais, Félix Camirand Lemyre, Pascal Mireault and Cameron D. Skinner
Journal of Analytical Toxicology **41** (4) (2017) 261-268 [1]

and

Procedure for the Selection and Validation of a Calibration Model II —Theoretical Basis
Brigitte Desharnais, Félix Camirand Lemyre, Pascal Mireault and Cameron D. Skinner
Journal of Analytical Toxicology **41** (4) (2017) 269-276 [2].

- Brigitte Desharnais developped the idea of a comprehensive statistical process for the selection and validation of a calibration model;
- Brigitte Desharnais performed the initial literature review to catalogue statistical tests used for heteroscedasticity testing, weighting model selection, calibration model order selection and validation;
- Brigitte Desharnais evaluated the performance of the different statistical tests and selected them according to performance to build a comprehensive selection process;

- Félix Camirand Lemyre developped an automated selection process for the weighting factor, rather than the visual evaluation developped by Brigitte Desharnais in the first iteration of the procedure;
- Félix Camirand Lemyre selected the model validation means (normality testing) and created the R script to perform it;
- Félix Camirand Lemyre wrote the final R script that automates the selection and validation of a calibration model (Appendix B.3);
- Brigitte Desharnais performed the *in silico* evaluation of the procedure performance, including generation of simulated data sets and their evaluation by means of R scripts (Appendix B.2);
- Gabrielle Daigneault and Marc-André Morel performed experimental data collection (Sections 2.1.3 and 2.2.3, "LC-MS/MS quantification") and analysis under the direction of Brigitte Desharnais;
- Brigitte Desharnais wrote the manuscripts;
- Félix Camirand Lemyre reviewed the manuscripts and suggested corrections to the statistics sections;
- Pascal Mireault and Cameron D. Skinner supervised the project, volunteered some ideas, pointed to areas of improvement, and reviewed the final manuscripts.

In the publication presented in Chapter 3:

A Tool for Automatic Correction of Endogenous Concentrations: Application to BHB Analysis by LC-MS/MS and GC-MS

Brigitte Desharnais, Marie-Jo Lajoie, Julie Laquerre, Stéphanie Savard, Pascal Mireault and Cameron D. Skinner

Journal of Analytical Toxicology **43** (7) (2019) 512-519 [3].

- Brigitte Desharnais developped the procedure to perform endogenous content correction;
- Cameron D. Skinner had the idea of performing a theoretical derivation of the error folded in by the correction process;

- Brigitte Desharnais derived the mathematical formulas calculating the error on the estimated concentrations for the corrected and uncorrected concentrations;
- Stéphanie Savard had the idea of evaluating the error brought by this process via the experimental route;
- Marie-Jo Lajoie and Stéphanie Savard performed experimental data collection;
- Brigitte Desharnais analyzed the production data collected;
- Julie Laquerre had the idea of developping a generalized version of the correction R script (robust to missing data points or different data input formats) and did some preliminary work on that script;
- Brigitte Desharnais coded the final version of the R scripts (Appendix D);
- Marie-Jo Lajoie wrote a first draft for Section 3.3.2 ("LC-MS/MS BHB analysis");
- Brigitte Desharnais wrote the manuscript;
- Pascal Mireault and Cameron D. Skinner pointed to areas of improvement and reviewed the final manuscript.

In the manuscript presented in Chapter 4.1:

Qualitative Method Validation and Uncertainty Estimation Via the Binary Output
I —Validation Guidelines and Statistical Foundations

Félix Camirand Lemyre[1], Brigitte Desharnais, Julie Laquerre, Marc-André Morel, Cynthia Côté, Pascal Mireault and Cameron D. Skinner

Submitted to *Journal of Analytical Toxicology* (2019) submission number JAT-19-2881 [4].

- Brigitte Desharnais developped the idea of qualitative method validation based on binary results;
- Brigitte Desharnais performed the literature review to catalogue qualitative method validation procedures, with some initial help from Maxime Gosselin;

[1]Félix Camirand Lemyre and Brigitte Desharnais contributed equally to the manuscript and are listed in alphabetical order.

- Brigitte Desharnais developped a qualitative decision point validation procedure applicable in forensic toxicology;
- Brigitte Desharnais wrote the R scripts performing data modeling and simulating decision point qualitative methods behavior (Appendix E.2);
- Félix Camirand Lemyre and Julie Laquerre discussed ideas and pointed to potential issues or improvements at the mathematical and analytical chemistry levels, respectively;
- Félix Camirand Lemyre reviewed the data modeling and contributed to data interpretation;
- Julie Laquerre, Marc-André Morel and Cynthia Côté developped the multi-analyte extraction and LC-MS/MS analysis method validated in this manuscript and performed the experimental data collection;
- Brigitte Desharnais wrote the manuscript;
- Pascal Mireault and Cameron D. Skinner volunteered some ideas, pointed to areas of improvement, and reviewed the final manuscript.

In the manuscript presented in Chapter 4.3:

Qualitative Method Validation and Uncertainty Estimation Via the Binary Output
II —Application to a Multi-Analyte LC-MS/MS Method for Oral Fluid

Brigitte Desharnais[2], Marie-Jo Lajoie, Julie Laquerre, Pascal Mireault and Cameron D. Skinner

Submitted to *Journal of Analytical Toxicology* (2019) submission number JAT-19-2882 [5].

- Marie-Jo Lajoie and Julie Laquerre performed the initial literature review;
- Julie Laquerre developped the oral fluid extraction and analysis method;
- Brigitte Desharnais and Julie Laquerre developed the method validation plan;

[2]Brigitte Desharnais, Marie-Jo Lajoie and Julie Laquerre contributed equally to the manuscript and are listed in alphabetical order.

- Julie Laquerre and Marie-Jo Lajoie performed experimental data collection and analysis;
- Brigitte Desharnais supervised the design of experiment (DoE) optimization of oral fluid extraction and interpreted the results in collaboration with Julie Laquerre;
- Marie-Jo Lajoie and Julie Laquerre drafted the majority of the manuscript;
- Brigitte Desharnais designed all data visualization instances;
- Brigitte Desharnais supervised redaction of the manuscript and performed extensive editing;
- Brigitte Desharnais, Marie-Jo Lajoie and Julie Laquerre settled on alphabetical order of the authors by mutual agreement due to the critical nature of every party's work. Without the qualitative validation framework, DoE optimization, data visualization work and extensive writing work of Brigitte Desharnais, this manuscript would not exist. Without the literature review, method development, and experimental work of Marie-Jo Lajoie and Julie Laquerre, this manuscript would not exist. These three authors thus agreed alphabetical order with a mention of equal contribution was satisfying;
- Pascal Mireault supervised the project and reviewed the final manuscript;
- Cameron D. Skinner reviewed the final manuscript.

In the manuscript presented in Chapter 5:

Postmortem Estimation of Metabolic Capacity Through Cytochrome P450 Enzyme Characterization and Quantification —A Proof of Concept

Brigitte Desharnais, Pascal Mireault and Cameron D. Skinner [6].

- Brigitte Desharnais and Cameron D. Skinner developped the idea of postmortem metabolic capacity estimation via a proteomics route;
- Brigitte Desharnais developed experimental procedures and performed the data analysis;

- Brigitte Desharnais carried out all of the experimental work from July 2013 to January 2017;
- Lucie Vaillancourt and Brigitte Desharnais shared the execution of the experimental work from January 2017 to December 2018, with Brigitte Desharnais carrying out the ultracentrifugation and LC-MS/MS analysis while Lucie Vaillancourt carried out the homogeneization and microsomal fraction treatment part, executing protocols written by Brigitte Desharnais;
- Brigitte Desharnais wrote the manuscript;
- Pascal Mireault and Cameron D. Skinner supervised the project, volunteered some ideas, pointed to areas of improvement, and reviewed the final manuscript.

Contents

List of Figures

List of Tables

List of Abbreviations

- AA: amino acid;
- AAFS: American Academy of Forensic Sciences;
- ACN: acetonitrile;
- ANOVA(-LoF): analysis of variance (- lack of fit);
- AQUA: absolute quantification;
- ASA: American Statistical Association;
- BHB: β-hydroxybutyric acid;
- CE: collision energy;
- CO: cut-off;
- CNS: central nervous system;
- CVM: Cramer-von Mises;
- CXP: cell exit potential;
- CYP: cytochrome P450;
- Da: Dalton;
- DFSA: drug facilitated sexual assault;
- DNA: deoxyribonucleic acid;
- DoE: design of experiments;

- DP: decision point;
- DP: declustering potential;
- DUID: driving under the influence of drugs;
- emPAI: exponentially modified protein abundance index;
- EP: entrance potential;
- ER: endoplasmic reticulum;
- ESI: electrospray ionization;
- FDA: Food and Drug Administration;
- FN(R): false negative (rate);
- FP(R): false positive (rate);
- GC-MS: gas chromatograph coupled to a mass spectrometer;
- GHB: γ-hydroxybutyric acid;
- GRAVY: grand average of hydropathy;
- HbCO: carboxyhemoglobin;
- HLM: human liver microsome;
- HPLC: high performance liquid chromatography;
- ICAT: isotope coded affinity tag;
- IS: internal standard;
- ISO: International Standards Organization;
- KS: Komogorov-Smirnov;
- LC-MS(/MS): liquid chromatograph coupled to a (tandem) mass spectrometer;
- (L)LOQ: (lower) limit of quantification;
- LURL: lower unreliability zone limit;
- MPPGL: microsomal protein per gram of liver;

- MRM: multiple reaction monitoring;
- MS: mass spectrometry;
- OF: oral fluid;
- PEG: polyethylene glycol;
- PMI: postmortem interval;
- PTM: post-translational modifications;
- Q1 (Q3): first or last quadrupole of an MS/MS system;
- QC: quality control;
- QconCAT: quantitative concatamer;
- RLR: reliability rate;
- %RSD: relative standard deviation, expressed as a percentage;
- SIL(-IS): standard isotope-labeled (- internal standard);
- SLR: selectivity rate;
- SNR: sensivitiy rate;
- SPE: solid phase extraction;
- SRM: single reaction monitoring;
- SWGTOX: Scientific Working Group for Forensic Toxicology;
- TCEP: tris(2-carboxyethyl)phosphine;
- TIC: total ion chromatogram;
- THC: Δ^9-tetrahydrocannabinol;
- THC-COOH: nor-9-carboxy-$\Delta 9^9$-tetrahydrocannabinol;
- THC-OH:11-hydroxy-$\Delta 9^9$-tetrahydrocannabinol;
- TN: true negative;
- TP: true positive;

- (U)LOQ: (upper) limit of quantification;
- UM: uncertainty of measurement;
- UR: unreliability zone;
- UURL: upper unreliability zone limit;
- WT: wild type.

1

General Introduction

1.1 Forensic toxicology

Forensic toxicology can be defined as the branch of science dedicated to the study of drugs, toxins and poisons for the purpose of answering questions arising during judicial proceedings [7, 8]. On a day-to-day basis, forensic toxicology departments analyze different biological matrices (blood, urine, vitreous humour, gastric content, organs, etc.) to determine the presence and quantity of legal and illegal drugs, poisons, as well as levels of certain key biochemical parameters (e.g., glucose, urea, creatinine) [8]. Results from these analyses are used, amongst other purposes, to contribute to the evidence of driving while impaired, explain circumstances surrounding a sexual assault and determine causes and/or circumstances of death.

Postmortem forensic toxicology is concerned with demonstrating one cause of death only: fatal intoxication (although the toxicant itself can vary). On the other hand, several relevant circumstances surrounding death might be highlighted, such as altered mental state or non-compliance with prescribed medication. One such interesting circumstance is divergent metabolism and reaction to drug intake in certain individuals [9–11].

1.2 Drug metabolism and metabolic capacity

1.2.1 Metabolism and cytochrome P450 (CYP) enzymes

Drugs, be they legal or illegal, are foreign substances entering the human biological system. As such, they can also be identified as xenobiotics [12]. Once in the human body, these compounds will undergo chemical transformations grouped under the general rubric of "metabolism", which aims to reduce the toxicity of the compound or favour its elimination by rendering it more hydrophilic [13]. In the first step, also denoted as Phase I metabolism, the lipophilic substrate is converted to a functionalized metabolite by oxidation, reduction or hydrolysis. In humans, these transformations are mostly carried out by a class of enzymes known as the cytochrome P450 (CYP) enzymes [13, 14]. In Phase II metabolism, the functionalized metabolite is coupled to a highly hydrophilic compound, such as a glucuronide, by several conjugating enzymes (UDP-glucuronosyltransferase, sulfotransferases, glutathione S-transferases, etc.) [13]. For example, as shown in Figure 1.2.1, one of codeine's[1] metabolism pathways, is to first undergo O-demethylation to morphine by the CYP 2D6 enzyme, and then conjugation to a glucuronide. Excretion will occur at different stages of the metabolism.

CYP enzymes are responsible for metabolizing dietary constituents, endogenous chemicals and xenobiotics [13, 16]. This family of enzymes carries out about 75% of all Phase I drug metabolism [16]. CYPs can be found in different locations in the body such as the uterus, breasts, ovaries, testes, prostate, adrenal glands, liver and intestines [16]. However, the most important drug metabolizing CYPs are essentially situated in the centrilobular area of the liver and found embedded in the intracellular smooth endoplasmic reticulum (ER) membrane [16]. CYP 2C9, 2C19, 2D6, 2E1 and 3A4 play the most important roles in drug metabolism [14] (Figure 1.2.2). All of these enzymes are polymorphic, meaning that one or more specific variations in the DNA sequence can be found in at least 1% of the population [17]. Identified CYP polymorphisms are numerous, with at least 105 known for CYP 2D6 alone [18]. These gene variations can lead to amino acid changes in the final enzyme, splicing defects or gene duplication that, in turn, can lead to enzymes with decreased or increased function, total dysfunction, absence of expression or overexpression of the enzyme [18]. Epigenetic or environmental factors such as age, sex, disease state, habituation and presence of inducers or inhibitors can also lead to a modified CYP expression level [9, 16, 19–22]. In both cases, this can have an important impact on the rate at which

[1]The interested reader will find structures of, and relevant information for, the small molecules named throughout this thesis in Appendix A.

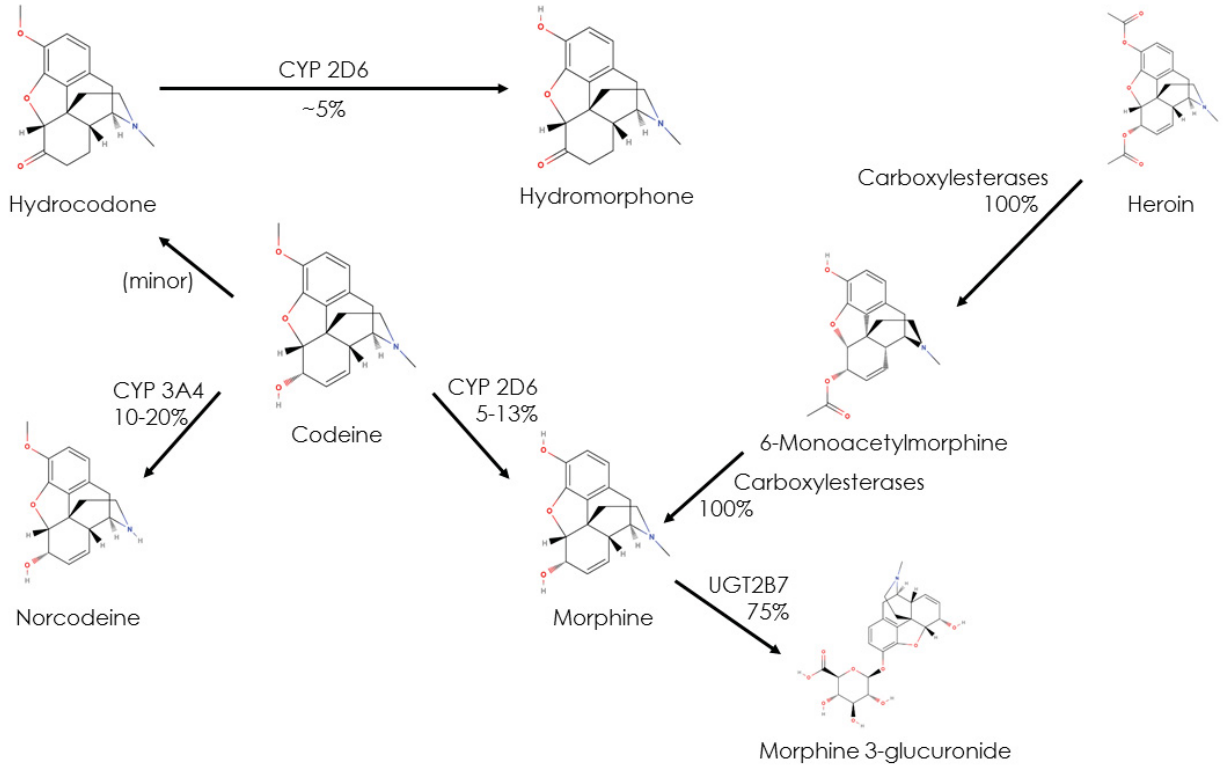


Figure 1.2.1: Opiates metabolism, including percent transformation and metabolizing enzymes, adapted from [15].

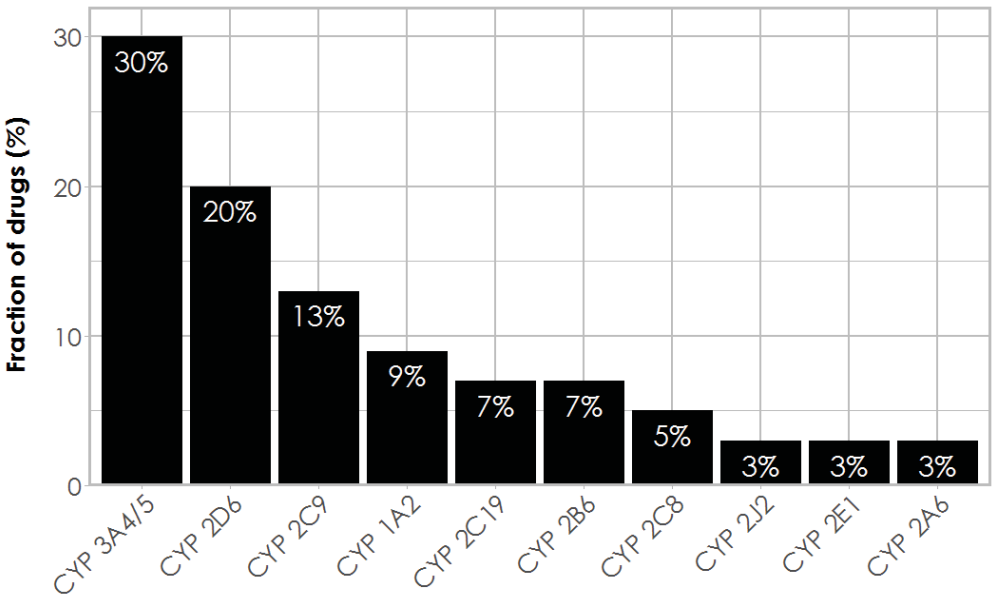


Figure 1.2.2: Fraction of drugs metabolized by CYP isoforms, adapted from [14].

metabolism of specific drugs is performed by individuals.

1.2.2 Metabolic capacity

Metabolic capacity refers to the rate of xenobiotic transformation, or metabolism, in the human body. This is a feature of interest for clinical and forensic toxicologists alike, because an individual's whose metabolic capacity is significantly different from the average could be more vulnerable to a fatal intoxication event [11, 23]. For example, a patient with a very low metabolic capacity will see the blood concentration of the drug rise as doses are given, eventually reaching a toxic level, because biotransformation and elimination occurs more slowly than the intake. A rapid metabolizer might face a similar fate when dosed with a pro-drug, such as codeine, which generates the active compound morphine when metabolized (Figure 1.2.1). Thus, knowledge of the metabolic capacity can provide forensic toxicologists with insight in fulfilling their role of establishing the relevant circumstances surrounding death [9–11].

What solutions have then been found to measure metabolic capacity? In the clinical domain, this is typically done by administering a set of innocuous probes and monitoring the pharmacokinetic curves of the probes and their metabolites [24]. For obvious reasons, this is not an option in postmortem cases. Instead, the current approach in the forensics context is

to perform genotyping of the individual to identify the presence of one or more deleterious mutations in the CYP encoding gene [11, 23]. According to the number and impact of the mutations, individuals are classified according to their metabolizer type (from slowest to fastest): poor, intermediate, extensive (normal/typical) or ultra-rapid metabolizer [23] (see Chapter 5, Figure 5.2.1a for more details). The genotyping approach can be first traced back to 1990 in pharmaceutical research [25], and has grown to cover nearly all the drug metabolizing CYPs: 1A2 [26], 2A6 [27], 2C8 [28], 2C9 [29], 2C19 [30], 2D6 [31], 2E1 [32], 3A4 and 3A5 [33]. These techniques were introduced into forensics in 1999 [9], and several forensic toxicology papers have been published on the topic [9, 10, 17, 33–42]. Yet, use of this technique remains extremely marginal in the field.

Two reasons might explain this fact. First and foremost, the genotype obtained via DNA analysis is certainly a relevant factor in predicting phenotype – the characteristics exhibited by an individual such as eye colour, height or metabolic capacity [43]. But while DNA might be the body's instruction manual, the actual outcome dictating the phenotype is the proteome (which is dependent on the genome, but is also highly dependent on several other factors). Thus, the correlation between the genetically determined metabolizer type and the observed outcome (metabolic capacity) is poor [11]. Second, most forensic toxicology laboratories do not have the instrumental, financial or human resources necessary to develop a DNA genotyping method and perform method validation according to accreditation standards, as it is outside of the scope of their usual practice and expertise (i.e., small molecule analysis). Globally, forensic toxicology laboratories thus have very little incentive to divert precious resources into a method where it is unclear if the result will be helpful. However, the inherent problems with the genotyping approach reveal an opportunity to kill two birds with one stone: developing a method yielding a more accurate estimation of the metabolic capacity via a proteomics approach and one that can be run on standard forensic toxicology instrumentation.

Thus, the general aim of this thesis is to demonstrate, at the proof-of-principle level, a liquid chromatography – tandem mass spectrometry (LC-MS/MS) method to characterize and quantify CYP enzymes in postmortem liver tissue in order to estimate antemortem metabolic capacity in deceased individuals. Characterizing the CYP enzymes will allow information about genetic polymorphisms to be collected, while quantifying the CYP enzymes will provide information on the expression level. Taken together, both pieces should allow for a more accurate estimation of the metabolic capacity than genotyping alone. Interestingly, this approach is also one that could be carried out relatively easily by forensic toxicology

laboratories, since it relies on their standard knowledge and instrumentation: liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). Although ultimately, a comprehensive metabolic capacity determination method would cover all xenobiotic metabolizing enzymes (CYPs and others), as a reasonable first step CYP 2D6 and CYP 3A4 enzymes were targeted due to their importance in the drug metabolizing process (as shown in Figure 1.2.2).

1.3 Cytochrome P450 analysis

1.3.1 CYP enzymes and their classification

CYPs are enzymes, i.e., proteins with catalytic properties [13]. These chains of amino acids are identified by a family number followed by a subfamily letter and gene number (e.g., CYP 3A4) [16]. CYPs sharing greater than 40% sequence identity belong to the same family, while those sharing greater than 55% sequence identity also belong to the same subfamily [16]. A common gene number is attributed when the enzymes have the same function and high gene conservation [16]. Being located in the rough endoplasmic reticulum membrane, these hemoproteins are typically labeled as highly hydrophobic [44] (Figure 1.3.1). In fact, with an average GRAVY score (grand average of hydropathy, ranging from the most hydrophilic, -2, to the most hydrophobic, +2) hovering around -0.3, the amino acid sequence hardly justifies this label. The key to this apparent contradiction is that membrane proteins are "highly hydrophobic" only in their folded states. Considering their location in the cell, they will need a good hydrophobic anchor, but will also require several hydrophilic parts to carry out their functions. Nevertheless, the hydrophobic character of CYPs prior to denaturation and solubilization does complicate the extraction and analysis process, as most of the protocols were developed for more hydrophilic proteins.

Traditional approaches for CYP quantification, such as Western blotting and enzyme activity assays, are not suitable for the proposed application. Antibodies are challenging and expensive to develop, and often are not highly specific [46, 47]. Additionally, this approach is not readily amenable to high throughput. As for enzyme activity assays, there is no telling how it might have been affected by the postmortem changes such as denaturation, cell lysis and pH decrease [48].

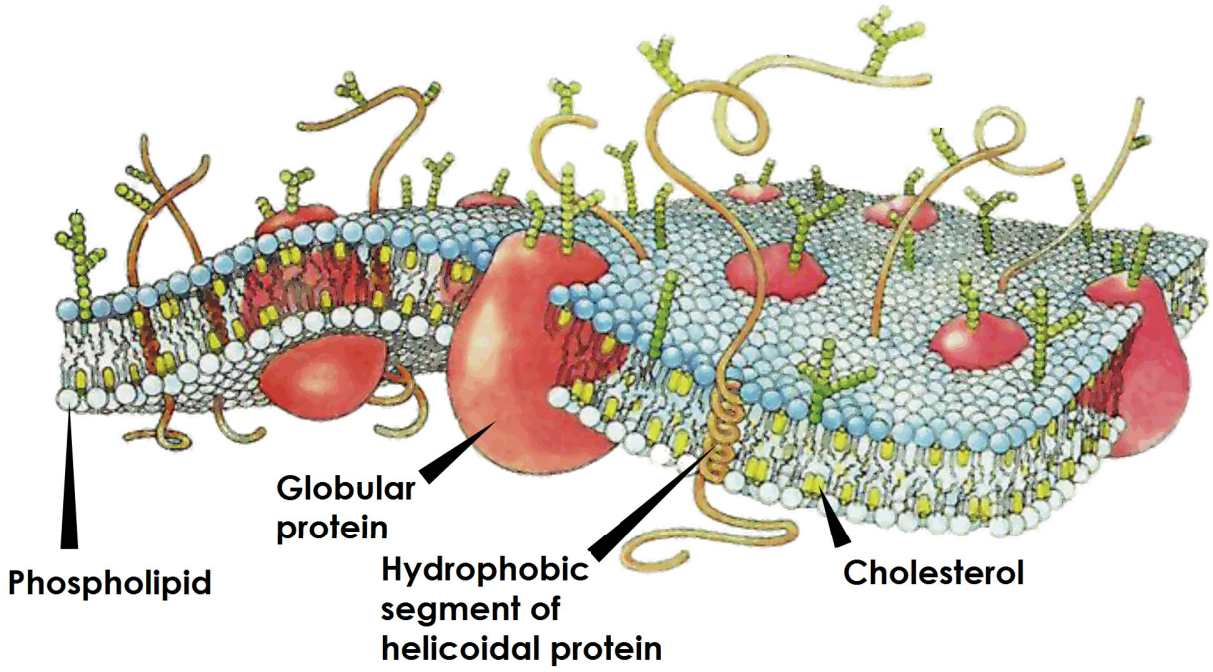


Figure 1.3.1: Cell membrane and membrane proteins, adapted from [45].

1.3.2 Protein extraction and isolation

For CYPs, a mass spectrometry approach seems to be the most efficient route to characterize and quantify the enzymes using tools readily available to the forensic toxicologist. However, these proteins need to be first extracted from the liver tissue. This is typically done by tissue homogenization which is often accompanied by cell lysis. Several common approaches are available. Tissue homogenization can be performed using a mechanical blender such as the Polytron®; a Potter-Elvehjem device, which compresses the tissue between a Teflon pestle and the glass wall of the container; sonication, which uses high-frequency energy to disrupt membranes by cavitation; cryo-grinders, which reduce frozen samples to a powder using a ball and mill process [49]. To complete cell lysis, two common options are the use of a salt gradient (water enters the cells until they rupture) and detergents (disruption of the hydrophobic membrane components) [49]. These tissue preparation steps also release proteases from cell compartments, hence protease inhibitors and a cold environment are necessary to preserve the integrity of proteins until further isolation is performed [50].

Following tissue homogenization, some form of protein isolation is required as very few analytical methods are capable of dealing with the complexity of the entire proteome com-

ing from a tissue sample, much less in a quantification context. In the case of CYPs, it is possible to take advantage of the fact that as the cell is disrupted fragments of the ER form small micelle-like structures called microsomes. Isolation of CYP rich microsomes typically requires an ultracentrifugation step due to the very small density difference between the microsomes and surrounding cell contents [51]. Another possible avenue is the use of immunoaffinity pulldown [52] which targets CYP enzymes directly. However this option was not favoured since it was more arduous to establish in a forensic toxicology laboratories and is accompanied by a hefty price tag (several hundred dollars per analysis). Once isolation is performed, chaotropes (e.g., guanidine hydrochloride, urea, thiourea, guanidine thiocyanate) and detergents (e.g., TritonTM, CHAPS, Brij®, octyl β-D-glucopyranoside, n-dodecyl-β-D-maltoside, OTG, SDS) can help solubilize the hydrophobic proteins by destabilizing their hydrogen bonds and denaturing the protein, thus exposing the more hydrophilic sites of the protein to the solution [53].

1.3.3 Protein analysis by tandem mass spectrometry

Protein analysis by mass spectrometry can be divided into two main categories. In top-down proteomics, intact proteins are injected into the mass spectrometer for analysis. This allows certain interference problems to be circumvented, and post-translational modifications to be characterized [54]. In bottom-up, or shotgun proteomics, proteins undergo a proteolytic digestion in order to cleave the protein into short chains of amino acids (peptides), which are then analyzed by mass spectrometry [54]. Due to the instrumentation selected (LC-MS/MS, unit resolution) and the targeted goal of identifying point mutations on the protein, bottom-up proteomics was more appropriate.

Several digestion enzymes are available for this purpose, such as chemotrypsin, thermolysin, pepsin, elastase, lysine-C and-N, and the most widely used protease in bottom-up analyses: trypsin [55]. Trypsin cleaves the peptide bond (see Chapter 5, Figure 5.4.1a) after the lysine and arginine amino acids, except if they are followed by a proline amino acid (due to steric hindrance) [55]. The amino acid sequence and mass of peptides obtained following digestion can thus be predicted based on the known behaviour of the protease, and informatic tools exist to carry out this prediction [56]. However, they cannot predict the presence of missed cleavages, which occur at 15 to 30% of the cleavage sites, nor the presence of chemical modifications (although some amino acids are known to be more prone to chemical modifications than others). Because proteases are themselves enzymatic proteins, the digestion environment is limited in terms of detergent, organic solvent, chaotropic agent

and salt concentration. Outside of these limits, excessive denaturation, inactivation of the protease and autolysis prevents successful protein analysis.

Successful MS detection of the proteolytically produced peptides requires that they are presented to the MS in a simplified matrix, preferentially one that enhances ionization, minimizes matrix effects and is isolated from the detergents and salts needed for digestion. The latter are particularly problematic because they suppress the analyte signal and are deleterious to the mass spectrometry instrumentation. In LC-MS/MS, the chromatography allows separation of the analytes from the matrix components based on their affinity for the stationary phase [57]. The effluent is constituted primarily of the mobile phase, buffer salts, the analytes and co-retained concomitant species. This mixture is fed to an electrospray ionization source (ESI), in which a positive (ESI+) or negative (ESI-) charge is applied to droplets formed by nebulization. Solvent evaporation by contact with heated gas causes the charge to be transferred to the analytes and any co-eluted species (via Coulombic explosions) [58]. The charged species are then swept into the mass spectrometer itself (Figure 1.3.2a), where a first quadrupole mass analyzer will filter out molecules not having the mass to charge ratio of the targeted analyte. These transmitted molecules will be fragmented in a collision cell by an inert gas, and transmitted to another quadrupole mass analyzer. Fragments generated by the targeted peptide or analyte are known, either via the analysis of a certified reference material (in the case of small molecule analysis), or via in silico prediction (in the case of peptide analysis, where fragmentation follows well defined patterns, Figure 1.3.2b). In single/multiple reaction monitoring (S/MRM) mode, only fragments with a mass corresponding to known fragments of the analyte are allowed through to the ion detector, generating a high specificity signal for the analyte targeted [59].

1.3.4 Protein quantification

Protein quantification can be separated into two broad areas. Relative quantification is typically used to compare the abundance of a vast array of proteins in normal and altered states in a single experiment [60, 61]. For example, hepatocytes (liver cells) could be prepared in a standard growth medium and one supplemented with an inducer. Following lysis and digestion, relative quantification should reveal a large increase in the concentration of CYPs in the medium with the inducer. Here, absolute concentrations are not measured; rather the fold difference relative to the normal state is estimated. Because our goal is to develop a CYP quantification method where results can be compared from one batch to the next

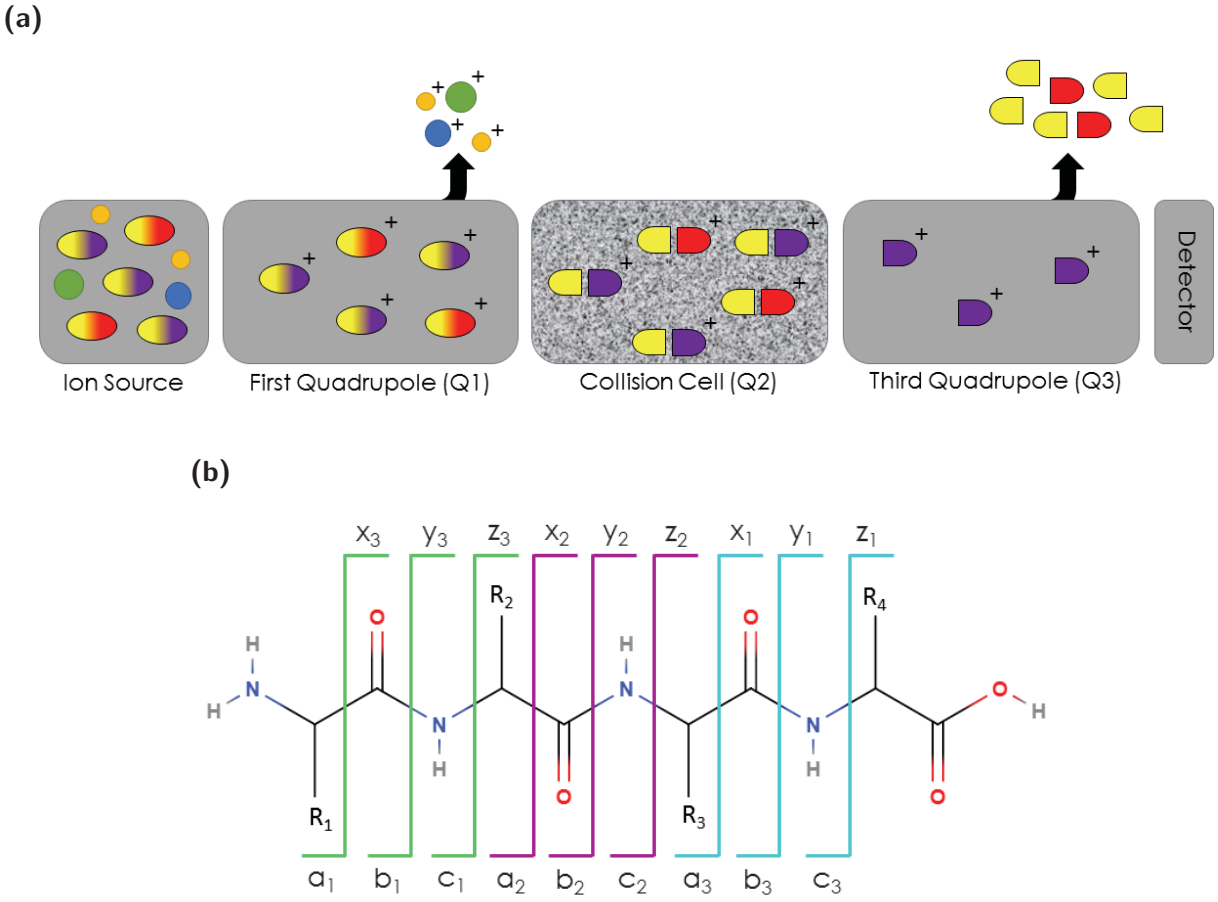


Figure 1.3.2: Tandem mass spectrometry and peptide analysis. **(a)** Schematic of a tandem mass spectrometer operated in single/multiple reaction monitoring (SRM/MRM) mode. Molecules are ionized in the source, filtered according to the mass-to-charge ratio in the first quadrupole, fragmented in the collision cell, and filtered again according to mass-to-charge ratio of the fragment in the third quadrupole. **(b)** Peptide fragmentation pattern. A peptide will fragment in the collision cell along the lines indicated here, with the charge being retained on the N-terminal side of the peptide (a, b, c ions) or the C-terminal side of the peptide (x, y, z ions). y-ions are typically the most abundant and relied upon for quantification and identification of peptides.

and between forensic toxicology laboratories, relative quantification is not relevant for this project. Absolute quantification needs to be used instead, where the output is an absolute concentration in a defined set of units. Absolute quantification can be achieved through a variety of calibration strategies, as shown in Table 1.3.1.

Table 1.3.1: Absolute quantification strategies, adapted from [62]

Stable Isotope-Labeled Solutions
Absolute quantification with stable isotope-labeled internal standards (AQUA)
Quantitative concatamer (QconCAT)
Equimolarity through equalizer peptide (EtEP)
Protein standard absolute quantification (PSAQ)
Full length expressed stable isotope-labeled proteins for quantification (FLEXIQuant)
Absolute stable isotope labeling through amino acids in cell culture (SILAC)
Protein epitope signature tags (PrEST)

Label-Free Solutions
Exponentially modified protein abundance index (emPAI)
Absolute protein abundance index (APEX)
Normalized spectral abundance factor (NSAF)
Normalized spectral intense peptides per protein (SIN)
Intensity-based absolute quantification (IBAQ)

Although label-free techniques are much less expensive to implement, they generally suffer from much lower precision, accuracy, sensitivity and reproducibility compared to stable isotope-labeled techniques in complex biological matrices [62]. Since the projected method was to be validated with strict precision, accuracy and reproducibility criteria, label-free solutions were a non-starter. In SIL solutions, AQUA uses synthetic versions of peptides targeted for quantification purposes. These peptides are synthesized with stable isotope-labeled amino acids resulting in an increase in molecular weight (typically +4 to +10 Da) that is readily differentiated by the mass spectrometer [62]. These internal standards (IS) will co-elute with the native peptides, and the peak area of the analyte divided by the peak area of the IS (peak area ratio) is a response signal that is proportional to the concentration of native peptide (and, with some limitations, the protein concentration). This technique is thus suited to focused, quantitative analysis of protein expression and modifications [63]. Even though it is expensive compared to label-free techniques, it is widely available, com-

mercial, and yields a more accurate result than other SIL techniques when used with an external (as opposed to internal) calibration [64]. Concatenated peptides (QconCATs) are artificial proteins consisting of up to about 50 concatenated peptide sequences produced by expression of a synthetic gene [62]. This protein is co-digested with the analyte protein(s), generating equimolar SIL peptides. QconCAT can be very advantageous when multiplexing is required, but their cost is still high and complete proteolysis must be confirmed. Equalizer peptides (EtEP) are standard SIL-IS with an additional N-terminal sequence, the equalizer peptide, which will be cleaved upon digestion [62]. This technique avoids the need for quantification of the SIL-IS, as long as internal calibration is used. Full-length SIL standards are also available, with proteins produced *in vitro* (PSAQ), *in vivo* (SILAC), in wheat germ extracts (FLEXIQuant) [62]. A full-length SIL protein would appear to be the gold standard, compensating for differential digestion efficiency and possessing nearly identical physicochemical characteristics to the target protein. However, these proteins are expensive, and their promise relies on the fact that the structure and modification status is the same as the native protein, which is far from guaranteed with a membrane protein like CYPs. This might be why recent studies have shown that full length proteins can actually perform worse than AQUA in terms of accuracy [64]. Of the stable isotope methods, the AQUA approach was more adapted to the projected method due to the limited multiplexing needs and limited access to molecular biology facilities to produce full length labelled proteins. The obvious question is which tryptic peptides should be selected for quantification purposes amongst the >30 produced by the tryptic digestion of a CYP enzyme. Rules for selection of quantotypic are discussed in Table 5.4.1.

1.3.5 CYP analysis in a forensic setting

The available literature on the topic of CYP quantification was limited, but unanimously reported that CYP quantification was achievable, albeit through different methodologies [44, 47, 65–70] (a more in depth literature review is available in Section 5.2). At the outset, the largest preoccupation for the projected method seemed to be which adjustments would be necessary to apply these methodologies to samples differing in one important aspect: they were collected postmortem instead of ante or perimortem. This included investigating protein degradation events, as detailed in Section 5.4.3.

Furthermore, a CYP analysis methodology aimed for application in the forensics field will have not only to use tools and techniques common to the trade, but also be as simple and as fast as possible to be compatible with the heavy production burden faced by laboratories.

Additionally, since the majority of forensic toxicology laboratories nowadays are accredited under one or more set of norms, such as ISO 17025 ("General requirements for the competence of testing and calibration laboratories") [71, 72] and CAN-P-1578 ("Guidelines for the accreditation of forensic testing laboratories", a Canadian norm) [73], methods must meet stringent validation requirements. This includes satisfying criteria with regards to accuracy, intra- and inter-day precision, absence of interferences and process efficiency. To adequately demonstrate that these demands are met, some challenges in method development and validation must be addressed.

1.4 Mathematical aspects of experimental quantification

1.4.1 Calibration curves

The basic principle on which quantification operations rely is the proportionality of the measured signal with the analyte concentration [74]. In chromatographic methods, the measured signal used is typically the area of the chromatographic peak (Figure 1.4.1a). When a mass spectrometer is used as a detector, this corresponds to the number of ions that trigger the detector.

For any measurement (e.g., signal units) made on a sample to be converted to units useful for interpretation and comparison purposes (e.g., concentration), a calibration model is built. Ideally, samples blank of the targeted analyte(s) are used to prepare the standards by spiking the analyte(s) at different known concentrations. The dynamic range spans from the lower limit of quantification (LLOQ, lowest useful calibration standard) to the upper limit of quantification (ULOQ, highest useful calibration standard) [75]. Two related concepts are the detection limit (DL or LOD), which is the "smallest measure that can be detected with reasonable certainty" [76], and the limit of quantification (LOQ), which is the lowest concentration which can be detected and quantified with predefined bias and precision goals [75]. From these calibration standards, a mathematical model describing this data set of (concentration, response) pairs can be obtained via least squares regression (Figure 1.4.1b), which minimizes the sum of squared differences with the signal predicted by the mathematical model [77]. This mathematical model in turn allows estimation of the concentration in an unknown sample in which a signal y_i is measured, by isolation of the x_i term. This standard, unweighted linear regression model assumes (in order of importance) that the true function is linear, that measurement errors are independent from one sample to the next, that the error is homoscedastic and normally distributed.

Ideally, the calibration standards used to build the calibration curve would be prepared in the same matrix type as the unknown samples [74]. For example, if cocaine is to be measured in postmortem blood samples, calibration standards should be prepared in postmortem blood from which cocaine is absent (blank postmortem blood). This practice, known as matrix-matched standards, compensates for the majority of matrix effects from concomitant components (metabolites, lipids, etc.) [74, 78]. Thus the calibration standards and the unknown samples should produce similar responses for similar concentrations, yielding a more accurate quantification [74, 78].

Another good practice in calibration operations is to prepare and measure quality control (QC) samples. QC samples are, to the best ability of the analyst, samples of known concentration identical to unknown samples, generally prepared at $\approx 20\%$, $\approx 50\%$ and $\approx 80\%$ of the calibration dynamic range [75, 79]. In practice, QC samples are prepared from a matrix matched blank sample spiked with the analyte(s) at the desired levels. These samples are then treated as unknown samples would, and the quantification result is compared with the expected result to obtain a bias value. Bias values considered as acceptable will vary from one reference to the other [75, 79–81], but SWGTOX states that the maximum acceptable bias is $\pm 20\%$ at all concentration levels.

1.4.2 Internal standards

Simply put, internal standards are trackers added in equal concentration to all calibration standards, quality controls and unknown concentration samples [74] (Figure 1.4.1a). These trackers permit correction for various manipulation errors and experimental variations (e.g., adsorption to the walls of a container, sensitivity changes, ionization suppression etc.) [82]. To track adequately all of these phenomena, an internal standard should behave as closely as possible to the targeted analyte, i.e., possess identical physico-chemical characteristics. Although this level of perfection is hard to attain, one solution which comes close to it is the use of stable isotope-labeled (SIL) internal standards [82] (Figure 1.4.2). Typically, hydrogen (^{1}H) or carbon (^{12}C) atoms in the analyte's structure are replaced with a heavier isotope (e.g., ^{2}D or ^{13}C) in the internal standard. These IS thus have almost identical physico-chemical characteristics, but are distinguishable by mass spectrometry by virtue of their increased mass.

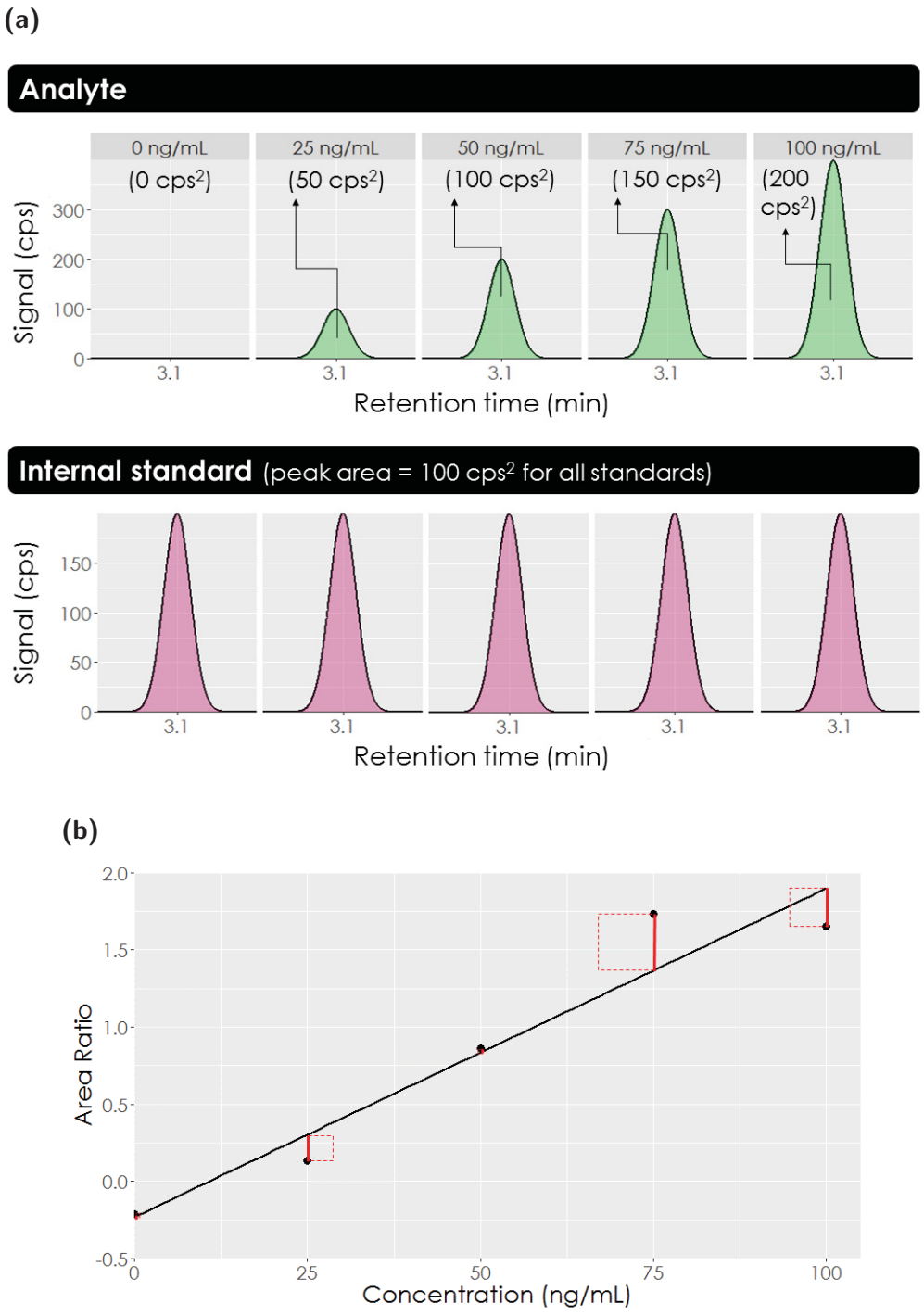


Figure 1.4.1: Chromatographic measurements in a series of standards and derived calibration curve (simulated data). **(a)** Chromatographic peak areas (in parentheses) for the analyte and IS in a series of spiked standards. Area ratio for e.g., 75 ng/mL standard, is 150 $cps^2/100$ cps^2=1.5. **(b)** Calibration curve and least squares regression.

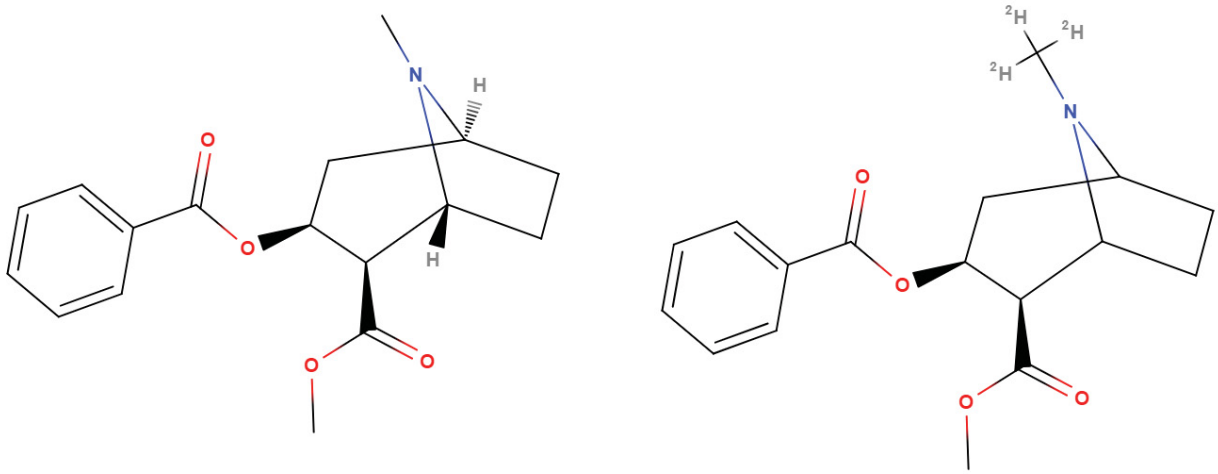


Figure 1.4.2: Cocaine (targeted analyte, 304 Da) and its stable isotope-labeled internal standard (SIL-IS), cocaine-D_3 (307 Da).

Although the use of SIL-IS is considered the gold standard, for costs and availability reasons, this might not always be possible. An imperfect alternative is to use analogues, which can take different forms: for example, fluoroamphetamine can be used as an IS for amphetamine [82]. In the realm of proteomics, a reporter protein not present in the targeted sample such as glutamate dehydrogenase can be used for liver CYPs [83]. Critically, the analogue IS used should not be endogenous. Ideally, it would also display as close physico-chemical characteristics to the analyte (hydrophobicity, ionization sensitivity) to the analyte as possible, and be commercially available at a reasonable price. Satisfying all these requirements for CYPs is no mean feat, and could be satisfied by a non-endogenous membrane protein with some reasonably hydrophobic sequences. Although a necessarily imperfect solution, glutamate dehydrogenase strikes somewhat of a balance between all of these requirements.

Whatever the avenue selected, this tracker is used to normalize the signal measured from the targeted analyte. Provided that a constant amount of IS is used throughout the analysis, the ratio is proportional to the analyte concentration (Figure 1.4.1a). Instead of using the peak area, the ratio of the analyte peak area to the internal standard peak area [74] is used as the response variable for calibration and quantification. As before, a least-squares regression can be performed on the responses versus the known, spiked, concentrations to establish the mathematical relationship between response and concentration.

1.4.3 Adjustments to the standard least squares regression

However, the run-of-the-mill least squares regression can and should be adjusted to take into account further knowledge developed about the data at hand. First amongst this is the behaviour of the error across the calibration range. Error (noise), which can be defined as the undesired "fluctuations in the desired signal which obscure [the] measurement" [84], can be found in different forms. Homoscedastic (constant absolute error) and heteroscedastic (absolute error varies with concentration) can both be encountered in analytical methods [84] (although heteroscedastic data is more frequent in modern mass spectrometry analysis methods). A full description of the sources and causes of error is beyond the scope of this thesis, but three types need attention. The first, and the simplest to consider mathematically, is constant noise superimposed upon the signal. The standard linear regression approach assumes this case and seeks to minimize the overall error. The literature calls this type of linear regression unweighted regression, but a more accurate view is to call it uniformly weighted regression since the difference between the expected signal (calibration line) and the observed signal is given equal weight for all standards in the regression [85]. The second noise type increases in a root proportional relationship to the signal while the last increases in a proportional relationship to the signal [84]. These noise behaviours can originate from fundamental processes, such as shot noise, or from non-fundamental processes such as flicker noise [84]. Of importance here is that different calibration standards have different precisions. Such a behaviour should be taken into account when calculating the calibration curve regression constants. Additional gains in accuracy, compared to uniform weighted regression, are possible by appropriate weighting of the data [85]. Noise characterization and selection of weighting is further introduced in Section 2.1.4.

Another adjustment which can be made is switching from a linear calibration model ($y = b_1x + b_0$) to a quadratic one ($y = b_2x^2 + b_1x + b_0$) (increasing the model order) [86]. In the simplest theory, analyte responses used in analytical chemistry such as ultraviolet, visible or infrared absorption, flame ionization detector, nitrogen phosphorus detector and mass spectrometry are linear (or linearized in the case of absorbance) [74]. That is, the increase in measured signal is proportional to the increase in analyte concentration. However, several chemical and physical processes that are not included in the simplistic models describing analyte response, such as saturation in atmospheric pressure ionization sources, can lead to non-linear responses [87], and therefore quadratic calibration curves are justified and appropriate. If such a behaviour is known, then again, taking it into account when performing a least squares regression will lead to increased accuracy. This has long been recognized by analytical instruments companies, which typically provide the user several

model orders (linear, quadratic) to compensate for non-linearity, and weighting (1, $1/x$, $1/x^2$) to accommodate for the noise type present in the data.

However, with regards to selecting the correct calibration model, this thesis will show that some of the bioanalysis "conventional wisdom" is downright detrimental. For example, with regards to model order selection, the most common instruction is, as stated by SWGTOX, that one should "choose the simplest model that fits the concentration – response relationship" [75]. The problem is, computer simulations show that this method performs quite poorly in selecting the actual weight and order underlying the data, which will impact quantification accuracy [1, 2, 88]. This approach of going for the simplest possible model was initially brought about by a desire to avoid overfitting, which is the production of an unnecessarily complex model with a poor prediction power for future data [89]. This was indeed a reasonable process at a time when the mathematical and technological tools did not allow for quick and accurate assessment of the real model underlying the data. But now that such tools exist, the need to always adopt the simplest model in fear of overfitting disappears, and rather we should use these techniques to make better model/weighting determinations to get an answer closer to the underlying data.

Published guidelines in forensic toxicology, and bioanalysis in general, are not as informative as they could be on the topic [75, 79–81]. Some suggest procedures to determine the order of the model, or the weighting, but not both. Others propose a way to validate the model, but are silent on how to choose it and, more importantly, what to do if validation fails. In other documents, the reader is offered a range of potential tests for each question, without any consideration as to which one works best, and under which conditions exactly. For example, SWGTOX guidelines suggest using a standardized residuals plot to visually evaluate the data, although they cite ANOVA lack of fit (for linear unweighted models), significance of the second order term (for quadratic models) and coefficient of determination (for linear models) as "other appropriate alternatives" [75]. This is typical of bioanalysis guidelines. Quite often, as is the case with SWGTOX guidelines [75], calculations are not even detailed, leaving the reader to figure out how to adapt these often generic tests to a calibration situation – and even less is provided in the case of a weighted or quadratic least squares regression. Prescriptive validation guidelines dictating the use of a single statistical procedure for the purpose of selecting and validating a calibration model are not necessarily desirable, since flexibility is required to fit the very diverse array of bioanalysis methods. However, for a complete validation leading to the most accurate results to take place, a method of efficiently and accurately selecting and validating the calibration model is required. To that end, prac-

tionners would be well served by explicit benchmarked tools with highly specific instructions.

1.4.4 Quantification bias introduced by endogenous analytes

Even with a method to choose and validate the correct calibration model, some situations remain prone to introducing bias in quantification procedures. One of these cases is the so called "endogenous substances" situation. As opposed to xenobiotics, substances foreign to the body which are ingested, endogenous substances originate within the body [12]. Forensic toxicologists can be concerned with quantifying endogenous substances such as gamma-hydroxybutyric acid (GHB), a neurotransmitter and drug of abuse [90], insulin (a regulator of carbohydrates metabolism, a therapeutic drug and potential poison) [91], steroids (hormones which can be abused) [92] and of course, in the context of this thesis, CYP enzymes.

Quantification of an endogenous substance will pose some issues during method validation and beyond. Indeed, if matrix-matched calibration standards are used (as generally recommended in validation guidelines), the laboratory will be hard pressed to find a matrix blank from the targeted analyte in which to spike calibration standards [93]. By nature, all matching matrices will contain the targeted analyte, albeit at different levels. This basal analyte level will generate a signal in all calibration standards, shifting the calibration curve upwards with $b_0 > 0$. Unknown samples analyzed using the mathematical equation derived from these calibration samples will suffer from a bias equal to the endogenous concentration present in the calibration matrix.

Potential solutions to this problem are reviewed in more detail in Section 3.2, but in essence none of them provides a comprehensive correction for the bias in matrix matched calibration standards compatible with routine analyses [93–100]. In a typical method operated in production mode in a forensic toxicology laboratory, calibration standards, quality controls and certain system checks will be analyzed alongside a varying number of case samples (1 to 80 and more, typically at least 30) that contain an unknown concentration of the targeted analyte. Any correction procedure thus needs to be efficient, and if possible, not require any additional experimental manipulation on the unknown samples.

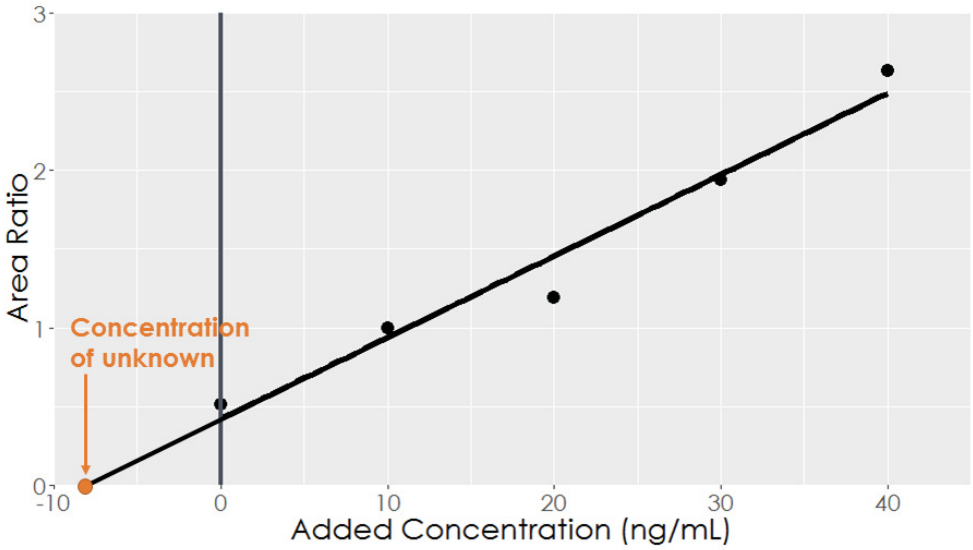


Figure 1.4.3: Standard addition technique: the unknown is measured, then repeatedly spiked with an aliquot of the analyte and remeasured. This results in a set of calibration measurements in the sample matrix. If we make the assumption that a blank sample would yield zero signal then the concentration of the sample can be calculated. As shown, this is done by calculating the absolute value of the x-intercept.

1.4.5 Standard addition

Standard addition is a quantification technique designed to take into account even the most severe matrix effect, since it is in essence the most perfect matrix matched calibration. In standard addition, known concentrations of the targeted analyte are added to aliquots of the sample in which we want to quantify the analyte of interest [101] (Figure 1.4.3). The concentration of analyte in the unknown sample is then calculated as the absolute value of the x-intercept ($x = |{}^{b_0}/_{b_1}|$). This technique, in principle, produces a high accuracy concentration estimation, but is labour intensive. This is daunting, especially when large numbers of unknown samples need to be quantified. The number of samples to be extracted and analyzed is essentially multiplied by the number of calibration levels, often 5 or more.

Although performing unknown sample quantification via standard addition for all casework is simply out of the question from a production efficiency standpoint, the standard addition system can still be co-opted to determine the endogenous concentration present in the calibration matrix and correct for it in the unknown samples, with no additional experimental work needed. Here, matrix matched calibration standards are prepared in the same "blank" matrix that contains some unknown endogenous level of the analyte. After the responses for the calibration standards are collected, they are treated like a standard addition

experiment to calculate the endogenous concentration. Subsequently, the known concentrations of the calibration standards are updated to reflect the true analyte concentration, prior to using them for regression and calibration for the unknowns. This strategy is explored further in Chapter 3, including a consideration of the potential for increased error in the results.

1.5 Qualitative determination methodology

1.5.1 Qualitative methods

Mathematical challenges are readily apparent in quantitative methods, as outlined above. But even though qualitative methods are often taken for granted in that respect, like every technique relying on a measurement prone to uncertainty, there are worthwhile questions to be asked. Qualitative methods are defined by the binary outcome of the analysis performed. They exist in two distinct flavours: qualitative identification methods, yielding a detected/non-detected result, and qualitative decision point methods, yielding an above/below cut-off result [4].

Decision point qualitative methods are useful when rules or legislation mandate a fixed limit or concentration for a controlled species. Some examples would include threshold concentrations for doping agents in biological samples [102], or for pesticides in food [103]. In forensic toxicology, this situation often arises with so called "per se" legislations, which establish a blood (or plasma, or oral fluid) concentration for one or more xenobiotics above which it is an infraction to drive a motor vehicle (the best known example in Canada is the 80 $mg\%$ or 0.08 $g/100\ mL$ for alcohol (ethanol), but recent C-46 legislation instituted per se levels for two additional drugs in blood [104]). Decision point qualitative methods are vulnerable to error, especially when using instrumentation in which the signal (sensitivity) varies significantly on a daily basis such as LC-MS/MS. A method run with such an instrument might issue a positive finding one day, and a negative finding the next for the same sample, only by virtue of the shift in sensitivity. While this might be acceptable for some applications, in forensic toxicology, this poses real questions as to equity between cases. For example, we would not want two DUID cases with exactly the same concentration in blood or urine to end up with divergent positive and negative results only because the instrument sensitivity shifted between the two analyses. Without performing a full quantification of the targeted analyte(s), this problem can be solved by putting in place a correctly validated qualitative decision point method.

1.5.2 Measurements and output in qualitative decision point methods

In these methods, the decision point is typically enunciated in units we can interpret intuitively, i.e., concentration units. To apply this decision point and generate the expected binary outcome, a conversion is needed from the cut-off concentration to its equivalent signal, i.e., what is the threshold signal corresponding to the threshold concentration to which we can compare unknown samples? Unless the method has a fixed conversion formula (for example in spectrophotometry), the standard approach is to analyze a matrix matched sample spiked at the threshold concentration alongside the unknown samples that are examined. The unknown samples' signals are then compared to the threshold or cut-off sample to be classified as "above cut-off" or "below cut-off", a binary outcome. This type of method is thus fundamentally different from quantification methods, since the magnitude of the signal, within certain limits, is irrelevant. The only pertinent element is if it is larger or smaller than the cut-off. This fact should be reflected in the procedure used for validation.

SWGTOX validation guidelines [75] for qualitative methods indicate that the precision of the measurement (area, area ratio) should be estimated at cut-off, as well as at $> 50\%$ and $< 150\%$ of the cut-off concentration. If the measured percent relative standard deviation (%RSD) at these levels is smaller than 20%, and if measurements at cut-off plus or minus two times the standard deviation (2σ) do not overlap those at 50% and 150%, then the method is considered to be validated. In other sources in the literature and guidelines, qualitative methods are little studied, and recommendations are heteroclite [105–107].

There are two significant issues with SWGTOX's qualitative method validation recommendations. First, they apply what is clearly a derivative of quantitative method validation procedures to a method which is inherently non-quantitative. Qualitative methods yield binary, yes or no, above or below results. If the goal of method validation is to define method performance under production conditions, then those same binary results, and their associated statistics, are what should be used for validation purposes, not quantitative (i.e., precision) parameters.

Second, validation procedures, which so clearly recognize the presence of measurement uncertainty, are not at all consistent with current production practices, which do not take it into account at all. In the current state, once in production, the measurement of an unknown sample is compared to the measurement at the cut-off concentration. But we know that upon repeated measurements we would observe a distribution of values, often normal, for both the sample and the cut-off. Therefore, it is entirely predictable that any sample

with a concentration close to the cut-off, analyzed multiple times, will yield different results: sometimes above, sometimes below cut-off. This is entirely attributable to the expected measurement uncertainty. However, presentation of the results as being purely binary masks the presence of measurement uncertainty to the scientists. A new qualitative method validation and production framework, relying entirely on the binary nature of the results, needs to be developed and is presented in this thesis.

1.6 Mathematical method validation tools for application to a proteomics approach of postmortem metabolic capacity estimation

The long term goal of this research is the development of a method to characterize and quantify CYP enzymes in postmortem liver tissue. This is needed to enable calculation of the metabolic capacity of an individual. However, as it has just been discussed, there were some foreseeable issues expected in the validation stage. In order to work on the main goal it was clear that these other neglected areas would need urgent attention. As a result, this thesis has largely become focused on addressing these urgent problems and laying the groundwork for a validated CYP method rather than solely focused on the difficult problem of quantitative CYP proteomics.

The following material is intended to prepare the reader for the remainder of the thesis.

In Chapter 2, an automated procedure for the selection of least squares regression parameters (weighting, model order) and validation of the final calibration model is presented. Section 2.1 provides a description of the solution developed and its application to an LC-MS/MS method, with cocaine and naltrexone being used as primary examples. Section 2.2 provides more in depth statistical theory underlying the choices made in the developed procedure, including the different avenues tested for the intended purpose.

In Chapter 3, the theory and tools to deal with endogenous presence of an analyte in the calibration matrix are developed and applied to a BHB quantification method in a production environment. Automated correction routines were built with R, including a calculation of the additional error folded in by the correction process.

In Chapter 4, sound validation guidelines for decision point qualitative methods are de-

veloped and tested. In Section 4.1, raw measurements and binary output of qualitative decision point methods are studied and modeled. From that information, validation guidelines relying on the binary output are proposed. In Section 4.3, these guidelines are used to validate a method for 92 analytes in oral fluid. Xenobiotics can be found in oral fluid as a result of partitioning from the circulatory system or by direct deposit in the oral cavity (e.g., smoking) [108, 109]. Over the last decades, oral fluid has grown as a biological sample of choice for driving under the influence of drugs investigation [110–113], and the method presented and validated in Section 4.3 is directed towards this application.

Chapters 2 to 4 of this thesis present packages of mathematical tools that can be applied to several bioanalysis methods, such as xenobiotics analysis, but that have been developed while keeping in mind the needs of the CYP analysis method: quantification of enzymes in liver tissue and detection of mutations (a qualitative measurement). In Chapter 5, a CYP 2D6 and CYP 3A4 characterization and quantification method developed for postmortem estimation of metabolic capacity is presented. The method could not be fully validated but rather demonstrates a proof of concept for application in the forensic toxicology domain.

This thesis explores two seldom studied interfaces of forensic toxicology: protein analysis and mathematics (statistics). The forensic toxicology field has long been focused on small molecules (i.e., legal and illegal drugs) analysis, with the occasional incursion in other types of methods. In order to make this research accessible to all forensic toxicologists, a large importance is given to the didactic aspect in the papers, making them accessible to scientists of different backgrounds.

2

Procedure for the Selection and Validation of a Calibration Model

This chapter is the integral text from[1]*:*

Procedure for the Selection and Validation of a Calibration Model I —Description and Application

Brigitte Desharnais, Félix Camirand Lemyre, Pascal Mireault and Cameron D. Skinner

Journal of Analytical Toxicology **41** (4) (2017) 261-268 [1]

and

Procedure for the Selection and Validation of a Calibration Model II —Theoretical Basis

Brigitte Desharnais, Félix Camirand Lemyre, Pascal Mireault and Cameron D. Skinner

Journal of Analytical Toxicology **41** (4) (2017) 269-276 [2].

[1]Throughout this thesis, small modifications have been made to the integral texts presented to correct minor typos, standardise section names, abbreviations and figures appearance.

2.1 Description and application

2.1.1 Abstract

Calibration model selection is required for all quantitative methods in toxicology and more broadly in bioanalysis. This typically involves selecting the equation order (quadratic or linear) and weighting factor correctly modelizing the data. A mis-selection of the calibration model will generate lower quality control (QC) accuracy, with an error up to 154%. Unfortunately, simple tools to perform this selection and tests to validate the resulting model are lacking. We present a stepwise, analyst-independent scheme for selection and validation of calibration models. The success rate of this scheme is on average 40% higher than a traditional "fit and check the QCs accuracy" method of selecting the calibration model. Moreover, the process was completely automated through a script (Appendix B.3) running in RStudio (free, open-source software). The need for weighting was assessed through an F-test using the variances of the upper limit of quantification and lower limit of quantification replicate measurements. When weighting was required, the choice between $1/x$ and $1/x^2$ was determined by calculating which option generated the smallest spread of weighted normalized variances. Finally, model order was selected through a partial F-test. The chosen calibration model was validated through Cramer–von Mises or Kolmogorov–Smirnov normality testing of the standardized residuals. Performance of the different tests was assessed using 50 simulated data sets per possible calibration model (e.g., linear-no weight, quadratic-no weight, linear-$1/x$, etc.). This first of two papers describes the tests, procedures and outcomes of the developed procedure using real LC-MS/MS results for the quantification of cocaine and naltrexone.

2.1.2 Introduction

Every toxicologist performing quantitative method development eventually faces the challenge of choosing a calibration model for the analyte. Most data acquisition and processing software (e.g., Agilent's ChemStation or Mass Hunter, AB Sciex's Analyst®) offer options with regards to forcing the calibration equation through the origin, applying a weight and model order (e.g., quadratic or linear). When only the most common weighting (none, $1/x$, $1/x^2$) and model order (linear, quadratic) options are taken into account, there are six possible calibration models per analyte.

Although in principle, all systems should have a linear response to the concentration and generate linear calibration curves, in reality, some physical and chemical phenomenon can

create quadratic calibration curves. Processes such as competition in the LC-MS ionization process or saturation of the detector will create saturation phenomenon at high concentrations, even if this is imperceptible to the naked eye. It is important to properly identify occurrences of quadraticity in the data, because this can have a large impact on quality control (QC) accuracy. Simulations using experimentally obtained calibration curves showed a 12% average improvement in QC accuracy when properly using the quadratic calibration model with uniformly weighted data (Appendix B.1). In a similar fashion, Gu et al. [88] demonstrated that there is a notable improvement in QC accuracy when the proper weighting is used for the calibration curve. Identifying the correct calibration model is, therefore, a crucial step in method validation that will have impacts on QC accuracy in production.

The Scientific Working Group for Toxicology (SWGTOX) guidelines state that "ultimately, the best approach is to use the simplest calibration model that best fits the concentration response relationship" [75]. The SWGTOX recommends that the fit be evaluated using a standardized residuals plot. Although this type of graph is a very useful tool to roughly estimate the fit, the visual interpretation of the data renders model selection very analyst-dependent and therefore subjective. The SWGTOX validation guidelines also mention that the correlation coefficient (r) alone cannot be used to evaluate the fit, and that other alternatives can be used, such as analysis of variance - lack of fit (ANOVA-LoF), significance of the second order term and the coefficient of determination. However, the calculations for these tests are not detailed in the validation guidelines, and there are no recommendations with regards to the circumstances in which they should be applied. Additionally, as is shown in Section 2.2, the ANOVA-LoF and significance of the second order term techniques have significant issues in terms of performance or ease of use.

In order to address these issues, we have developed a stepwise, systematic method to choose and validate the calibration model for an analyte. This method is not biased by the interpretation of the analyst since conclusions are reached by comparing test results to a cut-off. Furthermore, the testing and interpretation has been automated using a script in RStudio, allowing scientists with limited knowledge or comfort in statistics to perform these tests easily, reliably and quickly. As an example, a method validation for 60 analytes required 1 hour of data treatment time to objectively select the calibration model. Using 2610 calibration data sets spread over different calibration models, curvature levels (magnitude of the x^2 term) and %RSD values, this automated scheme was shown to have a success rate in average 40% higher than the traditional method of fitting with more complex models until QC accuracy is acceptable (Appendix B.1). This vast improvement in the exactness of

calibration model selection will ultimately result in higher QC accuracy in production. The selection method was developed by testing different approaches on data sets both from 50 analytes quantified by LC-MS/MS and simulated data sets with varying numbers of replicates. In this paper, we detail the calculations and interpretation steps that constitute the developed process. As practical examples, two different analytes were chosen to demonstrate this protocol: cocaine and naltrexone. The theoretical basis underlying the choice of each test, as well as the mathematical considerations for different aspects of the scheme (including data collection, outliers and forcing calibration through the origin), is covered in Section 2.2.

2.1.3 Materials and methods

LC-MS/MS quantification

Cocaine and naltrexone (Cerilliant, Round Rock, TX, USA) were spiked in bovine blood at concentrations of 5, 10, 15, 50, 75, 100, 400, 500 and 1000 ng/mL to produce calibration standards. Considering that most of the samples analyzed with this method fall in the low concentration range (i.e., therapeutic concentrations), it is appropriate to place more calibration levels at the lower end of the working range. Cocaine-D_3 and codeine-D_3 (Cerilliant) were used as internal standards (IS, 5 and 100 ng/mL, respectively). Solid phase extraction of the standards was performed using Oasis cartridges (HLB 3cc, product WAT094226, Waters, Mississauga, ON, Canada). A 2 mL volume of blood was extracted and reconstituted in 100 μL of 15:85 methanol:ammonium formate (10 mM). The samples were analyzed on an Agilent 1200 HPLC equipped with an AB Sciex 4000 QTrap mass spectrometer. An aliquot (5 μL) was injected and separated on an Agilent Zorbax Eclipse C18 column (100×2.1 mm, 3.5 μm) using a 25 minute step/ramp gradient from 10 mM ammonium formate + 0.2% formic acid to methanol. Quantitative analysis was performed with m/z transition 305.2/183.0 Da for cocaine (the ^{13}C-containing species was used to reduce the signal and remove saturation at the upper levels of the working range) and m/z transition 342.1/212.0 Da for naltrexone. The peak area ratio of the analyte to the IS was used as the response. This method has been validated according to ISO 17025 [71] and CAN-P-1578 [73] guidelines and is currently used as a routine quantification method. Five injections of each extracted standard were performed in order to create measurements replicates on which to base the statistical analysis. The selection of this experimental setup to obtain replicate measurements is explained in Section 2.2.4. Chromatographic data analysis was performed with Multiquant^{TM} (AB Sciex, Framingham, MA, USA).

Simulated data sets

To validate the calibration model selection accuracy for all six types of possible models (linear-no weight, quadratic-no weight, linear-$1/x$, etc.), simulated data were produced using a script written and run in RStudio (RStudio, Boston, MA, USA). R (programming environment, https://www.r-project.org/) and RStudio (graphical interface, https://www.rstudio.com/) are free open-source statistical software tools. The script for simulated data generation is available in Appendix B.2.

Using experimental LC-MS/MS calibration data for 50 analytes, intervals spanning the maximum and minimum calibration parameter values for b_0, b_1 and b_2 for quadratic models were established. Synthetic calibration data were generated using calibration parameters chosen at random from within these intervals. For the present study, interval boundaries were 9×10^{-3} to 5×10^{-1} for b_0, 3×10^{-3} to 8×10^{-1} for b_1 and -7×10^{-5} to -7×10^{-8} for b_2. Using these parameters, the predicted signal ($\hat{y}_i$) for each concentration level was calculated.

For every weighting scheme (none, $1/x$, $1/x^2$), each data set was assigned a maximal %RSD value at random between 1% and 20%. This 20% upper boundary was chosen by keeping in mind the SWGTOX guidelines that state precision values should not be higher than 20% [75]. From the randomly assigned %RSD at the lower limit of quantification (LLOQ), the standard deviations for other concentration levels were calculated according to the chosen weighting pattern.

Using the calibration parameters and calculated standard deviations, 50 data sets, each with 5, 7 or 10 normally distributed replicate measurements, were generated at each concentration level for each of the six calibration models tested here.

Heteroscedasticity testing

Only the description of the calculations and/or R functions used to carry out tests will be described in this section[2]. The purpose of the tests and interpretation of the results will be described in Section 2.1.4. All calculations required to choose and validate a calibration model were performed using an R script running in RStudio. All required R scripts as well as instructions for their use, including a video tutorial, are available in Appendix B.3.

[2]The reader can refer to Figure 2.1.3 for an overview of the complete process.

The presence of heteroscedasticity (a change in variance across concentration levels) was determined by calculating the probability that the variance of measurements at the upper limit of quantification (ULOQ) was equal to or smaller than the variance of measurements at the LLOQ using an F-test. This unilateral F-test was performed using the following RStudio formula:

```
var.test(MeasurementsLLOQ, MeasurementsULOQ, alternative = "less")
```
(2.1)

with the probability being stored in the P value element of the output list.

Variance evaluation for weight selection

Variance evaluation was performed by first applying each weighting scheme (w_i) to the measurements to calculate the concentration levels' normalized weighted variances (V_w^i), which were used to calculate the total normalized weighted variance (V_w) [114].

$$S = \sum_i \sqrt{w_i} \tag{2.2}$$

$$V_w^i = \frac{Var\left\{y_{i1}, y_{i2}, \ldots, y_{ij}\right\} \times w_i}{S^2} \tag{2.3}$$

$$V_w = Var\left\{V_w^1, V_w^2, \ldots, V_w^i\right\} \tag{2.4}$$

where S is the scaling factor, w_i was the weighting applied at the i^{th} concentration level (e.g., for a $1/x$ weighting at the 5 ng/mL concentration level, w_5 will be $1/5 = 0.2$), V_w^i is the weighted and normalized variance at the i^{th} concentration level for weighting scheme w and concentration level i, Var is the variance operator, which calculates the variance of the elements inside the braces, y_{ij} is the measurement at the i^{th} concentration level and the j^{th} replicate and V_w is the total normalized weighted variance for weight w. Three values of V_w should be obtained using this calculation, one for each possible weight (uniform or "no weight" ($w_i = 1$), $1/x_i$ and $1/x_i^2$).

Partial F-test for model order selection

To perform the partial F-test, the sum of squares for the linear ($y_L = b_1x + b_0$) and quadratic ($y_Q = b_2x^2 + b_1x + b_0$) models were calculated by [86]

$$SS_{reg,Q} = \sum_{n=1}^{i} w_i \times n_j \times (\hat{y}_i - \bar{y})^2 = \sum_{n=1}^{i} w_i \times n_j \times \left(\left\{b_2{x_i}^2 + b_1x_i + b_0\right\} - \bar{y}\right)^2 \tag{2.5}$$

$$SS_{reg,L} = \sum_{n=1}^{i} w_i \times n_j \times (\hat{y}_i - \bar{y})^2 = \sum_{n=1}^{i} w_i \times n_j \times \left(\left\{b_1x_i + b_0\right\} - \bar{y}\right)^2 \tag{2.6}$$

where $SS_{reg,Q}$ and $SS_{reg,L}$ are the sum of squares of the regression of the quadratic and linear models, respectively, w_i is the weighting applied at the i^{th} concentration level, n_j is the number of measurement replicates (here, 5) per concentration level, $\hat{y}_i$ is the predicted measurement at the i^{th} concentration level (obtained by inserting the value of x in the calibration equations, which need to be previously determined) and $\bar{y}$ is the average of measurements over all samples analyzed (nine concentration levels $\times$ five replicates = 45 measurements).

The sum of the residuals squared ($SS_{res,Q}$) was calculated for the quadratic model from

$$SS_{res,Q} = \sum_{n=1}^{i \times j} w_i \times (y_{ij} - \hat{y}_i)^2 \tag{2.7}$$

where y_{ij} was the j^{th} measurement at the i^{th} concentration level.

The F statistic was then calculated by

$$F_{calc} = \frac{SS_{reg,Q} - SS_{reg,L}}{\left(\frac{SS_{res,Q}}{n-3}\right)} \tag{2.8}$$

where n is the total number of measurements ($i \times j$, here 45).

The probability (P) associated with the calculated F statistic was found using the RStudio command

```
1-pf (F_calc, 1, (n-3))
```
(2.9)

Normality of the residuals

Normality of the standardized residuals was evaluated through the Kolmogorov–Smirnov (KS) and Cramer-von Mises (CVM) tests. The calculations necessary for these tests are fairly complicated and are covered in Section 2.2. The probability values (output of the tests) were collected in the `.txt` file created by the script.

2.1.4 Results and discussion

Raw data

Raw data resulting from the replicate analysis ($j = 5$) of the nine calibration standards for cocaine and naltrexone are presented in Table 2.1.1. Calibration curves and variance graphs for both analytes are shown in Figure 2.1.1.

Heteroscedasticity testing

The purpose of testing for heteroscedasticity was to determine if weighted least-squares regression was necessary. Data are heteroscedastic if the absolute error (standard deviation of the replicates) varies systematically across concentration levels. Figure 2.1.2 shows the calibration curve of simulated homoscedastic (a) and heteroscedastic (b and c) data sets. In least-squares regression, the best model parameters (e.g., slope and intercept for linear models) are found by minimization of the sum of the squared error between the measured values and the values predicted by the model (squared residuals). In unweighted (also called uniformly weighted) least-squares regression, the default regression, all squared errors are treated equally in the summation. On the other hand, when data are heteroscedastic, there is greater confidence in the measured values that have the smallest error. This greater certainty should be used advantageously by giving a greater weight to the values with the smallest error in the summation of errors and therefore a greater influence in fixing the calibration parameters [85].

The F-test was applied to the measurements at the LLOQ and the ULOQ, where the difference in variance (error) is the largest inside the calibration range for heteroscedastic data sets of the $1/x$ and $1/x^2$ type. The P-value obtained represents the probability that variance at the ULOQ is smaller than or equal to the variance at the LLOQ (null hypothesis).

Table 2.1.1: Area ratio (analyte/IS) at nine concentration levels with five measurement replicates for cocaine and naltrexone.

Concentration (ng/mL)	Area ratio (analyte/internal standard) Naltrexone				
	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$
5	0.131	0.127	0.131	0.126	0.130
10	0.256	0.249	0.244	0.249	0.249
15	0.340	0.333	0.328	0.331	0.311
50	1.235	1.257	1.224	1.234	1.225
75	1.596	1.663	1.710	1.656	1.613
100	2.055	2.046	2.109	2.033	2.127
400	7.733	7.727	7.964	7.687	7.747
500	9.688	9.447	9.557	9.476	9.346
1000	16.298	15.575	15.807	15.926	15.420

Concentration (ng/mL)	Area ratio (analyte/internal standard) Naltrexone				
	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$
5	0.052	0.054	0.047	0.049	0.053
10	0.101	0.104	0.103	0.099	0.102
15	0.133	0.132	0.136	0.131	0.135
50	0.528	0.510	0.515	0.497	0.503
75	0.669	0.676	0.649	0.682	0.639
100	0.923	0.909	0.924	0.964	0.964
400	3.419	3.451	3.497	3.673	3.595
500	4.426	4.455	4.458	4.529	4.600
1000	8.656	9.092	9.139	9.110	9.269

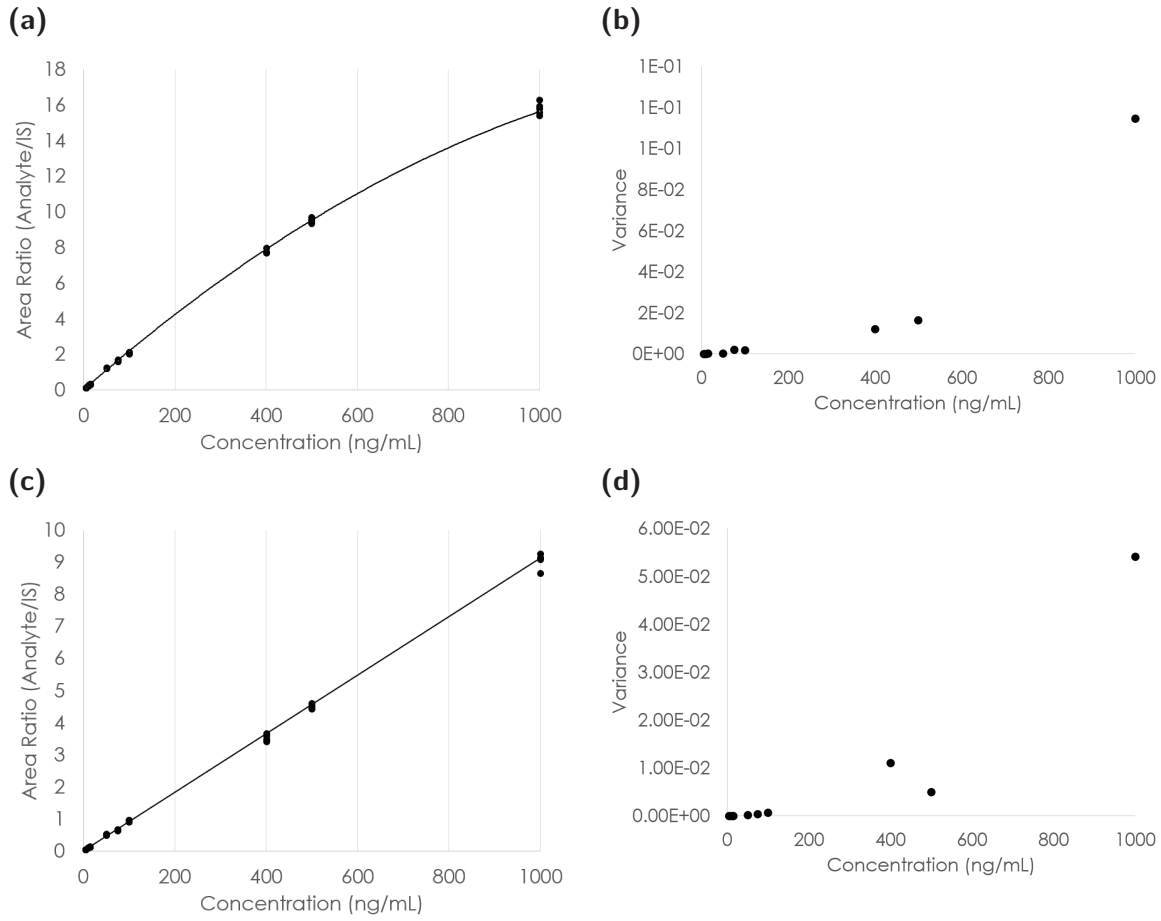


Figure 2.1.1: Calibration curves and variance graphs of cocaine and naltrexone. **(a)** Calibration curve of cocaine, equation: $y = 7 \times 10^{-6}x^2 + 0.0224x + 0.0174$, where y is the area ratio and x is the concentration in ng/mL. **(b)** Variance graph of cocaine. **(c)** Calibration curve of naltrexone, equation: $y = 0.0091x + 0.0059$, where y is the area ratio and x is the concentration in ng/mL. **(d)** Variance graph of naltrexone.

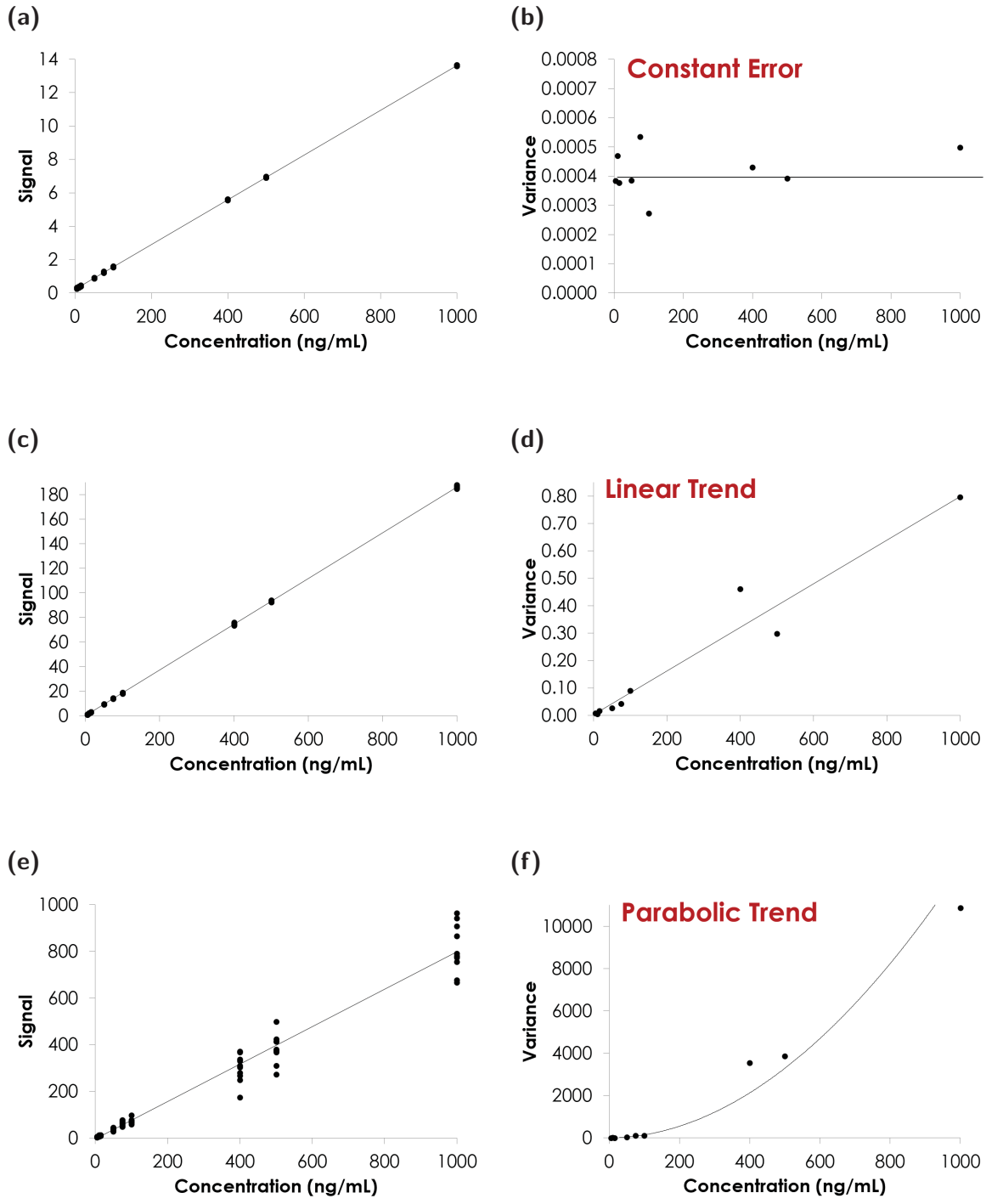


Figure 2.1.2: Calibration curves and variance graphs of linear models, created to show as well as possible the patterns of changing variance. **(a)** Calibration curve with homoscedastic data. **(b)** Variance graph with homoscedastic data. **(c)** Calibration curve with heteroscedastic data, $1/x$ weight. **(d)** Variance graph with heteroscedastic data, $1/x$ weight. **(e)** Calibration curve with heteroscedastic data, $1/x^2$ weight. **(f)** Variance graph with heteroscedastic data, $1/x^2$ weight.

If $P > 0.05$, this null hypothesis is accepted and data are considered to be homoscedastic (constant variance across concentrations), therefore no weighting is required. On the other hand, if $P < 0.05$, the null hypothesis is rejected and we accept the alternative hypothesis, which states that the variance at the ULOQ is larger than the variance at the LLOQ. This means that data are heteroscedastic and a weighting factor, which will be decided using the variance evaluation, should be used.

For cocaine and naltrexone, the F-test yielded P-values of 7×10^{-9} and 7×10^{-8}, respectively, indicating that the data sets were heteroscedastic and weighting should be applied in both calibration procedures.

Application to simulated data showed this test is robust, with an average success rate of 98% for all types of calibration model utilizing five replicate measurements (Table 2.1.2). The success rate represents the percentage of data sets that was correctly classified (e.g., declared homoscedastic when it was indeed homoscedastic). Because the P-value threshold is set at 0.05 (5%), a 95% success rate is expected. Since the observed success rates for all models are near the expected rate, as is the average rate, this test is considered robust.

Variance evaluation for weight selection

Most data analysis software offers unweighted or uniform regression (weighting factor = 1) as the default as well as weighted regression using $1/x$ and $1/x^2$ weighting factors. The theoretical basis for these three common weighting factors is beyond the scope of this paper, but is a result of the type of noise that dominates over the calibration range [84, 115, 116]. Examination of the variance plot (variance of replicates vs. concentration) provides confirmation of the heteroscedasticity test results and is also suggestive of the appropriate weighting factor. The variance plot is provided as a PDF output when the R script is executed. Constant error across the calibration range from the LLOQ to the ULOQ is indicative that unweighted regression was appropriate (see Figure 2.1.2d). Weighted regression was justified through heteroscedasticity testing and is apparent as increasing error across the variance plot. Plots which exhibited a linear trend, where variance increased proportionally to the concentration, indicated that $1/x$ weighting should be selected (see Figure 2.1.2e). If a parabolic trend was found, where variance increased proportionally to the square of the concentration, a $1/x^2$ weighting factor should be used [88] (see Figure 2.1.2f).

These characteristics were the basis for an automated, analyst independent selection

Table 2.1.2: Success rate of different tests in the process of calibration model selection and validation for 5, 7 or 10 simulated measurement replicates; 50 data sets were generated for each weighting/order combination.

	5 replicates					
Model order	**Linear**	**Quad.**	**Linear**	**Quad.**	**Linear**	**Quad.**
Weighting	**1**	**1**	$1/x$	$1/x$	$1/x^2$	$1/x^2$
F-test (Heteroscedasticity) (%)	98	92	100	98	100	100
Variance test (Weight selection) (%)	100	98	58	70	100	100
Partial F-test (Order selection) (%)	96	96	98	88	90	50
Validation (CVM) (%)	100	100	100	100	100	100

	7 replicates					
Model order	**Linear**	**Quad.**	**Linear**	**Quad.**	**Linear**	**Quad.**
Weighting	**1**	**1**	$1/x$	$1/x$	$1/x^2$	$1/x^2$
F-test (Heteroscedasticity) (%)	94	98	100	100	100	100
Variance test (Weight selection) (%)	100	100	86	90	100	100
Partial F-test (Order selection) (%)	98	100	100	94	98	48
Validation (CVM) (%)	100	100	98	100	100	100

	10 replicates					
Model order	**Linear**	**Quad.**	**Linear**	**Quad.**	**Linear**	**Quad.**
Weighting	**1**	**1**	$1/x$	$1/x$	$1/x^2$	$1/x^2$
F-test (Heteroscedasticity) (%)	100	98	100	100	100	100
Variance test (Weight selection) (%)	100	100	98	92	98	100
Partial F-test (Order selection) (%)	98	100	88	90	94	58
Validation (CVM) (%)	100	100	100	98	98	98

of the required weighting using a variance evaluation. Indeed, properly weighted variances should be constant across the calibration range. For example, if the raw variances increase linearly with concentration (x), multiplying all variances by the appropriate weighting, $1/x$, will result in constant weighted variances across all the calibration range. Conversely, multiplying by an inappropriate weighting factor, say $1/x^2$, will produce changing weighted variances across the calibration range. Therefore, the weighting factor producing the most uniform set of weighted variances, as evaluated by taking the variance of the weighted and normalized variances for the different concentration levels, is the closest to the proper weight and should be used. Variance evaluation also acts as a double-check of the F-test result, building a healthy redundancy in the weighting selection for the calibration curve.

Calibration data and variance plots obtained for cocaine and naltrexone are shown in Figure 2.1.1. Both variance plots show a parabolic pattern, although this pattern was subjectively less clear for naltrexone. Variance test scores for cocaine ($V_{w_1} = 1.7 \times 10^{-6}$; $V_{w_{1/x}} = 2.0 \times 10^{-9}$; $V_{w_{1/x^2}} = 6.2 \times 10^{-12}$) and naltrexone ($V_{w_1} = 3.8 \times 10^{-7}$; $V_{w_{1/x}} = 5.0 \times 10^{-10}$; $V_{w_{1/x^2}} = 2.6 \times 10^{-12}$) confirmed that a $1/x^2$ weighting factor should be used to build calibration models for both analytes, since this weight produced the smallest spread of weighted variances. Both the plot and weighted variance evaluation provide confirmation of the heteroscedasticity F-test results. Nearly, all LC-MS/MS analyses spanning a few concentration orders of magnitude can be expected to produce data with this weighting [88].

It is important to note here that sampling statistics govern the variance estimation at each concentration level. Thus, the smaller the number of replicates, the more likely the variance estimation is to be erroneously large or small. This estimation error propagates into the weighted variances as a bias toward an erroneous weighting and can result in incorrect selection of the weight. Tests with simulated data show that this happens up to 42% of the time for $1/x$ data with five replicates (Table 2.1.2). To overcome this fundamental limitation in the data requires increasing the number of replicates, which increased the success rate in identifying the proper weighting factor to 86% for 7 replicates and 92% for 10 replicates. For this reason, the authors suggest that the use of seven measurement replicates when selecting and validating the calibration model, which provides improved performance with the tests compared to the five measurement replicates suggested by the SWGTOX guidelines. In general, improved performance occurs for all tests with increased replicates, but it is the most marked in the weight selection step. For diverse practical reasons, analysts may justifiably use five measurement replicates and, with the aid of the calibration model selection scheme presented here, produce validated calibration models. However, they need to realize that the

trade-off will be an increased frequency of incorrect weight and/or order selection that can ripple through to lower accuracy and precisions in the results.

Partial F-test for model order selection

With the weighting factor chosen, the next step was to select the model order (i.e., linear or quadratic). The recommended practice by the SWGTOX and the FDA is to choose the model with the lowest order that adequately describes the calibration system under study [75, 117]. Often times in bioanalysis laboratories, this is done using the "Fit and Test" strategy [88], meaning that the lowest order yielding standard and QC accuracies below the 15% or 20% bar is chosen. However, rather than choosing the model which is "good enough", the partial F-test can be used to improve the likelihood of selecting the true model order underlying the measurements.

Selection of the appropriate model was done by performing a partial F-test. Here, the test was applied to establish if the quadratic calibration model significantly improved the captured variance of the data compared to a linear model [86]. Linear or quadratic calibration responses, which are typically encountered in toxicology validation work, were the two models compared. However, it is noteworthy that the partial F-test allows alternate calibration models to be compared. This test compares the improvement in the sum of squares of the regression when switching from a linear to a quadratic calibration model ($SS_{reg,Q} - SS_{reg,L}$) to the sum of squares of the residuals in the quadratic model ($^{SS_{res,Q}}/_{n-3}$). If there is a significant increase in the variance explained by the quadratic regression, then using a quadratic model is justified [86].

For cocaine and naltrexone, the P-values obtained from the partial F-test were 1×10^{-13} and 0.20, respectively. In the case of cocaine, since $P < 0.05$, the increase in the sum of squares of the regression when switching to a quadratic model was significant, therefore a quadratic (second order) calibration model should be used for this analyte. On the other hand, the P-value for naltrexone was > 0.05, which means the quadratic model does not capture a significantly greater portion of the measurements' variance. Therefore, a linear model should be used for naltrexone.

Tests with simulated data showed that the erroneous outcome of selecting a quadratic model when in fact the underlying data model was linear happens in 4% of the cases on average (Table 2.1.2), near the expected value of 5%. The opposite error (selecting linear when quadratic is the correct model) happens far more often, with an average success rate of

80% (Table 2.1.2). Erroneous selection of a linear model mainly happened when the second order term was small relative to the error at the upper concentration levels, for example when increasing variance (heteroscedastic data, especially $1/x^2$) masked the curvature. Indeed, quadratic, $1/x^2$ data with $n = 7$ show a 78% success rate for model order selection when %RSD = 2.5%, but the success rate decreased to 26% when %RSD = 20% (Appendix B.4). Large and/or increasing variance can mask curvature present in the data and result in an undetectable improvement in the fit obtained when using a quadratic model. Unfortunately, increasing the number of measurement replicates will not have a marked effect in this case. Satisfyingly, when the partial F-test fails, the result is to err on the side of caution advocated by the SWGTOX and the FDA: the lowest order model fitting the data (linear) is selected. Ultimately, when the curvature is masked, using a linear model instead of a quadratic one will not have an appreciable impact on the accuracy of the results.

Normality of the residuals

After having chosen the calibration model that best represents the data (weight and order), its validation was required. In principle, the correct model should describe all the systematic trends in the data with only random error remaining in the residuals [85]. Therefore, the residual errors are expected to follow a normal distribution. Both the CVM and KS procedures can be used to test whether the standardized residual distribution is significantly different from a normal distribution. In practice, to adhere correctly to statistical procedures and for clarity of decision, the user is expected to choose only one of them as a validation test. The authors favor the use of the CVM normality test, because the results obtained in simulations demonstrate that it is stricter than the KS test (lower P-values obtained) and therefore has greater ability to detect departure from normality.

The CVM test produced P-values of 0.865 and 0.992 for cocaine and naltrexone, respectively. In both cases, KS and CVM produced P-values > 0.05 suggesting that the standardized residuals did not depart significantly from a normal distribution. The calibration model was, therefore, considered validated. When $P < 0.05$, the distribution of the standardized residuals is significantly different from a normal distribution. This indicates that the calibration model chosen did not accurately account for all systematic trends in the data and therefore should not be validated.

When the CVM test was applied to the simulated data, the success rate was more than 98% across all six calibration models and higher than the expected rate of 95% (Table 2.1.2). Again, low numbers of measurement replicates and/or high variance will negatively impact

the ability of the tests to detect departure from normality (i.e., inappropriate models producing non-normally distributed residuals). This too points toward the benefits of using a higher number of measurement replicates as a better practice.

When a model fails to pass the validation step, the analyst should attempt to understand why, so that the fundamental problem can be addressed. A detailed exploration of all possible problems is well beyond the scope of this paper but certainly systematic errors or instrument drift should be investigated. Where appropriate, the method should be modified so an adequate model for the data can be found. This might involve a change of IS, a modification of the MS/MS transition(s), a reduction of the dynamic range or a move toward more exotic calibration models (e.g., logarithmic) when justified by the expected analyte/instrument response. The analyst should also be wary of methods with excessively high %RSD since, paradoxically, these methods are easier to validate but are inherently less precise and potentially less accurate.

2.1.5 Conclusions

We developed a general procedure to select and validate quantitative calibration models (Figure 2.1.3). The two model analytes, cocaine and naltrexone, were quantified by LC-MS/MS. The F-test demonstrated that both data sets were heteroscedastic and required weighting in the calibration process. Variance evaluation indicated that the spread of weighed normalized variances was the lowest for $1/x^2$ weighting. Visual examination of the variance graph and evaluation of the variance confirmed the F-test results. A weight of $1/x^2$ was, therefore, chosen for both analytes. A partial F-test demonstrated a significant increase in the sum of squares of the regression when switching from a linear to a quadratic model for cocaine, but not for naltrexone. Therefore, a quadratic calibration model was adopted for cocaine but a linear model was retained for naltrexone. Both calibration models were validated through CVM normality testing of the residuals. Analysis of simulated data sets showed good performance level of all tests; but it also pointed to benefits of increased replicate analysis ($n = 7$) in accurate selection of the calibration model.

Choosing the correct calibration model can have tremendous impact on the accuracy of the QCs. The process of selection and validation of a calibration model explained here is a stepwise, biasfree alternative to other less rigorous methods such as visual inspection of the standardized residuals graph. Simulations using experimentally determined calibration curves have shown that this approach performs much better than a more traditional approach

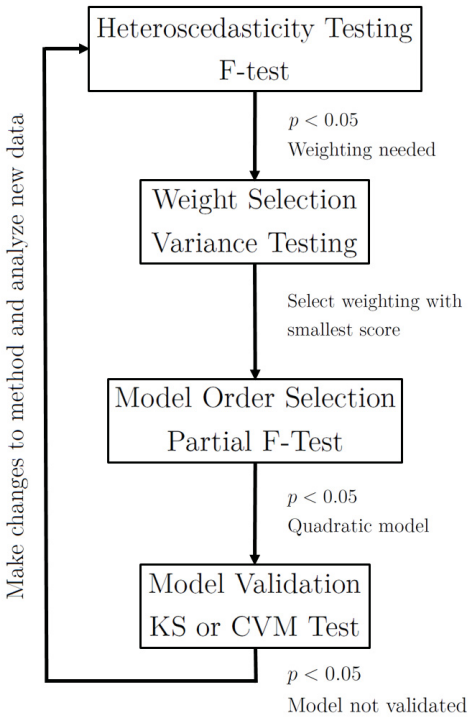


Figure 2.1.3: Flowchart for the selection and validation of the calibration model.

of fitting increasingly complex models until QC accuracy is satisfying. Additionally, the calculations and interpretation of tests results have been automated through the use of RStudio scripts made available to all readers in Appendix B.3. Experimental workload is not modified by the use of this scheme, and only a minute or two per analyte are added to the data treatment time, making this a very efficient option to remove the subjectivity in calibration model selection. This tool is intended to aid analysts in better calibration model selection in toxicology and bioanalysis.

2.1.6 Funding

Brigitte Desharnais, Félix Camirand Lemyre and Cameron D. Skinner gratefully acknowledge support of the National Sciences and Engineering Research Council of Canada. Brigitte Desharnais also gratefully acknowledges the support of the Fonds de recherche du Québec – Nature et technologies.

2.1.7 Acknowledgements

The authors wish to thank Cynthia Côté for her thorough review and comments as well as Sue-Lan Pearring and Chris H. House for their opinion on key matters. The authors are also

grateful to Gabrielle Daigneault, Lucie Vaillancourt, Julie Laquerre and Marc-André Morel for their technical work.

2.2 Theoretical basis

2.2.1 Abstract

In Section 2.1, an automated, stepwise and analyst independent process for the selection and validation of calibration models was put forward and applied to two model analytes. This second part presents the mathematical reasoning and experimental work underlying the selection of the different components of this procedure. Different replicate analysis designs (intra/inter-day and intra/inter-extraction) were tested and their impact on test results was evaluated. For most methods, the use of intra-day/intra-extraction measurement replicates is recommended due to its decreased variability. This process should be repeated three times during the validation process in order to assess the time stability of the underlying model. Strategies for identification of heteroscedasticity and their potential weaknesses were examined and a unilateral F-test using the lower limit of quantification and upper limit of quantification replicates was chosen. Three different options for model selection were examined and tested: ANOVA-lack-of-fit (LoF), partial F-test and significance of the second-order term. Examination of mathematical assumptions for each test and LC-MS/MS experimental results lead to selection of the partial F-test as being the most suitable. The advantages and drawbacks of ANOVA-LoF, examination of the standardized residuals graph and residuals normality testing (Kolmogorov-Smirnov or Cramer-von Mises) for validation of the calibration model were examined with the last option proving the best in light of its robustness and accuracy. Choosing the correct calibration model improves QC accuracy, and simulations have shown that this automated scheme has a much better performance than a more traditional method of fitting with increasingly complex models until QC accuracies pass below a threshold.

2.2.2 Introduction

Choosing an appropriate calibration model is an important part of quantitative method validation. The analyst has several choices to make: forcing the calibration through the origin, choosing a weighting factor and selecting model order (e.g., quadratic or linear). Recommendations as to how to make these choices are frequently vague and do not address all necessary decisions [75, 118–120], although some recent papers have begun to address these

issues [88, 121–124].

The first decision concerns which experimental design to employ to obtain the replicates necessary for data analysis. The goal is to mimic the process used in the production setting once validation is complete. It was shown in the first part of this paper that seven or more replicates are beneficial to improve the accuracy of model selection. One must also consider whether replicate measurements should be performed on different days or not, and with samples extracted separately or several injections of the same extract. Some sources suggest the use of an inter-day/inter-extracts setup [75, 116], but this recommendation does not appear to be supported by a reasoned decision based on mathematical concepts or experimental results.

An option available to the analyst is to force the calibration through the origin. Although no paper actually advocates this procedure and some explicitly discourage it [75, 116, 121], this is a software option available and has been used in papers [125–127]. This practice therefore needs to be clearly addressed.

The absolute error of replicate measurements can be independent of concentration (homoscedastic data) or scale with concentration (heteroscedastic data) [121, 128]. Regression minimizes the sum of squared error (difference between the measured values and predicted values) by selecting optimal calibration coefficients (slope(s) and intercept). When regression is performed on heteroscedastic data, greater importance should be given to the data with the smaller absolute error [85]. Choosing the proper weighting level from the common options (uniform weighting ["no weight"], $1/x$ or $1/x^2$) is an important part of obtaining a calibration robust to normal changes in individual measurement values [85, 88, 129]. Whereas SWGTOX guidelines points toward the use of residuals graph to select proper weighting [75], others use different tests to confirm the presence of heteroscedasticity [121, 122, 130] and select the proper weighting factor [88].

Model order selection is frequently necessary since non-linear behavior is expected with some methodologies (e.g., LC-MS/MS). Appropriate calibration models capture the systematic behavior of the instrument's response but do not model the random error. Excessive model order ("overfitting") results in inclusion of the random error in the model and actually reduces accuracy when the model is used [86]. This topic has been covered more thoroughly in the literature than weight selection, with suggestions of using the analysis of variance lack-of-fit (ANOVA-LoF) [75, 86, 122], residuals graph [75, 85, 116, 121], partial F-test [86]

and significance of the second-order term [75, 86, 122] for appropriate model order selection.

Once the calibration model is selected, a good but often overlooked step is to validate that the model describes only the systematic behavior of the data. ANOVA-LoF [75, 86, 115, 116, 121, 122, 128] and examination of the residuals graph [75, 85, 116] have been suggested as methods for validating final calibration models.

In Section 2.1, we outlined a generalized method for selecting and validating a calibration model. The procedure first tested for heteroscedasticity using an F-test on the lower limit of quantification (LLOQ) and the upper limit of quantification (ULOQ) measurements. Weight selection was performed using variance evaluation to examine which weighting (no weight, $1/x$ or $1/x^2$) produced the smallest spread in weighted normalized variances. A partial F-test was then used to select the model order (quadratic or linear). Finally, the model was validated through testing the residuals for normality (Kolmogorov-Smirnov (KS) or Cramer-von Mises (CVM) test). This procedure is an automated, analyst-independent approach to selection and validation of calibration models.

In choosing each part of this procedure, the main considerations were accuracy of the result, robustness, ease of use, mathematical soundness and how adequately it fit with real situations faced in toxicological analyses. In this paper, the different procedures tested are detailed. The mathematical reasoning justifying the selected procedures, buttressed by experimental work with 50 analytes quantified by LC-MS/MS are presented.

2.2.3 Materials and methods

LC-MS/MS quantification

Fifty analytes were spiked in bovine blood at concentrations of 5, 10, 15, 50, 75, 100, 400, 500 and 1000 ng/mL to produce a set of calibration standards. The analytes were obtained from Cerilliant (Round Rock, TX, USA) and belonged to the benzodiazepine, opiate, cocaine and amphetamine families and are listed in Appendix C.1. Amphetamine-D_8, benzoylecgonine-D_3, clonazepam-D_4, cocaetylene-D_8, cocaine-D_3, ephedrine-D_3, diazepam-D_5, MDEA-D_5, oxycodone-D_3, methamphetamine-D_5 and codeine-D_3 (Cerilliant) were used as internal standards for the analytes (concentrations and internal standard assignation are available in Appendix C.1). Sample preparation and analysis details can be found in Section 2.1.3. This method has been validated according to ISO 17025 [71] and CAN-P-1578 [73]

guidelines and is currently used as a routine quantification method. Although in the first paper, seven measurement replicates were found to improve the success rate, we present here data collected in accordance with current SWGTOX guidelines which dictate five measurement replicates. To generate inter-day, inter-extraction data, a set of standards was extracted and analyzed in the same day for 5 different days. To generate intra-day, inter-extraction data, five aliquots of the set of standards were extracted separately and injected on the same day. To generate intra-day, intra-extraction data, a set of standards was extracted and five injections of the extract were performed on the same day. Finally, inter-day, intra-extraction data were generated by extracting one set of standards and injecting an aliquot of the extract each day for 5 different days. Data analysis was performed with Multiquant™ (AB Sciex, Framingham, MA, USA).

Simulated data sets

The procedure used to generate simulated data sets for the six different calibration models was described in Section 2.1.3.

Heteroscedasticity testing

The procedure and calculations used to perform an F-test for heteroscedasticity testing were described in Section 2.1.3.

Tests for weight selection

Examination of the variance graph A plot of the variance as a function of the concentration was generated for each analyte. The variance for the five measurements at each concentration level was obtained in Excel 2010 (Microsoft, Redmond, WA, USA) by the formula `= VAR.S(Measurements)`. Constant variance across the calibration range indicated unweighted regression, while a linear increase in variance indicated $1/x$ and a parabolic increase indicated that a $1/x^2$ weighting factor should be selected [88, 116, 121]. Weight selection was thus based on visual inspection of the graph by the analyst.

Variance evaluation The procedure and calculations used to perform the variance evaluation for weight selection were described in Section 2.1.3.

Tests for model order

Analysis of variance lack-of-fit Coefficients for linear ($y_L = b_1x + b_0$) and quadratic ($y_Q = b_2x^2 + b_1x + b_0$) equations were obtained for each analyte using the data for all replicates ($n = 45$). The mean square for the pure (or experimental) error (MS_{PE}) was first calculated by [86]

$$MS_{PE} = \frac{SS_{PE}}{\text{dof}_{PE}} = \frac{\sum_i \sum_j w_i \times (y_{ij} - \overline{y_i})^2}{n_{ij} - k} \tag{2.10}$$

where SS_{PE} was the sum of squares of the pure error, dof_{PE} was the number of degrees of freedom of the SS_{PE} value, w_i was the weighting applied at the i^{th} concentration level (e.g., a $1/x$ weighting at the 5 ng/mL concentration level will be $1/5 = 0.2$), y_{ij} was the j^{th} measurement at the i^{th} concentration level, $\overline{y_i}$ was the average of all measurements at the i^{th} concentration level, n_{ij} was the total number of measurements (9 concentration levels $\times$ 5 replicates = 45 measurements) and k was the number of concentration levels (maximum value of i, here 9).

The mean square for the lack-of-fit (MS_{LoF}) was then calculated by

$$MS_{LoF} = \frac{SS_{LoF}}{\text{dof}_{LoF}} = \frac{\sum_i \sum_j w_i \times (\overline{y_i} - \hat{y}_i)^2}{k - z} \tag{2.11}$$

where SS_{LoF} was the sum of squares of the lack-of-fit, dof_{LoF} was the degrees of freedom of the SS_{LoF} value, $\hat{y}_i$ was the predicted measurement at the i^{th} concentration level (obtained by inserting the concentration in the calibration equation) and z was the number of regression parameters ($z = 3$ for quadratic models and 2 for linear models).

The F value was obtained through

$$F_{model} = \frac{MS_{LoF}}{MS_{PE}} \tag{2.12}$$

and the probability (P) associated with the F statistic was calculated using the Excel 2010 function `= F.DIST.RT(Fmodel; (k - z); (n_ij - k))`.

For each analyte, these calculations were performed for the quadratic and linear calibration models. The P_{quad} (P-value for the quadratic model) was compared with the P_{linear},

and the model order with the largest P-value was retained for calibration.

PARTIAL F-TEST The procedure and calculations used to perform a partial F-test for selection of model order are described in Section 2.1.3.

TESTS FOR VALIDATION OF THE CHOSEN MODEL

STANDARDIZED RESIDUALS GRAPH Standardized residuals were calculated for each measurement by [131][3]

$$s_{e_{ij}} = \frac{(y_{ij} - \hat{y}_i)}{s_e \times \sqrt{1 - h_{ij}}} \tag{2.13}$$

where $s_{e_{ij}}$ was the standardized residual for the j^{th} measurement at the i^{th} concentration level, $\hat{y}_i$ was the predicted measurement at the i^{th} concentration level, s_e was the estimate of the standard deviation of the residuals and h_{ij} was the leverage value for observation ij.

The estimate of the standard deviation of the residuals s_e was calculated by

$$s_e = \sqrt{\frac{\sum_1^{i \times j} (y_{ij} - \hat{y}_i)^2}{n - z}} \tag{2.14}$$

where z was the number of regression parameters, i.e., $z = 2$ for linear regressions or 3 for quadratic regressions.

The leverage value for each observation h_{ij} was calculated by

$$h_{ij} = \frac{1}{n} - \frac{(x_{ij} - \overline{x})^2}{\sum_1^{i \times j} (x_{ij} - \overline{x})^2} \tag{2.15}$$

where x_{ij} was the concentration associated to the measurement y_{ij} , n was the total number of measurements and $\overline{x}$ was the average concentration over all standards measured (9 concentration levels x 5 replicates = 45 concentration values).

[3]Note that the model weights are used in this equation via the fitted values $\hat{y}_i$. According to the theory of linear regression, $y_{ij} - \hat{y}_i$ should approach a normal distribution. Across i, the residuals $y_{ij} - \hat{y}_i$ for heteroscedastic data will have changing variance. However this section is a visual evaluation as to whether the residuals are randomly distributed around $y = 0$ line. Multiplying $y_{ij} - \hat{y}_i$ by a weighting factor will only change the spread, but not the sign, nor the conclusion of this visual evaluation.

All standardized residuals were then plotted against their concentration level (i) [75]. If the calibration model is appropriate, the residuals should be distributed randomly around the $y = 0$ line [85, 132].

Analysis of variance - lack-of-fit The ANOVA-LoF procedure can be used as a test for both model order and model validation. Once a model order had been chosen using the procedure presented above in Section 2.2.3, the P-value could be (re-)used to validate the calibration model. A P-value above the 0.05 threshold indicated that the error attributable to the LoF was not significant compared with the experimental or pure error [85]; therefore, the model was validated. A P-value below 0.05 marked an LoF error significantly larger than the experimental error; the model used was therefore not an adequate fit to the experimental data.

Normality of the residuals Calculations for testing the normality of the standardized residuals will not be detailed here since they are beyond the scope of this paper and require a specialized knowledge of statistics [133]. However, a general description of the operations performed is provided here. Readers interested in a fuller understanding can consult the R scripts available in Appendix B.3 which details all calculations.

When the chosen calibration model accurately describes all systematic trends in the data, the residuals correspond to pure random error. This means the residuals should be randomly distributed around zero (i.e., follow a normal distribution). Thus, normality testing of the residuals is an appropriate means of validating a calibration model [132]. This approach is conceptually similar to the standardized residuals plot, but produces a definitive, analyst-independent result.

To evaluate whether the residuals are normally distributed, the distribution function of the residuals is compared with the distribution function obtained from the expected normal distribution (see Figure 2.2.1).

A distribution function plots the proportion of data points that are smaller or equal to all values present in the distribution. Normally distributed residuals produce a sigmoidal distribution curve. In Figure 2.2.1a for example, at residual = 0 (when the experimental data value = predicted value by the model), the normal distribution function has $y = 50\%$;

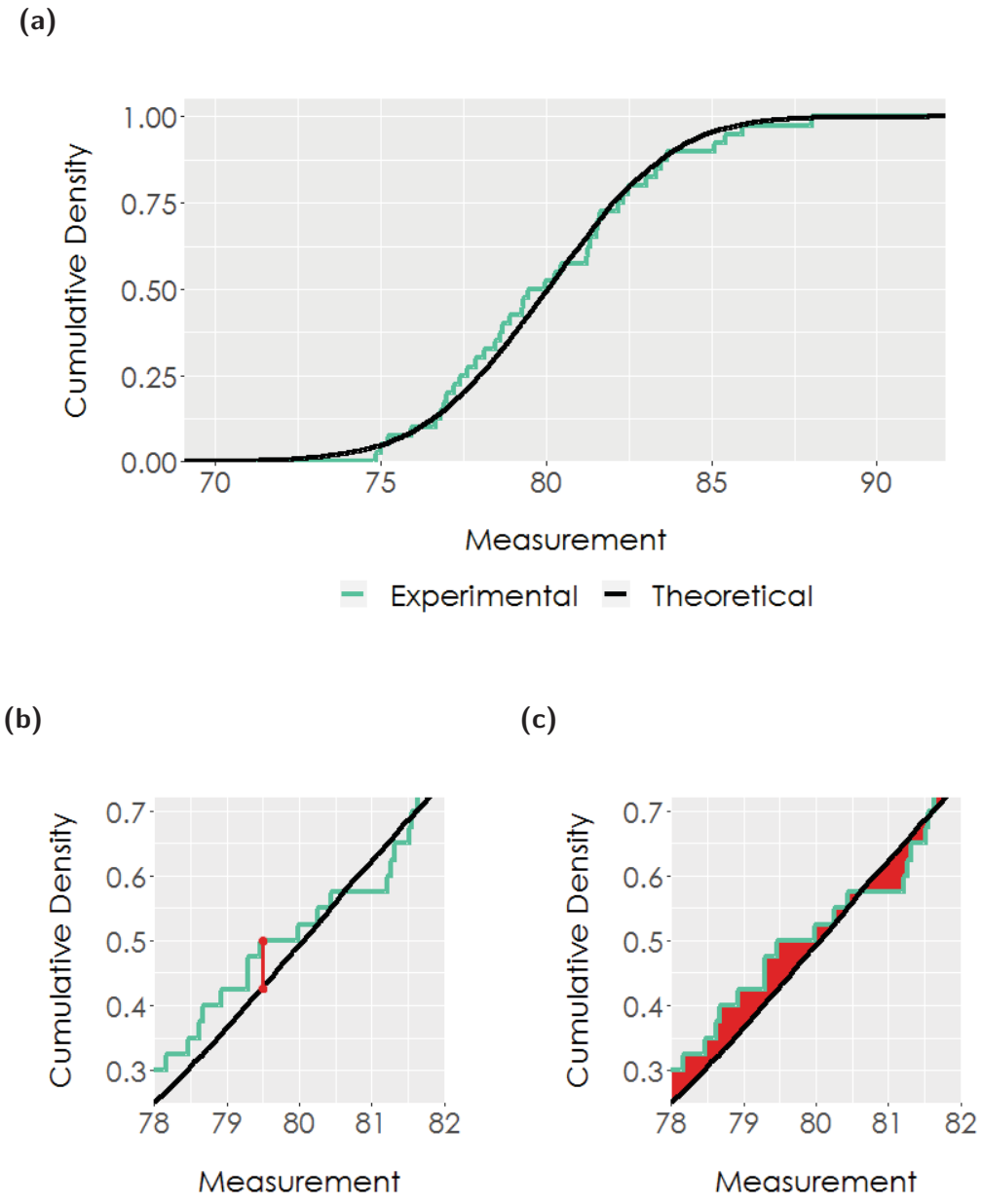


Figure 2.2.1: Graphical representation of normality testing process. **(a)** Graph of an expected normal distribution function and an experimental residuals distribution function. **(b)** Illustration of the calculation of the KS statistic. **(c)** Illustration of the calculation of the CVM statistic.

therefore, 50% of all the residuals are smaller than or equal to 0.

Two commonly used statistics exist to estimate whether there is a significant difference between the expected (normal) distribution and the experimental distribution of the residuals: CVM and KS. The CVM statistic is the integral of the squared difference between both

distribution functions, meaning the area between both curves as seen in Figure 2.2.1b [133]. The KS statistic is the maximal vertical distance observed between both distribution functions, as shown in Figure 2.2.1c [134].

Once the test statistic is obtained for the experimental residuals data set, a P-value needs to be calculated. It represents the probability of obtaining a statistic of that value when the residuals are indeed normally distributed. The distribution of the CVM and KS statistics is governed by complex laws. The best approach is therefore to estimate those distributions using a bootstrap approach [135]. Briefly, this numerical method synthesizes large numbers of residuals data sets by employing resampling of the experimental residuals [136]. Then, y_{ij} values associated with these synthetic residuals sets are utilized to determine parameters b_0, b_1 and b_2, which in turn permit the calculation of residuals. Since these residuals are derived from bootstrapped data, they are pseudo-residuals. The CVM or KS statistic from each set of pseudo-residuals is then calculated. Inclusion of a correction factor is necessary when calculating the results of the CVM or KS laws from pseudo-residuals [137]. Using a large ensemble (here, 1000) of the derived CVM or KS statistic allows the probability of observing the experimental statistic to be determined.

2.2.4 Results and discussion

Raw data necessary and replicate analysis

Validation guidelines from the Scientific Working Group for Toxicology (SWGTOX) state that "at least six different non-zero concentrations" should be used to choose the calibration model, but that additional calibration levels may be required to characterize properly higher order models [75]. This is a generally agreed upon standard [118, 120, 138].

Regarding the number of replicate measurements necessary for an adequate calibration model selection and validation process, the SWGTOX suggests that "a minimum of five replicates per concentration is required" [75]. Other authors suggest increasing the number of replicates up to nine [122]. In the opinion of the authors, in the context of method validation in toxicology, seven replicates is a good compromise between adequate statistical test performance and the amount of work that must be done in the laboratory (see Table 2.1.2). It is important to understand that with larger %RSD, more replicate measurements are needed before the estimation from the data converges to the real variance value.

The SWGTOX guidelines suggest that an experimental setup using inter-day and inter-extraction replicate analysis should be used [75]. However, using such a setup results in an increased variability in the data compared with an intra-day and/or intraextraction setup. The less robust the method is, the more this is true. Good analytical methods aim to curb this inter-day and interextraction variability by instrument maintenance and suitability tests as well as the use of stable isotope-labeled internal standards. Despite all of these measures, it is expected that there will always be some inter-day and inter-extraction variability left. The problem with this increased variability is that, when it is included in the calibration, it can mask patterns in the data, such as quadraticity. This will result in a higher mis-selection rate for the order of the model. It is important to point out that this is not a failure of this particular algorithm for the selection of the calibration model. Masking of the patterns occurs in the data set itself, and any calibration model selection scheme will suffer from this.

The purpose of any calibration model selection scheme is to find out the model underlying the data generated in production. Therefore, the authors suggest that calibration model selection should be performed on data obtained with an experimental setup that represents what will occur during that production process. If the method states that two series of standards are to be extracted and analyzed with each batch to create the calibration curve (intra-day, inter-extraction), then this is the data that should be used for selection of the calibration model during the validation process. More often than not, the method calls for one series of standards to be extracted and analyzed (sometimes in duplicate) with each batch to create the calibration curves. This was the case with the LC-MS/MS method presented here. In this situation, intra-day, intra-extraction data should be used to select the calibration model.

In our algorithm, pattern masking through increased variability creates problems on two fronts: the test for model order selection and the test for validation of the chosen model. Both showed a lower performance due to increased variability introduced by the inter-day/ inter-extraction setup.

Using simulated quadratic data sets, we observed that increased variability can result in a masking of the quadratic nature of the data. A data set that is, in fact, quadratic can therefore be identified as linear when performing the partial F-test for selecting model order. The experimental design combining inter-day/inter-extraction data can therefore incorporate systematic variability, which is then treated mathematically as random error during regression. Therefore, any improvement afforded by a quadratic fit can be obscured by the

artificially large variability of the measurements and mask the underlying nature of the data. An increased variability will have no impact when the instrument's pattern of response is linear, since the null hypothesis of the test is that the model is linear. These predictions are demonstrated by the experimental results obtained. When intra-day, intra-extraction data were used, the partial F-test concluded that a quadratic model should be used for 47 out of 50 analytes (Appendix C.1). On the other hand, this number dropped to 40, 43 and 41 out of 50 when intra-day/inter-extraction, inter-day/intra-extraction and inter-day/inter-extraction data are used, respectively. This phenomenon was confirmed using simulated data based on experimentally obtained calibration curves (Appendix B.1). These results show that as the variability (%RSD) increases in quadratic models with greater variability at the high end of the curve ($1/x$ or $1/x^2$ weighting), the success rate of the partial F-test for order selection drops when the curvature is weak, as expected.

Similarly, normality testing of the residuals can have a hard time pinpointing the inadequacy of a model when the apparent variability of the measurements is increased by using inter-day/inter-extraction data. Departures from normality would have to be much greater to be considered significant. Therefore, a model that would be identified as incorrect using intra-day, intra-extraction data could justifiably be validated if an inter-day and/or interextraction data set is used, which would mean an inadequate model would be missed by the test.

That being said, we do need to consider sources of variability that might change the instrument's pattern of response over time if we want to select a robust model that can be used over a long time period. The way to do this is not to use inter-day/inter-extraction data, as one might first think. Rather, the procedure for selection and validation of a calibration model should be performed multiple times throughout the whole validation process (which will most likely span two or more months). A minimum of three times is recommended by the authors, taking care to change (as appropriate) analysts, instruments, standard lots, spiking matrix lots, etc. If all procedures choose the same calibration model, it is a robust model and should be kept for future analyses. In our experience, when results alternate between linear and quadratic model and there is no instrument stability issue, then it is because results of the partial F-test are near the threshold (P-value is near 0.05), and often the curvature of the quadratic equation is small (quadratic term is close to zero). In this situation, both models would give similar results in terms of accuracy and precision, and we recommend that the simplest model (linear) should be used.

To summarize, we recommend the use of intra-day/intra-batch data (if this is the calibra-

tion scheme that will be used in production) in order to obtain the most accurate selection of the calibration model as possible, and repetition of this selection process over three different days to take into account the inter-batch/inter-day variability. Using inter-batch and/or inter-day data to feed to our calibration model selection and validation scheme is possible, however, the user has to be aware that some tests will have a lower performance as a result.

Forcing a calibration equation through the origin

Several data analysis packages offer the option of forcing the calibration function through the origin. In principle, the reasoning behind this option is that the signal is expected to be zero when the concentration is zero. However, the SWGTOX states in their validation guidelines that "the origin shall not be included as a calibration point" [75] which we agree is the correct approach. Forcing the calibration function through the origin means creating data that have not been measured. Although theoretically, the calibration function is expected to go through the origin, there can be several valid reasons why it may not do so experimentally. These reasons may vary from technique to technique but would include blank contamination, undetected non-linear behavior near the blank, insufficient background/baseline removal etc. Additionally, it is in conflict with the other statistical approaches used to choose and validate a calibration model, since it artificially modifies the error at the low end of the calibration curve and alters the calibration coefficients. Due to these mathematical concerns, the approach of forcing a calibration function through the origin was not used and is not recommended.

Heteroscedasticity testing

With heteroscedastic data, the observed variance and the interrelated precision of the measurement change across the concentration range. In standard least-squares regression, the "best fit" occurs when the sum of the squared errors between the data and the calibration function is minimized. In a heteroscedastic situation, the data with the largest variance will make the largest contribution to the sum of the squared errors and dominate selection of the coefficients - which is opposite to what is appropriate. Reliability of the parameter estimates will be reduced [85]. This problem can be largely "overcome by introducing weighting factors inversely proportional to the variance" [85]. This scaling increases the contribution to the squared error of the high precision data and reduces the contribution of the low precision data when establishing the calibration coefficients.

The presence of heteroscedasticity can be tested in different ways via the F-test, Cochran, Hartley and Bartlett tests [122]. The F-test was chosen for its simplicity of application. It compares the variance of two groups of data, in this application the LLOQ and ULOQ [85].

In many analytical chemistry experimental systems, there is a sound theoretical basis for heteroscedastic data [84, 115, 139], which can mostly be simplified to two primary types of noise (variance): additive (constant) and multiplicative (scales with concentration). Methods where multiplicative noise is the dominant source of error will produce heteroscedastic data [84]. The expectation that the variance increases with concentration when using MS detection dictates that a unilateral F-test should be applied (alternative hypothesis being that the variance at the ULOQ level is greater than the variance at the LLOQ level).

Two situations can create false negatives when the F-test is applied this way. In the first case, the heteroscedasticity pattern is more exotic than what is expected in the simple multiplicative noise case and the variance is not significantly different at the ULOQ and the LLOQ but higher or lower in the center of the calibration range. This is extremely atypical of toxicology methodologies, and standard data analysis software should not be used in this situation since there is no appropriate weighting scheme. In any event, a quick examination of the variance plot, produced as a PDF document when the R script is run, should ensure that this situation does not go unnoticed.

The second situation that might produce a false negative would be the presence of an outlier at the LLOQ. This can artificially increase the observed variance at the LLOQ and mask the heteroscedasticity. Analysts should always be attentive to data points vastly different (4σ or more away from the rest of the group), and understand that inclusion will affect their results. Of course, if the selection and validation of calibration model analysis is repeated more than once during the validation process, then the probability that this issue would occur each time is extremely small.

All of the 50 analytes studied in the LC-MS/MS method tested positive for heteroscedasticity, with P-values from 2.2×10^{-10} to 2.2×10^{-4} (Appendix C.1). When large dynamic ranges are analyzed on LC-MS/MS, this type of heteroscedasticity is quite common [118, 140–142].

Selection of weight

Once a data set had been shown to be heteroscedastic, the next step was to select an appropriate weighting factor. To correct completely for the heteroscedasticity issue, the inverse of the variance at each concentration level should be used as the weighting factor [85]. However, the variance typically follows set patterns across the calibration range: it increases in proportion to the concentration (x) or with the square of the concentration (x^2) [116, 121]. As these are the most common variance patterns, they are included as weighting options in almost all data analysis software. The use of the experimentally measured variance pattern (that might even vary over time) does not appear a good choice.

The first test option evaluated was visual examination of the variance plot by the analyst. This procedure worked well, but it was a manual, analyst-dependent process, which was exactly what we were trying to avoid with this new protocol. When analyzing data for the 50 analytes, the analyst often ended-up questioning his or her own weighting choice, especially hesitating between $1/x$ and $1/x^2$ when outliers disturbed the trend in variance.

The second option considered was variance evaluation, where the weighting model with the smallest spread of normalized weighted variances was chosen. This test was automated and the result was analyst independent. Under this test, all 50 analytes were deemed to have $1/x^2$ (Appendix C.1). This result was expected given that samples were analyzed on an LC-MS/MS with calibrations spanning a few orders of magnitude [88, 116]. It is, however, interesting to note that removing a few upper calibration levels would obscure the power relationship and result in an apparently linear variance plot, therefore changing the weighting factor. Using the simulated data for different weighting factors gave a better idea of the performance of this test. The outcome, described in Section 2.1.4, showed good test performance (2% failure) with 10 measurement replicates.

Selection of model order

In the SWGTOX guidelines for method validation, selection of model order is merged with validation of the model and tested with the examination of the standardized residuals plot [75]. However, it also stated that "there are other appropriate alternatives to evaluate calibration models (i.e., ANOVA-LoF test for unweighted linear models, checking for the significance of the second-order term in quadratic models, assessment of coefficient of determination for linear models)" [75]. We examined three different options for model order

selection: ANOVA-LoF, significance of the second-order term and the partial F-test. Visual examination of the standardized residuals plot was not considered due to its manual, analyst-dependent nature.

The most common form of ANOVA-LoF test is intended for unweighted linear models, as stated by the SWGTOX [75]. However, in the version presented here, the calculations have been adapted to assess weighted models of n^{th} order [86]. In this test, as with the standardized residuals, selection of model order is somewhat merged with model validation. The result is the selection of the model order based on comparison of two P-values. From a statistical point of view, this comparison is a precarious procedure. Statistical tests should be treated as having a binary outcome (acceptance or rejection of the null hypothesis). Moreover, the P-values are approximations and are therefore not exact values - especially in the extremes. For these reasons, comparing P-values that are nearly identical and/or are both at extreme values is not without risk. Nevertheless, when this procedure was applied to select model order, all but two analytes (naltrexone and phenylpropanolamine) were identified as quadratic models (Appendix C.1).

Calculations for the significance of the second-order term were not performed. The first thing to highlight is that calculations for this test are more involved than one might think [86]. An analyst might be tempted to establish confidence intervals of $(\pm^{t \times s}/_{\sqrt{n}})$ around the five b_2 values resultant from the five regressions, but this is a faulty strategy. Rather, calculation of the significance of the second-order term uses the variance-covariance matrix of the regression coefficients. Unfortunately, this test relies heavily on an assumption of normality, which is precarious[4]. With the limited data sets obtained for selection and validation of the calibration model, it is probable that normality has not kicked off yet. A bootstrap approach could be used to circumvent this issue, but would require significantly more involved calculations. Taking into account all of these considerations, we have decided not to perform calculations for this test of model order selection and to select another test in the final procedure.

The partial F-test is conceptually and mathematically simple, more so than the ANOVA-LoF and significance of the second-order term tests. Three analytes out of the fifty were classified as linear using the partial F-test: hydroxy-alprazolam, naltrexone and phenyl-

[4]Although this might not be the case in large sample sizes, where the number of replicates is > 10, several numerical experiments were carried out and suggested that a bootstrap approach worked better for designs with a restricted number of data.

propanolamine (Appendix C.1). This number was higher than with the ANOVA-LoF test because the partial F-test is more stringent against overfitting. Another advantage of this test was that it clearly separated model order selection from model validation, leading to a reduced degree of confusion. Given its several advantages over the other available tests, we decided to retain the partial F-test for model order selection.

Tests for validation of the chosen model

The main model validation metric test suggested by the SWGTOX is ANOVA-LoF [75]. This test was applied to the 50 analytes of the LC-MS/MS method (Appendix C.1). When using intra-day/intra-extraction data, ANOVA-LoF systematically rejected the chosen calibration model. Given the fact that the accuracies and precisions were acceptable for all analytes, this was, on the face of it, a surprising result. It turns out this test is very sensitive to experimental design, in particular the number of replicates and/or the number of calibration levels. As a result, it is difficult to produce a P-value above 0.05 given the typical variances observed in analytical methodologies. The excessive sensitivity of this test limits its practical applicability.

On the other hand, residual normality testing with either the KS or CVM test accepted the chosen calibration model for all 50 analytes. Between 0 and 2% of the simulated data sets saw their chosen calibration model rejected (see Table 2.1.2). Simulated non-normal data sets (not shown) were also rejected. Therefore, this test can detect most problems with the calibration model, but is more robust than ANOVA-LoF.

Residuals normality testing is based on the same underlying principle as the examination of the standardized residuals graph, which qualitatively verifies the randomness of residuals. Again, the major issue with this visual technique is the lack of automation and the analyst dependency of the result. The problem is that the visual differences between a residuals graph with a distribution deemed normal by the CVM test and a residuals graph with a non-normal distribution can be very slim indeed (see Appendix C.2). Knowing the results of the normality tests, the analyst can, *a posteriori*, find reasons for why the calibration model for Analyte A (Figure C.2.1b) was rejected and not for Analyte B (Figure C.2.1a), but these examples clearly emphasize the lack of rigor of the visual examination of standardized residuals plot and highlight the potential for analyst bias.

Residuals normality testing by KS or CVM was therefore selected for its bias-free, binary

output which is a great advantage over subjective examination of the residuals graph. The drawback to this test is the complexity of the calculations involved, but the tool we have created with RStudio makes implementation of this test fairly easy.

2.2.5 Conclusions

The impact of different types of replicate analysis were evaluated. The use of inter-day/inter-extraction validation schemes was found to introduce inappropriate variability into the data set, which masked some underlying trends (quadraticity, departure from normality). It is therefore suggested that validation data should mimic what will be done in production (i.e., use intra-day/intra-extraction data where daily calibration with one set of standards is to be used in production). Ideally, the process of selection and validation of the calibration model will be repeated at multiple points in the global validation process.

Forcing the calibration function through the origin was not applied or tested due to its artificial alteration of the error and calibration coefficients.

Heteroscedasticity was evaluated through the unidirectional F-test applied using the LLOQ and the ULOQ calibration levels. Increasing variance with concentration was the overwhelmingly observed pattern and is encountered with many instrumental analytical methods, which justifies the general use of weighted regression methods.

Selection of the weight was found to be best performed through a variance evaluation. Examination of the variance graph was rejected because of analyst dependency.

The partial F-test was chosen to perform the selection of model order for its conceptual and mathematical simplicity and validity. ANOVA-LoF and significance of the second-order term were examined, but rejected because of their strong dependence on normality of the data, utilization of the dubious P-value comparison procedure and calculation complexity.

Residual normality testing was chosen as the calibration model validation procedure because of its robustness. ANOVA-LoF, the main alternative metric, is too sensitive to experimental design (replicates, calibration levels) to be truly useful. On the other hand, examination of the standardized residuals graph is not a bias-free approach.

Through examination of different testing alternatives, we have created a selection and

validation procedure for calibration models that can be used in the overall validation of all quantitative methods in toxicology. This stepwise, bias-free method is a simple automated tool toward better calibration model selection.

2.2.6 Funding

Brigitte Desharnais, Félix Camirand Lemyre and Cameron D. Skinner gratefully acknowledge support of the National Sciences and Engineering Research Council of Canada. Brigitte Desharnais also gratefully acknowledges the support of the Fonds de recherche du Québec - Nature et technologies.

2.2.7 Acknowledgments

The authors wish to thank Cynthia Côté for her thorough review and comments as well as Sue-Lan Pearring and Chris H. House for their opinion on key matters. The authors are also grateful to Gabrielle Daigneault, Lucie Vaillancourt, Julie Laquerre and Marc-André Morel for their technical work.

2.3 Calibration models: small molecules, bioanalysis and beyond

Sections 2.1 and 2.2 presented different techniques to select a calibration model, including the weighting and order, evaluated those means with a mixture of experimental data and *in silico* simulations, and packaged the optimal combination in a user-friendly R script performing automated calibration model selection and validation. Selecting the correct calibration model allows for a more accurate quantification during the validation procedures and in production operations, indeed, selecting an incorrect weight can lead to "one order of magnitude of precision loss in the low concentration region" [88].

This procedure starts by testing homoscedasticity of the data before testing non-linearity (instead of the opposite) for two reasons. First, most linearity tests require that the weighting is already known to be carried out. Second, calibration curves in analytical chemistry are generally expected to be heteroscedastic but only weakly non-linear.

The majority of the procedure relies on the application of hypothesis tests. Null hypothesis significance testing (NHST) yields a P-value, which is defined as the probability of data as, or more, extreme occurring if the null hypothesis (typically of no difference) is true. Thus, when the null hypothesis is true (e.g., data is actually homoscedastic), it is expected that the test will return an incorrect result of significant difference (P-value < 0.05) 5% of the time. This is the Type-I error rate. The Type-II error rate, the probability of the test returning a non-significant finding if the alternative hypothesis (of difference) is true, actually depends on the *effect size*, e.g., how much heteroscedasticity there is in the data, or how curved the calibration curve is. Table 2.1.2 shows results from an applied form of power testing, where heteroscedasticity and curvature values relevant to the type of analysis were evaluated.

The previous sections demonstrated the process using small molecule (xenobiotic) analysis, which is the focus of forensic toxicology. Experimental data was thus readily accessible for those analytes. Additionally, papers concerned with small molecule analysis (such as Chapter 2, which has been published [1, 2]) are more likely to come to the attention of, and have an impact in the targeted community. But the developed process is more widely applicable than the confines of forensic toxicology and xenobiotic analysis. Any quantification method relying on the use of a calibration curve can benefit from this study, which thus has a broad application field.

Because CYP 2D6 and 3A4 enzymes will be quantified in liver tissue by means of calibration on LC-MS/MS instrumentation, the procedure presented above is directly applicable and will help achieve a more accurate result.

But even if the calibration model is correctly chosen, the accuracy of CYP quantification will be challenged by another problem: the near impossibility of finding analyte free (blank) liver tissue to prepare the matrix matched standards needed for calibration. Indeed, except for poor metabolizers, liver tissue in individuals will always contain CYP enzymes, albeit at different levels. This will introduce a bias in the quantification process.

Different measures are possible to side-step the problem, but it would be better if there was a solution to comprehensively address this challenge, one that satisfies criteria of accuracy and efficiency. This is what Chapter 3 presents.

3

A Tool for Automatic Correction of Endogenous Concentrations: Application to BHB Analysis by LC-MS/MS and GC-MS

This chapter is the integral text from:

A Tool for Automatic Correction of Endogenous Concentrations: Application to BHB Analysis by LC-MS/MS and GC-MS

Brigitte Desharnais, Marie-Jo Lajoie, Julie Laquerre, Stéphanie Savard, Pascal Mireault, Cameron D. Skinner

Journal of Analytical Toxicology **43** (7) (2019) 512-519 [3]

3.1 Abstract

Several substances relevant for forensic toxicology purposes have an endogenous presence in biological matrices: β-hydroxybutyric acid (BHB), γ-hydroxybutyric acid (GHB), steroids and human insulin, to name only a few. The presence of significant amounts of these endogenous substances in the biological matrix used to prepare calibration standards and quality

control samples (QCs) can compromise validation steps and quantitative analyses. Several approaches to overcome this problem have been suggested, including using an analog matrix or analyte, relying entirely on standard addition analyses for these analytes, or simply ignoring the endogenous contribution provided that it is small enough. Although these approaches side-step the issue of endogenous analyte presence in spiked matrix-matched samples, they create serious problems with regards to the accuracy of the analyses or production capacity. We present here a solution that addresses head-on the problem of endogenous concentrations in matrices used for calibration standards and quality control purposes. The endogenous analyte concentration is estimated via a standard-addition type process. This estimated concentration, plus the spiked concentration are then used as the de facto analyte concentration present in the sample. These de facto concentrations are then used in data analysis software (MultiQuant, Mass Hunter, etc.) as the sample's concentration. This yields an accurate quantification of the analyte, free from interference of the endogenous contribution. This de facto correction has been applied in a production setting on two BHB quantification methods (GC-MS and LC-MS/MS), allowing the rectification of BHB biases of up to 30 $\mu g/mL$. The additional error introduced by this correction procedure is minimal, although the exact amount will be highly method-dependent. The endogenous concentration correction process has been automated with an R script. The final procedure is therefore highly efficient, only adding four mouse clicks to the data analysis operations.

3.2 Introduction

Almost all bioanalytical method validation procedures involve spiking authentic matrices to evaluate figures of interest, such as precision, accuracy and matrix effects [75]. When endogenous analytes are targeted for quantification, the usual validation protocol can be difficult or impossible to apply. The authentic matrices (e.g., blood, urine or hair) which are typically used to spike known concentrations of analytes cannot be found blank of the targeted substance. Ignoring the endogenous amount, invariably of unknown magnitude, present in these authentic matrices introduces a systematic bias in all analyses. Not only does this pose a problem at the method validation stage, but standard production operations are compromised as well. The accepted practice in bioanalysis methods is to spike calibration standards and quality control samples (QCs) with internal standards in authentic matrices (or a pool of authentic matrices) to control for potential matrix effects. Again in this situation, the endogenous analyte will introduce a bias in the calibration equation, compromising, to various degrees, the quantitative accuracy of the result.

A number of analytes of interest in forensic toxicology are endogenous compounds. A few common examples include γ-hydroxybutyrate (GHB), cyanide, carboxyhemoglobin (HbCO), insulin and steroids. Some other analytes, such as alcohol consumption markers or caffeine are not endogenous [143–145] but can be highly prevalent in the population [144, 146]; thus finding authentic matrices free from them can also be difficult. The case of the endogenous compound β-hydroxybutyric acid (BHB) is specifically examined here. BHB acts as a diabetic or alcoholic ketoacidosis biomarker, and is a hyperglycemia indicator [147, 148]. Concentrations up to 50 $\mu g/mL$ in blood are generally considered normal, whereas concentrations from 50 to 200 $\mu g/mL$ are considered to be elevated [147, 148]. The decision point with regards to ketoacidosis is generally between 200 and 250 $\mu g/mL$ [147, 148].

Some publications have highlighted the problem posed by these analytes in the method validation and production contexts [93]. A number of possible solutions to this issue have been suggested. A comprehensive, albeit time and resource-consuming option, is to perform quantification of those analytes via standard addition methods [93–95, 101]. While this effectively eliminates the issue of endogenous presence in the spiking matrix, its use in a high-volume production context is not sustainable. The use of a surrogate matrix [93, 96–98], e.g., synthetic blood, or a surrogate analyte [93, 98–100], e.g., a stable isotope-labeled version of the targeted analyte can also be considered. In the end, these methods yield a more or less accurate approximation of the figure sought. For example, the accuracy of the target analyte in real blood cannot be assured, since the assumption that the targeted analyte behaves in synthetic blood in the same way as it would in real blood is probably untenable for a wide variety of analytes due to different analyte specific matrix effects [93]. In that sense, while bioanalysts might settle for a surrogate matrix or analyte, it is clearly a less than perfect solution. Finally, it is possible to ignore entirely the endogenous concentration in the spiking matrix, provided that it is small enough and can be dismissed as a low-level interference (e.g., <10% of lowest calibration standard) [75, 94]. Of course, this approach is quite appealing in its simplicity. The downside is that a bias, however small, will systematically be included in all the quantitative analyses performed. Additionally, this process might require the screening of several matrix sources (e.g., blood lots) to find one with an endogenous level low enough to fit the acceptance criteria, using precious time and resources.

In reality, these methods have been designed to side-step the issue of endogenous analytes in spiking matrices, but they each have their significant drawback and do not address the problem head-on. In this paper, we suggest a simple and comprehensive automated solution to correct for the presence of endogenous analytes in spiking matrices.

3.3 Materials and methods

3.3.1 GC-MS BHB analysis

Calibration standards and quality control samples were spiked in ante-mortem human whole blood (Utak, 44600-WB(F), Valencia, CA, USA) at concentrations of 0.00, 10.0, 20.0, 50.0, 100, 200, 425 and 500 $\mu g/mL$ (standards) and 60.0, 150 and 375 $\mu g/mL$ (QCs) using BHB (Sigma-Aldrich, H6501, Saint-Louis, MO, USA).

In a conical tube, 50.0 μL of blood and 25.0 μL of internal standard solution (400 $\mu g/mL$ BHB-D_4 dissolved in methanol, CDN isotopes, D-6088, Pointe-Claire, QC, Canada) were mixed. Protein precipitation was performed by adding 225 μL of acetonitrile and vortexing immediately. After centrifugation (9 600$\times g$, 7 min), 100 μL of supernatant was evaporated to dryness under nitrogen, reconstituted in 250 μL ethanol ($\geq$99.9%, Commercial Alcohols by Greenfield Global, P210EAAN, Brampton, ON, Canada) and reevaporated to dryness under nitrogen. Derivatization was performed for a maximum of 10 vials at a time to prevent evaporation by adding 70.0 μL acetonitrile (HPLC Grade, $\geq$99.9%, EMD millipore corporation, AX0156-1, Billerica, MA, USA) and 70.0 μL BSTFA+TMCS (99:1) (Cerilliant, B-023, Round Rock, TX, USA) and vortexing. Incubation was tested during method development but showed no significant advantage and was therefore not used [149, 150].

After centrifugation (3 200$\times g$, 5 min), 1 μL of extract was separated on an Agilent HP-5MS column (15 m, 0.25 mm *i.d.*, 0.25 μm film thickness) using a 16.25 minutes separation. The injector was kept at 280 °C and operated in pulsed split mode, with a split ratio of 3:1 (MS Agilent MSD 5975 C) or 20:1 (MS Agilent MSD 5977B HES). The GC was operated with a 1 mL/min helium flow and an oven program of 70 °C to 110 °C over 8 minutes followed by an increase to 280 °C over 4.25 minutes and a plateau of 4 minutes. Pressure at the head of the column was 1.9 psi. The mass spectrometer parameters were set as follows: mass range of 41 to 400 m/z, scan rate of 3.95 scan/min, solvent delay of 2.50 min, acquisition time of 16.25 min, ionization source temperature of 230 °C, quadrupole temperature of 150 °C, transfer capillary temperature of 280 °C and electron impact energy of 70 eV. The method was validated on Agilent gas chromatographs 7890 A and 7890B, equipped with Agilent MSD 5975 C or 5977B HES and automated injector 7693. The data acquisition and analysis software used was MassHunter® B.07.04.2260, B.08.00 build 8.0.598.0.

The method has been validated under SWGTOX [75], ISO 17025:2005 [71] and CAN-P 1578 [73] standards.

3.3.2 LC-MS/MS BHB analysis

Calibration standards and quality control samples were prepared in ante-mortem human whole blood (Utak, 44600-WB(F), Valencia, CA, USA) at concentrations of 0.00, 3.00, 6.00, 30.0, 60.0, 150, 255 and 300 $\mu g/mL$ (standards) and 9.0, 120 and 240 $\mu g/mL$ (QCs) using BHB (Sigma-Aldrich, H6501, Saint-Louis, MO, USA).

In a 96-well plate (2 mL square wells, Fisher Scientific, AB-0932, Ottawa, ON, Canada), 100 μL of blood and 10.0 μL of internal standard solution (330 $\mu g/mL$ GHB-D_6 dissolved in methanol, CND isotopes, D-5462, Pointe-Claire, QC, Canada) were mixed. Note that calibration dynamic range and internal standard differ between the GC-MS and LC-MS/MS methods; each of these validated methods were already in use in the laboratory and the differences are attributable to various historical and practical constraints. The samples were diluted by adding 100 μL of MeOH:0.2% formic acid in water (50:50 v:v) solution (methanol: EMD Millipore corporation, MX0486-1, Billerica, MA,USA; formic acid: Fisher Scientific, A117, Fair Lawn, NJ, USA). Proteins were then precipitated by adding 400 μL of acetone:ACN (30:70 v:v, room temperature) solution (acetone: Fisher Scientific, A949, Fair Lawn, NJ, USA; and acetonitrile: EMD Millipore corporation, AX0156-1, Billerica, MA USA). Samples were thoroughly vortexed after each mixing step. After centrifugation (3200×g, 5 min), 20.0 μL of supernatant was transferred to a second 96-well plate, diluted with 200 μL of MeOH:0.2%formic acid in water (10:90 v:v) and vortexed. An Agilent HPLC 1200 series, or 1260 Infinity, coupled to a Sciex MS/MS 5500 QTrap was used for a 8.75 minute separation of 5 μL of the extract on an Agilent Zorbax Eclipse Plus C18 column (2.1 x 100 mm, 3.5 μm) kept at 50 °C. A step/ramp gradient with 2:98 methanol:10 mM ammonium formate pH 3.0 (mobile phase A) and methanol (mobile phase B) was used. Using a flow of 600 $\mu L/min$, the percentage of mobile phase A was brought from 100% (0.0 min), to 85% (0.3 min), to 80% (3.0 min), to 40% (3.5 min), to 20% (6.0 min), to 0% (6.5 min) for a 0.7 min wash followed by a 0.9 min re-equilibration period. BHB retention time under these conditions was 0.95 min. Acquisition was carried out in negative multiple reaction monitoring (MRM) mode, with a quantitative Q1/Q3 transition of 103 Da/59 Da (collision energy of $-18V$ and collision cell exit potential of $-7V$) and a qualitative Q1/Q3 transition of 103 Da/41 Da (collision energy of $-30V$ and collision cell exit potential of $-10V$). Other mass spectrometer parameters were as follows: MRM detection window of

60 sec, target scan time of 0.25 sec, settling time of 50 msec, break of 3.00 msec, curtain gas (CUR) of 30.00, collision gas (CAD) of 10.00, ion spray voltage (IS) of $-4500\ V$, source heater temperature (TEM) of 700 °C, ion source nebulizer gas (GS1) of 60.00, heater gas (GS2) of 65.00. Data was acquired using Analyst® 1.6.2 build 8489 software and analyzed with Multiquant® 3.0.1 (Version 3.0.6256.0).

The method has been validated under SWGTOX [75], ISO 17025:2005 [71] and CAN-P 1578 [73] standards.

3.3.3 Correction for the endogenous concentration

Several known amounts of analyte were added to distinct aliquots (i.e., a calibration curve in the matrix was prepared) to determine the endogenous concentration of the analyte. This was performed by a two-step process. First, the endogenous analyte concentration in the calibration standards was determined via a standard addition process[1] using the known concentrations that were added to the matrix. Second, this endogenous concentration was then added to the nominal concentrations of the standards and QC samples to generate the de facto analyte concentration present.

If the calibration model selected during method validation [1, 2] was linear, the calibration equation was expressed as:

$$y = b_1 x + b_0 \tag{3.1}$$

The endogenous concentration (x_e) of the analyte in the biological matrix is the x-intercept; in this case, it is calculated as

$$x_e = \left| \frac{b_0}{b_1} \right| \tag{3.2}$$

If the calibration model selected was quadratic, i.e.,

[1] Note that the standard addition process assumes a blank sample would produce zero signal, thus the calibration equation actually simplifies to an intercept equal to zero. In samples with endogenous analyte, the y-intercept is the best estimate of the signal associated with this analyte concentration and therefore the regression should never be forced through the (0, 0) point.

$$y = b_2x^2 + b_1x + b_0 \tag{3.3}$$

the endogenous concentration was calculated as[2]

$$x_e = \left| \frac{-b_1 + \sqrt{{b_1}^2 - 4b_2b_0}}{2b_2} \right| \tag{3.4}$$

The corrected concentration (x_c), i.e., the de facto concentration to be used for standards and QCs, was calculated by adding the endogenous concentration (x_e) to the spiked concentration (x_s):

$$x_c = x_s + x_e \tag{3.5}$$

To implement these corrections, and remove the bias from the calibration, the tables of calibration standards' and QCs' expected concentrations (or "actual concentration" in MultiQuant software) were reassigned using x_c instead of x_s. This effectively "shifts" the calibration curve to the right and achieves the desired correction; accurate concentrations will be calculated for unknown samples.

This correction process has been automated by a script written and run in RStudio (RStudio, Boston, MA, USA). R (programming environment, https://www.r-project.org/) and RStudio (graphical interface, https://www.rstudio.com/) are free open-source statistical software tools. This automated tool requires that the user copy to the clipboard (Ctrl+C) a table with the concentrations (x_s) in the first column and the signals (e.g., area ratio) in the second column and run ("source") the script in RStudio. Once the script has run, the corrected concentrations (x_c) are automatically stored in the clipboard, ready to be pasted into the data analysis software of the user. The full correction script and instructions for use (including a link to a user video tutorial) are available in Appendix D.1.

[2]The quadratic equation has two potential solutions as a result of the addition or subtraction of the square root term. However, the squared linear term (${b_1}^2$) is generally much larger than the $-4b_2b_0$ term since the non-linearity (the quadratic term (b_2)) is generally very small in analytical calibration. Therefore, the solution shown in Equation 3.3 is the only logical solution for the type of systems studied.

3.3.4 Calculation of the additional error introduced by the endogenous concentration correction

The purpose of this endogenous concentration correction is to remove the systematic bias (or deterministic error) from the analysis. However, it is equally important to appreciate that the random error associated with determining the endogenous concentration needs to be considered when assessing the error associated with the samples. The exact magnitude of the error folded in by the correction operation can be computed using the formulas for the standard error of the prediction [85].

Since the spiked concentration[3] (x_s) and the endogenous concentration (x_e) are both calculated from the same data set, their errors are dependent on one another. In this situation, the summation of errors should be used for the corrected concentration (x_c) [151]:

$$s_{x_c} = s_{x_s} + s_{x_e} \tag{3.6}$$

rather than the standard error propagation formula $\sqrt{s_a^2 + s_b^2}$, which is reserved for addition of independent errors [151].

Calculation of the components s_{x_s} and s_{x_e} will differ depending on whether the data studied is homoscedastic or heteroscedastic.

For homoscedastic data, those items are calculated as [85]

$$s_{x_e} = \frac{s_{err}}{b_1}\sqrt{\frac{1}{n} + \frac{\overline{y}^2}{{b_1}^2 \sum (x_i - \overline{x})^2}} \tag{3.7}$$

$$s_{x_s} = \frac{s_{err}}{b_1}\sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(\overline{y_s} - \overline{y})^2}{{b_1}^2 \sum (x_i - \overline{x})^2}} \tag{3.8}$$

with

[3] Here, the quality control samples are used to model how much error would be added by the process. The spiked QC concentration is set by experimental design and known with high precision, but in effect, we are ignoring this knowledge to examine the efficiency of our analytical method. Moreover, this equation is a conservative estimate of the error for the sum, and not the actual standard error for the sum.

$$s_{err} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}} \tag{3.9}$$

Where

s_{x_e} is the error on the estimated endogenous analyte concentration;

s_{x_s} is the error on the estimated uncorrected analyte concentration;

s_{err} is the square root of the residual variance, representing the spread of the measurements around the fitted regression line (variance of y given x, often called the error of the regression S_r);

b_1 is the estimated slope of the linear regression;

n is the number of calibration levels (calibration standards);

m is the number of measurement replicates of the analyzed sample;

$\overline{y}$ is the average measurement for all calibration standards;

$\overline{y_s}$ is the average measurement for the analyzed sample;

y_i is the measurement for the calibration standard at concentration level x_i ;

$\hat{y}_i$ is the predicted measurement at concentration level x_i (obtained by inserting the value of x_i in the calibration equation);

x_i is the concentration at level i;

$\overline{x}$ is the average concentration for all calibration standards (including blanks).

For heteroscedastic data, the components s_{x_s} and s_{x_e} are calculated as

$$s_{x_e} = \frac{s_{err}}{b_1}\sqrt{\frac{1}{\sum w_i} + \frac{\overline{y}_w^2}{{b_1}^2\ (\sum w_i {x_i}^2 - \sum w_i \overline{x}_w^2)}} \tag{3.10}$$

$$s_{x_s} = \frac{s_{err}}{b_1}\sqrt{\frac{1}{w_s m} + \frac{1}{\sum w_i} + \frac{(\overline{y}_s - \overline{y}_w)^2 \sum w_i}{b_1^2 \left(\sum w_i \ \sum w_i x_i^2 - (\sum w_i x_i)^2\right)}} \tag{3.11}$$

with

$$s_{err} = \sqrt{\frac{\sum w_i \left(y_i - \hat{y}_i\right)^2}{n-2}} \tag{3.12}$$

$$\overline{y}_w = \frac{\sum w_i y_i}{\sum w_i} \tag{3.13}$$

$$\overline{x}_w = \frac{\sum w_i x_i}{\sum w_i} \tag{3.14}$$

Where
w_i is the weighting factor applied to concentration level i;
w_s is the weighting factor applied to the sample measurement (concentration level s);
$\overline{x}_w$ is the weighted mean concentration;
$\overline{y}_w$ is the weighted mean measurement;
s_{x_e}, s_{x_s}, s_e, b_1, n, m, $\overline{y}$, $\overline{y_s}$, y_i, $\hat{y}_i$ and x_i are as described above.

Once again, these calculations have been automated via an RStudio script. The user is required to build a table with a "Name", "Type", "Spiked_Conc" and "Measure" headers that store the type of sample ("QC" or "Cal"), the spiked concentration and the measure (e.g., area or area/response ratio). This table is copied to the clipboard (Ctrl+C), the script is run, and the estimated error for each QC level is displayed in the console. The full error evaluation script and instructions for use (including a link to a user video guide) are available in Appendix D.2.

3.4 Results and Discussion

3.4.1 Correction of the endogenous concentration

With "well-behaved" analytes in a matrix devoid of endogenous content, calibration curves should produce a y-intercept (b_0) approximately equal to zero (within the limit of the experimental error). On the other hand, as shown in Figure 3.4.1a, a calibration curve for an analyte with a detectable endogenous concentration will show an upward shift. This shift is generated by the constant signal produced, throughout the entire calibration range, by the endogenous concentration. The observed signal is then just the sum of the endogenous signal and the one generated by the analyte spiked into the matrix. Calibration curves with endogenous analytes will therefore present a y-intercept (b_0) significantly different from zero. In order to correct for the systematic bias introduced by endogenous content in the matrix, its contribution can be either ignored or corrected for.

Conceptually, the simplest route is to set the b_0 parameter in the calibration equation to zero, i.e., to ignore b_0 entirely (note that this is not the same thing as forcing the intercept

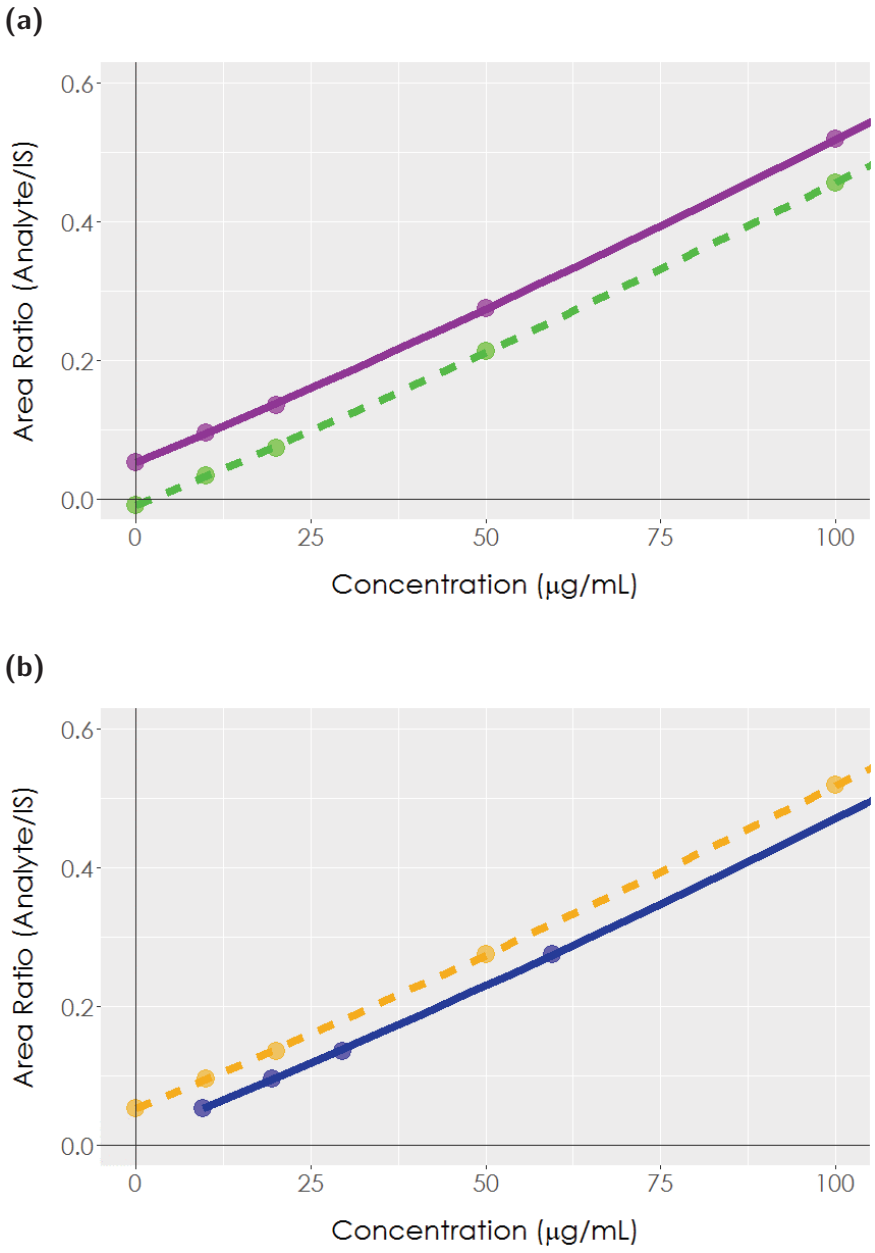


Figure 3.4.1: Calibration curves in different settings. **(a)** Dashed line: calibration curve for a regular analyte, with no endogenous contribution: the y intercept $b_0 \approx 0$. Solid line: calibration curve for an analyte with an endogenous concentration: the y intercept $b_0 > 0$. **(b)** Dashed line: uncorrected calibration curve for an endogenous analyte. Solid line: calibration curve for the endogenous analyte following the correction process using the de facto concentration in the matrix.

through zero, which would modify the b_1 parameter). Another equivalent way to achieve this correction would be to subtract b_0, or the average signal obtained for the 0 $\mu g/mL$ standard(s), from all calibration standards and QCs. These solutions are all adequate, and some software (e.g., Sciex OS, Mass Hunter's Blank Subtraction Add-In) offer tools to carry one of them out. However, the inflexibility of several data analysis software in routine use might render general application of these solutions difficult. If this procedure is to be applied in high throughput production settings, the best solution will be the one which operates seamlessly within the confines of the data analysis software.

The most practical procedure to meet those requirements involves modifying the expected ("actual") concentration provided to the data analysis software. Since the expected concentration is one of the few adjustable parameters in the analysis tables, this must be used as the tool to effect the correction. As stated in Equation 3.5, the corrected concentration for standards and QCs is one reflecting the real, or de facto, analyte concentration (x_c) in the sample: what was spiked (x_s) plus the amount which was already present to begin with, the endogenous concentration (x_e). Using the real, corrected concentrations to describe calibration standards and QCs will shift the calibration curve (Figure 3.4.1b), achieving the desired correction. Once the de facto concentration of analyte is used to build the calibration curve, it can perform bias-free quantification of any casework and QC samples.

The correction process can be used in conjunction with either linear or quadratic calibration. To demonstrate this, 10 000 calibration data sets (heteroscedastic, $1/x^2$) were generated using the range of b_0, b_1 and b_2 values observed over 14 LC-MS/MS experimental calibration curves with various known levels of "endogenous" analyte. The endogenous concentration was back-calculated using a quadratic, $1/x^2$ regression or using a linear, $1/x^2$ regression from which the top two standards (255 and 300 $\mu g/mL$) were removed to satisfy linear calibration requirements [1, 2]. The bias in the calculated endogenous concentration (vs. the known endogenous concentration) was calculated for every data set in both the linear and quadratic regression cases. The histograms of the resulting differences show that the distribution is more symmetrical, with a median closer to zero when the quadratic regression is used rather than the linear one. The complete results, including the R script used to perform this *in silico* evaluation, are available in Appendix D.3. Based on these results, the authors recommend using the calibration model selected and validated during the validation process to perform the correction for the endogenous concentration, whether it is quadratic or linear, unweighted, $1/x$ or $1/x^2$ weighted.

In order to achieve seamless integration into production work, an R script automating this correction process has been developed and is available in Appendix D.1, along with a user video tutorial. This script is easily configured to fit the needs of the user. Its use in a production setting adds four mouse clicks to the entire analysis process, taking less than 30 seconds to perform.

3.4.2 BHB endogenous concentrations estimated in a production setting

In our production operations, calibration standards and QCs have all been spiked in authentic human blood matrix and corrected for the presence of endogenous content using the R script available in Appendix D.1. Since these methods have been put in production, 17 GC-MS and 83 LC-MS/MS batches have been analyzed, using 2 and 10 different blood lots, respectively. In the GC-MS batches, the estimated endogenous BHB concentrations ranged from 1.1 to 13 $\mu g/mL$, with a median of 7.5 $\mu g/mL$. In the LC-MS/MS batches, the estimated BHB concentrations ranged from 0.080 to 30 $\mu g/mL$, with a median of 4.1 $\mu g/mL$. A box plot representation of the results per matrix lot is shown in Figure 3.4.2. This shows that results are generally coherent within a single matrix lot, despite the occasional outlier. These remote results might even create distortions in the boxplot; for example, lot B appears different in the GC-MS and LC-MS/MS analyses despite similar compositions and medians of 1.5 and 1.18 $\mu g/mL$ respectively. Extreme values could be attributable to analyte instability or neoformation during storage [152]. Indeed, the same matrix lot has been used for analysis for up to 169 days, potentially long enough to observe degradation. Within the scope of this research, we were unable to ascertain the exact mechanism generating these extreme values. An analysis of five separate GC-MS calibrations, prepared in the same matrix lot, extracted and analyzed in a close timeframe gave coherent endogenous concentrations (x_e). Estimated endogenous concentrations ($x_e \pm s_{x_e}$) were 11.32 ± 0.14, 11.87 ± 0.18, 11.64 ± 0.17, 11.76 ± 0.16 and 11.89 ± 0.18 $\mu g/mL$, which supports the conclusion that results within a same matrix lot are consistent.

These results show that, in the case of BHB, endogenous concentrations present in the spiking matrix can be significant. Indeed, at the maximum endogenous concentration measured (30 $\mu g/mL$), the endogenous analyte content contributes 15% of the decision point value (200 $\mu g/mL$) to quantifications performed. Applying this correction thus becomes important to uphold accuracy of the results. But regardless of the exact endogenous concentration, performing the correction for the endogenous content will yield a more accurate

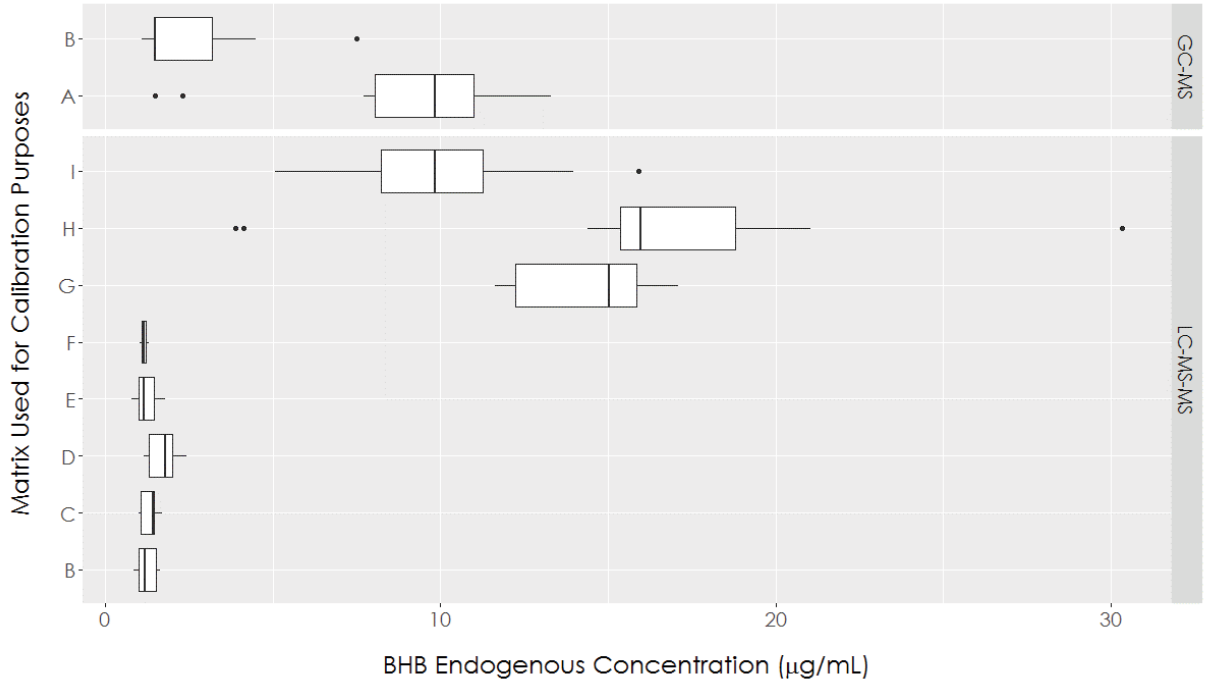


Figure 3.4.2: Distribution of the estimated BHB concentrations for each matrix lot. The vertical line within the box is the median value for each lot, while the ends of the box show the upper and lower quartiles. The end point of the lines (whiskers) are the highest and lowest values, excluding the outliers (black dots). This graph is based on all batches analyzed with the different matrix lots (lot A, $n = 10$; lot B GC-MS, $n = 8$; lot B LC-MS/MS, $n = 10$; lot C, $n = 7$; lot D, $n = 8$; lot E, $n = 7$; lot F, $n = 6$; lot G, $n = 11$; lot H, $n = 17$; lot I, $n = 12$).

quantification, at very little cost to productivity, since the R script developed is run in less than 30 seconds.

3.4.3 Additional error folded in by the endogenous concentration correction

Applying the endogenous concentration correction process will remove the systematic error from the analysis but will necessarily fold in additional indeterminant error. The error on the estimated endogenous concentration (s_{x_e}) must be combined with the error from the estimated uncorrected analyte concentration in the sample (s_{x_s}) to yield the total error for the corrected concentration, s_{x_c}. The magnitude of the error folded in will depend on a multitude of factors, as demonstrated by Equations 3.7 to 3.14. The weighting factor, number of replicate measurements, position of the calibration standards on the dynamic range and precision will all play a role in determining the total error and the contribution from the estimated endogenous concentration (s_{x_e}). The multifactorial nature of error magnitude estimation renders instinctual estimation very difficult. In order to help analysts estimate the amount of error involved in the correction procedure, a second automated R script has been written and is made available to the reader in Appendix D.2.

This script has been applied to all 17 GC-MS and 83 LC–MS-MS batches analyzed. For the GC-MS batches, the error added (s_{x_e}) ranges from 0.014 to 0.71 $\mu g/mL$ (median 0.15 $\mu g/mL$) compared to 6.54 $\mu g/mL$ as the median error of the uncorrected concentration (s_{x_s}) for the mid-level QC. For LC–MS-MS batches, the error added (s_{x_e}) ranges from 0.0078 to 2.7 $\mu g/mL$ (median 0.095 $\mu g/mL$) compared to 7.28 $\mu g/mL$ as the median error of the uncorrected concentration (s_{x_s}). Interestingly, although the absolute error increases after concentration (i.e., $s_{x_c} > s_{x_s}$), the relative (%) error is reduced by applying the endogenous concentration correction (i.e., ${}^{s_{x_c}}/_{x_c} < {}^{s_{x_s}}/_{x_s}$) in an overwhelming proportion of the cases (95%).

At first glance, this was a surprising result, so we decided to confirm its accuracy by running a simulation of the error generated in the calibration and endogenous correction processes. The GC-MS method parameters (calibration standards and QC levels, experimentally observed calibration curve parameters b_0 and b_1) were used for this procedure. In a simulation such as this one, we know the real value of all parameters (x_e, b_0, b_1, $\hat{y}_i$, etc.) and generate "measured" y_i values according to a normal distribution and SWGTOX recommendations (maximum tolerable QC precision of 20%). The complete R script is available

(a)

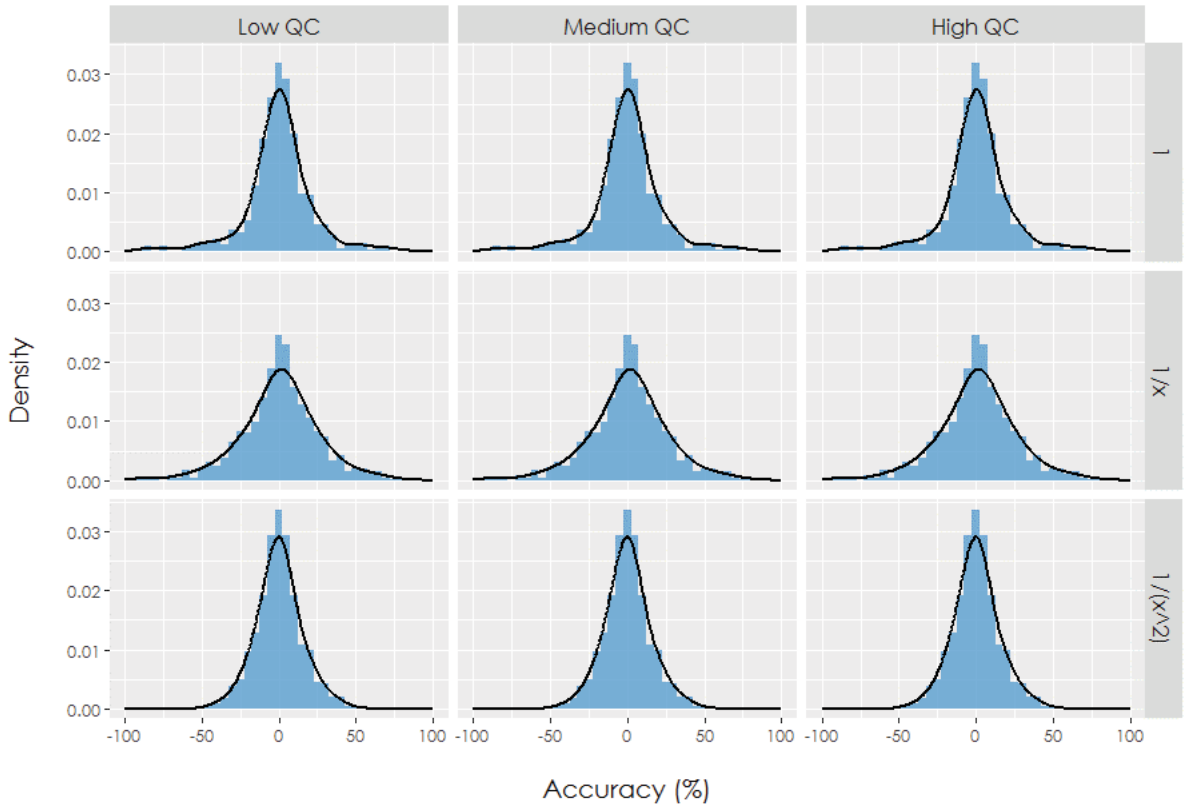


Figure 3.4.3: *In silico* evaluation of the endogenous concentration correction process and estimation of the error. **(a)** Relative error of the endogenous concentration, calculated as (Calculated Endogenous Concentration – Known Endogenous Concentration) / Known Endogenous Concentration $\times$ 100. 1000 simulations per QC level/weight combination have been performed (total 9000 simulations). 1.1% of data points (outliers) are omitted from this graphical representation for clarity purposes (these mostly have $>$ 100% relative error). Black lines are density function estimations.

in Appendix D.4. By comparing known and estimated x_e values, this *in silico* evaluation of the endogenous correction process allowed us to confirm that the procedure performs a generally accurate estimation of the endogenous concentration. The histograms of the percentage difference between the calculated and real endogenous concentration, accompanied by the density function, for the different QC levels and weighting schemes, are shown in Figure 3.4.3a.

As for the error observed throughout the procedure, the results observed in the *in silico*

(b)

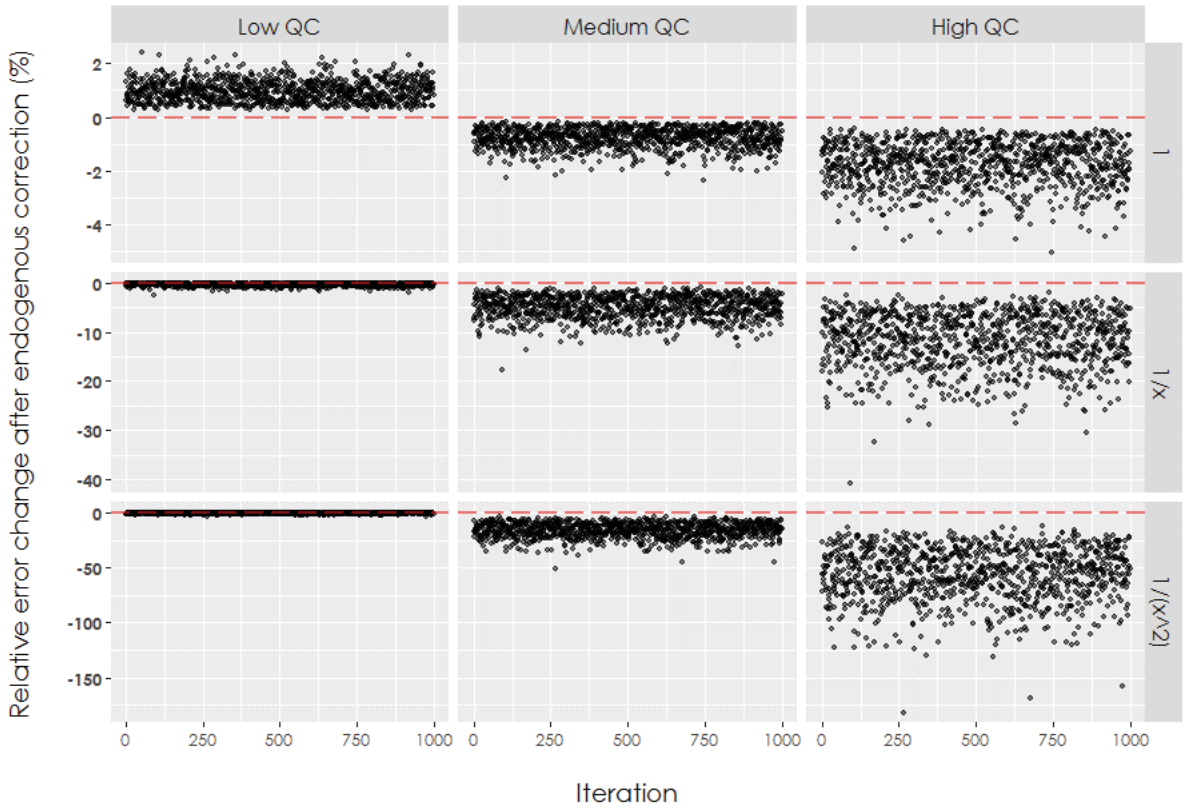


Figure 3.4.3: *In silico* evaluation of the endogenous concentration correction process and estimation of the error. **(b)** Modeling of the change in the relative error following the endogenous concentration correction process, shown for all QC level/weight combinations. The relative error change is calculated as the relative error on the corrected concentration ($^{s_{x_s}}/_{x_s}$ %) minus the relative error on the uncorrected concentration ($^{s_{x_c}}/_{x_c}$ %). 1000 simulations per QC level/weight combination have been performed (total 9000 simulations). Note the different magnitude of the y axis depending on the weighting scheme studied.

simulations match the experimental observations. The absolute error is systematically increased by an amount equal to the estimation error on the endogenous concentration (s_{x_e}). However, the relative (%) error is almost always reduced by applying the endogenous concentration correction, as shown in Figure 3.4.3b. The only situation where this does not happen is with homoscedastic (unweighted) calibrations and lower concentration samples. Not only is the relative error reduced in all other situations, but the higher the concentration of the sample, the greater the drop in relative error. Mathematically, this happens because s_{x_c}/x_c is smaller than s_{x_s}/x_s. Heteroscedastic data ($1/x^2$ model in particular) will also experience a higher drop in relative error following the correction procedure. In these cases, the correction produces a more accurate concentration with only a marginal increase in absolute error due to the heteroscedasity. Taken together, this improves the relative error. This explains the experimental observations, since the data for both the GC-MS and LC-MS/MS method are heteroscedastic with variance increasing proportionally to the square of the concentration ($1/x^2$ model). It is therefore expected to observe mostly a decrease in the relative error following the correction procedure.

3.5 Conclusions

We developed a tool to automatically correct for the presence of endogenous analyte content in matrices used for calibration and QC samples. Instead of using different strategies to more or less sidestep the issue - for example using only standard addition to quantify endogenous analytes, using an analog matrix or analyte, ignoring entirely the endogenous concentration and the bias created in the quantification process - this method can fully take into account the endogenous contribution and correct for it.

This procedure adjusts the expected analyte concentration to match the real, or de facto, concentration present in the calibration standard, i.e., the amount of analyte added plus the endogenous amount. The endogenous concentration is estimated via the calibration standards by treating them as a standard-addition experiment/calibration. Although there are several different ways to perform such a correction mathematically, adjusting the expected concentration allows seamless integration of the correction procedure with production operations performed in commercial data analysis software such as Mass Hunter or MultiQuant.

We have described in detail how the correction procedure can be applied in a production setting to correct for the endogenous concentration present in the spiking matrix. In val-

idation procedures, addressing the endogenous concentration problem head-on will require somewhat more work. The endogenous concentration will need to be determined for every matrix used in the validation procedures, and results corrected accordingly. The endogenous concentration for each matrix can be determined with the tool developed here (Appendix D.1). Ultimately, this supplementary work at validation time, combined with the endogenous concentration correction during the production operations, will yield an unbiased and more reliable quantification of the analyte for the long term.

The necessity to perform such a correction will vary based on the targeted application and the endogenous concentration of the analyte. However, at any detectable endogenous presence, correcting for the presence of analyte in the spiking matrix will generate a more accurate quantification.

We have applied this correction procedure in a production setting to two methods used for BHB quantification; one GC-MS and one LC-MS/MS method. Endogenous BHB concentrations ranging from 0 to 30 $\mu g/mL$ have been found in whole human blood matrices. Correcting for this content prevents a bias of up to 15% of the BHB decision point for hyperglycemia (200 $\mu g/mL$). It can therefore be important to correct for endogenous concentrations, which can be significant.

This procedure does increase the estimation error on the concentration for quantified unknowns. The exact magnitude of the error added by this process very much depends on specific method parameters, such as the number and distribution of calibration standards, the number of replicate measurement, homo or heteroscedasticity of the data, etc. We have developed an R script (Appendix D.2) which allows users to get an estimation of the error added with their specific method. For the BHB quantification methods presented in this paper, the additional error folded in by the correction procedure amounts to a few percent ($<3\%$) at most. The small error added by the procedure, especially when compared to the potential for removing systematic bias makes this an attractive solution to deal with endogenous analytes in validation and production settings.

3.6 Acknowledgements

The authors wish to thank Cynthia Côté for her questions and ideas.

3.7 Mathematical tools for quantitative method validation

This chapter presented a mathematical tool to correct for endogenous bias in matrix-matched calibration. BHB was used as a case study to verify this tool's performance and efficiency in a production setting. In the production environment, correcting for the endogenous amount of BHB present in sample used for the calibration standard(s) improves the accuracy for BHB in all other matrices (unknown cases). At the validation stage, each matrix used for validation purposes must also undergo the correction process before accuracy, precision, recovery, matrix effects and other standard parameters are evaluated. Other endogenous substances could benefit from a similar treatment in correcting endogenous presence bias in toxicology cases, such as cyanide [153], steroids [92] and of course, CYP enzymes in liver tissue.

The mathematical tools developed in Chapter 2 and Chapter 3 will increase the accuracy of the quantitative CYP analysis. But the projected method relies not only on quantification, but also on characterization of mutated CYP enzymes, that is to say a qualitative method.

Although a lot less attention is generally paid to qualitative method validation procedures, they are as important and have some subtleties that are rarely dealt with in the literature. To develop a reliable CYP analysis method compliant with ISO 17025:2017 norms [72], some work will be required about decision point qualitative methods. This is covered in Chapter 4.

4

Qualitative Method Validation and Uncertainty Evaluation Via the Binary Output

This chapter is the integral text from:

Qualitative Method Validation and Uncertainty Evaluation Via the Binary Output

I —Validation Guidelines and Theoretical Foundations

Félix Camirand Lemyre[1], Brigitte Desharnais, Julie Laquerre, Marc-André Morel, Cynthia Côté, Pascal Mireault and Cameron D. Skinner

Accepted for publication to *Journal of Analytical Toxicology* (2019) submission number JAT-19-2881 [4].

and

[1]The first two authors contributed equally to the manuscript and are listed in alphabetical order.

Qualitative Method Validation and Uncertainty Estimation Via the Binary Output

II —Application to a Multi-Analyte LC-MS/MS Method for Oral Fluid

Brigitte Desharnais[2], Marie-Jo Lajoie, Julie Laquerre, Pascal Mireault and Cameron D. Skinner

Accepted for publication to *Journal of Analytical Toxicology* (2019) submission number JAT-19-2882 [5].

4.1 Validation guidelines and theoretical foundations

4.1.1 Abstract

Qualitative methods have an important place in forensic toxicology, filling central needs in, amongst others, screening and analyses linked to *per se* legislation. Nevertheless, bioanalytical method validation guidelines either do not discuss this type of method, or describe method validation procedures ill adapted to qualitative methods. The output of qualitative methods are typically categorical, binary results such as "presence"/"absence" or "above cut-off"/"below cut-off". Since the goal of any method validation is to demonstrate fitness for use under production conditions, guidelines should evaluate performance by relying on the discrete results, instead of the continuous measurements obtained (e.g., peak height, area ratio).

We have developed a tentative validation guideline for decision point qualitative methods by modeling measurements and derived binary results behaviour, based on the literature and experimental results. This preliminary guideline was applied to an LC-MS/MS method for 40 analytes, each with a defined cut-off concentration. The standard deviation of measurements at cut-off (s) was estimated based on 10 spiked samples. Analytes were binned according to their %RSD (8.00%, 16.5%, 25.0%). Validation parameters calculated from the analysis of 30 samples spiked at $-3s$ and $+3s$ (false negative rate, false positive rate, selectivity rate, sensitivity rate and reliability rate) showed a surprisingly high failure rate. Overall, 13 out of the 40 analytes were not considered validated. Subsequent examination found that this was attributable to an appreciable shift in the standard deviation of the area ratio between different batches of samples analyzed. Keeping this behaviour in mind when setting the validation concentrations, the developed guideline can be used to validate qualitative decision point methods, relying on binary results for performance evaluation and

[2]The first three authors contributed equally to the manuscript and are listed in alphabetical order.

taking into account measurement uncertainty.

An application of this method validation scheme is presented in Section 4.3.

4.1.2 Introduction

Qualitative methods are best described by contrasting them with quantitative methods. The output of qualitative methods is categorical (or discrete) in nature, typically binary: presence or absence (qualitative identification methods), above or below threshold (qualitative decision point methods). On the other hand, quantitative methods produce concentration estimates on a continuous scale.

There is an abundant literature dealing specifically with quantitative methods and their validation procedures [75, 79, 80, 118, 121, 128, 139, 154, 155]. However, the literature and guidelines dealing with qualitative methods is much sparser.

SWGTOX [75] and AAFS Standards Board [81] both contain recommendations for decision point qualitative methods. According to these guidelines, LC-MS/MS qualitative decision point method validation should include interference and carryover studies, dilution integrity and stability if necessary, as well as precision evaluation. Precision of the measured signal should be evaluated at $\geq 50\%$ of the decision point (DP) or cut-off concentration, at the DP concentration and at $\leq 150\%$ of the DP concentration. The method is considered to be validated if $\%\text{RSD} \leq 20\%$ and $(\bar{x}_{50\%} + 2s_{50\%}) < \bar{x}_{DP} < (\bar{x}_{150\%} - 2s_{150\%})$, i.e., the mean $\pm$ two standard deviations at 50% and 150% of the decision point do not overlap with the mean measurement at cut-off.

Two potential weak points can be identified. First, these procedures fail to use the categorical or binary nature of qualitative methods' output, employing instead procedures derived from quantitative method validation which relies on continuous data. One reason likely explaining this state of affairs is the confusion induced by the fact that continuous measurements (area, height, area ratio, luminescence, etc.) are transformed into binary results. Moreover, quantitative method validation guidelines are so well developed that they are almost second nature to forensic toxicologists and bioanalysts. It therefore feels natural and safe to fall back on them for the related but distinct problem of qualitative method validation.

A second weak point is the absence of a clear framework for evaluation of the method's

uncertainty of measurement (UM). The requirement for UM evaluation in qualitative methods has been recently introduced in the ISO 17025:2017 [72] standard, therefore its absence from published guidelines is not surprising. Nonetheless, given this new requirement, adequate UM evaluation procedures for qualitative methods are required.

In any method validation, the goal is to demonstrate the quality of the analytical method by producing objective proof that predefined performance criteria are met [75, 81]. Importantly, this verification of the fitness for use has to occur under the same preparation, analysis and data processing procedures which will be used for analysis (production) [75, 81]. The same holds true for qualitative method validation. Accordingly, binary results (presence/absence, above/below cut-off) yielded by the method should be used to measure the adequacy of its performance, since this is the result ultimately produced in a production setting.

If the binary output of qualitative methods is to be used, what are the appropriate validation guidelines and the associated minimal performance thresholds, and how should UM be evaluated and taken into account in the final results?

In order to answer such questions, the behaviour of the response variable (area ratio, luminescence, etc.) in relation to the encoded, binary outcome must be understood. This subject is touched upon sparingly in the literature [105–107] where diverse validation procedures are suggested.

In this paper, we draw upon these various sources, computer simulations and experimental data to study the behaviour of the binary above/below threshold output of qualitative decision point methods, in order to put forward a tentative validation guideline. This guideline is heavily based on the performance evaluation of another kind of categorical test: medical tests for the presence of a diseased state [156, 157]. This validation process is then evaluated by using an LC-MS/MS method for 40 analytes in blood. Results guided modifications to the guidelines. The final version of the qualitative decision point method validation guidelines is applied to an LC-MS/MS method for 97 analytes in oral fluid in Section 4.3.

4.1.3 Materials and methods

Analytical method

The experimental data used for prospective and confirmatory studies of qualitative decision point methods were derived from a high throughput whole blood LC-MS/MS analysis method for 40 qualitative analytes and 60 quantitative analytes. The quantitative analytes were validated separately [158] and will not be discussed in this paper. For every qualitative analyte, a cut-off concentration was selected based on analytical (sensitivity across multiple LC-MS/MS systems) and toxicological (relevant concentrations for effects) considerations. The full list of substances and their designed cut-off is available in Appendix E.1.

Sample preparation

Samples were brought to room temperature over 1 hour. Following vortex mixing for 10 seconds, 100 μL of blood was transferred using a positive displacement pipette into a 96 well-plate with 2 mL square wells (Fisher Scientific, AB-0932, Ottawa, Ontario, Canada).

For the purpose of this study, blood samples were spiked at the cut-off concentration or its multiples (e.g., 50%, 150%, 200%, etc.). Postmortem cardiac and femoral blood with negative screening results, as well as antemortem blood purchased from UTAK (Valencia, California, USA) were used. All compounds used for spiking purposes were purchased from Cerilliant (Round Rock, Texas, USA), except for 3-hydroxy bromazepam and N-desmethyl diphenydramine which were purchased from Toronto Research Chemicals (North York, Ontario, Canada). 10 μL of stable isotope-labeled internal standards solution (IS, Cerilliant, Round Rock, Texas, USA), at concentrations indicated in Appendix E.1, were added to the blood sample and mixed using vortexing.

In order to obtain a more finely granular precipitate, 100 μL of methanol:0.2% formic acid in water (50:50 v:v) solution was mixed into the blood sample. Then 400 μL of acetone:acetonitrile (30:70 v:v) mixture was used to precipitate the proteins. Following mixing, the plate was centrifuged at 3200 $\times$ g for 5 minutes. A 25 μL aliquot of the supernatant was then transferred to a second 96 well-plate with 1 mL round bottom wells (Canadian Life Science, RT96PPRWU1mL, Peterborough, Ontario, Canada). This extract was diluted with 180 μL 0.2% formic acid in water and vortexed.

LC-MS/MS analysis

A 5 μL aliquot of diluted extract was separated on an Agilent Zorbax Eclipse Plus C18 column (2.1 $\times$ 100 mm, 3.5 μm) using a step/ramp gradient starting from 2:98 methanol:10 mM ammonium formate (pH 3.0) to 50% acetonitrile. The flow from the HPLC (Agilent 1200 or 1260 Infinity) was directed to a Sciex 5500 QTrap triple quadrupole mass spectrometer. Detailed analytical parameters with regards to the liquid chromatography and mass spectrometry acquisition are available in Appendix E.1.

Preliminary validation guidelines

Based on the picture of the behaviour of binary results in qualitative decision point methods described in the literature and our exploratory experimental data analysis (described in Section 4.1.4), a preliminary set of validation guidelines was determined and applied.

The standard deviation was estimated by analyzing a minimum of 10 different samples all spiked at the cut-off and calculating the standard deviation of the response variable used, in this case the ratio of analyte peak area to IS peak area.

Probability curves (akin to the one shown in Figure 4.1.1b) plotting the positivity rate (or above cut-off rate) as a function of concentration, were built by spiking 10 or more samples at regular concentration intervals from -4 to +4 times the estimated sample standard deviation (s), e.g., $-4s$, $-3s$, $-2s$, $-1s$, cut-off, $+1s$, $+2s$, $+3s$ and $+4s$. This facultative step assumed a linear response and a blank response of zero.

The core of the validation procedure consisted of analyzing 30 samples spiked at $-3s$ and $+3s$ (upper (UURL) and lower (LURL) unreliability limits, see explanation in Section 4.1.4), which were used to calculate the method's validation parameters. Care was taken to ensure that these samples were prepared, injected and analyzed as they would be in a production setting, including for this particular method two replicates of a sample spiked at cut-off, used to establish the threshold measurement and permit classification of samples as being above or below cut-off.

Ion ratio dependability for identification purposes was estimated as the percentage of all samples for which the ion ratio fell within $\pm$30% of the ion ratio measured in reference sample(s).

Carryover and interference studies should be carried out, as well as stability evaluation if deemed necessary. Although several procedures and guidelines exist, SWGTOX's practices are recommended in forensic toxicology [75]. If applicable, dilution integrity can be verified by repeating the main validation procedures (standard deviation estimation, evaluation of performance parameters on diluted samples). These studies were carried out for the method presented here, but will not be discussed since they are not the main focus of this paper.

Calculation of validation parameters

The validation parameters (false negative rate (FNR), false positive rate (FPR), reliability rate (RLR), selectivity rate (SLR) and sensitivity rate (SNR)) are calculated as follows [105, 106, 156]:

$$FNR = \frac{FN}{FN + TP} \times 100 \tag{4.1}$$

$$FPR = \frac{FP}{FP + TN} \times 100 \tag{4.2}$$

$$RLR = \frac{TP + TN}{n} \times 100 = 100 - FPR - FNR \tag{4.3}$$

$$SLR = \frac{TN}{TN + FP} \times 100 \tag{4.4}$$

$$SNR = \frac{TP}{TP + FN} \times 100 \tag{4.5}$$

Where:

TN is the number of true negative results;

TP is the number of true positive results;

FN is the number of false negative results;

FP is the number of false positive results;

n is the total number of results.

The reliability (RLR) represents the overall method's ability to correctly identify the samples as above or below cut-off; the sensitivity (SNR) evaluates the percentage of samples actually above cut-off that are indeed identified as such, and the selectivity (SLR) measures the percentage of samples actually below cut-off that are indeed identified as such.

A qualitative decision point method validated under these production conditions (2 measured cut-off samples, rates estimated over 30 samples) can be considered fit for purpose if the observed $FNR \leq 7\%$, $FPR = 0\%$, $RLR \geq 93\%$, $SLR = 100\%$, $SNR \geq 93\%$ (Figure 4.1.1d) and ion ratio, carry-over and interference studies are successful. These expected performance levels were calculated using the RStudio script presented in Appendix E.2. Expected performance levels under a different number of measured cut-offs and number of samples for rate estimation can be computed from the same R script: readers are encouraged to use it to define criteria under their own production and validation conditions.

Computer simulations

In preparation for the simulations, a set of 30 different samples spiked at cut-off were extracted and analyzed to determine their area ratios (analyte peak area/internal standard peak area). Application of the Cramer-von Mises normality test did not show significant departures from normality in all but two of the 40 analytes ($0.032 < P < 0.922$). The R script used to perform this analysis and the set of complete results are available in Appendix E.3. This is in accordance with the implicit statement consensus in the literature, namely that measurements (e.g., area, area ratios), including those made on an LC-MS/MS instrument, can be approximated by a normal distribution [105–107, 159].

Based on these results, response values for simulations were modeled using RStudio's normally distributed random number generator `rnorm(n, mean, sd)`, where `n` is the number of measurements to be generated, `mean` is the known true value of the measurement, and `sd` is the standard deviation. The number of measurements simulated per concentration level varied as needed between 1 and 100. The known true value of the measurement (area ratio) at cut-off was set to vary between 0.008 and 1.050, based on the experimentally observed area ratios at the cut-off concentration. A linear function describing the relationship between response and concentration was set as $y = b_1 x$, where b_1 was dictated by the known true

concentration of the cut-off selected. Unless otherwise stated, a standard deviation equivalent to 15% of the cut-off response value was applied. R scripts used to carry out these simulations are available in Appendix E.2.

4.1.4 Results and discussion

Theoretical behaviour of binary results

When thinking about decision point or threshold methods, the first reflex is often to assume (or hope) that results behave akin to what is displayed in Figure 4.1.1a. Instinct dictates that all samples with a concentration below the threshold, or cut-off, will produce a low response and therefore score negative every single time they are analyzed, and samples with a concentration higher than cut-off will similarly score positive (or above cut-off) systematically.

While this would be incredibly helpful, it is unfortunately impossible. A sample at a given concentration subjected to experimental manipulations and measured by a device that has some degree of imprecision will always produce a range of measured values, typically with a normal distribution. If this measurement error is ignored, important bias will ensue [160]. Thus, as can be seen in Figure 4.1.2a, repeated measurements on a sample with a concentration exactly equal to the cut-off will yield a normal distribution with an average response equal to the cut-off. 50% of these measurements will be reported as "below cut-off" and 50% as "above cut-off" (50% positivity rate).

If the sample analyzed has a concentration far enough away from the cut-off, e.g., if the mean measurement for that concentration is 3σ above or below from the cut-off value (Figure 4.1.2b), then almost all responses ($> 99.7\%$) will be reported as "above cut-off", or "below cut-off" respectively.

Logically, at a point between these two extremes, the normal distribution of responses will overlap to varying degrees with the threshold response, generating an intermediate positivity rate (Figure 4.1.2c). Samples with a concentration generating a mean response between the cut-off and $+3\sigma$ above the cut-off will yield positivity rates between 50.0% and 99.7%. The converse also applies to samples with responses below the cut-off with positivity rates

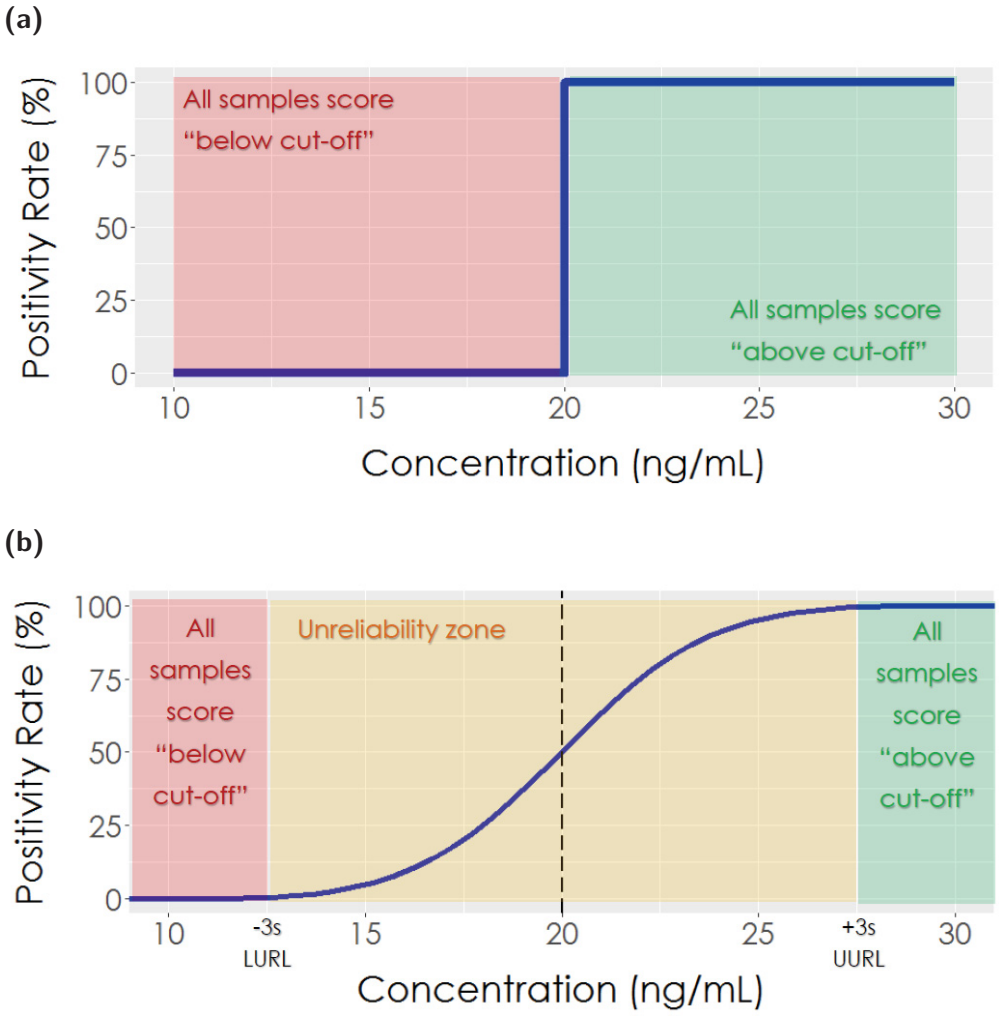


Figure 4.1.1: Positivity curves under different models. **(a)** Idealized behaviour of decision point qualitative methods. **(b)** Positivity curve for normally distributed measurements compared to a 20 ng/mL threshold.

decreasing as the concentration decreases.

The positivity curve for normal measurements compared to a threshold thus takes a sigmoidal form (Figure 4.1.1b). The uncertainty of measurement associated with qualitative methods is evident in this figure. Surrounding the cut-off is a range of concentrations where repeated measurement of the same sample will not always yield the same classification result (and the positivity rate takes an intermediate value). This unreliability (UR) zone, stemming from the uncertainty of measurement, is an ontological characteristic of qualitative decision point methods, and there is no possible way to avoid it. While some might reflexively believe

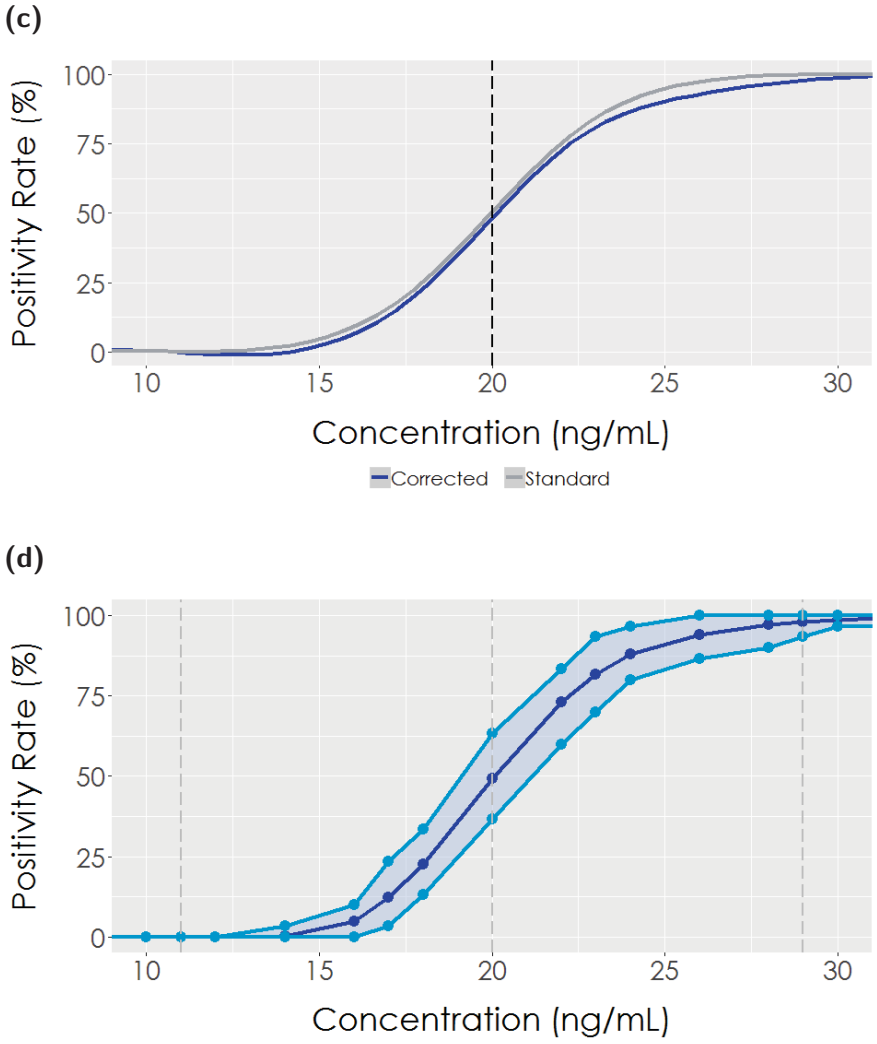


Figure 4.1.1: Positivity curves under different models. (c) Corrected positivity curve, accounting for heteroscedasticity and sampled threshold. (d) Average positivity rate when 30 spiked samples are measured and compared to a sampled cut-off (two measurements to establish threshold). 90% of positivity rate results fall within the shaded area (5% to 95% quantiles).

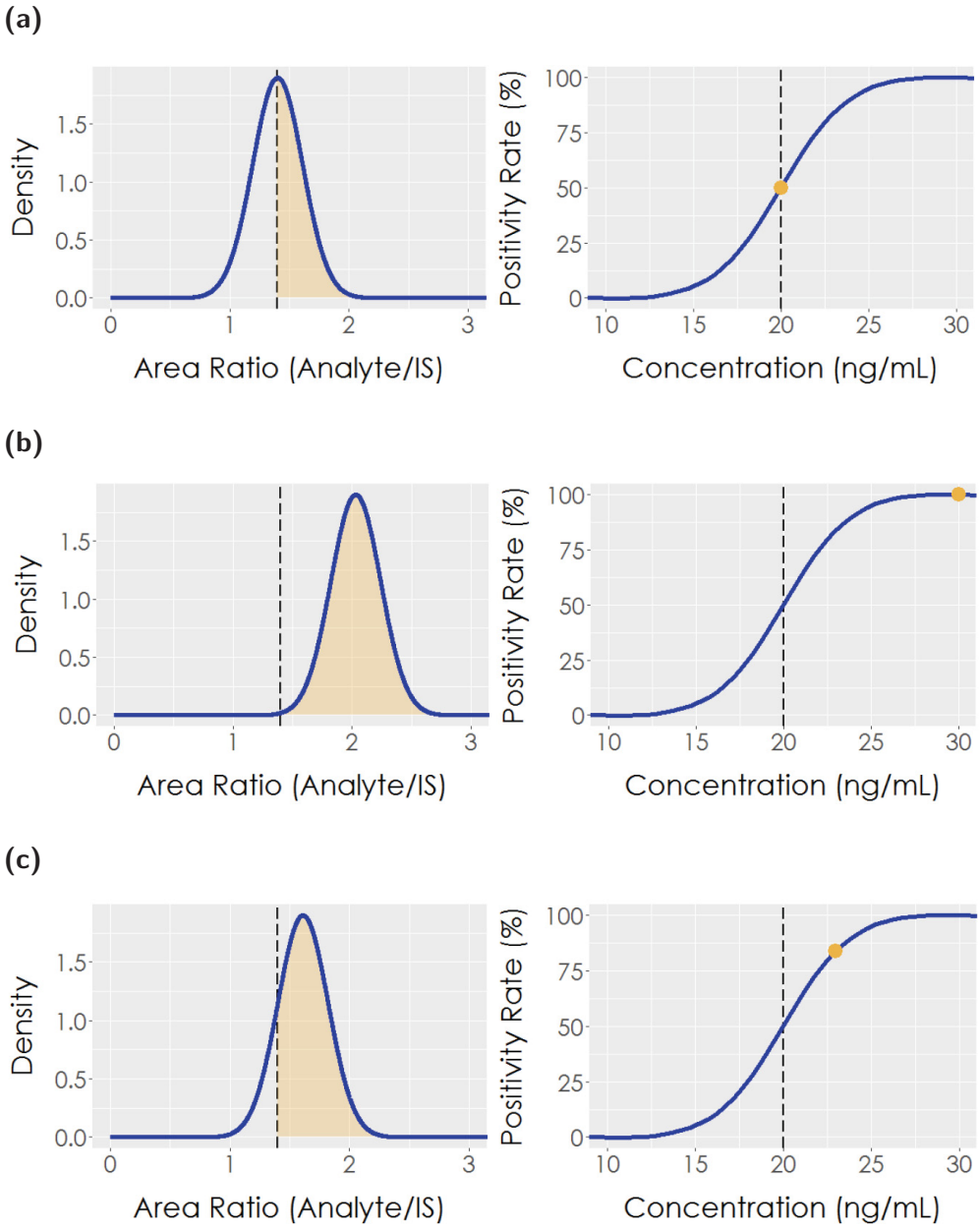


Figure 4.1.2: Normally distributed measurements in relation to a fixed threshold. **(a)** Sample spiked at the threshold concentration (20 ng/mL): distribution of measurements (density plot, left) and positivity curve (right). Exactly 50% of measurements are above the threshold measurement, resulting in a 50% positivity rate. **(b)** Sample spiked at $> 3\sigma$ above the cut-off concentration (30 ng/mL): distribution of measurements (density plot, left) and positivity curve (right). The whole distribution is far from the measurement at cut-off, yielding a 100% positivity rate. **(c)** Sample spiked at an intermediate concentration (between the cut-off concentration and 3σ above the cut-off concentration) (23 ng/mL): distribution of measurements (density plot, left) and positivity curve (right). The distribution of measurements overlaps with the threshold measurement, yielding an intermediate positivity rate (84%).

that moving the cut-off concentration could avoid this unreliability[3], it is important to understand that such a strategy is destined to fail since the unreliability zone would just follow right along with it. In much the same way that one cannot avoid measurement uncertainty in quantitative methods, the unreliability zone of qualitative decision point methods is here to stay and needs to be acknowledged, identified and estimated, not fought. The only viable strategy to minimize the magnitude of the unreliability zone is to minimize the standard deviation of the overall analytical process.

The positivity curve shown in Figure 4.1.1b is the one typically reported in the literature. But in order to adequately represent realistic (LC-MS/MS) data, it must be modified to take into account two important factors. First, real measurements are typically heteroscedastic, even over small concentration ranges. This is clearly demonstrated in Figure 4.1.3, which displays normal distribution curves based on 30 spiked samples replicates at different concentrations for buprenorphine. The variance increases with increasing concentrations, as made evident by the decreasing distribution maxima. This must be accounted for in validation guidelines.

Second, the standard positivity curve presumes that the response (measurement) at cut-off is a known and fixed value. But of course, this is not the case; in a production setting, this value is estimated based on a few measurements (typically 1 to 3) made on a sample spiked at the cut-off concentration. In other words, the threshold value is sampled, not fixed, and this means an unknown error of variable size is attached to the estimated value. This implies that the estimated threshold will move from experiment to experiment, which has a domino effect on which samples get called "above" or "below" cut-off, and thus on the positivity curve.

Fortunately, these measurement characteristics can be modeled and taken into account in establishing validation criteria, i.e., their impact on the positivity curve can be calculated. Details of the modeling performed in RStudio can be found in Appendix E.2. The resulting positivity curve is shown in Figure 4.1.1c. Notable differences result, particularly at the high concentration end, which affects the expected false negative rate and other parameters relying on it (reliability and sensitivity rates).

This software tool can also be used in establishing appropriate validation criteria by

[3]Note that moving the cut-off concentration is very different from moving the hypothesis test threshold, and is more akin to moving a measured value and its confidence interval.

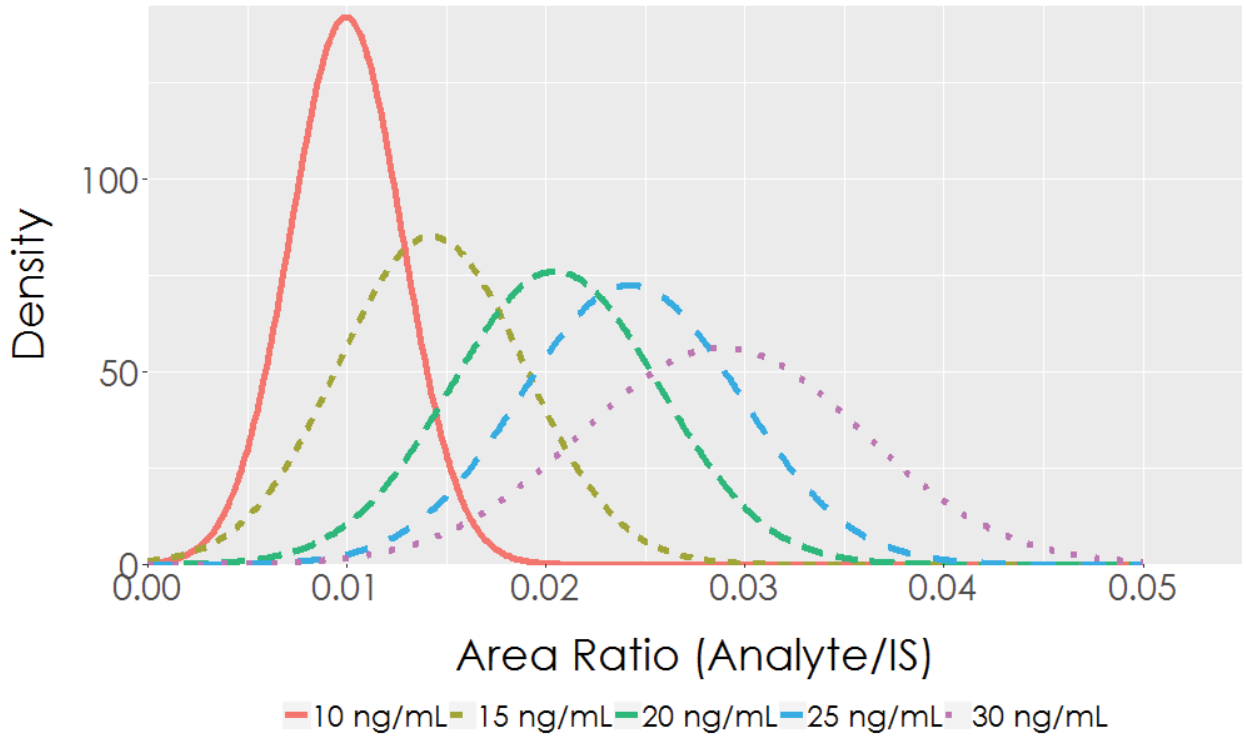


Figure 4.1.3: Fitted normal distribution curves for buprenorphine samples ($n = 30$) spiked at different concentrations. The low area ratios stem from the use of an internal standard primarily intended for quantification of an analogue (in the quantitative part of the method) combined with a low cut-off concentration for the qualitative analyte.

modelling realistic behaviour (heteroscedastic data, cut-off with sampling error) of the measurements and the derived binary results.

Derived method validation guidelines

The derived method validation guidelines presented in Section 4.1.3 utilize the performance of the actual method's output, the binary "above cut-off"/"below cut-off" results. With these guidelines we recognize the presence of measurement unreliability due to measurement error. Consequently, evaluation of the method's performance and validation needs to be performed outside of the unreliability zone but provides the most pertinent figures of merit when performed near these boundaries (lower and upper unreliability limits). However, depending on the purpose of the method, laboratories might find it pertinent to precisely evaluate the size of the unreliability zone, or might be satisfied by performing validation well outside of it through an overestimation of the UR size (e.g., ±50% of the cut-off concentration). But ultimately, this measurement error must be acknowledged in the production

setting as well. Method validation will confirm reliable performance for measurements below $-3s$ (LURL)and above $+3s$ (UURL), but what about measurements between these limits? These fall in the unreliability zone and must be identified and reported as such. Measurements between $-3s$ (LURL) and the measurement at cut-off should be reported as "likely below cut-off", and those between the measurement at cut-off and $+3s$ (UURL) should be reported as "likely above cut-off". This will reflect the fact that repeated measurements on these samples might yield different results, and will adequately convey measurement uncertainty in the final analysis report.

Validation of the qualitative decision point LC-MS/MS method

The method validation guideline initially developed was used in an attempt to validate an LC-MS/MS qualitative decision point method for 40 analytes. Standard deviation estimation was performed based on the analysis of 10 samples spiked at cut-off. Each analyte produced a unique standard deviation and therefore an UR zone of unique size. It follows, in principle, that the concentrations used for all subsequent validation steps should also be unique to each analyte. For the probability curves alone, 40 analytes x 10 samples x 9 concentration levels = 3 600 spiked samples would need to be analyzed, an unmanageable workload for the laboratory. Instead, analytes were classified as belonging to one of three standard deviation (%RSD) bins: 8%, 16.5% or 25% (Table 4.1.1). This binning process reduced the requirements to 3 bins x 10 samples x 9 concentration levels = 270 spiked samples, a reduction by a factor of 13. Samples spiked at $-4s$, $-3s$, $-2s$, $-1s$, cut-off, $+1s$, $+2s$, $+3s$ and $+4s$ were analyzed, and a smoothed conditional mean was fitted to the calculated positivity rate. Generally, all analytes produced the expected sigmoidal curve outcome, with some expected deformations attributed to the binning process.

To measure performance parameters and ion ratio reliability, 30 samples spiked at the LURL and UURL for each of the %RSD bins were analyzed. For example, MDEA (cut-off = 20 ng/mL, %RSD = 8%) was spiked at 15 and 25 ng/mL. Results show that numerous performance parameters fall outside of the expected range (greyed-out cells in Table 4.1.1). On the other hand, ion ratio, carryover and interference studies were all found to be satisfactory. Overall, 27 out of 40 analytes were considered to be validated.

The validation failure of so many analytes was quite surprising, and, in principle, should not have occurred if the theoretical model of measurements and derived binary results was

Table 4.1.1: LC-MS/MS validation results.

Analyte	Cut-Off (ng/mL)	s (ng/mL)	Bin	LURL (ng/mL)	UURL (ng/mL)	FNR	FPR	RLR	SLR	SNR	Validated?
α-Hydroxyalprazolam	20	0.13	0.165	10	30	0%	0%	100%	100%	100%	YES
Aripiprazole	10	0.1	0.08	8	12	30%	0%	85%	100%	70%	NO
3-Hydroxy Bromazepam	20	0.17	0.165	10	30	0%	0%	100%	100%	100%	YES
Buprenorphine	5	0.31	0.25	1	9	3%	0%	98%	100%	97%	YES
Hydroxybupropion	20	0.07	0.08	15	25	3%	0%	98%	100%	97%	YES
N-Desmethylcitalopram	20	0.08	0.08	15	25	3%	0%	98%	100%	97%	YES
N-Desmethylclobazam	20	0.11	0.08	15	25	20%	0%	90%	100%	80%	NO
Cocaethylene	20	0.08	0.08	15	25	0%	0%	100%	100%	100%	YES
Norcodeine	20	0.08	0.08	15	25	27%	0%	87%	100%	73%	NO
N-Desmethylcyclobenzaprine	20	0.16	0.165	10	30	0%	0%	100%	100%	100%	YES
Dextrorphan	20	0.11	0.08	15	25	7%	0%	97%	100%	93%	YES
Nordiazepam	20	0.09	0.08	15	25	3%	0%	98%	100%	97%	YES
N-Desmethyl diphenydramine	20	0.11	0.08	15	25	0%	0%	100%	100%	100%	YES
Duloxetine	20	0.15	0.165	10	30	7%	0%	97%	100%	93%	YES
Norfentanyl	0.5	0.13	0.08	0	1	43%	0%	78%	100%	57%	NO
7-Aminoflunitrazepam	20	0.09	0.08	15	25	40%	0%	80%	100%	60%	NO
N-Desmethylflunitrazepam	20	0.12	0.08	15	25	30%	0%	85%	100%	70%	NO
Norfluoxetine	20	0.22	0.25	5	35	0%	0%	100%	100%	100%	YES
2-Hydroxyethylflurazepam	20	0.09	0.08	15	25	80%	0%	60%	100%	20%	NO
Norketamine	20	0.1	0.08	15	25	10%	0%	95%	100%	90%	NO
Lorazepam-glucuronide	40	0.24	0.25	10	70	10%	0%	95%	100%	90%	NO
mCPP	20	0.1	0.08	15	25	0%	0%	100%	100%	100%	YES
MDEA	20	0.08	0.08	15	25	7%	0%	97%	100%	93%	YES
MDPV metabolite	20	0.1	0.08	15	25	17%	0%	92%	100%	83%	NO
Normeperidine	40	0.08	0.08	30	50	3%	0%	98%	100%	97%	YES
α-Hydroxymidazolam	20	0.09	0.08	15	25	23%	0%	88%	100%	77%	NO
N-desmethylmirtazapine	20	0.12	0.08	15	25	0%	0%	100%	100%	100%	YES
6-Acetylmorphine	5	0.1	0.08	4	6	3%	0%	98%	100%	97%	YES
Morphine-6β-D-glucuronide	100	0.31	0.25	25	175	7%	0%	97%	100%	93%	YES
Naloxone	20	0.08	0.08	15	25	3%	0%	98%	100%	97%	YES
Naltrexone	20	0.09	0.08	15	25	7%	0%	97%	100%	93%	YES
Desmethylolanzapine	20	0.3	0.25	5	35	0%	10%	95%	90%	100%	YES
Oxazepam-glucuronide	20	0.2	0.165	10	30	10%	0%	95%	100%	90%	NO
Phenylpropanolamine	30	0.09	0.08	23	37	3%	0%	98%	100%	97%	YES
Norpseudoephedrine	30	0.08	0.08	23	37	7%	0%	97%	100%	93%	YES
Norquetiapine	20	0.23	0.25	5	35	3%	0%	98%	100%	97%	YES
7-Hydroxyquetiapine	20	0.1	0.08	15	25	0%	0%	100%	100%	100%	YES
Temazepam-glucuronide	20	0.19	0.165	10	30	3%	0%	98%	100%	97%	YES
α-Hydroxytriazolam	20	0.13	0.165	10	30	0%	0%	100%	100%	100%	YES
N-Desmethylzopiclone	20	0.09	0.08	15	25	7%	0%	97%	100%	93%	YES

correct. It therefore seemed that something was not taken into account by the model. Further investigation revealed that the average response (area ratio) at cut-off and, more importantly, its standard deviation, shifted on a batch to batch basis, with an even more marked difference between days (Appendix E.4). The measurement error was therefore not adequately characterized or controlled. This type of analysis thus displays a two-part heteroscedasticity: the standard deviation changes with the concentration, which is properly accounted for by the model presented here; and the second part is, apparently unstructured and unrelated to other factors. This was the key to understanding the disappointing validation results and should be taken as a precautionary warning. The standard deviation changed between batches and daily, which meant that the size and edges (LLURL, UURL) of the unreliability zone also varied daily. Appendix E.4 shows, for example, an average 6-fold increase in the standard deviation between two batches. Therefore, while we thought we were measuring method validation parameters (FNR, FPR, SNR, SLR, RLR) at the binned $-3s$ and $+3s$ for each analyte, we might actually have been making these measurements significantly away from the edges, either inside or outside of the unreliability zone on that particular day. This will, naturally, have a major impact on the positivity rate, the method performance and its measurement uncertainty.

Modified method validation guidelines

Having a reliable estimation of the unreliability zone is important to apply adequate criteria on validation parameters, but also to properly take into account measurement uncertainty in production operations and accurately classify samples as below cut-off/likely below cut-off/likely above cut-off/above cut-off. Knowing that the size of this unreliability zone varies on a daily basis, the next obvious question is: can the position of its edges (LURL, UURL) be estimated with each batch?

The problem with this approach is that accurate estimation of the standard deviation is more difficult than accurate estimation of an average, since this parameter converges more slowly than the mean. To obtain a standard deviation estimation with lower than 20% average error, one would have to analyze at least 10 samples spiked at cut-off per batch/day. Given the constraints of a production setting, this is impractical.

Therefore, until better mathematical predictive tools are developed, an accurate estimate of the size of the unreliability zone seems out of reach. For the moment, the best that can be done is to proceed conservatively. Either perform several estimations of its size on different

batches/days and use the largest one, or, based on experience, choose one that insures that the validation points are outside the unreliability zone, e.g., at ±50% of the cut-off concentration. Note that this is essentially what is done with immunoassay method validations [75].

For validation of this LC-MS/MS method, all analytes in the 8% RSD bin were pushed up into the 16.5% RSD bin, yielding LURL and UURL at 50% and 150%. Using these relaxed conditions, all analytes satisfied the validation criteria.

4.1.5 Conclusions

Qualitative methods yield categorical, binary outputs very different in nature from quantitative methods, and validation guidelines should employ these categorical results to evaluate method performance, not the continuous measurements collected in the process. We have developed a tool to model the measurements and the derived binary results based on the literature and experimental data.

A tentative validation guideline was developed and applied to an LC-MS/MS qualitative decision point method for 40 analytes. Results also demonstrated a previously unreported behaviour of this type of measurements: the average area ratio and its variance changes on a daily basis, leading to significant variations of the unreliability zone size which is critical for method validation.

Considering this behaviour, we offer the following validation guidelines:

1. Decide on the validation points to be used above and below cut-off (LURL, UURL). If the size of the uncertainty of measurement is important, the standard deviation can be repeatedly evaluated on different days and the largest one used to establish conservative $\pm 3s$ validation points[4]. If not, a conservatively large size such as cut-off ±50% can be used.

2. 30 samples should be spiked at the validation points (above and below cut-off) and treated as they would in a production setting (i.e., analyze those samples as they

[4]Note that this is a conservative option keeping in line with the forensic toxicology practice. Other areas of application might be satisfied with alternatives such as pooling variance data from several days, although this will mean that on some days, the actual variance will be larger than the (pooled) estimated one.

would be in a real batch, with the same number of extracted cut-offs as will be used in production, and generate binary results). The validation parameters should satisfy the following criteria (for 2 injected cut-offs and 30 samples used for rate estimation): observed $FNR \leq 7\%$, $FPR = 0\%$, $RLR \geq 93\%$, $SLR = 100\%$, $SNR \geq 93\%$, ion ratio adequacy $> 95\%$. For other conditions, the R script in Appendix E.2 can be used to estimate expected performance.

3. Carryover and interference studies according to SWGTOX's practices should be performed and satisfy pre-established criteria.

4. If appropriate, dilution integrity can be assessed by repeating Step 2) with the desired dilution and verification that validation parameters continue to satisfy the above criteria.

5. In production, samples whose response falls below the low validation point are reported as "below cut-off", samples with a measurement between the low validation point (LURL) and the cut-off as "likely below cut-off", samples with a measurement between cut-off and the high validation point (UURL) as "likely above cut-off" and samples with a measurement above the high validation point as "above cut-off".

Using this validation guideline and method of reporting results not only produces a performance evaluation more in line with the definition of method validation, but also takes into account measurement uncertainty, as required by the new ISO 17025:2017 [72] validation guidelines. In Section 4.3, this framework was successfully applied to a method covering 92 analytes in saliva collected using a Quantisal® device.

4.1.6 Acknowledgements

The authors are grateful to Maxime Gosselin for his contribution to the literature review in the early days of the project. Brigitte Desharnais, Félix Camirand-Lemyre and Cameron D. Skinner gratefully acknowledge support of the National Sciences and Engineering Research Council of Canada. Brigitte Desharnais also gratefully acknowledges the support of the Fonds de recherche du Québec - Nature et technologies. This research was undertaken thanks in part to the funding from Canada First Research Excellence Fund and the Australian Research Council DP #140100125.

4.2 Legislation as a driving factor for analytical methods in forensic toxicology

In Section 4.1, experimental data and *in silico* simulations were used to derive validation guidelines for qualitative decision point methods, in particular for LC-MS/MS instrumentation, which was found to present characteristics which needed to be taken into account. For example, pronounced heteroscedasticity of the data was observed. Buprenorphine, as shown in Figure 4.1.3, is the perfect example of this phenomenon, but all analytes monitored display this behaviour. This is why it was necessary to integrate this characteristic, and others, into the data modelling tools. Although a multi-analyte LC-MS/MS method was used and validated using this process, these guidelines had yet to be applied to a full and linear validation setting. This is accomplished in Section 4.3, where a qualitative decision point method is validated for 92 analytes in oral fluid.

This method is a great example of how legislation can be central to method development and validation in forensic toxicology. In October 2018, the Cannabis Act [161] was enacted in Canada, allowing legal production, distribution and usage of cannabis in the country. Cannabis legalization also came with a reform to the drugs and driving legislation, generally referred to as Bill C-46 [104]. Under this new legislation, police officers can now administer an oral fluid drug test using a point-of-collection screening device approved by the Attorney General of Canada. These devices, which are new on Canadian soil, must detect at least one of these, and only these, drugs at the following thresholds: tetrahydrocannabinol (THC, 25 ng/mL), cocaine (50 ng/mL) and methamphetamine (50 ng/mL), in addition to including a reader and printed record of analytical results which avoids subjective interpretation by users.

Another major change introduced by C-46 was facilitated access to blood samples from suspects, accompanied by *per se* levels for 10 drugs (THC, GHB, cocaine, methamphetamine, 6-monoacetylmorphine, phencyclidine (PCP), ketamine, lysergic acid diethylamide (LSD), psilocin and psilocybin).

With these two aspects of the new legislation relying on the application of cut-off concentrations, it is easy to see how the work presented in Section 4.1 is timely and takes an important place in the Canadian forensic toxicology context. A planned roadside survey, with collection of oral fluid samples, provided the perfect opportunity (and necessity) to apply the tools developed to a new analytical method, as presented in Section 4.3.

4.3 Application to a multi-analyte LC-MS/MS method for oral fluid

4.3.1 Abstract

A study of impaired driving rates in the province of Québec is currently planned following the legalization of recreational cannabis in Canada. Oral fluid (OF) samples are to be collected with a Quantisal® device and sent to the laboratory for analysis. In order to prepare for this project, a qualitative decision point analysis method monitoring for the presence of 97 drugs and metabolites in OF was validated according to the guidelines presented in Section 4.1.

This high throughput method uses incubation with a precipitation solvent (acetone:acetonitrile 30:70 v:v) to boost drug recovery from the collecting device and improve stability of benzodiazepines (e.g., α-hydroxyalprazolam, clonazepam, 7-aminoclonazepam, flunitrazepam, 7-aminoflunitrazepam, N-desmethylflunitrazepam, nitrazepam). The Quantisal ® device has polyglycol in its stabilizing buffer but timed use of the mass spectrometer waste valve proved sufficient to avoid the glycol interferences for nearly all analytes. Interferences from OF matrices and 140 potentially interfering compounds, carryover, ion ratios, stability, recovery, reproducibility, robustness, false positive rate, false negative rate, selectivity, sensitivity and reliability rates were tested in the validation process. Five of the targeted analytes (olanzapine, oxazepam, 7-aminoclonazepam, flunitrazepam and nitrazepam) did not meet the set validation criteria but will be monitored for identification purposes (no comparison to a cut-off level).

Blind internal proficiency teting was performed, where six OF samples were tested and analytes were classified as "negative", "likely positive" or "positive" with success. The final validated OF qualitative decision point method covers 92 analytes, and the presence of 5 additional analytes is screened in this high throughput analysis.

4.3.2 Introduction

Scientific papers on the use of oral fluid (OF) in forensic toxicology were published as early as 1965 [162, 163], but the use of this alternative matrix has become more widespread specifically since the 2000s [164, 165]. There is no doubt that the numerous advantages

of OF play a role in its increased use. Indeed, OF collection is easy and minimally invasive [108, 109, 165–170], allowing roadside collection and testing [109, 166, 168–170], and reduces the legal burden to obtain a biological sample. Moreover, the risk of sample adulteration is considered to be lower than with urine [108, 165, 166, 168–170], which can be diluted or modified with an adulterant [108, 166, 168]. In contrast to urine, from which only past use can be inferred, OF analysis will inform the forensic toxicologist about recent use [108, 109, 166, 168–170]. Indeed, OF has a similar detection window to blood [165, 168], but has the noteworthy advantage of being easily collected at the roadside, shortly after arrest. Three main disadvantages are recognized with regards to OF use. First, contamination of the oral cavity with substances in direct contact with the mouth is to be expected, for example THC from smoked cannabis [108, 171]. Second, low saliva production in certain drug users might complicate an otherwise easy collection process and result in artificially increased concentrations [169]. Finally, drug concentrations in OF are not as well documented as they are in blood, complicating the interpretation process for the forensic toxicologist.

Nevertheless, on the whole, OF might be the matrix of choice to test for individuals driving under the influence of drugs. This alternative matrix represents a worthwhile compromise between ease of collection and toxicological relevance to impairment, which is often associated with recency of use [108, 109, 166–169]. With this application in mind, several OF screening devices [109, 166, 170] and OF collection devices for laboratory confirmation [108, 109, 165–167, 169, 170] have been developed since 1990 [168, 172]. The selection of a particular device relies on several considerations including cost, effectiveness and intended use [167].

The recent legalization of recreational cannabis in Canada [161] has been accompanied by several modifications to the impaired driving legislation [104]. Of particular interest here is the introduction of OF drug testing using point-of-collection screening devices [173] as an investigation tool in driving under the influence of drugs (DUID) cases. These instruments must be approved for use in Canada by the General Attorney and be able to detect at least one of the following: THC at 25 ng/mL, cocaine at 50 ng/mL and/or methamphetamine at 50 ng/mL [174]. Canada is not the first country to use these devices in this particular context and is following in the footsteps of Australia, Belgium, Finland, France, Germany, Italy and specific states in the United States [108, 166–168].

In the midst of these legislative modifications, a provincial study here in the province of Québec of impaired driving rates and the relevance of OF point-of-collection screening

devices is planned. At the checkpoint, the drivers are to be tested with a device currently approved for use in Canada [173, 175]. An OF sample is also to be collected from all drivers passing through this checkpoint and sent to the laboratory for testing. The Quantisal® collection device from Immunalysis, which has already demonstrated its effectiveness for drugs of abuse and therapeutics analysis [165, 167, 169], was selected for this purpose. This type of study has been carried out in several other jurisdictions such as Italy, Norway and the United States (Wisconsin) [110–113].

This paper presents the development and validation of a qualitative decision point method fit for the aforementioned study. The method needed to be high-throughput, since 2 500 OF samples would be received for analysis over 28 days. It covers an extensive set of analytes (97 in total), including cocaine, benzoylecgonine, amphetamines, benzodiazepines, cannabis (THC) and opioids, as they are the most prevalent DUID findings in Québec [176]. Samples were prepared using dilution with an organic solvent, and analyzed by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS).

A qualitative decision point method was developed rather than a quantitative method, for three main reasons. First, toxicological interpretation of OF drug concentration(s) is not yet well-established; second, only a short time frame could be allotted for method development and validation, and finally, it permits comparison with the point-of-collection screening device(s). Validation of this method was performed according to the guidelines presented in Section 4.1 and ISO 17025:2005 [71] requirements. Following method optimization, the absence of interferences and carryover was confirmed. False negative, false positive, selectivity, sensitivity and reliability rates were determined and confirmed to be reproducible and robust. Ion ratios, stability and recovery were also validated. The production-ready method permits classification of samples as being below cut-off, likely above cut-off or above cut-off by taking into account the uncertainty of measurement (as required by ISO 17025:2017) [72].

4.3.3 Materials and methods

Development and optimization

Polyglycol interferences

Polyglycols are known to be present in the Quantisal® stabilizing buffer [177] and can be deleterious to the mass spectrometer and accuracy of the results. The presence and behaviour of polyethyleneglycol (PEG) under the selected chromatographic conditions was investigated.

A blank OF sample was collected with the Quantisal® device, extracted as described below and analyzed using the two chromatographic methods in full scan mode rather than target multiple reaction monitoring (MRM) mode. The resulting total ion chromatograms (TIC) were investigated for the presence of characteristic PEG profiles, i.e., an envelope of peaks spaced 44 Da apart (mass of an ethylene glycol unit) [178].

Maximizing recovery

Analyte recovery for the 97 compounds targeted (see Appendix F.1 for a complete list) was estimated via the area ratios (analyte to internal standard) for spiked OF. Samples taken up with the collector stick and put in the Quantisal® tube were compared to the same reference OF added directly to the stabilizing buffer in the Quantisal® tube. Recovery was calculated as:

$$\text{Recovery}(\%) = \frac{\text{Area ratio with collector}}{\text{Area ratio directly in stabilizing buffer}} \times 100 \tag{4.6}$$

Recoveries $> 80\%$ for any given analyte were considered acceptable.

Recoveries were evaluated for several experimental conditions: collector stick equilibration times of 24, 48, 72, 96 and 120 hours (with the collector stick in the stabilizing buffer); inclusion of a 10 minute sonication step; addition of the organic precipitation solvent directly to the Quantisal® tube still containing the collector stick; added solvent volumes from 0.5 mL to 8.5 mL and solvent incubation times from 17 to 127 hours.

Final analytical method

Preparation of control and threshold samples

Samples of OF were collected from voluntary laboratory employees and anonymized. Aliquots (1.5 mL) were then spiked at the required concentrations with the 97 targeted analytes (see Appendix F.1 for threshold (cut-off) concentrations). All compounds were purchased from Cerilliant (Round Rock, Texas, USA), except for N-desmethyl diphenhydramine, procyclidine and rolicyclidine which were obtained from Toronto Research Chemicals (North York, Ontario, Canada) and 3,4-methylenedioxypyrovalerone metabolite which was secured from Cayman Chemicals (Ann Harbor, Michigan, USA). The OF samples were spiked in borosilicate glass tubes 16 $\times$ 100 mm (Fisher Scientific, 14-961-29, Fair Lawn, New Jersey, USA).

The Quantisal® collector pad was inserted into the glass tube, in contact with the OF, until blue coloration on the collector indicated saturation of the pad with OF (1 $mL \pm 10\%$, according to manufacturer's documentation). At this point, the collector was transferred into the Quantisal® tube (Immunalysis, QS-0025, Pomana, California, USA) containing 3 mL of stabilizing buffer. Samples were stored as is at 4 °C overnight prior to sample extraction to simulate the expected sample shipment delay of future checkpoint studies.

Sample extraction

In the Quantisal® tube, 4.5 mL of acetone:acetonitrile (30:70 v:v) (acetone: HPLC Grade, Fisher Scientific, A949, Fair Lawn, New Jersey, USA; and acetonitrile: HPLC Grade, $\geq$ 99.9%, EMD Millipore corporation, AX0156-1, Billerica, Massachusetts, USA) organic solvent was added, without removing the collection device. Tubes were capped, mixed by inversion and incubated for 72 h at 4 °C. Following incubation and vortexing, 600 μL of extract was transferred to a 2 mL square well 96-well plate (Fisher Scientific, AB-0932, Ottawa, Ontario, Canada). Stable isotope internal standards (IS) solution (10.0 μL) were added to each well; the compounds and concentrations are detailed in Appendix F.1. Following mixing (1 minute at 1500 rpm on Thermomixer, Eppendorf, Mississauga, Ontario, Canada) and centrifugation (5 minutes at 3200 $\times$ g), two different supernatant dilutions were prepared for analysis with the two chromatographic methods. The first chromatographic method (general) covered 96 of the 97 analytes targeted; for this purpose, 25.0 μL of supernatant was transferred to a different 96-well plate equipped with 1 mL round bottom wells (Canadian Life Science, RT96PPRWU1mL, Peterborough, Ontario, Canada) and diluted with 180 μL of 0.2% formic acid (Fisher Chemical, A117-50, Fair Lawn, New Jersey, USA). The second chromatographic method (cannabinoid) was designed specifically for cannabinoids analysis; for this purpose, 200 μL of supernatant was transferred to an identical, round bottom, 96-well plate and diluted with 50 μL of 1.5% formic acid.

LC-MS/MS analysis

For the general chromatography, 5 μL of the extract was separated in 13 minutes on an Agilent Zorbax Eclipse Plus C18 (2.1 $\times$ 100 mm, 3.5 μm) maintained at 50 °C. Mobile phase A was an aqueous solution of ammonium formate (pH 3.0):methanol (98:2). Mobile phase B was methanol (EMD Millipore corporation, MX0486-1, Billerica, MA, USA). A 650 $\mu L/min$ step/ramp from the A to the B mobile phase was used. Detailed gradient, analyte retention

times, Q1/Q3 identification and confirmation transitions, source and mass spectrometer parameters are available in Appendix F.1.

For the cannabinoid chromatography, a 10 μL aliquot of the extract was separated using the above conditions, except the column was 50 mm long and a 550 $\mu L/min$ flow rate was used for the 6.5 minute separation. Analytical details are available in Appendix F.1.

In both cases, an Agilent HPLC 1200 or 1260 Infinity coupled to a Sciex 5500 QTrap mass spectrometer were used. The data acquisition software used was Analyst® 1.6.2 build 8489. Data was analyzed using Multiquant® 3.0.1 (Version 3.0.6256.0) software.

Validation procedures

Method validation was performed according to the principles established in Section 4.1, while considering SWGTOX's [75] recommendations and conforming to ISO 17025:2005 [71] and CAN-P-1578 [73] requirements.

Interferences

The presence of interferences from the matrix or exogenous compounds was assessed. To evaluate interferences from the matrix (specificity), 15 blank OF samples were analyzed with the final analytical method, including both chromatographies. Similarly, 140 potentially interfering compounds including caffeine, nicotine, cannabidiol, cannabinol, Δ8-THC, exo-THC and cannabichromen were added to blank OF and analyzed using both chromatographic methods; the full list of compounds tested and their concentrations are available in Appendix F.2. Interferences were considered to be present, from the matrix or potentially interfering compounds, if the analyte peak area ratio in the blank was greater than 25% of the average analyte peak area ratio of the cut-off sample; or if internal standard peak area in the blank was greater than 5% of the average internal standard area in cut-off samples.

Carryover

Three OF samples were prepared: one spiked with all compounds at a concentration 50 times higher than the cut-off; the second (analyte blank) was spiked with internal standards only and the third (full blank) was not spiked. These samples were extracted and analyzed

in pairs (spiked followed by analyte blank or full blank) on three separate days with two replicates, for a total of 6 carryover tests. Carryover was considered to be present for an analyte when the analyte peak area ratio, in the analyte blank, was greater than 25% of the average area ratio of the cut-off samples; or for an internal standard when the internal standard peak area, in the full blank, was greater than 5% of the average internal standard area in cut-off samples.

Ion ratios

The ion ratio for each analyte was calculated as:

$$\text{Ion ratio} = \frac{\text{Area}_{\text{Transition 1}}}{\text{Area}_{\text{Transition 2}}} \times 100 \tag{4.7}$$

The ion ratio in all available samples spiked at 100% and 150% of the cut-off concentration (rates/reproducibility/robustness experiments, total of 60 samples from 15 different OF donors) were compared for each analyte to the average area ratio of 4 OF reference samples at 25 times the cut-off concentration. Ion ratios were expected to be within ±30% of the average ion ratio measured in the reference samples in ≥ 90% of cases.

Stability

Analyte stability at the cut-off concentration was evaluated both in the Quantisal® stabilizing buffer (i.e., the collector pad filled with OF in the Quantisal® tube, as the sample would be if it was shipped from the collection point to the laboratory) and in the organic solvent supplemented Quantisal® (i.e., after the addition of organic solvent for extraction purposes). In both cases, stability was evaluated at 4 °C over four weeks. Stability was calculated using the average of three OF samples as:

$$\text{Stability}(\%) = \frac{\overline{\text{Area ratio}}_{t=X} - \overline{\text{Area ratio}}_{t=0}}{\overline{\text{Area ratio}}_{t=0}} \times 100 \tag{4.8}$$

with a target stability of 0% (a negative stability indicating degradation) and $t = 0$ wk being freshly spiked and extracted samples. To be considered stable at time $t = X$ wk, the analyte's stability should be $> -20\%$. However, in interpreting the stability results, the pattern of stability over the course of the study, not just one time point, was considered. A single out-of-criteria point at, e.g., 2 weeks might be more reflective of a biased accuracy on that day when the next two time points showed stabilities closer to the 0% target.

Recovery

Recovery was evaluated according to the procedure described above using 5 OF samples spiked at the cut-off concentration.

False negative (FNR), false positive (FPR), selectivity (SLR), sensitivity (SNR) and reliability (RLR) rates

Method performance parameters were calculated on OF samples spiked with all analytes at 50% and 150% of the cut-off concentration. Results for the 15 samples at 50% of the cut-off were classified as true negative or false positive, whereas results for the 15 samples at 150% of the cut-off were classified as true positive or false negative. Results were used to calculate the performance parameters according to the equations detailed in Section 4.1. The acceptance criteria for the rates were as follows: observed $FNR \leq 7\%$, $FPR = 0\%$, $RLR \geq 93\%$, $SLR = 100\%$, $SNR \geq 93\%$.

Reproducibility and robustness

Reproducibility and robustness were evaluated by carrying out an evaluation of the rates as described in the previous paragraph on three different days, changing the solutions and mobile phases lots, HPLC columns, LC-MS/MS instruments and technical staff. The same acceptance criteria for the different rates apply.

Internal proficiency testing

In production, 5 samples of OF spiked at the cut-off concentration and 3 samples of OF spiked at 150% of the cut-off concentration were used to establish the classification bins. The average area ratio ($\overline{AR}_{CO}$) measured in the cut-off samples acted as the measurement threshold. Unknown samples with $AR_{UNK} < \overline{AR}_{CO}$ were classified as negative; samples with $\overline{AR}_{CO} \leq AR_{UNK} \leq \overline{AR}_{150\%}$ were classified as likely positive; and samples with $AR_{UNK} > \overline{AR}_{150\%}$ were classified as positive, as suggested in Section 4.1.

In order to test that the procedure performed as expected, six OF samples were fortified with different compounds at various concentrations, adsorbed onto the collector pad and

stored in the Quantisal® tube. A laboratory member who was "blinded" to their concentrations analyzed these samples. The results (negative/likely positive/positive) were checked for accuracy by the laboratory member who performed sample spiking. This blind analysis was carried out in triplicate over two different batches.

4.3.4 Results and Discussion

Development and optimization

Because of the short time frame available for method development and validation, the most efficient path was to modify an existing LC-MS/MS targeted screening method [158] to suit the needs of this project (new matrix, qualitative analysis). In doing this, most of the method development concerns centered on the impact of using this new matrix: the OF in a Quantisal® stabilizing buffer. The Quantisal® device was selected based on literature reports and communications with other forensic toxicology laboratories. No other collection device was evaluated for this project, due to the short timeframe.

Polyglycol interferences

Although the exact contents of the Quantisal® stabilizing buffer are proprietary, the presence of polyglycol compounds such as polyethylene glycol (PEG) has been surmised in a few other studies [177, 179]. This family of compounds is known to be problematic for mass spectrometers by creating large ionic suppression and contaminating the quadrupoles [180]. It was crucial that this project, being secondary to the main laboratory production activity, did not jeopardize the performance of the mass spectrometry instrumentation. Presence of PEGs was, therefore, the first item examined in the course of method development. Full scan analyses of extracted blank OF in the Quantisal® stabilizing buffer were performed for both the general and cannabinoid chromatographic methods. Resulting total ion chromatograms (TICs) did not reveal the presence of PEGs in the general method; most likely, they eluted during the wash where the flow was diverted to waste. On the other hand, a typical polyethylene glycol pattern was identified in the cannabinoid method (Figure 4.3.1). The elution of PEGs in this separation coincided with the elution of 11-nor-9-carboxy-Δ9-tetrahydrocannabinol (THC-COOH) and 11-hydroxy-Δ9-tetrahydrocannabinol (THC-OH). The chromatographic conditions could have been modified to attempt to separate PEG interferences from THC-COOH and THC-OH; however, these metabolites have only a moderate importance to toxicological interpretation in OF [171]. Although the presence of THC-

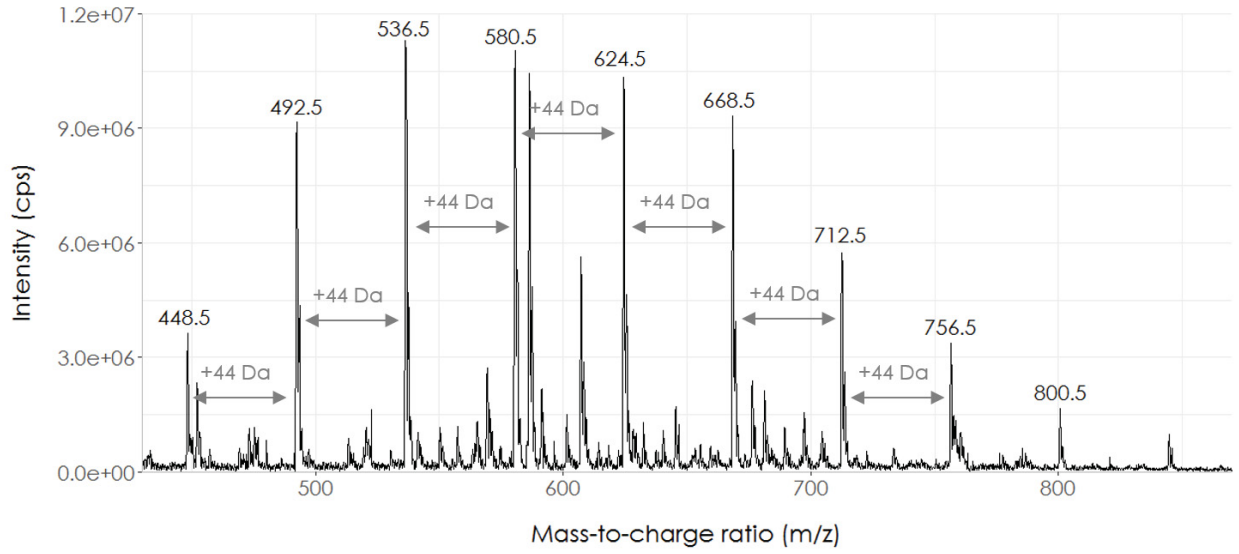


Figure 4.3.1: Polyethylene glycol (PEG) mass spectrum identified in the cannabinoid chromatography.

COOH in particular is relevant, it is present at very low levels (pg/mL) [181] in OF which, in any event, would be hard to detect with the present type of method. A separate, optimized method most likely using a pre-concentration step would need to be used [181]. Taking all of these elements into account, the final decision was to eliminate THC-COOH and THC-OH from the assay, which allowed the effluent to be diverted to waste during PEG's elution.

Maximizing recovery

Recovery from the whole collection device, including the collector pad and stabilizing buffer, should be estimated to be representative of analyzed samples [165, 166, 169]. Using this approach, it was quickly obvious that recovery from the device was an issue (Figure 4.3.2), with 55 out of 97 compounds achieving a recovery below the 80% threshold. Others have faced the same issue and increased the recovery by using a plunger to destroy the collection pad [166] or a serum separator to compress the collection pad [165]. The number of samples and time frame envisioned for the roadside study here called for a higher throughput solution, where individual collection pad crushing for each collected oral fluid sample was not required. Several solutions were tested, including longer incubation times, addition of a sonication step and incubation in the organic solvent. Figure 4.3.2 shows analyte recovery density plots for a standard one day preparation, with a 10 minute sonication step, a three day incubation at 4 °C prior to sample preparation, and a three day incubation with the or-

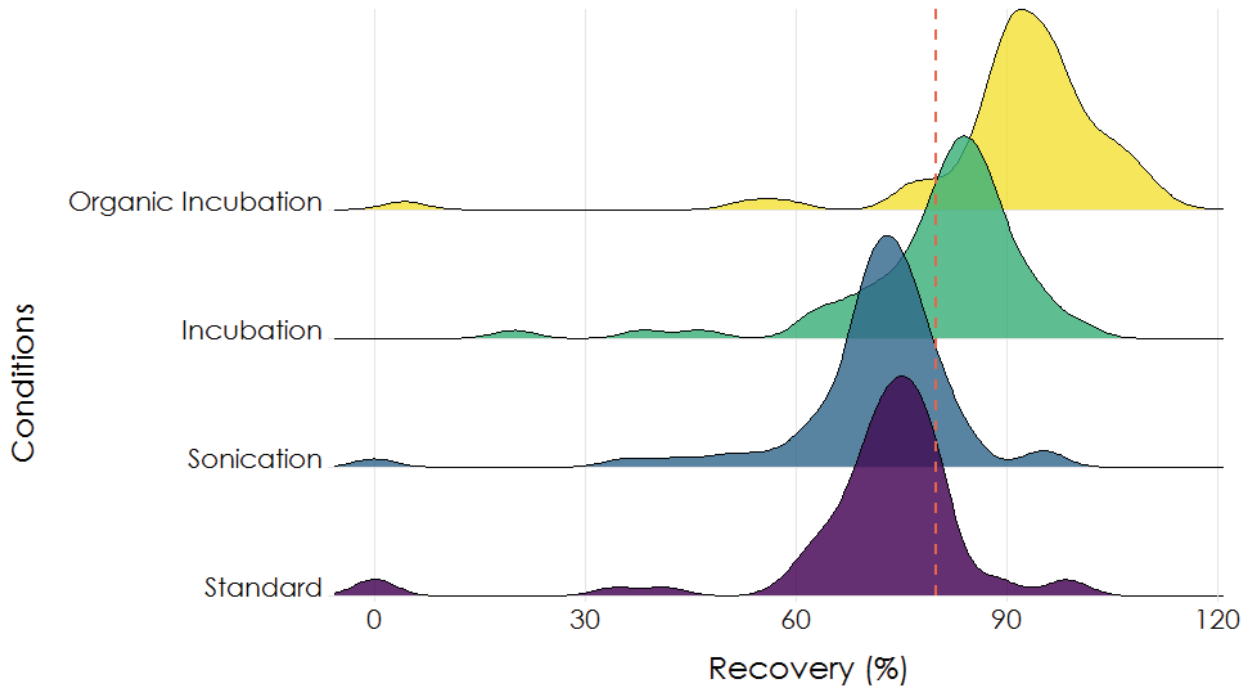


Figure 4.3.2: Density plots of the measured recoveries for all analytes under different preparation conditions.

ganic solvent. Peak density exceeds the 80% recovery threshold with an incubation period, and incubation with the organic solvent shows an additional recovery gain. Further optimization of the time and volume parameters for the organic solvent incubation was carried out to obtain the final extraction conditions.

Method validation

The main method validation results are summarized below, and complete results are available in Appendix F.2.

Interferences

Interferences from the OF matrix were observed for 7 analytes (aripiprazole, clobazam, cocaine, 7-amino-flunitrazepam, fluoxetine, lorazepam and N-desmethyl-mirtazapine), in 1 to 4 of the OF samples tested. In all cases however, the interferences were below the critical threshold (25% of the average area ratio in the cut-off samples). Interferences were also observed in all OF samples for THC's identification transition, but 0.05 minutes or more

after THC's retention time and thus outside of the integration window. No significant interferences from the potentially interfering compounds tested were observed; all interferences observed were either below the set thresholds or chromatographically separated. Despite being below the set threshold and narrowly chromatographically separated, an interference of nitrazepam-D_5 on oxazepam's identification transition was noted. Thus overall, no significant interferences were noted for the analytes and internal standards under study.

Carryover

Over the six blank samples analyzed, only one sample had analytes which did not satisfy the criteria for carryover from the high concentration sample (50 times the cut-off concentration). The analytes where carryover exceeded the threshold of 25% of the average area ratio in cut-off were: 7-amino-clonazepam (30%), benzoylecgonine (36%), 7-amino-flunitrazepam (33%) and O-desmethylvenlafaxine (35%). Given that the average carryover for each of these analytes was below the set threshold, none of the analytes were considered to have significant carryover. It is however noteworthy that 7-amino-flunitrazepam and oxycodone produce below threshold, but constant, carryover after injection of a high concentration sample. Samples immediately following those with a high concentration of these two analytes should be examined for potential carryover.

Ion ratios

The ion ratio of threshold and positive samples (100% and 150% of the cut-off concentration) meet the identification criteria for 99% of analyzed samples on average for all analytes except one: oxazepam. In this case, only 63% of the 60 samples analyzed meet the ion ratio criteria. This poor identification rate was attributable to the nitrazepam-D_5 interference on the identification transition, influencing the ion ratio for this analyte. Nitrazepam-D_5 was kept as an internal standard (IS) despite this interference because it is part of an IS mix used for other methods in the laboratory.

Stability

Stability of the OF samples, stored at 4°C, was evaluated over a period of 4 weeks. This time frame was longer than strictly necessary, since it is anticipated that samples will be

fully analyzed within 3 weeks of collection.

Several species exhibit an instability pattern in the Quantisal® stabilizing buffer alone, starting after 1 to 3 weeks of storage: α-hydroxyalprazolam, chlordiazepoxide, clonazepam, 7-aminoclonazepam, flunitrazepam, 7-aminoflunitrazepam, N-desmethyl-flunitrazepam, nitrazepam, olanzapine, N-desmethylolanzapine, zopiclone and N-desmethylzopiclone.

The good news is, addition of the organic solvent addresses this instability problem for most analytes. After addition of the organic solvent, the only analytes exhibiting an instability pattern are chlordiazepoxide (starting at 3 weeks), olanzapine (3 weeks) and N-desmethyl-olanzapine (2 weeks), while 7-aminoflunitrazepam and N-desmethylflunitrazepam exhibited some small instability pattern which did not cross the -20% threshold. These results are not surprising, since olanzapine and its metabolite [182] and chlordiazepoxide [183] are known to suffer from instability issues in most biological matrices and extraction solvents. In all cases, treating samples with the organic solvent less than one week after collection, and analyzing them within the following week should suffice to avoid potential instability.

Recovery

The recovery experiment was performed for information purposes only, i.e., a recovery $< 80\%$ would still be considered acceptable if all other validation criteria were met. Indeed, if the rates experiments, characterizing the quality of the output of this method, meet expected criteria despite a low recovery, this indicates internal standards and matrix matched cut-off samples correct sufficiently for the bias introduced by the low recovery. Only four analytes were found to have a recovery below the 80% threshold, albeit close to it: flunitrazepam (73.9%), norfluoxetine (74.3%), nitrazepam (70.3%) and N-desmethylolanzapine (79.4%).

Rates, reproducibility and robustness

False negative (FNR), false positive (FPR), sensitivity (SNR), selectivity (SLR) and reliability (RLR) rates were obtained for 45 different OF samples divided across three experiments, each run on a separate day. In the first experiment, olanzapine was out-of-specification for FNR, SNR and SLR. In the second experiment, out-of-specification results were obtained for olanzapine (FNR, SNR) and oxazepam (FPR, RLR, SLR). In the third experiment, out-

of-specification results were obtained for 7-aminoclonazepam (FNR, SNR), flunitrazepam (FNR, SNR) and nitrazepam (FNR, SNR). Combining results for the three batches, olanzapine, flunitrazepam and nitrazepam still did not meet the set criteria.

Olanzapine, 7-aminoclonazepam, flunitrazepam and nitrazepam all failed either the overall rates criteria or the reproducibility/robustness portion, in addition to having stability issues. A likely hypothesis is that the stability issues generate a higher rate of false negatives, in turn affecting the SNR. Oxazepam failed the reproducibility/robustness portion, in addition to the ion ratio specifications. The likely explanation here is that the interference from nitrazepam-D_5 contributed to increasing the number of false positives, in turn affecting the RLR and SLR.

Considering these results, olanzapine, 7-aminoclonazepam, flunitrazepam, nitrazepam and oxazepam were removed from the qualitative decision point method scope. These analytes are still monitored by LC-MS/MS, but the results are reported as "detected" or "not detected", i.e., positioning the sample with respect to a cut-off concentration cannot be reliably achieved. Rather, samples can only be screened. That being said, these drugs have low prevalence in the geographical region targeted by the roadside study (province of Québec, Canada), thus the impact of shifting those drugs from a qualitative decision point method to a screening only method is negligible.

Internal proficiency testing

Once the OF analysis method was validated for 92 analytes with a decision point and 5 analytes in screening mode, internal proficiency testing was performed. Spiked samples were treated as unknown samples (blind to the analyst) in exactly the same setting and conditions as in a production setting. 48 analytes were spiked at various levels around the cut-off concentrations in six different samples. Five extracted cut-off samples and three samples at 150% of the cut-off concentration (positive control) were analyzed as they would be in a production batch. Unknown samples, were classified as "negative" ($AR_{UNK} < \overline{AR}_{CO}$), "likely positive" ($\overline{AR}_{CO} \leq AR_{UNK} \leq \overline{AR}_{150\%}$) or "positive" ($AR_{UNK} > \overline{AR}_{150\%}$). The classification "likely negative" suggested in Section 4.1 was dropped due to its lower usefulness in a roadside survey study. Precautions against false positives should be taken, but false negatives in this context entails less consequences than false positive results. In samples analyzed (Appendix F.2), some below cut-off and $> 150\%$ cut-off samples did score as "likely positive", which was to be expected given the uncertainty of measurement and the concen-

trations used[5]. More importantly though, no false negative nor false positive results were detected, thus confirming the validity of the method.

4.3.5 Conclusions

We have successfully developed and validated a high throughput OF qualitative decision point analysis method for 92 analytes (and screening for 5 additional analytes). Amongst other possible applications, this atypically wide-scope method is suitable for OF roadside surveys where a large number of samples are collected and sent to the laboratory on the same day. Given the sample preparation and analysis time requirements, a total capacity for 350 to 400 samples per week is possible using one LC-MS/MS platform.

The extraction procedure was developed to maximize recovery from the collection device while remaining efficient in terms of laboratory labour and time required. Incubation with the precipitation organic solvent mixture achieves a high analyte recovery, while simultaneously curbing the stability issue of several benzodiazepines.

While solid phase extraction (SPE) is typically used for the analysis of such samples to deal with the polyglycol content of the stabilizing buffer, we have found that a dilution approach can be suitably deployed, enabling higher throughput. A judicious chromatographic method can divert the polyglycols before they reach the mass spectrometer. In the present case, co-elution of polyethylene glycol with THC-COOH and THC-OH prevented their inclusion in the final method but, the more important THC was successfully analyzed.

The validation guidelines presented in Section 4.1 allowed for efficient validation in compliance with ISO 17025 [71] requirements. Only four experiments were required to complete the validation stage. As an added benefit, using this type of validation and production analysis produces information of additional value to the client, introducing a notion of measurement uncertainty in the final result reported (i.e., "likely positive" vs. "positive", in accordance with ISO 17025:2017 [72] requirements).

[5]Note that performance rates were not measured on the "likely positive" limit, since its role is to act as the upper boundary of the unreliability zone and not as a cut-off. In the same way we are not measuring a confidence interval on the upper boundary of a confidence interval, we are not measuring rates for the "likely positive" boundary.

4.3.6 Acknowledgements

The authors wish to thank Cynthia Côté and Marc-André Morel for their work on the LC-MS/MS methods which inspired the one presented here and their thorough review of the manuscript, Marie-Pierre Taillon for her quality assurance work on this method, and Édith Viel for her review and comments. Brigitte Desharnais and Cameron D. Skinner gratefully acknowledge support of the National Sciences and Engineering Research Council of Canada. Brigitte Desharnais also gratefully acknowledges the support of the Fonds de recherche du Québec - Nature et technologies.

4.4 Mathematical method validation tools

Section 4.3 demonstrates clearly that the categorization scheme developed in Section 4.1 to deal with uncertainty of measurement (negative, probably negative, probably positive, positive) can be adapted to different casework requirements. In the case of oral fluid analysis, a false negative result was of lesser concern, but it was important to deal with false positive results. Thus the "likely negative" category was not implemented. In other types of casework, the opposite might happen, and laboratories will have the necessary flexibility to adapt the production settings as they see fit.

All the mathematical method validation tools, which were developed in the preceding chapters, were foreseen as necessary to validate quantitative and qualitative aspects of a CYP characterization and quantification method. What is presented next in Chapter 5 is the development, optimization and testing of said CYP method. Even though it is the last experimental chapter, work on CYP analysis started before any of the previous chapters and was carried out in parallel with the aforementioned tools.

5

Postmortem Estimation of Metabolic Capacity Through Cytochrome P450 Enzyme Characterization and Quantification — A Proof of Concept

5.1 Abstract

Metabolic capacity, an estimation of the rate at which xenobiotics are transformed by drug metabolizing enzymes in the human body, can be a relevant piece of information for forensic toxicologists, allowing for example differentiation between a medical error and an accidental intoxication. Metabolic capacity is typically estimated postmortem via enzyme genotyping. However, this classification of individuals according to their metabolizer type suffers from poor correlation with the actual metabolic capacity due to several factors affecting enzyme expression. As an alternative to the genomics based estimation, we describe a proteomics approach to characterization and quantification of cytochrome P450 (CYP) enzymes which has the potential to yield a more accurate estimation of metabolic capacity. CYP-rich

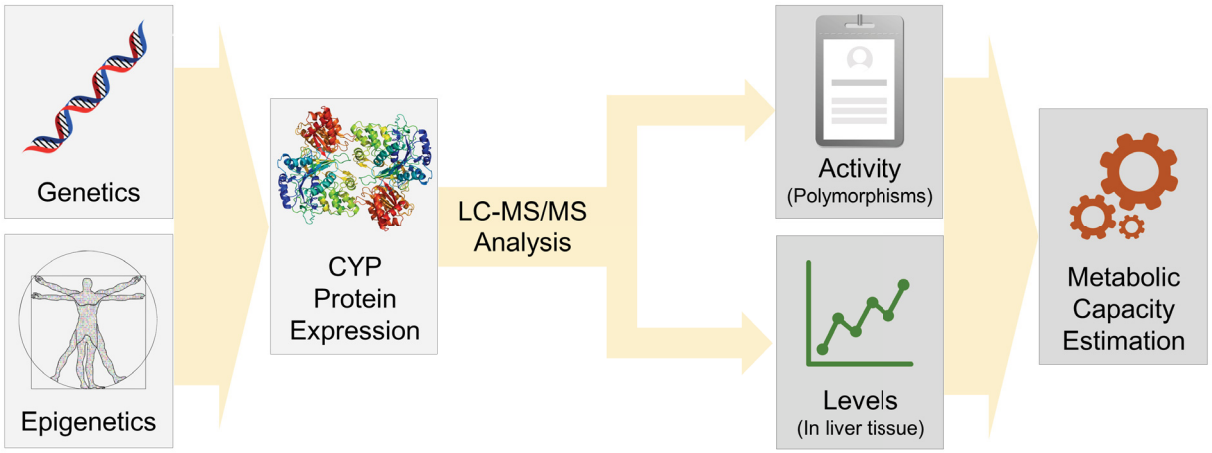


Figure 5.1.1: Graphical abstract for "Postmortem Estimation of Metabolic Capacity Through Cytochrome P450 Enzyme Characterization and Quantification — A Proof of Concept".

liver microsomes were isolated by mechanical homogenization followed by ultracentrifugation. The microsomal pellet was resuspended, reduced, alkylated and digested with trypsin. The resulting tryptic digest was purified by solid phase extraction, evaporated and reconstituted for analysis on a liquid chromatography system coupled to tandem mass spectrometry (LC-MS/MS). Method parameters were optimized by design of experiments to maximize sensitivity. Prevalent and deleterious CYP 2D6 and CYP 3A4 polymorphisms were monitored via analysis of the peptides expressing the genetic mutations. Simultaneously, the enzyme expression level in the tissue was estimated by quantification of selected CYP 2D6 and CYP 3A4 peptides. Combination of the activity and expression components should allow for a more accurate estimation of metabolic capacity. While application of this method on polymorphic human liver microsome samples show results matching the expected pattern, sample complexity remains a barrier to full validation and wide application in the forensic toxicology field.

5.1.1 Highlights

- A genomics approach to metabolizer type categorization suffers from poor correlation to the metabolic capacity;
- A proteomics approach has the potential to yield a better estimate of the metabolic capacity by combining both characterization of cytochrome P450 (CYP) enzyme mutations and CYP quantification in liver tissue;

- Human liver microsomes were isolated from liver homogeneate by ultracentrifugation, tryptically digested and analyzed by LC-MS/MS, the procedure was optimized for maximal sensitivity by design of experiments (DoE);
- Application of this method on recombinant CYPs and polymorphic human liver microsome samples resulted in the expected profile being observed (both in terms of peptides monitored and abundance);
- Liver sample complexity generates large matrix effects, sensitivity and instrumental issues, therefore sample treatment will require improvement before this method can be widely applied.

5.2 Introduction

Metabolic capacity can be defined as a figure representing the rate at which xenobiotics are transformed by metabolizing enzymes in the human body. This parameter is largely determined by the largest and most relevant enzyme family for forensic toxicology, cytochrome P450 (CYP) (e.g., CYP 2C9, CYP 2D6, CYP 3A4) [17, 184].

The relevance of possessing knowledge about metabolic capacity has been discussed by several authors. Most notably, the impact on markers of acute vs chronic consumption [9, 34–37] (e.g., enontiomeric ratio and metabolite to parent drug ratio) and the role in intoxication cases has been studied [10, 33, 34, 38, 39]. For example, individuals with a severely compromised metabolic capacity might undergo a fatal accumulation of a chronically dosed drug [23, 185]. Knowledge about the metabolic capacity should allow the forensic toxicologist to differentiate, for example, between medical error and voluntary or negligent overdose. It is therefore a relevant piece of information to shed light on the circumstances surrounding death.

The clinical approach to estimation of metabolic capacity, i.e., injection of innocuous probes and monitoring their pharmacokinetic curves [24], is impossible to apply with post-mortem cases. Forensic toxicologists have therefore relied on the traditional pharmacogenomics approach: genotyping [11, 23]. CYP encoding genes are analysed to detect the presence of one or more mutations. Changes in the DNA sequence can have different impacts on the CYP protein: prevention of its expression in the cell through various mechanisms (gene deleted, insertion of a stop codon, splicing defects) or modification of an amino acid in the enzyme, which can be deleterious to its activity or have no impact at all. These CYP enzyme

polymorphisms are compiled from publications by an international consortium, the Pharmacogene Variation Consortium [18]. An unmutated, wild type (WT) enzyme is labelled as CYP*1, while *2, *3, *4 and higher notations are attributed to mutated versions of the enzymes in the order they were discovered. The genotype of an individual, i.e., the polymorphic state of the CYP encoding gene(s) of the allele inherited from the mother and the allele inherited from the father, are used to classify individuals according to their metabolizer type (Figure 5.2.1a): poor, intermediate, extensive (normal) and ultrarapid metabolizers [23]. Poor and ultrarapid metabolizer types have been linked to fatal intoxications with morphine [40, 186], fluoxetine [41] and doxepin [42]. More broadly, ensembles of intoxication cases (n = 11 to 53) with methadone [10], codeine [34], citalopram [38], fentanyl [33] and oxycodone [39] have been evaluated under the pharmacogenomics lens. All studies found a possible genetic effect in a portion of the cases.

Nevertheless, genotyping for metabolizer type remains a sparingly used tool in forensic toxicology. This might be explained, in part, by the low genotype (metabolizer type) to phenotype (observed metabolic capacity) correlation [11]. Indeed, an individual classified as an extensive metabolizer might actually exhibit a metabolic capacity closer to a poor metabolizer (Figure 5.2.1b). This discrepancy can, at least in part, be linked to the wide range of CYP enzyme expression, i.e., the enzyme concentration in organ tissues, which can vary by a factor of 20 or more inside a metabolizer group [44, 65]. The transcription level gap can be caused by the presence of inducers or inhibitors [9, 16], age [19, 20], gender [19, 21, 22] or physiological state [187, 188] (e.g., liver disease). If proteins are the product of the DNA instructions set, their quantitative analysis combined with activity information should yield a metabolic capacity estimation closer to the phenotype than genotyping alone can. In the postmortem setting, liver tissue is freely available for protein analysis, a fact that can be taken advantage of.

Thus, our aim is to develop an LC-MS/MS proteomics method to estimate metabolic capacity postmortem via the characterization and quantification of CYP enzymes in liver tissue. Literature on CYP analysis reports a number of approaches to identify or quantify (Table 5.2.1) CYP enzymes in human liver collected during surgical interventions [44, 47, 65–67], purchased pooled human liver microsomes [47, 65], mouse [68, 69] or pig liver [70]. Concurrent work on postmortem liver samples was also carried out by Hansen *et al.* [189]. A large number of CYP subfamilies were quantitated absolutely [44, 47, 65–67, 189] or relatively [44, 68–70] (i.e., by comparing normal and induced state). Different preparation (e.g.,

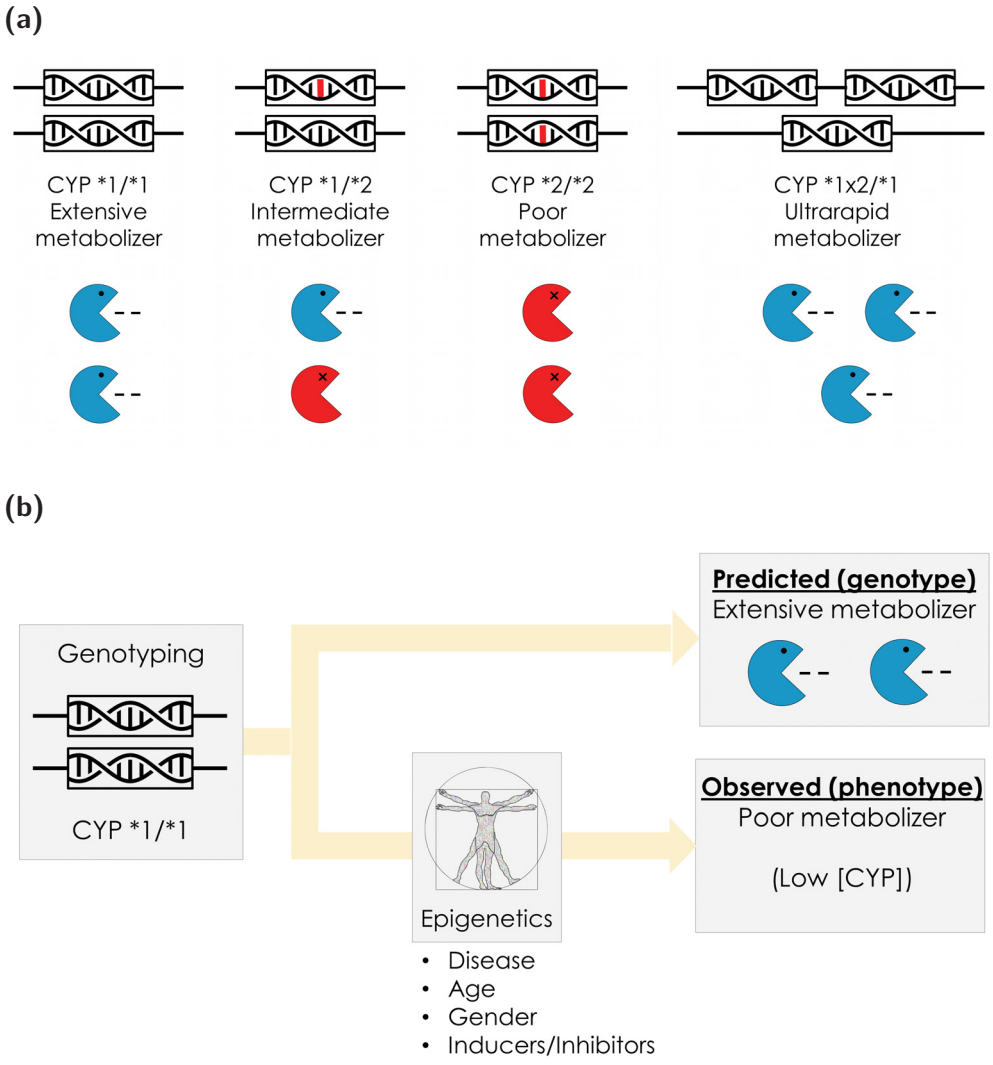


Figure 5.2.1: Classification of genotype, phenotype and metabolizer type. **(a)** Categorization of individuals by metabolizer type according to genotype (black: WT CYP encoding gene, red: deleterious mutation in the CYP encoding gene) and the predicted impact on CYP enzyme population (blue: functional enzymes, red: non-functional or absent enzyme). **(b)** Distinction between predicted metabolic capacity according to genotype and actual observed phenotype influenced by epigenetic factors.

gel electrophoresis [44, 66, 70], in solution [47, 65]) and quantification (e.g., absolute quantification with isotope-labelled internal standards (AQUA) [44, 47, 65, 67, 190], exponentially modified protein abundance index (emPAI) [70], isotope coded affinity tags (ICAT) [68] and ^{18}O labelling [69]) techniques have been used. Based on this body of literature, the only remaining step seemed to be minor adjustments and simplifications to match forensic toxicology samples and standard practices. In the end, transfer to the postmortem domain proved much more difficult than anticipated. Notwithstanding the complications that re-

main to be addressed, we present a proof of concept method for proteomic quantification and characterization of CYP enzymes in postmortem cases to estimate metabolic capacity. This proof of concept was developed for CYP 2D6 and CYP 3A4 enzymes, given their importance in the metabolism of major xenobiotics routinely encountered in forensic toxicology.

Table 5.2.1: Comparison of published CYP absolute quantification techniques.

1. Wang et al. [47]
   - Quantification of CYPs 3A4 and 3A5
   - Liver tissue preparation: N/A, only microsomes
   - Microsomes preparation: SDS-PAGE gel compared to in-solution digestion
   - Mass spectrometry instrument: unit resolution, API 4000
   - Chromatography length: 5.2 minutes
   - Number of peptides and transitions: 1 peptide per CYP, 1 transition per peptide
   - Quantification approach: external calibration with peptides, SIL-IS
2. Langenfeld et al. [67]
   - Quantification of CYP 2D6
   - Liver tissue preparation: homogeneization, differential/ultracentrifugation
   - Microsomes preparation: gel electrophoresis
   - Mass spectrometry instrument: unit resolution
   - Chromatography length: 40 minutes
   - Number of peptides and transitions: 4 peptides per CYP, 3 transitions per peptide
   - Quantification approach: external calibration with recombinant CYPs, SIL-IS
3. Seibert et al. [44]
   - Quantification of CYPs 1A2 and 2E1
   - Liver tissue preparation: grinding frozen tissue and differential centrifugation
   - Microsomes preparation: 1D-gel electrophoresis
   - Mass spectrometry instrument: unit resolution LCQ^{duo}
   - Chromatography length: 58 minutes
   - Number of peptides and transitions: 1 peptide per CYP
   - Quantification approach: internal calibration with SIL-IS
4. Kawakami et al. [65]
   - Quantification of CYPs 1A2, 2A6, 2B6, 2C8, 2C9, 2C19, 2D6, 2E1, 3A4/43, 3A5, 3A43
   - Liver tissue preparation: homogeneization in Potter-Elvehjem, differential centrifugation
   - Microsomes preparation: in-solution digestion
   - Mass spectrometry instrument: unit resolution, API 5000

- Chromatography length: 50 minutes
- Number of peptides and transitions: 1 peptide per CYP, 4 transitions per peptide
- Quantification approach: external calibration with peptides, SIL-IS

5. Williamson et al. [190]
   - Quantification of CYPs 1A2, 2B6 and 3A4
   - Liver tissue preparation: N/A, hepatocytes or microsomes only
   - Microsomes preparation: in-solution digestion
   - Mass spectrometry instrument: unit resolution, AB SCIEX 4000 QTRAP
   - Chromatography length: 15 minutes
   - Number of peptides and transitions: 3 peptides per CYP, 3 transitions per peptide
   - Quantification approach: internal calibration with SIL-IS
6. Hansen et al. [189]
   - Quantification of CYPs 1A2 and 3A4
   - Liver tissue preparation: motor-driven Potter-Elvehjem, differential centrifugation
   - Microsomes preparation: in-solution digestion
   - Mass spectrometry instrument: unit resolution, Waters Xevo TQ-S
   - Chromatography length: 16 minutes
   - Number of peptides and transitions: 2 peptides, 2 transitions per peptide
   - Quantification approach: internal calibration with SIL-IS

5.3 Material and methods

5.3.1 Human liver microsome isolation

The human liver microsomes (HLM) isolation process is illustrated in Figure 5.3.1. A 1 g piece of washed postmortem liver tissue was cut into small pieces of $\approx 3\ mm^3$ using a scalpel and put in a 50 mL polypropylene conical tube (Sarstedt, 62.547.205, Nümbrecht, Germany). 50 μL of protease inhibitor cocktail (Sigma, P8340-5mL, Oakville, Ontario, Canada) as well as 2 mL of deionized water was added to the liver tissue, and the mixture incubated at 4 °C for 24 hours. This hypotonic treatment is designed to generate cell lysis via differential salt concentration [191, 192]. While faster alternatives could be implemented, for different logistical reasons this long lysis did not hinder the process as a whole and was chosen for its simplicity. Following incubation, 50 μL of Triton X-100 (Sigma, T9284-100mL, Oakville, Ontario, Canada), 750 μL of 1 M NaCl (Sigma, S9888-500G, Oakville, Ontario, Canada),

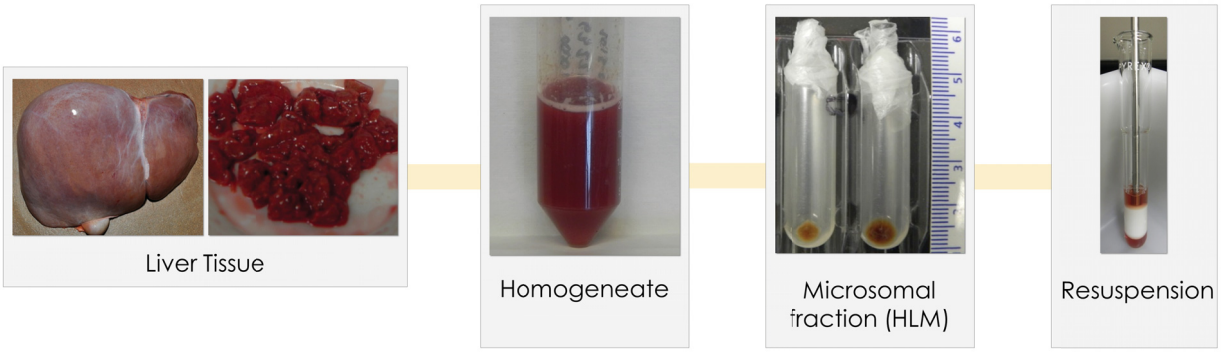


Figure 5.3.1: Human Liver Microsome (HLM) preparation process.

250 μL of 1 M, pH 7.4, Tris buffer (Sigma, T5941-500G (hydrochloride) and T6066-100G (base), Oakville, Ontario, Canada) and 1.9 mL of deionized water was added to the sample which was then vortexed. A Brinkmann Polytron® Homogenizer (Kinematica AG, Luzern, Switzerland) mechanical blender was then used to homogenize the sample completely. The S9 cellular fraction (also called post-mitochondrial fraction) was isolated by centrifuging the polypropylene tube and its content at 10 000 $\times$ g at 4 °C for 20 minutes (centrifuge 5810R, Eppendorf, Hamburg, Germany). The supernatant (S9 fraction) was further centrifuged for 60 minutes at 100 000 $\times$ g at 4 °C to isolate the microsomal pellet (Optima L-100 XP ultra-centrifuge, Beckman Coulter, Brea, California, USA). The supernatant was removed using a Pasteur pipette and discarded. The microsomal fraction was then lyophilized (LabConco, Kansas City, Missouri, USA) and homogenized into a powder.

The pooled human liver microsomes (XenoTech, H0630, Lenexa, Kansas, United States) and the genotyped human liver microsomes (XenoTech, H2D6.MA, Lenexa, Kansas, United States) extracts were purchased and used as is with no further isolation.

5.3.2 Sample preparation and tryptic digestion

Stable isotope-labelled (SIL) internal standard (IS) solutions

All isotopically labelled (heavy) peptides were purchased from Life Technologies Inc., Burlington, Ontario, Canada and dissolved to a final concentration of 1.125 $pmol/\mu L$ using the solvent specified in Table 5.3.1.

Table 5.3.1: Internal Standards (IS) and their solvent.

Identifier	Peptide sequence[1]	Solvent[2]
IS_2D6_Q1	$S_{116}QGVFLA(\mathbf{R})_{123}$	MPA
IS_2D6_Q2	$F_{134}SVSTL(\mathbf{R})_{140}$	MPA
IS_2D6_Q3	$A_{246}FLTQLDELLTEH(\mathbf{R})_{259}$	DMSO
IS_2D6_Q4	$V_{331}QQEIDDVIGQV(\mathbf{R})_{343}$	MPA
IS_2D6_Q5	$D_{381}IEVQGF(\mathbf{R})_{388}$	MPA
IS_2D6*4/10_WT1	$Y_{33}PPGPLPLPGLGNLLHVDFQNTPYCFDQL(\mathbf{R})_{62}$	DMSO
IS_2D6*4/17_WT2	$E_{89}ALVTHGEDTADRPPVPITQILGFGP(\mathbf{R})_{115}$	DMSO
IS_2D6*9_WT	$D_{270}LTEAFLAEME(\mathbf{K})_{281}$	MPA
IS_3A4_Q1	$L_{36}GIPGPTPLPFLGNILSYH(\mathbf{K})_{55}$	MPB
IS_3A4_Q2	$V_{71}WGFYDGQQPVLAITDPDMI(\mathbf{K})_{91}$	DMSO
IS_3A4_Q3	$E_{244}VTNFL(\mathbf{R})_{250}$	MPA
IS_3A4_Q4	$L_{366}FPIAM(\mathbf{R})_{372}$	MPA
IS_3A4_Q5	$G_{391}VVVMIPSYALH(\mathbf{R})_{403}$	MPA
IS_3A4*8_WT	$S_{131}LLSPTFTSG(\mathbf{K})_{141}$	MPA

[1]A peptide sequence noted "$S_{116}QGVFLA(\mathbf{R})_{123}$" indicates that the first amino acid of the peptide, serine (S), is the 116^{th} amino acid in the sequence of the protein (here, CYP 2D6). The last amino acid of the peptide, arginine (R) is the 123^{rd} amino acid in the protein sequence. Standard one letter codes for amino acids are used. Amino acids in parenthesis and bold, e.g., **(R)**, indicates that the heavy amino acid (+10 Da for arginine, +8 Da for lysine) was used for synthesis.

[2]MPA: mobile phase A, 0.1% formic acid in water; MPB: mobile phase B, 0.1% formic acid in acetonitrile; DMSO: dimethylsulfoxide (Sigma, 472301-100ML, Oakville, Ontario, Canada). Note that MPA was used as the first solvent, as suggested by the peptide manufacturer, followed by MPB failing proper solubilization, and finally DMSO as a last recourse. No comparative solubility test was carried out, and therefore some peptides might only dissolve partially in MPA. Further experiments could be carried out to select an optimal solvent for each peptide.

CYP calibration standards preparation

Serial dilutions (1:10, 1:100, 1:1000) of recombinant CYP 2D6 and CYP 3A4 (1 $nmol/mL$, XenoTech, CYP/EZ007 and CYP/EZ002, Lenexa, Kansas, United States) were prepared in 50 mM NH_4HCO_3 (Sigma, A6141-500G, Oakville, Ontario). These solutions were used to

prepare calibration standards with concentrations of 2.5, 5.0, 25, 50, 125, 250, 500, 1250, 2000, 2500 $pmol/mL$, as well as quality control samples of 37.5, 1000 and 1750 $pmol/mL$. Volume was brought to 150 μL using 50 mM NH_4HCO_3 buffer. 10 μL of each SIL-IS solution (Table 5.3.1) was added to each sample, followed by 205 μL of 1.2 M thiourea/4.8 M urea (Sigma, U5128-100G (urea) and T7875-5G (thiourea), Oakville, Ontario), and 205 μL of 1.8% CHAPS (Sigma, C5070-1G, Oakville, Ontario). Urea and thiourea are chaotropic agents, molecules that disrupt the hydrogen bonding network, thus destabilizing the three-dimensional structure of proteins. Samples were incubated at 45 °C in a water bath for 30 minutes to allow protein denaturation (unfolding) under the combined effect of the chaotropic agents, detergent and heat.

Preparation of a microsomal fraction isolated from postmortem liver

To all the powdered microsomal fraction transferred quantitatively from the ultracentrifugation vessel, 585 μL of 1.2 M thiourea/4.8 M urea, 200 μL of 1.8% CHAPS, 10 μL of each IS solution and 2440 μL of 50 mM NH_4HCO_3 buffer were added. The microsomal fraction was resuspended by performing approximately 10 passes of the pestle in a Potter-Elvehjem apparatus. The homogeneate was transferred to a 5 mL LoBind Eppendorf tube (Eppendorf, 0030108302, Hamburg, Germany). Samples were denatured in a 45 °C water bath for 30 minutes.

Preparation of purchased human liver microsomes (HLMs)

In a 5 mL LoBind Eppendorf tube, 100 μL of purchased HLMs, 585 μL of 1.2 M thiourea/4.8 M urea, 200 μL of 1.8% CHAPS, 10 μL of each IS solution and 2440 μL of 50 mM NH_4HCO_3 buffer were combined, vortexed and denatured in a 45 °C water bath for 30 minutes.

Reduction, alkylation and digestion

13 μL of 125 mM dithiothreitol (DTT) (prepared from 1 M DL-dithiothreitol solution, Sigma, 43816-10ML, Oakville, Ontario, Canada) was added to each sample, vortexed and incubated at room temperature for 30 minutes to reduce the disulfide bonds (bridges) in the proteins [193]. These bonds occur between two cysteine (C) side-chains and contribute to the overall three-dimensional structure of the proteins [194].

To prevent reformation of the disulfide bridges, the sulfur atoms were capped via an alkylation process [195]. Following the addition of 38 μL of 50 mM iodoacetamide (Sigma, A3221-1VL, Oakville, Ontario, Canada), each sample was vortexed and incubated at room temperature for 30 minutes. The incubation was carried out in the dark due to the light sensitivity of iodoacetamide. Excess alkylating agent was quenched by adding 15 μL of 250 mM DTT to every sample and vortexing.

Digestion with trypsin was initiated by adding 10 μg of trypsin (Promega, V511A, Madison, Wisconsin, USA) reconstituted in 50 mM acetic acid (Promega, V542A, Madison, Wisconsin, USA) to each sample and vortexing. The protein:protease ratio of the mid-level QC was 5.6. Samples were incubated for 5 hours at 37 °C with vortexing every 30 minutes. At the completion of the digestion step, trypsin was quenched by addition of 50 μL of formic acid (Fisher, A117-50, Ottawa, Ontario, Canada) and vortexing.

5.3.3 Tryptic digest clean-up

Specimens were centrifuged at 1258 $\times$ g for 10 minutes to pellet the suspended material (e.g., protein aggregates and particulates). Solid phase extraction (SPE) clean-up was carried out on centrifuge. SPE cartridges (Oasis HLB 3 cc, Waters, WAT094226, Mississauga, Ontario, Canada) were conditioned with 1 mL of acetonitrile (EMD Millipore, AX0156-1, Etobicoke, Ontario, Canada) followed by 1 mL of ammonium formate/formic acid buffer (50 mM, pH 3.0) (ammonium formate: Sigma, 70221, Oakville, Ontario, Canada; formic acid: Fisher, A117-50, Ottawa, Ontario, Canada). Cartridges were loaded with a mixture of 1000 μL of ammonium formate/formic acid buffer (50 mM, pH 3.0) and 1000 μL of sample; this loading step was performed 3 times to load a total of 3 mL of specimen. Cartridges were rinsed with two 1 mL volumes of ammonium formate/formic acid buffer (50 mM, pH 3.0). Retained material was eluted by using two 500 μL volumes of 80:20 v:v acetonitrile:buffer. Eluant was evaporated (Xcel Vap, Horizon Technologies, Salem, New Hampshire, USA) to dryness under N_2 at 30 °C and reconstituted in 200 μL of 80:20 0.1% formic acid in water:0.1% formic acid in acetonitrile. Extracts were stored in deactivated glass inserts (6 mm diameter, 300 μL, conical bottom with spring, deactivated, Phenomenex, AR0-4626-12, Torrance, California, USA).

5.3.4 LC-MS/MS analysis

A 20 μL aliquot of the extract was injected on an Advance Bio Peptide Mapping column (2.1 $\times$ 150 mm $\times$ 2.7 μm, 120 Å pores, 653750-902, Agilent, Mississauga, Ontario, Canada) for a 45 minute separation. An Agilent 1200 HPLC (Mississauga, Ontario, Canada) coupled to a Sciex 5500 QTrap (Concord, Ontario, Canada) was used to perform this analysis. All instrumental parameters (mass transitions, potentials, gas settings, chromatography) are available in Appendix G.1. Table 5.3.2 identifies the peptides monitored, in addition to the internal standards presented in Table 5.3.1. A peptide was considered as identified when all transitions had a signal-to-noise ratio $> 3:1$, ion ratios within $\pm 30\%$ of the reference samples, and a retention time within 0.1 min of the SIL-IS.

5.3.5 *In vitro* postmortem decomposition

The Research Ethics and Compliance Unit from Concordia's Office of Research approved the secondary use of these liver tissues. A ≈ 10 g block of liver tissue collected during a full autopsy was separated in 10 pieces of ≈ 1 g each. Each piece of liver was stored in 15 mL polypropylene tubes (Sarstedt, 62.554.002, Nümbrecht, Germany). One sample was frozen immediately at $-20\,^{\circ}\mathrm{C}$ as a reference. Other samples were left at room temperature for a duration of 1, 2, 3, 5, 7, 10, 14, 21 and 28 days, respectively, and frozen at $-20\,^{\circ}\mathrm{C}$ pending analysis. A $-20\,^{\circ}\mathrm{C}$ freezer was favoured over a $-80\,^{\circ}\mathrm{C}$ freezer because it is commonplace in forensic toxicology laboratories and achieves the goal of halting bacteriological degradation; preservation of enzyme functionality (activity) is not required in this case. Microsome isolation, sample preparation, tryptic digestion and sample clean-up were performed as described above.

Table 5.3.2: Monitored quantotypic and mutation site peptides.

Identifier[1]	Mutation[2]	Sequence
2D6_Q1		S_{116}QGVFLAR$_{123}$
2D6_Q2		F_{134}SVSTLR$_{140}$
2D6_Q3		A_{246}FLTQLDELLTEHR$_{259}$
2D6_Q4		V_{331}QQEIDDVIGQVR$_{343}$
2D6_Q5		D_{381}IEVQGFR$_{388}$
2D6*10_WT		Y_{33}PPGPLPLPGLGNLLHVDFQNTPYCFDQLR$_{62}$
2D6*10_MUT	P34S	Y_{33}SPGPLPLPGLGNLLHVDFQNTPYCFDQLR$_{62}$
2D6*17_WT		E_{89}ALVTHGEDTADRPPVPITQILGFGPR$_{115}$
2D6*17_MUT	T107I	E_{89}ALVTHGEDTADRPPVPIIQILGFGPR$_{115}$
2D6*9_WT		D_{270}LTEAFLAEMEK$_{281}$
2D6*9_MUT	K281del	D_{270}LTEAFLAEMEAK$_{283}$
3A4_Q1		L_{36}GIPGPTPLPFLGNILSYHK$_{55}$
3A4_Q2		V_{71}WGFYDGQQPVLAITDPDMIK$_{91}$
3A4_Q3		E_{244}VTNFLR$_{250}$
3A4_Q4		L_{366}FPIAMR$_{372}$
3A4_Q5		G_{391}VVVMIPSYALHR$_{403}$
3A4*8_WT		S_{131}LLSPTFTSGK$_{141}$
3A4*8_MUT	R130Q	L_{129}QSLLSPTFTSGK$_{141}$
3A4*11_WT		A_{343}PPTYDTVLQMEYLDMVVNETLR$_{365}$
3A4*11_MUT	T363M	A_{343}PPTYDTVLQMEYLDMVVNEMLR$_{365}$
3A4*13_WT		F_{414}LPER$_{418}$
3A4*13_MUT	P416L	F_{414}LLER$_{418}$

[1]Peptides identified as "XXX_QX" are quantotypic peptides, used for quantification purposes. Peptides identified as "XXX*YY_WT" or "XXX*YY_MUT" are mutation site peptides, used to identify the presence of a genetic polymorphism affecting the enzyme activity. "*YY" identifies the mutation studied, while "_WT" indicates the wild-type or normal version of the peptide, and "_MUT" indicates the mutated version of the peptide, where one amino acid has changed as a result of a change in the genetic code.

[2]This column identifies the precise impact of the genetic mutation on the amino acid sequence analyzed. For example, "P34S" indicates that the 34th amino acid in the protein is changed from a proline (P) to a serine (S) due to the genetic mutation (here, *10).

5.4 Results and discussion

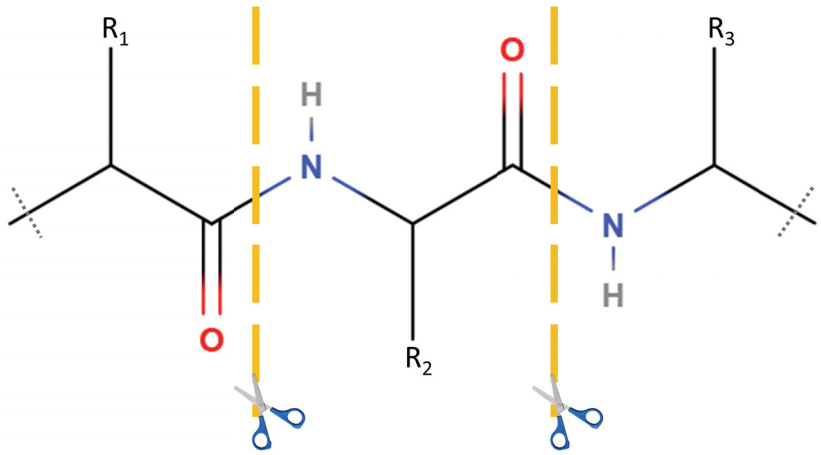
5.4.1 Monitored peptides

The CYP 2D6 and 3A4 enzymes targeted for analysis are proteins, i.e., very long strings of amino acids (497 for CYP 2D6 and 503 for CYP 3A4) [196, 197]. Proteins can be analysed whole (top-down) or cut into smaller chains of amino acids called peptides (bottom-up) [54]. In the context of the method discussed in this paper, since information is required not only on the enzyme as a whole (quantification) but also on composition at specific locations (characterization of the mutations), bottom-up analysis subsequent to a protein digestion seemed the logical choice.

Protein digestion can be carried out by proteases such as trypsin, chymotrypsin, pepsin, thermolysin and others, which cut the protein at the peptide bond (Figure 5.4.1a) after specific amino acids [55]. Trypsin, the most popular digestion enzyme in proteomics analysis, will cut after lysine and arginine amino acids, except if they are followed by a proline amino acid due to steric hindrance [62]. The structure, and mass, of peptides obtained following digestion can therefore be predicted *in silico*. This prediction was carried out for tryptic digestion of CYP 2D6 and 3A4 (Figure 5.4.1b and 5.4.1c) using the ExPASy PeptideCutter (`https://web.expasy.org/peptide_cutter/`).

Amongst the tryptic peptides, six peptides unaffected by polymorphisms for each CYP were selected to act as quantotypic peptides, i.e., to be used for quantification purposes (identified as "_Q1" to "_Q6" in Table 5.3.2). These peptides were selected by adhering as much as possible to the rules (Table 5.4.1) laid out by Ludwig *et al.* [62]. These guidelines ensure that peptides selected are detectable by mass spectrometry and are minimally affected by incomplete digestion or chemical modifications (oxidation, deamidation, incomplete alkylation). Because of the high degree of homology between the 57 known human CYP enzymes [185], particular care must be taken to select peptides with a unique amino acid sequence. Otherwise, if for example a peptide common to CYP 3A4 and CYP 3A7 is selected, the quantification result would represent the sum of the concentrations of the two enzymes. The National Center for Biotechnology Information's Basic Local Alignment Search Tool (`https://blast.ncbi.nlm.nih.gov/Blast.cgi`) can be used to search the human proteome for the intended quantotypic peptide and confirm its uniqueness, at least in liver tissue. In the present case, uniqueness against the full human proteome was confirmed

(a)



(b)

CYP 2D6 *9, *10, *17

MGLEALVPLAVIVAIFLLLVDLMHRRQRWAARYPPGPLPLPGLGNLLHVDFQN
TPYCFDQLRRRFGDVFSLQLAWTPVVVLNGLAAVREALVTHGEDTADRPPVPI
TQILGFGPRSQGVFLARYGPAWREQRRFSVSTLRNLGLGKKSLEQWVTEEAAC
LCAAFANHSGRPFRPNGLLDKAVSNVIASLTCGRRFEYDDPRFLRLLDLAQEG
LKEESGFLREVLNAVPVLLHIPALAGKVLRFQKAFLTQLDELLTEHRMTWDPA
QPPRDLTEAFLAEMEKAKGNPESSFNDENLRIVVADLFSAGMVTTSTTLAWGL
LLMILHPDVQRRVQQEIDDVIGQVRRPEMGDQAHMPYTTAVIHEVQRFGDIVP
LGVTHMTSRDIEVQGFRIPKGTTLITNLSSVLKDEAVWEKPFRFHPEHFLDAQ
GHFVKPEAFLPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPTGQPRPSH
HGVFAFLVSPSPYELCAVPR

Figure 5.4.1: Tryptic digestion of a generic peptide bond, CYP 2D6 and CYP 3A4. **(a)** Proteins are cleaved by trypsin and other proteases at the peptide bond. R_1 through R_3 are the amino acid side chains. **(b)** Tryptic peptides produced by the digestion process of CYP 2D6; bold underlined red letters are mutation sites of 2D6*9, *10 and *17. Peptides are coloured in alternating blue and black to facilitate differentiation.

for the selected peptides.

Adhering to the gold standard in quantitative proteomics, five quantotypic peptides

(c)

CYP 3A4 *8, *11, *13

MALIPDLAMETWLLLAVSLVLLYLYGTHSHGLFKKLGIPGPTPLPFLGNILSY
HKGFCMFDMECHKKYGKVWGFYDGQQPVLAITDPDMIKTVLVKECYSVFTNRR
PFGPVGFMKSAISIAEDEEWKRLRSLLSPTFTSGKLKEMVPIIAQYGDVLVRN
LRREAETGKPVTLKDVFGAYSMDVITSTSFGVNIDSLNNPQDPFVENTKKLLR
FDFLDPFFLSITVFPFLIPILEVLNICVFPREVTNFLRKSVKRMKESRLEDTQ
KHRVDFLQLMIDSQNSKETESHKALSDLELVAQSIIFIFAGYETTSSVLSFIM
YELATHPDVQQKLQEEIDAVLPNKAPPTYDTVLQMEYLDMVVNETLRLFPIAM
RLERVCKKDVEINGMFIPKGVVVMIPSYALHRDPKYWTEPEKFLPERFSKKNK
DNIDPYIYTPFGSGPRNCIGMRFALMNMKLALIRVLQNFSFKPCKETQIPLKL
SLGGLLQPEKPVVLKVESRDGTVSGA

Figure 5.4.1: Tryptic digestion of a generic peptide bond, CYP 2D6 and CYP 3A4. **(c)** Tryptic peptides produced by the digestion process of CYP 3A4; bold underlined red letters are mutation sites of 3A4*8, *11 and *13.

with 5 selected mass transitions were monitored [62] (Figure 5.4.2). This ensured that interferences weren't misidentified as the analytical target, and that quantification could be considered reliable, since all peptides/transitions should yield the same result within their respective uncertainties of measurement. Unless otherwise specified, the derived concentration is the average over the first transition of each of the quantitative peptides. The first transition was selected for this purpose due to its sensitivity and minimum amount of interferences. Transitions and associated potentials (DP, EP, CE, CXP) can be predicted for a given peptide with a high degree of reliability using the Skyline software (https://skyline.ms/wiki/home/software/Skyline/).

In addition to the quantotypic peptides, tryptic peptides expressing a deleterious mutation, present in 1% or more of a given sub-population [198–200], were also monitored, both in their wild-type and mutated format (Figure 5.4.2). This allowed detection of the mutations affecting enzyme activity in an individual.

As can be seen in Figure 5.4.2, all quantotypic and wild-type marker peptides can be

Table 5.4.1: Rules for the selection of CYP quantotypic peptides, adapted from [62].

- **Unique amino acid sequence in the proteome:** peptide sequence should not be found in other human protein; or at least, in no other protein present in the target tissue (here, the liver);
- **MS detectability:** obtained from empirical data or prediction tools;
- **Peptide length:** ideally between 8 and 25 amino acids;
- **Optimal mass range:** +2 and +3 charged peptides should have a m/z in the operating range covered by the instrument;
- **Complete proteolysis:** For trypsin, avoid a) Dibasic cleavage sites: -K-K-, -R-R-, -R-K-, -K-R-; b) Acidic residues (D and E) close to the cleavage site; c) N-terminal prolines: -K-P-, -R-P-;
- **No biological or chemical modifications:** Avoid a) Glycosylation motif: -N-X-S/T-; b) N- and C-terminal peptides susceptible to degradation; c) M and W, susceptible to oxidation; d) C, susceptible to incomplete alkylation; e) -N-G- or -Q-G-, susceptible to deamidation.

detected in recombinant CYP 2D6 and 3A4 enzyme samples. In addition, synthetic versions of the mutated peptides were detected in solution. However, the sensitivity is low for some of the marker peptides, due to their high mass ($>$ 25 amino acids length) and hydrophobic nature.

5.4.2 Optimization of the method parameters

Preparation and digestion of CYP enzymes proved to be more difficult than for, say, most plasma proteins due to their highly hydrophobic nature. The CYPs are membrane proteins, nearly all bound in the cells' endoplasmic reticulum [185], thus by definition hydrophobic. Trypsin requires access to the protein to be effective, therefore CYPs must be unfolded (denatured) and solubilized [62] prior to analysis. Performing absolute quantification of proteins requires a method with adequate sensitivity, precision, accuracy and reproducibility. All of these figures of merit will vary as a function of the protocol for liver homogeneization, cell fractionation, resuspension, denaturation, alkylation, reduction, digestion, clean-up, chro-

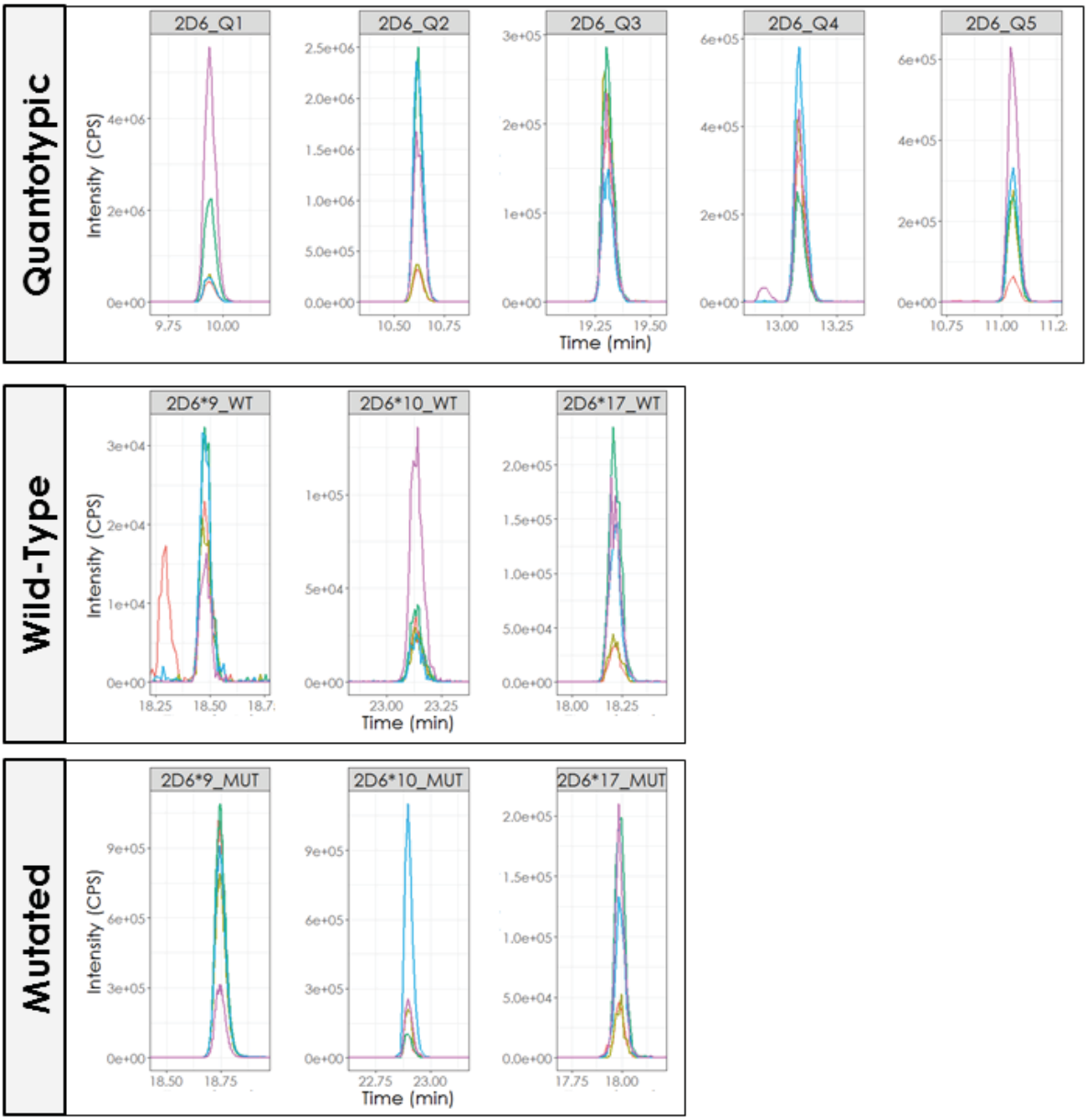


Figure 5.4.2: Monitored quantotypic and mutation bearing peptides for CYP 2D6, with five transitions (fragments) per peptide. Quantotypic and wild-type peptides displayed were analyzed in a recombinant CYP digest (1 mg/mL), whereas mutated peptides were synthetic peptides dissolved in solution (2D6*9_MUT: 26 $pmol/\mu L$, 2D6*10_MUT: 145 $pmol/\mu L$, 2D6*17_MUT: 159 $pmol/\mu L$).

matography and mass spectrometric detection. Furthermore, each of these steps contains a number of variables, or factors that can take several levels. Optimization of this kind of multi-variable problem is therefore a particularly challenging task.

Design of Experiments (DoE) is a methodology particularly well suited to efficiently screen and optimize a large number of experimental factors. In this approach, the level (or

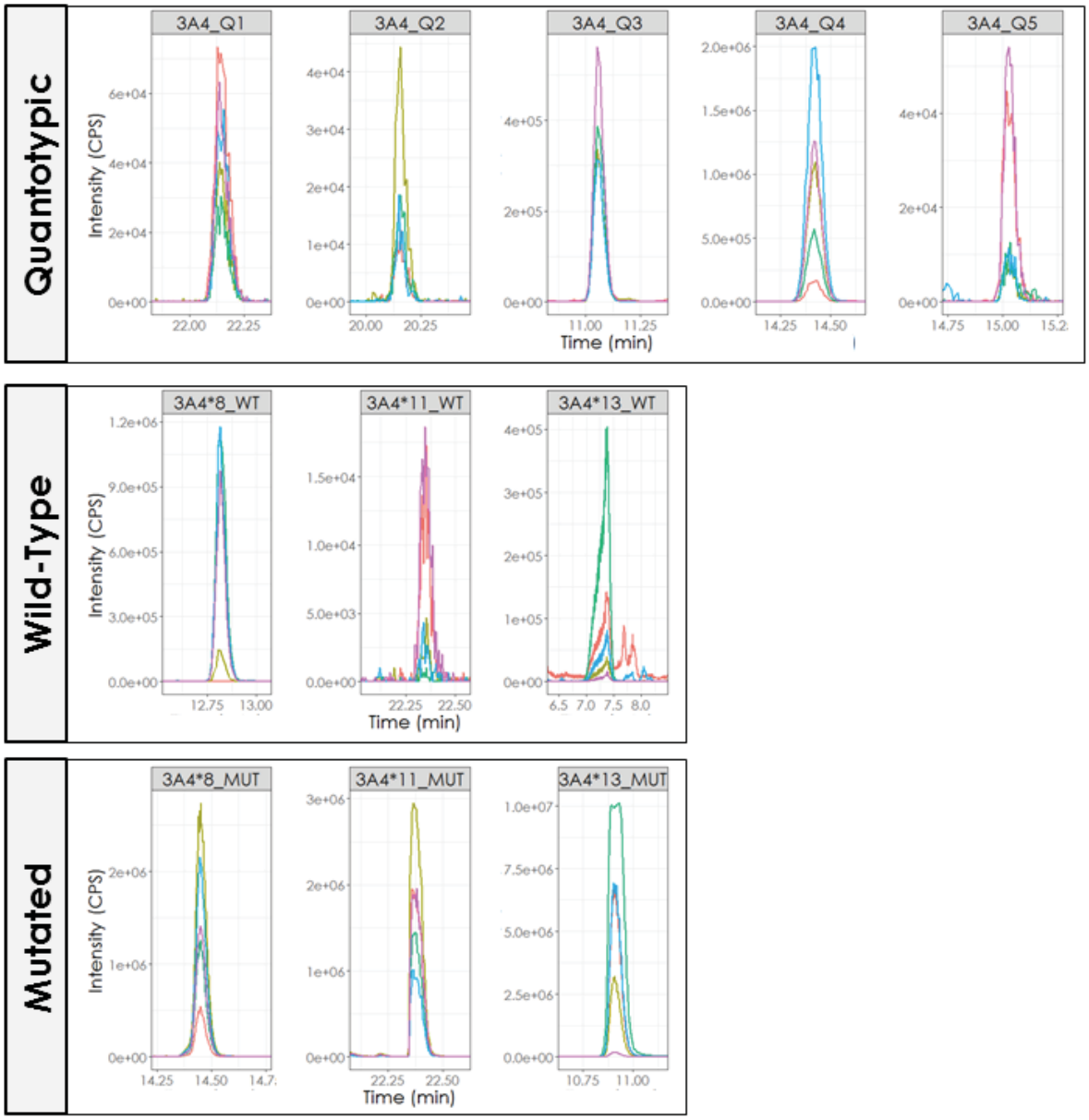


Figure 5.4.3: Monitored quantotypic and mutation bearing peptides for CYP 3A4, with five transitions (fragments) per peptide. Quantotypic and wild-type peptides displayed were analyzed in a recombinant CYP digest (1 mg/mL), whereas mutated peptides were synthetic peptides dissolved in solution (3A4*8_MUT: 3 $pmol/\mu L$, 3A4*11_MUT: 125 $pmol/\mu L$, 3A4*13_MUT: 7 $pmol/\mu L$).

value) of several factors (e.g., time, temperature, identity of a reagent) are varied in systematic patterns to explore all the experimental space. It is a more complete version of the latin squares method. Mathematical modelling, such as the response surface method, allows the most significant factors and their optimal value (that maximizes the studied response) to be identified. This framework also allows the optimal levels of factors to be found more accurately than the traditional "one factor at a time" approach while at the same time reducing

the number of experiments required to reach this conclusion. The interested reader can find a more thorough introduction to principles and practice in Eriksson *et al.* [201].

The method presented here was optimized for maximal peak area in order to maximize detection of peptides with low response (chiefly the mutation bearing peptides). The overall method was broken down into distinct experiments, each with a manageable number of factors. Microsomal fraction resuspension and denaturation were optimized with regards to total protein concentration, detergent (CHAPS, C7BzO, Brij® C10, SB3-10, ASB-14, octyl β-D-glucopyranoside, n-dodecyl-β-D-maltoside, OTG, TritonTM X-100, C13E10, SDS, RapiGest, Protease Max), temperature, chaotropic agent (guanidine hydrochloride, urea, urea/thiourea), denaturating organic solvent (none, methanol, acetonitrile), reduction agent (dithiothreitol, tris-2-carboxyethyl-phosphine-hydrochloride) and vortexing time. Reduction and alkylation were optimized with regards to the reducing agent (dithiothreitol, 2-mercaptoethanol, tributylphosphine, tris-2-carboxyethyl-phosphine-hydrochloride) and its concentration (0.5 to 10 mM), the alkylating agent concentration (2× to 4× the reducing agent concentration) and the quenching agent concentration (none to 4× the alkylating agent). The digestion was optimized with regards to the total time (1 h to 12 h), vortexing frequency, the buffer (Tris pH 7.8 or ammonium bicarbonate pH 7.8) and its concentration (25 to 50 mM) (Appendix G.3).

In most cases, derivation of highly reliable mathematical models proved to be difficult due to peak area variability in the replicates and the presence of several qualitative multi-level factors. An example is presented in Appendix G.2. Nevertheless, exploration of the experimental space revealed a number of high impact factors. The use of a Potter-Elvehjem apparatus and urea/thiourea mixture as chaotropic agent provided a signal increase of $> 300\%$. On the other hand, the use of TCEP as a reducing agent or organic solvent for denaturation hindered target detection. The method presented in Section 5.3 resulted from this optimization process.

Additional aspects of the procedure such as homogenization, cell fractionation (microsomes isolation) and sample clean-up could benefit from further optimization. Experiments to maximize response as a function of ultracentrifugation speed, time and composition of the centrifugation medium were planned, but not pursued in light of the sample purity issues discussed in Section 5.4.4. Once sample purity has been improved, method parameters should be (re-)optimized for relevant targets (high sensitivity, low matrix effects, precision, etc.).

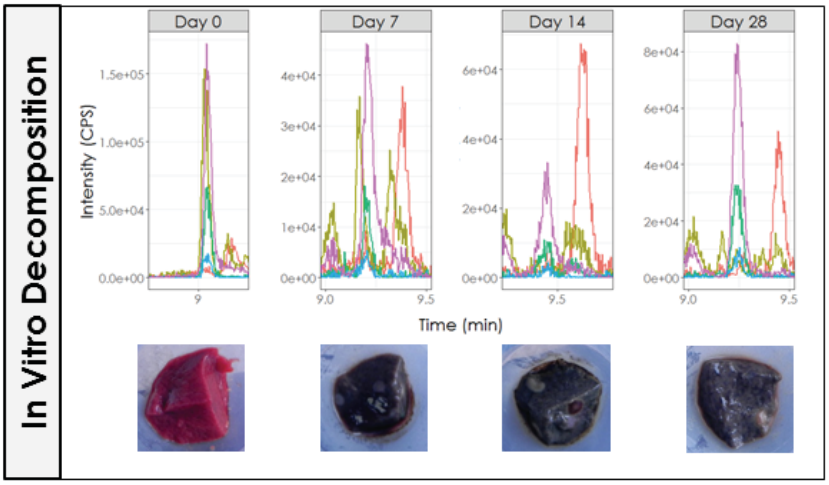
5.4.3 Postmortem degradation

One important and obvious question with regards to postmortem estimation of metabolic capacity using a proteomics approach is whether degradation of CYP proteins occurs postmortem, and if so, how it will affect the estimation method. Degradation of proteins can indeed occur extremely quickly postmortem; one well known example is insulin, with its disappearance only a few hours after death [202, 203]. But other proteins degrade over a much longer time frame, to the extent that they can still be identified in mummified corpses thousands of years old [204, 205]. Postmortem protein degradation involves, amongst others, hydrolysis of peptide bonds, which can involve proteolytic enzymes [206]. In the case of CYP enzymes, the same characteristics that complicated their analysis may also save them from rapid degradation: these are hydrophobic membrane proteins that are hard to solubilize and denature.

Nevertheless, it is prudent to take precautions against unintended degradation during the sample work-up. Notably, a protease inhibitor cocktail was added to the sample during the homogenization step to prevent the proteolytic action of the most common endogenous proteases. Additionally, tryptic peptides derived from the extremities of the CYP sequences were avoided, since these locales are generally more susceptible to proteolytic action than the core of the protein [62]. However, the postmortem interval cannot be controlled or modified; its influence must therefore be evaluated.

Hence, a preliminary experiment was performed to evaluate whether CYP proteins in postmortem liver degraded significantly over time spans commonly encountered as postmortem intervals. The goal was to establish if the CYP enzyme would disappear completely from liver tissue over the studied time frame. Of course, the conditions used in this experiment are not typical of a postmortem setting, but rather a worst-case scenario. The liver used for this experiment was collected from a Caucasian male whose cause of death was asphyxia. Results shown in Figure 5.4.4a show that CYP 2D6_Q1 peptide was present at all time points. Although ion ratios are known to have increased variation at low concentrations, the measured ion ratios for these samples still satisfy the identification criteria. Furthermore, although the data is not shown here, other CYP 2D6 and 3A4 peptides were identified at each time point. This screening experiment revealed that CYP enzymes can still be detected at significant levels after 28 days of *in vitro* decomposition at room temperature, although the human liver microsomes mass yield decreased over the time course.

(a)



(b)

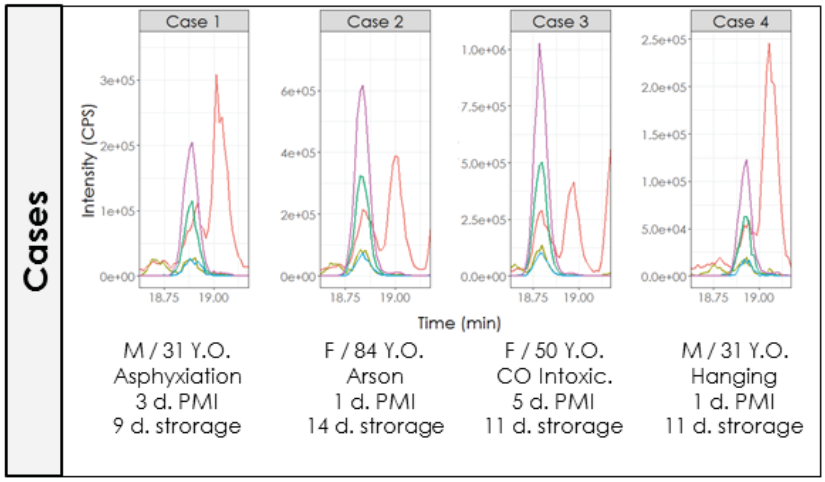


Figure 5.4.4: Postmortem degradation of CYP 2D6 *in vitro* and *in vivo*. **(a)** CYP 2D6_Q1 peptide in human liver microsomes following *in vitro* degradation of 0, 7, 14 and 28 days. Note that the measured signal is subject to sensitivity shifts, as is typically observed for LC-MS/MS instrumentation. **(b)** CYP 2D6_Q1 peptide in four human liver microsome samples each with different postmortem intervals (at unknown temperature) and storage delays (at $4\,^{\circ}\mathrm{C}$). An earlier version of the chromatographic method (compared to the one described in Appendix G.1) was used for this analysis, with the following steps: 0 min-500 μL-min/99% MPA, 5 min-500 μL/min-99% MPA, 50 min-500 μL/min-20% MPA, 50.5 min-1200 μL/min-1% MPA, 55 min-1200 μL/min-1% MPA, 55.5 min-500 μL/min-99% MPA, 60 min-500 μL/min-99% MPA. This resulted in different retention times and additional interference(s) compared to panel (a). Despite this interference, which would affect potential quantification, identification was possible here, which was the goal of the experiment.

While the *in vitro* decomposition results were promising, *"in vivo"* conditions differ in several important respects from these decomposition conditions. For example, initially the temperature would be close to 37 °C. To check whether significant CYP degradation occurred *in situ*, samples with varying postmortem intervals and storage times were collected and analysed. Samples with a postmortem interval ranging between 1 and 5 days and storage time ranging between 9 and 14 days support the conclusion of the screening experiment (Figure 5.4.4b), since again CYP enzymes were detected in all cases at significant levels. These results are also corroborated by the work of Hansen *et al.*, who found relatively stable CYP 1A2 and CYP 3A4 protein abundances in stored liver tissue for up to seven days at 4 °C [189]. However, Hansen *et al.* did find variable and case specific decay patterns after multi-day storage at room temperature, although CYPs could still be identified for most samples.

Taken together, all of this information supports the conclusion that yes, CYP proteins undergo a degradation process postmortem. However, this decay is not rapid nor as drastic as insulin's. Work remains to be done to map out the degradation behaviour as a function of time, temperature and environment over a large number of postmortem cases. Ideally, such work would identify quantitative degradation marker(s) for CYPs which could be used to account for the decay process which took place. In the meantime, the data on hand indicates that degradation will not be a problem for short postmortem intervals of 0 to 3 days and storage of up to 7 days.

5.4.4 Quantification of CYP enzymes in liver tissue

After optimization using quantotypic and mutation peptides, and the verification that the impact of postmortem degradation was manageable, the next step was to assess the method performance criteria: limit of detection (LOD), limit of quantification (LOQ), precision, accuracy, matrix effects, carryover and ion ratio parameters. Results of this assessment are shown in Table 5.4.2. LOD/LOQ for quantitative peptides varied between 0.1 and 1 *pmol* on column (similar to other published quantification methods [44, 65]), and between 0.1 and 2.50 *pmol* on column for qualitative, mutation bearing peptides (not shown). The accuracy was well within the generally accepted ±15% limits, while the precision for all but one quantotypic peptide was below the 20% threshold. Identity confirmation using the four ion ratios generated (1-to-2, 1-to-3, 1-to-4 and 1-to-5) was successful on average 94% of the time in actual samples where the peptide was above the LOD, using an acceptance criteria of ±30%. A blank sample injected after the upper limit of quantification (ULOQ) standard showed

no detected carryover. Therefore, all method validation parameters seemed adequate, except for the matrix effect which indicated important signal suppression in pooled human liver microsomes samples. Table 5.4.3 shows additional matrix effect results in postmortem liver samples. This result prompted further investigation and led to the discovery of several inter-related problems.

Table 5.4.2: Quantification method performance parameters.

Peptide	LOD/LOQ[1] (*pmol* on column)	Precision[2] (%)	Accuracy[3] (%)	Matrix Effect[4] (%)
2D6_Q1	0.5	15	98	32
2D6_Q2	1	7	98	24
2D6_Q3	0.5	14	97	18
2D6_Q4	0.5	13	113	33
2D6_Q5	0.5	6	103	11
3A4_Q1	0.5	26	99	43
3A4_Q2	1	10	111	22
3A4_Q3	0.1	11	102	55
3A4_Q4	0.5	17	110	17
3A4_Q5	0.5	NA	115	17

[1]The LOD, defined as the concentration for which all monitored transitions display a 3:1 signal-to-noise ratio (peak-to-peak), and the LOQ, defined as the concentration for which the most intense quantitative transition displayed a 10:1 signal-to-noise ratio (peak-to-peak). Given the ion ratios, the LOD is equal to the LOQ.

[2]Average %RSD calculated from 3 replicate measurements of a pooled human liver microsomes sample and 5 postmortem human livers. Sample/peptides with a signal-to-noise ratio < 3 were not considered.

[3]Target accuracy of 100% ($[\,]_{\text{measured}}/[\,]_{\text{expected}} \times 100$). Based on the first (quantitative) transition for each quantotypic, averaged over recombinant samples spiked at 20 and 35 *pmol* on column.

[4]Calculated as (IS Area in HLM)/(IS Area in Neat Solution) $\times$ 100, therefore the target matrix effect was 100%.

First, matrix effects were not limited to microsome digests but also occurred in recombinant CYP digest samples. A concentration dependent pattern of signal suppression was observed by tracking the area of the internal standards, reaching 57% on average in the ULOQ (2500 $pmol/mL$). Nevertheless, the accuracy of spiked samples remained adequate, demonstrating that AQUA-IS corrected properly for this effect.

Table 5.4.3: Matrix effects (%) in postmortem liver samples.

Peptide	Liver 1	Liver 2	Liver 3	Liver 4	Liver 5
2D6_Q1	5	9	4	3	35
2D6_Q2	6	11	8	8	36
2D6_Q3	32	64	58	43	115
2D6_Q4	21	36	31	24	53
2D6_Q5	6	10	8	7	21
3A4_Q1	77	123	108	85	198
3A4_Q2	NA	47	40	37	98
3A4_Q3	4	8	5	4	16
3A4_Q4	1	7	3	3	47
3A4_Q5	16	28	20	15	72

Calculated over 3 replicates and all 5 transitions as (IS Area in Postmortem Liver)/(IS Area in Neat Solution) $\times$ 100, therefore the target matrix effect was 100%.

A problem related to signal suppression, unsurprisingly, was the loss of detection for some of the lowest sensitivity peptides, in particular the mutation bearing peptides. For example, in one pooled and 5 postmortem HLM samples, loss of detection of wild-type mutation bearing peptides occurred in 33% of cases, mostly with CYP 2D6 WT peptides which were less sensitive than CYP 3A4 WT peptides.

A problem demonstrating mass spectrometer overload, likely leading to ion suppression, was the impact of injecting a large number of samples in a row. After the injection of 60 postmortem liver microsomes digested samples, the sensitivity dropped by 98% and charging on the first quadrupole (Q1) was observed. A Q1 MS scan experiment on blank samples, recombinant CYPs, pooled and postmortem HLM samples did not reveal the presence of a single major contributor to this large matrix effect. A further experiment monitoring for the detergent mass showed that it eluted in the chromatographic window where the divert valve of the MS was set to waste; hence, the detergent is not responsible for the matrix effect or suppression.

The presence of this overpowering matrix was also evident in the shifting retention times, varying as much as 10% between samples. This manifestation in otherwise properly shaped and symmetrical peaks indicated that the chromatography column was also overloaded.

All of these issues are likely symptoms of the same problem: despite ultracentrifugation

and SPE clean-up, sample purity needed to be increased to remove the numerous interfering compounds. While other in-solution HLM digestion CYP analysis methods have been published before [47, 65], none was accompanied by a matrix effect evaluation and it is probable that this issue was missed entirely. Moreover, there is a possibility that postmortem livers, where collection often occurs well after death (i.e., PMI $>$ 1 day) yield a more complicated microsomal fraction than fresh liver. In any event, additional sample purification is indicated, and this could be done via gel electrophoresis or immunoaffinity preparation, for example [44, 66, 70, 207].

Another quantification concern was identified when comparing the CYP concentrations measured in pooled and postmortem samples via the different quantotypic peptides (Table 5.4.4). The quantification results do not agree with one another within the margin of their respective errors. One can wonder whether the calibration strategy (external calibration with AQUA-IS) was to blame for those results. However, calibration using the label-free approach (no IS), with a reporter (surrogate) protein as an IS (glutamate dehydrogenase), or standard addition technique performed worse, both at the level of method performance parameters and peptide quantification discrepancy. These results are troublesome if we claim to be performing absolute quantification. However, it is debatable whether accurate quantification is even actually achievable for proteins [64], in particular for CYPs. In much the same way that absolute quantification of drugs in hair samples is always subject to the unknown extraction efficiency from the hair matrix, a quantification method such as this one will always be subject to the unknown variations of the extraction/recovery efficiency of CYPs from microsomes derived from liver tissue, and from digestion efficiency. In that spirit, Table 5.4.4 also displays quantification results relative to the pooled HLM sample, i.e., the absolute quantification results for each case divided by the absolute quantification result for the pooled HLM sample for each quantitative peptide independently. These figures are somewhat more coherent than the absolute quantification results, but unfortunately, some discrepancy remains.

This quantification discrepancy between peptides could result from either differences in digestion efficiencies, recovery or interferences from one sample to the other. This finding appears at odds with the initial assessment that reliable CYP quantification techniques were available in the literature [44, 47, 65–67], and only minor adjustments and simplifications to fit forensic toxicology samples and standard practices would be required. A survey of published CYP quantification methods quickly showed that, in the vast majority of cases,

Table 5.4.4: Quantification discrepancy between quantotypic peptides.

Absolute Quantification (*pmol* on column)						
Peptide	Pooled HLM	Case A	Case B	Case C	Case D	Case E
2D6_Q1	8	7	28	92	154	4
2D6_Q2	3	3	16	27	33	2
2D6_Q3	3	5	35	45	49	5
2D6_Q4	0	3	18	31	37	2
2D6_Q5	0	0	14	22	23	4

Relative Quantification						
Peptide	Pooled HLM	Case A	Case B	Case C	Case D	Case E
2D6_Q1	1.0	0.8	3.4	11.2	18.7	0.5
2D6_Q2	1.0	1.0	4.9	8.6	10.3	0.8
2D6_Q3	1.0	2.0	13.1	17.1	18.5	1.9
2D6_Q4	NA	NA	NA	NA	NA	NA
2D6_Q5	NA	NA	NA	NA	NA	NA

only one quantotypic peptide was used for concentration measurement [44, 47, 65], often with calibration curves made of spiked synthetic peptide rather than recombinant CYP enzymes [47, 65]. In cases where more than one quantotypic peptide was used, discrepancies were observed [67, 189].

The method, as it stands now, is reproducible but presents some obvious issues at the quantification level. These issues were likely present in other published CYP quantification methods but went undetected. Given the importance of CYP abundance estimation in liver tissue for proper phenotypic evaluation of individuals, there is therefore an undeniable need to either further develop the method to make digestion efficiency uniform between samples, in a way that at least allows relative quantification, or to map out the impact of such a phenomenon and figure out an adequate response. In the meantime, area ratio can be considered a very rough estimation of the CYP level in liver tissue, and since we are primarily interested in fold differences present between individuals [44, 65], we can still perform a preliminary applicability experiment on HLM samples of known genotype.

5.4.5 Application of the method to genotyped human liver microsomes samples

This preliminary method aimed at combining enzyme abundance and activity information to estimate postmortem metabolic capacity was therefore run on genotyped human liver microsome samples to find out whether results matched the expected profile and could correctly classify individuals as poor or intermediate metabolizers, for example. Genotyping is not a perfect method to estimate metabolic capacity, but it does have its usefulness and is, at least in the extremes, relevant (i.e., when the gene is suppressed and no enzyme is present, the metabolic capacity is null). The goal is to see whether this proof-of-principle method can, at least, reproduce the *de facto* gold standard.

The *5 mutation is a pure deletion of the CYP 2D6 encoding gene, hence homozygous bearers will not produce any CYP 2D6 enzyme. The *6 mutation introduces a frameshift from amino acid 118 and on. In all likelihood the protein produced, having nothing to do with CYP 2D6 and no utility, is quickly degraded by the cell. The *4 mutation is an interesting case, and introduces several point mutations (e.g., P34S, H94R, S486T), but also a splicing defect/frameshift. Although point mutation bearing peptides were initially monitored, analysis of *4 homozygous individuals make it quite obvious that if any CYP 2D6 enzyme is produced, it is quickly destroyed by the cell. No *4 mutation bearing peptide therefore needs to be monitored, only the global enzymes level.

Results shown in Figure 5.4.5 show that although CYP 3A4 was present in all cases but one, no CYP 2D6 was detected in *4/*4 and *4/*6 poor metabolizers, and reduced levels were observed in *1/*4 and *1/*5 intermediate metabolizers. The proteomics approach using mass spectrometry detection presented here would therefore have generated a correct estimation of the metabolic capacity, coherent with the known genotype. Hence the proposed method can properly estimate metabolic capacity for genotypes acting on enzyme levels, but what about mutations acting on enzyme activity?

To properly account for these, mutated versions of peptide bearing mutation sites would need to be identified. Unfortunately, in real HLM samples, the sample purity problem decreased sensitivity, as described above, and prevented this (Figure 5.4.6). Whereas the wild-type *9 peptide was detected in recombinant CYP, neither the wild-type nor the mutated *9 peptides were detected in *1/*1 or *6/*9 human liver microsomes.

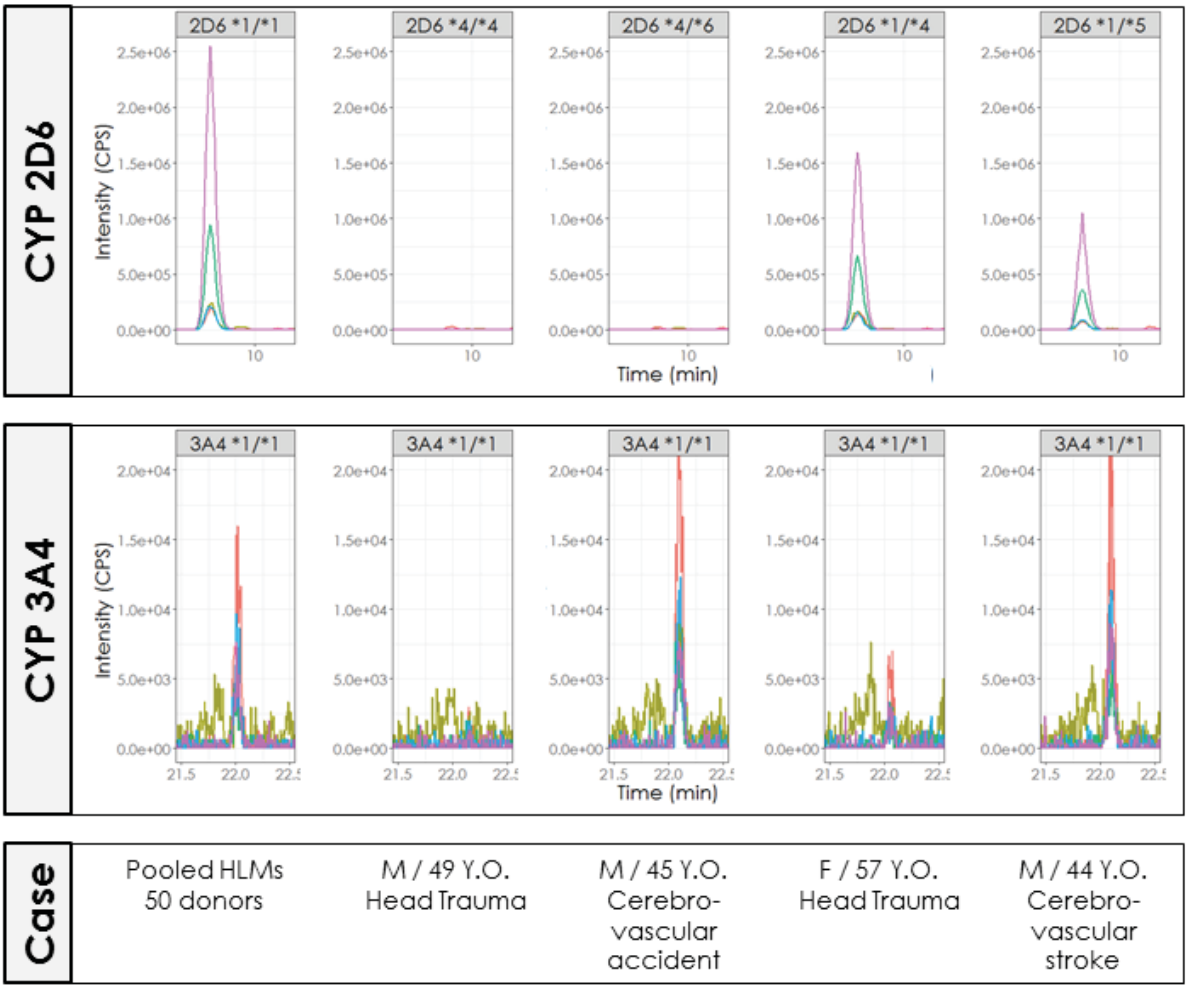


Figure 5.4.5: CYP 2D6 and 3A4 levels in genotyped human liver microsome samples.

However, a situation where the matrix effect is far less intense is shown by the analysis of recombinant CYP 2D6*10 (Figure 5.4.7). The mutated enzyme displays the mutated *10 peptide, but not the wild-type *10 peptide, as expected. This indicates that screening mutation bearing peptides will allow determination of the mutations displayed by an enzyme and its activity level. These analyses on genotyped human liver microsomes demonstrated conclusively that while improvements to sample purity and quantification processes are necessary, a method to estimate metabolic capacity postmortem via characterization and quantification of metabolizing cytochrome P450 enzymes would yield highly relevant results.

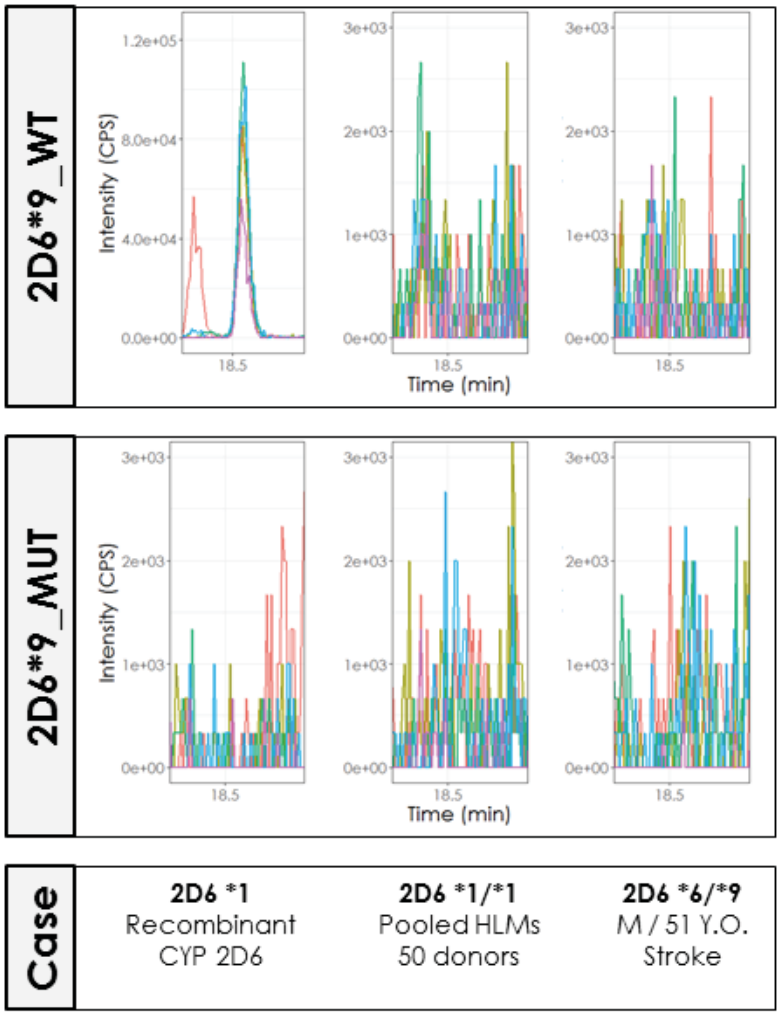


Figure 5.4.6: CYP 2D6*9 wild-type and mutated peptides in recombinant CYPs and genotyped human liver microsomes (1 mg/mL). If recombinant CYP 2D6*9 was analyzed, levels around 1×10^5 would be expected. As for genotyped human liver microsomes, it is impossible to know what peak intensity would be without correcting the sensitivity problem.

5.5 Conclusion

The method developed in this paper isolated human liver microsomes from postmortem liver tissue samples and subjected them to tryptic digestion. The peptides, measured via liquid chromatography coupled to tandem mass spectrometry, were used to characterize mutations and quantify cytochrome P450 2D6 and 3A4 enzymes. Quantification was based on 5 quantotypic peptides per protein and utilized external calibration curves of recombinant CYPs and stable isotope-labelled internal standards. Mutations were identified by monitoring mutation bearing peptides in their wild type and mutated versions.

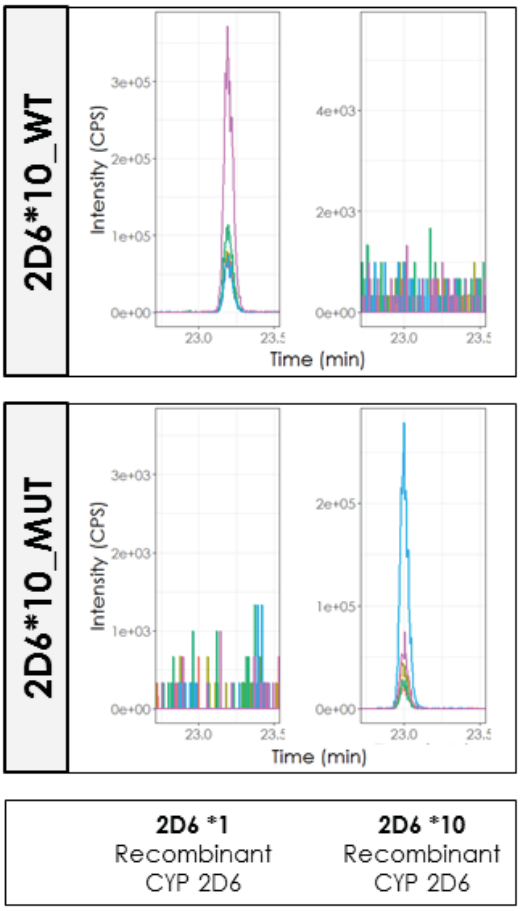


Figure 5.4.7: CYP 2D6*10 wild-type and mutated peptides in wild-type and mutated recombinant enzyme (1 mg/mL).

A method such as this one should, in principle, allow for a more accurate estimation of metabolic capacity than genotyping alone since it accounts for differences in protein expression, affecting the phenotype. Having a single method including both characterization and quantification of CYPs would allow for straightforward metabolic capacity estimation using LC-MS/MS instrumentation that is readily available in the vast majority of forensic toxicology laboratories – which is currently not the case with genotyping.

Issues identified in the pre-validation stage (matrix effect, retention time shifts, Q1 charging) prevented full method validation. At the heart of the problem is the need for increased sample purity. A more selective sample preparation technique, such as gel electrophoresis or immunoaffinity preparation could be developed to solve this issue. Additionally, discrepancies in concentrations measured via the different quantotypic peptides were noticed in human liver microsome samples. Optimization of the digestion method directed not towards sensitivity, but rather levelling digestion efficiency between samples, or properly accounting

for it, for example using a full length SIL-IS protein standard [64], will be necessary before a full validation can be performed and succeed.

A comprehensive method such as this one has multiple, interconnected moving parts. Our review of the available literature seemed to indicate that these parts were already in a working state and only needed minor adjustments. As such, optimizations were started on the method as a whole, only to realize that there are significant issues with more than one of these parts that will require specific and dedicated attention. Solving these problems will require collaboration between different specialized areas: genomics, molecular biology, proteomics, analytical chemistry and forensic toxicology.

Nevertheless, the results presented here offer a proof of concept that characterization and quantification of CYP enzymes to estimate metabolic capacity postmortem is a sound avenue. Hansen *et al.* corroborate this conclusion with their work on CYP quantification in postmortem liver samples. Analysis of genotyped HLM revealed CYP 2D6 and 3A4 expression patterns matching the expected ones, while the predicted wild-type and mutated peptides were observed in recombinant samples. Hence, while further method development is still needed, this type of procedure is a new doorway to more accurate estimation of metabolic capacity postmortem.

5.6 Funding

Brigitte Desharnais and Cameron D. Skinner gratefully acknowledge support of the National Sciences and Engineering Research Council of Canada. Brigitte Desharnais also gratefully acknowledges the support of the Fonds de recherche du Québec - Nature et technologies.

5.7 Acknowledgements

The authors wish to thank Lucie Vaillancourt for her extensive technical work, and Julie Laquerre for her ideas and contribution to data analysis.

5.8 Postmortem metabolic capacity estimation

The metabolic transformation of a substrate into its products by an enzyme depends on multiple factors. CYP enzymes are typically selective in the transformation that they carry-out but are not specific to a single substrate. Therefore, a single xenobiotic is usually metabolized by more than one CYP enzyme, its transformation being split across several metabolic pathways. Thus the overall metabolic capacity for a given compound is the sum of all the individual enzymatic capacities. While it was not possible to achieve a completely functional and validated CYP quantification and characterization method, it is possible to develop the framework for how a postmortem metabolic capacity could be calculated and applied in forensic cases.

The general metabolic capacity for a particular drug could be hypothesised as follows:

$$MC_{\text{drug}} = \sum_{e=1}^{E} \gamma_e \times \alpha_e \times \rho_e \times [e] \times m \tag{5.1}$$

Where:
MC_{drug} is the metabolic capacity of the individual for the drug studied;
e are all enzymes participating in the metabolism of the drug;
γ_e is a correction factor accommodating for the postmortem loss (degradation) of enzyme e;
α_e is the activity level of enzyme e with respect to the WT version (= 1);
ρ_e is the proportion of drug metabolized by enzyme e;
$[e]$ is the concentration of enzyme e in the liver tissue;
m is the mass of the liver.

The γ_e term represents an estimation of the postmortem loss through degradation of the enzyme in the case studied. Initial work suggests that for the majority of cases where the body is recovered, cooled, examined and sampled quickly, this factor would be close to 1 (Section 5.4.3). As this method is further developed, experimentally determined values for CYP under alternate and less favorable conditions could be developed. It is expected, from the food production literature [208], that degradation markers could be identified, perhaps via quantification of the fraction of selected peptides to their proteins. However, development of this method would constitute a large undertaking on its own, most likely requiring the use of high resolution mass spectrometry.

The value of the α_e term depends on the genotype. Most commonly, individuals would

bear the extensive (normal) genotype and the α_e term would be assigned as 1. On the other hand, for example, the mutation CYP 2D6*17, which is highly prevalent in individuals of African descent, would be assigned an α_e value of 0.15 [209].

The values for ρ_e can also be obtained from the literature, as there are many pharmacokinetic studies where the relative contributions to the metabolism of common licit and illicit drugs are reported, e.g., [210–212].

Last, but not least, $[e] \times m$ is the concentration of the CYP times the liver mass to yield the total amount of CYP present. Thus far, accurate determination of the postmortem CYP enzyme concentration has proved to be difficult. However, literature, as well as variability of the observed area ratios (Figure 5.4.5), indicates that concentration of enzymes can vary by several orders of magnitude. This natural variability is likely linked to epigenetic factors, and can be compounded by the presence of liver disease, one prominent example being cirrhosis. Normal male livers have a mass of $\approx 1500\ g$ but, the normal range encompasses a rather large 840 to 2590 g making it essential to include in the calculation [213].

While it is not currently possible to carry-out a complete metabolic capacity calculation, it is possible to explore how it might play out for a set of selected drugs. In the following examples, the metabolic capacity towards venlafaxine and fentanyl will be compared in several hypothetical cases. As the body of the victim is often recovered shortly after death, γ_e, the correction factor accommodating for the postmortem loss (degradation), will be assumed to be 1. Figure 5.8.1 shows the primary metabolism routes for both drugs and the enzymes carrying them out.

Let's say that in all of these cases, a lethal concentration ($> 10\ \mu g/ml$ blood [210, 215]) of venlafaxine is detected. The likely cause of death is thus an intoxication. However, in cases A, B, and C, the metabolic capacity is below what is normally expected. As such, the forensic toxicology report should include information to the effect that the deceased had a decreased (or lower than average) metabolic capacity for venlafaxine. This may mean that regular therapeutic doses of venlafaxine, an antidepressant, might have led to an accumulation of the drug, adverse effects, and could be contributory to the death. If, rather than a fatal venlafaxine concentration, a fatal fentanyl concentration is found (40 ng/mL [210, 215]), cases E and F would be reported with a similar caveat, while metabolic capacity would not

(a)

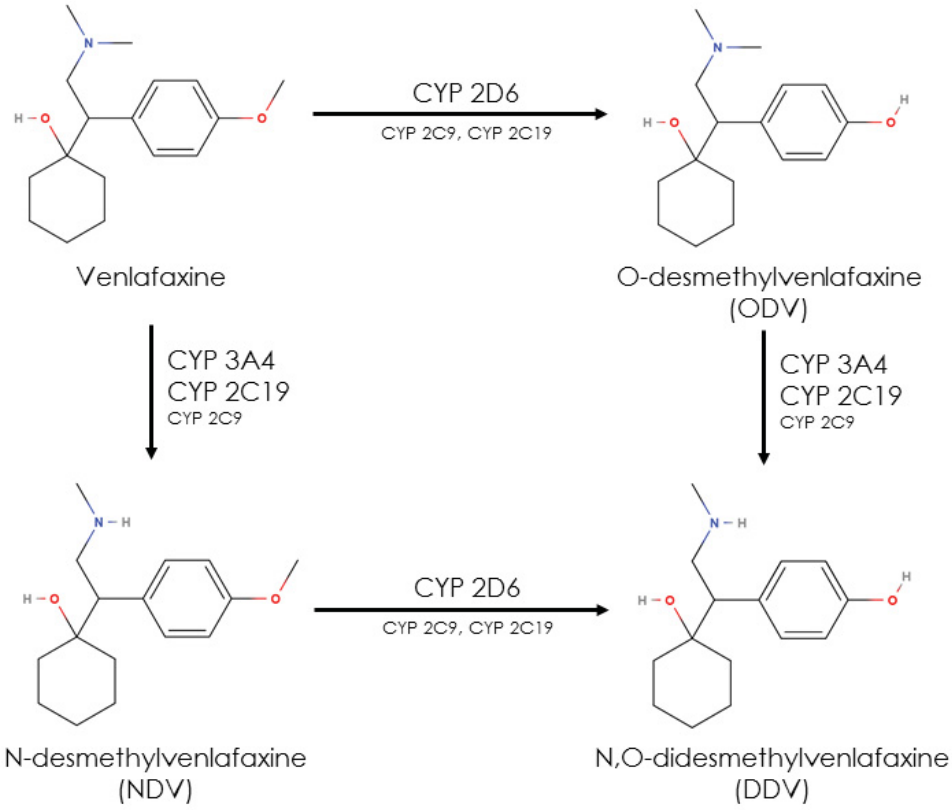


Figure 5.8.1: Metabolism of selected drugs. **(a)** Metabolism of venlafaxine [214].

(b)

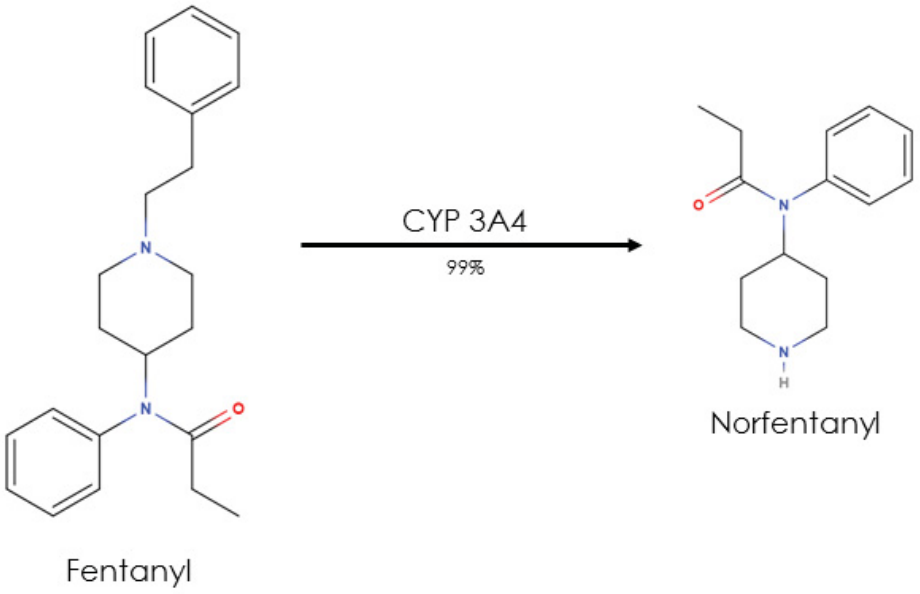


Figure 5.8.1: Metabolism of selected drugs. **(b)** Metabolism of fentanyl [15].

be a consideration for cases A to D. Each of these findings influence the interpretation of the toxicology results and provide information pertaining to the circumstances surrounding these deaths by venlafaxine intoxication. While the cases presented in Table 5.8.1 put an emphasis on the impact of the genotype as dictating the concentration of CYP enzymes in the liver tissue, measured concentrations vary significantly due to epigenetic factors (e.g., health, habituation, injury) with much the same results.

Building metabolic capacity estimators for the vast array of drugs encountered in the forensics context would require meticulous work of sifting through the literature to collect data on the relative activity of different mutated versions of polymorphic enzymes (information which might not even be fully available yet, or not quantitatively) as well as the proportion of metabolism carried out by each CYP enzyme (again, information which is not always available). In some cases, a large part of the metabolism is carried out by non-CYP enzymes (carboxylesterases, alcohol dehydrogenase, etc.), and those drugs would require a special dedicated treatment. This picture is even more complex in the case of drugs like tramadol [15], which is mostly metabolized to an inactive metabolite by CYP 3A4, but generates a metabolite via CYP 2D6 which is hundreds of times more active than the parent drug.

Table 5.8.1: Metabolic capacity in different hypothetical cases.

Case	Genotype	Activity α_e	Measured concentration $[e]$[2]	Metabolic capacity (MC)[3,4]
A[1]	2D6*4/*6	0	0	FEN: 100
	3A4*1/*1	1	100	VEN: 15
B[1]	2D6*1/*5	0	0	FEN: 100
	3A4*1/*1	1	100	VEN: 15
C	2D6*10/*10	0.5	100	FEN: 100
	3A4*1/*1	1	100	VEN: 45
D	2D6*1x2/*1x2	1	200	FEN: 100
	3A4*1/*1	1	100	VEN: 135
E	2D6*1/*1	1	100	FEN: 50
	3A4*8/*8	0.5	100	VEN: 68
F	2D6*1/*1	1	100	FEN: 20
	3A4*1/*1	1	20	VEN: 63

[1]Akin to the cases analyzed in Section 5.4.5.

[2]To simplify, it is assumed that some mutations prevent any CYP enzyme from being produced, hence $[e] = 0$, or generate excess production hence $[e] = 200$. For non-mutants the normal amount corresponds to $[e] = 20$ to 100, within the 20-fold differences in enzyme abundance noted in the literature [44, 65].

[3]Proportion (ρ_e) for fentanyl (FEN) is $> 99\%$ CYP 3A4 [15], for venlafaxine (VEN) is $> 60\%$ CYP 2D6 and $< 15\%$ CYP 3A4 [214]

[4]For example, in Case A, $MC_{\text{fentanyl}} = \sum_{e=1}^{E} \alpha_e \times \rho_e \times [e] = (1 \times 1 \times 100) = 100$ and $MC_{\text{venlafaxine}} = \sum_{e=1}^{E} \alpha_e \times \rho_e \times [e] = (0 \times 0.6 \times 0) + (1 \times 0.15 \times 100) = 15$. The mass of the liver was considered constant for all cases and omitted from the calculation.

While the limitations of the available information noted above are acknowledged, they are not, in of themselves, disastrous. Much of the necessary information for the most prevalent drugs are available, and as more cases are examined it will also be possible to develop empirical metabolic capacity distribution densities. Thus, there is still clearly a large amount of work to be done to make this a routine tool for forensic toxicologists but, it is also clear that it will be a valuable tool.

6

General Conclusion and Future Directions

The overall goal of this research is to develop a method for the characterization and quantification of CYPs for postmortem evaluation of the metabolic capacity. While it was not possible to develop a fully functional and validated method for postmortem evaluation of the metabolic capacity, important contributions to the field of analytical forensic toxicology were made. Significantly, the proof of principle has been demonstrated using selected CYPs, the problems identified and a clear route forward established.

Four different research projects exploring interfaces of forensic toxicology with other areas of specialization in science were presented in this thesis. The necessity of these subsidiary projects became evident considering the validation requirements for the intended CYP analysis method and for forensic toxicology applications in general. Chapters 2 through 4 [1–5] developed and described mathematical method validation tools, and Chapter 5 details the progress made in a postmortem CYP quantification and characterization method [6].

6.1 Selection and validation of a calibration model

Selection of an appropriate calibration model, including weighting and regression order, is critical to obtaining accurate results out of all bioanalysis methods. Indeed, poor model selection can lead to quantification biases of up to 154% [1, 2]. Inadequacies in previously established methodologies for model selection and validation were identified early in the work, an especially concerning issue for forensic toxicology where accurate results are critical.

In Chapter 2, methodologies for calibration model selection and validation were systematically assessed in the context of an LC-MS/MS multi-analyte method. Based on the results of this investigation, an automated, analyst independent tool for selection and validation of a calibration model was put forward. In this R script, the need for weighting was first assessed via an F-test on the lower and upper limits of quantification (LLOQ and ULOQ) measurement replicates. Subsequent variance testing allowed for the selection of the most appropriate weighting. Regression order was chosen based on the results of a partial F-test. Finally, the selected model was validated by normality testing of the residuals, either via the Cramer-von Mises or Kolmogorov-Smirnov approach. Best practices with regards to the experimental setup (number of replicates, intra/inter batch) were also established.

Calibration curves are a staple of quantitative methods, in bioanalysis as well as in forensic toxicology. Despite the ubiquitousness of calibration curves, literature and validation guidelines on the topic are not as developed as one would hope. Hypothesis tests were suggested piecemeal, for selection of one part or the other of the model. A number of validation guidelines presented several competing tests for the same purpose. Some suggested tools were entirely subjective, relying on the user to make a judgment call about the result. Others were simply inapplicable to standard bioanalysis methods, being too sensitive to the number of calibration levels or measurement replicates. Often, calculation examples were absent, or shown only for the simplest, linear unweighted, model. Knowing that quadraticity and weighting can radically modify statistical test calculations, not guiding the user better in this process is a recipe for errors. Considering this initial environment, there was an obvious need for a comprehensive calibration model selection and validation package, which, while not being to only way to achieve the goal of calibration model selection and validation, would at least provide explicit benchmarked tools to the practitionners. The reception of the papers, published as an Editor's Choice in the Journal of Analytical Toxicology [1, 2], demonstrates this in itself: this procedure is now cited as an update to validation procedures in forensic toxicology [155] and has been adopted as a method development and validation

tool by several laboratories [216–220].

To improve its adoption, this tool was developed with an easy-to-use interface to perform accurate, appropriate and complex statistical calculations. The tool is simple by virtue of decoupling the need for an understanding of the calculations and the operation. It can thus be applied by any forensic toxicologist, or bioanalyst, regardless of their level of knowledge in statistics. This will allow analysts and laboratories to take advantage of the significantly better reliability of this procedure over the traditional "fit and check" approach. To highlight this, simulations demonstrate that the traditional approach would succeed on average 47% of the time in selecting the correct weighting and 61% of the time in picking the right order. Under similar test conditions, the developed approach elevates those numbers to 96% and 90%, respectively. This result is a testament to the leaps which can be made when research groups of different background (statistics and analytical chemistry) collaborate. The collaboration established here yielded a final product that would not exist (or at least, not in such a comprehensive form) if one collaborator or the other was absent.

Despite its clear advantages, the developed tool has limitations that need to be acknowledged. The algorithm has a harder time correctly identifying 1/x heteroscedastic data, with an 88% success rate on average. It also has difficulty identifying quadratic data when variance increases proportionally with concentration ($1/x^2$ weighting, 48% success rate), undoubtedly due to the increased variability at the higher concentration levels masking the curvature. These difficulties would be endemic to any model selection process, including those relying on the analyst's visual judgement. Additionally, although the tool is simple to operate, its use requires additional work at the method validation stage. However, this was limited as much as possible by design, and is of course a small price to pay for the additional accuracy and confidence in the model gained.

At the outset of this project, all approaches, tests and procedures suggested in bioanalysis guidelines and publications were considered.. Several were tested and their efficiency compared in Section 2.2. However, some solutions were not explored by design, although this is not necessarily spelled out for the reader (due to manuscript space constraints). For example, we focused on Kolmogrov-Smirnov and Cramer-Von Mises normality testing for the residuals, without exploring the well-known Shapiro-Wilk test. This is because the latter asseses whether a set of independent and identically distributed (iid) random variables are normally distributed. In the case studied, the goal was to test whether the (unobserved) regression errors $e_ij = y_ij - f(x_i)$, where $f()$ is the true (thus unknown) underlying regression model,

are normally distributed. However, as mentioned, the regression errors are not observable: we need to estimate them (these estimates are often called regression residuals). Since the residuals are estimated, they are not iid anymore since they are calculated using the fitted coefficients, which are themselves calculated from the data. Therefore, the Shapiro-Wilk test is not appropriate in this case. Since a high premium was placed on automation and ease-of-use, approaches were not tested if they did not fit the type of procedure targeted. The starting point and history of this project means that not all of the broad statistical literature on how to conduct model selection was covered. Other paths such as likelihood ratio tests or the Aikaike Information Criteria could be tested as potential alternatives. Some tests which were set aside for the previously stated reasons, such as the confidence interval of the quadratic term, could be assessed to see if the benefits outweigh the problematic aspects. Nonetheless, the developed approach has the advantage that it is readily understandable by practical bioanalysts, requires litlle statistical training and uses the already familiar P-values as a model selection tool. Moreover, only the most common weighting patterns (none, $1/x$, $1/x^2$) in data analysis software were considered. However, several software packages also offer additional options which could be considered, such as $1/y$ or $1/y^2$.

One area of application where greater clarity remains to be obtained is guidance on how to properly use and interpret results when several instruments are to be used in the production setting. How should an analyst deal with diverging results, e.g., a linear, $1/x$ model has been selected on one system and a quadratic, $1/x^2$ model has been selected on a second one? This was superficially covered in Section 2.2, but should be investigated further, as it is common for larger laboratories to have several instruments operating in parallel.

A plethora of questions remain to be answered about calibration curves to reach the ultimate goal of analytical chemists: discovering the actual model of the signal measured as a function of concentration, while expending as little experimental work as necessary. This underlying model information is essential to obtain truly accurate and precise estimations of the unknown sample's analyte concentration. Until now, it has been dogmatic to always use the simplest model but, should we start using models other than linear or quadratic? What is the most efficient distribution of the concentration levels of standards and quality control samples? What replicate pattern would serve us best for the different standard levels of a calibration curve? What about the replicate pattern for quality control samples giving us the most accurate information on the uncertainty of measurement? How can we do all of that while minimizing the burden on production operations?

6.2 Quantitative correction of the endogenous concentration in the calibration matrix

In Chapter 3, a tool to automatically correct for endogenous analytes concentration in the calibration matrix was presented [3].The presence of an analyte, at some endogenous concentration in the authentic calibration matrix will bias quantification results. The mathematical tools developed allow correction for this bias, first by calculating the endogenous concentration via the x-intercept of the calibration curve in a standard addition like procedure, then by adding this calculated concentration to the spiked concentrations for all standards and quality controls, generating a de facto concentration to be used in all calculations. This effectively shifts the calibration curve to the right, correcting for the bias introduced by the presence of the endogenous analytes.

Analytes displaying an endogenous content are common in forensic toxicology. γ-hydroxybutyrate (GHB), β-hydroxybutyrate (BHB), cyanide, carboxyhemoglobin (HbCO), insulin and steroids are but a few examples of such molecules [90–92, 147, 153, 221]. To those, we can add common xenobiotics found in the population such as caffeine or alcohol consumption markers [143–145], for which finding authentic blank matrices can also be difficult. Clearly, this is a hurdle regularly faced in forensic toxicology. However, proposed solutions are either manually intensive, time consuming, potentially inaccurate, or simply ignore the problem [93–100].

Apart from improving accuracy, the main advantages of the developed tool are its conceptual simplicity and seamless integration in the production environment of the bioanalysis laboratory. This last advantage is a crucial requirement for mathematical tools to be accepted and used by bioanalysts and forensic toxicologists. The demands of a high throughput environment do not allow for extra operations to be performed manually or outside of the standard software environment. Moreover, the target user is the analyst, not necessarily a research scientist with an enthusiasm for the detailed mathematical concepts. Therefore, usage of the tool was, once again, decoupled from a thorough understanding of its underpinnings.

The downside of this approach is that it required the use of a suboptimal mathematical approach from the perspective of errors introduced. Some alternative approaches would result in less error being introduced in the final calculated concentration. Unfortunately, direct integration of those with the commonly used data analysis software is not currently

possible. An exact calculation of the error folded in by the correction process was performed, and an error evaluation tool was developed to help guide users. In the end, the relative error is generally decreased by the correction process, although the absolute error increases.

Eventually, this tool or its underlying mathematical calculations could be integrated in the data analysis software. Already, Sciex and Agilent have made steps in this direction with Sciex OS' Standard Addition Processing feature and Mass Hunter's Blank Subtraction Add In. Full integration in the software ecosystem would mean correction could systematically be applied to endogenous analytes, with no associated productivity cost. Hopefully the publication of this paper will highlight its need to the software designers.

Both this chapter and the previous one presented automated statistical tools to deal with a specific method validation or production issue. When these tools were developed, a premium was placed on making the end experience easy for the scientist manipulating it. This is because we did not want statistical knowledge to be a barrier to the use of better and more relevant tools. But these tools are not a blackbox by any means. In both cases, the calculations were clearly laid out in the published papers. Moreover, the full R scripts are available free of charge to anybody who might want to access them to probe the calculation further or perform their own modifications. In the end though, to convince a large number of laboratories to use these tools, ease of use is an essential ingredient. Dissemination of these improved practices occurs by different means: the published papers, conferences, e-mail exchanges, explicative papers in non-peer reviewed outlets, social media, etc.

In the quantification aspect of the CYP analysis method, matrix matched calibration curves will be prepared to quantify CYP 2D6 and 3A4 enzymes in postmortem liver tissue. The biological matrix used to prepare these spiked samples will contain an endogenous amount of CYP enzymes (unless fastidious and impractical means are taken to avoid it), yielding biased calibration curves. The tool developed in Chapter 3 will allow for complete correction of this bias, while retaining the correction advantages of matrix matching, yielding an overall more accurate quantification of CYP enzymes.

6.3 Qualitative decision point method validation

Qualitative methods are characterized by their binary outcomes, such as "presence" or "absence", or "above" or "below cut-off" [4]. Unfortunately, the fact that continuous mea-

surements underpin those binary outputs has long created confusion. The rare validation guidelines [75] which cover this type of method, and not only the quantitative ones, typically adapt quantitative method validation procedures rather than presenting specific procedures for qualitative methods. Moreover, new requirements from the International Organization for Standardization compounds the need for better validation guidance. Indeed, the most recent version of the ISO 17025 guidelines (2017) requires taking into account uncertainty of measurement for all methods, rather than only the quantitative ones [72]. This requirement is further bolstered by the position of the American Statistical Association (ASA) which states that all forensic expertise reports should include "the limitations and uncertainty associated with measurements, and the inferences that could be drawn from them" [222]. Current guidance on qualitative methods is clearly lacking and does not fulfill those requirements.

In Chapter 4, we presented a series of papers tackling this very question. Despite its seemingly trivial nature, early analyses quickly revealed that this was not the case, and that available literature suffered from notable gaps [4]. Refinements to the standard model of normally distributed measurements were developed, namely accounting for heteroscedasticity and sampled threshold [4]. These two factors by themselves modify the expected rates of false negative, reliability and sensitivity in particular, hence employing the correct model can make the difference between an analyte passing the validation stage or not. The method validation guidelines developed specifically for qualitative decision point methods were successfully applied to a multi-analyte LC-MS/MS method in oral fluid in an efficient and thorough validation process [5].

Qualitative decision point methods are common in a number of application areas: anti-doping, workplace drug testing, food safety, and of course forensic toxicology [102–104]. In all of these domains, a threshold concentration is often dictated by the law. In forensic toxicology in particular, *per se* regulations are commonplace and establish a maximum drug concentration in a specific biological matrix (e.g., blood) which a driver is allowed to have. Additional *per se* regulations came into effect in Canada in 2018, as part of the cannabis legalization initiative [104] and prompted the project presented in Chapter 4.3.

The developed approach is much more coherent both with the definition of method validation and the characteristics of qualitative decision point methods. The goal of method validation procedures is to demonstrate the performance characteristics of a method as it will be used. Since the categorical (binary) output from qualitative methods is intended to

be used as the method output in the production setting, performance indicators used for method validation should rely on those same binary results to estimate the adequacy of the method, not on the continuous measurements. Additionally, the improved data modeling developed is a marked advantage over previously published material about qualitative methods.

An obvious comment can be made regarding the increased workload required by the application of such a framework, which is higher than what would be required by the current SWGTOX's guidelines [75]. It is indeed a shift in perspective. We are used to considering qualitative methods as easy to validate, or at least, as requiring less work than quantitative methods. But this myth needs to be dispelled. Qualitative methods generate less detailed information than quantitative methods do, thus adequately characterizing their performance requires a greater number of analyzed samples.

One limitation of the developed approach is the inability to estimate the precise size of the unreliability zone on a daily basis, due to shifting instrument variance. To address this issue, overestimates of the unreliability zone size are used. But ideally, mathematical modelling tools would allow us to estimate precisely the boundaries of this confidence interval and compare results to it.

Another area which has yet to be explored is how to properly validate and operate qualitative identification methods, which yield a detected/not detected binary result. Although several characteristics are similar, or identical, to the decision point methods already discussed, one major difference can be pointed out. Whereas qualitative decision point methods generally present continuous measurement distributions, qualitative identification methods have semi-continuous measurement distributions, i.e., the probability of a measurement being equal to zero will be non-null, but conditionally to the measurement being > 0, its distribution is continuous (also called data with excess zeros). This modification to the fundamental distribution of the data means modelling has to be carried out all over again to properly predict the behavior of qualitative identification methods. This has natural effects on the validation guidelines and production framework.

Thus, numerous aspects of qualitative methods remain to be explored, but the work presented in Chapter 4 has established a stable ground for those investigations. Additionally, it will be used for the characterization aspect of the CYP analysis method, by providing a validation framework for wild-type and mutated peptides. These peptides are not quantified, but rather their presence or absence above a biologically relevant and analytically feasible

threshold is sought, i.e., a decision point qualitative method is required.

6.4 Proteomics approach to postmortem estimation of the metabolic capacity

In Chapter 5, we presented the proof of concept that a proteomics approach to postmortem estimation of the metabolic capacity is possible. The activity level of CYP 2D6 and CYP 3A4 enzymes were estimated via monitoring of point mutation bearing peptides. The most deleterious and prevalent point mutations were monitored for both these important model enzymes. This piece of information, when combined with quantification results obtained in liver tissue, will yield an estimation of the metabolic capacity. This approach, relying on the analysis of liver tissue, is particularly well suited to the postmortem setting where it is fully available for analysis (by contrast with the clinical domain where access is greatly restricted).

Although not yet fully developed and validated, this proof of concept method was applied to actual human liver microsome samples for which the genotype was known, in order to demonstrate that the correct metabolizer group could be inferred from the analysis. Although results were inconclusive for one sample due to sensitivity issues, results from four other samples (CYP 2D6 *4/*4, *4/*6, *1/*4, *1/*5) matched the expected pattern, qualitatively and quantitatively, thus demonstrating the usefulness of such an approach.

Estimating the metabolic capacity of an individual postmortem can be useful to the forensic toxicologist on two main levels. First, it can aid in distinguishing between an accidental overdose, a suicide attempt or a medical error, for example. Hence, it helps establish the circumstances surrounding death with more clarity. Second, it could help the forensic toxicologist to elucidate abnormal drug findings, e.g., outlier metabolite to parent drug ratios could be explained [10, 33, 34, 38, 39] when correlated to a metabolic capacity significantly different from average. At first sight, one attractive option would seem to be testing the collected liver tissue with metabolic assays using CYP specific probes [24]. Even though these probes are not circulating, as they would be in an antemortem assay, surely this methodology hits close to the desired target when it comes to metabolic capacity? The problem here is that enzymatic activity is quickly lost in the postmortem setting, due for example to changes in pH, ionic strength and depletion of cofactors [223]. Despite being inactive, proteins can still be intact in the studied tissue. This is what Hansen et al. found in their postmortem study of CYPs: although activity disappears within 1 day, relatively stable enzymes levels

are found for at least up to 7 days of storage [189].

The presented method is built around two main advantages. First, by taking into account not only the genotype of an individual (i.e., deleterious point mutations), but also the expression profile of an individual, the obtained result is closer to the phenotype, i.e., the actual metabolic capacity, than genotyping alone ever can be. A higher predictive value is therefore expected from this method. Second, the procedure uses, by design, tools and techniques present in the vast majority of forensic toxicology laboratories, such as LC-MS/MS. Taken together, these two advantages should lead to broader use of metabolic capacity determination by forensic toxicologists due to increased ease of analysis and relevance of the results.

As this project was started, some five papers described CYP absolute quantification methods [44, 47, 65, 67, 190] using unit resolution instruments [44, 47, 65, 67, 190]. Some quantified CYPs in solution [47, 65, 190], others in liver samples collected either antemortem or perimortem [44, 65, 67]. The proposed methodology was an in-solution digestion of postmortem liver microsomes followed by analysis on a unit resolution instrument. CYP 2D6, which was a common target of the literature [65, 67], was also selected for our purposes, as was CYP 3A4, which is less commonly analyzed but not completely novel either [44, 47, 190]. The largest step from the available literature was to be the (likely) increased complexity of postmortem samples, due to postmortem degradation of tissues (including proteins) [206].

Several years of experiments later, we find that the sample complexity issues, sensitivity and instrument fouling prevented completion of method development and successful validation. Moreover, as the method stands now, quantification results obtained differ for one quantotypic peptide to the next, even in relative amounts. Other tested quantification methods, such as standard addition, do not fare better. Since the SIL-IS peptides appear to be correcting properly for the matrix effects, the remaining logical explanation is therefore that digestion efficiency changes between samples. The method, in its current state, can give very rough approximations of the amount of enzyme present, especially in the extreme (absence/presence), but cannot be considered a tool for absolute quantification. This state of affairs is at odds with the initial assessment of the complexity of the project. The pertinent question is: where is this discrepancy stemming from?

First of all, it needs to be acknowledged that some elementary mistakes were due to the lack of experience in proteomics, even more so in mass spectrometry based quantitative

proteomics. For example, most synthetic peptides were dissolved in a 0.1% formic acid solution in water. The peptides were thus detectable by mass spectrometry, but there is a high likelihood that this solvent system did not completely dissolve the peptides, and the presence of an organic solvent would have been helpful. Another example is the centrifugal force used after digestion quenching ($1000 \times g$), was too low to attain the goal of precipitating proteins in solution [224]. One could also point out to the wide ion ratio criteria ($\pm$30%) applied for identification, which according to recent guidelines should be lowered to $\pm$20% [79]. Absolute quantitative proteomics, a relatively recent and wide-ranging domain has limited consolidated information, such as textbooks, available [225–227]. Papers, on the other hand, present information for those already expert in the area, which makes replication from non-specialized laboratories arduous as there seems to be information gaps. All in all, a collaboration with a group specialized in absolute quantitative proteomics, and potentially also with a group specialized in peptide and protein synthesis, would have been beneficial to the project. This would have allowed knowledge transfer, but also the use of specialized tools and techniques such as high resolution instrumentation with protein libraries, white rooms, sample cleanup techniques (both pre and post-digestion). For all of the other chapters of this thesis, such a collaboration existed with a group specialized in statistics, and these projects were much more successful.

Second, it is noteworthy that other CYP absolute quantification methods published either lack the required data to evaluate if similar problems to ours were observed, or the same issues occurred but were ignored. Matrix effects experiments are rarely reported; when they are, large matrix effects are indeed present [189]. Quantification is often done using a single quantotypic peptide [44, 47, 65]; when several are used and a discordant result is obtained, the odd peptide is removed from the method [67]. Digested standards or QCs are not always used, but are replaced by spiked solutions of peptides [47, 65]. Thus, nothing tells us we are experiencing issues which are out of the ordinary; rather, it might be that we are the first to specifically test for them or report them. This was not obvious after the first reading, especially as a newcomer to the domain. But when these papers are examined, using the harsh light of our findings, it becomes clear that there is serious work to be done to bring CYP analysis techniques to the point where they could be considered absolute quantification.

Knowing what we know today, how can we achieve our goal of quantification and characterization of CYP enzymes in the postmortem human liver tissue? This is a vast enterprise, which will likely require several years of work from several individuals in different areas of specialty. There is a need to start back to the simplest level, focus on small, well thought-out

designs to incrementally increase the complexity of the method. Considering the complexity and delicate nature of the absolute quantification problem, comparing and contrasting several different approaches will also be required, e.g., in solution microsomes vs gel electrophoresis isolated CYPs vs immunoaffinity pulldown, or internal calibration vs external calibration vs standard addition, etc.

First of all, it would be relevant to perform development operations on a high-resolution MS instrument. In this thesis, all experimental work was performed on a unit resolution instrument, first because a high-resolution MS was not available in the laboratory, second, the literature indicated that this was sufficient analytically, and third, this type of instrument is more widely available in forensic toxicology laboratories than a high-resolution instrument is. Our experience indicates that a high-resolution instrument would be useful in the development setting, if only to give more target flexibility (i.e., look at detergents, other proteins and molecules in addition to CYPs). Once the bulk of the method development is completed (e.g., preparation conditions fixed), then a transfer onto a unit resolution instrument could be done for dissemination into the forensic toxicology laboratories.

With this more versatile instrumentation in hand, and starting back at the simplest level, means developing a quantification and characterization method for recombinant CYP enzymes. With the goal of absolute quantification in mind, even at this stage, we want to obtain a complete digestion. A common strategy appears to be the digestion of samples with a constant protein concentration, obtained either by dilution of the sample or addition of an extraneous protein (e.g., bovine serum albumin) [64]. This allows the protein to protease ratio, as well as, hopefully, the digestion efficiency, to be stable. Digestion efficiency should be consistent and quantifiable, for example by the addition of a quality control with quantifiable, non-peptide tags [189]. Acceptability criteria will need to be set on digestion efficiency, and trend monitoring performed. In this thesis, no recovery experiments were performed – by the time the required materials (light and heavy peptides, proteins, etc.) were ordered and received, it had already become clear that there were issues at the sensitivity and quantification levels, thus rendering the recovery experiments moot. However, such recovery experiments are essential in a complete method assessment and validation. At this stage (working on recombinant enzymes in solution), if complete digestion is confirmed, the only recovery experiment required concerns peptides after the post-digestion clean-up process (either by SPE, as carried out in this thesis, or any other selected technique). Furthermore, whereas little matrix effect is expected at this stage (especially if a constant protein or peptide concentration is achieved in the sample), this factor should still be monitored and

satisfy standard method validation criteria.

Mass spectrometry analysis should be more expansive from that point on, keeping several options open. First, all detected peptides specific to a given CYP enzyme (not aliased with another human protein) should be kept, along with all the detected transitions from the peptides. This will allow flexibility later on in the development – for example, if one peptide or transition needs to be dropped due to matrix interferences, or if higher sensitivity transitions need to be dropped to bring relative uniformity to the signal intensities across peptides. Probable chemical modifications to the peptides should be monitored [228]. One might not want to immediately abandon peptides with chemical modifications, as this will likely change with the human matrix. With so many peptides and transitions to consider, using a unit resolution mass spectrometer would obviously be a stretch, bringing the cycle time up so much that the number of points per peak would dip under the acceptable minimum. However, a high-resolution instrument is suited for that type of work, particularly if one uses the data dependent or independent acquisition modes (rather than the reaction monitoring equivalent). The sensitivity of this type of instrument is typically a factor of 10 lower than a triple quadrupole instrument (for similar generations), but this can be construed as an advantage in the sense that transfer onto a unit resolution instrument should not generate sensitivity issues.

Multiplexing of the method, i.e., covering all relevant CYP enzymes and mutations, should be started right at this stage for a better global project efficiency. Overall there are a limited number of CYP enzymes relevant for forensic toxicology purposes: 10 enzymes, with half of them metabolizing more than 5% of drugs [14]. Not all of them are polymorphic [18],and none are polymorphic as CYP 2D6 is. In light of the very large number (>105) of polymorphism for this enzyme [18], it was decided to focus analysis on mutations present in $> 1\%$ of a given sub-population. Several mutations result in the absence of the enzyme from the liver tissue (e.g., gene deletion, splicing defects) [9], which are covered by the qualitative aspects of the method. Targeted mutations should be decided at this stage, which will require extensive literature research and would be a project in its own. It is definitely feasible to cover 10 enzymes in a single method, especially if all of them are included in the development from the first day.

Optimization will need to be performed at this early stage, at least for selected aspects of the procedure while some will need to be deferred until later e.g., microsome extraction. This is a complex operation, fraught with pitfalls. One of the pitfalls we did see and address

was the problematic use of a "one factor at a time" approach to optimization [201, 229]. This is why the more robust design of experiments (DoE) was selected. While this decision was correct, we encountered issues. First, large designs with several qualitative, multi-level factors are rather uninformative and tend to yield low quality models. Second, a great deal more thought than anticipated needs to go into defining the response variable(s) used for optimization. In this thesis, the large number of measurements generated by the analysis of a single sample (see Tables G.2.2 and G.3.2) were combined into a single value, the "Normalized Averaged Peak Area". This was an attempt to reduce the dimensionality of the data and achieve a conceptually simpler picture of the optimization process. However, in doing so, important pieces of information might be missed. First, individual peptides might express different response patterns to the studied factors. If these patterns go in opposite directions, the experimenter might decide that optimizing certain particular peptides (e.g., mutation site peptides) is more important than others (as long as they satisfy a minimum outcome). This can become a constrained optimization problem [201], and experience with these tools would need to be developed. Moreover, an important lesson of our work with DoE is "it does a wonderful optimization job, but be careful what you wish for". In other words, beyond the dimensionality of the data, the nature of the response variable must be carefully chosen. Here, peak areas were used as the underlying measure. However, this might not reflect, in all cases, an optimal method. For example, in Section G.2, the goal was to maximize the quantity of CYP denatured and solubilized. However, the optimal detergents in the design solubilized the CYPs but yielded ion suppression during the mass spectrometry analysis, thus erroneously orienting the investigation path. One could always argue that in some sense, since the MS signal is the final measurement, this is what we should be optimizing for, regardless of what causes it to vary. In any event, this needs to be carefully considered and reasoned: optimization for what exactly is desired, and what is the best measurement for this?

This method for CYPs in solution is the foundation of what needs to follow. The next step would be application to commercially available human liver microsomes. These samples, collected postmortem or perimortem, were found to be much cleaner than those isolated in-house from postmortem cases. This indicates that either the isolation technique used by this company performs better than what we have been using, or their samples are cleaner to begin with, possibly due to the absence, or removal, of degradation by-products. The digestion method should be optimized with these commercially available microsome samples. Complete digestion should be achieved, or, failing that, a means of compensating for it should be set up. Recovery, at least from the cleanup step (be it solid phase extraction or

any other mean) should be measured. Transitions representing undesired interferences can either be removed from the MS method or avoided by modifying the chromatography. At this point (and further along), the addition of an additional cleanup step such as gel electrophoresis can be considered. Ideally, we would like to avoid this kind of step: it is longer and more involved than manipulations in solutions and the addition of more steps inevitably leads to lower recovery levels. Also, forensic toxicology laboratories have no experience with this technique, and seeing it as part of this method might dissuade them for implementing it. Nonetheless, it is a powerful method of protein isolation and clean-up and if it is the solution that allows the whole setup to work, it would be worth it. In the end, a technique can be learned, and forensic toxicology laboratories are likely to want to analyze only a very limited percentage of their cases with this method, thus lessening the speed of production requirement.

In what could be a parallel project in itself, an isolation method of human liver microsomes from postmortem liver tissue should be developed. This could be based on human or animal human liver tissue, although given the wide availability of human liver tissue at forensics laboratories, there is no reason to go for an animal model when the target is actually human. The method presented in this thesis combines chemical and mechanical lysis to ultracentrifugation in order to isolate human liver microsomes. This is a standard approach in the literature [189], and is used by commercial companies producing human liver microsomes. However, it is time consuming, contains many steps potentially reducing the yield and operates with uncertain recovery. Also, it requires the use of an ultracentrifuge, a non-standard piece of equipment in a forensics lab. Hansen et al. presented an approach avoiding ultracentrifugation, though still proceeding by a lysis/centrifugation combination [189]. Immunological pulldown might be an interesting avenue. Following cell lysis, the microsomes, which are small micelles, are bound to magnetic particles via a CYP specific antibody and pulled out of the solution. Different lysis techniques and parameters (e.g., incubation time, chemicals concentration) should be investigated and optimized. In the end, one must always keep a balance between sample cleanliness and practicality. A practical outcome of this project would a catalogue of methods classified according to their different advantages and weaknesses. Should one of these methods fail at a later stage, e.g., due to lower than necessary sample cleanliness, we could always go back to the catalogues to pick another optimized method. In developing these microsome isolation techniques, the microsomal yield per gram of liver tissue should be recorded, as is standard in the literature [189]. This was not done in the method presented here, for different reasons. First, the liver weight is highly influenced by blood perfusion in the liver tissue, thus introducing a large uncertainty in the final yield

measurement. Second, this technique assumes that the microsomal protein per gram of liver (MPPGL) should be constant from one individual to the other, which does not seem to have been conclusively demonstrated, and is even more of a risky assumption for postmortem cases where degradation intervenes. However, it is a standard practice in the literature, and for better or worse, having this data on hand could help comparison with the current body of knowledge. Also, despite all its flaws, it could be a useful diagnostic tool. Choosing the optimization variable for the isolation technique is not a simple decision. MPPGL could be targeted, but this might encourage low purity samples with lots of extraneous proteins. If perimortem liver tissue is used, then CYP activity assays results could be used as the targeted variable. This is more involved, but a much better proxy for the desired goal. In a perfect world, we would be able to calculate and optimize for the recovery of CYPs (or microsomes alone) from the liver tissue. But this is a complex problem in and of itself.

Indeed, experiments to evaluate the absolute recovery of CYP enzymes – or even the microsomal fraction – from liver tissue will be extremely hard to design and execute. How can one generate liver tissue with known levels of CYP enzymes, or microsomal fraction? Absolute recovery figures rely on the assumption of this knowledge. Even recovery of CYPs from the microsomal fraction is problematic. In both cases, how do you either create a realistic matrix (liver tissue or microsomal fraction) with a known amount of CYPs? An alternative would be to measure the complete amount of CYPs left behind following the procedure, but how can you make sure this has been achieved and no CYPs are left behind? This is a serious challenge which might have no solution, in which case the only possible proxy would be relative recoveries (instead of absolute ones).

All of this preliminary work lays a solid basis for stepping over to the postmortem application of a quantification and characterization method of CYP enzymes in liver tissue. Here, a review of the already assessed parameters will be needed, to confirm that application to this type of matrix is still functional and implement any changes which might be required. Complete digestion will need to be confirmed. Recovery of the cleanup step should be measured. Above all, an evaluation of the matrix effect will be required. This was one of the major issues in the version of the method presented in this thesis.From the experimental work, no single component of the method was identified for the matrix effect. The most likely culprit, the detergent, was shown to elute when the flow was diverted to waste. A parent ion (Q1) scan of extracted samples, blanks and deconstructed blanks showed no single problematic source. A likely hypothesis is that the MS matrix effect stems from the sheer mass and number of concomitant peptides and small molecules injected. It probably

is, more globally, a sample purity issue. Of course, this would be easier to confirm using a high resolution mass spectrometer with spectral libraries that would allow identification of the different species. Principal component analysis techniques, for example, could be useful in identifying the key variables during purification optimization required to minimize these significant sources of matrix effects. If required, additional purification techniques such as gel electrophoresis could be implemented, as discussed above.

Once method development and optimization is completed, we will be able to utilize the mathematical method validation and production tools presented in Chapters 2 to 4. First, with the quantitative aspect of the method. Recombinant CYP enzymes spiked in authentic human liver matrix (whole tissue or microsomes) will be used to generate a biased calibration curve due to the endogenous CYP concentration present. Fortunately, as described in Chapter 3, we now have an automated tool to correct for the endogenous concentration in calibration matrices, albeit the special case of high level endogenous analytes will need to be considered further. A calibration model will need to be selected for this corrected calibration curve; and this can be done efficiently and accurately using the selection and validation tool presented in Chapter 2. As for the qualitative aspect of the method, that is, the detection of point mutation site bearing peptides, a coherent validation procedure is described in Section 4.1. This validation procedure can correctly evaluate the method's performance based on its natural output, the method's binary results. All of the required tools are thus at our disposal for the method validation process.

There is much more work to be done once this method is validated. First of all, it would be worth verifying the central premise of this project, i.e., that CYP protein abundance in liver tissue correlates with activity. Clearly, it holds true in the extremes: if the enzyme expression is entirely suppressed, no metabolism by this enzyme can occur and the metabolic capacity will be null. But, in-between this all or nothing, there remains to be conclusive proof of a tight correlation. There are studies demonstrating *in vitro* support for this theory [190, 230–233] but no *in vivo* confirmation as of yet. This should be studied in more detail. An animal model could be used, but this would mean redeveloping the method for the animal's specific CYP sequences; and even then, findings in animal models are not always transferable to humans. A more efficient approach would be the use of clinical or perimortem samples, where the metabolic capacity would be estimated both via the proteomics approach and the probe monitoring and/or activity approach.

There is a non-negligible probability that any clinical probe monitoring results would not

correlate that closely with the result of the proteomics approach, and part of the explanation would certainly be the high complexity of metabolism activities. There still is much to be learned about how metabolism proceeds. In recent years, it has been discovered, for example, that the role of active transporters is much bigger than initially thought [13]. These proteins are not included at all in the present project. Should they be, and how? It is hard to say at the moment, given the nascent literature on the topic. Nonetheless, we should keep in mind that the model presented here is one complexity step above the genotyping approach (and hopefully, one accuracy step above it as well, addressing some of the issues of correlation between proteotype and phenotype), but there might be several ones above that remain to be climbed as well.

Another area of concern is the spatial distribution of CYPs in the liver tissue. Although the metabolic capacity estimation presented in Section 5.8 assumes a uniform distribution by multiplying the full liver weight by the concentration measured in the sample analyzed, it is far from being certain that this is the case. There could be spatial patterns of concentration distribution, and they could even vary from one CYP to the other. In the postmortem case in particular, there could be spatial degradation patterns influencing our collection and sampling choices. Once a quantitative method is established, this becomes an easy hypothesis to test, analyzing several samples per liver from multiple individuals.

An additional potential layer of complexity are the impacts of post-translational modifications (PTMs). These chemical modifications occur after synthesis and can lead to changes in protein function (activity, half-life or localization in the cell) [228]. A large number of PTMs exist, such as acetylation, biotinylation, phosphorylation, sulfation, glycylation, sumoylation [228]. Of course, should such a PTM be necessary for CYP metabolic activity, or should another PTM remove all metabolic activity, this would have major impacts on the metabolic capacity. The active version(s) of the CYP would need to be distinguished from the inactive(s) one(s). PTMs are highly localized on the protein, thus taking them into account while maintaining good protein quantification practices would be challenging. Ideally, this information could be added to the current model, refining it further. Currently, the literature on PTMs in CYPs is extremely limited, relying on studies of a few enzymes only [234]. These early studies suggest that there might be a possible role of PTMs, but it is not fully characterized [234]. Hence, the current state of knowledge does not allow assessment of the role of PTMs or how this affects the feasibility of the proposed workflow. Note that testing the central premise of the project, the correlation between the CYP concentration and metabolic capacity, as suggested above, would at the same time answer the

PTMs question (i.e., in presence of a good correlation, PTMs are not a major concern).

Further work will also be required to examine postmortem stability of the enzymes in liver tissue. Preliminary work presented here [6], and elsewhere [189], indicated at least a few days of post-collection stability at 4 °C. Longer studies on larger populations and under more diverse postmortem conditions are indicated. We have carried out an *in vitro* study experiment at room temperature, but of course this is not representative of the conditions which will occur in casework. After death, the liver, protected by its outer envelope, undergoes a temperature drop and change in biochemical conditions [206] which is not only impossible to accurately simulate in the laboratory, but varies depending on the death scene (e.g., inside vs outside, different cause of death, different medical conditions). Potentially, this could be studied using samples collected on so called "body farms". This work should allow stability of enzymes during the postmortem interval to be established (i.e., the liver is still in the body on the scene of death), while the body is stored in the morgue (which generally does not last more than one week), and in collected liver tissue stored cold (4 °C, −20 °C and −80 °C). Should there be a desire to extend the time window of applicability beyond the measured stability delay, investigation into degradation markers for CYP enzymes could be carried out. This would involve pinpointing the degradation products of the CYP enzymes, and from the concentration of these degradation products, performing a retro-calculation of the original enzyme concentration. This is highly exploratory work, and there is a high probability that it would not be achievable, or at least would not widen the time window as much as desired. Nonetheless, this type of knowledge would be highly transferable to other postmortem proteomics methods.

The only remaining piece then is to carry out larger population studies. In the current genotyping framework, individuals are classified according to metabolizer type [9], a categorical classification. The developed methodology would yield a continuous estimate of the metabolic capacity, allowing finer comparison of metabolism between individuals. A population study would establish a metabolic capacity distribution curve for the racial and socio-economic population served by the forensics laboratory. Incoming cases could then be compared to the distribution curve to place the metabolic capacity for the studied case within the larger population context: does the individual in this case have a metabolic capacity significantly higher or lower than the population? More broadly, what is significant in terms of metabolic capacity? These sorts of questions can only begin to be addressed once a reliable analytical method is in place and has been adopted by multiple labs.

A posteriori, it is clear that there was an underestimation of the complexity of the overall project. This was in part due to incomplete data in the literature, and in part due to the lack of experience in quantitative proteomics. The acute preoccupation to develop a simple method (simple sample preparation, basic instrumentation) that the majority of forensic toxicology laboratories could apply easily unintentionally compounded this problem. It is important to think about applicability and diffusion of the techniques, but in the pyramid of needs, it comes after a fully functional method. The proteomics approach is complex, more than initially thought, but it holds great promise. If properly developed and validated, it would take into account several epigenetic factors that affect the protein abundance, such as the gender, age and disease state. Incorporation of these factors measured into the estimation of the metabolic capacity would be more accurate than by genotyping alone. To reach this goal however, a large amount of work is still required.

6.5 Bringing it all together

If all of these projects teach us one thing, it is that problems are generally less trivial than we would like to believe on first approach. Correcting for endogenous concentration is easy! Yes, but how do you integrate that seamlessly in production operations? And what additional error have you introduced? Selecting a calibration model is elementary! You just pick the simplest or use one of the numerous tests available! Okay, but are those tests actually appropriate for your data set? Are they actually better than your subjective evaluation? How do you deal with multiple instruments? What experimental setup should you use? Qualitative methods are simple to validate! No, not if done correctly. You need more data than your preconceived ideas might lead you to believe. And even while developing a validation framework for binary results might seem relatively elementary, an in-depth study of real data brings surprises that are not so simple to tackle: heteroscedasticity, sample cut-off and random instrument shifts hidden within the changing variance of the area ratio. CYP quantification methods have been published, only adjustment to the postmortem setting remains! Well, several central parameters were overlooked in these published methods and pose problems for appropriate method development and validation.

The interfaces probed in this thesis are still wide open for exploration and discovery. In retrospect, the one overarching theme of this thesis has been to aid the forensic toxicologist in achieving the fundamental goal to help explaining the circumstances surrounding death and shedding light on questions arising during judicial proceedings (driving under the influence,

sexual assault). This cannot be accomplished without also simultaneously incorporating the fundamental basis of science: make judgements based on evidence from which the degree of certainty has been quantified. Which interestingly, is reminiscent of the motto of Dr. Wilfrid Derome (1877–1931), founder of the Laboratoire de sciences judiciaires et de médecine légale: “n’avance rien que tu ne sois capable de prouver” [235] (“do not state anything that you can’t prove”).

References

[1] B. Desharnais, F. Camirand-Lemyre, P. Mireault, C. D. Skinner, Procedure for the selection and validation of a calibration model I — Description and application, Journal of Analytical Toxicology 41 (4) (2017) 261–268. doi:10.1093/jat/bkx001.

[2] B. Desharnais, F. Camirand-Lemyre, P. Mireault, C. D. Skinner, Procedure for the selection and validation of a calibration model II — Theoretical basis, Journal of Analytical Toxicology 41 (4) (2017) 269–276. doi:10.1093/jat/bkx002.

[3] B. Desharnais, M.-J. Lajoie, J. Laquerre, S. Savard, P. Mireault, C. D. Skinner, A tool for automatic correction of endogenous concentrations: application to BHB analysis by LC–MS-MS and GC-MS, Journal of Analytical Toxicology 43 (7) (2019) 512–519. doi:10.1093/jat/bkz024.

[4] F. Camirand Lemyre, B. Desharnais, J. Laquerre, M.-A. Morel, C. Côté, P. Mireault, C. D. Skinner, Qualitative method validation and uncertainty evaluation via the binary output I — Validation guidelines and theoretical foundations, Journal of Analytical Toxicology (2019) submission number JAT–19–2881.

[5] B. Desharnais, M.-J. Lajoie, J. Laquerre, P. Mireault, C. D. Skinner, Qualitative method validation and uncertainty estimation via the binary output II — Application to a multi-analyte LC-MS/MS method for oral fluid, Journal of Analytical Toxicology (2019) submission number JAT–19–2882.

[6] B. Desharnais, P. Mireault, C. D. Skinner, Postmortem estimation of metabolic capacity through cytochrome P450 enzyme characterisation and quantification — A proof of concept.

[7] A. Moffat, M. Osselton, B. Widdop, S. Jickells, A. Negrusz, Introduction to forensic toxicology, in: A. Negrusz, G. Cooper (Eds.), Clarke's Analytical Forensic Toxicology, 2nd Edition, Pharmaceutical Press, London, United Kingdom, 2013, Ch. 1, pp. 1–10.

[8] B. A. Goldberger, D. Lee, D. G. Wilkins, Analytical and forensic toxicology, in: C. D. Klaassen (Ed.), Casarett & Doull's Toxicology: The Basic Science of Poisons, 9th Edition, McGraw-Hill Education / Medical, New York, United States, 2018, Ch. 32, pp. 1511–1529.

[9] H. Druid, P. Holmgren, B. Carlsson, J. Ahlner, Cytochrome P450 2D6 (CYP2D6) genotyping on postmortem blood as a supplementary tool for interpretation of forensic

toxicological results, Forensic Science International 99 (1) (1999) 25–34. doi:10.1016/S0379-0738(98)00169-8.

[10] S. H. Wong, M. A. Wagner, J. M. Jentzen, C. Schur, J. Bjerke, S. B. Gock, C.-C. Chang, Pharmacogenomics as an aspect of molecular autopsy for forensic pathology/toxicology: does genotyping CYP 2D6 serve as an adjunct for certifying methadone toxicity?, Journal of Forensic Sciences 48 (6) (2003) 1406–1415. doi:10.1520/JFS2002392.

[11] A. Sajantila, J. Palo, I. Ojanperä, C. Davis, B. Budowle, Pharmacogenetics in medicolegal context, Forensic Science International 203 (1-3) (2010) 44–52. doi:10.1016/j.forsciint.2010.09.011.

[12] P. Wexler, A. N. Hayes, The evolving journey of toxicology: A historical glimpse, in: C. D. Klaassen (Ed.), Casarett & Doull's Toxicology: The Basic Science of Poisons, 9th Edition, McGraw-Hill Education / Medical, New York, United States, 2018, Ch. 1, pp. 3–23.

[13] A. Parkinson, B. W. Ogilvie, D. B. Buckley, F. Kazmi, O. Parkinson, Biotransformation of xenobiotics, in: C. D. Klaassen (Ed.), Casarett & Doull's Toxicology: The Basic Science of Poisons, 9th Edition, McGraw-Hill Education / Medical, New York, United States, 2018, Ch. 6, pp. 193–399.

[14] U. M. Zanger, M. Schwab, Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation, Pharmacology & Therapeutics 138 (1) (2013) 103–141. doi:10.1016/j.pharmthera.2012.12.007.

[15] A. DePriest, B. Puet, A. Holt, A. Roberts, E. Cone, Metabolism and disposition of prescription opioids: a review, Forensic Science Review 27 (2) (2015) 115–145.

[16] M. Ingelman-Sundberg, Human drug metabolising cytochrome P450 enzymes: properties and polymorphisms, Naunyn-Schmiedeberg's Archives of Pharmacology 369 (1) (2004) 89–104. doi:10.1007/s00210-003-0819-z.

[17] P. Holmgren, J. Ahlner, Pharmacogenomics for forensic toxicology: Swedish experience, in: S. H. Y. Wong, M. W. Linder, R. Valdes Jr. (Eds.), Pharmacogenomics Proteomics Enabling the Practice of Personalized Medicine, AACC Press, Washington, United States, 2006, Ch. 28, pp. 295–299.

[18] A. Gaedigk, M. Ingelman-Sundberg, N. A. Miller, J. S. Leeder, M. Whirl-Carrillo, T. E. Klein, P. S. Committee, The Pharmacogene Variation (PharmVar) Consortium: incorporation of the human cytochrome P450 (CYP) allele nomenclature database, Clinical Pharmacology & Therapeutics 103 (3) (2018) 399–401. doi:10.1002/cpt.910.

[19] C. M. Hunt, W. R. Westerkam, G. M. Stave, Effect of age and gender on the activity of human hepatic CYP3A, Biochemical Pharmacology 44 (2) (1992) 275–283. doi:10.1016/0006-2952(92)90010-G.

[20] Z. E. Barter, J. E. Chowdry, J. R. Harlow, J. E. Snawder, J. C. Lipscomb, A. Rostami-Hodjegan, Covariation of human microsomal protein per gram of liver with age: absence of influence of operator and sample storage may justify interlaboratory data pooling, Drug Metabolism and Disposition 36 (12) (2008) 2405–2409. doi:10.1124/dmd.108.021311.

[21] R. Wolbold, K. Klein, O. Burk, A. K. Nüssler, P. Neuhaus, M. Eichelbaum, M. Schwab, U. M. Zanger, Sex is a major determinant of CYP3A4 expression in human liver, Hepatology 38 (4) (2003) 978–988. doi:10.1002/hep.1840380424.

[22] Y. Shao, X. Yin, D. Kang, B. Shen, Z. Zhu, X. Li, H. Li, L. Xie, G. Wang, Y. Liang, An integrated strategy for the quantitative analysis of endogenous proteins: A case of gender-dependent expression of P450 enzymes in rat liver microsome, Talanta 170 (2017) 514–522. doi:10.1016/j.talanta.2017.04.050.

[23] B. Madea, P. Saukko, A. Oliva, F. Musshoff, Molecular pathology in forensic medicine – introduction, Forensic Science International 203 (1-3) (2010) 3–14. doi:10.1016/j.forsciint.2010.07.017.

[24] M. Bosilkovska, C. Samer, J. Déglon, A. Thomas, B. Walder, J. Desmeules, Y. Daali, Evaluation of mutual drug–drug interaction within Geneva Cocktail for cytochrome P450 phenotyping using innovative dried blood sampling method, Basic & clinical pharmacology & toxicology 119 (3) (2016) 284–290. doi:10.1111/bcpt.12586.

[25] D. W. Nebert, Role of genetics and drug metabolism in human cancer risk, Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 247 (2) (1991) 267–281. doi:10.1016/0027-5107(91)90022-G.

[26] N. Djordjevic, D. D. Milovanovic, M. Radovanovic, I. Radosavljevic, S. Obradovic, M. Jakovljevic, D. Milovanovic, J. R. Milovanovic, S. Jankovic, CYP1A2 genotype affects carbamazepine pharmacokinetics in children with epilepsy, European Journal of Clinical Pharmacology 72 (4) (2016) 439–445. doi:10.1007/s00228-015-2006-9.

[27] M. Nakajima, R. Yoshida, T. Fukami, H. L. McLeod, T. Yokoi, Novel human CYP2A6 alleles confound gene deletion analysis, FEBS Letters 569 (1-3) (2004) 75–81. doi:10.1016/j.febslet.2004.05.053.

[28] C.-C. Chang, P.-C. Lin, C.-H. Lin, K.-T. Yeh, H.-Y. Hung, J.-G. Chang, Rapid identification of CYP2C8 polymorphisms by high resolution melting analysis, Clinica Chimica Acta 413 (1-2) (2012) 298–302. doi:10.1016/j.cca.2011.10.005.

[29] B. D. Swar, S. R. Bendkhale, A. Rupawala, K. Sridharan, N. J. Gogtay, U. M. Thatte, N. A. Kshirsagar, Evaluation of cytochrome P450 2C9 activity in normal, healthy, adult Western Indian population by both phenotyping and genotyping, Indian Journal of Pharmacology 48 (3) (2016) 248. doi:10.3109/09537104.2015.1095875.

[30] J. H. Lee, S. G. Ahn, J.-W. Lee, Y. J. Youn, M.-S. Ahn, J.-Y. Kim, B.-S. Yoo, S.-H. Lee, J. Yoon, J. Kim, et al., Switching from prasugrel to clopidogrel based

on Cytochrome P450 2C19 genotyping in East Asian patients stabilized after acute myocardial infarction, Platelets 27 (4) (2016) 301–307. doi:10.3109/09537104.2015.1095875.

[31] S. Ben, R. M. Cooper-DeHoff, H. K. Flaten, O. Evero, T. M. Ferrara, R. A. Spritz, A. A. Monte, Multiplex SNaPshot—a new simple and efficient CYP2D6 and ADRB1 genotyping method, Human Genomics 10 (1) (2016) 11. doi:10.1186/s40246-016-0073-3.

[32] C. Innocenti, A. Accorsi, V. Cerreta, V. Mantovani, F. Violante, Fast CYP2E1 genotyping using automated fluorescent detection, La Medicina del Lavoro 97 (6) (2006) 799–804.

[33] M. Jin, S. B. Gock, P. J. Jannetto, J. M. Jentzen, S. H. Wong, Pharmacogenomics as molecular autopsy for forensic toxicology: genotyping cytochrome P450 3A4* 1B and 3A5* 3 for 25 fentanyl cases, Journal of Analytical Toxicology 29 (7) (2005) 590–598. doi:10.1093/jat/29.7.590.

[34] J. Frost, A. Helland, I. S. Nordrum, L. Slørdal, Investigation of morphine and morphine glucuronide levels and cytochrome P450 isoenzyme 2D6 genotype in codeine-related deaths, Forensic Science International 220 (1-3) (2012) 6–11. doi:10.1016/j.forsciint.2012.01.019.

[35] Y. He, J. Brockmöller, H. Schmidt, I. Roots, J. Kirchheiner, CYP2D6 ultrarapid metabolism and morphine/codeine ratios in blood: was it codeine or heroin?, Journal of Analytical Toxicology 32 (2) (2008) 178–182. doi:10.1093/jat/32.2.178.

[36] M. Kingbäck, L. Karlsson, A.-L. Zackrisson, B. Carlsson, M. Josefsson, F. Bengtsson, J. Ahlner, F. C. Kugelberg, Influence of CYP2D6 genotype on the disposition of the enantiomers of venlafaxine and its major metabolites in postmortem femoral blood, Forensic Science International 214 (1-3) (2012) 124–134. doi:10.1016/j.forsciint.2011.07.034.

[37] A. Levo, A. Koski, I. Ojanperä, E. Vuori, A. Sajantila, Post-mortem SNP analysis of CYP2D6 gene reveals correlation between genotype and opioid drug (tramadol) metabolite ratios in blood, Forensic Science International 135 (1) (2003) 9–15. doi:10.1016/S0379-0738(03)00159-2.

[38] P. Holmgren, B. Carlsson, A.-L. Zackrisson, B. Lindblom, M.-L. Dahl, M. G. Scordo, H. Druid, J. Ahlner, Enantioselective analysis of citalopram and its metabolites in postmortem blood and genotyping for CYD2D6 and CYP2C19, Journal of Analytical Toxicology 28 (2) (2004) 94–104. doi:10.1093/jat/28.2.94.

[39] P. J. Jannetto, S. H. Wong, S. B. Gock, E. Laleli-Sahin, B. C. Schur, J. M. Jentzen, Pharmacogenomics as molecular autopsy for postmortem forensic toxicology: genotyping cytochrome P450 2D6 for oxycodone cases, Journal of Analytical Toxicology 26 (7) (2002) 438–447. doi:10.1093/jat/26.7.438.

[40] G. Koren, J. Cairns, D. Chitayat, A. Gaedigk, S. J. Leeder, Pharmacogenetics of morphine poisoning in a breastfed neonate of a codeine-prescribed mother, The Lancet 368 (2006) 704. doi:10.1016/S0140-6736(06)69255-6.

[41] F. R. Sallee, C. L. DeVane, R. E. Ferrell, Fluoxetine-related death in a child with cytochrome P-450 2D6 genetic deficiency, Journal of Child and Adolescent Psychopharmacology 10 (1) (2000) 27–34. doi:10.1089/cap.2000.10.27.

[42] A. Koski, I. Ojanperä, J. Sistonen, E. Vuori, A. Sajantila, A fatal doxepin poisoning associated with a defective CYP2D6 genotype, The American Journal of Forensic Medicine and Pathology 28 (3) (2007) 259–261. doi:10.1097/PAF.0b013e3180326701.

[43] R. D. Dowell, O. Ryan, A. Jansen, D. Cheung, S. Agarwala, T. Danford, D. A. Bernstein, P. A. Rolfe, L. E. Heisler, B. Chin, et al., Genotype to phenotype: a complex problem, Science 328 (5977) (2010) 469. doi:10.1126/science.1189015.

[44] C. Seibert, B. R. Davidson, B. J. Fuller, L. H. Patterson, W. J. Griffiths, Y. Wang, Multiple-approaches to the identification and quantification of cytochromes P450 in human liver tissue by mass spectrometry, Journal of Proteome Research 8 (4) (2008) 1672–1681. doi:10.1021/pr800795r.

[45] W. Crochot, Diagram of the Cell Membrane's Structures and Their Funtion, Figure (2014).
URL https://en.wikipedia.org/wiki/Cell_membrane#/media/File:Cell_membrane_drawing-en.svg

[46] P. Picotti, R. Aebersold, Selected reaction monitoring–based proteomics: workflows, potential, pitfalls and future directions, Nature Methods 9 (6) (2012) 555.

[47] M. Z. Wang, J. Q. Wu, J. B. Dennison, A. S. Bridges, S. D. Hall, S. Kornbluth, R. R. Tidwell, P. C. Smith, R. D. Voyksner, M. F. Paine, et al., A gel-free MS-based quantitative proteomic approach accurately measures cytochrome P450 protein concentrations in human liver microsomes, Proteomics 8 (20) (2008) 4186–4196. doi:10.1002/pmic.200800144.

[48] B. Rathgeber, J. Boles, P. Shand, Rapid postmortem pH decline and delayed chilling reduce quality of turkey breast meat, Poultry Science 78 (3) (1999) 477–484. doi:10.1093/ps/78.3.477.

[49] Traditional Methods of Cell Lysis, Handbook, ThermoFisher Scientific (2019).
URL https://www.thermofisher.com/ca/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/traditional-methods-cell-lysis.html

[50] Protein Preparation Handbook, Handbook, ThermoFisher Scientific (2016).
URL https://assets.thermofisher.com/TFS-Assets/BID/Handbooks/protein-preparation-handbook.pdf

[51] M. K. Rasmussen, B. Ekstrand, G. Zamaratskaia, Comparison of cytochrome P450 concentrations and metabolic activities in porcine hepatic microsomes prepared with two different methods, Toxicology In Vitro 25 (1) (2011) 343–346. doi:10.1016/j.tiv.2010.10.007.

[52] Y. Nakatani, V. Ogryzko, Immunoaffinity purification of mammalian protein complexes, Methods in Enzymology 370 (2003) 430–444. doi:10.1016/S0076-6879(03)70037-8.

[53] Y. Hatefi, W. Hanstein, Solubilization of particulate proteins and nonelectrolytes by chaotropic agents, Proceedings of the National Academy of Sciences 62 (4) (1969) 1129–1136. doi:10.1073/pnas.62.4.1129.

[54] F. Lottspeich, Top down and bottom up analysis of proteins (focusing on quantitative aspects), in: T. Letzel (Ed.), Protein and Peptide Analysis by LC-MS: Experimental Strategies, Royal Society of Chemistry, London, United Kingdom, 2011, Ch. 1, pp. 1–10. doi:10.1039/9781849733144-00001.

[55] J. Barbour, S. Wiese, H. E. Meyer, B. Warscheid, Mass spectrometry, in: J. von Hagen (Ed.), Proteomics Sample Preparation, Wiley, Hoboken, United States, 2008, Ch. 4, pp. 41–128. doi:10.1002/9783527622832.ch4.

[56] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel, A. Bairoch, ExPASy: the proteomics server for in-depth protein knowledge and analysis, Nucleic Acids Research 31 (13) (2003) 3784–3788. doi:10.1093/nar/gkg563.

[57] D. A. Skoog, F. J. Holler, S. R. Crouch, Liquid chromatography, in: Principles of Instrumental Analysis, 6th Edition, Thomson Higher Education, Belmont, United States, 2006, Ch. 28, pp. 816–855.

[58] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, C. M. Whitehouse, Electrospray ionization–principles and practice, Mass Spectrometry Reviews 9 (1) (1990) 37–70. doi:10.1002/mas.1280090103.

[59] V. Lange, P. Picotti, B. Domon, R. Aebersold, Selected reaction monitoring for quantitative proteomics: a tutorial, Molecular Systems Biology 4 (1) (2008) 222. doi:10.1038/msb.2008.61.

[60] A. A. Dowle, J. Wilson, J. R. Thomas, Comparing the diagnostic classification accuracy of iTRAQ, peak-area, spectral-counting, and emPAI methods for relative quantification in expression proteomics, Journal of Proteome Research 15 (10) (2016) 3550–3562. doi:10.1021/acs.jproteome.6b00308.

[61] C. Lindemann, N. Thomanek, F. Hundt, T. Lerari, H. E. Meyer, D. Wolters, K. Marcus, Strategies in relative and absolute quantitative mass spectrometry based proteomics, Biological Chemistry 398 (5-6) (2017) 687–699. doi:10.1515/hsz-2017-0104.

[62] C. Ludwig, R. Aebersoldab, Getting Absolute: Determining Absolute Protein Quantities via Selected Reaction Monitoring Mass Spectrometry, in: C. E. Eyers, S. Gaskell (Eds.), Quantitative Proteomics, Royal Society of Chemistry, London, United Kingdom, 2014, Ch. 4, pp. 80–109. doi:10.1039/9781782626985-00080.

[63] S. A. Gerber, J. Rush, O. Stemman, M. W. Kirschner, S. P. Gygi, Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS, Proceedings of the National Academy of Sciences 100 (12) (2003) 6940–6945. doi:10.1073/pnas.0832254100.

[64] C. M. Shuford, J. J. Walters, P. M. Holland, U. Sreenivasan, N. Askari, K. Ray, R. P. Grant, Absolute protein quantification by mass spectrometry: not as simple as advertised, Analytical Chemistry 89 (14) (2017) 7406–7415. doi:10.1021/acs.analchem.7b00858.

[65] H. Kawakami, S. Ohtsuki, J. Kamiie, T. Suzuki, T. Abe, T. Terasaki, Simultaneous absolute quantification of 11 cytochrome P450 isoforms in human liver microsomes by liquid chromatography tandem mass spectrometry with in silico target peptide selection, Journal of Pharmaceutical Sciences 100 (1) (2011) 341–352. doi:10.1002/jps.22255.

[66] C. Lane, S. Nisar, W. Griffiths, B. Fuller, B. Davidson, J. Hewes, K. Welham, L. Patterson, Identification of cytochrome P450 enzymes in human colorectal metastases and the surrounding liver: a proteomic approach, European Journal of Cancer 40 (14) (2004) 2127–2134. doi:10.1016/j.ejca.2004.04.029.

[67] E. Langenfeld, U. M. Zanger, K. Jung, H. E. Meyer, K. Marcus, Mass spectrometry-based absolute quantification of microsomal cytochrome P450 2D6 in human liver, Proteomics 9 (9) (2009) 2313–2323. doi:10.1002/pmic.200800680.

[68] R. E. Jenkins, N. R. Kitteringham, C. L. Hunter, S. Webb, T. J. Hunt, R. Elsby, R. B. Watson, D. Williams, S. R. Penningtonand, B. K. Park, Relative and absolute quantitative expression profiling of cytochromes P450 using isotope-coded affinity tags, Proteomics 6 (6) (2006) 1934–1947. doi:10.1002/pmic.200500432.

[69] C. S. Lane, Y. Wang, R. Betts, W. J. Griffiths, L. H. Patterson, Comparative cytochrome P450 proteomics in the livers of immunodeficient mice using ^{18}O stable isotope labeling, Molecular & Cellular Proteomics 6 (6) (2007) 953–962. doi:10.1074/mcp.M600296-MCP200.

[70] B. Achour, J. Barber, A. Rostami-Hodjegan, Cytochrome P450 pig liver pie: determination of individual cytochrome P450 isoform contents in microsomes from two pig livers using liquid chromatography in conjunction with mass spectrometry, Drug Metabolism and Disposition 39 (11) (2011) 2130–2134. doi:10.1124/dmd.111.040618.

[71] General requirements for the competence of testing and calibration laboratories, Standard ISO/IEC 17025:2005, International Organization for Standardization, Geneva,

Switzerland (2005).
URL https://www.iso.org/standard/39883.html

[72] General requirements for the competence of testing and calibration laboratories, Standard ISO/IEC 17025:2017, International Organization for Standardization, Geneva, Switzerland (2017).
URL https://www.iso.org/standard/66912.html

[73] Guidelines for the Accreditation of Forensic Testing Laboratories, Standard CAN-P-1578, Standards Council of Canada (Conseil canadien des normes), Ottawa, Canada (2009).
URL https://www.scc.ca/en/about-scc/publications/scc-requirements-and-guidance-for-accreditation-for-forensic-testing-laboratories

[74] D. A. Skoog, F. J. Holler, S. R. Crouch, Introduction, in: Principles of Instrumental Analysis, 6th Edition, Thomson Higher Education, Belmont, United States, 2006, Ch. 1, pp. 1–24.

[75] Scientific Working Group for Forensic Toxicology, Scientific Working Group for Forensic Toxicology (SWGTOX) standard practices for method validation in forensic toxicology, Journal of Analytical Toxicology 37 (7) (2013) 452–474. doi:10.1093/jat/bkt054.

[76] Gold Book - Limit of Detection, Standard, International Union of Pure and Applied Chemistry, Research Triangle Park, United States (2014).
URL https://goldbook.iupac.org/terms/view/L03540

[77] M. Haenlein, A. M. Kaplan, A beginner's guide to partial least squares analysis, Understanding Statistics 3 (4) (2004) 283–297. doi:10.1207/s15328031us0304_4.

[78] G. R. Jones, Postmortem toxicology, in: A. Negrusz, G. Cooper (Eds.), Clarke's Analytical Forensic Toxicology, 2nd Edition, Pharmaceutical Press, London, United Kingdom, 2013, Ch. 7, pp. 189–214.

[79] Bioanalytical Method Validation – Guidance for Industry, Standard, Food and Drug Administration, Silver Springs, United States (May 2018).
URL http://www.fda.gov/downloads/Drugs/Guidances/ucm070107.pdf

[80] Guideline on Bioanalytical Method Validation, Standard, European Medicines Agency, London, United Kingdom (2011).
URL https://www.ema.europa.eu/documents/scientific-guideline/guideline-bioanalytical-method-validation_en.pdf

[81] Standard Practices for Method Validation in Forensic Toxicology (Draft), Standard, American Academy of Forensic Sciences Standards Board, Colorado Springs, United States (2018).
URL https://asb.aafs.org/wp-content/uploads/2018/09/036_Std_Ballot02.pdf

[82] E. Stokvis, H. Rosing, J. H. Beijnen, Stable isotopically labeled internal standards in quantitative bioanalysis using liquid chromatography/mass spectrometry: necessity or not?, Rapid Communications in Mass Spectrometry 19 (3) (2005) 401–407. doi:10.1002/rcm.1790.

[83] M. S. Halquist, H. T. Karnes, Quantification of Alefacept, an immunosuppressive fusion protein in human plasma using a protein analogue internal standard, trypsin cleaved signature peptides and liquid chromatography tandem mass spectrometry, Journal of Chromatography B 879 (11–12) (2011) 789–798. doi:10.1016/j.jchromb.2011.02.034.

[84] J. D. Ingle, S. R. Crouch, Signal-to-noise ratio considerations, in: Spectrochemical Analysis, Prentice Hall, Englewood Cliffs, United States, 1988, Ch. 5, pp. 135–163.

[85] D. Massart, B. Vandeginste, L. Buydens, S. De Jong, P. Lewi, J. Smeyers-Verbeke, Straight line regression and calibration, in: Handbook of Chemometrics and Qualimetrics: Part A, Vol. 20A of Data Handling in Science and Technology, Elsevier, Amsterdam, Netherlands, 1997, Ch. 8, pp. 171–230.

[86] D. Massart, B. Vandeginste, L. Buydens, S. De Jong, P. Lewi, J. Smeyers-Verbeke, Multiple and polynomial regression, in: Handbook of Chemometrics and Qualimetrics: Part A, Vol. 20A of Data Handling in Science and Technology, Elsevier, Amsterdam, Netherlands, 1997, Ch. 10, pp. 263–303.

[87] K. Tang, J. S. Page, R. D. Smith, Charge competition and the linear dynamic range of detection in electrospray ionization mass spectrometry, Journal of the American Society for Mass Spectrometry 15 (10) (2004) 1416–1423. doi:10.1016/j.jasms.2004.04.034.

[88] H. Gu, G. Liu, J. Wang, A.-F. Aubry, M. E. Arnold, Selecting the correct weighting factors for linear and quadratic calibration curves with least-squares regression algorithm in bioanalytical LC-MS/MS assays and impacts of using incorrect weighting factors on curve stability, data quality, and assay performance, Analytical Chemistry 86 (18) (2014) 8959–8966. doi:10.1021/ac5018265.

[89] M. A. Babyak, What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models, Psychosomatic Medicine 66 (3) (2004) 411–421.

[90] F. P Busardo, A. W Jones, GHB pharmacology and toxicology: acute intoxication, concentrations in blood and urine in forensic cases and treatment of the withdrawal syndrome, Current Neuropharmacology 13 (1) (2015) 47–70.

[91] S. M. Darby, M. L. Miller, R. O. Allen, M. LeBeau, A mass spectrometric method for quantitation of intact insulin in blood samples, Journal of Analytical Toxicology 25 (1) (2001) 8–14. doi:10.1093/jat/25.1.8.

[92] Y. Lood, A. Eklund, M. Garle, J. Ahlner, Anabolic androgenic steroids in police cases in Sweden 1999–2009, Forensic Science International 219 (1–3) (2012) 199–204. doi:10.1016/j.forsciint.2012.01.004.

[93] C. Hess, K. Sydow, T. Kueting, M. Kraemer, A. Maas, Considerations regarding the validation of chromatographic mass spectrometric methods for the quantification of endogenous substances in forensics, Forensic Science International 283 (2018) 150–155. doi:10.1016/j.forsciint.2017.12.019.

[94] S. M. Wille, F. T. Peters, V. Di Fazio, N. Samyn, Practical aspects concerning validation and quality control for forensic and clinical bioanalytical quantitative methods, Accreditation and Quality Assurance 16 (6) (2011) 279. doi:10.1007/s00769-011-0775-0.

[95] S. P. Elliott, Gamma hydroxybutyric acid (GHB) concentrations in humans and factors affecting endogenous production, Forensic Science International 133 (1-2) (2003) 9–16. doi:10.1016/S0379-0738(03)00043-4.

[96] M. Jemal, A. Schuster, D. B. Whigan, Liquid chromatography/tandem mass spectrometry methods for quantitation of mevalonic acid in human plasma and urine: method validation, demonstration of using a surrogate analyte, and demonstration of unacceptable matrix effect in spite of use of a stable isotope analog internal standard, Rapid Communications in Mass Spectrometry 17 (15) (2003) 1723–1734. doi:10.1002/rcm.1112.

[97] D. Ji, C.-G. Jang, S. Lee, A sensitive and accurate quantitative method to determine N-arachidonoyldopamine and N-oleoyldopamine in the mouse striatum using column-switching LC–MS–MS: use of a surrogate matrix to quantify endogenous compounds, Analytical and Bioanalytical Chemistry 406 (18) (2014) 4491–4499. doi:10.1007/s00216-014-7816-6.

[98] B. R. Jones, G. A. Schultz, J. A. Eckstein, B. L. Ackermann, Surrogate matrix and surrogate analyte approaches for definitive quantitation of endogenous biomolecules, Bioanalysis 4 (19) (2012) 2343–2356. doi:10.4155/bio.12.200.

[99] T. M. Binz, U. Braun, M. R. Baumgartner, T. Kraemer, Development of an LC–MS/MS method for the determination of endogenous cortisol in hair using $^{13}C_3$-labeled cortisol as surrogate analyte, Journal of Chromatography B 1033 (2016) 65–72. doi:10.1016/j.jchromb.2016.07.041.

[100] S. Kang, S. M. Oh, K. H. Chung, S. Lee, A surrogate analyte-based LC–MS/MS method for the determination of γ-hydroxybutyrate (GHB) in human urine and variation of endogenous urinary concentrations of GHB, Journal of Pharmaceutical and Biomedical Analysis 98 (2014) 193–200. doi:10.1016/j.jpba.2014.05.028.

[101] D. A. Skoog, F. J. Holler, S. R. Crouch, Selecting an analytical method, in: Principles of Instrumental Analysis, 6th Edition, Thomson Higher Education, Belmont, United States, 2006, Ch. 1E, pp. 17–21.

[102] D. Hughes, The world anti-doping code in sport: update for 2015, Australian Prescriber 38 (5) (2015) 167–170. doi:10.18773/austprescr.2015.059.

[103] F. P. Carvalho, Pesticides, environment, and food safety, Food and Energy Security 6 (2) (2017) 48–60. doi:10.1002/fes3.108.

[104] An Act to amend the Criminal Code (offences relating to conveyances) and to make consequential amendments to other Acts, Legislation S.C. 2018, c. 21, Bill C-46, Government of Canada, Ottawa, Canada (2018).
URL http://www.parl.ca/DocumentViewer/en/42-1/bill/C-46/royal-assent

[105] C. de Souza Gondim, O. A. M. Coelho, R. L. Alvarenga, R. G. Junqueira, S. V. C. de Souza, An appropriate and systematized procedure for validating qualitative methods: Its application in the detection of sulfonamide residues in raw milk, Analytica Chimica Acta 830 (2014) 11–22. doi:10.1016/j.aca.2014.04.050.

[106] M. I. López, M. P. Callao, I. Ruisánchez, A tutorial on the validation of qualitative methods: From the univariate to the multivariate approach, Analytica Chimica Acta 891 (2015) 62–72. doi:10.1016/j.aca.2015.06.032.

[107] E. Trullols, I. Ruisánchez, F. Rius, J. Huguet, Validation of qualitative methods of analysis that use control samples, Trends in Analytical Chemistry 24 (6) (2005) 516–524. doi:10.1016/j.trac.2005.04.001.

[108] V. Vindenes, B. Yttredal, E. Øiestad, H. Waal, J. Bernard, J. Mørland, A. Christophersen, Oral fluid is a viable alternative for monitoring drug abuse: detection of drugs in oral fluid by liquid chromatography-tandem mass spectrometry and comparison to the results from urine samples from patients treated with methadone or buprenorphine, Journal of Analytical Toxicology 35 (1) (2011) 32–39. doi:10.1093/anatox/35.1.32.

[109] S. M. Wille, V. Di Fazio, S. W. Toennes, J. H. van Wel, J. G. Ramaekers, N. Samyn, Evaluation of Δ9-tetrahydrocannabinol detection using DrugWipe5S® screening and oral fluid quantification after Quantisal™ collection for roadside drug detection via a controlled study with chronic cannabis users, Drug Testing and Analysis 7 (3) (2015) 178–186. doi:10.1002/dta.1660.

[110] L. D. Edwards, K. L. Smith, T. Savage, Drugged driving in Wisconsin: oral fluid versus blood, Journal of Analytical Toxicology 41 (6) (2017) 523–529. doi:10.1093/jat/bkx051.

[111] S. Strano-Rossi, E. Castrignanò, L. Anzillotti, G. Serpelloni, R. Mollica, F. Tagliaro, J. P. Pascali, D. Di Stefano, R. Sgalla, M. Chiarotti, Evaluation of four oral fluid devices (DDS®, Drugtest 5000®, Drugwipe 5+® and RapidSTAT®) for on-site monitoring drugged driving in comparison with UHPLC–MS/MS analysis, Forensic Science International 221 (1-3) (2012) 70–76. doi:10.1016/j.forsciint.2012.04.003.

[112] A. M. Veitenheimer, J. R. Wagner, Evaluation of Oral Fluid as a Specimen for DUID, Journal of Analytical Toxicology 41 (6) (2017) 517–522. doi:10.1093/jat/bkx036.

[113] H. Furuhaugen, R. E. Jamt, G. Nilsson, V. Vindenes, H. Gjerde, Roadside survey of alcohol and drug use among Norwegian drivers in 2016–2017: A follow-up of the 2008–2009 survey, Traffic Injury Prevention 19 (6) (2018) 555–562. doi:10.1080/15389588.2018.1478087.

[114] M. Davidian, P. D. Haaland, Regression and calibration with nonconstant error variance, Chemometrics and Intelligent Laboratory Systems 9 (3) (1990) 231–248. doi:10.1016/0169-7439(90)80074-G.

[115] H. T. Karnes, G. Shiu, V. P. Shah, Validation of bioanalytical methods, Pharmaceutical Research 8 (4) (1991) 421–426. doi:10.1023/A:1015882607690.

[116] P. Hubert, P. Chiap, J. Crommen, B. Boulanger, E. Chapuzet, N. Mercier, S. Bervoas-Martin, P. Chevalier, D. Grandjean, P. Lagorce, et al., The SFSTP guide on the validation of chromatographic methods for drug bioanalysis: from the Washington Conference to the laboratory, Analytica Chimica Acta 391 (2) (1999) 135–148. doi:10.1016/S0003-2670(99)00106-3.

[117] Bioanalytical Method Validation – Guidance for Industry, Standard, Food and Drug Administration, Silver Springs, United States (May 2001).
URL http://www.fda.gov/downloads/Drugs/Guidances/ucm070107.pdf

[118] F. T. Peters, O. H. Drummer, F. Musshoff, Validation of new methods, Forensic Science International 165 (2-3) (2007) 216–224. doi:10.1016/j.forsciint.2006.05.021.

[119] V. P. Shah, K. K. Midha, J. W. Findlay, H. M. Hill, J. D. Hulse, I. J. McGilveray, G. McKay, K. J. Miller, R. N. Patnaik, M. L. Powell, et al., Bioanalytical method validation - a revisit with a decade of progress, Pharmaceutical Research 17 (12) (2000) 1551–1557. doi:10.1023/A:1007669411738.

[120] F. T. Peters, Method validation using LC-MS, in: A. Polettini (Ed.), Applications of LC-MS in Toxicology, Pharmaceutical Press, London, United Kingdom, 2006, Ch. 4, pp. 71–96.

[121] C. Hartmann, J. Smeyers-Verbeke, D. Massart, R. McDowall, Validation of bioanalytical chromatographic methods, Journal of Pharmaceutical and Biomedical Analysis 17 (2) (1998) 193–218. doi:10.1016/S0731-7085(97)00198-2.

[122] W. Penninckx, C. Hartmann, D. Massart, J. Smeyers-Verbeke, Validation of the calibration procedure in atomic absorption spectrometric methods, Journal of Analytical Atomic Spectrometry 11 (4) (1996) 237–246. doi:10.1039/JA9961100237.

[123] J. Burrows, K. Watson, Linearity of chromatographic systems in drug analysis part I: theory of nonlinearity and quantification of curvature, Bioanalysis 7 (14) (2015) 1731–1743. doi:10.4155/bio.15.103.

[124] E. Pagliano, Z. Mester, J. Meija, Calibration graphs in isotope dilution mass spectrometry, Analytica Chimica Acta 896 (2015) 63–67. doi:10.1016/j.aca.2015.09.020.

[125] C. Moore, S. Rana, C. Coulter, Determination of meperidine, tramadol and oxycodone in human oral fluid using solid phase extraction and gas chromatography–mass spectrometry, Journal of Chromatography B 850 (1-2) (2007) 370–375. doi:10.1016/j.jchromb.2006.12.008.

[126] M. Cociglio, H. Peyriere, D. Hillaire-Buys, R. Alric, Application of a standardized coextractive cleanup procedure to routine high-performance liquid chromatography assays of teicoplanin and ganciclovir in plasma, Journal of Chromatography B: Biomedical Sciences and Applications 705 (1) (1998) 79–85. doi:10.1016/S0378-4347(97)00499-4.

[127] A. Gupta, B. Jansson, P. Chatelain, R. Massingham, M. Hammarlund-Udenaes, Quantitative determination of cetirizine enantiomers in guinea pig plasma, brain tissue and microdialysis samples using liquid chromatography/tandem mass spectrometry, Rapid Communications in Mass Spectrometry 19 (12) (2005) 1749–1757. doi:10.1002/rcm.1983.

[128] F. T. Peters, H. H. Maurer, Bioanalytical method validation and its implications for forensic and clinical toxicology – A review, Accreditation and Quality Assurance 7 (11) (2002) 441–449. doi:10.1007/s00769-002-0516-5.

[129] E. Rozet, A. Ceccato, C. Hubert, E. Ziemons, R. Oprean, S. Rudaz, B. Boulanger, P. Hubert, Analysis of recent pharmaceutical regulatory documents on analytical method validation, Journal of Chromatography A 1158 (1-2) (2007) 111–125. doi:10.1016/j.chroma.2007.03.111.

[130] D. Massart, B. Vandeginste, L. Buydens, S. De Jong, P. Lewi, J. Smeyers-Verbeke, Some important hypothesis tests, in: Handbook of Chemometrics and Qualimetrics: Part A, Vol. 20A of Data Handling in Science and Technology, Elsevier, Amsterdam, Netherlands, 1997, Ch. 5, pp. 93–120.

[131] R. D. Cook, S. Weisberg, Diagnostic methods using residuals, in: Residuals and Influence in Regression, Chapman and Hall, New York, United States, 1982, Ch. 2, pp. 10–100.

[132] A. G. González, M. Á. Herrador, A practical guide to analytical method validation, including measurement uncertainty and accuracy profiles, Trends in Analytical Chemistry 26 (3) (2007) 227–238. doi:10.1016/j.trac.2007.01.009.

[133] D. A. Darling, The Kolmogorov-Smirnov, Cramer-von Mises tests, The Annals of Mathematical Statistics 28 (4) (1957) 823–838.

[134] J. Miller, J. C. Miller, Statistics of repeated measurements, in: Statistics and Chemometrics for Analytical Chemistry, 6th Edition, Pearson Education, Harlow, England, 2010, Ch. 3, pp. –37–73.

[135] A. W. van der Vaart, J. A. Wellner, The bootstrap, in: Weak Convergence and Empirical Processes: With Applications to Statistics, Springer, New York, United States, 1996, Ch. 3.6, pp. 345–359.

[136] B. Desharnais, F. Camirand-Lemyre, P. Mireault, C. D. Skinner, Determination of confidence intervals in non-normal data: application of the bootstrap to cocaine concentration in femoral blood, Journal of Analytical Toxicology 39 (2) (2015) 113–117. doi:10.1093/jat/bku127.

[137] J. A. W. Wellner, A. W. van der Vaart, Empirical processes indexed by estimated functions, in: Asymptotics: Particles, Processes and Inverse Problems, Vol. 55 of Lecture Notes - Monograph Series, Institute of Mathematical Statistics, 2007, pp. 234–252. doi:10.1214/074921707000000382.

[138] Y. Zhao, G. Liu, J. X. Shen, A.-F. Aubry, Reasons for calibration standard curve slope variation in LC–MS assays and how to address it, Bioanalysis 6 (11) (2014) 1439–1443. doi:10.4155/bio.14.71.

[139] P. Hubert, J.-J. Nguyen-Huu, B. Boulanger, E. Chapuzet, N. Cohen, P.-A. Compagnon, W. Dewé, M. Feinberg, M. Laurentie, N. Mercier, et al., Harmonization of strategies for the validation of quantitative analytical procedures: A SFSTP proposal–Part III, Journal of Pharmaceutical and Biomedical Analysis 45 (1) (2007) 82–96. doi:10.1016/j.jpba.2007.06.032.

[140] K. Lanckmans, R. Clinckers, A. Van Eeckhaut, S. Sarre, I. Smolders, Y. Michotte, Use of microbore LC–MS/MS for the quantification of oxcarbazepine and its active metabolite in rat brain microdialysis samples, Journal of Chromatography B 831 (1-2) (2006) 205–212. doi:10.1016/j.jchromb.2005.12.003.

[141] C. Apostolou, Y. Dotsikas, C. Kousoulos, Y. L. Loukas, Development and validation of an improved high-throughput method for the determination of anastrozole in human plasma by LC–MS/MS and atmospheric pressure chemical ionization, Journal of Pharmaceutical and Biomedical Analysis 48 (3) (2008) 853–859. doi:10.1016/j.jpba.2008.06.006.

[142] L. B. Nilsson, G. Eklund, Direct quantification in bioanalytical LC–MS/MS using internal calibration via analyte/stable isotope ratio, Journal of Pharmaceutical and Biomedical Analysis 43 (3) (2007) 1094–1099. doi:10.1016/j.jpba.2006.09.030.

[143] V. Pirro, V. Valente, P. Oliveri, A. De Bernardis, A. Salomone, M. Vincenti, Chemometric evaluation of nine alcohol biomarkers in a large population of clinically-classified subjects: pre-eminence of ethyl glucuronide concentration in hair for confirmatory classification, Analytical and Bioanalytical Chemistry 401 (7) (2011) 2153. doi:10.1007/s00216-011-5314-7.

[144] V. L. Fulgoni III, D. R. Keast, H. R. Lieberman, Trends in intake and sources of caffeine in the diets of US adults: 2001–2010, The American Journal of Clinical Nutrition 101 (5) (2015) 1081–1087. doi:10.3945/ajcn.113.080077.

[145] H. Andresen-Streichert, A. Müller, A. Glahn, G. Skopp, M. Sterneck, Alcohol biomarkers in clinical and forensic contexts, Deutsches Ärzteblatt International 115 (18) (2018) 309. doi:10.3238/arztebl.2018.0309.

[146] National Health and Nutrition Examination Survey, Data, National Center for Health Statistics, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Hyattsville, United States (2018).
URL http://www.cdc.gov/nchs/nhanes.htm

[147] S. Elliott, C. Smith, D. Cassidy, The post-mortem relationship between β-hydroxybutyrate (BHB), acetone and ethanol in ketoacidosis, Forensic Science International 198 (1-3) (2010) 53–57. doi:10.1016/j.forsciint.2009.10.019.

[148] E. Osuna, G. Vivero, J. Conejero, J. M. Abenza, P. Martínez, A. Luna, M. D. Pérez-Cárceles, Postmortem vitreous humor β-hydroxybutyrate: its utility for the post-mortem interpretation of diabetes mellitus, Forensic Science International 153 (2-3) (2005) 189–195. doi:10.1016/j.forsciint.2004.09.105.

[149] S. Savard, C. Lapointe, M. Lamarche, P. Mireault, Development of a New Method for Simultaneous Quantitative BHB and GHB Analysis by GC-MS, Poster presentation, International Association of Forensic Sciences (IAFS) 2017 Meeting, Toronto, Canada (2017).

[150] Instructions: BSTFA + TMCS N,O-bis(Trimethylsilyl)trifluoroacetamide with Trimethylchlorosilane, Handbook, ThermoFisher Scientific (2017).
URL https://fscimage.fishersci.com/images/D00369~.pdf

[151] W. Navidi, Propagation of error, in: Statistics for Engineers and Scientists, Mc Graw Hill Education, New York, United States, 2014, Ch. 3, pp. 164–199.

[152] S. Zörntlein, A. Kopp, J. Becker, T. Kaufmann, J. Röhrich, R. Urban, *In vitro* production of GHB in blood and serum samples under various storage conditions, Forensic Science International 214 (1-3) (2012) 113–117. doi:10.1016/j.forsciint.2011.07.030.

[153] B. Desharnais, G. Huppé, M. Lamarche, P. Mireault, C. D. Skinner, Cyanide quantification in post-mortem biological matrices by headspace GC–MS, Forensic Science International 222 (1–3) (2012) 346–351. doi:10.1016/j.forsciint.2012.06.017.

[154] O. González, M. E. Blanco, G. Iriarte, L. Bartolomé, M. I. Maguregui, R. M. Alonso, Bioanalytical chromatographic method validation according to current regulations, with a special focus on the non-well defined parameters limit of quantification, robustness and matrix effect, Journal of Chromatography A 1353 (2014) 10–27. doi:10.1016/j.chroma.2014.03.077.

[155] S. M. Wille, W. Coucke, T. De Baere, F. T. Peters, Update of standard practices for new method validation in forensic toxicology, Current Pharmaceutical Design 23 (36) (2017) 5442–5454. doi:10.2174/1381612823666170714154444.

[156] R. Parikh, A. Mathai, S. Parikh, G. C. Sekhar, R. Thomas, Understanding and using sensitivity, specificity and predictive values, Indian Journal of Ophthalmology 56 (1) (2008) 45. doi:10.4103/0301-4738.37595.

[157] D. G. Altman, J. M. Bland, Diagnostic tests. 1: Sensitivity and specificity, British Medical Journal 308 (6943) (1994) 1552. doi:10.1136/bmj.308.6943.1552.

[158] C. Côté, B. Desharnais, M.-A. Morel, J. Laquerre, M.-P. Taillon, G. Daigneault, C. D. Skinner, P. Mireault, High Throughput Protein Precipitation: Screening and Quantification of 106 Drugs and their Metabolites using LC-MS/MS, Oral presentation, 2017 Society of Forensic Toxicologists Meeting (SOFT) and 55th Annual Meeting of the International Association of Forensic Toxicologists (TIAFT), Boca Raton, United States (January 2018).

[159] E. Trullols, I. Ruisanchez, F. X. Rius, Validation of qualitative analytical methods, Trends in Analytical Chemistry 23 (2) (2004) 137–145. doi:10.1016/S0165-9936(04)00201-8.

[160] G. Y. Yi, Statistical Analysis with Measurement Error or Misclassification: Strategy, Method and Application, Springer, New York, United States, 2017.

[161] Cannabis Act, Legislation S.C. 2018, c. 16, Government of Canada, Ottawa, Canada (2018).
URL https://laws-lois.justice.gc.ca/eng/acts/C-24.5/

[162] J. Borzelleca, H. Cherrick, The excretion of drugs in saliva. Antibiotics., Journal of Oral Therapeutics and Pharmacology 2 (3) (1965) 180.

[163] M. G. Horning, L. Brown, J. Nowlin, K. Lertratanangkoon, P. Kellaway, T. E. Zion, Use of saliva in therapeutic drug monitoring, Clinical Chemistry 23 (2) (1977) 157–164.

[164] O. H. Drummer, Drug testing in oral fluid, Clinical Biochemist Reviews 27 (3) (2006) 147.

[165] O. Quintela, D. J. Crouch, D. M. Andrenyak, Recovery of drugs of abuse from the Immunalysis Quantisal™ oral fluid collection device, Journal of Analytical Toxicology 30 (8) (2006) 614–616. doi:10.1093/jat/30.8.614.

[166] M. H. Tang, C. Ching, S. Poon, S. S. Chan, W. Ng, M. Lam, C. Wong, R. Pao, A. Lau, T. W. Mak, Evaluation of three rapid oral fluid test devices on the screening of multiple drugs of abuse including ketamine, Forensic Science International 286 (2018) 113–120. doi:10.1016/j.forsciint.2018.03.004.

[167] M. Gröschl, Saliva: a reliable sample matrix in bioanalytics, Bioanalysis 9 (8) (2017) 655–668. doi:10.4155/bio-2017-0010.

[168] A. Doyon, L. Paradis-Tanguay, F. Crispino, A. Lajeunesse, Les analyses médico-légales de salives: expertise vis-à-vis l'analyse des drogues, Canadian Society of Forensic Science Journal 50 (2) (2017) 90–102. doi:10.1080/00085030.2017.1303254.

[169] C. Cohier, B. Mégarbane, O. Roussel, Illicit drugs in oral fluid: Evaluation of two collection devices, Journal of Analytical Toxicology 41 (1) (2017) 71–76. doi:10.1093/jat/bkw100.

[170] A. J. Krotulski, A. L. Mohr, M. Friscia, B. K. Logan, Field detection of drugs of abuse in oral fluid using the Alere™ DDS® 2 mobile test system with confirmation by liquid chromatography tandem mass spectrometry (LC–MS/MS), Journal of Analytical Toxicology 42 (3) (2017) 170–176. doi:10.1093/jat/bkx105.

[171] S. M. Wille, V. Di Fazio, N. Samyn, La salive dans les investigations toxicologiques: considérations pratiques et analytiques, in: P. Kintz (Ed.), Traité De Toxicologie Médico-judiciaire, Elsevier Masson, Paris, France, 2012, Ch. 8, pp. 219–255.

[172] A. G. Verstraete, Oral fluid testing for driving under the influence of drugs: history, recent progress and remaining challenges, Forensic Science International 150 (2-3) (2005) 143–150. doi:10.1016/j.forsciint.2004.11.023.

[173] Approved Drug Screening Equipment Order, Legislation SOR/2018-179, Government of Canada, Ottawa, Canada (2018).
URL https://laws-lois.justice.gc.ca/PDF/SOR-2018-179.pdf

[174] Drug Screening Equipment – Oral Fluid Standards and Evaluation Procedures, Standard, Canadian Society of Forensic Science Drugs and Driving Committee, Ottawa, Canada (2017).
URL https://www.csfs.ca/wp-content/uploads/2017/11/Approval-Standards-for-Drug-Screening-Equipment.pdf

[175] Order Amending the Approved Drug Screening Equipment Order, Legislation, Government of Canada, Ottawa, Canada (2019).
URL http://www.gazette.gc.ca/rp-pr/p1/2019/2019-04-20/html/reg4-eng.html

[176] E. Viel, E. Blais, P. Mireault, Statistical Overview of Drug Findings in Urine Samples from the DRE Program in the Province of Québec, Canada, Poster presentation, 2013 Society of Forensic Toxicologists Meeting (SOFT), Orlando, United States (2013).

[177] D. Menasco, C. Summit, J. Neifeld, S. Marin, L. Williams, E. Gairloch, Practical Considerations using Quantisal Oral Fluid Collection Devices & SPE Method Development by Polymeric Mixed-Mode Cation Exchange, Poster presentation, Annual Congress in Clinical Mass Spectrometry, Palm Springs, United States (2018).
URL http://www.weber.hu/Downloads/SPE/Posters/P179_oral_fluid_evolute_CX.pdf

[178] R. Gudihal, S. Babu CV, N. Tang, S. Palaniswamy, U. S, S. Basingi, Analysis of Polyethylene Glycol (PEG) and a Mono and Di-PEGylated Therapeutic Protein Using HPLC and Q-TOF Mass Spectrometry, Application Note 5991-1509EN, Agilent Technologies, United States (2012).
URL https://www.agilent.com/cs/library/applications/5991-1509EN.pdf

[179] N. Fabresse, H. Aouad, A. Knapp, C. Mayer, I. Etting, I. A. Larabi, J.-C. Alvarez, Development and validation of a liquid chromatography-tandem mass spectrometry

method for simultaneous detection of 10 illicit drugs in oral fluid collected with FLOQSwabs™ and application to real samples, Drug Testing and Analysis 11 (6) (2019) 824–832. doi:10.1002/dta.2563.

[180] R. Leverence, M. J. Avery, O. Kavetskaia, H. Bi, C. E. Hop, A. I. Gusev, Signal suppression/enhancement in HPLC-ESI-MS/MS from concomitant medications, Biomedical Chromatography 21 (11) (2007) 1143–1150. doi:10.1002/bmc.863.

[181] D. Lee, G. Milman, A. J. Barnes, R. S. Goodwin, J. Hirvonen, M. A. Huestis, Oral fluid cannabinoids in chronic, daily cannabis smokers during sustained, monitored abstinence, Clinical Chemistry 57 (8) (2011) 1127–1136. doi:10.1373/clinchem.2011.164822.

[182] E. Saar, D. Gerostamoulos, O. H. Drummer, J. Beyer, Assessment of the stability of 30 antipsychotic drugs in stored blood specimens, Forensic Science International 215 (1-3) (2012) 152–158. doi:10.1016/j.forsciint.2011.02.022.

[183] D. C. Mata, Stability of 26 sedative hypnotics in six toxicological matrices at different storage conditions, Journal of Analytical Toxicology 40 (8) (2016) 663–668. doi:10.1093/jat/bkw084.

[184] A. Parkinson, B. W. Ogilvie, D. B. Buckley, F. Kazmi, M. Czerwinski, O. Parkinson, Biotransformation of xenobiotics, in: C. D. Klaassen (Ed.), Casarett & Doull's Toxicology: The Basic Science of Poisons, 8th Edition, McGraw-Hill Education / Medical, New York, United States, 2013, Ch. 6, pp. 185–366.

[185] F. P. Guengerich, Cytochromes P450, drugs, and diseases, Molecular Interventions 3 (4) (2003) 194. doi:10.1124/mi.3.4.194.

[186] Y. Gasche, Y. Daali, M. Fathi, A. Chiappe, S. Cottini, P. Dayer, J. Desmeules, Codeine intoxication associated with ultrarapid CYP2D6 metabolism, New England Journal of Medicine 351 (27) (2004) 2827–2831. doi:10.1056/NEJMoa041888.

[187] J. George, C. Liddle, M. Murray, K. Byth, G. C. Farrell, Pre-translational regulation of cytochrome P450 genes is responsible for disease-specific changes of individual P450 enzymes among patients with cirrhosis, Biochemical Pharmacology 49 (7) (1995) 873–881. doi:10.1016/0006-2952(94)00515-N.

[188] F. P. Guengerich, C. G. Turvy, Comparison of levels of several human microsomal cytochrome P-450 enzymes and epoxide hydrolase in normal and disease states using immunochemical analysis of surgical liver samples, Journal of Pharmacology and Experimental Therapeutics 256 (3) (1991) 1189–1194.

[189] J. Hansen, J. Palmfeldt, K. W. Pedersen, A. D. Funder, L. Frost, J. B. Hasselstrøm, J. R. Jornil, Postmortem protein stability investigations of the human hepatic drug-metabolizing cytochrome P450 enzymes CYP1A2 and CYP3A4 using mass spectrometry, Journal of Proteomics 194 (2019) 125–131. doi:10.1016/j.jprot.2018.11.024.

[190] B. L. Williamson, S. Purkayastha, C. L. Hunter, L. Nuwaysir, J. Hill, L. Easterwood, J. Hill, Quantitative protein determination for CYP induction via LC-MS/MS, Proteomics 11 (1) (2011) 33–41. doi:10.1002/pmic.201000456.

[191] A. Bowdler, T. Chan, The time course of red cell lysis in hypotonic electrolyte solutions, The Journal of Physiology 201 (2) (1969) 437–452. doi:10.1113/jphysiol.1969.sp008765.

[192] L. E. Westerman, P. E. Jensen, Liposomes Composed of Reconstituted Membranes for Induction of Tumor-Specific Immunity, Methods in Enzymology 373 (2003) 118–127. doi:10.1016/S0076-6879(03)73008-0.

[193] W. Konigsberg, Reduction of disulfide bonds in proteins with dithiothreitol, Methods in Enzymology 25 (1972) 185–188. doi:10.1016/S0076-6879(72)25015-7.

[194] S. F. Betz, Disulfide bonds and the stability of globular proteins, Protein Science 2 (10) (1993) 1551–1558. doi:10.1002/pro.5560021002.

[195] B. Herbert, M. Galvani, M. Hamdan, E. Olivieri, J. MacCarthy, S. Pedersen, P. G. Righetti, Reduction and alkylation of proteins in preparation of two-dimensional map analysis: Why, when, and how?, Electrophoresis 22 (10) (2001) 2046–2057. doi:10.1002/1522-2683(200106)22:10<2046::AID-ELPS2046>3.0.CO;2-C.

[196] Human cytochrome P450 IID6 (CYP2D6) gene, complete cds, Data, National Center for Biotechnology Information (NCBI), Bethesda, United States (1994).
URL http://www.ncbi.nlm.nih.gov/nuccore/M33388.1

[197] CYP3A4 - Cytochrome P450 3A4 - Homo sapiens (Human) - CYP3A4 gene & protein, Data, National Center for Biotechnology Information (NCBI), Bethesda, United States.
URL https://www.uniprot.org/uniprot/P08684

[198] H.-G. Xie, R. B. Kim, A. J. Wood, C. M. Stein, Molecular basis of ethnic differences in drug disposition and response, Annual Review of Pharmacology and Toxicology 41 (1) (2001) 815–850. doi:10.1146/annurev.pharmtox.41.1.815.

[199] Bradford, L DiAnne, CYP2D6 allele frequency in european caucasians, asians, africans and their descendants, Pharmacogenomics 3 (2) (2002) 229–243. doi:10.1517/14622416.3.2.229.

[200] J. K. Lamba, Y. S. Lin, K. Thummel, A. Daly, P. B. Watkins, S. Strom, J. Zhang, E. G. Schuetz, Common allelic variants of cytochrome P4503A4 and their prevalence in different populations, Pharmacogenetics and Genomics 12 (2) (2002) 121–132.

[201] L. Eriksson, E. Johansson, N. Kettaneh-Wold, C. Wikström, S. Wold, Design of Experiments: Principles and Applications, Umetrics Academy, Umea, Sweden, 2008.

[202] P. R. Cook, C. Glenn, A. Armston, Effect of hemolysis on insulin determination by the Beckman Coulter Unicell DXI 800 immunoassay analyzer, Clinical Biochemistry 43 (6) (2010) 621–622. doi:10.1016/j.clinbiochem.2010.01.002.

[203] I. Ojanperä, A. Sajantila, L. Vinogradova, A. Thomas, W. Schänzer, M. Thevis, Post-mortem vitreous humour as potential specimen for detection of insulin analogues by LC–MS/MS, Forensic Science International 233 (1-3) (2013) 328–332. doi:10.1016/j.forsciint.2013.10.009.

[204] A. Corthals, A. Koller, D. W. Martin, R. Rieger, E. I. Chen, M. Bernaski, G. Recagno, L. M. Dávalos, Detecting the immune system response of a 500 year-old Inca mummy, PloS one 7 (7) (2012) e41244. doi:10.1371/journal.pone.0041244.

[205] C. Wadsworth, M. Buckley, Proteome degradation in fossils: investigating the longevity of protein survival in ancient bone, Rapid Communications in Mass Spectrometry 28 (6) (2014) 605. doi:10.1002/rcm.6821.

[206] M. Lamare, R. G. Taylor, L. Farout, Y. Briand, M. Briand, Changes in proteasome activity during postmortem aging of bovine muscle, Meat Science 61 (2) (2002) 199–204. doi:10.1016/S0309-1740(01)00187-5.

[207] J. R. Whiteaker, L. Zhao, L. Anderson, A. G. Paulovich, An automated and multiplexed method for high throughput peptide immunoaffinity enrichment and multiple reaction monitoring mass spectrometry-based quantification of protein biomarkers, Molecular & Cellular Proteomics 9 (1) (2010) 184–196. doi:10.1074/mcp.M900254-MCP200.

[208] R. Lametsch, P. Roepstorff, E. Bendixen, Identification of protein degradation during post-mortem storage of pig meat, Journal of Agricultural and Food Chemistry 50 (20) (2002) 5508–5512. doi:10.1021/jf025555n.

[209] M. Oscarson, M. Hidestrand, I. Johansson, M. Ingelman-Sundberg, A combination of mutations in the CYP2D6*17 (CYP2D6Z) allele causes alterations in enzyme function, Molecular Pharmacology 52 (6) (1997) 1034–1040. doi:10.1124/mol.52.6.1034.

[210] R. Baselt, Disposition of Toxic Drugs and Chemicals in Man, 11th Edition, Biomedical Publications, Seal Beach, United States, 2017.

[211] Compendium of Pharmaceuticals and Specialties, Canadian Pharmacists Association, Ottawa, Canada, 2019.

[212] A. C. Moffat, M. D. Osselton, B. Widdop, Clarke's Analysis of Drugs and Poisons, 4th Edition, Vol. 2, Pharmaceutical Press, London, United Kingdom, 2011.

[213] D. K. Molina, V. J. DiMaio, Normal organ weights in men: Part II—the brain, lungs, liver, spleen, and kidneys, The American Journal of Forensic Medicine and Pathology 33 (4) (2012) 368–372. doi:10.1097/PAF.0b013e31823d29ad.

[214] P. Magalhães, G. Alves, A. LLerena, A. Falcão, Clinical drug-drug interactions: focus on venlafaxine, Drug Metabolism and Personalized Therapy 30 (1) (2015) 3–17. doi:10.1515/dmdi-2014-0011.

[215] M. Schulz, S. Iwersen-Bergmann, H. Andresen, A. Schmoldt, Therapeutic and toxic blood concentrations of nearly 1,000 drugs and other xenobiotics, Critical Care 16 (4) (2012) R136. doi:10.1186/cc11441.

[216] A. Salomone, E. Gerace, D. Di Corcia, E. Alladio, M. Vincenti, P. Kintz, Hair analysis can provide additional information in doping and forensic cases involving clostebol, Drug Testing and Analysis 11 (1) (2019) 95–101. doi:10.1002/dta.2469.

[217] O. Deltombe, T. Mertens, S. Eloot, A. G. Verstraete, Development and validation of an ultra-high performance liquid chromatography–high resolution mass spectrometry method for the quantification of total and free teicoplanin in human plasma, Clinical Biochemistry 65 (2019) 29–37. doi:10.1016/j.clinbiochem.2018.12.010.

[218] E. Amante, E. Alladio, A. Salomone, M. Vincenti, F. Marini, G. Alleva, S. De Luca, F. Porpiglia, Correlation between chronological and physiological age of males from their multivariate urinary endogenous steroid profile and prostatic carcinoma-induced deviation, Steroids 139 (2018) 10–17. doi:10.1016/j.steroids.2018.09.007.

[219] E. Alladio, G. Biosa, F. Seganti, D. Di Corcia, A. Salomone, M. Vincenti, M. R. Baumgartner, Systematic optimisation of ethyl glucuronide extraction conditions from scalp hair by design of experiments and its potential effect on cut-off values appraisal, Drug Testing and Analysis 10 (9) (2018) 1394–1403. doi:10.1002/dta.2405.

[220] C. Bozzolino, S. Vaglio, E. Amante, E. Alladio, E. Gerace, A. Salomone, M. Vincenti, Individual and cyclic estrogenic profile in women: structure and variability of the data, Steroids (2019) [Advanced article]doi:10.1016/j.steroids.2019.108432.

[221] L. Vaillancourt, B. Desharnais, N. Goudreau, P. Mireault, Interference of fetal hemoglobin in the determination of carboxyhemoglobin by spectrophotometry, Canadian Society of Forensic Science Journal 49 (2) (2016) 69–77. doi:10.1080/00085030.2015.1115692.

[222] American Statistical Association Position on Statistical Statements for Forensic Evidence, Position statement, American Statistical Association, Alexandria, United States (January 2019).
URL https://www.amstat.org/asa/files/pdfs/POL-ForensicScience.pdf

[223] M. Bauer, I. Gramlich, S. Polzin, D. Patzelt, Quantification of mRNA degradation as possible indicator of postmortem interval—a pilot study, Legal Medicine 5 (4) (2003) 220–227. doi:10.1016/j.legalmed.2003.08.001.

[224] D. Wessel, U. Flügge, A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids, Analytical Biochemistry 138 (1) (1984) 141–143. doi:10.1016/0003-2697(84)90782-6.

[225] S. Sechi, Quantitative Proteomics by Mass Spectrometry, 2nd Edition, Humana Press, Totowa, United States, 2016.

[226] K. Marcus, Quantitative Methods in Proteomics, Humana Press, Totowa, United States, 2012.

[227] C. E. Eyers, S. Gaskell, Quantitative Proteomics, Royal Society of Chemistry, London, United Kingdom, 2014.

[228] R. G. Krishna, F. Wold, Post-translational modifications of proteins, in: K. Imahori, F. Sakiyama (Eds.), Methods in Protein Sequence Analysis, Springer, Boston, United States, 1993, pp. 167–172. doi:10.1007/978-1-4899-1603-7_21.

[229] V. Czitrom, One-factor-at-a-time versus designed experiments, The American Statistician 53 (2) (1999) 126–131. doi:10.1080/00031305.1999.10474445.

[230] B. Achour, J. Barber, A. Rostami-Hodjegan, Expression of hepatic drug-metabolizing cytochrome P450 enzymes and their intercorrelations: a meta-analysis, Drug Metabolism and Disposition 42 (8) (2014) 1349–1356. doi:10.1124/dmd.114.058834.

[231] S. Michaels, M. Z. Wang, The revised human liver cytochrome P450 "Pie": absolute protein quantification of CYP4F and CYP3A enzymes using targeted quantitative proteomics, Drug Metabolism and Disposition 42 (8) (2014) 1241–1251. doi:10.1124/dmd.114.058040.

[232] A. Cieślak, I. Kelly, J. Trottier, M. Verreault, E. Wunsch, P. Milkiewicz, G. Poirier, A. Droit, O. Barbier, Selective and sensitive quantification of the cytochrome P450 3A4 protein in human liver homogenates through multiple reaction monitoring mass spectrometry, Proteomics 16 (21) (2016) 2827–2837. doi:10.1002/pmic.201500386.

[233] M. Cronin, M. Pho, D. Dutta, J. C. Stephans, S. Shak, M. C. Kiefer, J. M. Esteban, J. B. Baker, Measurement of gene expression in archival paraffin-embedded tissues: development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay, The American Journal of Pathology 164 (1) (2004) 35–42. doi:10.1016/S0002-9440(10)63093-3.

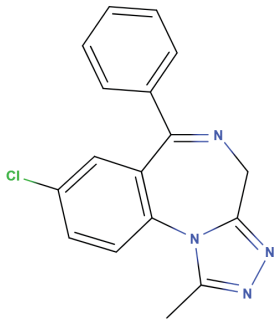
[234] M. Aguiar, R. Masse, B. F. Gibbs, Regulation of cytochrome P450 by posttranslational modification, Drug Metabolism Reviews 37 (2) (2005) 379–404. doi:10.1081/DMR-46136.

[235] J. Côté, Wilfrid Derome: Expert en homicides, Les éditions du Boréal, Montréal, Canada, 2003.

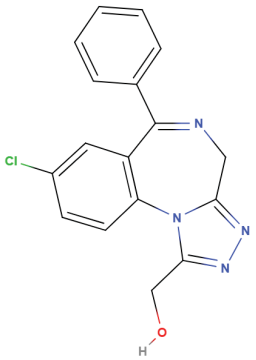
A

Structure and Relevant Information of Molecules Discussed in This Work

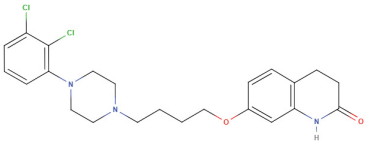
Unless otherwise stated, information was taken from *Clarke's Analysis of Drugs and Poisons – Volume 2* [212] and from *Disposition of Toxic Drugs and Chemicals in Man* [210]. Molecules were generally placed in alphabetical order, but metabolites were placed right after their parent drug.



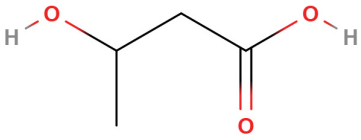
Name: Alprazolam
Family: Benzodiazepines
Use: Short-term management of anxiety disorders
Molecular weight: 308.8 $g\ mol^{-1}$
pKa: 2.4
Log P**:** 2.12
Half-life: 11 to 15 h



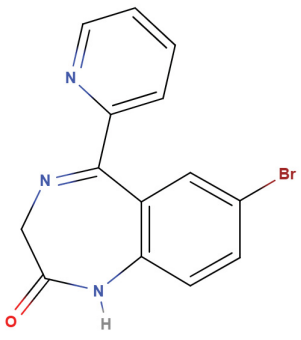
Name: α-Hydroxyalprazolam
Family: Benzodiazepines
Use: Active metabolite of alprazolam
Molecular weight: 324.8 $g\ mol^{-1}$
pKa: NA
Log P**:** NA
Half-life: NA



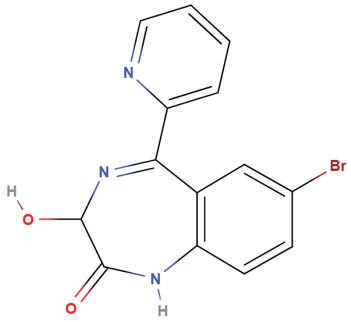
Name: Aripiprazole
Family: Antipsychotic
Use: Treatment of schizophrenia and bipolar disorder
Molecular weight: 448.4 $g\ mol^{-1}$
pKa: 7.46, 13.51
Log P**:** 4.6
Half-life: 47 to 68 h



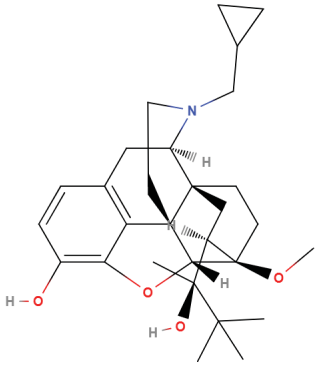
Name: β-hydroxybutyric acid (BHB)
Family: Biomarkers
Use: Ketoacidosis marker
Molecular weight: 104.1 $g\ mol^{-1}$
pKa: NA
Log P**:** -0.5
Half-life: NA



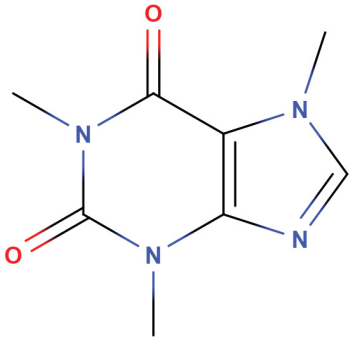
Name: Bromazepam
Family: Benzodiazepines
Use: Anti anxiety
Molecular weight: 316.2 $g\ mol^{-1}$
pKa: 2.9, 11.0
Log P**:** 2.05
Half-life: 8 to 19 h



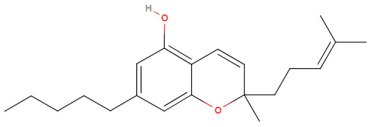
Name: 3-Hydroxy bromazepam
Family: Benzodiazepines
Use: Active metabolite of bromazepam
Molecular weight: 332.2 $g\ mol^{-1}$
pKa: NA
Log P**:** 1.1
Half-life: Similar to bromazepam



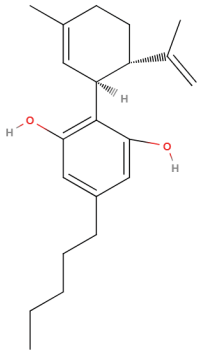
Name: Buprenorphine
Family: Opioids
Use: Treatment of opioid addiction, acute or chronic pain
Molecular weight: 467.6 $g\ mol^{-1}$
pKa: 8.5, 10.0
Log P**:** 4.98
Half-life: 1.2 to 7.2 h



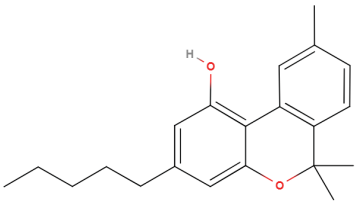
Name: Caffeine
Family: Simulants
Use: Food and drinks
Molecular weight: 194.2 $g\ mol^{-1}$
pKa: 10.4, 14.0
Log P**:** -0.07
Half-life: 2 to 10 h



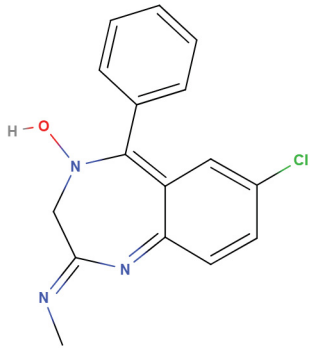
Name: Cannabichromen (CBC)
Family: Cannabinoids
Use: Part of the cannabis plant
Molecular weight: 314.5 $g\ mol^{-1}$
pKa: NA
Log P**:** NA
Half-life: NA



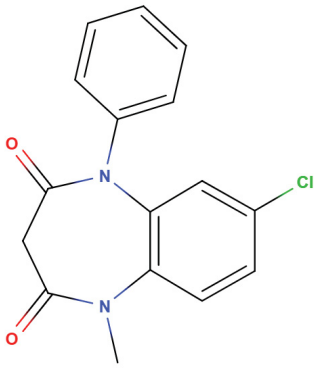
Name: Cannabidiol (CBD)
Family: Cannabinoids
Use: Part of the cannabis plant
Molecular weight: 314.5 $g\ mol^{-1}$
pKa: NA
Log P**:** NA
Half-life: NA



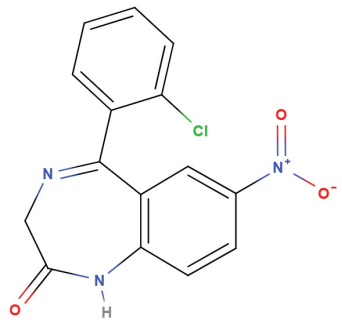
Name: Cannabinol (CBN)
Family: Cannabinoids
Use: Part of the cannabis plant
Molecular weight: 310.4 $g\ mol^{-1}$
pKa: NA
Log P**:** NA
Half-life: NA



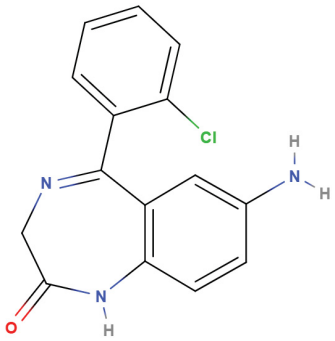
Name: Chlordiazepoxide
Family: Benzodiazepines
Use: Treatment of anxiety and insomnia
Molecular weight: 299.8 $g\ mol^{-1}$
pKa: 4.8
Log P**:** 2.44
Half-life: 5 to 30 h



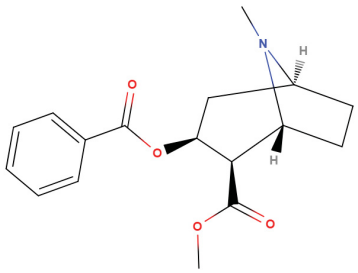
Name: Clobazam
Family: Benzodiazepines
Use: Anxioloytic, anticonvulsivant
Molecular weight: 300.8 $g\ mol^{-1}$
pKa: NA
Log P**:** 2.12
Half-life: 10 to 58 h



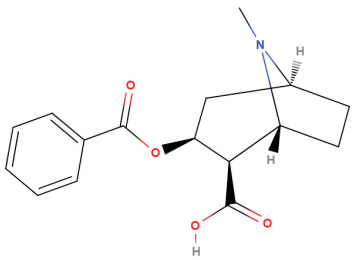
Name: Clonazepam
Family: Benzodiazepines
Use: Anticonvulsivant, treatment of panic disorder
Molecular weight: 315.7 $g\ mol^{-1}$
pKa: 1.5, 10.5
Log P**:** 2.41
Half-life: 20 to 40 h



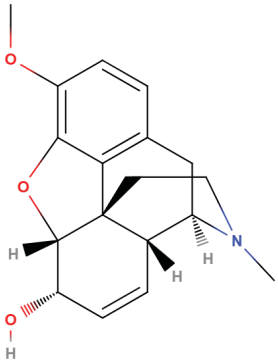
Name: 7-Aminoclonazepam
Family: Benzodiazepines
Use: Metabolite of clonazepam
Molecular weight: 285.7 $g\ mol^{-1}$
pKa: NA
Log P**:** 1.8
Half-life: NA



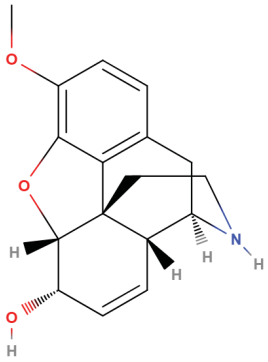
Name: Cocaine
Family: Stimulants
Use: Drug of abuse
Molecular weight: 303.4 $g\ mol^{-1}$
pKa: 8.7
Log P**:** 2.3
Half-life: 0.7 to 1.5 h



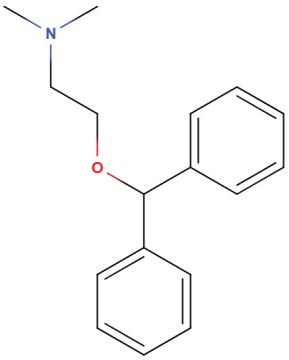
Name: Benzoylecgonine
Family: NA
Use: Metabolite of cocaine
Molecular weight: 289.3 $g\ mol^{-1}$
pKa: NA
Log P**:** -1.3
Half-life: NA



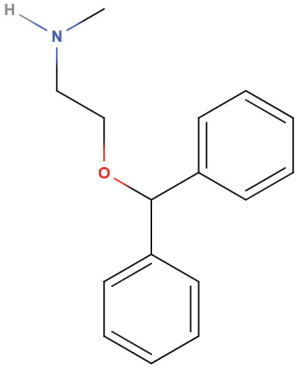
Name: Codeine
Family: Opiates
Use: Pain treatment, cough medicine
Molecular weight: 317.4 $g\ mol^{-1}$
pKa: 8.2
Log P**:** 0.6
Half-life: 2 to 4 h



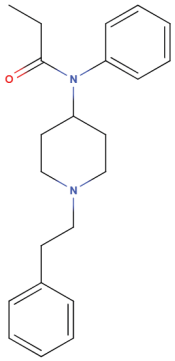
Name: Norcodeine
Family: Opiates
Use: Metabolite of codeine
Molecular weight: 285.3 $g\ mol^{-1}$
pKa: 9.23
Log P**:** 0.69
Half-life: NA



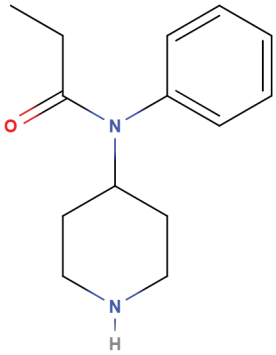
Name: Diphenhydramine
Family: NA
Use: Antihistamine, treatment of insomnia
Molecular weight: 255.4 $g\ mol^{-1}$
pKa: 8.98
Log P**:** 3.27
Half-life: 2.4 to 9.3 h



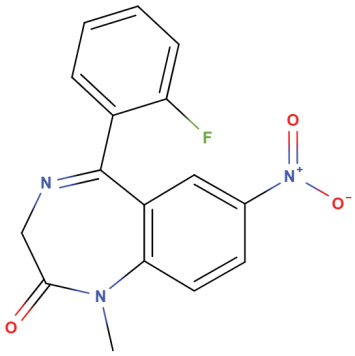
Name: N-Desmethyl diphenhydramine
Family: NA
Use: Metabolite of diphenhydramine
Molecular weight: 277.8 $g\ mol^{-1}$
pKa: NA
Log P**:** NA
Half-life: NA



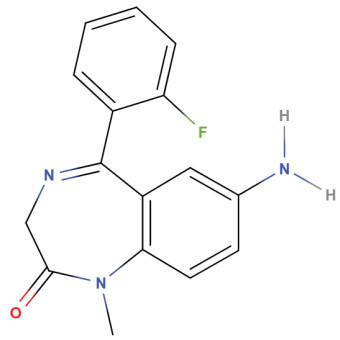
Name: Fentanyl
Family: Opioids
Use: Pain treatment, anesthesia
Molecular weight: 336.5 $g\ mol^{-1}$
pKa: NA
Log P**:** 2.3
Half-life: 3.7 h



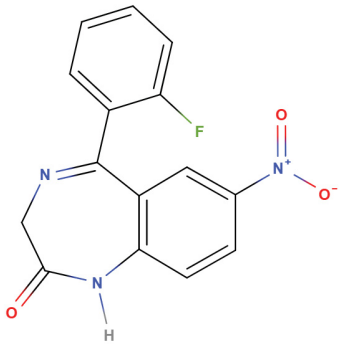
Name: Norfentanyl
Family: Opioids
Use: Metabolite of fentanyl
Molecular weight: 232.3 $g\ mol^{-1}$
pKa: NA
Log P**:** NA
Half-life: NA



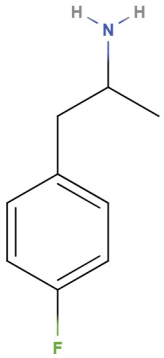
Name: Flunitrazepam
Family: Benzodiazepines
Use: Treatment of insomnia, purported use in DFSA
Molecular weight: 313.3 $g\ mol^{-1}$
pKa: 1.8
Log P**:** 2.1
Half-life: 16 to 35 h



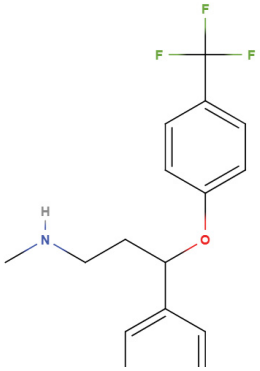
Name: 7-Aminoflunitrazepam
Family: Benzodiazepines
Use: Metabolite of flunitrazepam
Molecular weight: 283.3 $g\ mol^{-1}$
pKa: NA
Log P**:** 1.3
Half-life: NA



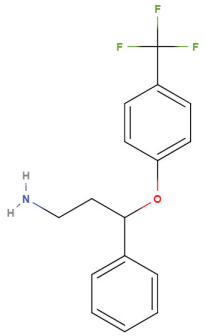
Name: N-Desmethylflunitrazepam
Family: Benzodiazepines
Use: Metabolite of flunitrazepam
Molecular weight: 299.3 $g\ mol^{-1}$
pKa: NA
Log P**:** 2.04
Half-life: NA



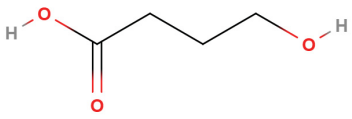
Name: 4-Fluoroamphetamine
Family: Stimulants
Use: Research chemical, drug of abuse, analogue
Molecular weight: 153.2 $g\ mol^{-1}$
pKa: NA
Log P**:** NA
Half-life: NA



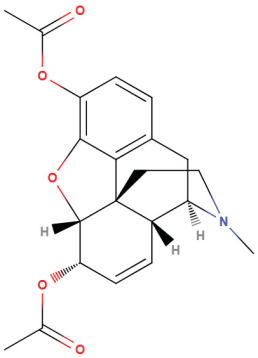
Name: Fluoxetine
Family: Antidepressant
Use: Treatment of severe depression, OCD
Molecular weight: 309.3 $g\ mol^{-1}$
pKa: 9.8
Log P**:** 4.05
Half-life: 4 to 6 days



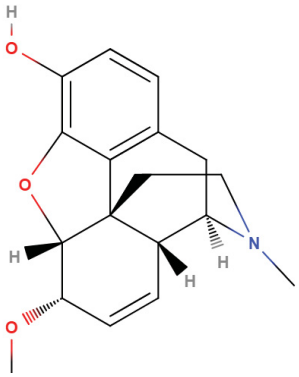
Name: Norlfuoxetine
Family: Antidepressant
Use: Active metabolite of fluoxetine
Molecular weight: 295.3 $g\ mol^{-1}$
pKa: 9.77
Log P**:** NA
Half-life: 4 to 16 days



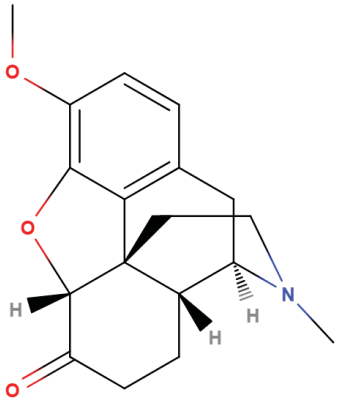
Name: γ-Hydroxybutyric Acid (GHB)
Family: CNS depressant
Use: Treatment of narcolepsy, drug of abuse
Molecular weight: 104.1 $g\ mol^{-1}$
pKa: 4.44
Log P**:** -0.51
Half-life: 0.3 to 1 h



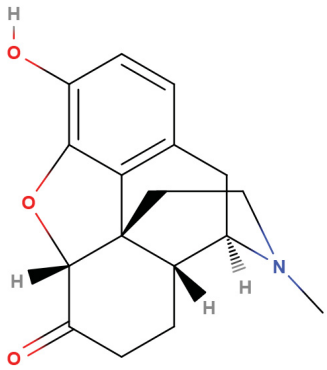
Name: Heroin
Family: Opiates
Use: Drug of abuse
Molecular weight: 369.4 $g\ mol^{-1}$
pKa: 7.6
Log P**:** 0.2
Half-life: 3 minutes



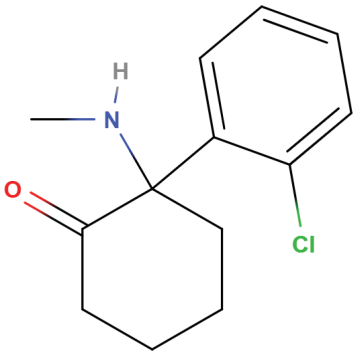
Name: 6-Monoacetylmorphine (6-MAM)
Family: Opiates
Use: Metabolite of heroin
Molecular weight: 327.4 $g\ mol^{-1}$
pKa: NA
Log P**:** NA
Half-life: NA



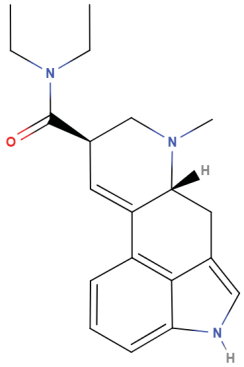
Name: Hydrocodone
Family: Opioids
Use: Treatment of severe pain, cough suppressant
Molecular weight: 299.4 $g\ mol^{-1}$
pKa: 8.3
Log P**:** 2.2
Half-life: $\approx$ 4 h



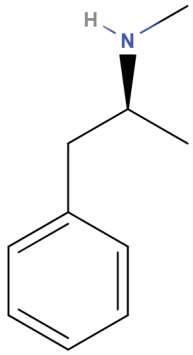
Name: Hydromorphone
Family: Opioids
Use: Treatment of severe pain
Molecular weight: 285.3 $g\ mol^{-1}$
pKa: 8.2
Log P**:** -4.0
Half-life: $\approx$ 2.5 h



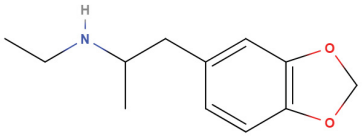
Name: Ketamine
Family: Anesthetic
Use: Anesthesia, chronic pain, drug of abuse
Molecular weight: 237.7 $g\ mol^{-1}$
pKa: 7.5
Log P**:** 3.1
Half-life: 2 to 3 h



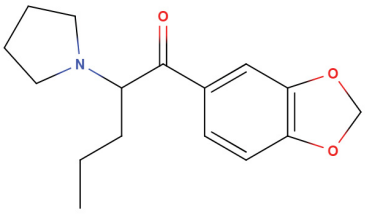
Name: Lysergic acid diethylamide (LSD)
Family: Hallucinogens
Use: Drug of abuse
Molecular weight: 323.4 $g\ mol^{-1}$
pKa: 7.5
Log P**:** 2.9
Half-life: 2.5 h



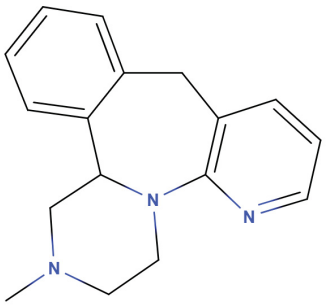
Name: Methamphetamine
Family: Stimulants
Use: Drug of abuse
Molecular weight: 149.2 $g\ mol^{-1}$
pKa: 9.87, 10.1
Log P**:** 2.07
Half-life: 9 h



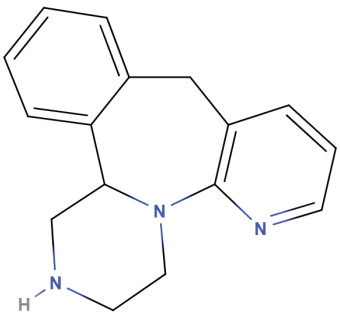
Name: Methylenedioxy-N-ethylamphetamine (MDEA)
Family: Stimulants
Use: Drug of abuse
Molecular weight: 207.3 $g\ mol^{-1}$
pKa: NA
Log P**:** 2.5
Half-life: $\approx$ 7.5 h



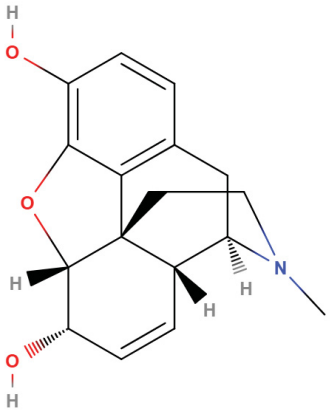
Name: 3,4-Methylenedioxypyrovalerone (MDPV)
Family: Cathinone derivative
Use: Drug of abuse
Molecular weight: 275.3 $g\ mol^{-1}$
pKa: 8.4
Log P**:** 3.4
Half-life: NA



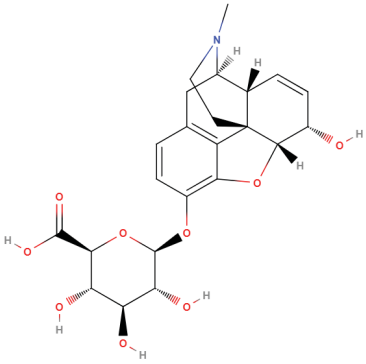
Name: Mirtazapine
Family: Antidepressant
Use: Treatment of depression and insomnia
Molecular weight: 265.4 $g\ mol^{-1}$
pKa: 7.1
Log P**:** 3.3
Half-life: 20 to 40 h



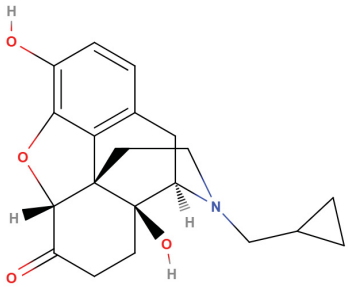
Name: N-Desmethylmirtazapine
Family: NA
Use: Metabolite of mirtazapine
Molecular weight: 265.4 $g\ mol^{-1}$
pKa: NA
Log P**:** NA
Half-life: NA



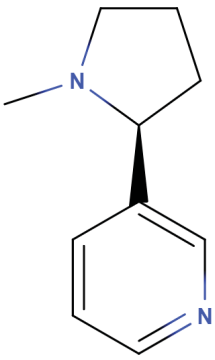
Name: Morphine
Family: Opiates
Use: Treatment of acute and chronic pain
Molecular weight: 303.4 $g\ mol^{-1}$
pKa: 8.0, 9.9
Log P**:** -0.1
Half-life: 2 to 3 h



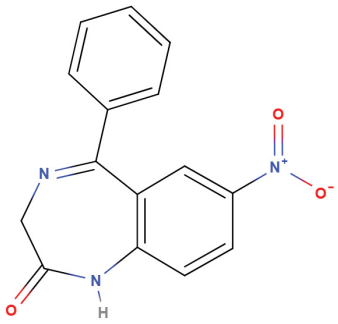
Name: 3-Morphine-glucuronide
Family: Opiates
Use: Metabolite of morphine
Molecular weight: 461.5 $g\ mol^{-1}$
pKa: NA
Log P**:** NA
Half-life: NA



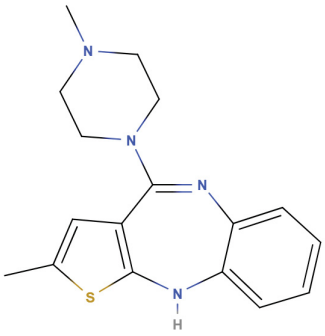
Name: Naltrexone
Family: Narcotic agonist
Use: Treatment of opioid dependence
Molecular weight: 341.4 $g\ mol^{-1}$
pKa: NA
Log P**:** 1.92
Half-life: 3 h



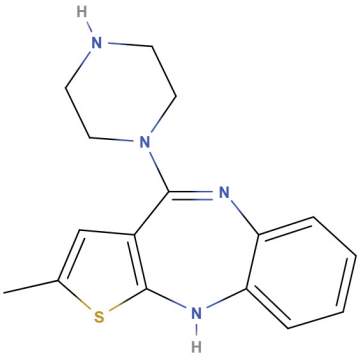
Name: Nicotine
Family: NA
Use: Part of the tobacco plant
Molecular weight: 162.2 $g\ mol^{-1}$
pKa: 3.2, 7.9
Log P**:** 1.2
Half-life: 0.5 to 2 h



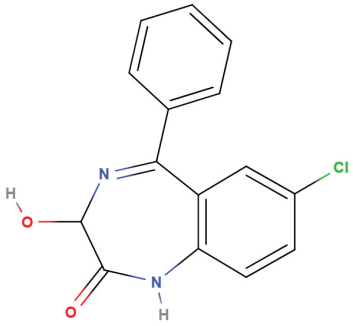
Name: Nitrazepam
Family: Benzodiazepines
Use: Treatment of anxiety and insomnia
Molecular weight: 281.3 $g\ mol^{-1}$
pKa: 3.2, 10.8
Log P**:** 2.25
Half-life: 18 to 38 h



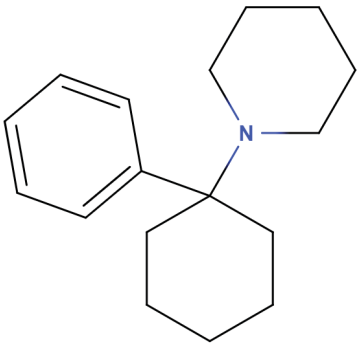
Name: Olanzapine
Family: Antipsychotic
Use: Treatment of schizophrenia and bipolar disorder
Molecular weight: 312.4 $g\ mol^{-1}$
pKa: 5.0, 7.4
Log P**:** 3.39
Half-life: 30 to 60 h



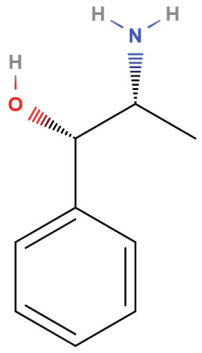
Name: N-Desmethylolanzapine
Family: NA
Use: Metabolite of olanzapine
Molecular weight: 298.4 $g\ mol^{-1}$
pKa: NA
Log P**:** 2.4
Half-life: NA



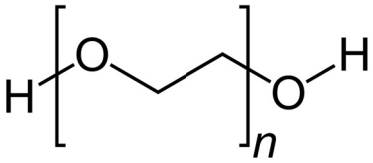
Name: Oxazepam
Family: Benzodiazepines
Use: Treatment of anxiety and insomnia
Molecular weight: 286.7 $g\ mol^{-1}$
pKa: 1.7 and 11.6
Log P**:** 2.24
Half-life: 4 to 15 h



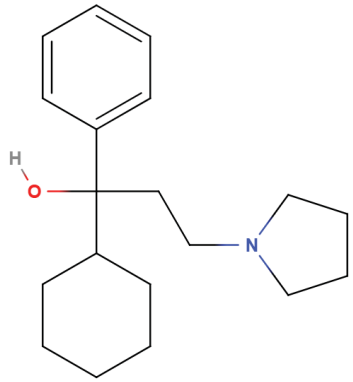
Name: Phencyclidine (PCP)
Family: Hallucinogens
Use: Drug of abuse
Molecular weight: 243.4 $g\ mol^{-1}$
pKa: 8.5
Log P**:** 4.7
Half-life: 7 to 46 h



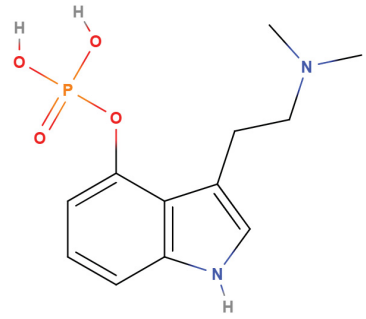
Name: Phenylpropanolamine
Family: NA
Use: Decongestant, appetite suppressant
Molecular weight: 151.2 $g\ mol^{-1}$
pKa: 9.4
Log P**:** 0.7
Half-life: 4 h



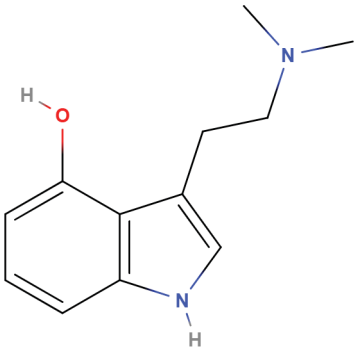
Name: Polyethyleneglycol (PEG)
Family: Polyethers
Use: Laxative, excipient, etc.
Molecular weight: $(44.05\ n + 18.02)\ g\ mol^{-1}$
pKa: NA
Log P**:** NA
Half-life: NA



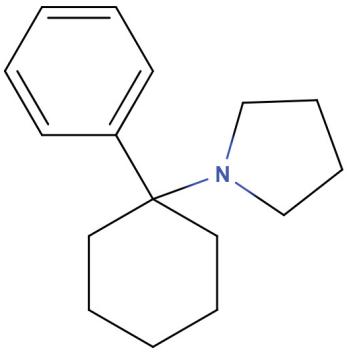
Name: Procyclidine
Family: NA
Use: Treatment of Parkinson and dystonia
Molecular weight: 287.4 $g\ mol^{-1}$
pKa: NA
Log P**:** 4.8
Half-life: 8 to 16 h



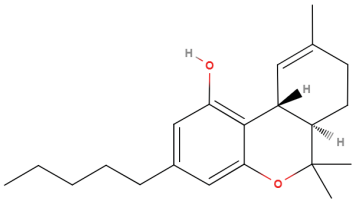
Name: Psilocybin
Family: Hallucinogens
Use: Drug of absuse
Molecular weight: 284.3 $g\ mol^{-1}$
pKa: NA
Log P**:** 1.0
Half-life: NA



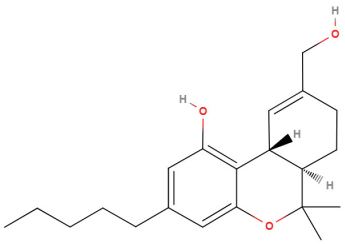
Name: Psilocin
Family: Hallucinogens
Use: Metabolite of psilocybin
Molecular weight: $g\ mol^{-1}$
pKa: NA
Log P**:** NA
Half-life: NA



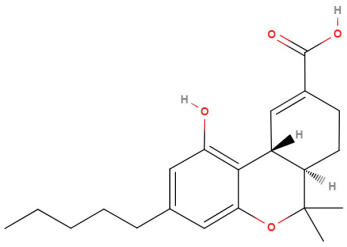
Name: Rolicyclidine
Family: Dissociative anesthetic
Use: Drug of abuse
Molecular weight: 229.4 $g\ mol^{-1}$
pKa: NA
Log P**:** NA
Half-life: NA



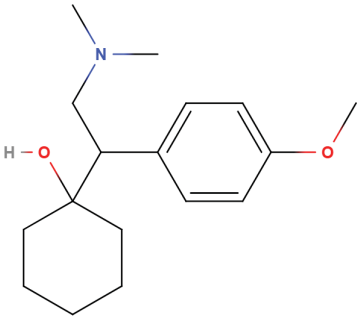
Name: Δ^9-Tetrahydrocannabinol (THC)
Family: Cannabinoids
Use: Drug of abuse, antiemetic
Molecular weight: 314.5 $g\ mol^{-1}$
pKa: 10.6
Log P**:** 7.6
Half-life: 2 to 59 h



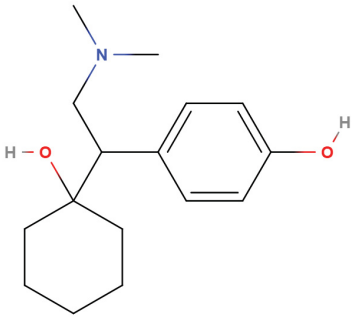
Name: 11-Hydroxy-tetrahydrocannabinol (THC-OH)
Family: Cannabinoids
Use: Active metabolite of THC
Molecular weight: 330.5 $g\ mol^{-1}$
pKa: 9.34
Log P**:** 4.71
Half-life: NA



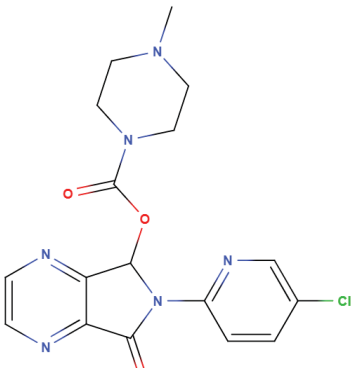
Name: 11-Nor-9-carboxy-tetrahydrocannabinol (THC-COOH)
Family: Cannabinoids
Use: Metabolite of THC
Molecular weight: 344.5 $g\ mol^{-1}$
pKa: 4.2
Log P**:** 5.14
Half-life: 5 to 6 days



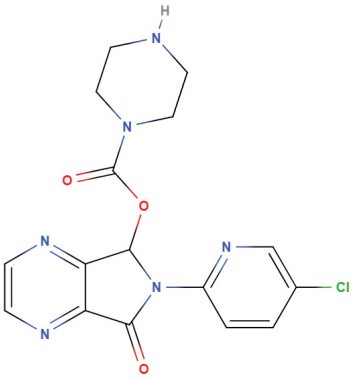
Name: Venlafaxine
Family: Antidepressant
Use: Treatment of major depression and anxiety disorders
Molecular weight: 277.4 $g\ mol^{-1}$
pKa: 14.42
Log P**:** 3.28
Half-life: 4 h



Name: O-Desmethylvenlafaxine
Family: Antidepressant
Use: Active metabolite of venlafaxine, prescribed on its own
Molecular weight: 263.4 $g\ mol^{-1}$
pKa: 9.45, 10.66
Log P**:** 2.6
Half-life: $\approx$ 11 h



Name: Zopiclone
Family: Hypnotic
Use: Treatment of insomnia
Molecular weight: 388.8 $g\ mol^{-1}$
pKa: 6.70
Log P**:** 0.8
Half-life: 3.5 to 6.5



Name: N-desmethylzopiclone
Family: NA
Use: Metabolite of zopiclone
Molecular weight: 374.8 $g\ mol^{-1}$
pKa: NA
Log P**:** NA
Half-life: NA

B

Supplementary data to "Procedure for the Selection and Validation of a Calibration Model I —Description and Application"

B.1 Supplementary data 1 – Improvement in accuracy and success rates

This spreadsheet is available for download at https://spectrum.library.concordia.ca/984860/3/Supplemental%20Data%201.xlsx.

B.2 Supplementary data 2 – R Script for simulated data generation

This R script is available for download at https://spectrum.library.concordia.ca/984860/4/Supplemental%20Data%202.pdf.

B.3 Supplementary data 3 – R scripts to perform the selection and validation of the calibration model

This collection of R scripts is available for download at https://spectrum.library.concordia.ca/984860/6/Supplemental%20Data%203.pdf.

B.4 Supplementary data 4 – Success rate under increasing %RSD

Table B.4.1: Success rate of the different tests in the process of calibration model selection and validation under increasing %RSD, using quadratic, $1/x^2$ data with 7 measurement replicates.

%RSD (LLOQ)	2.5%	5.0%	7.5%	10.0%	12.5%	15.0%	17.5%	20.0%
F-test	100%	100%	100%	100%	100%	100%	98%	100%
Variance test	100%	100%	100%	100%	100%	100%	100%	100%
Partial F-test	78%	68%	58%	54%	42%	34%	28%	26%
Validation	100%	100%	100%	100%	100%	100%	100%	100%
Final model	78%	68%	58%	54%	42%	34%	28%	26%

C

Supplementary Data to "Procedure for the Selection and Validation of a Calibration Model II —Theoretical Basis"

C.1 Supplementary data 1 – Analytical specifications and selection procedure results for all analytes

This spreadsheet is available for download at https://spectrum.library.concordia.ca/984861/2/Supplemental%20Data%201.xlsx.

C.2 Supplementary data 2 – Standardized residuals graph

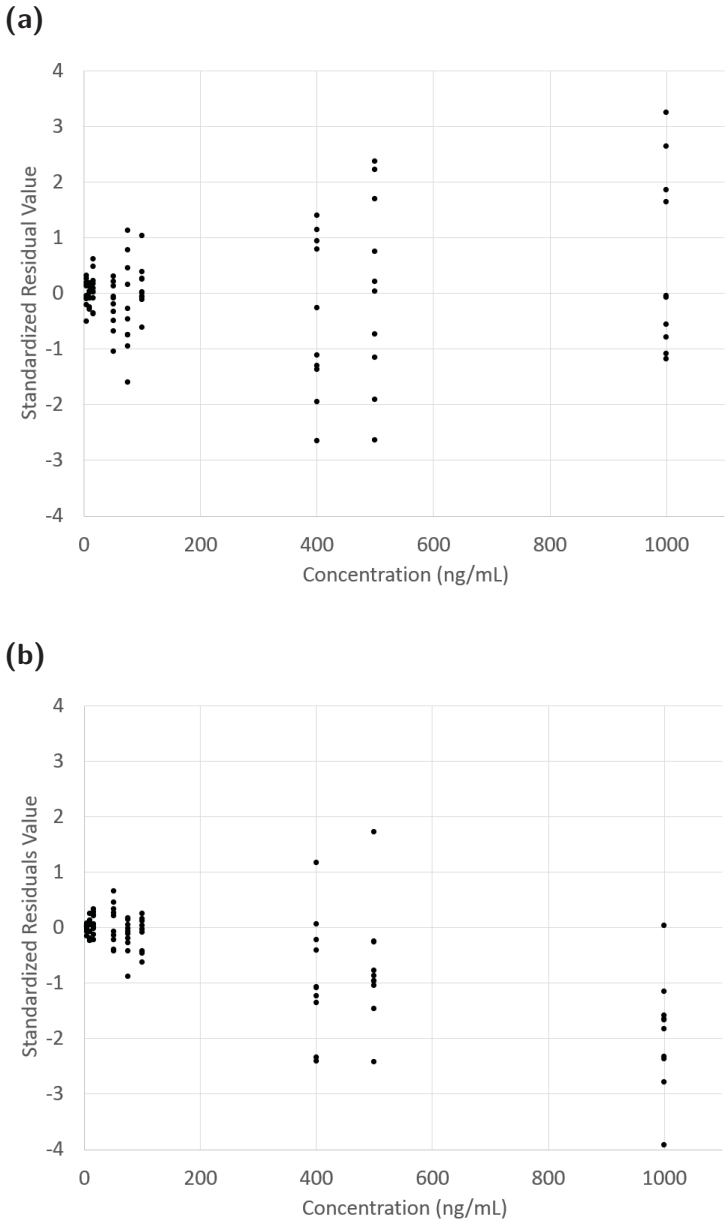


Figure C.2.1: Standardized residuals graph. **(a)** Residuals with a normal distribution according to the CVM test. **(b)** Residuals with a non-normal distribution according to the CVM test.

Note: Figure C.2.1b might also display a lack of fit to the linear model, since all but one of the residuals at 1000 ng/mL are below the $y = 0$ line.

D

Supplementary Data to "A Tool for Automatic Correction of Endogenous Concentrations: Application to BHB Analysis by LC-MS/MS and GC-MS"

The post-print version of this paper, including Supplementary Data, will be fully available to the public at https://spectrum.library.concordia.ca/985479/ on May 29th 2020 following the one year embargo period requested by Oxford Academic. In the meantime, the interested reader can find Supplementary Data on the publisher's website.

D.1 Supplementary data 1 – R Script for automatic correction of endogenous concentration

This R script with full instructions is available for download at https://academic.oup.com/jat/advance-article/doi/10.1093/jat/bkz024/5505413#supplementary-data.

D.2 Supplementary data 2 – R Script for automatic error calculation of the endogenous concentration correction

This R script with full instructions is available for download at https://academic.oup.com/jat/advance-article/doi/10.1093/jat/bkz024/5505413#supplementary-data.

D.3 Supplementary data 3 – Evaluation of the endogenous concentration using a quadratic and a linear regression

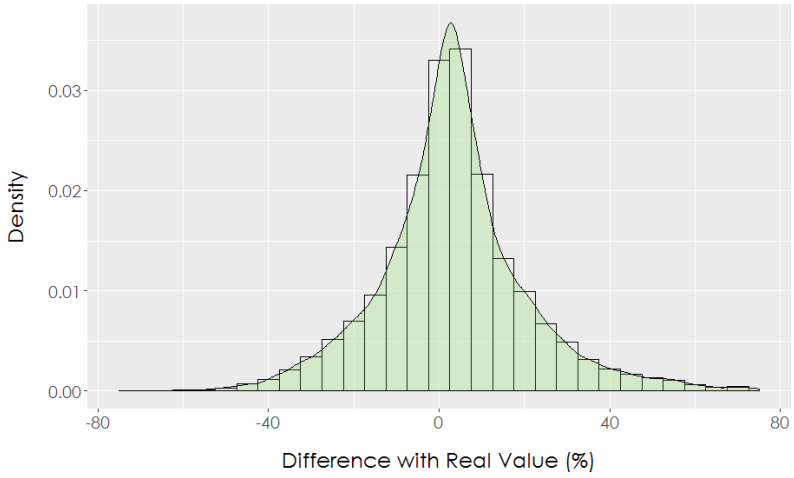


Figure D.3.1: Percent difference observed between the real endogenous value and the measured endogenous value when a linear calibration model is used on quadratic data.

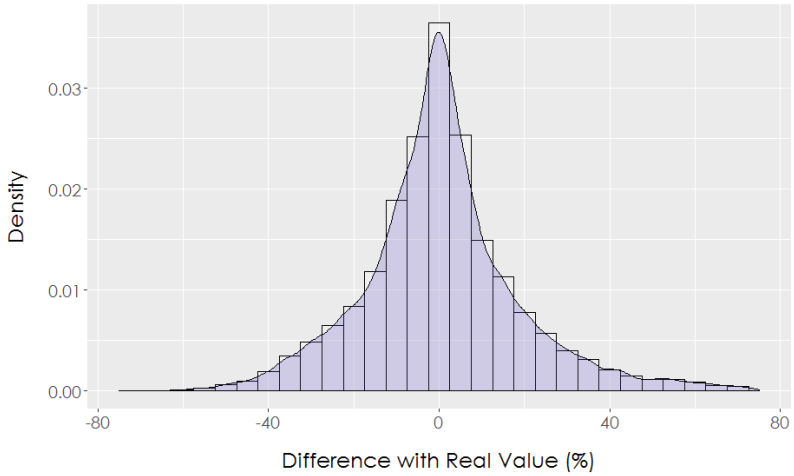


Figure D.3.2: Percent difference observed between the real endogenous value and the measured endogenous value when a quadratic calibration model is used on quadratic data.

Distribution density is more symmetrical, with a median closer to zero when the quadratic regression is used on quadratic data rather than the linear one.

The R script used to produce those histograms is available for download at https://academic.oup.com/jat/advance-article/doi/10.1093/jat/bkz024/5505413#supplementary-data.

D.4 Supplementary data 4 – *In silico* evaluation of the endogenous concentration correction process and modelisation of the error

This R Script is available for download at https://academic.oup.com/jat/advance-article/doi/10.1093/jat/bkz024/5505413#supplementary-data.

E

Supplementary Data to "Qualitative Method Validation and Uncertainty Evaluation Via the Binary Output I —Validation Guidelines and Theoretical Foundations"

The pre-print version of this paper, including Supplementary Data, is available free of charge at https://spectrum.library.concordia.ca/985532/.

E.1 Supplementary data 1 – Detailed analytical parameters

E.1.1 Cut-off and internal standards concentrations

Table E.1.1: Analyte and internal standards concentrations.

Analyte	Cut-off (ng/mL)	Internal Standard	[IS] (ng/mL)
α-Hydroxyalprazolam	20	Triazolam-D_4	100
Aripiprazole	10	Trazodone-D_6	50
3-Hydroxy Bromazepam	20	Bromazepam-D_4	450
Buprenorphine	5	Oxycodone-D_3	100
Hydroxybupropion	20	Ephedrine-D_3	100
N-Desmethylcitalopram	20	Citalopram-D_6	40
N-Desmethylclobazam	20	Flunitrazepam-D_7	20
Cocaethylene	20	Cocaine-D_3	50
Norcodeine	20	Codeine-D_3	100
N-Desmethylcyclobenzaprine	20	Amitriptyline-D_3	40
Dextrorphan	20	Oxycodone-D_3	100
Nordiazepam	20	Diazepam-D_5	100
N-Desmethyl diphenydramine	20	Diphenhydramine-D_3	200
Duloxetine	20	Amitriptyline-D_3	40
Norfentanyl	0.5	Fentanyl-D_5	10
7-Aminoflunitrazepam	20	7-Aminoclonazepam-D_4	100
N-Desmethylflunitrazepam	20	Desalkylflurazepam-D_4	100
Norfluoxetine	20	Amphetamine-D_8	100
2-Hydroxyethylflurazepam	20	Lorazepam-D_4	300
Norketamine	20	Ketamine-D_4	120
Lorazepam-glucuronide	40	Lorazepam-D_4	300
mCPP	20	Amphetamine-D_8	100
MDEA	20	MDMA-D_5	7
MDPV metabolite	20	MDPV-D_8	100
Normeperidine	40	Codeine-D_3	100
α-Hydroxymidazolam	20	Diazepam-D_5	100
N-desmethylmirtazapine	20	Amitriptyline-D_3	40
6-Acetylmorphine	5	Codeine-D_3	100
Morphine-6β-D-glucuronide	100	Morphine-D_6	100

Analyte	Cut-off (ng/mL)	Internal Standard	[IS] (ng/mL)
Naloxone	20	Oxymorphone-D_3	100
Naltrexone	20	Oxycodone-D_3	100
Desmethylolanzapine	20	Olanzapine-D_8	20
Oxazepam-glucuronide	20	Oxazepam-D_5	20
Phenylpropanolamine	30	Amphetamine-D_8	100
Norpseudoephedrine	30	Amphetamine-D_8	100
Norquetiapine	20	Quetiapine-D_8	40
7-Hydroxyquetiapine	20	Quetiapine-D_8	40
Temazepam-glucuronide	20	Diazepam-D_5	100
α-Hydroxytriazolam	20	Triazolam-D_4	100
N-Desmethylzopiclone	20	Zopiclone-D_4	100

E.1.2 Liquid chromatography method

Mobile phase A: methanol : 10 mM ammonium formate pH 3.0 (2:98 v:v)
Mobile phase B: acetonitrile
Rinsing solution: methanol : 1% formic acid in water : isopropanol (50:25:25 v:v:v)
Analytical column: Zorbax Eclipse Plus C18, 2.1 $\times$ 100 mm, 3.5 μm (Agilent)

LC pump gradient

Flow rate: 650 $\mu L/min$

Autosampler and thermostat settings

- Injection volume: 5 μL
- Wash time: 10 sec
- Autosampler temperature: 4 °C
- Column oven temperature: 50 °C

Table E.1.2: LC pump gradient.

Time (min)	A (%)	B (%)
0.00	100	0
0.50	97	3
2.00	95	5
4.00	80	20
5.50	75	25
10.00	50	50
10.20	0	100
11.30	0	100
11.31	100	0
13.00	100	0

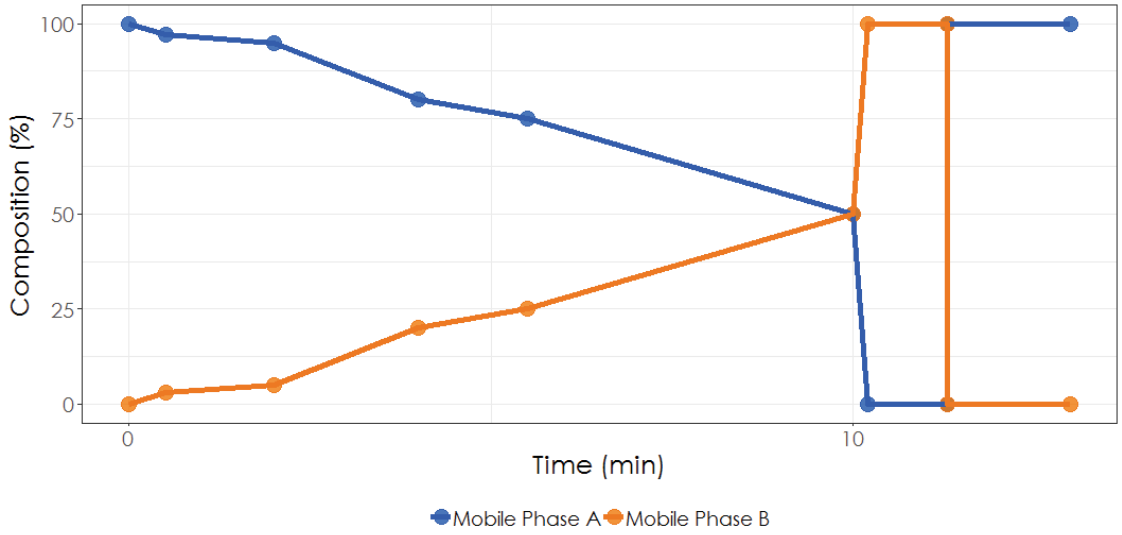


Figure E.1.1: LC pump gradient graph.

E.1.3 Mass spectrometry method

Divert valve settings

Voltages and other conditions

- Scan type: scheduled multiple reaction monitoring (MRM)
- Polarity: positive
- MRM detection window: 45 *sec*

Table E.1.3: Divert valve program.

Time (min)	Position
0	MS
0.1	Waste
0.7	MS
10.5	Waste
12.5	MS

- Ion source: Turbo Spray
- Curtain Gas (CUR): 30.0
- Collision Gas (CAD): 10
- Ion spray voltage (IS): 3000.0 V
- Source temperature: 700.0 °C
- Ion source gas 1 (GS1): 60.0
- Ion source gas (GS2): 65.0

Table E.1.4: Monitored MRM transitions.

Q1 (Da)	Q3 (Da)	Time (min)	Analyte	DP (V)	EP (V)	CE (V)	CXP (V)
281.2	117.0	8.3	Amitriptyline-D_3	76	10	32	11
144.1	127.0	3.4	Amphetamine-D_8	40	10	13	10
320.0	186.2	6.9	Bromazepam-D_4	100	10	45	18
331.2	234.0	7.2	Citalopram-D_6	85	10	40	15
290.1	254.0	4.8	7-Aminoclonazepam-D_4	80	10	30	18
308.2	185.0	5.3	Cocaine-D_3	30	10	21	15
303.3	215.0	3.5	Codeine-D_3	110	10	36	18
290.1	89.0	9.8	Diazepam-D_5	45	10	70	15
259.2	115.0	7.0	Diphenhydramine-D_3	55	10	80	10
169.1	136.0	2.9	Ephedrine-D_3	40	10	29	12
342.2	188.2	6.8	Fentanyl-D_5	110	10	34	17
321.1	275.1	8.7	Flunitrazepam-D_7	100	10	36	12
293.1	140.0	8.8	Desalkylflurazepam-D_4	110	10	40	15

Q1 (Da)	Q3 (Da)	Time (min)	Analyte	DP (V)	EP (V)	CE (V)	CXP (V)
244.1	131.0	4.6	Ketamine-D_4	64	10	43	12
327.1	281.0	8.3	Lorazepam-D_4	70	10	33	15
199.1	165.1	4.0	MDMA-D_5	60	10	20	15
285.2	175.1	5.3	MDPV-D_8	75	10	31	15
292.1	152.1	1.7	Morphine-D_6	100	10	79	14
321.1	261.1	4.0	Olanzapine-D_8	40	10	26	25
292.1	246.1	8.1	Oxazepam-D_5	60	10	30	14
319.2	259.0	3.8	Oxycodone-D_3	95	10	35	15
305.2	287.0	1.9	Oxymorphone-D_3	90	10	22	16
392.2	254.0	7.0	Quetiapine-D_8	70	10	55	15
379.2	182.1	6.2	Trazodone-D_6	90	10	35	14
347.1	208.0	8.7	Triazolam-D_4	100	10	70	18
393.1	245.0	4.9	Zopiclone-D_4	60	10	26	10
325.1	297.2	8.0	α-Hydroxyalprazolam 1	110	10	38	11
325.1	216.1	8.0	α-Hydroxyalprazolam 2	110	10	56	18
448.1	285.1	8.3	Aripiprazole 1	80	10	37	10
448.1	176.1	8.3	Aripiprazole 2	80	10	42	11
332	287.0	6.0	3-Hydroxy Bromazepam 1	60	10	30	15
332.0	303.1	6.0	3-Hydroxy Bromazepam 2	60	10	30	15
468.3	396.2	7.6	Buprenorphine 1	20	10	55	10
468.3	414.3	7.6	Buprenorphine 2	20	10	50	10
256.1	130.1	5.0	Hydroxybupropion 1	50	10	48	12
256.1	103.1	5.0	Hydroxybupropion 2	50	10	49	18
312.2	110.1	7.1	N-Desmethylcitalopram 1	70	10	30	15
312.2	263.1	7.1	N-Desmethylcitalopram 2	70	10	24	25
287.1	245.0	8.3	N-Desmethylclobazam 1	100	10	28	15
287.1	210.0	8.3	N-Desmethylclobazam 2	100	10	42	13
319.2	197.1	6.1	Cocaethylene 1	100	10	20	15
318.2	150.0	6.1	Cocaethylene 2	100	10	28	15
286.1	115.0	3.3	Norcodeine 1	100	10	90	15
286.1	128.0	3.3	Norcodeine 2	100	10	80	15
263.2	216.0	8.0	N-Desmethylcyclobenzaprine 1	60	10	51	17
263.2	232.0	8.0	N-Desmethylcyclobenzaprine 2	60	10	22	21
259.2	158.1	4.8	Dextrorphan 1	85	10	53	17

Q1 (Da)	Q3 (Da)	Time (min)	Analyte	DP (V)	EP (V)	CE (V)	CXP (V)
258.2	115.0	4.8	Dextrorphan 2	85	10	83	15
271.1	140.0	8.9	Nordiazepam 1	95	10	42	13
271.1	165.1	8.9	Nordiazepam 2	95	10	39	15
243.1	168.1	6.9	N-Desmethyl diphenydramine 1	40	10	20	15
243.1	153.1	6.9	N-Desmethyl diphenydramine 2	40	10	50	15
298.1	154.0	8.2	Duloxetine 1	60	10	9	9
298.1	44.0	8.2	Duloxetine 2	60	10	50	10
233.2	84.1	4.6	Norfentanyl 1	60	10	26	10
233.2	150.1	4.6	Norfentanyl 2	60	10	25	13
284.1	240.0	5.5	7-Aminoflunitrazepam 1	120	10	51	20
284.1	226.1	5.5	7-Aminoflunitrazepam 2	120	10	40	20
300.1	254.0	7.9	N-Desmethylflunitrazepam 1	85	10	36	15
300.1	198.1	7.9	N-Desmethylflunitrazepam 2	85	10	55	15
296.1	134.2	8.5	Norfluoxetine 1	65	10	10	10
296.1	296.1	8.5	Norfluoxetine 2	65	10	5	10
333.1	211.1	8.5	2-Hydroxyethylflurazepam 1	120	10	51	15
333.1	109.1	8.5	2-Hydroxyethylflurazepam 2	120	10	40	14
224.1	179.3	4.5	Norketamine 1	70	10	20	15
225.1	126.0	4.5	Norketamine 2	70	10	35	20
497.1	321.0	7.0	Lorazepam-glucuronide 1	70	10	20	24
497.1	275.1	7.0	Lorazepam-glucuronide 2	70	10	57	16
197.1	154.1	4.9	mCPP 1	75	10	20	13
197.1	118.0	4.9	mCPP 2	75	10	29	14
209.1	164.0	4.4	MDEA 1	60	10	14	15
209.1	136.1	4.4	MDEA 2	60	10	30	15
279.2	176.1	4.6	MDPV metabolite 1	50	10	23	20
279.2	127.2	4.6	MDPV metabolite 2	50	10	23	15
235.1	161.1	5.5	Normeperidine 1	60	10	22	15
234.1	115.1	5.5	Normeperidine 2	60	10	85	11
253.2	196.2	5.0	N-desmethylmirtazapine 1	70	10	31	15
252.2	209.2	5.0	N-desmethylmirtazapine 2	70	10	22	20
342.1	168.1	7.3	α-Hydroxymidazolam 1	110	10	55	15
342.1	140.0	7.3	α-Hydroxymidazolam 2	110	10	84	12
328.2	165.0	4.0	6-Acetylmorphine 1	80	10	55	10

Q1 (Da)	Q3 (Da)	Time (min)	Analyte	DP (V)	EP (V)	CE (V)	CXP (V)
328.2	211.0	4.0	6-Acetylmorphine 2	80	10	37	10
462.1	286.1	1.6	Morphine-6β-D-glucuronide 1	50	10	44	14
462.1	165.0	1.6	Morphine-6β-D-glucuronide 2	50	10	71	15
328.2	212.2	3.4	Naloxone 1	100	10	55	15
328.2	253.2	3.4	Naloxone 2	100	10	38	15
342.2	270.1	3.9	Naltrexone 1	100	10	38	15
342.2	212.0	3.9	Naltrexone 2	100	10	60	20
299.1	213.1	3.9	Desmethylolanzapine 1	140	10	30	11
299.1	198.1	3.9	Desmethylolanzapine 2	140	10	42	12
463.1	241.1	6.5	Oxazepam-glucuronide 1	80	10	55	20
463.1	287.1	6.5	Oxazepam-glucuronide 2	80	10	25	20
152.1	117.1	2.3	Phenylpropanolamine 1	40	10	23	12
152.1	91.0	2.3	Phenylpropanolamine 2	40	10	47	16
152.1	134.1	2.5	Norpseudoephedrine 1	35	10	14	10
152.1	117.0	2.5	Norpseudoephedrine 2	35	10	23	10
297.1	140.1	6.9	Norquetiapine 1	70	10	80	15
297.1	184.0	6.9	Norquetiapine 2	70	10	53	15
401.1	270.0	4.7	7-Hydroxyquetiapine 1	30	10	35	18
401.1	209.1	4.7	7-Hydroxyquetiapine 2	30	10	60	20
477.1	301.1	7.2	Temazepam-glucuronide 1	85	10	21	9
477.1	255.1	7.2	Temazepam-glucuronide 2	85	10	60	9
359	331.0	8.0	α-Hydroxytriazolam 1	100	10	40	10
359.0	176.0	8.0	α-Hydroxytriazolam 2	100	10	38	16
375.1	245.1	4.9	N-Desmethylzopiclone 1	70	10	25	10
375.1	217.0	4.9	N-Desmethylzopiclone 2	70	10	50	10

The method is validated on the following systems: HPLC Agilent 1200 series and 1260 Infinity, LC-MS/MS 5500 QTRAP Sciex.

The data acquisition software used is Analyst® 1.6.2 build 8489 and the data analysis software used is Multiquant® 3.0.1 (Version 3.0.6256.0)

E.2 Supplementary data 2 – Modeling of qualitative decision point methods

This R script is available for download at https://spectrum.library.concordia.ca/985532/12/Supplementary%20Data%202_VF.pdf.

E.3 Supplementary data 3 – Normality of measurements

The area ratio (analyte peak area to internal standard peak area) was measured for 30 different matrices (antemortem and postmortem blood) spiked at the cut-off concentration specified in Appendix E.1.

Following outlier removal subsequent to a Grubbs test for 3 analytes (buprenorphine, norcodeine and temazepam-glucuronide), normality testing using the Cramer-von Mises test was performed in RStudio.

```
# Normality_Ratios.R
# A script to test the normality of measurements for the 40 qualitative
    analytes studied.
# By Brigitte Desharnais, last modification 2019-01-08.

# Set working directory.
setwd("E:/RECHERCHE/QUALITATIF")

# Load necessary packages.
library(dplyr)
library(nortest)

# Import data from an Excel table copied in the clipboard.
Data <- read.delim("clipboard", header = TRUE, sep = "\t", dec = ".")
Data <- tbl_df(Data)

# Import the list of analytes copied in the clipboard.
Analytes <- read.delim("clipboard", header = FALSE, sep = "\t", dec = ".")
Analytes <- as.character(Analytes$V1)

# Create the empty results matrix.
Results <- matrix(nrow = 40, ncol = 1)

```

```r
# Perform CVM test for each analyte and store result in the matrix.
for(i in 1:40){
  # Create a temporary data frame storing only results for the studied analyte.
  Temp <- Data %>% filter(Analyte == Analytes[i])

  # Perform the CVM test.
  CVM <- cvm.test(as.numeric(Temp$Area.Ratio))

  # Store p-value in results matrix.
  Results[i, 1] <- CVM$p.value
}

# Create final results matrix by appending analyte names.
Results <- cbind(Analytes, Results)
```

The following results were obtained.

Table E.3.1: Cramer-von Mises normality test results.

Analytes	P-value
α-Hydroxyalprazolam	0.1059
Aripiprazole	0.1204
3-Hydroxy Bromazepam	0.8826
Buprenorphine	0.4012
Hydroxybupropion	0.4618
N-Desmethylcitalopram	0.2681
N-Desmethylclobazam	0.6332
Cocaethylene	0.5719
Norcodeine	0.1240
N-Desmethylcyclobenzaprine	0.4606
Dextrorphan	0.4596
Nordiazepam	0.6851
N-Desmethyl diphenydramine	0.5524
Duloxetine	0.7734
Norfentanyl	0.8879
7-Aminoflunitrazepam	0.3235
N-Desmethylflunitrazepam	0.0321

Analytes	P-value
Norfluoxetine	0.2267
2-Hydroxyethylflurazepam	0.5575
Norketamine	0.6207
Lorazepam-glucuronide	0.8237
mCPP	0.0749
MDEA	0.2117
MDPV metabolite	0.1189
Normeperidine	0.4355
α-Hydroxymidazolam	0.2231
N-desmethylmirtazapine	0.7960
6-Acetylmorphine	0.1912
Morphine-6β-D-glucuronide	0.0325
Naloxone	0.1174
Naltrexone	0.8822
Desmethylolanzapine	0.3207
Oxazepam-glucuronide	0.8373
Phenylpropanolamine	0.2928
Norpseudoephedrine	0.7128
Norquetiapine	0.3976
7-Hydroxyquetiapine	0.7979
Temazepam-glucuronide	0.2592
α-Hydroxytriazolam	0.9219
N-Desmethylzopiclone	0.8527

All but two analytes (N-desmethylflunitrazepam and morphine-6β-D-glucuronide) have $P < 0.05$, indicating that there is no significant departure from normality for the vast majority of analytes.

For the two remaining analytes, quantile-quantile plots are shown in Figure E.3.1.

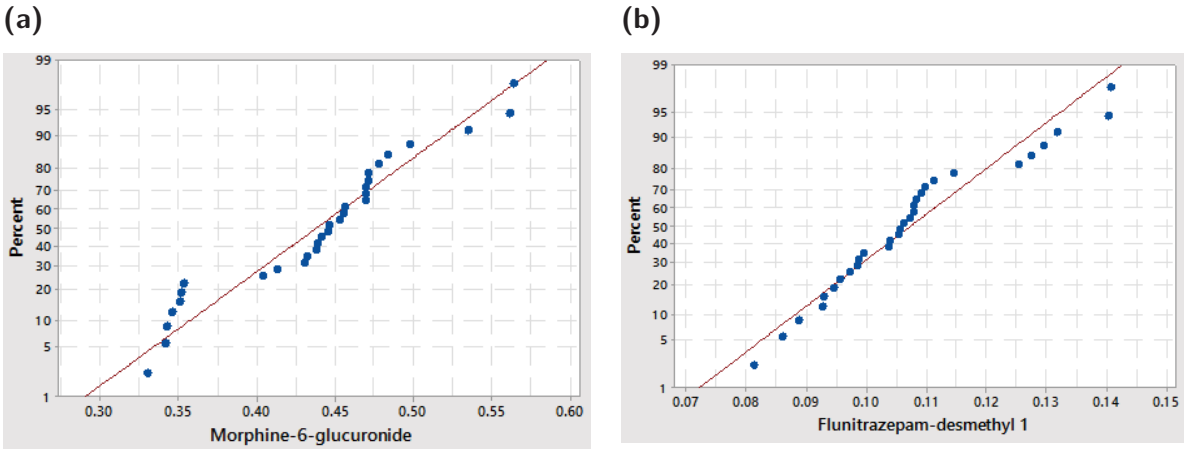


Figure E.3.1: Normal quantile-quantile plots. **(a)** Morphine-6β-D-glucuronide. **(b)** N-Desmethylflunitrazepam.

E.4 Supplementary data 4 – Changing variance of measurements at cut-off

The full spreadsheet is available for download at https://spectrum.library.concordia.ca/985532/25/Supplementary%20Data%204_VF.xlsx. A summary of results is presented here.

Levene's test was performed using the `levene.test()` function of the `lawstat` package in RStudio. $P < 0.05$ indicates a significant difference between the variances studied.

Table E.4.1: Area ratio variance over several days.

Analyte	Variance				Levene
	A	B	C	D	P-value
α-Hydroxyalprazolam	9.86×10^{-4}	2.58×10^{-3}	1.49×10^{-2}	1.98×10^{-4}	0.0100
Aripiprazole	4.64×10^{-3}	1.02×10^{-2}	5.16×10^{-2}	1.37×10^{-3}	0.0073
3-Hydroxy Bromazepam	2.57×10^{-3}	3.96×10^{-4}	1.39×10^{-3}	9.34×10^{-5}	0.2460
Buprenorphine	9.19×10^{-6}	3.04×10^{-5}	6.56×10^{-5}	3.38×10^{-6}	0.0025
Hydroxybupropion	4.34×10^{-4}	7.57×10^{-4}	2.52×10^{-3}	1.50×10^{-4}	0.0524
N-Desmethylcitalopram	3.16×10^{-4}	5.42×10^{-4}	1.71×10^{-3}	6.90×10^{-5}	0.0246
N-Desmethylclobazam	8.28×10^{-4}	1.28×10^{-3}	2.95×10^{-3}	1.09×10^{-4}	0.0302
Cocaethylene	2.86×10^{-3}	5.27×10^{-3}	2.90×10^{-2}	5.59×10^{-4}	0.0307
Norcodeine	1.42×10^{-4}	3.93×10^{-4}	8.48×10^{-5}	5.16×10^{-5}	0.0384

Analyte	Variance				Levene
	A	B	C	D	*P*-value
N-Desmethylcyclobenzaprine	3.39×10^{-4}	7.49×10^{-4}	5.64×10^{-3}	9.10×10^{-5}	0.0027
Dextrorphan	1.61×10^{-4}	3.96×10^{-4}	1.28×10^{-3}	4.79×10^{-5}	0.0294
Nordiazepam	3.55×10^{-3}	1.76×10^{-2}	8.68×10^{-2}	7.84×10^{-4}	0.0038
N-Desmethyl diphenydramine	5.25×10^{-5}	1.68×10^{-4}	8.84×10^{-4}	1.32×10^{-5}	0.0148
Duloxetine	2.56×10^{-4}	8.36×10^{-4}	7.52×10^{-3}	1.30×10^{-4}	0.0018
Norfentanyl	4.03×10^{-5}	6.69×10^{-5}	1.70×10^{-4}	1.13×10^{-5}	0.0090
7-Aminoflunitrazepam	6.95×10^{-5}	1.08×10^{-4}	4.62×10^{-4}	8.68×10^{-6}	0.0328
N-Desmethylflunitrazepam	2.44E-04	7.33×10^{-4}	1.11×10^{-3}	5.91×10^{-5}	0.0067
Norfluoxetine	5.84×10^{-5}	1.25×10^{-4}	5.18×10^{-4}	1.75×10^{-5}	0.0081
2-Hydroxyethylflurazepam	5.45×10^{-5}	1.73×10^{-4}	8.08×10^{-4}	3.04×10^{-5}	0.0022
Norketamine	2.85×10^{-5}	1.59×10^{-4}	6.78×10^{-4}	9.43×10^{-6}	0.0078
Lorazepam-glucuronide	4.55×10^{-6}	5.64×10^{-6}	2.47×10^{-5}	2.39×10^{-6}	0.1066
mCPP	6.80×10^{-5}	1.89×10^{-4}	7.69×10^{-4}	1.64×10^{-5}	0.0385
MDEA	3.42×10^{-4}	9.79×10^{-4}	6.64×10^{-3}	1.73×10^{-4}	0.0226
MDPV metabolite	4.48×10^{-4}	7.81×10^{-4}	3.42×10^{-3}	2.13×10^{-4}	0.0470
Normeperidine	3.02×10^{-3}	5.54×10^{-3}	2.22×10^{-2}	4.53×10^{-4}	0.0342
α-Hydroxymidazolam	9.06×10^{-4}	4.33×10^{-3}	2.79×10^{-2}	2.95×10^{-4}	0.0021
N-desmethylmirtazapine	1.29×10^{-4}	4.77×10^{-4}	2.62×10^{-3}	8.75×10^{-5}	0.0181
6-Acetylmorphine	1.90×10^{-4}	4.33×10^{-4}	9.12×10^{-4}	4.85×10^{-5}	0.0377
Morphine-6 -D-glucuronide	1.41×10^{-3}	6.18×10^{-3}	8.82×10^{-3}	7.64×10^{-4}	0.0013
Naloxone	1.81×10^{-4}	2.47×10^{-4}	1.13×10^{-3}	2.43×10^{-5}	0.0228
Naltrexone	2.80×10^{-4}	1.32×10^{-3}	2.38×10^{-3}	7.83×10^{-5}	0.0096
Desmethylolanzapine	7.67×10^{-2}	6.90×10^{-2}	6.50×10^{-2}	1.21×10^{-3}	0.0011
Oxazepam-glucuronide	7.78×10^{-5}	4.69×10^{-5}	2.15×10^{-4}	1.57×10^{-5}	0.0036
Phenylpropanolamine	NA	4.56×10^{-4}	1.63×10^{-3}	3.98×10^{-5}	0.0001
Norpseudoephedrine	1.19×10^{-3}	3.68×10^{-3}	1.15×10^{-2}	1.50×10^{-4}	0.0037
Norquetiapine	6.90×10^{-4}	8.28×10^{-4}	1.47×10^{-3}	3.82×10^{-4}	0.3506
7-Hydroxyquetiapine	6.14×10^{-4}	2.08×10^{-3}	1.04×10^{-2}	2.09×10^{-4}	0.0053
Temazepam-glucuronide	2.48×10^{-4}	6.25×10^{-4}	1.47×10^{-3}	1.69×10^{-4}	0.0415
α-Hydroxytriazolam	3.28×10^{-4}	1.31×10^{-3}	7.62×10^{-3}	2.41×10^{-4}	0.0001
N-Desmethylzopiclone	5.14×10^{-5}	6.82×10^{-5}	3.35×10^{-4}	8.98×10^{-6}	0.0331

F

Supplementary Data to "Qualitative Method Validation and Uncertainty Estimation Via the Binary Output II —Application to a Multi-Analyte LC-MS/MS Method for Oral Fluid"

The pre-print version of this paper, including Supplementary Data, is available free of charge at https://spectrum.library.concordia.ca/985534/.

F.1 Supplementary data 1 – Detailed analytical parameters

F.1.1 Cut-off and internal standards concentrations

In Table F.1.1, [IS] is the concentration of internal standard (IS) in the IS solution added to samples, not the final concentration in the sample.

Table F.1.1: Analyte and internal standards concentrations.

Analyte	Cut-off (ng/mL)	Internal Standard	[IS] (ng/mL)
Alprazolam	8	Lorazepam-D_4	3300
α-Hydroxyalprazolam	20	Triazolam-D_4	1100
Amitriptyline	30	Amitriptyline-D_3	440
Amphetamine	15	Amphetamine-D_8	1100
Aripiprazole	10	Trazodone-D_6	550
Bromazepam	75	Bromazepam-D_4	4950
Buprenorphine	5	Oxycodone-D_3	1100
Bupropion	15	Ephedrine-D_3	1100
Hydroxybupropion	20	Ephedrine-D_3	1100
Chlordiazepoxide	15	Chlordiazepoxide-D_5	1100
Citalopram	30	Citalopram-D_6	440
N-Desmethylcitalopram	20	Citalopram-D_6	440
Clobazam	15	Flunitrazepam-D_7	220
N-Desmethylclobazam	20	Flunitrazepam-D_7	220
Clonazepam	8	Flunitrazepam-D_7	220
7-Aminoclonazepam	8	7-Aminoclonazepam-D_4	1100
Cocaethylene	8	Cocaine-D_3	550
Cocaine	8	Cocaine-D_3	550
Benzoylecgonine	8	Cocaine-D_3	550
Codeine	5	Codeine-D_3	1100
Norcodeine	20	Codeine-D_3	1100
Cyclobenzaprine	30	Amitriptyline-D_3	440
N-Desmethylcyclobenzaprine	20	Amitriptyline-D_3	440
Demoxepam	15	Demoxepam-D_5	1100
Dextromethorphane	15	Oxymorphone-D_3	1100
Dextrorphane	20	Oxycodone-D_3	1100
Diazepam	8	Diazepam-D_5	1100
Nordiazepam	8	Diazepam-D_5	1100
Diphenhydramine	30	Diphenhydramine-D_3	2200
N-Desmethyldiphenhydramine	20	Diphenhydramine-D_3	2200
Duloxetine	20	Amitriptyline-D_3	440
EDDP	30	Diphenhydramine-D_3	2200

Analyte	Cut-off (ng/mL)	Internal Standard	[IS] (ng/mL)
Ephedrine	15	Ephedrine-D_3	1100
Fentanyl	0.5	Fentanyl-D_5	110
Acetyl fentanyl	1.5	Fentanyl-D_5	110
Norfentanyl	0.5	Fentanyl-D_5	110
Flunitrazepam	15	Flunitrazepam-D_7	220
7-Aminoflunitrazepam	20	7-Aminoclonazepam-D_4	1100
N-Desmethylflunitrazepam	20	Desalkylflurazepam-D_4	1100
Fluoxetine	30	Amitriptyline-D_3	440
Norfluoxetine	20	Amphetamine-D_8	1100
Flurazepam	6	Lorazepam-D_4	3300
Desalkylflurazepam	30	Desalkylflurazepam-D_4	1100
2-Hydroxyethylflurazepam	20	Lorazepam-D_4	3300
Hydrocodone	5	Oxycodone-D_3	1100
Hydromorphone	5	Oxymorphone-D_3	1100
Ketamine	30	Ketamine-D_4	1320
Norketamine	20	Ketamine-D_4	1320
Lorazepam	30	Lorazepam-D_4	3300
mCPP	20	Amphetamine-D_8	1100
MDA	20	Amphetamine-D_8	1100
MDEA	20	MDMA-D_5	77
MDMA	15	MDMA-D_5	77
MDPV	15	MDPV-D_8	1100
MDPV metabolite 1	20	MDPV-D_8	1100
Meperidine	15	Cocaine-D_3	550
Normeperidine	40	Codeine-D_3	1100
Methadone	10	Amitriptyline-D_3	440
Methamphetamine	15	Methamphetamine-D_5	110
Midazolam	15	Lorazepam-D_4	3300
α-Hydroxymidazolam	20	Diazepam-D_5	1100
Mirtazapine	30	Amitriptyline-D_3	440
N-Desmethylmirtazapine	20	Amitriptyline-D_3	440
Morphine	5	Morphine-D_6	1100
6-Acetylmorphine	2	Codeine-D_3	1100
Naloxone	20	Oxymorphone-D_3	1100

Analyte	Cut-off (ng/mL)	Internal Standard	[IS] (ng/mL)
Naltrexone	20	Oxycodone-D_3	1100
Nitrazepam	15	Nitrazepam-D_5	1100
7-Aminonitrazepam	15	7-Aminoclonazepam-D_4	1100
Nortriptyline	15	Amitriptyline-D_3	440
Olanzapine	15	Olanzapine-D_8	220
N-Desmethylolanzapine	20	Olanzapine-D_8	220
Oxazepam	8	Oxazepam-D_5	220
Oxycodone	5	Oxycodone-D_3	1100
Oxymorphone	5	Oxymorphone-D_3	1100
Paroxetine	15	Amitriptyline-D_3	440
PCP	6	PCP-D_5	220
Procyclidine	15	Citalopram-D_6	440
Pseudoephedrine	15	Ephedrine-D_3	1100
Quetiapine	30	Quetiapine-D_8	440
Norquetiapine	20	Quetiapine-D_8	440
7-Hydroxyquetiapine	20	Quetiapine-D_8	440
Risperidone	15	Trazodone-D_6	550
9-Hydroxyrisperidone	15	Trazodone-D_6	550
Rolicyclidine	9	Diphenhydramine-D_3	2200
Sertraline	9	Amitriptyline-D_3	440
Temazepam	30	Lorazepam-D_4	3300
THC	1.5	THC-D_3	110
Tramadol	10	Tramadol-D_3	2200
O-Desmethyl-cis-tramadol	10	Tramadol-D_3	2200
Trazodone	30	Trazodone-D_6	550
Triazolam	15	Triazolam-D_4	1100
α-Hydroxytriazolam	20	Triazolam-D_4	1100
Venlafaxine	30	Venlafaxine-D_6	2200
O-Desmethylvenlafaxine	30	Venlafaxine-D_6	2200
Zopiclone	9	Zopiclone-D_4	1100
N-Desmethylzopiclone	20	Zopiclone-D_4	1100

F.1.2 General method

Liquid chromatography

Mobile phase A: methanol : 10 mM ammonium formate pH 3.0 (2:98 v:v)
Mobile phase B: acetonitrile (LC-MS Grade)
Rinsing solution: methanol : 1% formic acid in water : isopropanol (50:25:25 v:v:v)
Analytical column: Zorbax Eclipse Plus C18, 2.1 × 100 mm, 3.5 μm (Agilent)
Flow rate: 650 $\mu L/min$

Table F.1.2: LC pump gradient.

Time (min)	A (%)	B (%)
0.00	100	0
0.50	97	3
2.00	95	5
4.00	80	20
5.50	75	25
10.00	50	50
10.20	0	100
11.30	0	100
11.31	100	0
13.00	100	0

Autosampler and thermostat Settings

- Injection volume: 5 μL
- Wash time: 10 sec
- Autosampler temperature: 4 °C
- Column oven temperature: 50 °C

Mass spectrometry method

- Scan type: scheduled multiple reaction monitoring (MRM)

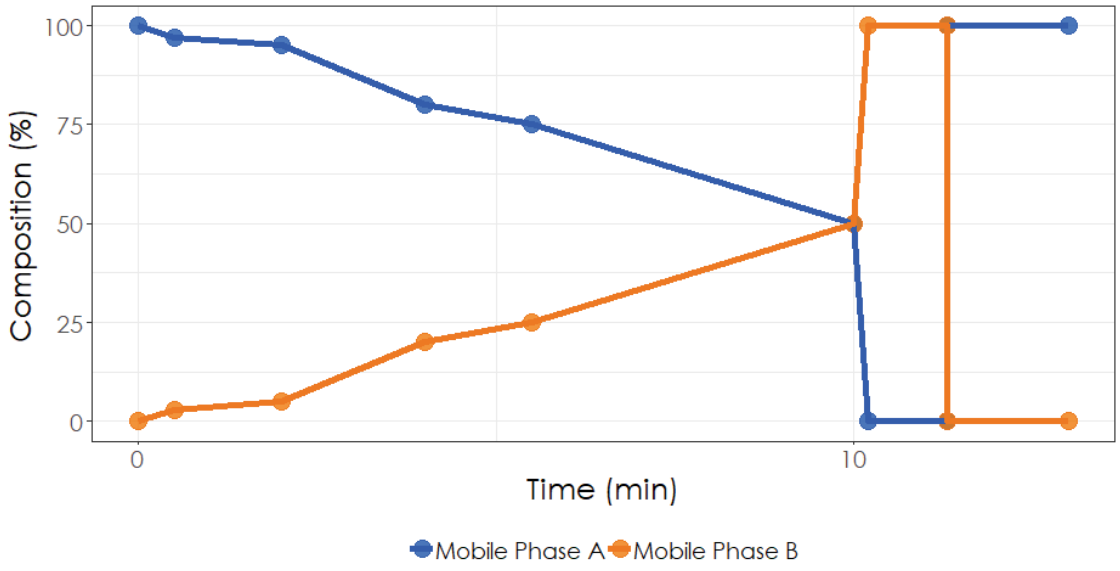


Figure F.1.1: LC pump gradient graph.

Table F.1.3: Divert valve program.

Time (min)	Position
0.0	MS
0.1	Waste
1.5	MS
10.5	Waste
12.5	MS

- Polarity: positive
- MRM detection window: 45 *sec*
- Target scan time: 0.6 *sec*
- Break: 3.00 *msec*
- Acquisition time: 13 minutes
- Ion source: Turbo Spray
- Curtain gas (CUR): 30.0
- Collision gas (CAD): 10.0
- Ion spray voltage (IS): 3000.0 V

- Source temperature (TEM): 700.0 °C
- Ion source gas 1 (GS1): 60.0
- Ion source gas (GS2): 65.0

Table F.1.4: Monitored MRM transitions.

Q1 (Da)	Q3 (Da)	Time (min)	Analyte	DP (V)	EP (V)	CE (V)	CXP (V)
309.1	205.0	8.6	Alprazolam 1	140	10	50	15
311.1	283.1	8.6	Alprazolam 2	140	10	37	12
325.1	297.2	8.0	α-Hydroxyalprazolam 1	110	10	38	11
325.1	216.1	8.0	α-Hydroxyalprazolam 2	110	10	56	18
279.2	118.0	8.3	Amitriptyline 1	76	10	32	11
279.2	106.0	8.3	Amitriptyline 2	76	10	30	11
136.1	119.0	3.5	Amphetamine 1	40	10	13	10
136.1	91.1	3.5	Amphetamine 2	40	10	28	10
448.1	285.1	8.3	Aripiprazole 1	80	10	37	10
448.1	176.1	8.3	Aripiprazole 2	80	10	42	11
290.1	105.0	4.5	Benzoylecgonine 1	70	10	26	10
290.1	82.0	4.5	Benzoylecgonine 2	70	10	25	10
316.0	182.2	7.0	Bromazepam 1	100	10	45	18
316.0	209.1	7.0	Bromazepam 2	100	10	38	18
240.1	139.0	5.7	Bupropion 1	55	10	29	10
240.1	103.0	5.7	Bupropion 2	55	10	45	10
256.1	130.1	5.0	Hydroxybupropion 1	50	10	48	12
256.1	103.1	5.0	Hydroxybupropion 2	50	10	49	18
300.1	227.1	6.5	Chlordiazepoxide 1	65	10	35	20
300.1	165.1	6.5	Chlordiazepoxide 2	65	10	70	13
326.2	235.0	7.2	Citalopram 1	85	10	40	15
326.2	263.1	7.2	Citalopram 2	85	10	22	15
312.2	110.1	7.1	N-Desmethylcitalopram 1	70	10	30	15
312.2	263.1	7.1	N-Desmethylcitalopram 2	70	10	24	25
301.1	224.1	9.1	Clobazam 1	100	10	47	15
301.1	259.0	9.1	Clobazam 2	100	10	20	15
287.1	245.0	8.3	N-Desmethylclobazam 1	100	10	28	15
287.1	210.0	8.3	N-Desmethylclobazam 2	100	10	42	13

Q1 (Da)	Q3 (Da)	Time (min)	Analyte	DP (V)	EP (V)	CE (V)	CXP (V)
316.1	270.0	8.3	Clonazepam 1	90	10	36	15
316.1	214.0	8.3	Clonazepam 2	90	10	53	19
286.1	250.1	4.9	7-Aminoclonazepam 1	80	10	30	18
286.1	94.1	4.9	7-Aminoclonazepam 2	80	10	56	13
305.2	183.1	5.3	Cocaine 1	30	10	21	15
304.2	91.0	5.3	Cocaine 2	30	10	38	12
319.2	197.1	6.1	Cocaethylene 1	100	10	20	15
318.2	150.0	6.1	Cocaethylene 2	100	10	28	15
300.2	215.0	3.4	Codeine 1	110	10	36	18
300.2	152.0	3.4	Codeine 2	110	10	91	18
286.1	115.0	3.3	Norcodeine 1	100	10	90	15
286.1	128.0	3.3	Norcodeine 2	100	10	80	15
277.2	232.1	8.0	Cyclobenzaprine 1	85	10	19	15
276.1	115.1	8.0	Cyclobenzaprine 2	85	10	33	12
263.2	216.0	8.0	N-Desmethylcyclobenzaprine 1	60	10	51	17
263.2	232.0	8.0	N-Desmethylcyclobenzaprine 2	60	10	22	21
287.1	241.0	7.4	Demoxepam 1	100	10	50	15
287.1	115.1	7.4	Demoxepam 2	100	10	33	15
273.2	216.1	6.7	Dextromethorphane 1	120	10	35	20
273.2	172.0	6.7	Dextromethorphane 2	120	10	52	15
259.2	158.1	4.8	Dextrorphan 1	85	10	53	17
259.8	115.0	4.8	Dextrorphan 2	85	10	83	15
285.1	89.0	9.8	Diazepam 1	45	10	70	15
285.1	223.1	9.8	Diazepam 2	45	10	37	12
298.1	154.0	8.2	Duloxetine 1	60	10	9	9
298.1	44.0	8.2	Duloxetine 2	60	10	50	10
271.1	140.0	8.9	Nordiazepam 1	95	10	42	13
271.1	165.1	8.9	Nordiazepam 2	95	10	39	15
256.2	115.0	7.0	Diphenhydramine 1	55	10	70	10
256.2	128.1	7.0	Diphenhydramine 2	55	10	70	10
243.1	168.1	6.9	N-Desmethyldiphenhydramine 1	40	10	20	15
243.1	153.1	6.9	N-Desmethyldiphenhydramine 2	40	10	50	15
279.2	250.1	7.5	EDDP 1	100	10	26	15
279.2	116.0	7.5	EDDP 2	100	10	95	13

Q1 (Da)	Q3 (Da)	Time (min)	Analyte	DP (V)	EP (V)	CE (V)	CXP (V)
166.1	133.0	2.9	Ephedrine 1	40	10	29	12
166.1	117.0	2.9	Ephedrine 2	40	10	29	10
337.2	188.2	6.8	Fentanyl 1	110	10	34	17
337.2	105.1	6.8	Fentanyl 2	110	10	55	14
233.2	84.1	4.6	Norfentanyl 1	60	10	26	10
233.2	150.1	4.6	Norfentanyl 2	60	10	25	13
323.2	188.2	5.9	Acetyl fentanyl 1	100	10	33	14
323.2	105.0	5.9	Acetyl fentanyl 2	100	10	55	12
314.1	268.1	8.8	Flunitrazepam 1	100	10	36	12
314.1	183.0	8.8	Flunitrazepam 2	100	10	75	15
284.1	240.0	5.5	7-Aminoflunitrazepam 1	120	10	51	20
284.1	226.1	5.5	7-Aminoflunitrazepam 2	120	10	40	20
300.1	254.0	7.9	N-Desmethylflunitrazepam 1	85	10	36	15
300.1	198.1	7.9	N-Desmethylflunitrazepam 2	85	10	55	15
310.1	148.0	8.6	Fluoxetine 1	40	10	10	15
310.1	44.0	8.6	Fluoxetine 2	130	10	23	10
296.1	134.2	8.5	Norfluoxetine 1	65	10	10	10
296.1	296.1	8.5	Norfluoxetine 2	65	10	5	10
389.2	316.0	7.1	Flurazepam 1	100	10	35	20
389.2	134.0	7.1	Flurazepam 2	100	10	75	16
289.1	140.0	8.8	Desalkylflurazepam 1	110	10	40	15
289.1	226.1	8.8	Desalkylflurazepam 2	110	10	40	12
333.1	211.1	8.5	2-Hydroxyethylflurazepam 1	120	10	51	15
333.1	109.1	8.5	2-Hydroxyethylflurazepam 2	120	10	40	14
300.2	128.2	3.9	Hydrocodone 1	90	10	78	15
300.2	115.0	3.9	Hydrocodone 2	90	10	85	14
286.1	185.1	2.2	Hydromorphone 1	120	10	40	15
286.1	157.0	2.2	Hydromorphone 2	120	10	58	20
239.1	126.0	4.6	Ketamine 1	64	10	43	12
239.1	208.1	4.6	Ketamine 2	64	10	21	16
224.1	179.3	4.5	Norketamine 1	70	10	20	15
224.1	126.0	4.5	Norketamine 2	70	10	35	20
321.0	275.0	8.3	Lorazepam 1	70	10	31	15
321.0	229.1	8.3	Lorazepam 2	70	10	72	15

Q1 (Da)	Q3 (Da)	Time (min)	Analyte	DP (V)	EP (V)	CE (V)	CXP (V)
197.1	154.1	4.9	mCPP 1	75	10	20	13
197.1	118.0	4.9	mCPP 2	75	10	29	14
180.1	79.0	3.7	MDA 1	50	10	40	12
180.1	135.0	3.7	MDA 2	50	10	26	12
209.1	164.0	4.4	MDEA 1	60	10	14	15
209.1	136.1	4.4	MDEA 2	601	10	30	15
195.1	164.1	4.0	MDMA 1	60	10	20	15
195.1	106.1	4.0	MDMA 2	60	10	35	10
277.2	176.1	5.3	MDPV 1	75	10	31	15
277.2	127.1	5.3	MDPV 2	75	10	35	15
279.2	176.1	4.6	MDPV metabolite 1	50	10	23	20
279.2	127.2	4.6	MDPV metabolite 2	50	10	23	15
249.2	22.1	5.4	Meperidine 1	80	10	31	20
249.2	175.2	5.4	Meperidine 2	80	10	27	13
235.1	161.1	5.5	Normeperidine 1	60	10	22	15
234.1	115.1	5.5	Normeperidine 2	60	10	85	11
311.2	266.0	8.4	Methadone 1	80	10	15	10
311.2	224.2	8.4	Methadone 2	80	10	31	20
151.1	120.1	3.8	Methamphetamine 1	65	10	15	10
150.1	91.0	3.8	Methamphetamine 2	40	10	15	10
326.1	291.0	7.0	Midazolam 1	85	10	30	12
326.1	249.0	7.0	Midazolam 2	85	10	53	12
342.1	168.1	7.3	α-Hydroxymidazolam 1	110	10	55	15
342.1	140.0	7.3	α-Hydroxymidazolam 2	110	10	84	12
267.2	210.1	5.2	Mirtazapine 1	90	10	33	10
266.2	115.0	5.2	Mirtazapine 2	90	10	50	15
253.2	196.2	5.0	N-Desmethylmirtazapine 1	70	10	31	15
252.2	209.2	5.0	N-Desmethylmirtazapine 2	70	10	22	20
286.1	152.1	1.7	Morphine 1	100	10	79	14
286.1	165.0	1.7	Morphine 2	100	10	60	15
328.2	165.0	4.0	6-Acetylmorphine 1	80	10	55	10
328.2	211.0	4.0	6-Acetylmorphine 2	80	10	37	10
328.2	212.2	3.4	Naloxone 1	100	10	55	15
328.2	253.2	3.4	Naloxone 2	100	10	38	15

Q1 (Da)	Q3 (Da)	Time (min)	Analyte	DP (V)	EP (V)	CE (V)	CXP (V)
342.2	270.1	3.9	Naltrexone 1	100	10	38	15
342.2	212.0	3.9	Naltrexone 2	100	10	60	20
282.2	236.0	8.0	Nitrazepam 1	100	10	36	15
282.2	207.0	8.0	Nitrazepam 2	100	10	50	11
252.1	146.0	3.6	7-Aminonitrazepam 1	100	10	40	14
253.1	122.1	3.6	7-Aminonitrazepam 2	100	10	37	15
265.2	234.2	8.1	Nortriptyline 1	80	10	20	10
265.2	192.0	8.1	Nortriptyline 2	80	10	35	10
314.1	257.0	4.0	Olanzapine 1	40	10	26	25
313.1	282.0	4.0	Olanzapine 2	40	10	38	10
299.1	213.1	3.9	N-Desmethylolanzapine 1	140	10	30	11
299.1	198.1	3.9	N-Desmethylolanzapine 2	140	10	42	12
287.1	241.1	8.1	Oxazepam 1	60	10	30	20
287.1	104.1	8.1	Oxazepam 2	60	10	45	12
316.2	256.1	3.7	Oxycodone 1	95	10	35	15
316.2	241.0	3.7	Oxycodone 2	95	10	42	15
302.1	284.0	1.9	Oxymorphone 1	90	10	22	16
302.1	227.0	1.9	Oxymorphone 2	90	10	41	12
330.2	70.0	7.8	Paroxetine 1	100	10	45	10
330.2	109.0	7.8	Paroxetine 2	100	10	75	10
244.2	159.1	6.3	PCP 1	50	10	20	20
244.2	91.0	6.3	PCP 2	50	10	35	10
152.1	117.1	2.3	Phenylpropanolamine 1	40	10	23	12
152.1	91.0	2.3	Phenylpropanolamine 2	40	10	47	16
288.2	42.0	7.7	Procyclidine 1	90	10	75	10
289.2	85.2	7.7	Procyclidine 2	90	10	30	10
166.1	133.0	3.1	Pseudoephedrine 1	45	10	29	12
166.1	117.0	3.1	Pseudoephedrine 2	45	10	28	11
385.2	248.1	7.1	Quetiapine 1	70	10	55	15
385.2	280.1	7.1	Quetiapine 2	70	10	40	20
297.1	140.1	6.9	Norquetiapine 1	70	10	80	15
297.1	184.0	6.9	Norquetiapine 2	70	10	53	15
401.1	270.0	4.7	7-Hydroxyquetiapine 1	30	10	35	18
401.1	209.1	4.7	7-Hydroxyquetiapine 2	30	10	60	20

Q1 (Da)	Q3 (Da)	Time (min)	Analyte	DP (V)	EP (V)	CE (V)	CXP (V)
411.2	110.0	6.1	Risperidone 1	90	10	56	15
411.2	163.0	6.1	Risperidone 2	90	10	68	15
427.2	179.2	5.7	9-Hydroxyrisperidone 1	110	10	60	15
427.2	165.2	5.7	9-Hydroxyrisperidone 2	110	10	60	15
230.2	72.0	6.0	Rolicyclidine 1	50	10	18	10
230.2	159.0	6.0	Rolicyclidine 2	50	10	18	12
306.1	123.1	8.7	Sertraline 1	50	10	75	15
306.1	275.1	8.7	Sertraline 2	50	10	12	15
301.1	177.0	9.0	Temazepam 1	65	10	43	17
302.1	256.1	9.0	Temazepam 2	65	10	23	15
264.2	42.0	5.0	Tramadol 1	80	10	75	10
265.2	59.0	5.0	Tramadol 2	80	10	45	10
250.2	58.0	4.0	O-Desmethyl-cis-tramadol 1	150	10	90	10
250.2	42.0	4.0	O-Desmethyl-cis-tramadol 2	70	10	105	14
373.2	177.1	6.2	Trazodone 1	50	10	25	14
373.2	149.1	6.2	Trazodone 2	90	10	37	15
343.1	204.2	8.7	Triazolam 1	100	10	70	18
343.1	177.2	8.7	Triazolam 2	100	10	90	15
359.0	331.0	8.0	α-Hydroxytriazolam 1	100	10	40	10
359.0	176.0	8.0	α-Hydroxytriazolam 2	100	10	38	16
278.2	147.1	6.0	Venlafaxine 1	60	10	25	13
279.2	216.2	6.0	Venlafaxine 2	60	10	24	16
264.2	201.2	4.6	O-Desmethylvenlafaxine 1	60	10	25	20
264.2	107.0	4.6	O-Desmethylvenlafaxine 2	60	10	27	15
389.1	245.0	4.9	Zopiclone 1	60	10	26	10
389.1	217.0	4.9	Zopiclone 2	60	10	47	16
375.1	245.1	4.9	N-Desmethylzopiclone 1	70	10	25	10
375.1	217.0	4.9	N-Desmethylzopiclone 2	70	10	50	10
281.2	117.0	8.3	Amitriptyline-D_3	76	10	32	11
144.1	127.0	3.4	Amphetamine-D_8	40	10	13	10
320.0	186.2	6.9	Bromazepam-D_4	100	10	45	18
305.1	232.2	6.4	Chlordiazepoxide-D_5	65	10	35	20
331.2	234.0	7.2	Citalopram-D_6	85	10	40	15
209.1	254.0	4.8	7-Aminoclonazepam-D_4	80	10	30	18

Q1 (Da)	Q3 (Da)	Time (min)	Analyte	DP (V)	EP (V)	CE (V)	CXP (V)
308.2	185.0	5.3	Cocaine-D_3	30	10	21	15
303.3	215.0	3.5	Codeine-D_3	110	10	36	18
292.1	246.2	7.3	Demoxepam-D_5	100	10	50	15
290.1	89.0	9.8	Diazepam-D_5	45	10	70	15
259.2	115.0	7.0	Diphenhydramine-D_3	55	10	80	10
169.1	136.0	2.9	Ephedrine-D_3	40	10	29	12
342.2	188.2	6.8	Fentanyl-D_5	110	10	34	17
321.1	275.1	8.7	Flunitrazepam-D_7	100	10	36	12
293.1	140.0	8.8	Desalkylflurazepam-D_4	110	10	40	15
244.1	131.0	4.6	Ketamine-D_4	64	10	43	12
327.1	281.0	8.3	Lorazepam-D_4	70	10	33	15
199.1	165.1	4.0	MDMA-D_5	60	10	20	15
285.2	175.1	5.3	MDPV-D_8	75	10	31	15
155.1	121.1	3.8	Methamphetamine-D_5	65	10	15	10
292.1	152.1	1.7	Morphine-D_6	100	10	79	14
287.1	185.1	8.0	Nitrazepam-D_5	100	10	50	10
321.1	261.1	4.0	Olanzapine-D_8	40	10	26	25
292.1	246.1	8.1	Oxazepam-D_5	60	10	30	14
319.2	259.0	3.8	Oxycodone-D_3	95	10	35	15
305.2	287.0	1.9	Oxymorphone-D_3	90	10	22	16
249.2	164.2	6.3	PCP-D_5	50	10	20	20
392.2	254.0	7.0	Quetiapine-D_8	70	10	55	15
268.2	42.0	5.0	Tramadol-D_3	80	10	75	10
379.2	182.1	6.2	Trazodone-D_6	90	10	35	14
347.1	208.0	8.7	Triazolam-D_4	100	10	70	18
284.2	147.0	6.0	Venlafaxine-D_6	60	10	34	13
393.1	245.0	4.9	Zopiclone-D_4	60	10	26	10

Q1: parent mass; Q3: fragment mass; DP: declustering potential; EP: entrance potential; CE: collision energy; CXP: collision cell exit potential. Transition 1 is the "identification transition", transition 2 is the "confirmation transition".

F.1.3 Cannabinoid method

Liquid chromatography

Mobile phase A: methanol : 10 mM ammonium formate pH 3.0 (2:98 v:v)
Mobile phase B: acetonitrile (LC-MS Grade)
Rinsing solution: methanol : 1% formic acid in water : isopropanol (50:25:25 v:v:v)
Analytical column: Zorbax Eclipse Plus C18, 2.1 × 50 mm, 3.5 μm (Agilent)
Flow rate: 550 $\mu L/min$

Table F.1.5: LC pump gradient.

Time (min)	A (%)	B (%)
0.00	65	35
0.50	65	35
1.00	50	50
2.50	45	55
3.50	20	80
4.50	20	80
4.51	0	100
5.50	0	100
5.51	65	35
6.50	65	35

Autosampler and thermostat settings

- Injection volume: 10 μL
- Wash time: 10 sec
- Autosampler temperature: 4 °C
- Column oven temperature: 50 °C

Mass spectrometry method

- Scan type: scheduled multiple reaction monitoring (MRM)

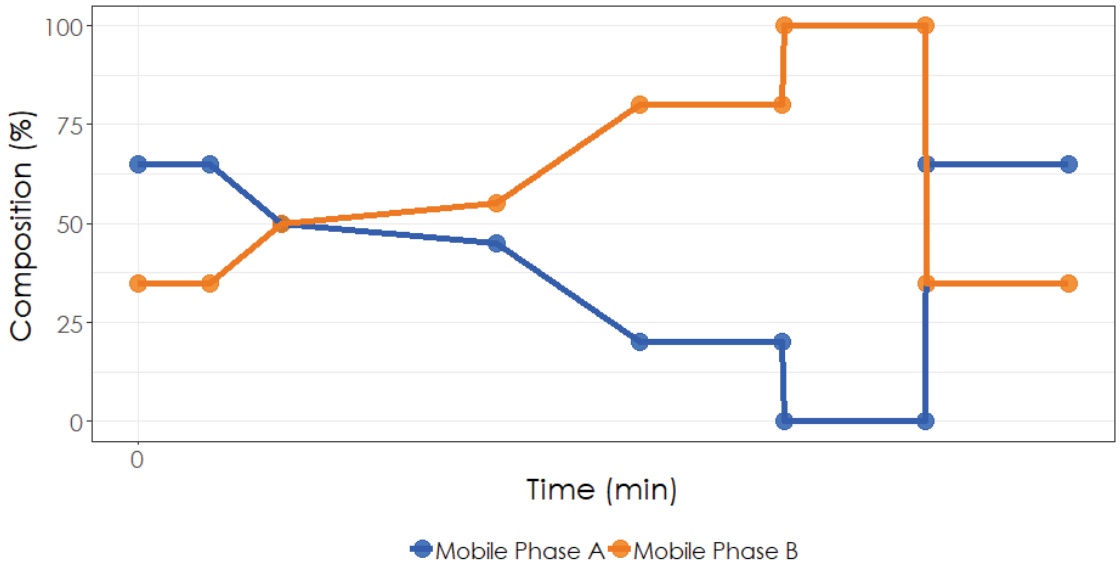


Figure F.1.2: LC pump gradient graph.

Table F.1.6: Divert valve program.

Time (min)	Position
0	MS
0.1	Waste
4.2	MS
5.3	Waste
6.5	MS

- Polarity: positive
- MRM detection window: 60 *sec*
- Target scan time: 0.6 *sec*
- Break: 3.00 *msec*
- Acquisition time: 6 minutes 30 seconds
- Ion source: Turbo Spray
- Curtain gas (CUR): 30.0
- Collision gas (CAD): 10.0
- Ion spray voltage (IS): 5500.0 V

- Source temperature (TEM): 700.0 °C
- Ion source gas 1 (GS1): 60.0
- Ion source gas (GS2): 65.0

Table F.1.7: Monitored MRM transitions.

Q1 (Da)	Q3 (Da)	Time (min)	Analyte	DP (V)	EP (V)	CE (V)	CXP (V)
315.2	193.1	4.8	THC 1	100	10	30	15
315.2	123.1	4.8	THC 2	100	10	47	10
318.3	196.1	4.8	THC-D_3	100	10	30	15

The method is validated on the following systems: HPLC Agilent 1200 series and 1260 Infinity, LC-MS/MS 5500 QTRAP Sciex.

The data acquisition software used is Analyst® 1.6.2 build 8489 and the data analysis software used is Multiquant® 3.0.1 (Version 3.0.6256.0)

F.2 Supplementary data 2 – Complete validation data

The full spreadsheet is available for download at https://spectrum.library.concordia.ca/985534/8/Supplementary%20Data%202_VF.xlsx.

G

Supplementary Data to "Postmortem Estimation of Metabolic Capacity Through Drug Metabolizing Enzyme Proteomics – A Proof of Concept"

G.1 Supplementary data 1 – CYP 2D6 and 3A4 LC-MS/MS analysis method

G.1.1 Liquid chromatography conditions

LC pump gradient

Flow rate: 500 $\mu L/min$

Table G.1.1: LC pump gradient.

Time (min)	A (%)	B (%)
0.0	98	2
25.0	60	40
25.1	1	99
35.0	1	99
35.1	98	2
45.0	98	2

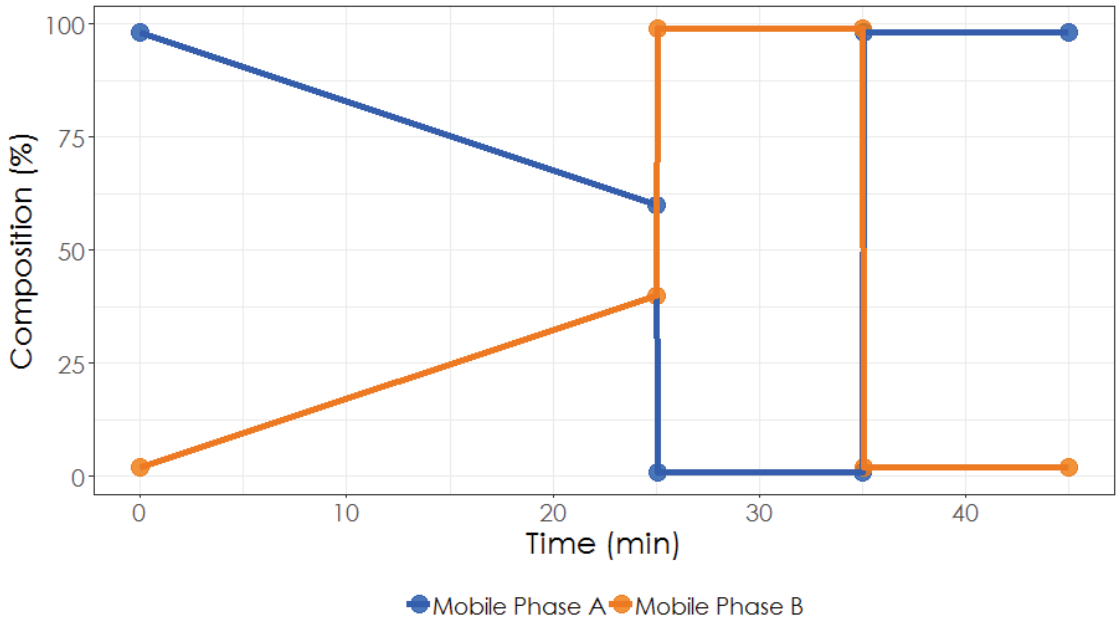


Figure G.1.1: LC pump gradient graph.

Autosampler and thermostat settings

- Injection volume: 20 μL
- Wash time: 10 *sec*
- Autosampler temperature: 4 °C
- Column oven temperature: 60 °C

Divert valve settings

Table G.1.2: Divert valve program.

Time (min)	Position
0	MS
0.1	Waste
3	Waste
3.1	MS
25	MS
25.1	Waste
35	Waste
35.1	MS
45	MS

G.1.2 Mass spectrometry conditions

- Scan type: scheduled multiple reaction monitoring (MRM)
- Polarity: positive
- MRM detection window: 240 *sec*
- Ion source: Turbo Spray
- Curtain gas (CUR): 30.0
- Collision gas (CAD): 5
- Ion spray voltage (IS): 2500.0 *V*
- Source temperature: 500.0 °C
- Ion source gas 1 (GS1): 60.0
- Ion source gas (GS2): 65.0

Table G.1.3: Monitored MS transitions.

Q1 (m/z)	Q3 (m/z)	Time (min)	Transition ID	DP (V)	EP (V)	CE (V)	CXP (V)
439.248	662.398	9.80	2D6_Q1.+2y6	63.1	10	24.7	18
439.248	506.309	9.80	2D6_Q1.+2y4	63.1	10	24.7	18
439.248	605.377	9.80	2D6_Q1.+2y5	63.1	10	24.7	18
439.248	359.240	9.80	2D6_Q1.+2y3	63.1	10	24.7	18
439.248	246.156	9.80	2D6_Q1.+2y2	63.1	10	24.7	18
405.229	575.351	10.49	2D6_Q2.+2y5	60.7	10	23.4	18
405.229	476.283	10.49	2D6_Q2.+2y4	60.7	10	23.4	18
405.229	662.383	10.49	2D6_Q2.+2y6	60.7	10	23.4	18
405.229	389.251	10.49	2D6_Q2.+2y3	60.7	10	23.4	18
405.229	175.119	10.49	2D6_Q2.+2y1	60.7	10	23.4	18
562.633	542.268	19.26	2D6_Q3.+3y4	72.1	10	28.2	18
562.633	655.352	19.26	2D6_Q3.+3y5	72.1	10	28.2	18
562.633	1012.506	19.26	2D6_Q3.+3y8	72.1	10	28.2	18
562.633	312.178	19.26	2D6_Q3.+3y2	72.1	10	28.2	18
562.633	897.479	19.26	2D6_Q3.+3y7	72.1	10	28.2	18

Q1 (m/z)	Q3 (m/z)	Time (min)	Transition ID	DP (V)	EP (V)	CE (V)	CXP (V)
500.267	459.267	12.96	2D6_Q4.+3y4	67.6	10	24.8	18
500.267	572.352	12.96	2D6_Q4.+3y5	67.6	10	24.8	18
749.897	901.474	12.96	2D6_Q4.+2y8	85.8	10	35.9	18
749.897	459.267	12.96	2D6_Q4.+2y4	85.8	10	35.9	18
749.897	1014.558	12.96	2D6_Q4.+2y9	85.8	10	35.9	18
482.248	735.378	10.91	2D6_Q5.+2y6	66.3	10	26.2	18
482.248	606.336	10.91	2D6_Q5.+2y5	66.3	10	26.2	18
482.248	507.267	10.91	2D6_Q5.+2y4	66.3	10	26.2	18
482.248	379.209	10.91	2D6_Q5.+2y3	66.3	10	26.2	18
482.248	175.119	10.91	2D6_Q5.+2y1	66.3	10	26.2	18
1146.915	1098.504	23.12	2D6*10_WT.+3y8	114.7	10	60.0	18
860.438	1098.504	23.12	2D6*10_WT.+4y8	93.8	10	44.4	18
860.438	838.388	23.12	2D6*10_WT.+4y6	93.8	10	44.4	18
860.438	678.357	23.12	2D6*10_WT.+4y5	93.8	10	44.4	18
860.438	1001.451	23.12	2D6*10_WT.+4y7	93.8	10	44.4	18
1143.575	1098.504	22.89	2D6*10_MUT.+3y8	114.5	10	59.8	18
1143.575	1199.551	22.89	2D6*10_MUT.+3y9	114.5	10	59.8	18
1143.575	416.262	22.89	2D6*10_MUT.+3y3	114.5	10	59.8	18
1143.575	838.388	22.89	2D6*10_MUT.+3y6	114.5	10	59.8	18
1143.575	970.441	22.89	2D6*10_MUT.+3y15+2	114.5	10	59.8	18
962.842	1198.694	18.19	2D6*17_WT.+3y11	101.3	10	50.0	18
962.842	1118.590	18.19	2D6*17_WT.+3y21+2	101.3	10	50.0	18
962.842	1237.643	18.19	2D6*17_WT.+3y23+2	101.3	10	50.0	18
962.842	1187.119	18.19	2D6*17_WT.+3y22+2	101.3	10	50.0	18
962.842	968.044	18.19	2D6*17_WT.+3y18+2	101.3	10	50.0	18
966.854	1124.608	17.98	2D6*17_MUT.+3y21+2	101.6	10	50.2	18
966.854	1243.661	17.98	2D6*17_MUT.+3y23+2	101.6	10	50.2	18
966.854	1193.137	17.98	2D6*17_MUT.+3y22+2	101.6	10	50.2	18
966.854	974.062	17.98	2D6*17_MUT.+3y18+2	101.6	10	50.2	18
966.854	1210.731	17.98	2D6*17_MUT.+3y11	101.6	10	50.2	18
698.837	938.465	18.41	2D6*9_WT.+2y8	82.1	10	34.0	18
698.837	867.428	18.41	2D6*9_WT.+2y7	82.1	10	34.0	18
698.837	607.276	18.41	2D6*9_WT.+2y5	82.1	10	34.0	18
698.837	720.360	18.41	2D6*9_WT.+2y6	82.1	10	34.0	18

Q1 (m/z)	Q3 (m/z)	Time (min)	Transition ID	DP (V)	EP (V)	CE (V)	CXP (V)
698.837	1067.508	18.41	2D6*9_WT.+2y9	82.1	10	34.0	18
734.355	938.465	18.75	2D6*9_MUT.+2y8	84.7	10	35.3	18
734.355	678.313	18.75	2D6*9_MUT.+2y6	84.7	10	35.3	18
734.355	1009.502	18.75	2D6*9_MUT.+2y9	84.7	10	35.3	18
734.355	791.397	18.75	2D6*9_MUT.+2y7	84.7	10	35.3	18
489.906	678.313	18.75	2D6*9_MUT.+3y6	66.8	10	24.2	18
712.073	931.500	22.16	3A4_Q1.+3y8	83.0	10	36.3	18
712.073	647.351	22.16	3A4_Q1.+3y5	83.0	10	36.3	18
712.073	284.172	22.16	3A4_Q1.+3y2	83.0	10	36.3	18
712.073	534.267	22.16	3A4_Q1.+3y4	83.0	10	36.3	18
1067.607	926.012	22.16	3A4_Q1.+2y17+2	108.9	10	47.3	18
798.400	603.317	20.14	3A4_Q2.+3y5	89.3	10	41.0	18
798.400	1003.513	20.14	3A4_Q2.+3y9	89.3	10	41.0	18
798.400	819.392	20.14	3A4_Q2.+3y7	89.3	10	41.0	18
798.400	932.476	20.14	3A4_Q2.+3y8	89.3	10	41.0	18
798.400	1116.597	20.14	3A4_Q2.+3y10	89.3	10	41.0	18
439.740	650.362	10.92	3A4_Q3.+2y5	63.2	10	24.7	18
439.740	435.271	10.92	3A4_Q3.+2y3	63.2	10	24.7	18
439.740	549.314	10.92	3A4_Q3.+2y4	63.2	10	24.7	18
439.740	288.203	10.92	3A4_Q3.+2y2	63.2	10	24.7	18
439.740	175.119	10.92	3A4_Q3.+2y1	63.2	10	24.7	18
424.247	587.333	14.37	3A4_Q4.+2y5	62.0	10	24.1	18
424.247	294.170	14.37	3A4_Q4.+2y5+2	62.0	10	24.1	18
424.247	377.197	14.37	3A4_Q4.+2y3	62.0	10	24.1	18
424.247	490.281	14.37	3A4_Q4.+2y4	62.0	10	24.1	18
424.247	175.119	14.37	3A4_Q4.+2y1	62.0	10	24.1	18
481.271	843.447	14.99	3A4_Q5.+3y7	66.2	10	23.8	18
721.403	843.447	14.99	3A4_Q5.+2y7	83.7	10	34.8	18
481.271	746.394	14.99	3A4_Q5.+3y6	66.2	10	23.8	18
481.271	659.362	14.99	3A4_Q5.+3y5	66.2	10	23.8	18
721.403	1087.572	14.99	3A4_Q5.+2y9	83.7	10	34.8	18
569.311	737.383	12.75	3A4*8_WT.+2y7	72.6	10	29.4	18
569.311	824.415	12.75	3A4*8_WT.+2y8	72.6	10	29.4	18
569.311	937.499	12.75	3A4*8_WT.+2y9	72.6	10	29.4	18

Q1 (m/z)	Q3 (m/z)	Time (min)	Transition ID	DP (V)	EP (V)	CE (V)	CXP (V)
569.311	196.611	12.75	3A4*8_WT.+2y4+2	72.6	10	29.4	18
569.311	640.330	12.75	3A4*8_WT.+2y6	72.6	10	29.4	18
689.882	1137.615	14.45	3A4*8_MUT.+2y11	81.4	10	33.7	18
689.882	824.415	14.45	3A4*8_MUT.+2y8	81.4	10	33.7	18
689.882	937.499	14.45	3A4*8_MUT.+2y9	81.4	10	33.7	18
689.882	737.383	14.45	3A4*8_MUT.+2y7	81.4	10	33.7	18
689.882	1050.583	14.45	3A4*8_MUT.+2y10	81.4	10	33.7	18
900.108	731.405	22.25	3A4*11_WT.+3y6	96.7	10	46.6	18
900.108	632.336	22.25	3A4*11_WT.+3y5	96.7	10	46.6	18
900.108	1076.541	22.25	3A4*11_WT.+3y9	96.7	10	46.6	18
900.108	830.473	22.25	3A4*11_WT.+3y7	96.7	10	46.6	18
900.108	961.514	22.25	3A4*11_WT.+3y8	96.7	10	46.6	18
910.106	761.397	22.30	3A4*11_MUT.+3y6	97.5	10	47.1	18
910.106	662.329	22.30	3A4*11_MUT.+3y5	97.5	10	47.1	18
910.106	1106.533	22.30	3A4*11_MUT.+3y9	97.5	10	47.1	18
910.106	860.466	22.30	3A4*11_MUT.+3y7	97.5	10	47.1	18
910.106	991.506	22.30	3A4*11_MUT.+3y8	97.5	10	47.1	18
331.187	401.214	7.27	3A4*13_WT.+2y3	55.3	10	20.8	18
331.187	175.119	7.27	3A4*13_WT.+2y1	55.3	10	20.8	18
331.187	201.111	7.27	3A4*13_WT.+2y3+2	55.3	10	20.8	18
331.187	304.162	7.27	3A4*13_WT.+2y2	55.3	10	20.8	18
331.187	514.298	7.27	3A4*13_WT.+2y4	55.3	10	20.8	18
339.203	417.246	10.92	3A4*13_MUT.+2y3	55.8	10	21.1	18
339.203	530.330	10.92	3A4*13_MUT.+2y4	55.8	10	21.1	18
339.203	175.119	10.92	3A4*13_MUT.+2y1	55.8	10	21.1	18
339.203	304.162	10.92	3A4*13_MUT.+2y2	55.8	10	21.1	18
339.203	265.668	10.92	3A4*13_MUT.+2y4+2	55.8	10	21.1	18
444.252	672.407	9.81	IS_2D6_Q1.+2y6	63.1	10	24.7	18
444.252	516.317	9.81	IS_2D6_Q1.+2y4	63.1	10	24.7	18
444.252	615.385	9.81	IS_2D6_Q1.+2y5	63.1	10	24.7	18
444.252	369.248	9.81	IS_2D6_Q1.+2y3	63.1	10	24.7	18
444.252	256.164	9.81	IS_2D6_Q1.+2y2	63.1	10	24.7	18
410.234	585.359	10.49	IS_2D6_Q2.+2y5	60.7	10	23.4	18
410.234	486.291	10.49	IS_2D6_Q2.+2y4	60.7	10	23.4	18

Q1 (m/z)	Q3 (m/z)	Time (min)	Transition ID	DP (V)	EP (V)	CE (V)	CXP (V)
410.234	672.391	10.49	IS_2D6_Q2.+2y6	60.7	10	23.4	18
410.234	399.259	10.49	IS_2D6_Q2.+2y3	60.7	10	23.4	18
410.234	185.127	10.49	IS_2D6_Q2.+2y1	60.7	10	23.4	18
565.969	552.276	19.24	IS_2D6_Q3.+3y4	72.1	10	28.2	18
565.969	665.360	19.24	IS_2D6_Q3.+3y5	72.1	10	28.2	18
565.969	1022.514	19.24	IS_2D6_Q3.+3y8	72.1	10	28.2	18
565.969	322.186	19.24	IS_2D6_Q3.+3y2	72.1	10	28.2	18
565.969	907.487	19.24	IS_2D6_Q3.+3y7	72.1	10	28.2	18
503.603	469.276	12.98	IS_2D6_Q4.+3y4	67.6	10	24.8	18
503.603	582.360	12.98	IS_2D6_Q4.+3y5	67.6	10	24.8	18
754.901	911.482	12.98	IS_2D6_Q4.+2y8	85.8	10	35.9	18
754.901	469.276	12.98	IS_2D6_Q4.+2y4	85.8	10	35.9	18
754.901	1024.566	12.98	IS_2D6_Q4.+2y9	85.8	10	35.9	18
487.252	745.387	10.94	IS_2D6_Q5.+2y6	66.3	10	26.2	18
487.252	616.344	10.94	IS_2D6_Q5.+2y5	66.3	10	26.2	18
487.252	517.276	10.94	IS_2D6_Q5.+2y4	66.3	10	26.2	18
487.252	389.217	10.94	IS_2D6_Q5.+2y3	66.3	10	26.2	18
487.252	185.127	10.94	IS_2D6_Q5.+2y1	66.3	10	26.2	18
1150.251	1108.512	23.12	IS_2D6*10_WT.+3y8	114.7	10	60.0	18
862.940	1108.512	23.12	IS_2D6*10_WT.+4y8	93.8	10	44.4	18
862.940	848.396	23.12	IS_2D6*10_WT.+4y6	93.8	10	44.4	18
862.940	688.365	23.12	IS_2D6*10_WT.+4y5	93.8	10	44.4	18
862.940	1011.459	23.12	IS_2D6*10_WT.+4y7	93.8	10	44.4	18
966.178	1208.703	23.12	IS_2D6*17_WT.+3y11	101.3	10	50.0	18
966.178	1123.594	18.20	IS_2D6*17_WT.+3y21+2	101.3	10	50.0	18
966.178	1242.647	18.20	IS_2D6*17_WT.+3y23+2	101.3	10	50.0	18
966.178	1192.123	18.20	IS_2D6*17_WT.+3y22+2	101.3	10	50.0	18
966.178	973.048	18.20	IS_2D6*17_WT.+3y18+2	101.3	10	50.0	18
702.844	946.479	18.43	IS_2D6*9_WT.+2y8	82.1	10	34.0	18
702.844	875.442	18.43	IS_2D6*9_WT.+2y7	82.1	10	34.0	18
702.844	615.290	18.43	IS_2D6*9_WT.+2y5	82.1	10	34.0	18
702.844	728.374	18.43	IS_2D6*9_WT.+2y6	82.1	10	34.0	18
702.844	1075.522	18.43	IS_2D6*9_WT.+2y9	82.1	10	34.0	18
714.745	939.514	22.16	IS_3A4_Q1.+3y8	83.0	10	36.3	18

Q1 (m/z)	Q3 (m/z)	Time (min)	Transition ID	DP (V)	EP (V)	CE (V)	CXP (V)
714.745	655.365	22.16	IS_3A4_Q1.+3y5	83.0	10	36.3	18
714.745	292.186	22.16	IS_3A4_Q1.+3y2	83.0	10	36.3	18
714.745	542.281	22.16	IS_3A4_Q1.+3y4	83.0	10	36.3	18
1071.614	930.019	22.16	IS_3A4_Q1.+2y17+2	108.9	10	47.3	18
801.071	611.331	20.13	IS_3A4_Q2.+3y5	89.3	10	41.0	18
801.071	1011.527	20.13	IS_3A4_Q2.+3y9	89.3	10	41.0	18
801.071	827.406	20.13	IS_3A4_Q2.+3y7	89.3	10	41.0	18
801.071	940.490	20.13	IS_3A4_Q2.+3y8	89.3	10	41.0	18
801.071	1124.611	20.13	IS_3A4_Q2.+3y10	89.3	10	41.0	18
444.744	660.370	10.94	IS_3A4_Q3.+2y5	63.2	10	24.7	18
444.744	445.280	10.94	IS_3A4_Q3.+2y3	63.2	10	24.7	18
444.744	559.323	10.94	IS_3A4_Q3.+2y4	63.2	10	24.7	18
444.744	298.211	10.94	IS_3A4_Q3.+2y2	63.2	10	24.7	18
444.744	185.127	10.94	IS_3A4_Q3.+2y1	63.2	10	24.7	18
429.251	597.342	14.38	IS_3A4_Q4.+2y5	62.0	10	24.1	18
429.251	299.174	14.38	IS_3A4_Q4.+2y5+2	62.0	10	24.1	18
429.251	387.205	14.38	IS_3A4_Q4.+2y3	62.0	10	24.1	18
429.251	500.289	14.38	IS_3A4_Q4.+2y4	62.0	10	24.1	18
429.251	185.127	14.38	IS_3A4_Q4.+2y1	62.0	10	24.1	18
484.607	853.455	14.99	IS_3A4_Q5.+3y7	66.2	10	23.8	18
726.407	853.455	14.99	IS_3A4_Q5.+2y7	83.7	10	34.8	18
484.607	756.403	14.99	IS_3A4_Q5.+3y6	66.2	10	23.8	18
484.607	669.371	14.99	IS_3A4_Q5.+3y5	66.2	10	23.8	18
726.407	1097.580	14.99	IS_3A4_Q5.+2y9	83.7	10	34.8	18
573.318	745.397	12.76	IS_3A4*8_WT.+2y7	72.6	10	29.4	18
573.318	832.429	12.76	IS_3A4*8_WT.+2y8	72.6	10	29.4	18
573.318	945.513	12.76	IS_3A4*8_WT.+2y9	72.6	10	29.4	18
573.318	200.618	12.76	IS_3A4*8_WT.+2y4+2	72.6	10	29.4	18
573.318	648.344	12.76	IS_3A4*8_WT.+2y6	72.6	10	29.4	18

G.2 Supplementary data 2 – Multifactorial design optimization (DoE) of the denaturation process

G.2.1 Factors selected

Following an initial factors screening process, the following factors/levels were retained for optimization:

- Chaotrope
  - 6 M urea
  - 1.2 M thiourea/4.8 M urea
  - None
- Detergent
  - Brij-C10
  - SDS
  - Triton X-100
  - CHAPS
  - OTG
- Temperature
  - 20 °C to 95 °C

A response surface method (RSM) optimization using a D-optimal design was carried out.

G.2.2 Experimental protocol

90 μL of pooled human liver microsomes (HLM, XenoTech, cat. no. H0630, Lenexa, Kansas, United States) are added in a 5 mL LoBind Eppendorf (cat. no. 0030108302, Eppendorf, Eppendorf, Hamburg, Germany). 40 μL of the selected chaotrope and 40 μL of the selected detergent (Table G.2.1) are added to the sample. Where no chaotrope is to be used, 40 μL of 50 mM NH_4HCO_3 is added instead. Following vortexing, the sample is heated at the indicated temperature for 15 minutes in the water bath.

Table G.2.1: Experimental design for optimization of the chaotrope, detergent and temperature for denaturation.

Sample ID	Run Order	Chaotrope	Detergent	Temperature (°C)
N1	17	6 M Urea	Brij-C10	20
N2	8	1.2 M Thiourea/4.8 M Urea	Brij-C10	20
N3	5	None	Brij-C10	20
N4	16	6 M Urea	SDS	20
N5	24	1.2 M Thiourea/4.8 M Urea	SDS	20
N6	22	None	SDS	20
N7	23	6 M Urea	Triton X-100	20
N8	13	1.2 M Thiourea/4.8 M Urea	Triton X-100	20
N9	14	None	Triton X-100	20
N10	20	6 M Urea	CHAPS	20
N11	15	1.2 M Thiourea/4.8 M Urea	CHAPS	20
N12	31	None	CHAPS	20
N13	2	6 M Urea	OTG	20
N14	30	1.2 M Thiourea/4.8 M Urea	OTG	20
N15	27	None	OTG	20
N16	19	1.2 M Thiourea/4.8 M Urea	Brij-C10	60
N17	4	None	OTG	60
N18	6	6 M Urea	Brij-C10	95
N19	21	1.2 M Thiourea/4.8 M Urea	Brij-C10	95
N20	33	None	Brij-C10	95
N21	25	6 M Urea	SDS	95
N22	9	1.2 M Thiourea/4.8 M Urea	SDS	95
N23	26	None	SDS	95
N24	18	6 M Urea	Triton X-100	95
N25	11	1.2 M Thiourea/4.8 M Urea	Triton X-100	95
N26	7	None	Triton X-100	95
N27	28	6 M Urea	CHAPS	95
N28	1	1.2 M Thiourea/4.8 M Urea	CHAPS	95
N29	3	None	CHAPS	95
N30	34	6 M Urea	OTG	95
N31	32	1.2 M Thiourea/4.8 M Urea	OTG	95
N32	12	None	OTG	60
N33	10	None	OTG	60
N34	29	None	OTG	60

1337 μL of 50 mM NH_4HCO_3 is added to all samples. After vortexing, 15 μL of 1 M DTT is added to all samples which are incubated at room temperature for 30 minutes. This performs the reduction of the disulfide bonds in the proteins.

Alkylation of the reduced cysteines is then performed by adding 20 μL of 2.25 M iodoacetamide to all samples. Following vortexing, samples are incubated in the dark at room temperature for 30 minutes.

The alkylating agent, iodoacetamide, is quenched by adding 8 μL of 1 M DTT to all samples and vortexing.

Digestion is carried out by adding 20 μg of trypsin reconstituted in 50 μL of 50 mM acetic acid and incubating for 5 hours at 37 °C with vortexing every 30 minutes. Trypsin activity is quenched by adding 30 μL of formic acid.

Tryptic digest clean-up and LC-MS/MS analysis were performed as described in the main paper.

Following peak integration, the normalized average peak area (NAPA) was calculated:

$$\text{NAPA} = \frac{\sum_j \text{NPA}_i}{j} \tag{G.1}$$

where:

$$\text{NPA} = \frac{A_{ij}}{\overline{A_j}} \tag{G.2}$$

A is a peak area, i is the sample and j is the transition number.

NPA therefore represents the peak area divided by the average peak area over all samples for one specific transition, and NAPA represents the average of the NPAs for all transitions of a given sample.

Data analysis was carried using the MODDE software (Umetrics/Sartorius Stedim Biotech, Umea, Sweden).

G.2.3 Results

Raw data

Raw measurements and calculated normalized peak areas are shown in Table G.2.2.

Table G.2.2: Measured peak areas.

	2D6_Q1 Peak Area (cps^2)			
Sample ID	**+2y6**	**+2y5**	**+2y4**	**+2y3**
N1	2.02E+06	1.83E+05	1.01E+06	1.72E+05
N2	2.31E+06	2.12E+05	1.14E+06	1.88E+05
N3	2.27E+06	2.12E+05	1.15E+06	1.93E+05
N4	2.15E+06	2.00E+05	1.07E+06	1.75E+05
N5	2.27E+06	2.14E+05	1.12E+06	1.87E+05
N6	2.16E+06	1.95E+05	1.04E+06	1.71E+05
N7	2.66E+06	2.49E+05	1.34E+06	2.54E+05
N8	2.63E+06	2.50E+05	1.32E+06	2.17E+05
N9	1.93E+06	1.84E+05	9.82E+05	1.63E+05
N10	3.06E+06	2.83E+05	1.54E+06	2.61E+05
N11	3.08E+06	2.85E+05	1.55E+06	2.55E+05
N12	2.77E+06	2.53E+05	1.41E+06	2.34E+05
N13	2.52E+06	2.33E+05	1.25E+06	2.08E+05
N14	3.49E+06	3.32E+05	1.79E+06	3.05E+05
N15	2.94E+06	2.77E+05	1.49E+06	2.46E+05
N16	3.36E+06	3.08E+05	1.67E+06	3.01E+05
N17	3.08E+06	2.90E+05	1.55E+06	2.78E+05
N18	3.43E+06	3.20E+05	1.71E+06	3.15E+05
N19	3.73E+06	3.49E+05	1.88E+06	3.30E+05
N20	3.00E+06	2.71E+05	1.51E+06	2.71E+05
N21	1.73E+06	1.56E+05	8.61E+05	1.43E+05
N22	2.79E+06	2.57E+05	1.37E+06	2.25E+05
N23	2.62E+06	2.38E+05	1.30E+06	2.12E+05
N24	3.06E+06	2.83E+05	1.53E+06	2.84E+05
N25	3.27E+06	3.02E+05	1.63E+06	2.98E+05
N26	3.24E+06	3.00E+05	1.62E+06	2.91E+05
N27	4.21E+06	3.84E+05	2.11E+06	3.59E+05
N28	3.88E+06	3.61E+05	1.90E+06	3.37E+05
N29	3.13E+06	2.94E+05	1.59E+06	2.71E+05
N30	3.89E+06	3.62E+05	1.99E+06	3.80E+05
N31	4.63E+06	4.34E+05	2.35E+06	4.13E+05
N32	3.11E+06	2.92E+05	1.56E+06	2.74E+05

N33	3.05E+06	2.82E+05	1.52E+06	2.72E+05	
N34	3.50E+06	3.35E+05	1.80E+06	3.16E+05	
			2D6_Q2 Peak Area (cps²)		
Sample ID	**+2y6**	**+2y5**	**+2y4**	**+2y3**	**+2y2**
N1	7.90E+05	1.65E+06	1.25E+06	1.86E+05	7.82E+04
N2	8.33E+05	1.75E+06	1.30E+06	2.04E+05	7.52E+04
N3	8.58E+05	1.80E+06	1.35E+06	2.01E+05	8.27E+04
N4	4.89E+05	1.07E+06	7.95E+05	1.16E+05	4.94E+04
N5	5.36E+05	1.17E+06	8.44E+05	1.23E+05	4.90E+04
N6	4.95E+05	1.13E+06	8.02E+05	1.19E+05	4.82E+04
N7	9.42E+05	2.00E+06	1.50E+06	2.27E+05	9.29E+04
N8	8.81E+05	1.91E+06	1.42E+06	2.26E+05	8.75E+04
N9	8.06E+05	1.74E+06	1.30E+06	2.06E+05	7.59E+04
N10	1.17E+06	2.50E+06	1.85E+06	2.76E+05	1.14E+05
N11	1.12E+06	2.37E+06	1.79E+06	2.72E+05	1.05E+05
N12	1.09E+06	2.34E+06	1.71E+06	2.69E+05	1.04E+05
N13	1.03E+06	2.19E+06	1.63E+06	2.47E+05	1.00E+05
N14	1.35E+06	2.78E+06	2.11E+06	3.20E+05	1.29E+05
N15	1.18E+06	2.48E+06	1.85E+06	2.89E+05	1.09E+05
N16	1.03E+06	2.19E+06	1.61E+06	2.41E+05	9.70E+04
N17	1.13E+06	2.38E+06	1.77E+06	2.77E+05	1.06E+05
N18	1.24E+06	2.58E+06	1.94E+06	2.86E+05	1.13E+05
N19	1.29E+06	2.70E+06	2.01E+06	3.12E+05	1.21E+05
N20	1.02E+06	2.18E+06	1.62E+06	2.50E+05	1.01E+05
N21	2.80E+05	5.95E+05	4.31E+05	6.51E+04	2.95E+04
N22	7.07E+05	1.43E+06	1.08E+06	1.56E+05	6.71E+04
N23	6.20E+05	1.31E+06	9.61E+05	1.45E+05	6.24E+04
N24	1.05E+06	2.20E+06	1.65E+06	2.49E+05	1.01E+05
N25	1.13E+06	2.40E+06	1.79E+06	2.75E+05	1.14E+05
N26	9.96E+05	2.09E+06	1.61E+06	2.42E+05	1.00E+05
N27	1.47E+06	3.10E+06	2.29E+06	3.57E+05	1.42E+05
N28	1.39E+06	2.95E+06	2.20E+06	3.36E+05	1.31E+05
N29	1.06E+06	2.23E+06	1.67E+06	2.58E+05	1.05E+05
N30	1.38E+06	2.89E+06	2.18E+06	3.27E+05	1.37E+05
N31	1.62E+06	3.29E+06	2.51E+06	3.86E+05	1.53E+05

N32	1.13E+06	2.42E+06	1.80E+06	2.54E+05	1.16E+05
N33	1.10E+06	2.32E+06	1.76E+06	2.75E+05	1.14E+05
N34	1.27E+06	2.63E+06	2.02E+06	3.06E+05	1.25E+05

2D6_Q3 Peak Area (cps²)					
Sample ID	**+2y4**	**+2y2**	**+3y5**	**+3y4**	**+3y2**
N1	3.02E+04	1.66E+04	1.41E+05	2.11E+05	1.34E+05
N2	3.38E+04	1.96E+04	1.52E+05	2.39E+05	1.37E+05
N3	4.67E+04	3.02E+04	2.10E+05	3.16E+05	1.94E+05
N4	0	0	0	0	0
N5	0	0	0	0	0
N6	0	0	0	0	0
N7	6.35E+04	3.74E+04	3.06E+05	4.74E+05	2.76E+05
N8	6.10E+04	4.25E+04	2.76E+05	4.16E+05	2.48E+05
N9	4.92E+04	2.22E+04	1.79E+05	2.74E+05	1.65E+05
N10	6.81E+04	4.68E+04	3.30E+05	4.97E+05	2.87E+05
N11	7.20E+04	4.59E+04	3.07E+05	4.64E+05	2.63E+05
N12	6.73E+04	4.37E+04	3.16E+05	4.31E+05	2.72E+05
N13	4.94E+04	3.31E+04	2.59E+05	3.84E+05	2.45E+05
N14	6.27E+04	4.55E+04	3.40E+05	5.00E+05	2.88E+05
N15	6.88E+04	4.15E+04	3.30E+05	4.99E+05	2.95E+05
N16	4.08E+04	2.99E+04	2.19E+05	3.12E+05	1.90E+05
N17	5.95E+04	4.10E+04	2.80E+05	4.31E+05	2.76E+05
N18	2.52E+04	1.97E+04	1.34E+05	2.05E+05	1.27E+05
N19	3.07E+04	1.93E+04	1.55E+05	2.12E+05	1.36E+05
N20	3.65E+04	2.42E+04	1.71E+05	2.62E+05	1.52E+05
N21	0	0	0	0	0
N22	0	0	0	0	0
N23	0	0	0	0	0
N24	3.45E+04	2.33E+04	1.71E+05	2.38E+05	1.47E+05
N25	4.03E+04	2.15E+04	1.76E+05	2.73E+05	1.66E+05
N26	7.03E+04	5.34E+04	3.08E+05	4.66E+05	2.84E+05
N27	4.41E+04	3.34E+04	1.78E+05	2.88E+05	1.74E+05
N28	4.63E+04	2.90E+04	1.95E+05	2.75E+05	1.74E+05
N29	5.66E+04	3.65E+04	2.45E+05	3.65E+05	2.11E+05
N30	2.93E+04	2.12E+04	1.50E+05	2.25E+05	1.17E+05

N31	4.17E+04	2.79E+04	1.77E+05	2.55E+05	1.52E+05
N32	5.89E+04	4.29E+04	2.94E+05	4.32E+05	2.57E+05
N33	5.91E+04	4.02E+04	2.74E+05	3.77E+05	2.45E+05
N34	6.10E+04	3.81E+04	3.05E+05	4.68E+05	2.83E+05
			2D6_Q4 Peak Area (cps²)		
Sample ID	**+2y8**	**+2y5**	**+2y4**	**+3y5**	**+3y4**
N1	2.84E+05	1.53E+05	2.23E+05	2.94E+05	3.50E+05
N2	3.51E+05	1.78E+05	2.61E+05	3.29E+05	4.00E+05
N3	3.83E+05	1.98E+05	2.90E+05	3.42E+05	4.43E+05
N4	6.20E+05	3.30E+05	4.72E+05	6.15E+05	7.91E+05
N5	6.95E+05	3.56E+05	5.12E+05	6.44E+05	8.35E+05
N6	6.41E+05	3.39E+05	4.56E+05	6.18E+05	8.26E+05
N7	4.01E+05	2.07E+05	3.16E+05	3.66E+05	4.76E+05
N8	4.27E+05	2.29E+05	3.30E+05	3.97E+05	5.06E+05
N9	2.27E+05	1.19E+05	1.84E+05	2.14E+05	2.77E+05
N10	5.00E+05	2.59E+05	3.86E+05	4.42E+05	5.60E+05
N11	5.82E+05	3.01E+05	4.30E+05	4.85E+05	6.36E+05
N12	5.06E+05	2.72E+05	3.88E+05	4.55E+05	5.90E+05
N13	3.80E+05	1.97E+05	2.85E+05	3.27E+05	4.27E+05
N14	5.55E+05	2.97E+05	4.41E+05	4.74E+05	5.90E+05
N15	4.22E+05	2.24E+05	3.30E+05	3.77E+05	4.80E+05
N16	3.82E+05	1.99E+05	3.02E+05	3.48E+05	4.40E+05
N17	4.63E+05	2.46E+05	3.60E+05	3.98E+05	5.11E+05
N18	3.73E+05	1.95E+05	2.97E+05	3.38E+05	4.45E+05
N19	4.14E+05	2.11E+05	3.29E+05	3.68E+05	4.92E+05
N20	3.44E+05	1.87E+05	2.63E+05	3.37E+05	4.21E+05
N21	3.56E+05	1.93E+05	2.61E+05	3.35E+05	4.11E+05
N22	5.53E+05	2.84E+05	4.03E+05	5.06E+05	6.39E+05
N23	6.10E+05	3.13E+05	4.45E+05	6.10E+05	7.94E+05
N24	3.50E+05	1.74E+05	2.82E+05	3.29E+05	4.49E+05
N25	3.80E+05	1.94E+05	3.17E+05	3.51E+05	4.54E+05
N26	3.70E+05	1.94E+05	2.89E+05	3.22E+05	4.27E+05
N27	5.53E+05	2.96E+05	4.29E+05	4.88E+05	6.38E+05
N28	4.61E+05	2.41E+05	3.46E+05	3.82E+05	5.00E+05
N29	4.31E+05	2.35E+05	3.22E+05	3.91E+05	5.19E+05

N30	4.88E+05	2.55E+05	3.75E+05	4.59E+05	5.72E+05
N31	5.73E+05	2.98E+05	4.46E+05	4.93E+05	6.36E+05
N32	4.01E+05	2.10E+05	3.06E+05	3.54E+05	4.55E+05
N33	4.36E+05	2.29E+05	3.32E+05	3.83E+05	5.09E+05
N34	4.94E+05	2.56E+05	3.80E+05	4.23E+05	5.34E+05
		2D6_Q5 Peak Area (cps^2)			
Sample ID	**+2y5**	**+2y6**	**+2y4**	**+2y3**	**+2y1**
N1	5.82E+05	1.20E+06	5.88E+05	3.84E+05	2.00E+05
N2	7.12E+05	1.45E+06	7.14E+05	4.62E+05	2.03E+05
N3	7.00E+05	1.44E+06	7.04E+05	4.65E+05	2.23E+05
N4	7.82E+05	1.69E+06	8.21E+05	5.31E+05	1.76E+05
N5	1.00E+06	2.11E+06	1.02E+06	6.58E+05	2.33E+05
N6	8.10E+05	1.71E+06	8.39E+05	5.46E+05	1.66E+05
N7	8.14E+05	1.64E+06	8.09E+05	5.30E+05	2.60E+05
N8	9.08E+05	1.87E+06	9.29E+05	5.99E+05	3.04E+05
N9	4.55E+05	9.31E+05	4.62E+05	3.03E+05	1.81E+05
N10	1.00E+06	2.01E+06	9.91E+05	6.49E+05	2.98E+05
N11	1.14E+06	2.31E+06	1.13E+06	7.51E+05	3.14E+05
N12	8.93E+05	1.79E+06	9.12E+05	5.77E+05	2.34E+05
N13	8.58E+05	1.75E+06	8.56E+05	5.57E+05	2.94E+05
N14	1.30E+06	2.64E+06	1.28E+06	8.30E+05	4.21E+05
N15	9.52E+05	1.94E+06	9.46E+05	6.21E+05	3.35E+05
N16	8.97E+05	1.84E+06	8.95E+05	6.02E+05	3.07E+05
N17	9.99E+05	2.02E+06	9.93E+05	6.56E+05	3.89E+05
N18	1.03E+06	2.11E+06	1.04E+06	6.87E+05	3.35E+05
N19	1.07E+06	2.17E+06	1.08E+06	6.98E+05	3.52E+05
N20	8.47E+05	1.76E+06	8.64E+05	5.52E+05	2.93E+05
N21	3.49E+05	7.54E+05	3.71E+05	2.36E+05	8.34E+04
N22	8.85E+05	1.85E+06	8.97E+05	5.84E+05	1.91E+05
N23	9.52E+05	1.99E+06	9.72E+05	6.37E+05	2.08E+05
N24	8.04E+05	1.68E+06	8.43E+05	5.39E+05	2.65E+05
N25	8.93E+05	1.88E+06	9.26E+05	6.11E+05	2.97E+05
N26	1.02E+06	2.07E+06	1.04E+06	6.78E+05	3.26E+05
N27	1.30E+06	2.65E+06	1.32E+06	8.53E+05	4.00E+05
N28	1.28E+06	2.67E+06	1.30E+06	8.59E+05	3.82E+05

N29	1.02E+06	2.08E+06	1.03E+06	6.74E+05	3.28E+05
N30	1.18E+06	2.43E+06	1.18E+06	7.80E+05	3.72E+05
N31	1.40E+06	2.91E+06	1.43E+06	9.23E+05	4.52E+05
N32	9.57E+05	1.95E+06	9.71E+05	6.25E+05	3.38E+05
N33	9.23E+05	1.86E+06	9.40E+05	6.13E+05	3.63E+05
N34	1.06E+06	2.15E+06	1.07E+06	6.94E+05	4.15E+05

2D6*410_WT1 Peak Area (cps^2)					
Sample ID	**+3y8**	**+3y6**	**+3y4**	**+3y3**	
N1	6.30E+03	2.70E+02	9.73E+02	1.60E+03	
N2	6.11E+03	5.59E+02	3.22E+02	1.19E+03	
N3	8.82E+03	5.09E+02	6.92E+02	5.49E+03	
N4	0	0	0	0	
N5	0	0	0	0	
N6	0	0	0	0	
N7	1.00E+04	9.03E+02	6.95E+02	3.32E+03	
N8	1.08E+04	2.09E+03	1.04E+03	3.06E+03	
N9	4.04E+03	4.51E+02	7.64E+02	1.95E+03	
N10	9.72E+03	1.39E+03	6.94E+02	3.40E+03	
N11	1.07E+04	3.67E+02	9.71E+02	3.26E+03	
N12	1.14E+04	1.88E+03	1.04E+03	5.00E+03	
N13	7.81E+03	1.02E+03	1.32E+03	2.69E+03	
N14	7.99E+03	1.11E+03	5.27E+02	3.26E+03	
N15	8.33E+03	1.74E+03	9.72E+02	5.14E+03	
N16	7.29E+03	9.73E+02	9.72E+02	2.01E+03	
N17	7.32E+03	2.15E+03	2.15E+03	4.86E+03	
N18	7.23E+03	5.55E+02	8.33E+02	2.92E+03	
N19	9.31E+03	5.55E+02	9.02E+02	4.38E+03	
N20	8.47E+03	1.08E+03	8.23E+02	5.00E+03	
N21	0	0	0	0	
N22	0	0	0	0	
N23	0	0	0	0	
N24	8.47E+03	5.56E+02	6.59E+02	3.26E+03	
N25	9.87E+03	1.39E+03	8.50E+02	3.25E+03	
N26	9.00E+03	1.11E+03	1.19E+03	7.78E+03	
N27	6.11E+03	1.04E+03	9.03E+02	3.33E+03	

N28	6.04E+03	9.72E+02	1.32E+03	2.64E+03	
N29	3.37E+03	4.17E+02	3.47E+02	1.39E+03	
N30	7.88E+03	7.46E+02	6.25E+02	4.58E+03	
N31	8.61E+03	1.04E+03	1.25E+03	7.64E+03	
N32	1.28E+04	1.39E+03	6.94E+02	3.26E+03	
N33	8.27E+03	1.18E+03	6.26E+02	1.56E+03	
N34	9.53E+03	1.60E+03	6.26E+02	5.22E+03	

	2D6*17_WT2 Peak Area (cps^2)				
Sample ID	**+3y23+2**	**+3y22+2**	**+3y21+2**	**+3y18+2**	**+3y15+2**
N1	4.22E+04	2.52E+04	5.04E+04	1.16E+04	9.05E+03
N2	4.32E+04	3.14E+04	5.67E+04	1.23E+04	7.17E+03
N3	5.66E+04	3.97E+04	6.18E+04	1.17E+04	1.03E+04
N4	0	0	0	0	0
N5	0	0	0	0	0
N6	0	0	0	0	0
N7	4.41E+04	3.27E+04	6.14E+04	1.49E+04	1.30E+04
N8	5.65E+04	4.24E+04	6.80E+04	1.52E+04	1.31E+04
N9	3.00E+04	2.20E+04	4.48E+04	1.19E+04	7.06E+03
N10	6.14E+04	4.68E+04	7.56E+04	1.58E+04	1.25E+04
N11	5.23E+04	4.40E+04	8.05E+04	1.93E+04	1.37E+04
N12	5.99E+04	4.31E+04	7.08E+04	1.81E+04	1.23E+04
N13	4.35E+04	3.73E+04	6.72E+04	1.32E+04	1.30E+04
N14	6.44E+04	4.87E+04	8.17E+04	2.01E+04	1.39E+04
N15	5.00E+04	4.66E+04	7.81E+04	1.55E+04	1.30E+04
N16	4.77E+04	3.46E+04	5.62E+04	1.35E+04	7.94E+03
N17	3.61E+04	3.69E+04	6.05E+04	1.45E+04	9.82E+03
N18	3.76E+04	3.20E+04	5.41E+04	1.39E+04	1.12E+04
N19	4.29E+04	3.37E+04	5.38E+04	1.19E+04	1.18E+04
N20	3.74E+04	2.42E+04	5.95E+04	8.39E+03	1.00E+04
N21	0	0	0	0	0
N22	0	0	0	0	0
N23	0	0	0	0	0
N24	4.19E+04	3.21E+04	5.06E+04	1.16E+04	6.95E+03
N25	4.56E+04	3.49E+04	5.99E+04	1.29E+04	1.05E+04
N26	3.93E+04	3.12E+04	5.64E+04	1.21E+04	1.19E+04

N27	4.86E+04	3.71E+04	6.64E+04	1.47E+04	1.13E+04
N28	4.32E+04	3.79E+04	6.40E+04	1.51E+04	1.07E+04
N29	4.01E+04	3.10E+04	5.35E+04	1.21E+04	9.49E+03
N30	4.50E+04	3.97E+04	5.35E+04	1.10E+04	1.16E+04
N31	4.68E+04	4.02E+04	6.87E+04	1.79E+04	1.32E+04
N32	4.82E+04	3.68E+04	5.50E+04	1.50E+04	7.95E+03
N33	4.26E+04	3.49E+04	5.41E+04	1.08E+04	1.08E+04
N34	4.31E+04	3.66E+04	5.65E+04	1.36E+04	1.02E+04

	3A4_Q6 Peak Area (cps^2)				
Sample ID	**+2y8**	**+3y8**	**+3y5**	**+3y14+2**	**+3y13+2**
N1	1.22E+06	3.62E+06	1.26E+06	8.93E+05	1.66E+06
N2	1.29E+06	3.68E+06	1.23E+06	9.24E+05	1.72E+06
N3	1.32E+06	3.30E+06	1.17E+06	8.12E+05	1.52E+06
N4	3.73E+05	7.93E+05	2.67E+05	2.93E+05	4.86E+05
N5	5.84E+05	1.25E+06	4.31E+05	4.47E+05	7.73E+05
N6	4.26E+05	8.90E+05	2.99E+05	3.22E+05	5.56E+05
N7	1.31E+06	3.75E+06	1.33E+06	9.00E+05	1.69E+06
N8	1.29E+06	3.36E+06	1.16E+06	7.99E+05	1.52E+06
N9	1.30E+06	4.10E+06	1.39E+06	1.04E+06	1.91E+06
N10	1.72E+06	4.15E+06	1.46E+06	9.98E+05	1.87E+06
N11	1.74E+06	4.13E+06	1.41E+06	1.00E+06	1.85E+06
N12	1.64E+06	3.76E+06	1.32E+06	9.20E+05	1.78E+06
N13	1.49E+06	3.53E+06	1.20E+06	9.16E+05	1.68E+06
N14	1.54E+06	3.93E+06	1.37E+06	9.71E+05	1.81E+06
N15	1.53E+06	3.99E+06	1.41E+06	1.05E+06	1.87E+06
N16	1.23E+06	3.47E+06	1.20E+06	7.97E+05	1.53E+06
N17	1.39E+06	3.29E+06	1.15E+06	7.68E+05	1.48E+06
N18	5.64E+05	1.31E+06	4.80E+05	3.39E+05	6.18E+05
N19	6.49E+05	1.65E+06	5.47E+05	4.05E+05	7.41E+05
N20	1.48E+06	4.12E+06	1.39E+06	1.01E+06	1.88E+06
N21	0	0	0	0	0
N22	3.53E+05	7.31E+05	2.37E+05	2.68E+05	4.55E+05
N23	4.98E+05	1.20E+06	4.09E+05	4.34E+05	7.71E+05
N24	4.41E+05	1.16E+06	4.20E+05	2.95E+05	5.25E+05
N25	5.91E+05	1.50E+06	5.27E+05	3.60E+05	6.72E+05

N26	1.55E+06	3.82E+06	1.33E+06	9.03E+05	1.68E+06
N27	6.90E+05	1.44E+06	5.09E+05	3.70E+05	6.88E+05
N28	8.28E+05	1.71E+06	6.06E+05	4.68E+05	8.80E+05
N29	1.98E+06	4.56E+06	1.56E+06	1.16E+06	2.07E+06
N30	6.55E+05	1.53E+06	5.56E+05	3.87E+05	7.08E+05
N31	7.54E+05	1.81E+06	6.53E+05	4.61E+05	8.13E+05
N32	1.33E+06	3.08E+06	1.09E+06	7.59E+05	1.41E+06
N33	1.38E+06	3.26E+06	1.13E+06	7.90E+05	1.51E+06
N34	1.47E+06	3.69E+06	1.28E+06	8.86E+05	1.67E+06
		3A4*8_WT Peak Area (cps²)			
Sample ID	**+2y7**	**+2y9**	**+2y8**	**+2y5**	**+2y4**
N1	1.27E+07	1.11E+07	1.34E+07	1.38E+06	1.80E+06
N2	1.43E+07	1.22E+07	1.50E+07	1.51E+06	2.01E+06
N3	1.49E+07	1.27E+07	1.55E+07	1.62E+06	2.09E+06
N4	1.21E+07	1.02E+07	1.27E+07	1.28E+06	1.69E+06
N5	1.10E+07	9.50E+06	1.16E+07	1.17E+06	1.52E+06
N6	1.19E+07	1.02E+07	1.28E+07	1.26E+06	1.64E+06
N7	1.58E+07	1.36E+07	1.64E+07	1.70E+06	2.20E+06
N8	1.69E+07	1.46E+07	1.77E+07	1.83E+06	2.37E+06
N9	1.43E+07	1.22E+07	1.50E+07	1.53E+06	2.02E+06
N10	1.90E+07	1.64E+07	1.98E+07	2.09E+06	2.70E+06
N11	1.93E+07	1.68E+07	2.04E+07	2.11E+06	2.76E+06
N12	1.84E+07	1.56E+07	1.90E+07	1.98E+06	2.58E+06
N13	1.64E+07	1.41E+07	1.72E+07	1.79E+06	2.32E+06
N14	1.96E+07	1.66E+07	2.01E+07	2.15E+06	2.77E+06
N15	1.79E+07	1.52E+07	1.85E+07	1.95E+06	2.54E+06
N16	1.61E+07	1.36E+07	1.69E+07	1.74E+06	2.27E+06
N17	1.72E+07	1.46E+07	1.78E+07	1.85E+06	2.40E+06
N18	1.17E+07	9.89E+06	1.22E+07	1.26E+06	1.63E+06
N19	1.29E+07	1.14E+07	1.36E+07	1.38E+06	1.82E+06
N20	1.45E+07	1.23E+07	1.50E+07	1.55E+06	2.00E+06
N21	8.73E+06	6.95E+06	8.41E+06	8.68E+05	1.07E+06
N22	1.26E+07	1.08E+07	1.33E+07	1.34E+06	1.73E+06
N23	1.36E+07	1.15E+07	1.43E+07	1.44E+06	1.89E+06
N24	1.06E+07	8.95E+06	1.10E+07	1.12E+06	1.49E+06

N25	1.17E+07	9.95E+06	1.22E+07	1.26E+06	1.66E+06
N26	1.46E+07	1.28E+07	1.54E+07	1.61E+06	2.08E+06
N27	1.36E+07	1.17E+07	1.42E+07	1.46E+06	1.92E+06
N28	1.41E+07	1.19E+07	1.45E+07	1.51E+06	1.94E+06
N29	1.63E+07	1.42E+07	1.74E+07	1.76E+06	2.33E+06
N30	1.37E+07	1.15E+07	1.41E+07	1.45E+06	1.89E+06
N31	1.50E+07	1.28E+07	1.58E+07	1.65E+06	2.13E+06
N32	1.66E+07	1.44E+07	1.72E+07	1.81E+06	2.35E+06
N33	1.72E+07	1.50E+07	1.79E+07	1.86E+06	2.42E+06
N34	1.78E+07	1.52E+07	1.87E+07	1.91E+06	2.51E+06

A rapid analysis shows no difference between overall NAPA, NAPA with only qualitative peptides and NAPA with only quantitative peptides. Consequently, the model using the overall NAPA (all peptides) as presented in Table G.2.3 will be used.

Table G.2.3: Normalized averaged peak areas (NAPA) measured.

Sample ID	NAPA	Sample ID	NAPA	Sample ID	NAPA	Sample ID	NAPA
N1	0.8252	N11	1.3050	N21	0.2832	N31	1.2758
N2	0.8740	N12	1.2755	N22	0.5431	N32	1.1554
N3	0.9944	N13	1.0999	N23	0.5912	N33	1.1180
N4	0.5178	N14	1.3714	N24	0.8488	N34	1.2667
N5	0.5772	N15	1.2768	N25	0.9706		
N6	0.5254	N16	1.0559	N26	1.2046		
N7	1.1099	N17	1.2455	N27	1.1474		
N8	1.1713	N18	0.9234	N28	1.1157		
N9	0.8321	N19	1.0056	N29	1.1060		
N10	1.2958	N20	1.0347	N30	1.0565		

Optimization

The replicate plot (Figure G.2.1) shows a variability of repeated experiments (green triangles) lower than the overall variability (purple circles), which is the desired outcome. Data analysis shows negative skewness in the data (deviation from normality), but negative log transformation does not improve the model produced and was therefore not applied. No

square terms were detected.

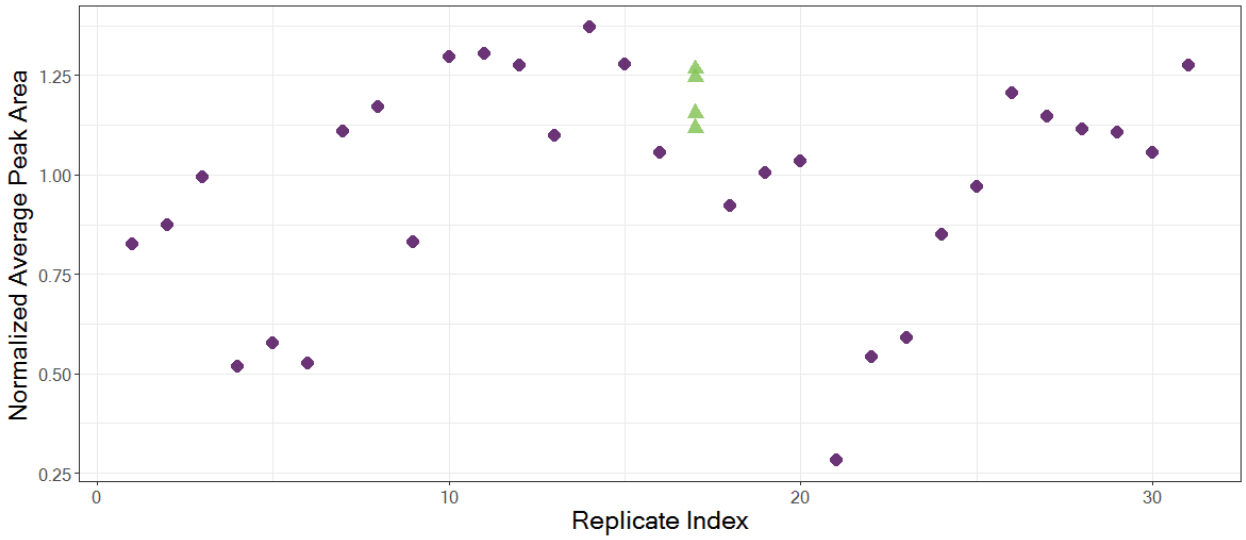


Figure G.2.1: Replicates plot of NAPA.

Basic model statistics are acceptable, with a model fit (R^2) of 0.888 and a future prediction precision (Q^2) of 0.802. Residuals normal probability graph also point to the adequacy of the model (Figure G.2.2), with residuals present on the diagonal and all but two within -2 and +2 standard deviations.

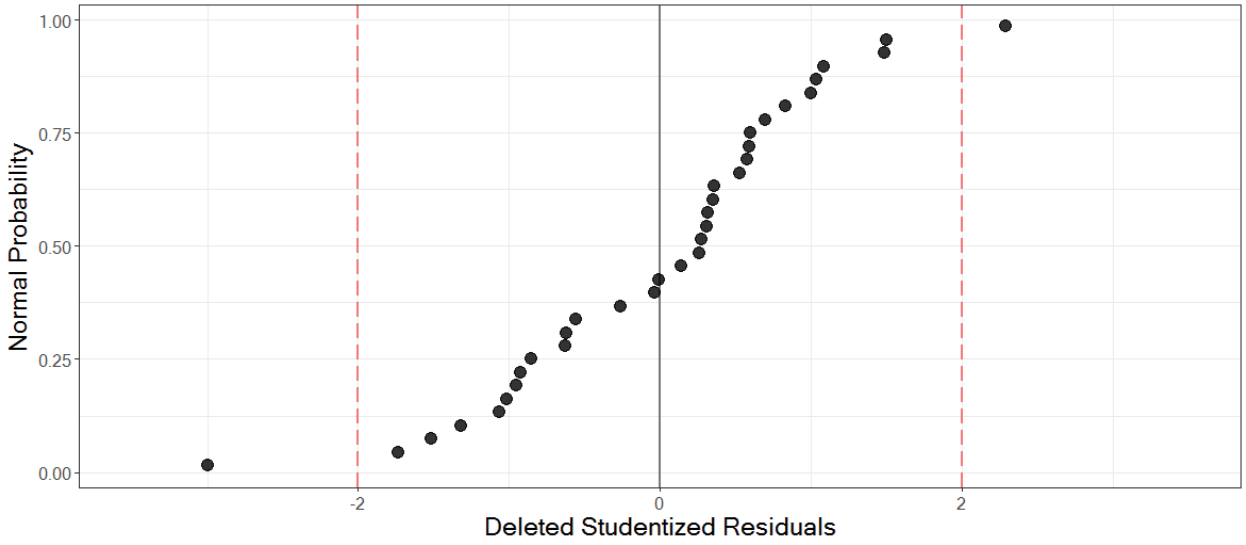


Figure G.2.2: Residuals normal probability

The final model predicting NAPA is described by the coefficients shown in Figure G.2.3.

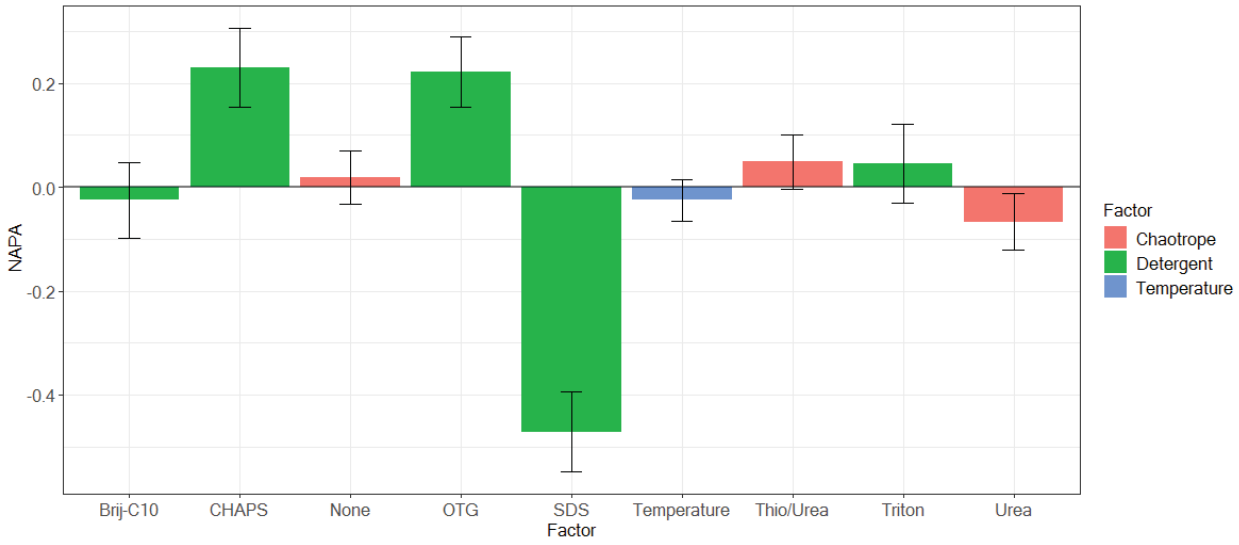


Figure G.2.3: Coefficients plot

G.2.4 Conclusions

The coefficients plot informs us that the best chaotrope is 1.2 M thiourea/4.8 M urea, with a close to significant impact on NAPA. Several detergents can be helpful, like OTG and CHAPS. This optimization confirms screening findings regarding temperature of denaturation: heating does not bring a significant advantage.

Running the optimizer confirms this qualitative evaluation, with the optimal settings being 1.2 M thiourea/4.8 M urea, 1.8% CHAPS as a detergent and 20 °C. The addition of 1.2 M thiourea/4.8 M urea is not attributed a significant coefficient, as observed in Figure G.2.3; however, it is on the very edge of being significant, and certainly better than any other chaotrope condition tested. This is why the optimizer selects this as the best chaotrope setting: it might not be significant overall, but it is the best choice. No weeding out of factors is carried out at the optimization stage, only the selection of the best (imperfect as it might be) settings.

G.3 Supplementary data 3 – Multifactorial design optimization (DoE) of the digestion process

G.3.1 Factors selected

Following an initial factors screening process, the following factors/levels were retained for optimization:

- Buffer system
  - Tris (pH 7.8)
  - NH_4HCO_3 (pH 7.8)
- Buffer concentration
  - 25 to 50 mM
- Time
  - 1 to 12 h
- Vortexing (every 30 min)
  - Yes
  - No

A response surface method (RSM) optimization using a full factorial design (2 levels) was carried out. The design space is illustrated in Figure G.3.1.

G.3.2 Experimental protocol

40 μL of pooled human liver microsomes (HLM, XenoTech, cat. no. H0630, Lenexa, Kansas, United States) are added in a 1.5 mL LoBind Eppendorf (cat. no. 0030108302, Eppendorf, Eppendorf, Hamburg, Germany). 20 μL of 1.2 M thiourea/4.8 M urea and 20 μL of 1.8% CHAPS are added to the sample. Following vortexing, the sample is heated at the indicated temperature for 15 minutes in the water bath.

After vortexing, 15 μL of 3 mM DTT is added to all samples which are incubated at room temperature for 30 minutes. This performs the reduction of the disulfide bonds in the

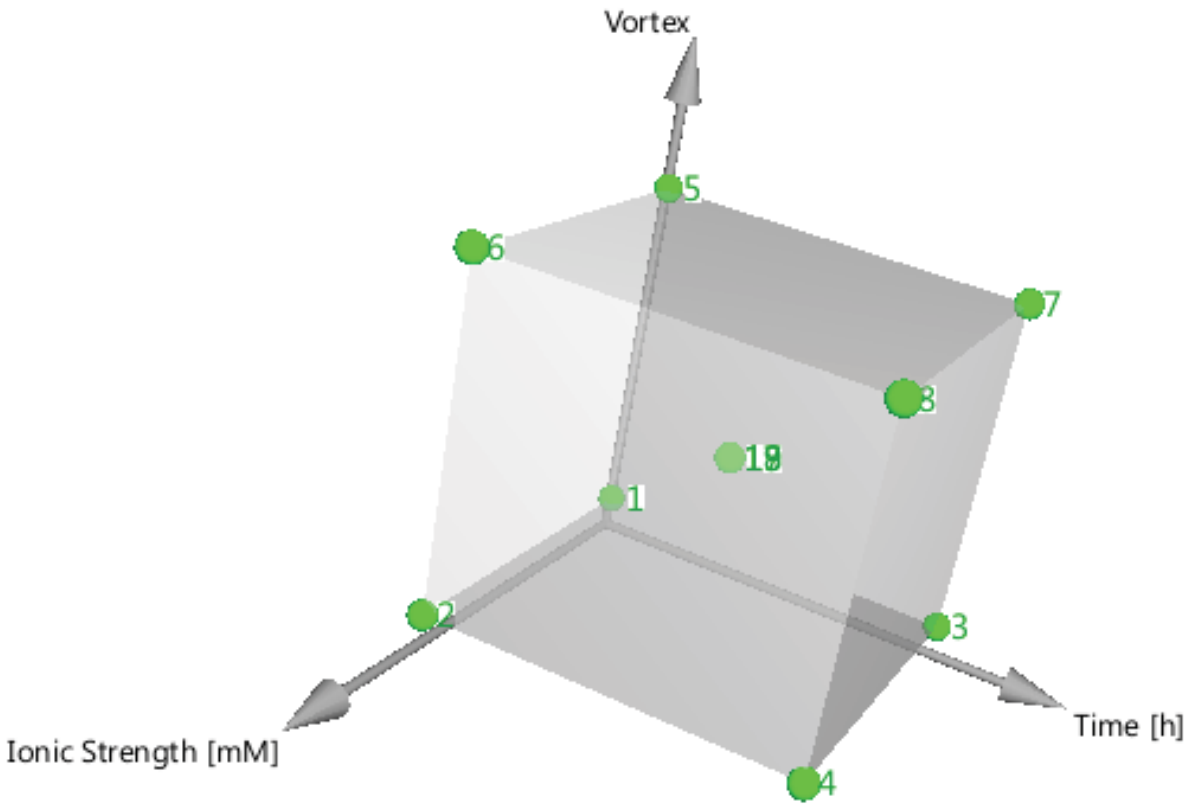


Figure G.3.1: Illustration of the full factorial design for digestion optimization.

proteins.

Alkylation of the reduced cysteines is then performed by adding 12 μL of 5 mM iodoacetamide to all samples. Following vortexing, samples are incubated in the dark at room temperature for 30 minutes.

The alkylating agent, iodoacetamide, is quenched by adding 13 μL of 10 mM DTT to all samples and vortexing.

Samples are diluted prior to digestion by adding 639 μL of the designated buffer (see Table G.3.1).

Digestion is carried out by adding 16 μg of trypsin reconstituted in 32 μL of 50 mM acetic acid and incubating for 5 hours at 37 °C with vortexing every 30 minutes. Trypsin activity is quenched by adding 8 μL of formic acid.

Tryptic digest clean-up and LC-MS/MS analysis were performed as described in the main

paper.

Table G.3.1: Experimental design for optimization of the human liver microsomes digestion.

Sample ID	Buffer ID	Buffer Concentration (mM)	Time (h)	Vortex
B-N1	Tris (pH 7.8)	25	1	No
B-N2	Tris (pH 7.8)	50	1	No
B-N3	Tris (pH 7.8)	25	1	No
B-N4	Tris (pH 7.8)	50	Overnight	No
B-N5	Tris (pH 7.8)	25	Overnight	Yes
B-N6	Tris (pH 7.8)	50	1	Yes
B-N7	Tris (pH 7.8)	25	1	Yes
B-N8	Tris (pH 7.8)	50	Overnight	Yes
B-N9	NH_4HCO_3 (pH 7.8)	25	Overnight	No
B-N10	NH_4HCO_3 (pH 7.8)	50	1	No
B-N11	NH_4HCO_3 (pH 7.8)	25	1	No
B-N12	NH_4HCO_3 (pH 7.8)	50	Overnight	No
B-N13	NH_4HCO_3 (pH 7.8)	25	Overnight	Yes
B-N14	NH_4HCO_3 (pH 7.8)	50	1	Yes
B-N15	NH_4HCO_3 (pH 7.8)	25	Overnight	Yes
B-N16	NH_4HCO_3 (pH 7.8)	50	Overnight	Yes
B-N17	Tris (pH 7.8)	37.5	6.5	Yes
B-N18	Tris (pH 7.8)	37.5	6.5	Yes
B-N19	Tris (pH 7.8)	37.5	6.5	Yes

Following peak integration, the normalized average peak area (NAPA) was calculated:

$$\text{NAPA} = \frac{\sum_j \text{NPA}_i}{j} \tag{G.3}$$

where:

$$\text{NPA} = \frac{A_{ij}}{\overline{A_j}} \tag{G.4}$$

A is a peak area, i is the sample and j is the transition number.

NPA therefore represents the peak area divided by the average peak area over all samples for one specific transition, and NAPA represents the average of the NPAs for all transitions of a given sample.

Data analysis was carried using the MODDE software (Umetrics/Sartorius Stedim Biotech, Umea, Sweden).

G.3.3 Results

Raw data

Raw measurements and calculated normalized peak areas are shown in Table G.3.2.

Table G.3.2: Measured peak areas.

	2D6_Q1 Peak Area (cps^2)				
Sample ID	**+2y2**	**+2y3**	**+2y4**	**+2y5**	**+2y6**
B-N1	1.86E+05	2.54E+05	1.04E+06	2.42E+05	2.50E+06
B-N2	1.68E+05	2.46E+05	9.62E+05	2.21E+05	2.35E+06
B-N3	4.85E+04	6.77E+04	2.92E+05	6.89E+04	7.04E+05
B-N4	7.65E+04	9.97E+04	4.39E+05	1.04E+05	1.07E+06
B-N5	1.68E+05	2.29E+05	9.23E+05	2.21E+05	2.25E+06
B-N6	1.67E+05	2.37E+05	9.67E+05	2.21E+05	2.28E+06
B-N7	4.85E+04	6.26E+04	2.70E+05	5.78E+04	6.35E+05
B-N8	8.10E+04	1.06E+05	4.68E+05	1.07E+05	1.11E+06
B-N9	1.68E+05	2.32E+05	9.70E+05	2.33E+05	2.32E+06
B-N10	1.70E+05	2.20E+05	9.75E+05	2.32E+05	2.33E+06
B-N11	1.03E+05	1.40E+05	5.93E+05	1.30E+05	1.43E+06
B-N12	1.09E+05	1.50E+05	6.22E+05	1.38E+05	1.50E+06
B-N13	1.74E+05	2.44E+05	1.02E+06	2.42E+05	2.45E+06
B-N14	1.76E+05	2.50E+05	1.03E+06	2.52E+05	2.50E+06
B-N15	1.02E+05	1.32E+05	5.53E+05	1.31E+05	1.33E+06
B-N16	1.14E+05	1.53E+05	6.39E+05	1.60E+05	1.61E+06
B-N17	1.43E+05	1.99E+05	8.58E+05	1.95E+05	2.04E+06
B-N18	1.44E+05	2.04E+05	8.37E+05	1.96E+05	2.03E+06
B-N19	1.29E+05	1.86E+05	7.76E+05	1.80E+05	1.82E+06
	2D6_Q2 Peak Area (cps^2)				
Sample ID	**+2y1**	**+2y3**	**+2y4**	**+2y5**	**+2y6**
B-N1	1.82E+05	1.91E+05	1.12E+06	1.23E+06	8.78E+05
B-N2	1.89E+05	1.72E+05	1.01E+06	1.10E+06	7.24E+05
B-N3	9.14E+04	4.52E+04	2.55E+05	2.92E+05	1.90E+05

B-N4	1.18E+05	8.39E+04	4.84E+05	5.28E+05	3.70E+05
B-N5	1.57E+05	1.58E+05	9.49E+05	1.00E+06	7.60E+05
B-N6	1.69E+05	1.55E+05	9.63E+05	1.01E+06	6.98E+05
B-N7	7.71E+04	4.35E+04	2.44E+05	2.83E+05	1.85E+05
B-N8	1.09E+05	8.74E+04	4.98E+05	5.56E+05	3.75E+05
B-N9	1.67E+05	1.71E+05	1.03E+06	1.03E+06	7.61E+05
B-N10	1.64E+05	1.75E+05	1.01E+06	9.95E+05	7.37E+05
B-N11	1.21E+05	9.29E+04	5.60E+05	6.31E+05	4.14E+05
B-N12	1.45E+05	1.13E+05	6.82E+05	7.16E+05	5.10E+05
B-N13	1.54E+05	1.81E+05	1.04E+06	1.10E+06	7.91E+05
B-N14	1.61E+05	1.61E+05	1.01E+06	1.08E+06	7.56E+05
B-N15	1.15E+05	8.27E+04	4.70E+05	5.22E+05	3.63E+05
B-N16	1.48E+05	1.19E+05	7.13E+05	7.90E+05	5.05E+05
B-N17	2.21E+05	2.26E+05	1.28E+06	1.41E+06	9.67E+05
B-N18	2.15E+05	1.92E+05	1.20E+06	1.32E+06	9.27E+05
B-N19	2.00E+05	1.86E+05	1.14E+06	1.23E+06	8.29E+05

	2D6_Q3 Peak Area (cps^2)				
Sample ID	**+3y2**	**+3y4**	**+3y5**	**+3y6**	**+3y8**
B-N1	1.19E+05	2.07E+05	1.93E+05	8.66E+04	1.43E+05
B-N2	1.09E+05	2.17E+05	2.06E+05	8.79E+04	1.48E+05
B-N3	7.20E+04	1.27E+05	1.18E+05	5.17E+04	8.25E+04
B-N4	8.50E+04	1.51E+05	1.21E+05	5.76E+04	1.01E+05
B-N5	1.11E+05	1.62E+05	1.64E+05	7.35E+04	1.17E+05
B-N6	1.16E+05	1.90E+05	1.72E+05	7.72E+04	1.25E+05
B-N7	7.07E+04	1.14E+05	1.04E+05	4.20E+04	7.77E+04
B-N8	7.83E+04	1.50E+05	1.34E+05	6.73E+04	9.90E+04
B-N9	1.19E+05	1.96E+05	1.86E+05	8.20E+04	1.44E+05
B-N10	1.10E+05	1.84E+05	1.79E+05	8.12E+04	1.17E+05
B-N11	1.12E+05	1.95E+05	1.85E+05	8.42E+04	1.44E+05
B-N12	8.94E+04	1.56E+05	1.50E+05	7.49E+04	1.14E+05
B-N13	1.23E+05	2.28E+05	1.91E+05	8.48E+04	1.44E+05
B-N14	1.22E+05	2.18E+05	2.12E+05	8.54E+04	1.33E+05
B-N15	9.70E+04	1.58E+05	1.64E+05	7.72E+04	1.20E+05
B-N16	1.01E+05	1.69E+05	1.77E+05	7.35E+04	1.31E+05
B-N17	1.29E+05	2.09E+05	2.10E+05	1.00E+05	1.53E+05

B-N18	1.17E+05	2.25E+05	2.18E+05	8.74E+04	1.42E+05
B-N19	1.10E+05	1.91E+05	1.64E+05	7.55E+04	1.29E+05
2D6_Q4 Peak Area (cps²)					
Sample ID	**+2y4**	**+2y8**	**+2y9**	**+3y4**	**+3y5**
B-N1	3.40E+05	3.47E+05	2.50E+05	3.69E+05	2.42E+05
B-N2	3.26E+05	3.12E+05	2.26E+05	3.40E+05	2.41E+05
B-N3	3.99E+05	4.08E+05	2.92E+05	4.32E+05	3.09E+05
B-N4	4.15E+05	4.32E+05	3.15E+05	4.46E+05	3.17E+05
B-N5	3.06E+05	3.16E+05	2.22E+05	3.20E+05	2.20E+05
B-N6	3.28E+05	3.33E+05	2.25E+05	3.48E+05	2.38E+05
B-N7	4.53E+05	4.42E+05	2.97E+05	4.53E+05	3.09E+05
B-N8	4.24E+05	4.38E+05	2.96E+05	4.63E+05	3.03E+05
B-N9	3.18E+05	3.17E+05	2.36E+05	3.29E+05	2.25E+05
B-N10	3.30E+05	3.21E+05	2.13E+05	3.29E+05	2.35E+05
B-N11	4.41E+05	4.40E+05	3.03E+05	4.67E+05	3.03E+05
B-N12	4.08E+05	4.03E+05	2.80E+05	4.20E+05	2.97E+05
B-N13	3.27E+05	3.29E+05	2.35E+05	3.41E+05	2.46E+05
B-N14	3.23E+05	3.25E+05	2.29E+05	3.41E+05	2.37E+05
B-N15	4.14E+05	4.19E+05	3.14E+05	4.23E+05	2.98E+05
B-N16	4.30E+05	4.46E+05	3.00E+05	4.33E+05	3.04E+05
B-N17	4.12E+05	4.24E+05	3.11E+05	4.42E+05	2.94E+05
B-N18	4.22E+05	4.15E+05	3.02E+05	4.32E+05	2.83E+05
B-N19	3.92E+05	3.78E+05	2.81E+05	3.97E+05	2.70E+05
2D6_Q5 Peak Area (cps²)					
Sample ID	**+2y1**	**+2y3**	**+2y4**	**+2y5**	**+2y6**
B-N1	2.91E+05	5.39E+05	6.01E+05	7.30E+05	1.30E+06
B-N2	2.57E+05	5.16E+05	5.66E+05	7.23E+05	1.30E+06
B-N3	1.42E+05	4.91E+05	5.43E+05	6.70E+05	1.18E+06
B-N4	1.58E+05	5.03E+05	5.29E+05	6.93E+05	1.28E+06
B-N5	2.36E+05	4.81E+05	5.40E+05	6.78E+05	1.16E+06
B-N6	2.60E+05	5.11E+05	5.82E+05	7.03E+05	1.26E+06
B-N7	1.49E+05	5.02E+05	5.74E+05	7.23E+05	1.27E+06
B-N8	1.76E+05	5.31E+05	5.81E+05	7.14E+05	1.29E+06
B-N9	2.50E+05	4.93E+05	5.65E+05	7.09E+05	1.27E+06
B-N10	2.44E+05	5.18E+05	5.86E+05	7.11E+05	1.28E+06

B-N11	1.89E+05	5.04E+05	5.46E+05	7.14E+05	1.28E+06
B-N12	1.77E+05	4.87E+05	5.40E+05	6.84E+05	1.21E+06
B-N13	3.16E+05	5.13E+05	5.80E+05	7.15E+05	1.27E+06
B-N14	2.40E+05	5.34E+05	6.05E+05	7.34E+05	1.33E+06
B-N15	1.58E+05	4.84E+05	5.34E+05	6.89E+05	1.22E+06
B-N16	1.64E+05	4.80E+05	5.56E+05	6.77E+05	1.25E+06
B-N17	2.57E+05	5.33E+05	5.80E+05	7.10E+05	1.27E+06
B-N18	2.31E+05	5.06E+05	5.79E+05	7.38E+05	1.23E+06
B-N19	2.17E+05	4.74E+05	5.21E+05	6.61E+05	1.16E+06
2D6_Q6 Peak Area (cps²)					
Sample ID	**+2y2**	**+2y4**	**+2y5**	**+2y6**	**+2y7**
B-N1	6.42E+03	2.52E+04	3.14E+03	4.17E+04	3.60E+04
B-N2	5.74E+03	1.95E+04	3.70E+03	3.62E+04	3.03E+04
B-N3	4.03E+04	2.18E+04	4.48E+03	4.74E+04	4.09E+04
B-N4	2.77E+04	3.49E+04	4.79E+03	5.07E+04	4.78E+04
B-N5	5.60E+03	2.48E+04	3.39E+03	4.32E+04	3.82E+04
B-N6	6.49E+03	2.32E+04	3.48E+03	4.14E+04	3.29E+04
B-N7	4.16E+04	2.77E+04	4.81E+03	4.00E+04	3.64E+04
B-N8	3.05E+04	3.42E+04	5.92E+03	5.52E+04	4.52E+04
B-N9	8.19E+03	2.51E+04	4.22E+03	4.63E+04	3.84E+04
B-N10	5.92E+03	1.99E+04	0.00E+00	3.78E+04	3.04E+04
B-N11	2.69E+04	3.86E+04	5.45E+03	6.59E+04	5.72E+04
B-N12	1.86E+04	3.39E+04	5.41E+03	6.19E+04	5.31E+04
B-N13	8.51E+03	3.04E+04	4.35E+03	5.36E+04	4.83E+04
B-N14	9.57E+03	2.55E+04	3.97E+03	4.48E+04	3.78E+04
B-N15	2.55E+04	3.90E+04	5.73E+03	6.13E+04	5.27E+04
B-N16	1.91E+04	3.16E+04	5.09E+03	5.78E+04	4.73E+04
B-N17	1.14E+04	4.11E+04	5.39E+03	7.09E+04	6.15E+04
B-N18	1.14E+04	3.89E+04	5.15E+03	6.30E+04	5.72E+04
B-N19	1.27E+04	4.30E+04	5.31E+03	7.20E+04	5.66E+04
3A4_Q1 Peak Area (cps²)					
Sample ID	**+2y17+2**	**+3y2**	**+3y4**	**+3y5**	**+3y8**
B-N1	1.68E+06	5.90E+05	4.28E+05	7.96E+05	7.06E+05
B-N2	1.51E+06	5.80E+05	3.78E+05	7.04E+05	7.36E+05
B-N3	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00

B-N4	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
B-N5	1.46E+06	5.62E+05	3.72E+05	6.89E+05	6.99E+05
B-N6	1.50E+06	5.52E+05	3.70E+05	6.51E+05	6.76E+05
B-N7	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
B-N8	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
B-N9	1.58E+06	6.03E+05	4.52E+05	7.67E+05	7.68E+05
B-N10	1.18E+06	4.40E+05	2.95E+05	5.03E+05	5.19E+05
B-N11	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
B-N12	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
B-N13	1.78E+06	6.09E+05	4.74E+05	7.70E+05	8.68E+05
B-N14	1.39E+06	4.78E+05	3.43E+05	6.25E+05	6.50E+05
B-N15	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
B-N16	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
B-N17	1.21E+06	4.21E+05	2.74E+05	5.34E+05	5.55E+05
B-N18	1.23E+06	4.01E+05	2.79E+05	5.31E+05	5.27E+05
B-N19	9.65E+05	3.61E+05	2.37E+05	4.77E+05	4.52E+05

	3A4_Q2 Peak Area (cps^2)				
Sample ID	**+3y10**	**+3y5**	**+3y7**	**+3y8**	**+3y9**
B-N1	9.31E+04	2.58E+05	1.20E+05	9.38E+04	1.27E+05
B-N2	9.97E+04	2.45E+05	1.27E+05	1.02E+05	1.14E+05
B-N3	7.28E+04	1.90E+05	8.66E+04	7.57E+04	9.64E+04
B-N4	7.18E+04	1.99E+05	9.86E+04	8.52E+04	1.13E+05
B-N5	8.61E+04	2.37E+05	9.90E+04	7.58E+04	1.12E+05
B-N6	7.48E+04	2.40E+05	1.11E+05	9.09E+04	1.09E+05
B-N7	7.18E+04	1.93E+05	8.77E+04	6.83E+04	9.61E+04
B-N8	9.65E+04	2.16E+05	1.12E+05	8.68E+04	1.08E+05
B-N9	8.23E+04	2.79E+05	1.31E+05	9.14E+04	1.21E+05
B-N10	8.88E+04	2.26E+05	8.45E+04	7.78E+04	1.10E+05
B-N11	1.04E+05	2.84E+05	1.18E+05	9.19E+04	1.47E+05
B-N12	9.52E+04	2.28E+05	1.08E+05	9.14E+04	1.24E+05
B-N13	9.89E+04	2.90E+05	1.22E+05	1.08E+05	1.36E+05
B-N14	9.73E+04	2.81E+05	1.15E+05	1.03E+05	1.32E+05
B-N15	1.03E+05	2.45E+05	1.15E+05	9.60E+04	1.30E+05
B-N16	9.67E+04	2.41E+05	1.07E+05	9.09E+04	1.19E+05
B-N17	1.32E+05	3.07E+05	1.44E+05	1.03E+05	1.61E+05

B-N18	1.14E+05	3.17E+05	1.33E+05	1.09E+05	1.48E+05
B-N19	1.01E+05	2.66E+05	1.27E+05	1.04E+05	1.24E+05

3A4_Q3 Peak Area (cps²)					
Sample ID	**+2y1**	**+2y2**	**+2y3**	**+2y4**	**+2y5**
B-N1	2.97E+06	3.03E+06	3.92E+06	3.02E+06	5.24E+06
B-N2	2.81E+06	2.90E+06	3.87E+06	2.92E+06	5.08E+06
B-N3	3.50E+06	3.71E+06	4.71E+06	3.50E+06	6.34E+06
B-N4	3.48E+06	3.53E+06	4.77E+06	3.55E+06	6.20E+06
B-N5	2.67E+06	2.78E+06	3.59E+06	2.61E+06	4.77E+06
B-N6	2.85E+06	2.95E+06	3.83E+06	2.80E+06	5.10E+06
B-N7	3.64E+06	3.72E+06	4.99E+06	3.67E+06	6.55E+06
B-N8	3.60E+06	3.71E+06	4.94E+06	3.73E+06	6.67E+06
B-N9	2.92E+06	3.04E+06	3.95E+06	2.99E+06	5.42E+06
B-N10	2.82E+06	2.96E+06	3.77E+06	2.81E+06	5.11E+06
B-N11	3.63E+06	3.90E+06	4.91E+06	3.70E+06	6.47E+06
B-N12	3.29E+06	3.45E+06	4.46E+06	3.33E+06	5.92E+06
B-N13	3.32E+06	3.37E+06	4.40E+06	3.35E+06	5.95E+06
B-N14	3.19E+06	3.33E+06	4.27E+06	3.27E+06	5.81E+06
B-N15	3.42E+06	3.66E+06	4.70E+06	3.53E+06	6.27E+06
B-N16	3.33E+06	3.57E+06	4.65E+06	3.48E+06	6.07E+06
B-N17	3.99E+06	4.04E+06	5.20E+06	3.87E+06	7.06E+06
B-N18	3.78E+06	3.98E+06	5.29E+06	3.90E+06	6.88E+06
B-N19	3.42E+06	3.51E+06	4.63E+06	3.43E+06	6.16E+06

3A4_Q4 Peak Area (cps²)					
Sample ID	**+2y1**	**+2y3**	**+2y4**	**+2y5**	**+2y5+2**
B-N1	8.64E+04	4.50E+05	2.08E+05	7.79E+05	5.02E+05
B-N2	7.50E+04	4.28E+05	2.06E+05	7.80E+05	4.89E+05
B-N3	0.00E+00	0.00E+00	0.00E+00	1.89E+05	0.00E+00
B-N4	0.00E+00	0.00E+00	0.00E+00	2.28E+05	0.00E+00
B-N5	6.82E+04	4.09E+05	2.02E+05	7.34E+05	4.63E+05
B-N6	7.43E+04	4.22E+05	2.06E+05	7.82E+05	4.89E+05
B-N7	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
B-N8	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
B-N9	7.81E+04	4.31E+05	2.21E+05	7.83E+05	4.64E+05
B-N10	7.53E+04	4.15E+05	1.97E+05	7.36E+05	4.74E+05

B-N11	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
B-N12	0.00E+00	4.08E+04	1.98E+04	7.80E+04	5.29E+04
B-N13	8.56E+04	4.40E+05	2.31E+05	8.38E+05	5.03E+05
B-N14	7.40E+04	4.74E+05	2.26E+05	7.59E+05	4.90E+05
B-N15	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
B-N16	0.00E+00	3.61E+04	2.14E+04	8.35E+04	5.20E+04
B-N17	5.49E+04	3.91E+05	2.02E+05	7.37E+05	4.65E+05
B-N18	6.47E+04	3.96E+05	1.98E+05	6.95E+05	4.46E+05
B-N19	5.33E+04	3.59E+05	1.86E+05	6.76E+05	4.22E+05
3A4_Q5 Peak Area (cps²)					
Sample ID	**+2y7**	**+2y9**	**+3y5**	**+3y6**	**+3y7**
B-N1	1.49E+06	3.73E+05	2.08E+05	2.29E+05	1.26E+06
B-N2	1.89E+06	4.49E+05	2.45E+05	3.08E+05	1.61E+06
B-N3	2.71E+04	8.19E+03	0.00E+00	0.00E+00	2.75E+04
B-N4	2.74E+04	4.80E+03	0.00E+00	0.00E+00	1.89E+04
B-N5	1.37E+06	3.09E+05	1.81E+05	2.08E+05	1.18E+06
B-N6	1.83E+06	4.16E+05	2.16E+05	2.89E+05	1.60E+06
B-N7	3.52E+04	5.74E+03	5.26E+03	4.89E+03	2.90E+04
B-N8	2.69E+04	7.18E+03	5.72E+03	3.91E+03	2.35E+04
B-N9	1.86E+06	4.14E+05	2.26E+05	2.62E+05	1.48E+06
B-N10	1.90E+06	4.60E+05	2.44E+05	2.96E+05	1.67E+06
B-N11	1.83E+04	0.00E+00	0.00E+00	0.00E+00	1.41E+04
B-N12	1.93E+04	0.00E+00	0.00E+00	0.00E+00	1.64E+04
B-N13	1.90E+06	4.62E+05	2.41E+05	3.12E+05	1.69E+06
B-N14	2.26E+06	5.26E+05	2.73E+05	3.55E+05	1.86E+06
B-N15	1.63E+04	0.00E+00	0.00E+00	0.00E+00	1.27E+04
B-N16	1.54E+04	0.00E+00	0.00E+00	0.00E+00	1.58E+04
B-N17	4.04E+06	9.14E+05	4.96E+05	5.79E+05	3.46E+06
B-N18	3.70E+06	8.76E+05	4.58E+05	5.68E+05	3.21E+06
B-N19	3.33E+06	7.39E+05	4.39E+05	5.33E+05	3.05E+06
3A4_Q6 Peak Area (cps²)					
Sample ID	**+3y1**	**+3y2**	**+3y5**	**+3y6**	**+3y8**
B-N1	5.30E+05	4.68E+05	4.37E+05	1.79E+05	2.16E+06
B-N2	4.66E+05	4.80E+05	4.15E+05	1.65E+05	2.00E+06
B-N3	3.34E+04	4.65E+04	3.30E+04	9.68E+03	1.08E+05

B-N4	5.84E+04	6.12E+04	5.60E+04	1.77E+04	1.90E+05
B-N5	4.36E+05	4.07E+05	3.57E+05	1.56E+05	1.80E+06
B-N6	4.86E+05	4.33E+05	4.10E+05	1.76E+05	2.13E+06
B-N7	4.03E+04	5.80E+04	2.45E+04	6.45E+03	1.02E+05
B-N8	8.13E+04	6.66E+04	4.32E+04	1.38E+04	1.96E+05
B-N9	4.74E+05	4.75E+05	4.18E+05	1.66E+05	2.22E+06
B-N10	5.18E+05	4.63E+05	3.89E+05	1.79E+05	2.02E+06
B-N11	1.17E+05	5.74E+04	6.16E+04	2.64E+04	2.41E+05
B-N12	1.90E+05	1.51E+05	1.42E+05	5.37E+04	6.91E+05
B-N13	6.02E+05	4.78E+05	4.23E+05	1.67E+05	2.23E+06
B-N14	4.77E+05	4.68E+05	4.18E+05	1.73E+05	2.12E+06
B-N15	6.28E+04	6.07E+04	5.02E+04	1.87E+04	2.18E+05
B-N16	1.49E+05	1.50E+05	1.33E+05	5.42E+04	6.62E+05
B-N17	4.61E+05	4.71E+05	3.90E+05	1.62E+05	1.92E+06
B-N18	4.78E+05	4.54E+05	3.62E+05	1.66E+05	2.08E+06
B-N19	4.33E+05	4.41E+05	3.58E+05	1.46E+05	1.80E+06

	2D6*9_WT Peak Area (cps^2)				
Sample ID	**+2y5**	**+2y6**	**+2y7**	**+2y8**	**+2y9**
B-N1	7.29E+04	6.56E+04	1.15E+05	1.18E+05	6.28E+04
B-N2	1.14E+05	9.31E+04	1.26E+05	1.36E+05	7.12E+04
B-N3	1.46E+05	1.40E+05	1.98E+05	2.30E+05	1.10E+05
B-N4	1.64E+05	1.57E+05	2.27E+05	2.20E+05	1.30E+05
B-N5	8.06E+04	6.39E+04	1.14E+05	8.83E+04	4.29E+04
B-N6	9.79E+04	8.09E+04	1.37E+05	1.14E+05	6.57E+04
B-N7	1.47E+05	1.50E+05	2.16E+05	2.29E+05	1.09E+05
B-N8	1.72E+05	1.74E+05	2.07E+05	2.47E+05	1.01E+05
B-N9	8.54E+04	8.97E+04	1.32E+05	1.07E+05	6.11E+04
B-N10	9.14E+04	8.60E+04	1.32E+05	1.08E+05	6.10E+04
B-N11	1.73E+05	1.37E+05	1.92E+05	2.35E+05	1.12E+05
B-N12	1.50E+05	1.44E+05	2.07E+05	1.96E+05	1.14E+05
B-N13	7.66E+04	7.84E+04	1.17E+05	1.18E+05	5.42E+04
B-N14	9.05E+04	8.74E+04	1.24E+05	1.32E+05	5.96E+04
B-N15	1.60E+05	1.37E+05	1.94E+05	2.31E+05	9.73E+04
B-N16	1.58E+05	1.45E+05	1.85E+05	2.00E+05	1.04E+05
B-N17	1.78E+05	1.51E+05	2.09E+05	2.14E+05	1.18E+05

B-N18	1.61E+05	1.54E+05	2.15E+05	2.01E+05	1.14E+05
B-N19	1.44E+05	1.26E+05	2.02E+05	1.89E+05	1.02E+05
3A4*11_WT Peak Area (cps²)					
Sample ID	**+3y5**	**+3y6**	**+3y7**	**+3y8**	**+3y9**
B-N1	2.10E+04	3.90E+04	2.00E+04	1.03E+04	2.39E+04
B-N2	2.47E+04	3.82E+04	1.39E+04	1.01E+04	2.30E+04
B-N3	4.42E+04	6.78E+04	3.45E+04	2.17E+04	5.32E+04
B-N4	6.90E+04	7.96E+04	3.71E+04	1.87E+04	5.51E+04
B-N5	2.19E+04	2.43E+04	1.38E+04	7.95E+03	2.12E+04
B-N6	2.24E+04	2.41E+04	1.34E+04	9.25E+03	2.61E+04
B-N7	4.25E+04	6.27E+04	2.89E+04	2.06E+04	4.59E+04
B-N8	6.59E+04	9.84E+04	3.86E+04	2.55E+04	8.00E+04
B-N9	2.45E+04	3.32E+04	1.51E+04	8.79E+03	2.32E+04
B-N10	2.13E+04	2.83E+04	1.39E+04	8.58E+03	2.17E+04
B-N11	5.99E+04	8.22E+04	3.93E+04	3.11E+04	6.94E+04
B-N12	4.20E+04	7.38E+04	3.95E+04	2.38E+04	5.47E+04
B-N13	2.04E+04	3.75E+04	1.52E+04	7.74E+03	2.66E+04
B-N14	3.39E+04	4.05E+04	2.23E+04	1.18E+04	2.47E+04
B-N15	4.53E+04	7.98E+04	3.60E+04	2.38E+04	5.19E+04
B-N16	6.21E+04	8.64E+04	3.22E+04	2.23E+04	5.73E+04
B-N17	9.42E+04	1.16E+05	5.34E+04	4.72E+04	8.17E+04
B-N18	7.36E+04	1.07E+05	4.72E+04	3.60E+04	7.77E+04
B-N19	7.53E+04	9.79E+04	4.76E+04	3.26E+04	7.40E+04
3A4*13_WT Peak Area (cps²)					
Sample ID	**+2y1**	**+2y2**	**+2y3**	**+2y3+2**	**+2y4**
B-N1	1.02E+07	2.73E+06	3.42E+07	6.29E+06	9.14E+05
B-N2	9.69E+06	2.60E+06	3.27E+07	6.07E+06	8.71E+05
B-N3	4.35E+06	1.12E+06	1.44E+07	2.74E+06	3.87E+05
B-N4	4.43E+06	1.15E+06	1.48E+07	2.93E+06	3.88E+05
B-N5	9.03E+06	2.41E+06	2.92E+07	5.37E+06	7.76E+05
B-N6	9.56E+06	2.55E+06	3.06E+07	5.65E+06	8.23E+05
B-N7	4.27E+06	1.13E+06	1.43E+07	2.66E+06	3.85E+05
B-N8	4.72E+06	1.24E+06	1.59E+07	3.13E+06	4.25E+05
B-N9	9.54E+06	2.58E+06	3.20E+07	5.89E+06	8.43E+05
B-N10	1.01E+07	2.71E+06	3.34E+07	6.17E+06	9.00E+05

B-N11	3.27E+06	8.10E+05	1.05E+07	2.15E+06	2.84E+05
B-N12	3.41E+06	8.85E+05	1.13E+07	2.31E+06	2.93E+05
B-N13	9.98E+06	2.66E+06	3.24E+07	6.03E+06	8.71E+05
B-N14	9.87E+06	2.63E+06	3.18E+07	5.90E+06	8.49E+05
B-N15	2.97E+06	7.58E+05	9.76E+06	2.08E+06	2.63E+05
B-N16	3.47E+06	8.97E+05	1.15E+07	2.35E+06	3.00E+05
B-N17	9.89E+06	2.62E+06	3.17E+07	5.91E+06	8.60E+05
B-N18	9.62E+06	2.59E+06	3.27E+07	5.97E+06	5.19E+05
B-N19	9.25E+06	2.45E+06	3.04E+07	5.55E+06	8.05E+05

	3A4*8_WT Peak Area (cps²)				
Sample ID	**+2y6**	**+2y7**	**+2y7+2**	**+2y8**	**+2y9**
B-N1	6.30E+05	6.86E+06	3.41E+06	5.75E+06	4.21E+06
B-N2	5.79E+05	6.33E+06	3.30E+06	5.57E+06	4.07E+06
B-N3	2.13E+05	2.16E+06	1.17E+06	1.95E+06	1.48E+06
B-N4	3.38E+05	3.67E+06	1.90E+06	3.24E+06	2.48E+06
B-N5	5.81E+05	6.36E+06	3.13E+06	5.34E+06	3.84E+06
B-N6	6.22E+05	6.56E+06	3.35E+06	5.65E+06	4.08E+06
B-N7	1.63E+05	1.70E+06	8.67E+05	1.46E+06	1.17E+06
B-N8	3.41E+05	3.69E+06	1.94E+06	3.27E+06	2.53E+06
B-N9	6.09E+05	6.64E+06	3.30E+06	5.66E+06	4.23E+06
B-N10	6.20E+05	6.45E+06	3.30E+06	5.63E+06	4.16E+06
B-N11	5.35E+05	5.96E+06	3.08E+06	5.22E+06	3.95E+06
B-N12	5.72E+05	6.03E+06	3.15E+06	5.40E+06	4.09E+06
B-N13	6.08E+05	6.59E+06	3.32E+06	5.67E+06	4.12E+06
B-N14	6.12E+05	6.89E+06	3.45E+06	5.86E+06	4.27E+06
B-N15	5.21E+05	5.62E+06	2.90E+06	4.99E+06	3.75E+06
B-N16	5.82E+05	6.26E+06	3.28E+06	5.49E+06	4.07E+06
B-N17	6.19E+05	6.94E+06	3.44E+06	5.97E+06	4.32E+06
B-N18	6.19E+05	6.79E+06	3.41E+06	5.81E+06	4.39E+06
B-N19	6.06E+05	6.65E+06	3.32E+06	5.71E+06	4.20E+06

A rapid analysis shows no difference between overall NAPA, NAPA with only qualitative peptides and NAPA with only quantitative peptides. Consequently, the model using the

overall NAPA (all peptides) as presented in Table G.3.3 will be used.

Table G.3.3: Normalized averaged peak areas (NAPA) measured.

Sample ID	NAPA
B-N1	1.2631
B-N2	1.2269
B-N3	0.5552
B-N4	0.6212
B-N5	1.1207
B-N6	1.1896
B-N7	0.5521
B-N8	0.6413
B-N9	1.2418
B-N10	1.1398
B-N11	0.7070
B-N12	0.6987
B-N13	1.3271
B-N14	1.2660
B-N15	0.6588
B-N16	0.7126
B-N17	1.4272
B-N18	1.3825
B-N19	1.2683

Optimization

The replicate plot (Figure G.3.2) shows a variability of repeated experiments (green triangles) lower than the overall variability (purple circles), which is the desired outcome.

Basic model statistics are acceptable, with a model fit (R^2) of 0.722 and a future prediction precision (Q^2) of 0.597. Residuals normal probability graph also point to the adequacy of the model (Figure G.3.3), with residuals present on the diagonal and all but two within -2 and +2 standard deviations.

The final model predicting NAPA is described by the coefficients shown in Figure G.3.4.

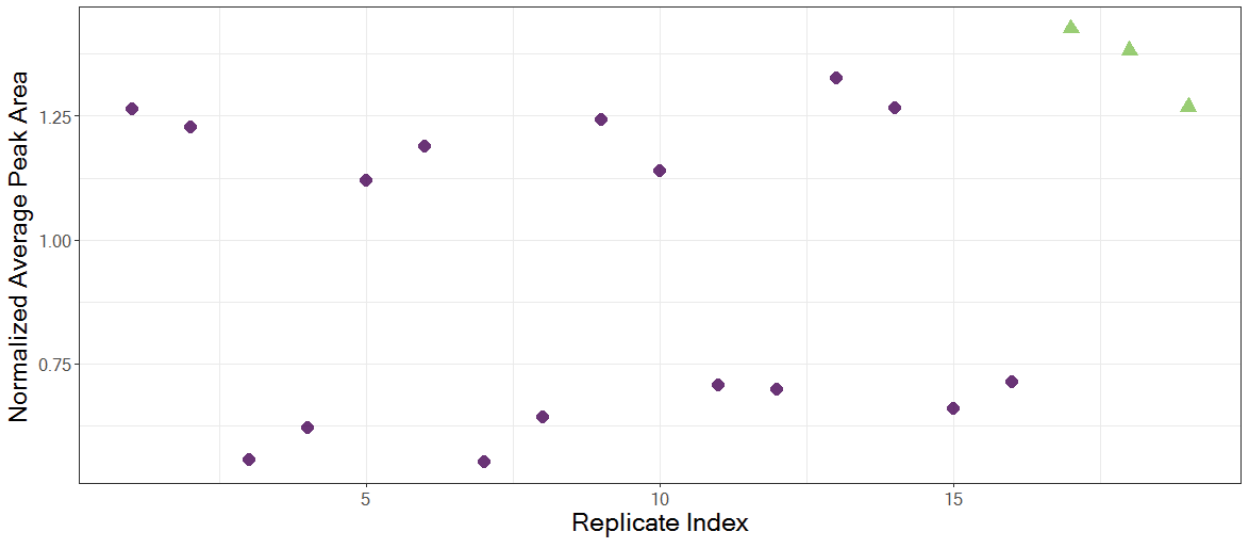


Figure G.3.2: Replicates plot of NAPA.

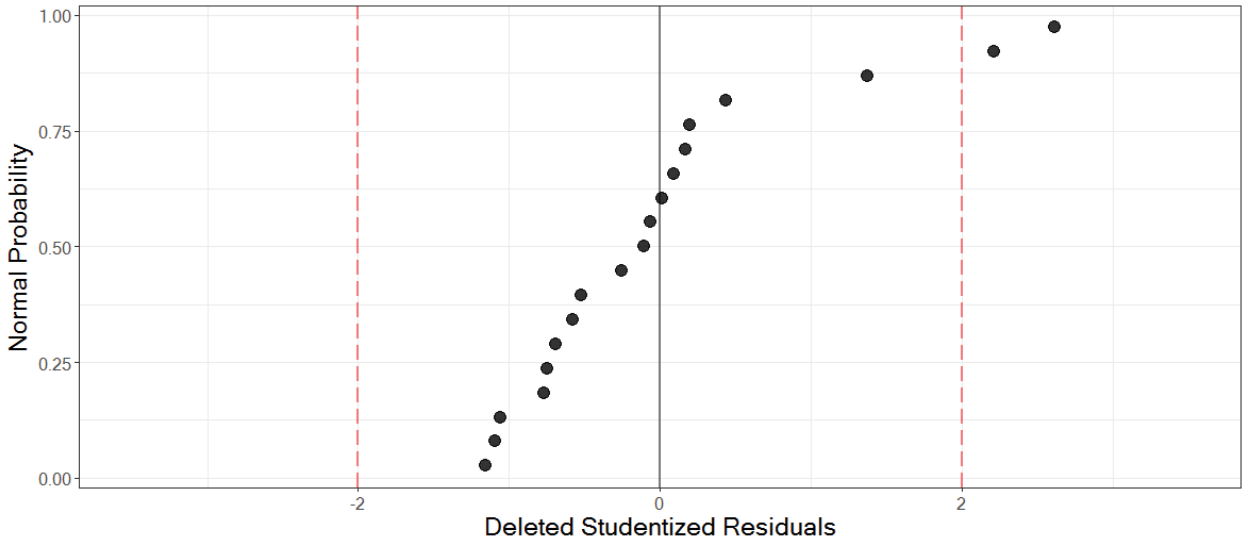


Figure G.3.3: Residuals normal probability.

G.3.4 Conclusions

The coefficients plot informs us that the only significant factor is time, with a negative impact. Running the optimizer confirms this qualitative evaluation that shortening the digestion time has the desired impact.

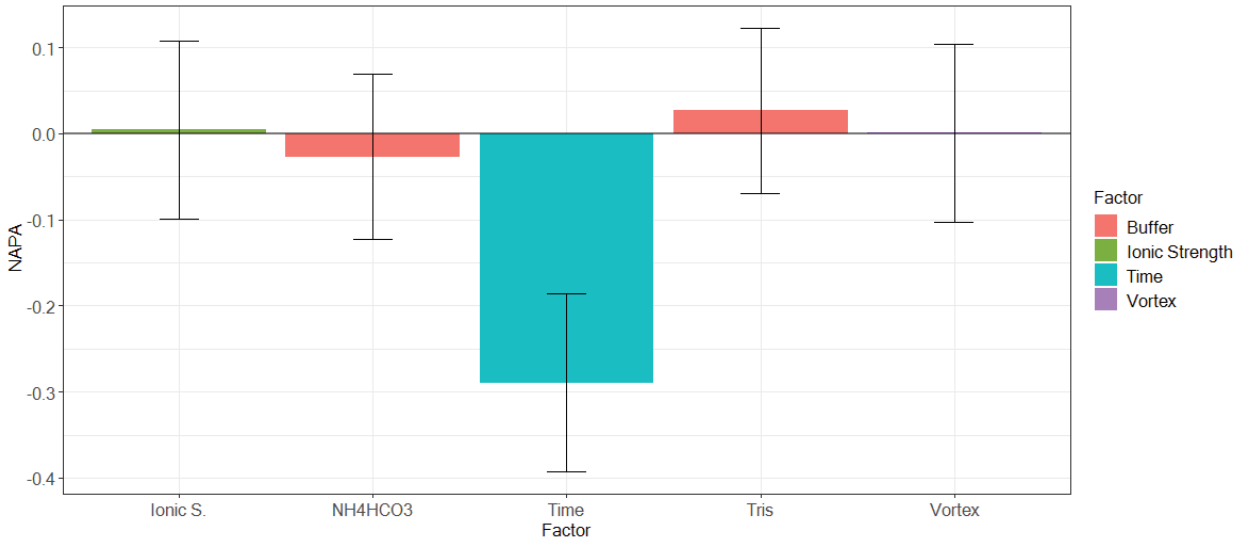


Figure G.3.4: Coefficients plot for digestion optimization.