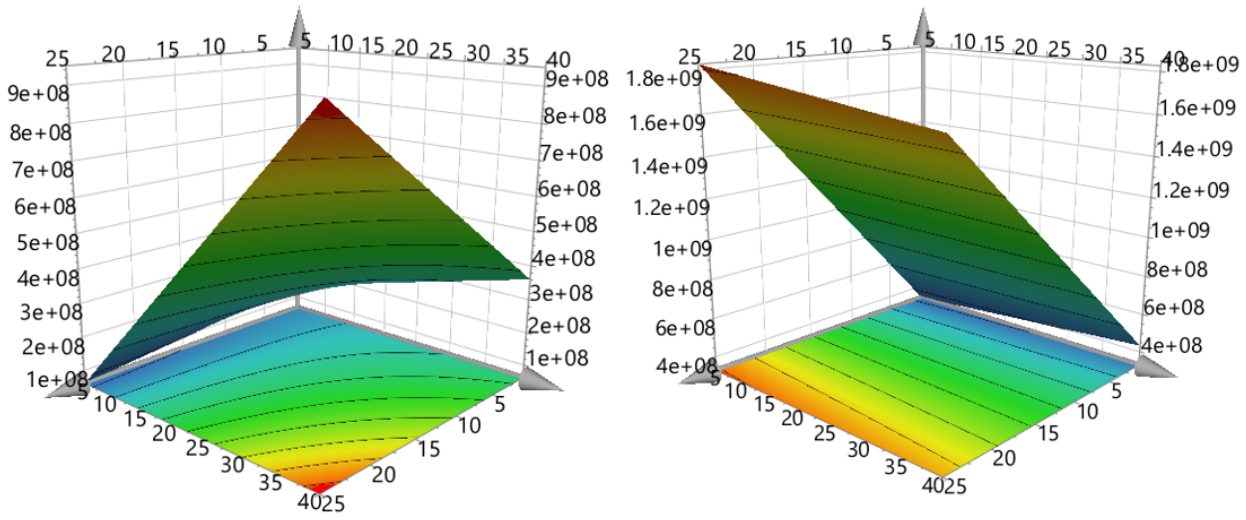
CHALMERS



Evaluation of signal linearity of Charged Aerosol Detector in Supercritical Fluid Chromatography

Master of Science Thesis 2023

Kinga Kamińska

Department of Chemistry and Chemical Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
ASTRAZENECA R&D
Göteborg, Sweden 2023

Master's Thesis 2023

Evaluation of signal linearity of Charged Aerosol Detector in Supercritical Fluid Chromatography

Master of Science Thesis

Kinga Kamińska

CHALMERS
UNIVERSITY OF TECHNOLOGY

Chemistry and Chemical Engineering
Chalmers Univesity of Technology
Göteborg, Sweden 2023

Evaluation of signal linearity of Charged Aerosol Detector in Supercritical Fluid Chromatography
Master of Science Thesis
KINGA KAMIŃSKA - kaminska@student.chalmers.se

Supervisors:
Astrid Buica, Early Chemical Development, Pharmaceutical Sciences, BioPharma R&D - AstraZeneca
Christoph Bauer, Data Science and Modelling, Pharmaceutical Sciences, AstraZeneca
Hanna Leek, Early Chemical Development, Pharmaceutical Sciences, BioPharma R&D - AstraZeneca
Kristina Öhlén, Early Chemical Development, Pharmaceutical Sciences, BioPharma R&D - AstraZeneca
Victor Spelling, Early Chemical Development, Pharmaceutical Sciences, BioPharma R&D - AstraZeneca

Examiner:
Aldo Jesorka, Chemistry and Chemical Engineering, Biophysical Technology Laboratory, Chalmers University of Technology

Chemistry and Chemical Engineering
Chalmers University of Technology
SE-412 96 Göteborg
Sweden

In collaboration with AstraZeneca R&D
SE-431 83 Göteborg
Sweden

Figures in the text, where no reference is given, are my own.

Cover: Response surface plots (MLR Model) of *ampicillin* (left) and *doxycycline* (right) for

IPA (modifier) combined with ACN (make-up). Obtained with MODDE®13.

Göteborg, Sweden 2023

Acknowledgements

First of all, I would like to express my sincerest thanks to my supervisors at AstraZeneca, Astrid Buica, Christoph Bauer, Hanna Leek, Kristina Öhlén and Victor Spelling. Thank you for giving me the opportunity to perform my master's thesis with you and for being a part of this project. Thank you for all instructions and answers to my questions. Without you, this work would never see the daylight, and I would lose as much as I gained thanks to this project.

Many thanks to Carolina Sanchez and Johanna Kollback for their help and support in operating TECAN.

I also want to thank Olivia Luige and Astrid Buica for all the lighthearted and random conversations that made my thoughts wander among various topics, giving my mind a rest. Thanks also to the other members of the SSL team for making me feel like a part of the crew. Thank you with all my heart, for every good word, and above all for being kind souls.

Help and support coming from my examiner from Chalmers University of Technology, Aldo Jesorka, will not be forgotten either. Thank you very much.

Kinga Kamińska

2023.05.26, Gothenburg

Evaluation of signal linearity of Charged Aerosol Detector in Supercritical Fluid Chromatography

KINGA KAMIŃSKA
Department of Chemistry and Chemical Engineering

This Master's Thesis project was performed at Separation Sciences Laboratory, Early Chemical Development, Pharmaceutical Sciences, R&D, AstraZeneca, Göteborg, Sweden.

Abstract

Charged Aerosol Detector (CAD) is growing in popularity due to its ability to detect non- or semi-volatile compounds, including those that do not have any UV-chromophores, therefore invisible to a UV/Vis detector. For CAD, the detection is based on the formation of aerosols and the response signal is found to be non-linear. Although the detector is designed for (Ultra) High Performance Liquid Chromatography ((U)HPLC), it has been successfully implemented in the Supercritical Fluid Chromatography (SFC) system. Due to the non-linear response of this detector, this study aimed to evaluate various analysis parameters and physichochemical characteristics of a wide range of compounds, to see what causes this non-linearity and if it's consistent.

Analyses were performed using a Waters Acquity UPC$^{2®}$. Instead of a chromatography column, a static mixer filled with beads was used, providing turbulent flow and efficient mixing of the SFC mobile phase, CO_2, modifier and an analyte solution. 34 compounds with different physicochemical properties were used in the study.

The physicochemical properties evaluated were cLogP, number of aliphatic and aromatic hydroxyl groups, number of rings within a molecule, number of amine substituents and number of hydrogen donors and acceptors. The analytical parameters studied were choice of modifier (MeOH, EtOH, IPA), modifier percentage (5, 15, 25, 40%), choice of make-up solvent (MeOH, EtOH, ACN) and volumetric flow rate of the make-up (0.5, 1.0, 1.5 mL/min). Design of Experiments (DOE) was used to exclude some of the aforementioned parameters, based on the results for two test compounds. Excluded were EtOH as a modifier and make-up and volumetric flow rates higher than 0.5 mL/min. Concentration span of analytes was 0.4-21 mM for test compounds and 0.5-25 mM for the remaining ones. Data collection for CAD response signal peak area was performed using Empower® 3 and the analysis in terms of correlation evaluation and orthogonality was performed in Python. Python has also been utilised to compute structural descriptors. Visualisation was performed with Python, MODDE® 13 and Excel.

To assess linearity, the coefficient of determination, R^2, was used. Detailed investigation was carried out for compounds and methods with $R^2<0.8$. The influence of combined analytical parameters was also performed. The contribution coming from solvents used to dissolve analytes was considered as well.

Good linearity was observed for methods where 40% MeOH as modifier was combined with MeOH and ACN as make-up and for 40% and 25% IPA as modifier combined with MeOH as make-up. Computed structural descriptors for which the CAD response signal showed to be linear are high number of aliphatic hydroxyl groups, high number of rings within a molecule as well as either high or low number of hydrogen donors and acceptors. These observations likely correlate with the analytes' solubilities in the mobile phase.

Keywords: Supercritical Fluid Chromatography, Charged Aerosol Detector, Pharmaceutical Industry, Linearity, Data Modeling

Contents

List of Abbreviations

ABPR	Automated Back Pressure Regulator
CAD	Charged Aerosol Detector
DOE	Design of Experiments
FLD	Fluorescence Detector
GC	Gas Chromatography
LC	Liquid Chromatography
MLR	Multiple Linear Regression
MVA	Multivariate data Analysis
MS	Mass Spectrometry
SFC	Supercritical Fluid Chromatography
SQD	Single Quadrupole Detector
UPC^2	Ultra high Performance convergence Chromatography
UV/Vis Detector	Ultraviolet/Visible Detector
QSAR	Quantitative Structure-Activity Relationship
ACN	Acetonitrile
CO_2	Carbon dioxide
DEA	Diethylamine
DMSO	Dimethyl sulfoxide
DOPE	Lipid, 1,2-Dioleoyl-sn-glycero-3-phosphoethanolamine
DSPC	Lipid, 1,2-Distearoyl-sn-glycero-3-phosphorylcholine
EtOH	Ethanol
FA	Formic acid
IPA	Isopropanol
MC3	Lipid, D-Lin-MC3-DMA
MeOH	Methanol
NH_3	Ammonia

1 Introduction

This section will start with a broad introduction where the importance of separation and detection in analytical chemistry will be described. Short presentation about SFC and diverse SFC-compatible detectors will be included followed by reasons for having such variety. In this section it will also be possible to read about the place where the experimental part of this study was carried out, as well as about the study's aim and objectives.

With each published article or book, our knowledge and understanding of the world around us and the phenomena that we observe every day grows [1]. One of the fields that contributes to this understanding is analytical chemistry, which uses various analytical instruments and methods that allow for separation and detection of matter [2]. The reason why these analytical techniques are important is that they enable qualitative and quantitative analysis. An important group of analytical methods is chromatography, which allows for separation of a mixture into their individual constituents. While detection can be performed without separation, successful separation cannot be recognised without detection. That is why it is essential to realise the importance of detection in analytical chemistry [3].

The most common detectors used in liquid chromatography are UV/Vis Detector, Refractive Index (RI) Detector, Fluorescence Detector (FLD) and Mass Spectrometry (MS) [4, 5]. However, detectors that have become more popular are aerosol detectors like CAD and Evaporative Light Scattering Detector (ELSD). Both are considered near-universal detectors for the detection of volatile and also semi-volatile solutes [6]. The reason why these novel detectors have grown in popularity is that they can detect analytes that do not contain chromophores, so called non-UV absorbers. The ability to detect non-UV absorbers is also becoming more important as more substances appear on the market with low or no UV-absorption, examples of which are lipids or the boronic acid-based inhibitor *arginase*, which was used for research in cancer immunotherapy [7].

SFC is a liquid chromatography (LC) technique that uses supercritical CO_2 (main mobile phase), mixed with organic co-solvents [8]. One of the strengths of SFC is the nature of this eluent which makes the SFC compatible with most detectors on the market [4]. This compatibility also applies to CAD and the interest to understand the SFC-CAD combination in more detail is growing, an example of which being a recently published article about a quantification strategy performed with CAD coupled to SFC [9]. The SFC-CAD combination has not been thoroughly explored as most approaches on CAD have been performed using either HPLC or UHPLC [10,11]. One of the aspects that has not been discussed in detail so far is the linearity of the CAD signal. What has been observed is that the response of this detector generally results in an overall concave-shaped signal-response curve but it also exhibits a wide concentration range over which it might be approximated as linear [12].

The experimental part of this study was carried out at the Separation Science Laboratory (SSL), one of the groups at the AstraZeneca Research and Development (R&D) site in Göteborg. On a daily basis, SSL works with purification in mg to kg scale, where analytical instruments like (U)HPLC or SFC are utilised during method development, characterisation and analysis of the target compounds [13]. One of many detectors used by SSL is the CAD, which might be used for purity analysis of non-UV absorbers, which SSL would be particularly interested in.

1.1 Aim and objectives

The aim of this study was to evaluate the linearity of the signal from CAD coupled to SFC, for a wide range of compounds in order to find trends in the operating parameters and/or structural characteristics of a molecule that affect the linearity of the signal. Finding these trends would make the CAD response signal better understood and could lead to a way of predicting the response of compounds whose characteristic features would be similar to those analysed in this study.

The aim of this study was achieved with the help of the following objectives:

- Select operational parameters and compounds, including concentrations.
- Test and evaluate all operational parameters (full factorial design of experiments) for a selection of compounds (two).
- Test the remaining compounds (34) on a number of operational parameters, selected based on the results for the initial screen, using full factorial design of experiments.
- Evaluate the results based on CAD signal, including linearity and solvent contribution.
- Correlate CAD signal to method parameters and/or analyte structure.

In the future, CAD could be used more effectively for analysing various types of compounds, and above all, non-UV absorbers. The results could also help to predict what analysis parameters would lead to a linear signal, and at the same time to the accurate quantification of an analyte. The findings of this study could be used for more effective detection at SSL in AstraZeneca when using CAD.

2 Background

This section will start with the presentation of AstraZeneca followed by a literature review covering topics like SFC, detectors compatible with SFC, as well as detailed explanation about working principles of CAD. Also included in this section is a brief description of the multivariate data analysis used.

2.1 AstraZeneca

AstraZeneca is a multinational pharmaceutical company with subsidiaries in many countries [13]. In Sweden, such sites are located in Göteborg (R&D centre) and Södertälje (production site). Some of the company's main research areas are oncology, rare diseases and biopharmaceuticals. The company employs over 83,000 people world-wide, including about 2,500 at the site in Göteborg. As one of the leading pharmaceutical companies, AstraZeneca wants the medicines they create to be of the highest quality for the well-being of patients [13]. That is why it is so important for this company to be able to detect and quantify the compounds synthesised. There are many departments within AstraZeneca, one of which is Early Chemical Development (ECD), which is further subdivided into smaller groups. One such group is the SSL, where separation and purification of candidate drugs, intermediates, and test compounds synthesised by chemistry departments within AstraZeneca is performed. SSL also collaborates with data science in order to handle, process and evaluate obtained chromatographic data.

2.2 Supercritical Fluid Chromatography - SFC

SFC is an LC technique in which the mobile phase used is in a supercritical state [14]. SFC is said to exhibit both gas chromatography (GC) and LC behaviour [15] due to the nature of the mobile phase used. The phase diagram of the most commonly used mobile phase, CO_2, is presented in Figure 1.

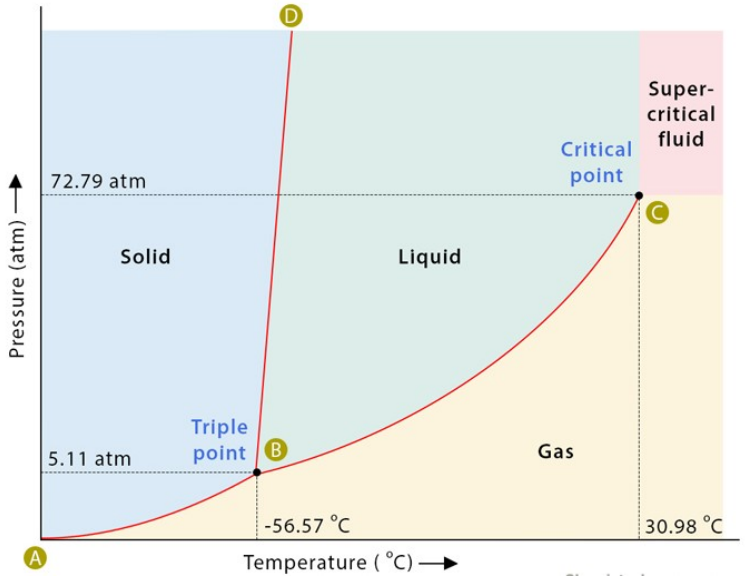


Figure 1: Phase diagram of CO_2 [16].

Analysing Figure 1 it can be seen that CO_2 can be in a solid, liquid or gaseous as well as supercritical state [16]. The reason it is called supercritical is that it is above its critical point

where it does not longer exist in gaseous or liquid form but it behaves similar to both states simultaneously. The supercritical state is reached for CO_2 at relatively low temperature and pressure, 30.98°C and 72.79 atm. In order to increase the polarity of the SFC mobile phase, co-solvents are used, which may also contain alkaline or acidic additives. SFC usually operates under normal-phase retention mechanism, which means that the stationary phase is more polar than the mobile phase.

Using CO_2 in SFC leads to many advantages such as being compatible for the analysis of non-polar compounds [14]. High solubilisation power, low viscosity and high diffusivity in the SFC mobile phase allows for a high flow rate which makes it possible to perform analysis faster than in LC, and also improves column efficiency [17]. The advantage over GC is that it can successfully separate compounds showing thermal instability [14]. The disadvantage is that the mobile phase used is highly non-polar and even with the addition of co-solvent, the analysis of highly polar compounds is difficult because they might not be soluble in the supercritical fluid [17].

SSL favours the use of SFC over other chromatography techniques because it is a more 'green' alternative due to the usage of CO_2 mixed with organic co-solvents as mobile phase. The CO_2 that is used for all systems at SSL is a waste product from other industries in the Göteborg area. Furthermore, the CO_2 used in preparative scale is automatically recycled within the systems. The use of CO_2 is also dictated by health precautions, where exposure to more hazardous apolar solvents used in normal phase LC systems is decreased.

2.3 Detectors

2.3.1 Charged Aerosol Detector - CAD

CAD, an aerosol-based detector, allows allows for detection of non- or semi-volatile compounds [4, 10]. The working principles of CAD are presented in Figure 2.

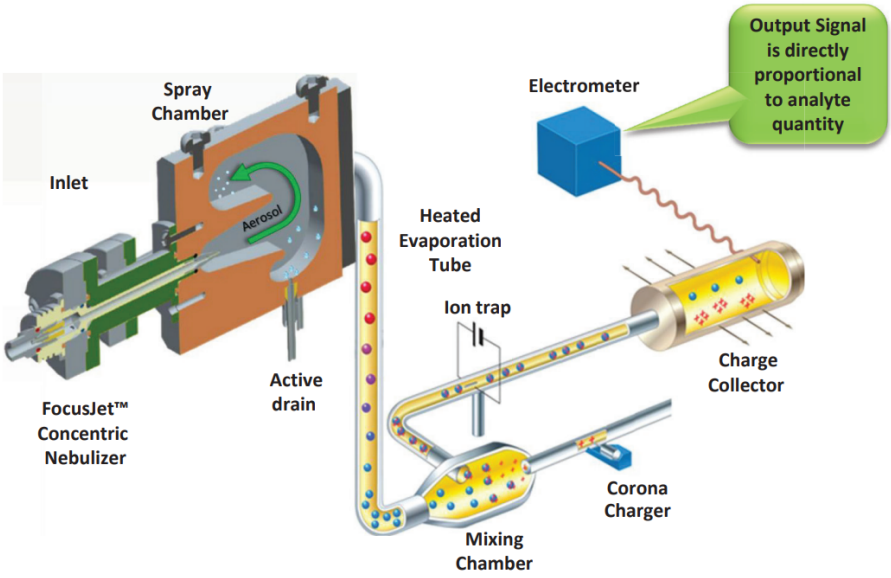


Figure 2: Working principles of the Charged Aerosol Detector [18].

Aerosol production begins with nebulisation in the *spray chamber*, where the SFC eluent is mixed with the nebulising gas, which is usually N_2 [10, 18]. At the nebuliser tip, the infusion of the eluent and the nebulising gas occurs, resulting in the formation of aerosols of various sizes. Larger aerosol droplets condense, and are then removed from the system. Smaller ones, on the

other hand, travel along with the stream of N_2 to the *heated evaporation tube*, where volatile species are evaporated, including any volatile solvent present in the sample. In this way, the dried aerosol particles enter the *mixing chamber*, where the surface diffusional charging process takes place. Positive charges obtained from secondary nitrogen stream by a corona discharge needle attach to the surface of the dried aerosol particles, thus not causing the dried aerosol particles to become ions. Charged in this way, dried particles travel to the *ion trap* where fast mobile particles, e.g. free ions, are collected and removed from the system. Then, the dried aerosol particles go to the *charge collector*, which is a sensitive *electrometer*, where the charges previously attached to the surface of the dried aerosol particles are counted. The amount of these charges is proportional to the surface area of the dried aerosol particles, which is directly proportional to the amount of an analyte injected.

The response from CAD is generally non-linear over wide concentration ranges [12, 19] and an example is presented in Figure 3.

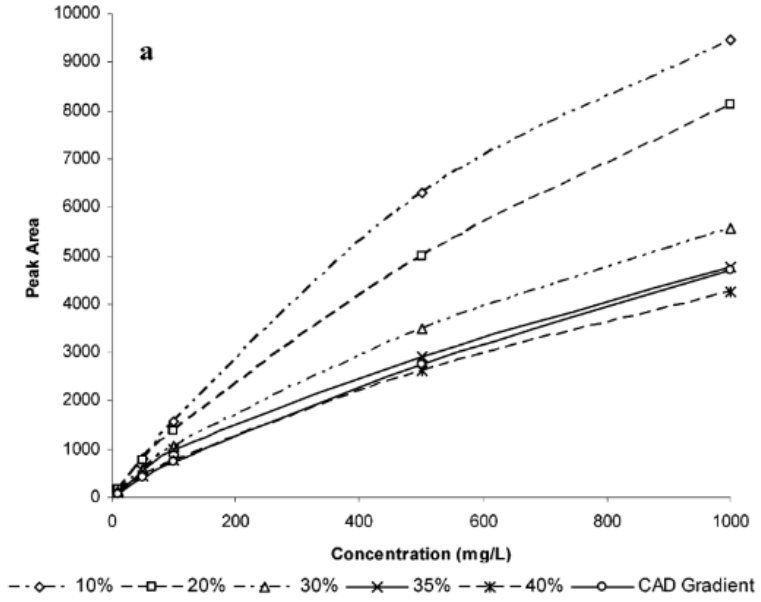


Figure 3: CAD response curves for different amounts of MeOH in mobile phase [19].

There might be several reasons why the CAD response curve exhibits an overall concave shape and why it might differ depending on the analytical parameters applied. One aspect that may influence the signal is the diameter of the created dried particles which depends on aerosol droplet diameter (related to density and viscosity of the mobile phase), density of the particle and concentration of the analyte [6]. Therefore aerosol creation might be either encouraged or suppressed leading to differences in the response curve of CAD, as can be seen in Figure 3 where the curves do not align with each other for different percentages of the modifier.

A possible explanation to the overall concave shaped response curve might be saturation of the detector, which means that no more than a certain amount of aerosol particles can be formed and detected [20]. At high analyte concentration, there is more of the analyte than there is of the solvent. In that case, interactions between the analyte molecules might be stronger than between the analyte and the solvent, which may result in partial particle agglomeration. This could cause aerosols to form in a wide distribution of sizes. Bigger droplets could then condense in the *spray chamber* and be removed and only smaller ones are detected. Higher sample concentration could also result in precipitation of an analyte in the system and therefore suppressed detection [21].

CAD has both advantages and disadvantages [4]. One of the biggest advantages marketed is that the detection ability is independent of chemical properties of an analyte. Another advantage, primarily over ELSD, is that it has broader dynamic response range where low (ng) and high (μg) amounts of analyte can be detected with high sensitivity [6]. Another, more obvious, advantage over UV is that solutes without chromophores can be detected, since CAD is not an absorption based technique.

Despite how it is marketed, one of CAD's main drawbacks, is the non-linear signal like the one presented in Figure 3, which can lead to inaccurate quantification if the limits where this linearity is exceeded have not been carefully investigated and defined. Another disadvantage is the variation of the response with mobile phase composition [6] which also makes it challenging to foresee the signal linearity.

2.4 Data evaluation

Given the scale of the project presented in this study, one should take into account the amount of information obtained after all the analyses were performed. Therefore, multivariate data analysis (MVA) software, such as Python was used to analyse obtained data. Python is the programming language that was utilised when using the internal web development environment, JupyterHub [22]. For DOE, MODDE® 13 (Sartorius AG, Göttingen, Germany) was used because it allows for successful prediction of the impact of many factors on a given observable. MODDE is also capable of creating response contour plots by fitting either Multiple Linear Regression (MLR) or Partial Least Square (PLS) models into obtained data [23].

MVA is a statistical tool that involves a study of two or more variables in a data set [24]. These variables can be operating parameters or other factors that are compared to the component of the experiment under control. In this way it is possible to identify trends and other events found in the data. MVA can be used to systematically analyse large chromatographic data, for example to predict purity and/or yield from input raw material or to model a detector response through regression analysis [25]. Multivariate regression analysis is an analysis in which multiple simple straight line models are compiled. Examples of multivariate regression models are MLR, PLS and Principal Component Analysis (PCA).

MLR or PLS can be used as tools for method development in analytical chromatography as they have been considered to be highly predictive quantitative structure-activity relationship (QSAR) models [26]. QSAR models have become more popular as the amount of obtained and processed data for large sets of molecules grew in recent years. Moreover, comparison of obtained data to physicochemical and structural parameters of the analytes have became more important.

In this study, an MLR model was used to create response contour plots of the response (peak area) as they visualise and predict the joint influence of more than one variable on the observable. The fit of the model into the data is described by the validation metrics, R^2 and Q^2, also called 'goodness of fit' and 'goodness of prediction', respectively [27]. The fit of the model is considered ideal when these values are close to each other and close to one.

3 Methodology

In this section, the methodology used in this study will be presented. It will also contain a detailed description of the experimental setup used during the analysis of compounds, with the distinction of constant and variable parameters. The next part of the methodology section will present the experimental design in which the variable parameters of the analysis will be specified as well as all analysed compounds including sample preparation prior to the analysis. This section will end with a description of data analysis, where the types of data collected and the ways of their processing will be specified.

3.1 Experimental setup

The analytical instrument used during laboratory work was an Acquity UPC$^{2®}$ (Waters^{TM}, Milford, USA) equipped with a binary solvent manager supplying the mobile phase and co-solvent; a sample manager with a 10 μL loop and two 48 position sample trays; a convergence manager which monitors and regulates the incoming flow rate and pressure of CO_2 into the system ; a column manager with 8 chromatography column positions, one position being replaced by a static mixer filled with beads (250 μL, internally provided) to suit the wide range of analysed compounds and provide good miscibility of eluent with co-solvents and analyte solution; a Photodiode Array (PDA) detector (range 210-400 nm); a Single Quadrupole Detector (SQD). Additionally, the system was equipped with an Adjustable LC-MS Post-Column Flow Splitter (MilliporeSigma, Massachusetts,USA) and Thermo Scientific^{TM} Dionex^{TM} Corona^{TM} Veo^{TM} CAD, (ThermoFisher Scientific^{TM}, Massachusetts, USA). The settings of the CAD are presented in Table 1.

Table 1: Settings applied to CAD.

Setting	Value
Power function	1
Data collection rate	100 Hz
Filter constant	5.0 s
Evaporation temperature	Low (35°C) ± 5 °C

For all runs, isocratic conditions have been used with a sample injection volume of 5 μL and time of analysis being either 0.6 or 1 min. Each sample concentration has been injected twice for statistical purposes. As a main mobile phase for the SFC system CO_2 was used, which was then mixed with different organic co-solvents. These organic co-solvents were a combination of a modifier and NH_3 as an additive, with a ratio of 100/20 mM. The total volumetric flow rate of the mobile phase was set to be constant at 3.5 mL/min. The temperature was set to 40°C. Automated back pressure regulator was set to 120 bar. Additional solvent flow, called *make-up* flow has been utilised during the analysis, in order to obtain a stable base line and prevent condensation and loss of solutes that could occur in the transfer line between ABPR and CAD [9]. Splitting ratio of the adjustable post-column flow splitter (Analyical Scientific Instruments Richmond, USA) was 20:1 (CAD:MS). Acquisition and processing of the UV, MS and CAD data was performed using Empower® 3 (Waters, Milford, USA) software. A schematic over the experimental configuration used in this study is presented in Figure 4.

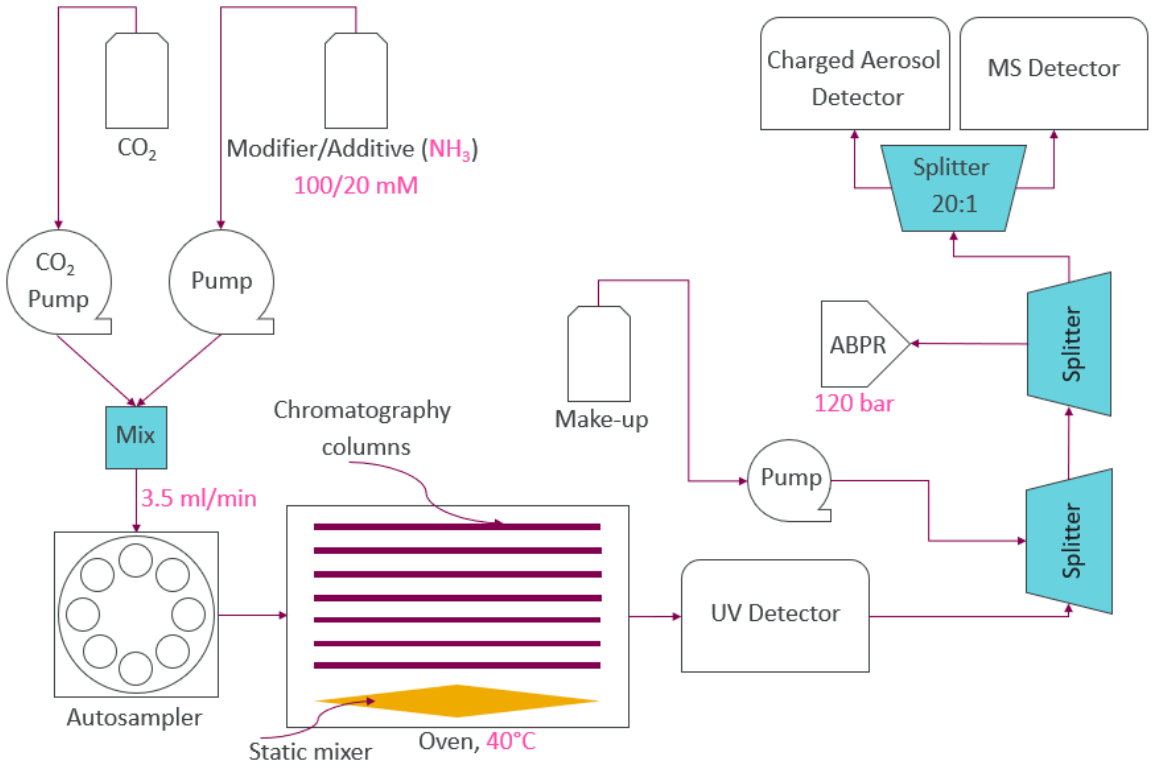


Figure 4: Schematic of the analytical setup for SFC-UV/MS/CAD experiments.

Parameters like *modifier* and *make-up* as well as the percentage of modifier and volumetric flow of the make-up are considered as independent *variables*.

3.2 Design of experiment

3.2.1 Variable selection

The initial selection of analytical parameters was based on commonly used solvents and conditions when using the SFC system. These variables are presented in Table 2.

Table 2: Initial analysis parameters for SFC-UV/MS/CAD experiments.

Variable	Descriptor
Modifier	MeOH, EtOH, IPA
Modifier [%]	5, 15, 25, 40
Make-up flow	MeOH, EtOH, ACN
Flow rate of make-up [ml/min]	0.5, 1.0, 1.5
Sample concentration [mM]	0.4, 1.7, 8.3, 15, 17, 21

With the help of MODDE a *full factorial* design of experiments using the variables, modifier choice, amount of modifier, make-up flow rate and make-up solvent, have been randomised and combined into 108 unique *methods* applied to each sample concentration. The methods were denoted as follows:

modifier_ %modifier_ make-up_ volumetric flow of make-up
example: MeOH_5_muMeOH_05

Due to the large number of selected analytical parameters, presented in Table 2, the exclusion of some of them was considered. To be able to do that, it was necessary to assess the significance

of these parameters' influence on the CAD signal. This assessment was made with the help of MODDE, by analysing coefficient plots where all parameters presented in Table 2, and combinations thereof, were included. To create such a plot, the MLR model was fitted into the data obtained with Empower. Two compounds, *ketoprofen* and *lidocaine*, were used as test substances, and the obtained coefficient plots are presented as attachments in Appendix A.

The decision to exclude a parameter was based on various factors. When analysing the impact of the different modifiers, MeOH, EtOH and IPA, it was decided that because their impact is similar, the factors deciding about the exclusion of one of them will be their molecular structure and their elution strength. EtOH, a structurally related solvent and with a weaker elution strength than MeOH [28], was excluded. Similar reasoning was applied when excluding EtOH as make-up. The impact of percentage modifier was varied, so none of them was excluded. The influence of volumetric flow of make-up was close to zero or negative, therefore 1.0 and 1.5 ml/min volumetric flows were excluded.

After the exclusion, the number of methods studied had decreased to 16. Analytical parameters that were used in further SFC-MS/UV/CAD experiments are shown in Table 3.

Table 3: Analysis parameters after variable exclusion based on MODDE coefficient plots for *ketoprofen* and *lidocaine*.

Factor	Variable
Modifier	MeOH, IPA
Modifier [%]	5, 15, 25, 40
Make-up flow	MeOH, ACN
Flow rate of make-up [ml/min]	0.5
Sample concentration [mM]	0.50, 6.63, 12.8, 18.9, 25.0
Notation for sample concentration	c_1, c_2, c_3, c_4, c_5

For the lowest concentration of *ketoprofen*, 0.4 mM, no signal could be detected, meaning that the concentration was below the limit of detection. Therefore the lowest concentration has been amended to be 0.5 mM. Obtained CAD peak areas for both *ketoprofen* and *lidocaine* resulted in linear signal, which means that the linearity limit for these test compound has not been reached. As the aim of this study was to investigate linearity and concentration ranges where this linearity is exceeded, also the highest concentration was increased to be 25 mM.

3.2.2 Chemical space

The solvents used in this study like methanol (MeOH), ethanol (EtOH), isopropanol (IPA), acetonitrile (ACN) and dimethyl sulfoxide (DMSO) have been purchased from MilliporeSigma (Massachusetts, USA), diethylamine (DEA) from Thermo Scientific™ Chemicals (former Alfa Aestar, Massachusetts, USA). Mobile phase used for SFC (CO_2) was supplied by Linde Gas AB (former AGA, Solna, Sweden). NH_3, the basic additive, has been purchased from Thermo Scientific Chemicals (Massachusetts, USA) as 2M in MeOH solution. The compounds used in this study as analytes, are commercially available and the supplier list is provided in Appendix B. The full list of the compounds, their chemical structures and acronyms used throughout this study as well as solvents used for dissolving them, are presented in Table 4.

Table 4: Compounds, their labels, chemical structures and solvents used for dissolving them. Chemical structures are taken from PubChem database [29]. The ratio of MeOH:DEA or DMSO:DEA was 19:1 for all amino acids except *L-isoleucine*, for which this ratio was 13:1.

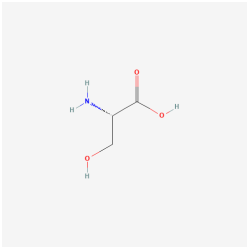
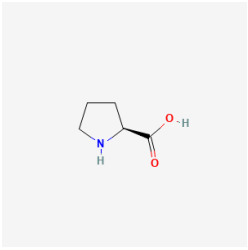
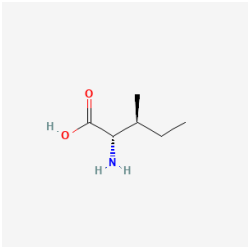
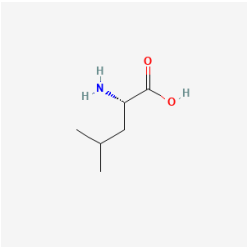
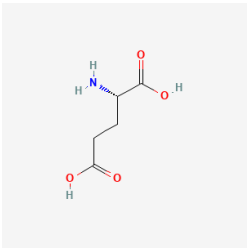
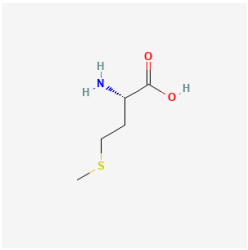
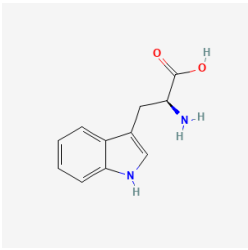
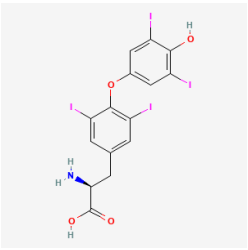
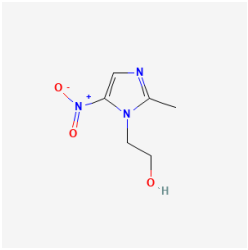
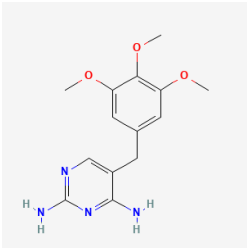
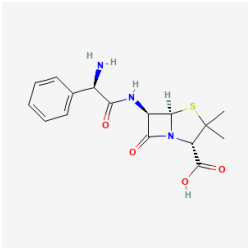
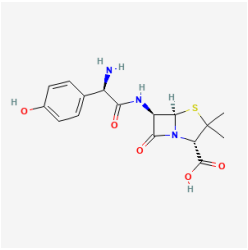
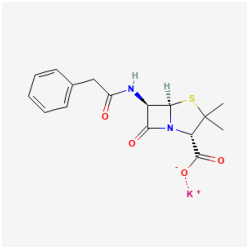
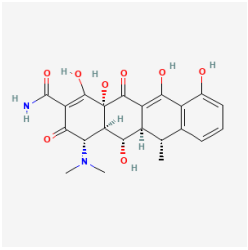
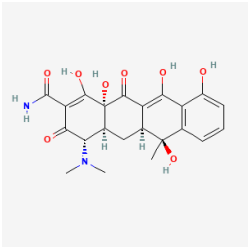
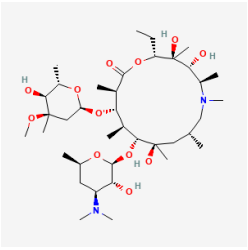
	Amino acids			
Chemical structure				
Name	L-Serine	L-Proline	L-Isoleucine	L-Leucine
Label	Ser	Pro	Iso	Leu
Solvent	MeOH:DEA	MeOH	MeOH:DEA	MeOH:DEA
Chemical structure				
Name	L-Glutamic acid	L-Methionine	L-Tryptophan	L-Thyroxine
Label	Glu	Met	Try	Trx
Solvent	DMSO:DEA	MeOH:DEA	MeOH:DEA	DMSO
	Antibiotics			
Chemical structure				
Name	Metronidazole	Trimethoprim	Ampicillin	Amoxicillin
Label	Metro	Tri	Amp	Amo
Solvent	MeOH	DMSO	MeOH	DMSO
Chemical structure				
Name	Penicillin G K	Doxycycline	Tetracycline	Azithromycin
Label	PenG	Dox	Tet	Azi
Solvent	MeOH	MeOH	MeOH	MeOH

Table 5: Continuation of Table 4.

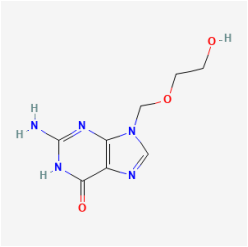
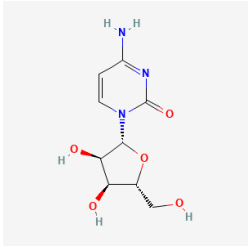
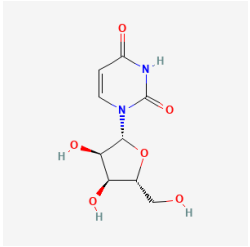
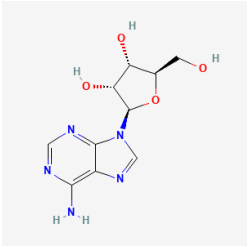
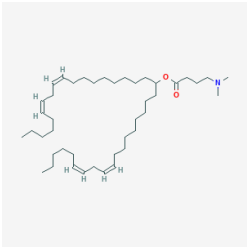
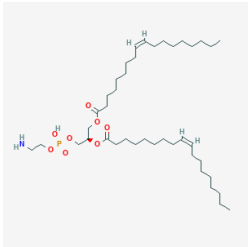
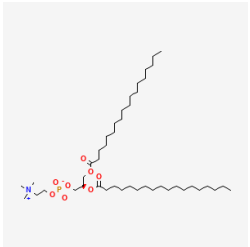
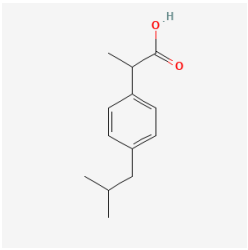
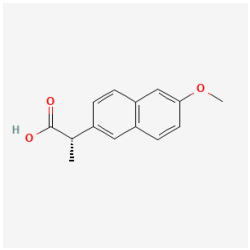
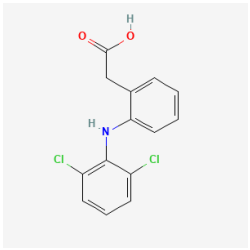
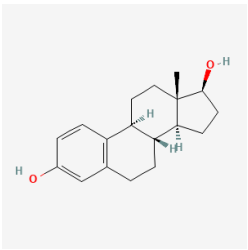
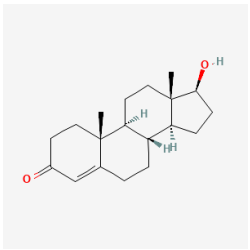
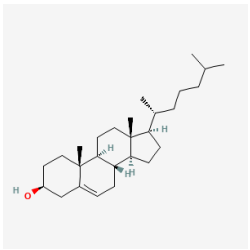
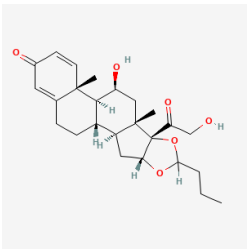
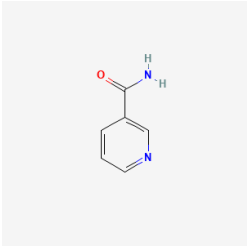
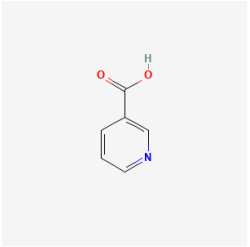
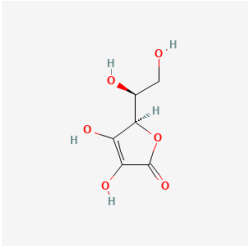
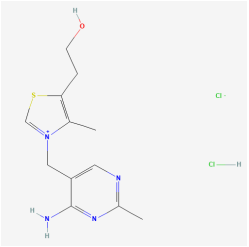
	Antiviral	Nucleosides		
Chemical structure				
Name	Acyclovir	Cytidine	Uridine	Adenosine
Label	Acy	Cyd	Uri	Ade
Solvent	DMSO	DMSO	DMSO	DMSO
	Lipids			
Chemical structure				
Name	MC3	DOPE	DSPC	
Label	MC3	DOPE	DSPC	
Solvent	MeOH	EtOH	EtOH	
	Pain relievers			
Chemical structure				
Name	Ibuprofen	Naproxen	Diclofenac acid	
Label	Ibu	Nap	Dic	
Solvent	MeOH	MeOH	MeOH	
	Steroids			
Chemical structure				
Name	Estradiol	Testosterone	Cholesterol	Budesonide
Label	Est	Tes	Cho	Bud
Solvent	MeOH	MeOH	EtOH	MeOH

Table 6: Continuation of Table 4.

	Vitamins			
Chemical structure				
Name	Nicotinamide	Nicotinic acid	Ascorbic acid	Thiamine HCl
Label	NAm	Nic	Asc	Thi
Solvent	MeOH	DMSO	MeOH	DMSO

Selected physicochemical properties of the analytes has been calculated with RDKit Open-Source Cheminformatics Software [30], using the internal web development environment, Jupyter-Hub [22], which utilises the Python programming language. All calculated properties are presented in Table 7. One of the physicochemical properties, *cLogP*, is the logarithm of n-octanol's and water's partition coefficient, $log(c_{octanol}/c_{water})$, and it is considered to be a measure of compound's hydrophilicity [31]. Presented *cLogP* values are not experimentally measured, but calculated, where different functional groups are taken into account [32]. The contribution of different functional groups is not strictly additive, which is why even individual functional groups have been considered.

Table 7: Selected physicochemical properties of analysed compounds. *Molecular weights* are taken from PubChem database [29]. *Rings* denotes number of rings within a molecule, *aliphatic -OH groups* - number of aliphatic hydroxy groups, *-H donors* - number of hydrogen bond donors, *-H acceptors* - number of hydrogen bond acceptors, NR_3 - number of tertiary amino groups, *R-*NH_2 - number of primary amino groups. *Number of rings* does not distinguish between 5- or 6-member rings or whether they have saturated or unsaturated bonds.

Label	Molecular weight [g/mol]	cLogP	Rings	Aliphatic -OH group	-H acceptor	-H donor	NR_3	R-NH_2
Ser	105.1	-1.609	0	1	3	3	0	1
Pro	115.1	-0.1770	1	0	2	2	0	0
Iso	131.2	0.4444	0	0	2	1	0	1
Leu	131.2	0.4444	0	0	2	2	0	1
Glu	147.1	-0.7369	0	0	3	3	0	1
Met	149.2	0.1514	0	0	3	2	0	1
Try	204.2	1.122	2	0	2	3	0	1
Trx	776.9	4.557	0	0	4	3	0	1
Metro	171.2	0.09202	1	1	5	1	3	0
Tri	290.3	1.258	2	0	7	2	2	2
Amp	349.4	0.02370	3	0	6	4	1	1
Amo	365.4	0.02370	3	0	6	4	1	1
PenG	372.5	-3.470	3	0	5	1	1	0
Dox	444.4	-0.5042	4	4	9	6	1	1
Tet	444.4	-0.3710	4	4	9	6	1	1
Azi	749.0	1.901	1	5	14	5	2	0
Acy	225.2	-1.332	2	1	7	3	3	1
Cyd	243.2	-2.563	2	3	8	4	2	1
Uri	244.2	-2.852	2	3	7	4	1	0
Ade	267.2	-1.980	3	3	9	4	4	1
MC3	642.1	13.65	0	0	3	0	1	0
DOPE	744.0	11.61	0	0	8	2	0	1
DSPC	790.1	12.17	0	0	8	0	1	0
Ibu	206.3	3.073	1	0	1	1	0	0
Nap	230.3	3.037	2	0	2	1	0	0
Dic	296.1	4.364	2	0	2	2	0	0
Est	272.4	3.609	4	1	2	2	0	0
Tes	288.4	3.879	4	1	2	1	0	0
Cho	386.7	7.389	4	1	1	1	0	0
Bud	430.5	2.717	5	2	6	2	0	0
NAm	122.1	0.1805	1	0	2	1	1	1
Nic	123.1	0.7798	1	0	2	1	1	0
Asc	176.1	-1.407	1	4	6	4	0	0
Thi	337.3	-1.967	2	1	5	2	3	1

3.2.3 Sample preparation

All compounds were dissolved in appropriate solvents, presented in Table 4. Stock solutions of each analyte were prepared with concentrations equal to or above 25 mM. Dilution of the

stock solutions to obtain c_1-c_5 was performed with TECAN Automated Liquid Handler (Tecan Group Ltd., Männedorf, Switzerland).

3.3 Data analysis

Empower allowed for data export in a form of the text file (.txt). The exported variables and observations as well as their categorisations are presented in Table 8.

Table 8: Empower output information categorised into variables, constants and observations. *Acq Method Set* denotes methods applied to all analysed compounds, *Injection Number* denotes first or second injection of the same concentration.

Variable				Constant	Observation
Categorical		Numeric			
Nominal	Ordinal	Continuous	Discrete		Quantitative
Compound Name Acq Method Set Modifier Make-up	Injection Number	Sample Concentration	Percentage Modifier	Additive Additive Concentration Mobile Phase Flow Rate	Area

Linearity evaluation of signal obtained with the CAD was carried out by using Python. Average values of the surface area between two injections of each concentration have been calculated. These average values were used to create signal-response curves. A signal-response curve was obtained for each method for each analyte (16 methods * 34 compounds), where the average surface area was plotted against sample concentration. A linear regression was then fitted to the signal-response curve using the *scikit-learn* module and the *linear_model* class, which enables machine learning with the use of linear models [33]. These linear regressions were then used to obtain the coefficients of determination, R^2. The R^2 coefficients were then used as the factors determining the linearity of a signal for a given method. Calculated R^2 values were also compared with the physicochemical descriptors of the analytes in order to categorise them as well as to determine the trends of the CAD signal.

Average area values were also used in MODDE, where MLR was fitted in order to create response contour plots, one for each of the analysed compounds. The validation metrics defining the fit and prediction of the model, R^2 and Q^2 [34], were then used to estimate whether the fitted model and response contour plots could be effectively used to predict influence of two or more variables on the CAD signal response.

4 Results & discussion

In this section, general characterisation of the chemical space analysed in this study will be presented. The linearity of the signal coming from CAD will be discussed in detail, considering which compounds and which methods lead to linearity. The signal strength for the classes of compounds will also be presented and discussed. In addition, with the help of physicochemical descriptors, an attempt will be made to find trends that lead to linearity of the signal. Subsequently, cases resulting in low linearity will be investigated in more detail. In the next part response contour plots will be discussed as a method to see the effect of the interaction of two variables in the CAD signal. There will also be a discussion on the solvents used to dissolve the tested chemicals in and their possible impact on the CAD signal.

4.1 Chemical space characterisation

Characterizing the compounds in the chemical space will facilitate interpretation of the observations in this study. By plotting the cLogP against the analytes' molecular weights, the diversity of the chemical space can be overviewed. The nature of the compounds (hydrophilic or hydrophobic) [30] and the relationships between different classes of compounds is presented in Figure 5.

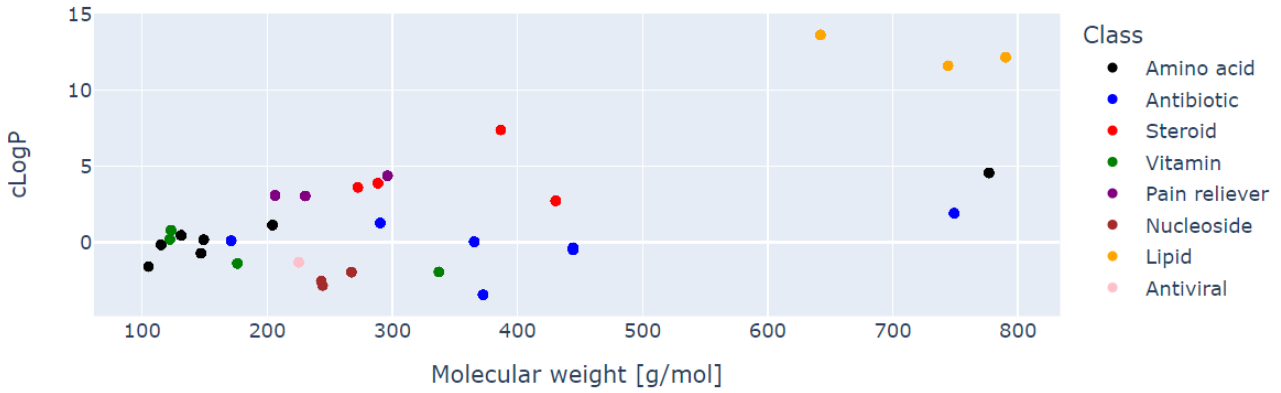


Figure 5: Scatter plot of the cLogP values of the 34 analytes making up the chemical space plotted against their molecular weights. Compounds have been assigned to their classes as presented in Table 4.

The first thing that can be seen in Figure 5 is the fact that the analytes in this study make up a diverse chemical space. A wide range of molecular weights can be observed, with larger grouping between 100-400 g/mol, where a group classified as small molecules can be identified. The analytes whose molecular weight is much higher, are primarily lipids, but also *L-thyroxine* and *azithromycin*. Considering the location on the chemical map, it can be seen that those belonging to the same class, are rather close to each other, especially amino acids (except *L-thyroxine*), steroids, vitamins, painkillers, nucleosides and lipids. Antibiotics, on the other hand, are more spread out, due to their structural diversity, as can be seen in Table 4.

Considering hydrophilicity, it can be seen in Figure 5 that a large portion of the analytes (12 out of 34) are polar (cLogP<0), while the rest are rather hydrophobic, e.g the lipids but also the steroidal analyte, *cholesterol*. This categorisation suggests that the hydrophobic group of analysed compounds will be compatible with the SFC system, where non-polar mobile phase is used. The fact that modifiers (MeOH, IPA) and basic additive, NH_3, were combined with

CO_2 in the mobile phase suggests that also hydrophilic compounds can be successfully analysed with the set-up utilised in this study.

4.2 Linearity evaluation

The area under the curve of the CAD signal was the only result evaluated for the linearity analysis. The reason why the signal intensity, i.e. the height of the peak, was not considered is due to the fact that when analysing the test compounds, *ketoprofen* presented split peak behaviour instead of a single peak, like *lidocaine* did. A comparison of the chromatograms for both test compounds is presented in Appendix C.

The reason for the split peak behaviour was not considered in detail, besides resulting in the exclusion of peak height as a response value. What was considered, however, was the way of integrating such chromatograms. Since only a static mixer was used in the set-up and no chromatography column, no chromatographic separation should be occurring. In addition, only analytes with a high purity were analysed. Therefore, the signal coming from the CAD had to reflect what was injected into the system. This fact resulted in the decision to include the area under the front peak when performing integration. This action was also applied during the integration of all CAD responses for which split peak could be observed.

4.2.1 General observations

As a first attempt to categorise the analysed compounds, a comparison of the coefficient of determination values, R^2, were made for each of them. Such comparison is presented in Figure 6.

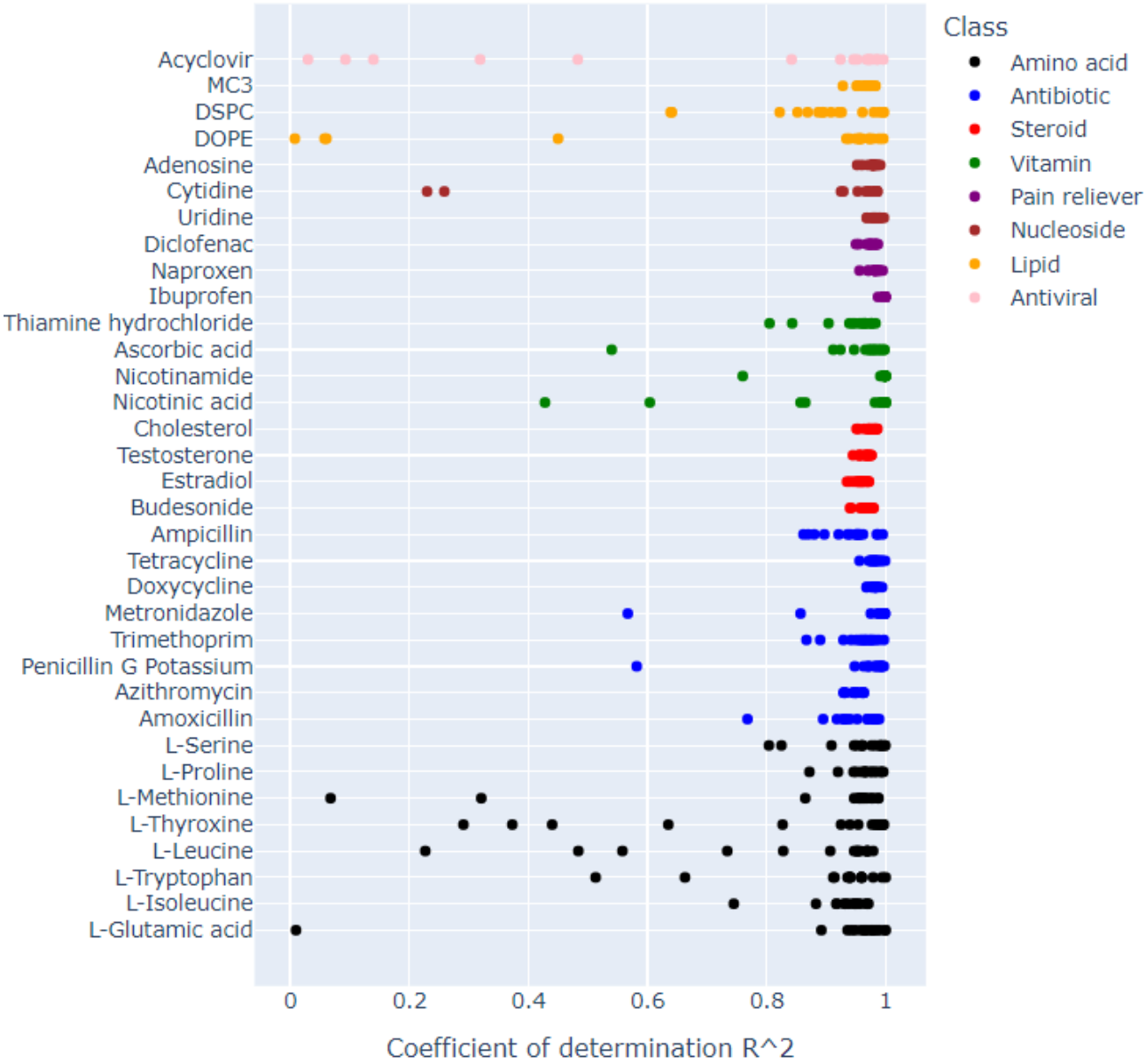


Figure 6: R^2 values plotted for each of 34 compounds. Each data point for a compound corresponds to one analytical method.

Figure 6 shows that not all methods for all of the analytes showed a linear CAD response signal (R^2>0.8). Most of the amino acids showed non-linear behaviour. Other polar analytes can also be included in this category, such as *acyclovir*, an antiviral. Among the vitamins, antibiotics, lipids and nucleosides, there are several compounds that yielded R>0.8. Compounds belonging to the groups of pain relievers and steroids showed a linear behaviour for all methods.

Considering the behaviour of the different constitutional isomers analysed, *L-isoleucine*, *L-leucine* and *doxycycline*, *tetracycline*, it can be seen from Figure 6 that the results were not always similar. While the linearity results for isomeric antibiotics were consistent, amino acids showed that minor differences in molecular structure can lead to diversity in CAD response signal.

What is also worth noting is that pain relievers and steroids, which showed linear behaviour, are hydrophobic analytes (cLogP>0). This confirms earlier hypothesis that compounds with these characteristics are more soluble in the CO_2 used in the SFC system. However, the possibility of analysing hydrophilic compounds should not be completely ruled out. Two of the analysed

nucleosides, *adenosine* and *uridine*, also showed linearity of the signal, and it should be noted that, based on clogP values, they are hydrophilic. *Penicillin G K*, having the lowest cLogP of the compounds tested, also showed linear behaviour for all but one method (MeOH_40_muACN). These results tell us that just looking at analyte clogP is not enough to predict whether a signal will be linear or not. A heatmap with the R^2 values for each method and each analyte with concentrations c_1-c_5 is presented in Figure 7.

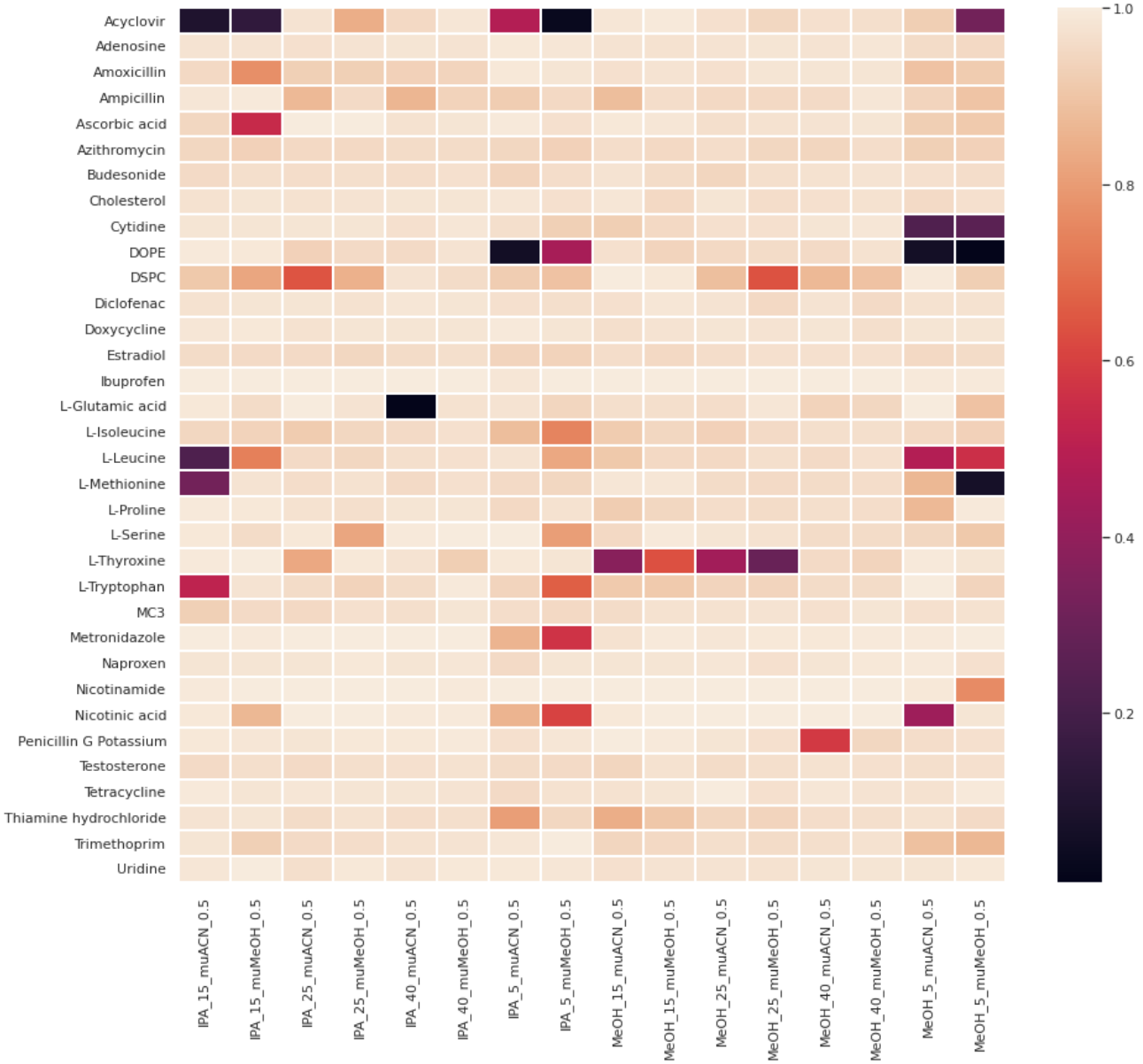


Figure 7: Heat map comparing values of R^2 for all analysed compounds as well as for all methods. Compounds are ordered alphabetically.

Focusing on the analysis of the methods used, it is possible, based on the heat map presented in Figure 7, to identify those that resulted in linearity of the signal. The methods that mainly contributed to low linearity results were those where only 5% of modifier was used. Among these methods, however, it can be distinguished that the use of MeOH as make-up resulted in more cases of lower linearity than when ACN was used as make-up. 15% IPA as modifier resulted in lower linearity more often than 15% of MeOH did. For higher percentages of modifiers, low linearity was observed in single cases. Some methods showed linearity for all compounds and these methods were IPA_25_muMeOH, MeOH_40_muMeOH and IPA_40_muMeOH.

A general trend observed was that a higher percentage of modifier correlated with linarity f the CAD signal. This proves that for most methods, where 40% of the modifier was used, the saturation limit of the CAD was not reached in the tested concentration range, c_1-c_5. This suggests that by using higher amounts of modifier, compounds with similar characteristics to those analysed in this study could also result in a linear signal of the CAD.

To better understand how strong of a signal that was detected by CAD, a comparison of the peak area and the R^2 values of all analytes was plotted and is presented in Figure 8.

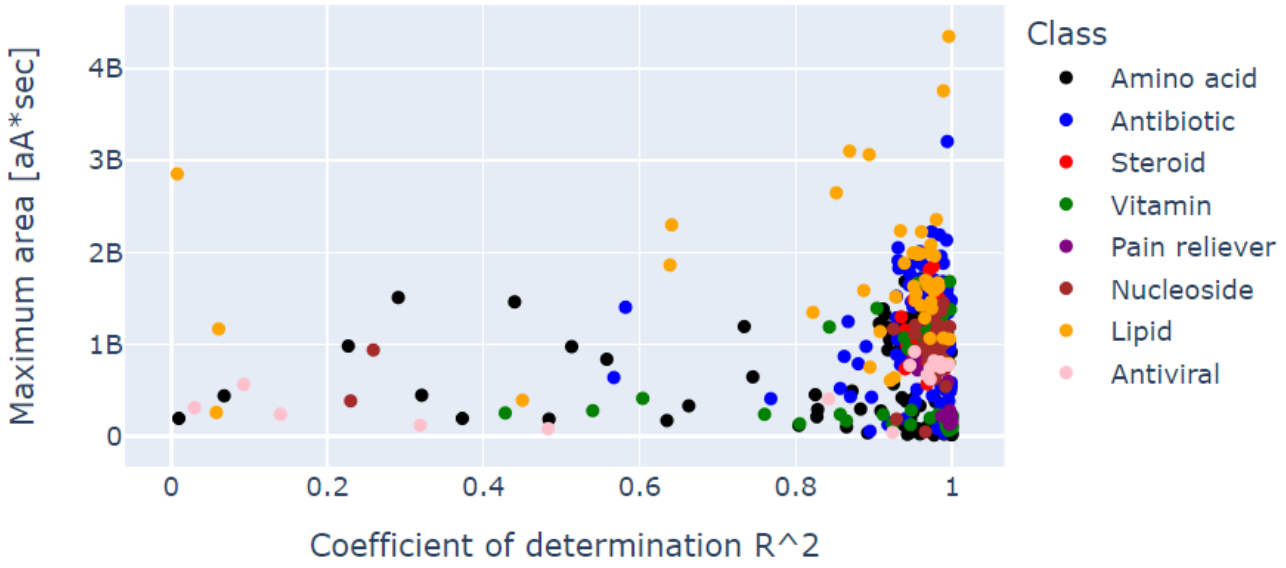


Figure 8: Maximum area of the CAD signal plotted against R^2 of each method for 34 analysed compounds.

What can be seen from Figure 8 is that the CAD signal varies from around 100s of millions to over 4 billion aA*sec (10^{-18} Ampere * seconds). The signal for amino acids, vitamins, pain relievers and the antiviral, classified as small molecules, ranged from the lowest to 45% of the highest overall value. The steroid and antibiotic groups resulted in slightly higher values, up to about 80% of the overall highest value. The largest peak areas were observed for lipids.

Given the way the CAD works and how the signal is measured, the results presented in Figure 8 were in line with expectations. Compounds in the group of 'larger compounds', i.e. with a higher molecular weight, should result in dried aerosol particles of larger areas. Larger area means that more positive charges produced in the CAD can be attached, thus allowing a larger signal to be received. Although only the largest peak area obtained for a given method was included in Figure 8 above, it should be mentioned that this doesn't mean that it corresponds to the highest concentration of an analyte. Therefore, it is difficult to draw more conclusions at this stage.

A more detailed visualisation of all peak areas is presented as a heat map in Appendix D. From there it can be found that there is no method that always resulted in high peak area, which leads to an interpretation that the CAD response peak area is highly compound specific. However, the high CAD response peak area for the lipids *DOPE* and *DSPC* suggests that they could be successfully analysed at concentrations lower than 0.5 mM and still be detected, especially using methods with a high percentage of modifier. This leads to the suggestion that the CAD response peak area is in fact influenced by the method parameters, as indicated in Figure 3.

4.2.2 Observations based on physicochemical properties

The first physicochemical descriptor analysed was cLogP, which has been plotted against the R^2 values of each method, as presented in Figure 9.

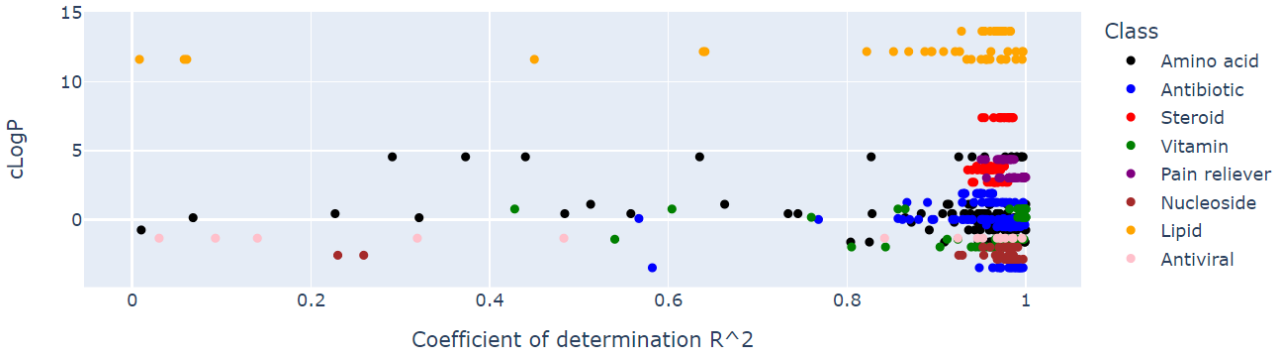


Figure 9: Plot of cLogP vs R^2 for all methods applied to 34 compounds.

The results presented in Figure 9 do not clearly explain the difference in signal linearity of the CAD for different compounds. The hydrophobicity of an analyte was not a good indicator of whether or not the results would be linear. However, this reasoning has been drawn based on the results obtained under supercritical conditions, while cLogP has been calculated under atmospheric conditions. It is therefore hard to decide whether cLogP should be used for such comparison or if another descriptor could be more appropriate.

Two structural descriptors, the number of aliphatic hydroxyl groups and the number of rings within a molecule, have been compared to the R^2 values for the tested analytes and the results are presented in Figure 10.

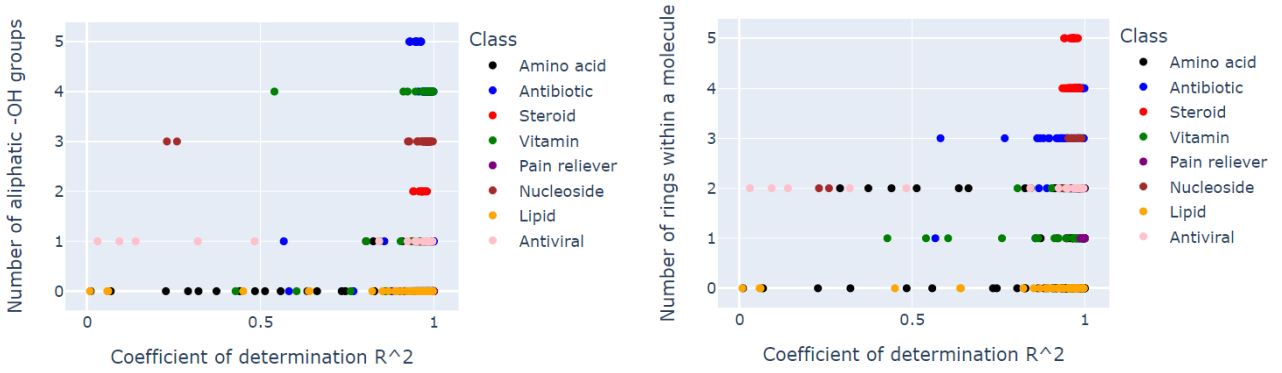


Figure 10: Plots showing number of the aliphatic -OH groups (left) and the number of rings (right) within each of the analysed molecules vs R^2 of each method.

By analysing Figure 10, general trends leading to signal linearity could be found. The presented comparison of the number of aliphatic hydroxyl groups allows to hypothesise that the greater number of these groups leads to higher linearity of the signal (higher R^2 values). Among analysed compounds, the highest number of aliphatic hydroxyl groups is attributed to the antibiotic *azithromycin*, which showed not only signal linearity but also a large signal peak area, as presented in Appendix D. To explain the observed situation, one can look for the compatibility of a large number of aliphatic hydroxyl groups with the modifiers used, MeOH and IPA. Since a large number of hydroxyl groups in a given compound makes it more soluble in highly polar solvents [35], MeOH and IPA modifiers with an elution strength higher than that of the supercritical CO_2 [28] would result in a good dissolution of a substance. In this way,

it can be indicated that a greater number of hydroxyl groups increases the chance of obtaining a linear CAD signal response, especially when using high amounts of modifier.

The number of rings in a molecule may also lead to greater linearity of the CAD signal, as can be seen in Figure 10. In this case, the analysed compounds with a large number of rings are primarily steroids, but also antibiotics. Compounds with more than three rings had a linear CAD response. A large number of rings in a given molecule might lead to its increased hydrophobic character and reduced solubility in polar solvents [36]. The reason why a large number of rings leads to a linearity of the CAD signal might be that in the study, the mobile phase had a high percentage of supercritical CO_2, which is non-polar, therefore increasing solubility of compounds with hydrophobic character.

Other structural descriptors presented in Table 7 are the number of tertiary and primary amino group in a molecule. Plots with the aforementioned descriptors compared to R^2 values are presented in Figure 11.

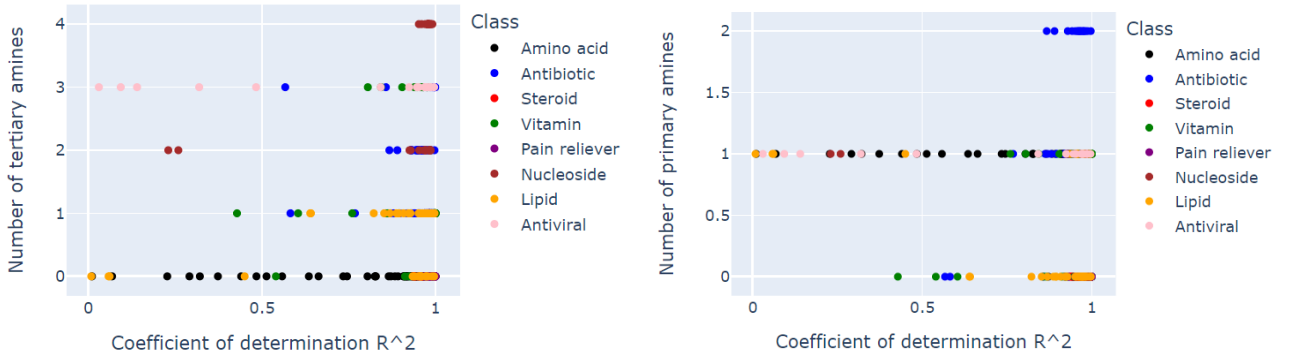


Figure 11: Plots showing number of tertiary (left) and primary (right) amines vs R^2 values.

From Figure 11 no clear trends are observed that relate the signal linearity to the number of tertiary amino groups. Tertiary amines are hydrogen bond acceptors and they can only form hydrogen bonds with a solvent in the surrounding if it has hydrogen donor ability [37]. The reason why the ability to form hydrogen bonds is so important is that they help to create stable interactions between the solvent and analyte [38]. Therefore, a large number of tertiary amino groups may be used to create many hydrogen bonds with used modifiers, which could allow the analyte to stay dissolved and not precipitate in the system and then reach the detector. It seems that the high amount of tertiary amines correlates with signal linearity but there might be also other factors that influence the signal, e.g size of the analytes. Small size of solutes might lead to more effective aerosol creation and therefore positively contribute to the signal linearity. This is the likely reason why *adenosine* (4 tertiary amino groups) and *thiamine HCl* (3 tertiary amino groups) showed high signal linearity. *Acyclovir*, also having three tertiary amino groups, however, did not show a similar behaviour, which means that the number of tertiary amines in a molecule may not be a good indicator of linearity of the CAD signal by itself.

Considering the number of primary amino groups, it can be seen in Figure 11 that the majority of the analysed compounds for which low linearity has been obtained, contain one such group. Analysed compounds without any primary amino group showed somewhat more linear results than for cases when one primary amino group was present. *Trimethoprim* has two primary amino groups, yet it presented linear behaviour. This implies that either the presence of primary amines are non-related to the signal linearity of CAD, or that they affect it negatively but other structural elements of *trimethoprim* compensate for it.

The analysis of the number of secondary amines as well as aromatic hydroxyl groups were omit-

ted in this study, because the analytes had either none or one such group and the contribution to the results could not be included in a trend.

The last two structural descriptors chosen in order to find trends in the CAD signal are the number of hydrogen donors and acceptors. The results of comparing these values with the values of R^2 are presented in Figure 12.

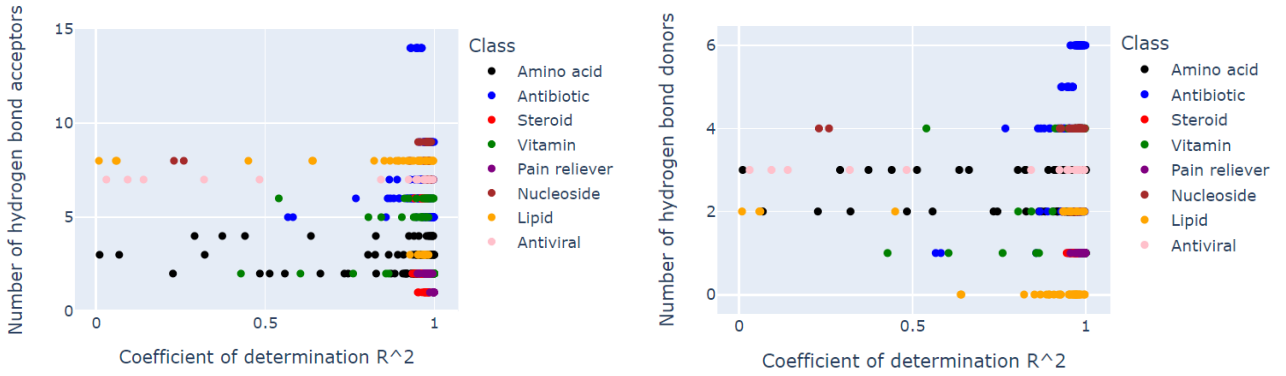


Figure 12: Plots showing the number of hydrogen bond acceptors (left) and donors (right) vs R^2 values.

In Figure 12 it can be observed that both a high number of hydrogen donors and acceptors could be related to a higher linearity of the CAD signal. Analytes with a high number of hydrogen donors and acceptors and which also showed linearity of the signal are some of the antibiotics, *azithromycin*, *doxycycline*, *tetracycline*, and the nucleoside, *adenosine*. This might be attributed to the analytes' abilities to form hydrogen bonds with the solvent. Thus, it can be expected that due to these interactions, analytes can remain dissolved during analysis when including a modifier with the supercritical CO_2. Keeping analytes dissolved and avoiding precipitation in the system is important for the linearity sake. Precipitation is more likely to occur for higher concentrations of analyte, as there is less solvent that could positively contribute to the analyte solubilisation by creation of hydrogen bonds. Precipitated analyte will not reach the detector leading to decreased detected amount and thus, a non-linear signal response being observed.

A small number of hydrogen donors and acceptors can lead to linearity of the CAD signal, as is the case for steroids and pain relievers. Looking for an explanation for the observed situation, one can come to the previously discussed observation, which is the fact that both groups exhibit a hydrophobic character (cLogP>0), and therefore are compatible with the mobile phase of the SFC. This compatibility ensures that they remain dissolved and do not separate from the solvent leading to precipitation in the system during their analysis.

4.2.3 Investigation of low linearity

Linearity results of R^2<0.8 was observed for 16 out of the 34 analysed compounds. Thirteen out of 16 methods were associated with lower linearity of these 16 analytes. Appendix E presents the CAD chromatograms obtained for all cases where R^2<0.8. In Figure 13, signal-response curves for each compound and method that contributed to the non-linearity of the CAD signal were presented.

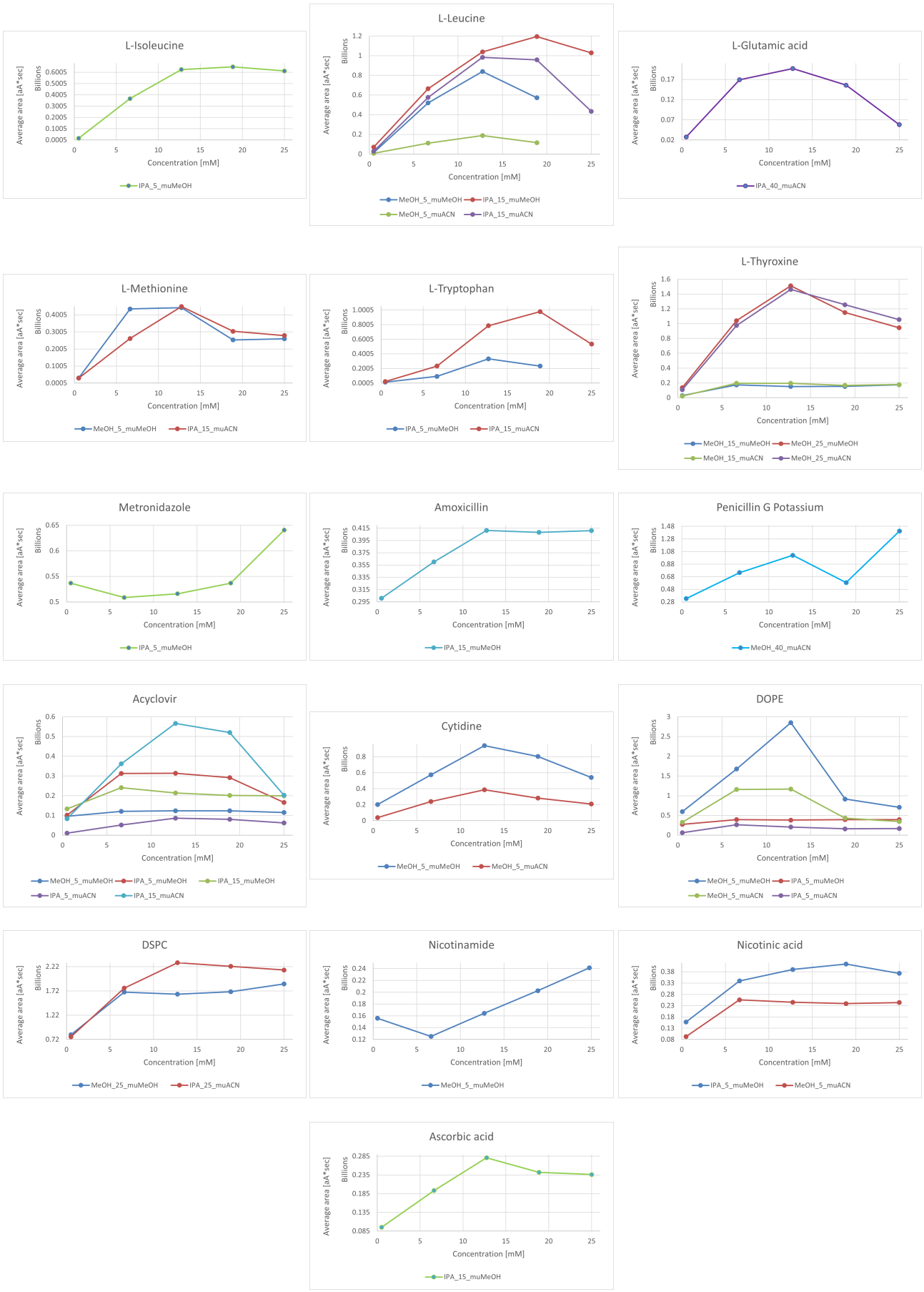


Figure 13: Signal-response curves for compounds where applied methods resulted in R^2<0.8.

By first analysing the shapes of the curves seen in Figure 13, three types can be distinguished: *concave-increasing*, where the area increases with increasing concentration of the substance and then reaches an almost diminishing plateau; *concave-decreasing*, where the area first increases with increasing sample concentration, and then decreases after reaching a critical concentration; *convex*, where the area first decreases and then increases with increasing analyte concentration. Among the curves presented, it can also be seen that the plot for *penicillin G K* contains a potential outlier, namely the resulting peak area for concentration c_4. Analysing the corresponding chromatogram presented in Table 22, it can be seen that the low peak area must have been the result of inaccurate integration of the CAD signal. Therefore, the previously mentioned concentration was considered an outlier, and its removal caused the MeOH_40_muACN method to reach the value R^2=0.965.

Analytes with concave-increasing behaviour were *L-isoleucine*, *amoxicillin*, *DSPC*, and *nicotinic acid* which all belong to different classes. This appearance of the signal-response curves is typical for CAD and is associated with reaching the limit of saturation in the detector. The linear signal is detected until it reaches a critical concentration, after which it stops at a plateau or increases slightly. The reason for this behaviour could be traced to CAD's ability to form aerosols. Higher analyte concentrations can cause larger aerosol droplets to be formed. Larger aerosol droplets can condense and then be removed from the system, thus not being detectable.

Compounds that have been observed to exhibit concave-decreasing behaviour include *L-leucine*, *L-glutamic acid*, *L-methionine*, *L-tryptophan*, *L-thyrosine*, *acyclovir*, *cytidine*, *DOPE*, *ascorbic acid*. Among them, many are in the amino acid group and single ones from other classes with various structural characteristics. The observed shape of the signal-response curve shows that the amount of compound detected by CAD, after exceeding a certain critical concentration, drops dramatically. The reason may be similar to the one explained for concave-increasing behaviour. Considering, however, the methods that contributed to achieving this behaviour, it can be seen that these are mainly methods where a low percentage of modifier was used. This may have caused polar compounds to precipitate in the system and therefore less production of aerosol. Alternatively, they could form larger agglomerations and thus contribute to the aerosol with larger droplets. This reasoning can be applied to *DOPE*, which is prone to create micelles in the case of insufficient polar solvent [15] as the interactions between hydrophilic heads of *DOPE* and the supercritical CO_2 are to be minimised [39].

Convex behaviour has been observed for *metronidazole* and *nicotinamide*. In this case, the reasons behind this behaviour are only speculation, as no actions were taken in order to investigate it in more detail. The lowest concentration for both compounds could not be identified as outliers, as the corresponding chromatograms, showed in Appendix E, do not present the signal being distinctive. A possible explanation for this observation could be different interactions, e.g between the analytes molecules. At the lowest concentration, c_1, these interactions might have been weaker than at higher concentrations, which could have lead to either more effective aerosol creation or smaller aerosol droplets. This would then result in a larger peak area. An alternative explanation is the possibility of contaminants that have been detected in the CAD for this concentration. These contaminants could be possible remnants of analyte that had precipitated in the system and were eluted in subsequent injections.

A complete list detailing analytes, methods, exact R^2<0.8 values and concentration span that was analysed, is presented in Table 30 in Appendix F. The reason why there is also a concentration range in this table is because some of the analytes have not been tested in the entire 0.5-25 mM range. The incomplete range is a consequence of precipitation of a given compound during the analysis, which caused clogging of the system and it stopping due to over-pressure. For a complete list of missing data, see Appendix G. Table 30 includes manually adjusted concentration ranges resulting in R^2>0.8. This was an attempt in order to determine the linearity

limits of the CAD signal.

4.2.4 Summary of linearity

A summary of the linearity of the CAD signal is shown in Table 9. As can be seen in Table 9, there are some methods for which the analysis of the entire concentration span, c_1-c_5 was not possible. As mentioned earlier, it was mainly the methods with a low modifier percentage that showed lower linearity of the signal. The same methods contributed to the incomplete analysis of the c_1-c_5 concentration. This is especially true for amino acids, but also for antibiotics (*metronidazole*, *amoxicillin*), the antiviral, nucleosides (*cytidine*), and vitamins. Groups of compounds that turned out to be compatible with all the methods used are pain relievers and steroids, the structural characteristics of which have been discussed earlier.

What is also worth noting in Table 9 is the fact that for some compounds the linear signal could not be achieved at all or it was only related to the span between three concentration points. Looking more closely at the methods that led to this behaviour, one can point out those that have the same factor in them. Examples of such methods are MeOH_15_muMeOH-MeOH_15_muACN and MeOH_25_muMeOH-MeOH_25_muACN used for the analysis of *L-thyroxine*. In this case, it was the 15% and 25% MeOH modifier that led to the aforementioned inability to determine the linearity. When analysing the chromatograms corresponding to these methods, presented in Table 19, pak broadening and tailing can be observed. This is probably due to *L-thyroxine* being insoluble in lower amounts of MeOH modifier. Similar trends can also be found for other analytes presented in Table 9.

Table 9: Ranges of linear behaviour of analysed compounds for all investigated methods. Legend: '√' corresponds to linear behaviour (R^2>0.8) across a concentration span of 0.5-25 [mM]; displayed concentration value corresponds to highest analysed concentration with linear behaviour; displayed concentration value with '*' describes the adjusted value where linear behaviour has been reached; 'x' corresponds to situations where linearity could not be achieved for more than two points. Crossed out concentration value corresponds to identified outlier.

	MeOH_5_muMeOH	MeOH_15_muMeOH	MeOH_25_muMeOH	MeOH_40_muMeOH	IPA_5_muMeOH	IPA_15_muMeOH	IPA_25_muMeOH	IPA_40_muMeOH	MeOH_5_muACN	MeOH_15_muACN	MeOH_25_muACN	MeOH_40_muACN	IPA_5_muACN	IPA_15_muACN	IPA_25_muACN	IPA_40_muACN
Ser	c_4	√	√	√	c_4	√	√	√	c_4	√	√	√	c_4	√	√	√
Pro	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Iso	√	√	√	√	c_4*	√	√	√	√	√	√	√	√	√	√	√
Leu	c_3*	√	√	√	c_3	c_4*	√	√	c_3*	√	√	√	c_3	c_4*	√	√
Glu	√	√	√	√	c_3	c_3	c_3	√	√	√	√	√	c_3	c_3	c_3	c_3*
Met	x*	√	√	√	c_4	√	√	√	√	√	√	√	c_4	c_4*	√	√
Try	x	√	√	√	c_3*	√	√	√	x	√	√	√	c_4	c_4*	√	√
Trx	c_4	x*	c_3*	√	c_4	c_4	√	√	c_4	x*	c_3*	√	c_4	c_4	√	√
Metro	√	√	√	√	c_2-c_4*	√	√	√	√	√	√	√	√	√	√	√
Tri	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Amp	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Amo	√	√	√	√	√	c_4*	√	√	√	√	√	√	√	√	√	√
PenG	√	√	√	√	√	√	√	√	√	√	√	~~c_3~~	√	√	√	√
Dox	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Tet	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Azi	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Acy	c_3*	√	√	√	x*	x*	√	√	c_4	√	√	√	c_4*	c_4*	√	√
Cyd	c_3*	√	√	√	√	√	√	√	c_3*	√	√	√	√	√	√	√
Uri	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Ade	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
MC3	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
DOPE	c_3*	√	√	√	x*	√	√	√	x*	√	√	√	x*	√	√	√
DSPC	√	√	x*	√	√	√	√	√	√	√	√	√	√	√	c_4*	√
Ibu	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Nap	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Dic	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Est	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Tes	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Cho	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Bud	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
NAm	c_2-c_5*	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Nic	√	√	√	√	c_4*	√	√	√	x*	√	√	√	√	√	√	√
Asc	√	√	√	√	√	c_3*	√	√	√	√	√	√	√	√	√	√
Thi	√	√	√	√	c_4	√	√	√	√	√	√	√	c_4	√	√	√

4.3 Interaction between variables

The methods used to predict changes in linearity and signal peak area of CAD are 4D (3D for four conditions) response contour of area plots created in MODDE, to which an MLR model has been fitted. These charts compare the choice of modifier and make-up, amount of modifier and analyte concentration all at once, while also showing predicted responses [23]. One 4D response contour plot was created for each compound, in which it was possible to compare all methods used in this study simultaneously. At this point, however, it should be noted that none of the acquired peak area data was removed when creating 4D response contour plots. An example of observed behaviours for various compounds is illustrated in Figure 14.

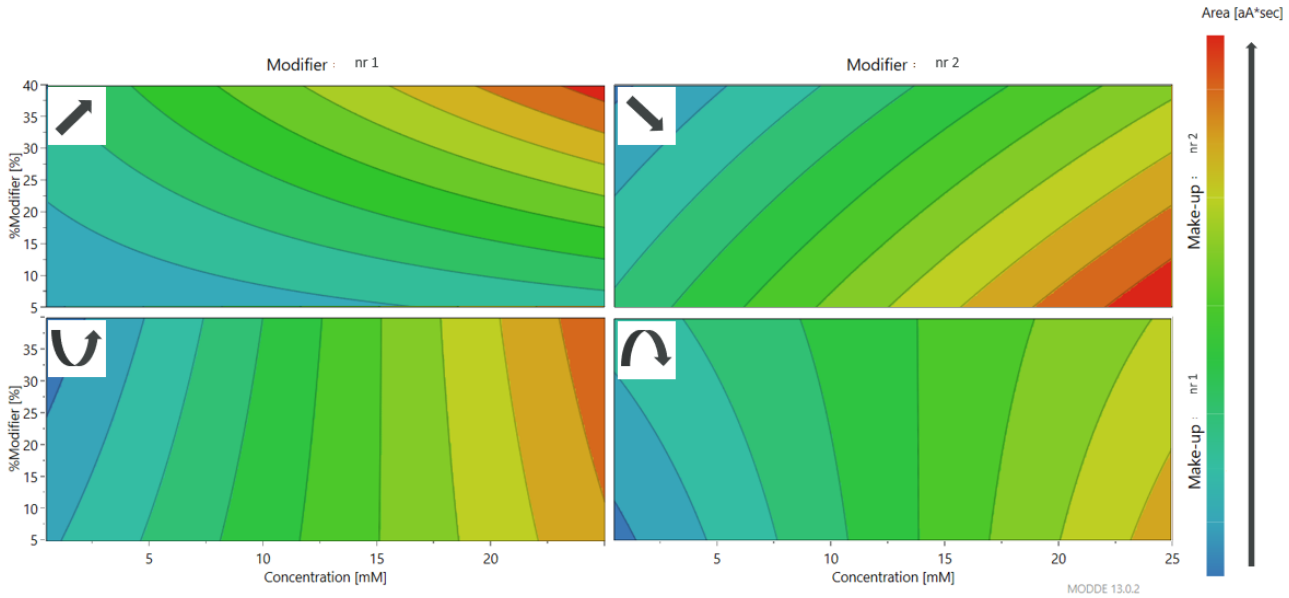


Figure 14: An example of observed behaviour of 4D response contour of area plots performed with MODDE. The x-axis shows both solute concentration and modifier used. The y-axis shows the percentage of modifier as well as the make-up solvent used. The colors show the magnitude of peak area of the CAD signal. Arrows: Pointing up - area is increasing with increasing percentage modifier; Pointing down - area is decreasing with increasing percentage modifier; Twisted pointing up - area is first decreasing and then increasing; Twisted pointing down - area is first increasing and then decreasing. Modifier nr 1/ nr 2 can be either MeOH or IPA; make-up nr 1/ nr 2 can be either MeOH or ACN.

Since different behaviours were observed for different compounds, it was also interesting to see what trends could be observed for different methods. For this purpose, pie-charts were created in which the analysed compounds were compared with each other, categorising them according to their increasing or decreasing peak area with the increasing percentage of a modifier. These diagrams are presented in Figure 15.

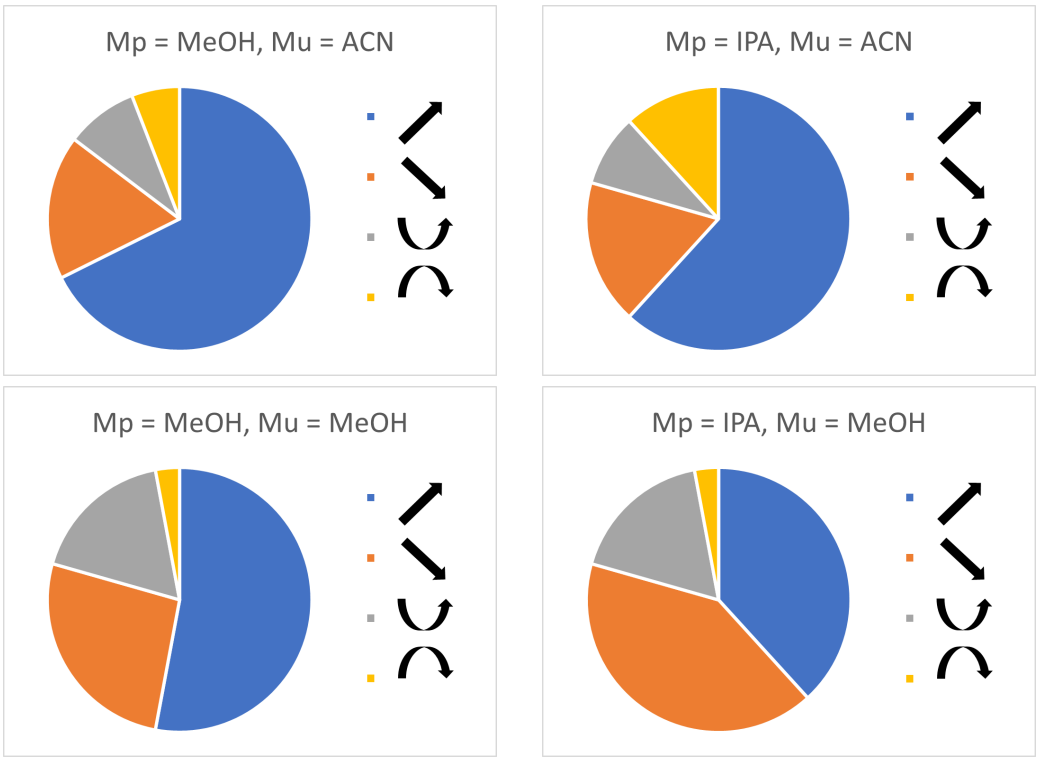


Figure 15: Pie charts showing trends in peak area behaviour with increasing percentage modifier. The charts have been created based on the response contour plots of the peak area analysis performed with MODDE.

Analysing Figure 15, it can be seen that the largest number of analysed compounds, in which the response peak area increases with the increasing percentage modifier, is observed for methods in which the combination of MeOH (modifier) and ACN (make-up) was used. The smallest number showing this behaviour was recorded for the combination of IPA (modifier) and MeOH (make-up). The smallest number of analysed compounds for which a decreasing behaviour was observed, corresponds to the combination of MeOH (modifier) and ACN (make-up), while the highest to IPA (modifier) and MeOH (make-up).

What is also noticeable in Figure 15 is that a large number of compounds exhibiting the behaviour of the peak area first decreasing and then increasing with increasing percentage modifier, have been recorded when MeOH is used as make-up. This behaviour counts as well as an increasing behaviour. When ACN has been used as make-up, a greater number of compounds with peak area first increasing and then decreasing with the amount of used modifier, could be observed, and these can be classified as decreasing behaviour.

With the aim of realising which of the analytes contributed to these observations, a comparison was created detailing the increasing, decreasing and mixed behaviour of the response peak area from CAD across increasing percentage modifier. This is presented in Table 10.

Table 10: Table with observed behaviours from MODDE 4D response contour of area plots fitted with MLR model.

Behaviour		
Increasing	Decreasing	Mixed
L-Serine	Trimethoprim	L-Isoleucine
L-Proline	Doxycycline	Metronidazole
L-Leucine	Tetracycline	Azithromycin
L-Glutamic acid	Uridine	Naproxen
L-Methionine	Adenosine	Diclofenac acid
L-Tryptophan	Ibuprofen	Estradiol
L-Thyroxine	Nicotinic acid	Nicotinamide
Ampicillin		Thiamine HCl
Amoxicillin		
Penicillin G K		
Acyclovir		
Cytidine		
MC3		
DOPE		
DSPC		
Testosterone		
Cholesterol		
Budesonide		
Ascorbic acid		

The first observation from Table 10 is the fact that the majority of the analytes show an increase in peak area with both compound concentration and percentage modifier. The number for which the response peak area decreased with increasing percentage modifier is small and comparable to mixed behaviour. Almost all analytes that presented increasing behaviour were amino acids, lipids and steroids. Antibiotics, nucleosides and vitamins have shown that their behaviour varies. *Ampicillin*, *amoxicillin* and *penicillin G K* - antibiotics with similar molecular structures, exhibited the same increasing behaviour. The isomers *doxycycline* and *tetracycline* belong to this group as well. This, however, is not the case for other analysed isomers, which are *L-leucine* and *L-isoleucine*, which belong to the increasing and mixed behaviour groups, respectively. This comparison does not specify any of the modifiers or make-up, and only shows general trends. A detailed summary of all observed behaviours for each compound is presented in Appendix H. Obtained 4D response contour of area plots where mixed behaviour was observed are presented in Appendix I.

4.3.1 Evaluation of MLR model fitting

Taking a closer look at the values defining the fit of the MLR model, R^2 and Q^2, might help to better understand the results and evaluate whether the presented grouping is indeed reliable. A summary of both these values for all analytes is presented in Figure 16.

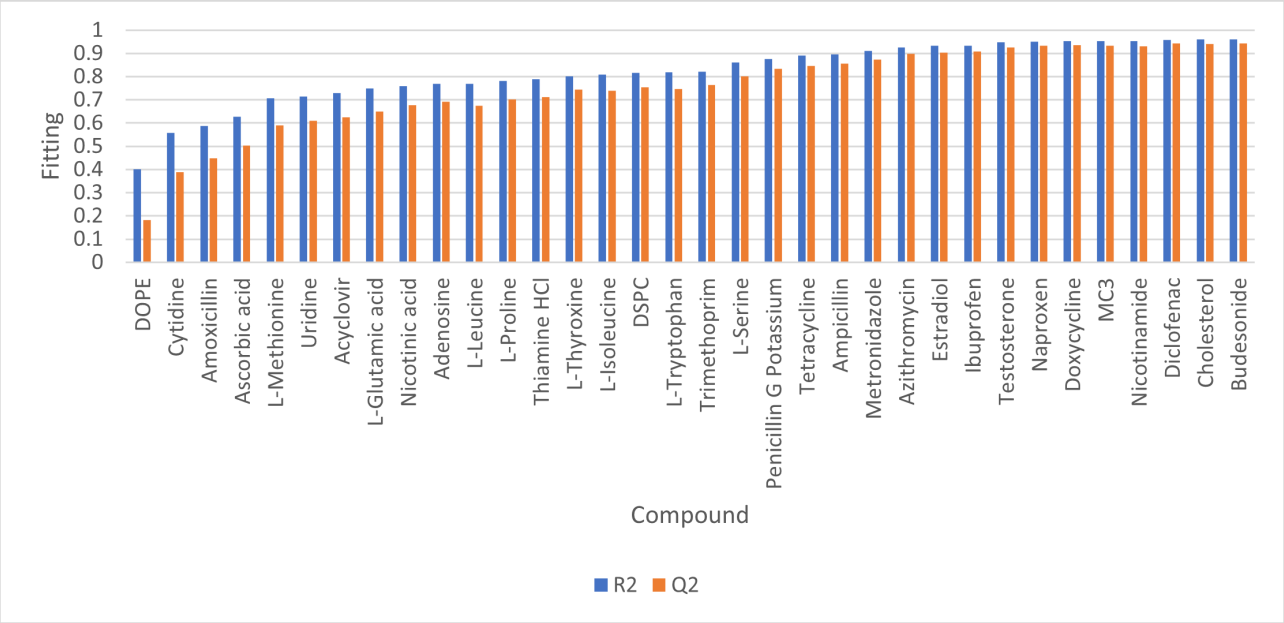


Figure 16: Summary of the fits for MLR model obtained with MODDE.

As can be seen from Figure 16, not all compounds had data that could be modeled with high R^2 and Q^2. Reconsidering that a value of 0.8 can be considered as the limit of good model fit and good model predictability, it can be concluded that predicting combined influence of variables on signal peak area for most compounds may prove challenging. However, the fact that the MLR model was fitted to all data collected during the SFC-CAD experiments, and to a partially incomplete data set, it is understandable why the fit is so low. Since the MLR fits many linear regressions to the input data, the previously discussed non-linearity seen for some compounds will also be seen in the fit of this model.

Something that could help to get a better fit and predictability of the MLR models for different compounds is to exclude data considered as outliers. This could be done only after deeper data analysis where these outliers could be identified. However, omitting a larger amount of data collected where non-linearity was observed could result in false conclusions being drawn. Disregarding the results for high concentrations of *DOPE*, for example, implies that the response peak area would increase as the percentage modifier increases, which, as previously discussed, would not be true. Therefore, it was decided not to exclude any of the acquired data, and the results obtained for each of the analysed compounds are classified as reliable and could be used when predicting joint-influence of more than one variables on the CAD signal response.

4.4 Solvent contribution

Something that has not yet been considered is the possible contribution from the solvents used to dissolve the analytes. As can be seen in Table 4, different solvents were used, mainly MeOH, EtOH and DMSO. A possible signal contribution from these solvents has not been subtracted from the total signal used in the results presented and discussed above. However, in order to check what impact they might have had on the CAD response peak area, it was decided to analyse them as well. For this purpose, blank injections of the above-mentioned solvents were made ten times for each of the 16 methods used in this study. The obtained CAD response peak areas are presented in Figure 17.

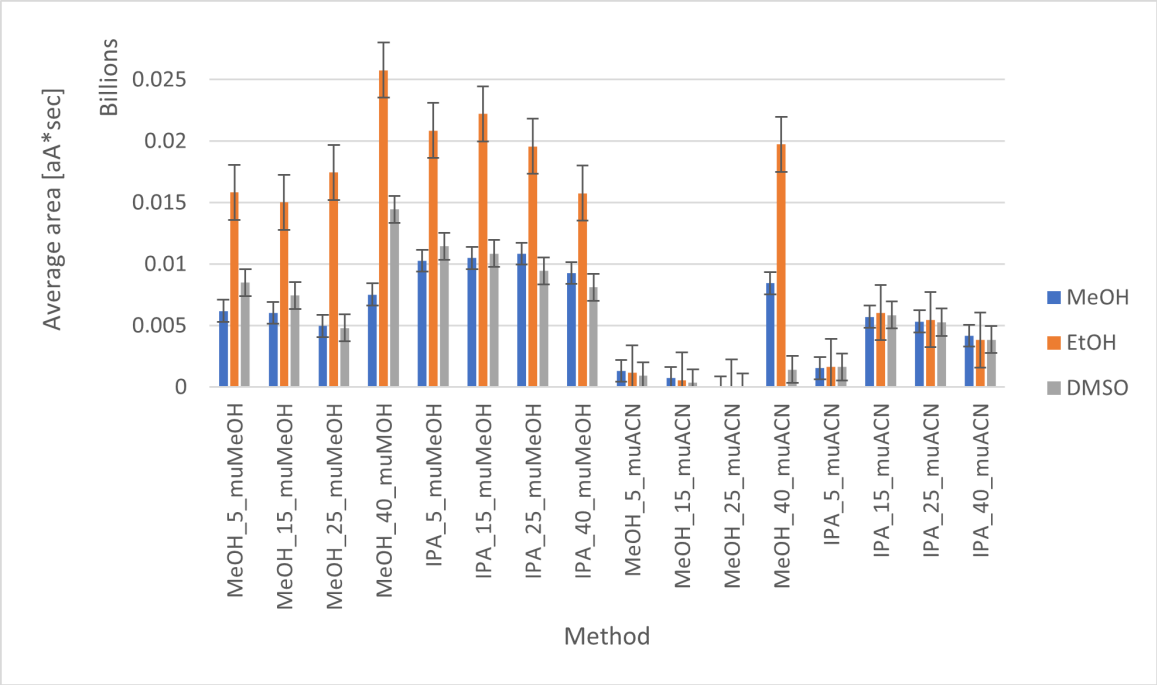


Figure 17: Comparison of CAD signal average area for solvents used in the study with displayed error bars. These error bars should be discarded for the method MeOH_25_muACN as no signal peak could be distinguished for integration.

As can be seen from Figure 17 the solvents used do not exhibit a completely zero signal in CAD, except for one method. The reason they are still detectable is that in the case of MeOH or EtOH, they most probably do not evaporate completely in the conditions used in CAD evaporation tube. The contribution of DMSO, having a high evaporation temperature, can be attributed most probably to the various sizes of aerosols created out of it, where only the smaller ones continue their way to the electrometer. The contribution of MeOH and DMSO is remarkably comparable, while that of EtOH exceeds them for most methods.

However, taking into account that the CAD signal peak areas for the analysed solvents ranges from undetectable to 100 millions, while the signal for the analysed compounds was much larger, it can be assumed that the contribution of solvents does not strongly affect the obtained CAD response peak area. It should also be noted that it is not known whether the CAD signal recorded for the sample blank would be comparable with the signal for the same amount of solvent in which a substance was dissolved. In this study, however, the difference between the CAD signal for the same analyte in two different solvents was not examined in order to verify whether the signal would be identical or different.

Another reason why the contribution from the solvent used to dissolve the substance was not subtracted is the fact that each time a given substance is in a crystal form, it must be dissolved in order to be analysed in liquid chromatography. Moreover, excluding the solvent contribution from the obtained CAD results would not affect the coefficient of determination, R^2, because a constant value for each method would be subtracted, resulting in the same spatial position in calibration curves. An argument against the above reasoning is the fact that not subtracting the contribution of solvents may contribute to an increased percentage error of the obtained results.

4.5 Other sources of error

Since the process that led to the presentation and categorisation of the results consisted of many steps, possible sources of error could be found in many of them. The automatic integration made in Empower, although it was thoroughly verified, could be one of the sources of errors. This was verified for *penicillin G K*, for which the MeOH_40_muACN method resulted in lower linearity, while the chromatogram for this specific case showed otherwise. Another source of error can be found in the way the compounds were tested, namely the injection sequence. At the very beginning, especially during the analysis of antibiotics, sample blanks with pure solvent were not used in order to eliminate the risk for carry-over effects. In this way, this could affect the signal recorded for subsequent injections. Another source of error could be the fact that the CAD signal for the blank solvent samples was not subtracted before the analysis and categorisation of the results for the 34 analysed compounds.

Something that should also be mentioned as a possible source of error is the dilution of analysed compounds to desired concentrations performed with TECAN. In this instrument, the liquid that cleans and rinses the needles before and after dilution is DMSO, which could, although in a rather small amount, contaminate the samples. Even though data obtained with UV/Vis and MS was not evaluated in this study, it was used as a helping tool in verification whether an analyte was contaminated and the obtained CAD signal was in fact reliable. For some of the analytes, the signal *157 (m/z)* could be seen in MS chromatograms. This signal corresponds most probably to the ion $[2*\text{DMSO}+\text{H}]^+$. This contribution would certainly have an impact on the results of compounds whose CAD responses turned out to be low.

5 Conclusions and future work

This part will summarise all findings presented in this study and well as possible sources of error. Improvements as well as things that could be further evaluated in the future will also be included in this section.

The aim of this study was to investigate the linearity of the signal coming from CAD coupled to SFC for a wide range of compounds using various analytical conditions. To evaluate the linearity, the coefficient of determination, R^2, was used. The results found could help to identify the analytical conditions that are most conductive in obtaining linearity of CAD signal. Combining the signal linearity with the physicochemical characteristics of a given molecule, would allow for a more effective use of CAD. Due to the large number of analytical conditions varied, DOE was used, where results of the test compounds, *ketoprofen* and *lidocaine* were analysed and allowed for exclusion of some of the parameters.

Wide range of compounds analysed in this study, allowed to find suggestions for physicochemical features of a given molecule that might have contributed to a linear CAD signal. These features were a large number of aliphatic hydroxyl groups, a large number of rings, but also either high or low number of hydrogen bond donors and acceptors. The reason why these features turned out to be significant is that they contribute to better solubilisation of the solutes in both CO_2 but also in the organic co-solvents, MeOH/NH_3 and IPA/NH_3.

Since CO_2 was used as the main mobile phase, hydrophobic analytes from the groups of pain relievers and steroids showed high linearity for each of the methods used, due to their compatibility with the supercritical CO_2. Phospholipids with the highest cLogP values, however, resulted in signal linearity for most methods. *DOPE* showed linearity only for methods with higher than 5% of modifier, while *DSPC* for all methods except two. Almost all analysed antibiotics also showed linear CAD signal, which is assigned to the presence of modifier and additive. The nucleosides and vitamins analysed in this study showed linearity of the signal for all methods, except for one, where the percentage modifier was equal to or higher than 15%. The presence of the modifier and additive did not seem to greatly facilitate amino acid and antiviral analysis, which for the highest and only a few lower percentage modifiers resulted in a linear signal (*L-glutamic acid* being an exception). Four methods, mostly with high amounts of modifier, allowed for obtaining a linear signal for the entire chemical space. A comparison of lower percentage modifiers showed that the use of IPA, with a lower elution strength than MeOH, gave non-linear results. However, comparing the make-up solvents used and their impact on the CAD response area, they were found similar in the conditions used.

By matching the MLR model to the obtained CAD data, it was possible to create a 4D surface contour of peak area plots, which would serve as a prediction of the behaviour of combined influence of variables on CAD signal peak area. Fitted MLR models for each of the analysed compounds showed that a diverse response can be expected when it comes to the obtained CAD signal peak area itself. Here, however, it was more difficult to find any clear trends apart from one concerning lipids, which was manifested by the growing CAD area response along with the increasing percentage modifier and analyte concentration. A similar trend can be observed for almost all amino acids analysed, where *L-isoleucine* was the one that exhibited mixed behaviour.

Analysis of the solvent contribution showed that used in this study solvents do not strongly affect the magnitude of compounds' peak areas. The highest solvent peak area of CAD signal was found for EtOH, while MeOH and DMSO showed to be smaller than for EtOH and similar

to each other.

As an extension of the knowledge presented in this study and as plans for the future, several things could be suggested. The first is the possibility of using a acidic additive, for example, formic acid (FA). Due to time constraints of the study, only the one additive was used and it would be interesting to see if better compatibility of acidic compounds could be observed when acidic additive is present. In addition, the possibility of a non-neutral make-up, for example MeOH:FA, can be considered and assessed whether it would affect the received CAD signal. Something that could also be explored more extensively is EtOH, which was excluded at the very beginning of this study, as a modifier and make-up. Perhaps the analysis of lipids such as *DOPE* or *DSPC* would be more successful when using EtOH. Considering the analysis of such a large number of data, one could also be tempted to use supervised machine learning, which is a combination of machine learning and artificial intelligence. In this way, an application could be created that, based on the data collected in this study, would be able to select those methods that would lead to a linear CAD signal for other compounds with properties similar to those analysed. In addition to linearity, CAD signal response area and those methods that allowed to obtain a high value of it could also be taken into account. What could also be tested in the future is the behaviour of the tested compounds when using column chromatography and comparing whether the CAD results are identical or if any correction factor could be identified. Another improvement could be to compare the signal from the CAD to that from ELSD, which is also a type of detector in which detection depends on aerosol formation. Comparison of the signal could be also done with UV and MS, which responses are better known.

References

[1] N. R. Council *et al.*, "Sharing publication-related data and materials: Responsibilities of authorship in the life sciences," 2003.

[2] "What role does analytical chemistry play in medicine?." https://www.azolifesciences.com/article/What-Role-does-Analytical-Chemistry-Play-in-Medicine.aspx.

[3] P. Worsfold, A. Townshend, C. F. Poole, and M. Miró, *Encyclopedia of analytical science*. Elsevier, 2019.

[4] G. J. Langley, S. Cancho-Gonzalez, and J. M. Herniman, "Different detectors used with sfc," in *Separation Science and Technology*, vol. 14, pp. 299–324, Elsevier, 2022.

[5] M. Swartz, "Hplc detectors: a brief review," *Journal of Liquid Chromatography & Related Technologies*, vol. 33, no. 9-12, pp. 1130–1150, 2010.

[6] T. Vehovec and A. Obreza, "Review of operating principle and applications of the charged aerosol detector," *Journal of Chromatography A*, vol. 1217, no. 10, pp. 1549–1556, 2010.

[7] B. Borek, T. Gajda, A. Golebiowski, and R. Blaszczyk, "Boronic acid-based arginase inhibitors in cancer immunotherapy," *Bioorganic & Medicinal Chemistry*, vol. 28, no. 18, p. 115658, 2020.

[8] B. C. Sousa, M. J. Wakelam, and A. F. Lopez-Clavijo, "Chapter 2 - methods of lipid analysis," in *Biochemistry of Lipids, Lipoproteins and Membranes (Seventh Edition)* (N. D. Ridgway and R. S. McLeod, eds.), pp. 53–83, Elsevier, seventh edition ed., 2021.

[9] D. Papp, T. Rukkijakan, D. Lebedeva, T. Nylander, M. Sandahl, J. S. Samec, and C. Turner, "Single-standard quantification strategy for lignin dimers by supercritical fluid chromatography with charged aerosol detection," *Analytical Chemistry*, 2023.

[10] T. F. Scientific, "Charged aerosol detection in hplc and uhplc analysis." https://www.thermofisher.com/se/en/home/industrial/chromatography/chromatography-learning-center/liquid-chromatography-information/hplc-system-components/how-hplc-detectors-work/charged-aerosol-detection.html.

[11] S. Almeling, D. Ilko, and U. Holzgrabe, "Charged aerosol detection in pharmaceutical analysis," *Journal of pharmaceutical and biomedical analysis*, vol. 69, pp. 50–63, 2012.

[12] D. Thomas, B. Bailey, M. Plante, and I. Acworth, "Charged aerosol detection and evaporative light scattering detection–fundamental differences affecting analytical performance. poster presentation at hplc."

[13] "Our company." https://www.astrazeneca.com/our-company.html.

[14] H. Yuan, S. V. Olesik, and C. West, "Supercritical fluid chromatography," *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation*, pp. 1–19, 2006.

[15] S. Marrink and A. Mark, "Molecular view of hexagonal phase formation in phospholipid membranes," *Biophysical journal*, vol. 87, no. 6, p. 3894–3900.

[16] "Phase diagram of carbon dioxide (co2)." https://www.chemistrylearner.com/co2-phase-diagram.html.

[17] L. Chen, B. Dean, and X. Liang, "A technical overview of supercritical fluid chromatography-mass spectrometry (sfc-ms) and its recent applications in pharmaceutical research and development," *Drug Discovery Today: Technologies*, vol. 40, pp. 69–75, 2021.

[18] Q. Zhang, B. Bailey, D. Thomas, M. Plante, and I. Acworth, "Determination of polymerized triglycerides by high pressure liquid chromatography and corona veo charged aerosol detector."

[19] C. Brunelli, T. Gorecki, Y. Zhao, and P. Sandra, "Corona-charged aerosol detection in supercritical fluid chromatography for pharmaceutical analysis," *Analytical chemistry*, vol. 79, no. 6, pp. 2472–2482, 2007.

[20] P. H. Gamache, "Charged aerosol detection for liquid chromatography and related separation techniques," 2017.

[21] V. Desfontaine, D. Guillarme, E. Francotte, and L. Nováková, "Supercritical fluid chromatography in pharmaceutical analysis," *Journal of pharmaceutical and biomedical analysis*, vol. 113, pp. 56–71, 2015.

[22] F. Pérez and B. Granger, "Jupyterhub." `https://jupyter.org/hub`.

[23] S. S. D. A. AB, "In modde® 12 user guide," 2017.

[24] W. Black and B. J. Babin, "Multivariate data analysis: Its approach, evolution, and impact," in *The Great Facilitator: Reflections on the Contributions of Joseph F. Hair, Jr. to Marketing and Business Research*, pp. 121–130, Springer, 2019.

[25] "Introduction to multivariate data analysis in chemical engineering." `https://www.controleng.com/articles/introduction-to-multivariate-data-analysis-in-chemical-engineering/`.

[26] A. Saxena and P. Prathipati, "Comparison of mlr, pls and ga-mlr in qsar analysis," *SAR and QSAR in Environmental Research*, vol. 14, no. 5-6, pp. 433–445, 2003.

[27] "Multiple linear regression (mlr) definition, formula, and example." `https://www.investopedia.com/terms/m/mlr.asp`.

[28] I. Smallwood, *Handbook of organic solvent properties*. Butterworth-Heinemann, 2012.

[29] NationalLibraryofMedicine, "Pubchem." `https://pubchem.ncbi.nlm.nih.gov/`, 2023.

[30] RDKit, "Open-source cheminformatics software rdkit.org: Github, sourceforge." `https://www.rdkit.org/`.

[31] OrganicChemistryPortal, "Clogp calculation." `https://www.organic-chemistry.org/prog/peo/cLogP.html`, 2023.

[32] S. A. Wildman and G. M. Crippen, "Prediction of physicochemical parameters by atomic contributions," *Journal of chemical information and computer sciences*, vol. 39, no. 5, pp. 868–873, 1999.

[33] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[34] N. R. Draper and H. Smith, *Applied regression analysis*, vol. 326. John Wiley & Sons, 1998.

[35] A. Vishwakarma and J. M. Karp, *Biology and engineering of stem cell niches*. Academic Press, 2017.

[36] T. J. Ritchie and S. J. Macdonald, "The impact of aromatic ring count on compound developability–are too many aromatic rings a liability in drug design?," *Drug discovery today*, vol. 14, no. 21-22, pp. 1011–1020, 2009.

[37] J. Clark, "Basic properties of amines." https://chem.libretexts.org/Bookshelves/Organic_Chemistry/Supplemental_Modules_(Organic_Chemistry)/Amines/Properties_of_Amines/Basic_Properties_of_Amines.

[38] M. J. Yunta, "It is important to compute intramolecular hydrogen bonding in drug design," *Am. J. Model. Optim*, vol. 5, no. 1, pp. 24–57, 2017.

[39] L. Cui, K. Kittipongpittaya, D. J. McClements, and E. A. Decker, "Impact of phosphoethanolamine reverse micelles on lipid oxidation in bulk oils," *Journal of the American Oil Chemists' Society*, vol. 91, pp. 1931–1937, 2014.

Appendix

A Coefficients plots made with MODDE full factorial design

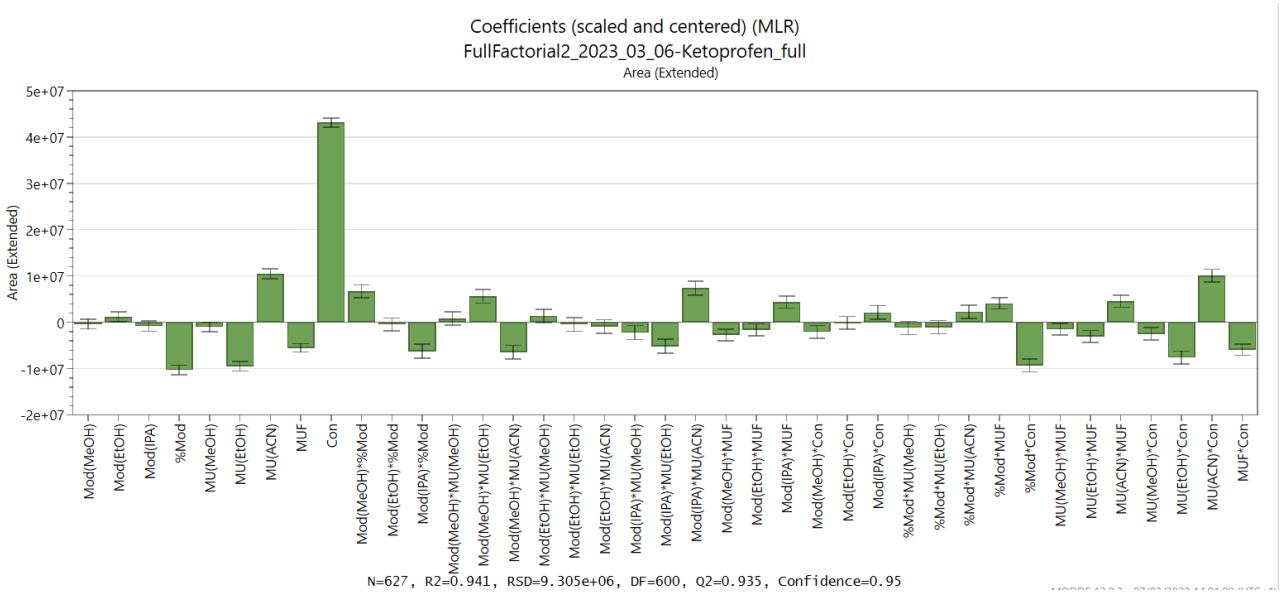


Figure 18: Coefficient plot for *ketoprofen* for full factorial design performed with MODDE. Nonsignificant variables has been excluded. *Mod* describes modifier, *%Mod* - modifier percentage, *Mu* - make-up, *MUF* - flow rate of make-up, *Con* - sample concentration.

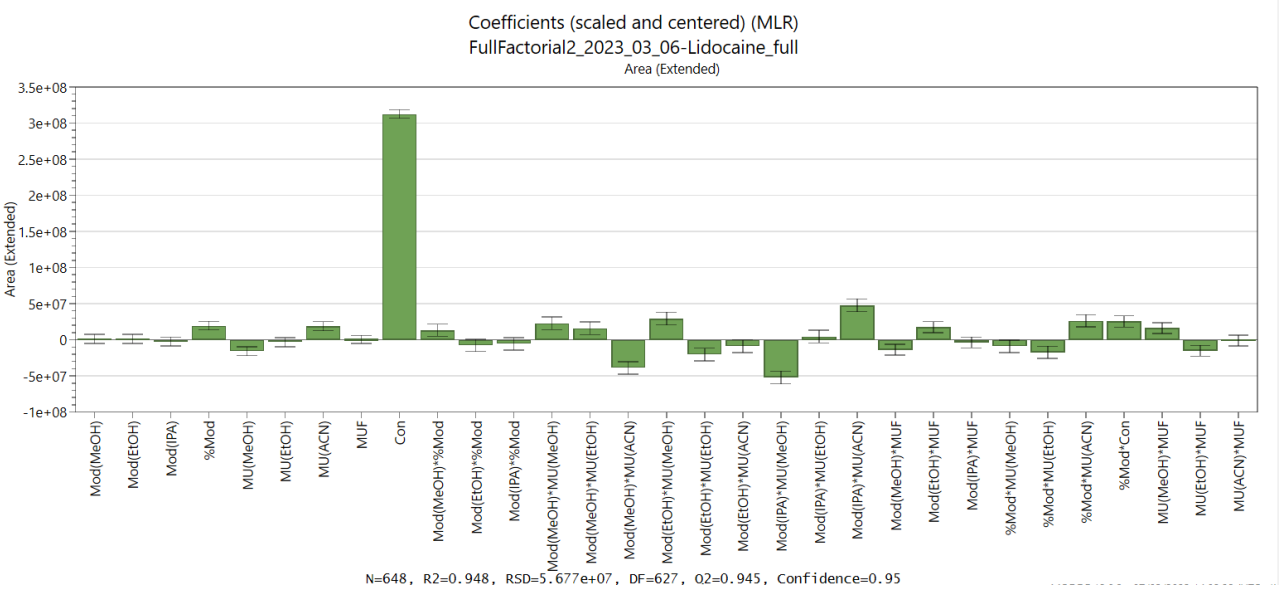


Figure 19: Coefficient plot for *lidocaine* for full factorial design performed with MODDE. Nonsignificant variables has been excluded. *Mod* describes modifier, *%Mod* - modifier percentage, *Mu* - make-up, *MUF* - flow rate of make-up, *Con* - sample concentration.

B Compound suppliers

Table 11: List of suppliers for compounds used in this study. Compounds purity is 99%+, otherwise stated.

Compound name	Supplier	Comments
Acyclovir	MilliporeSigma	
Adenosine	MilliporeSigma	
Amoxicillin	MilliporeSigma	
Ampicillin	MilliporeSigma	
Ascorbic acid	Thermo Scientific™ (former Acros Organics)	L-Ascorbic acid
Azithromycin dihydrate	Fluorochem	Calculations omitted water
Budesonide	AstraZeneca AB	
Cholesterol	MilliporeSigma	
Cytidine	MilliporeSigma	
DOPE	*Internally provided*	
DSPC	*Internally provided*	
Diclofenac acid	AstaTech Inc	Purity: 97%
Doxycycline	Clontech	
Estradiol	MilliporeSigma	
Ibuprofen	MilliporeSigma	
Ketoprofen	MilliporeSigma	
Lidocaine	MilliporeSigma	
L-Glutamic acid	MilliporeSigma	
L-Isoleucine	MilliporeSigma	
L-Leucine	MilliporeSigma	
L-Methionine	MilliporeSigma	
L-Proline	Janssen Chimica	
L-Serine	MilliporeSigma	
L-Thyroxine	MilliporeSigma	
L-Tryptophan	MilliporeSigma	
MC3	BLDpharm	
Metronidazole	Thermo Scientific™	
Naproxen	MilliporeSigma	
Nicotinamide	MilliporeSigma	
Nicotinic acid	FLUKA™	
Penicillin G Potassium	FLUKA™	
Testosterone	SERVA	
Tetracycline HCl	Calbiochem™	Calculations omitted hydrochloride
Thiamine HCl	MilliporeSigma	
Trimethoprim	Selleck	
Uridine	MilliporeSigma	

C Chromatograms of test compounds

Table 12: *Ketoprofen* chromatogram for concentrations 0.4 mM, 1.7 mM showing presence of a front peak.

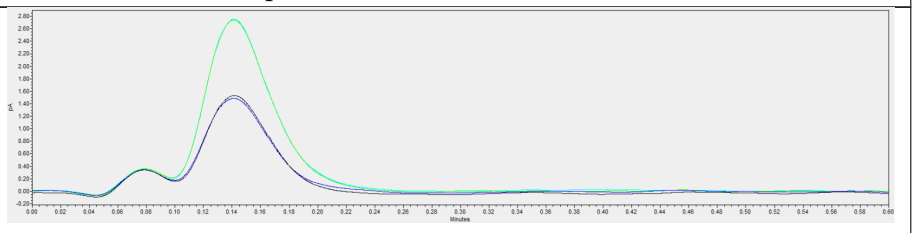
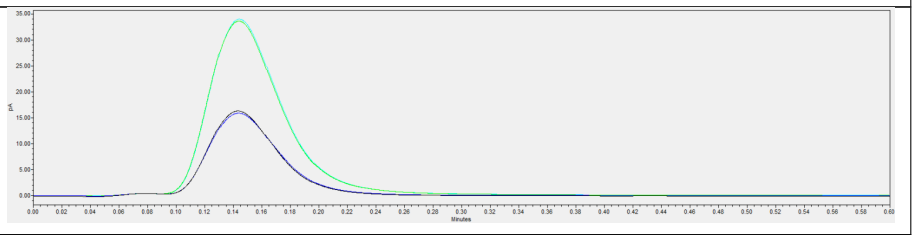
Ketoprofen	
IPA_25_muMeOH_05	

Table 13: *Lidocaine* chromatogram for concentrations 0.4 mM, 1.7 mM showing single response peak.

Lidocaine	
IPA_25_muMeOH_05	

D Heat map of CAD signal response peak area

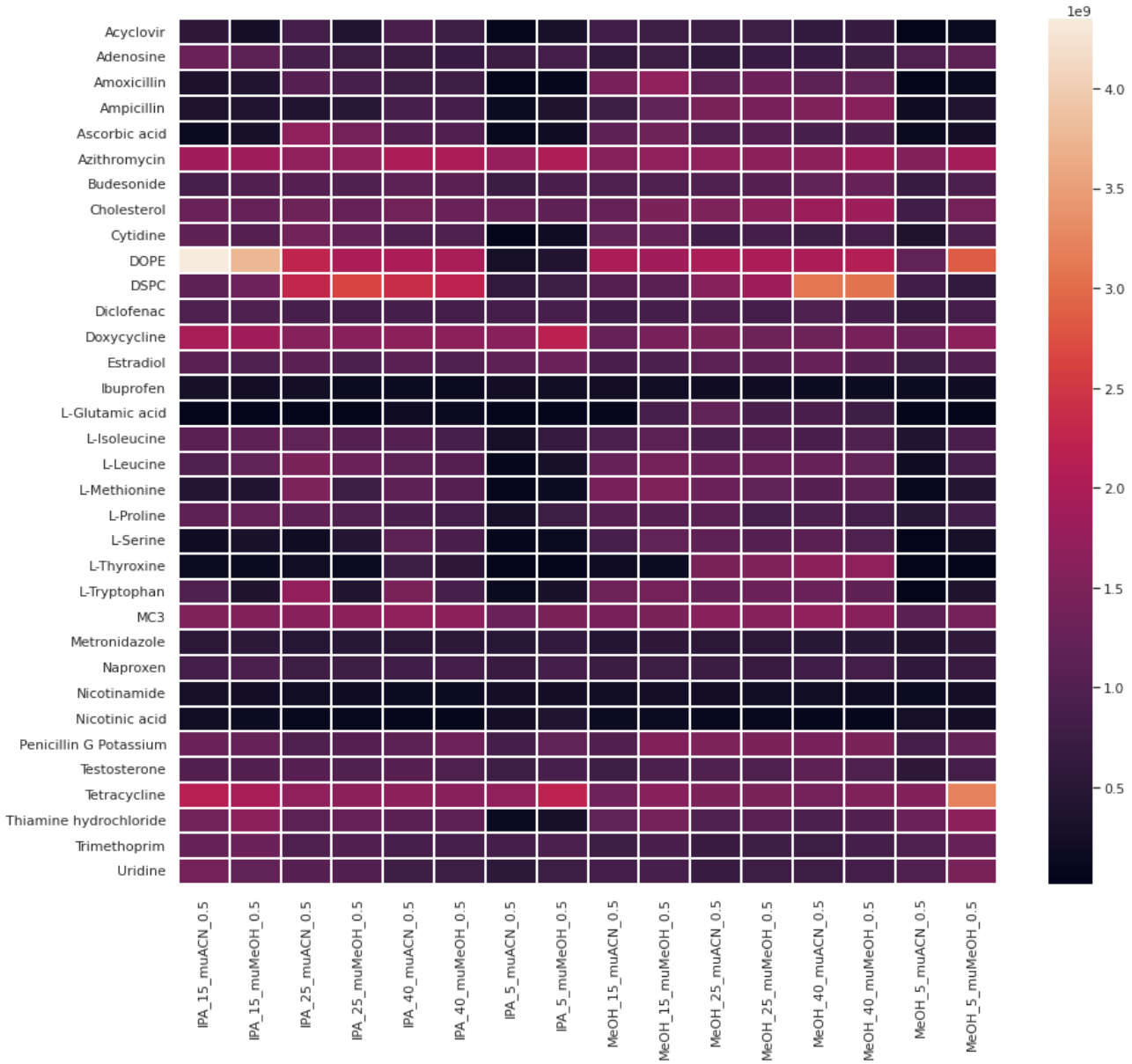


Figure 20: Heat map comparing values of maximum signal peak area for each method for all analysed compounds as well as for all methods. Compounds are segregated in alphabetic order by default. Area displayed in billions [aA*sec].

E Chromatograms for compounds and methods with R^2<0.8

Table 14: *L-Isoleucine* chromatogram for a method with R^2<0.8.

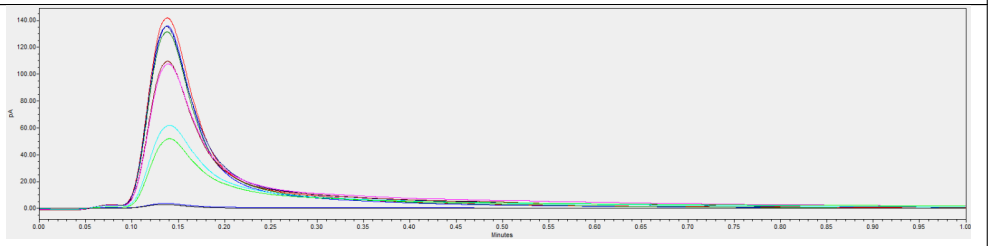
L-Isoleucine	
IPA_5_muMeOH	

Table 15: *L-Leucine* chromatogram for methods with R^2<0.8.

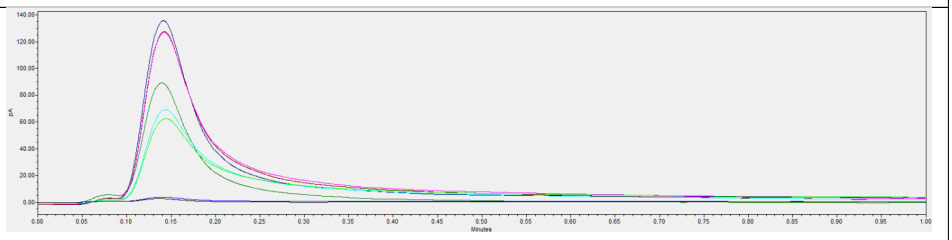
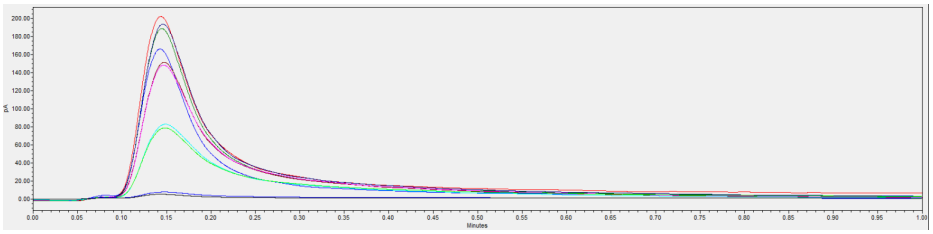
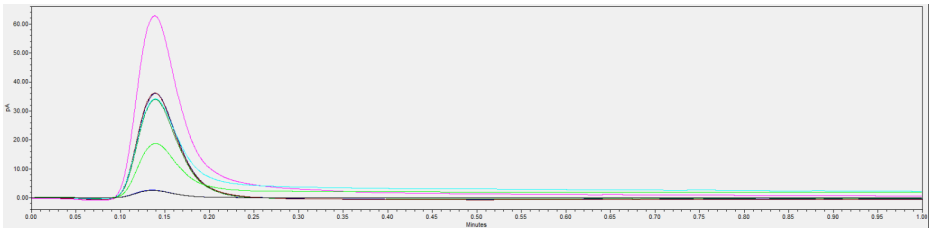
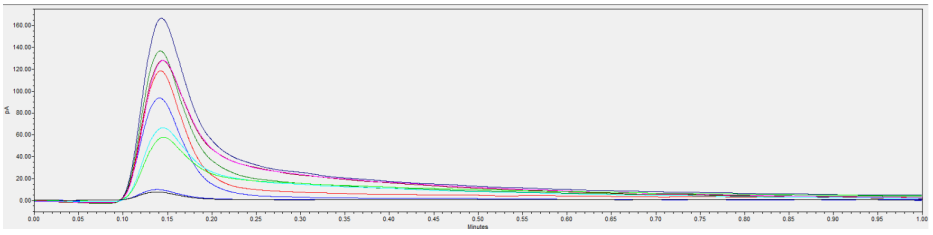
L-Leucine	
MeOH_5_muMeOH	
IPA_15_muMeOH	
MeOH_5_muACN	
IPA_15_muACN	

Table 16: *L-Glutamic acid* chromatogram for a method with R^2<0.8.

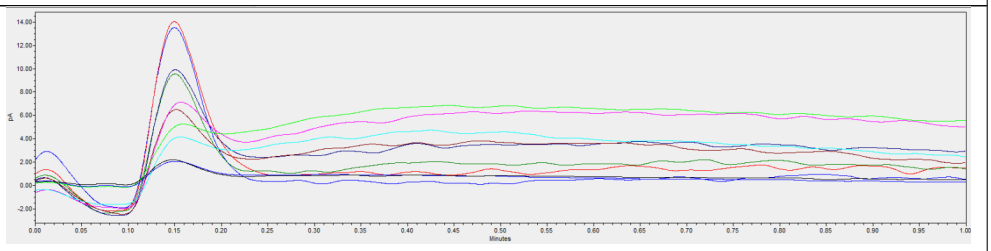
L-Glutamic acid	
IPA_40_muACN	

Table 17: *L-Methionine* chromatograms for methods with R^2<0.8.

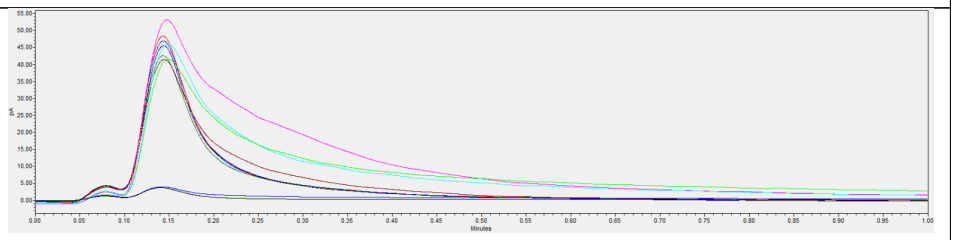
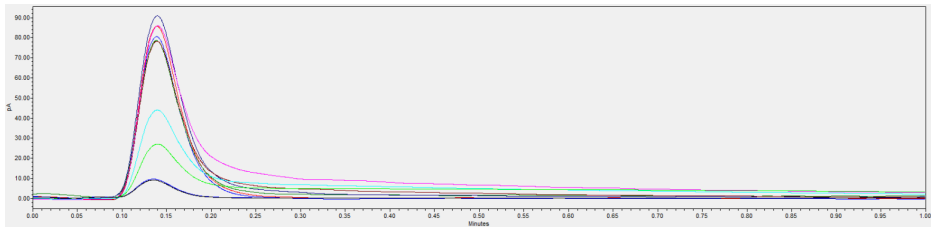
L-Methionine	
MeOH_5_muMeOH	
IPA_15_muACN	

Table 18: *L-Tryptophan* chromatograms for methods with R^2<0.8.

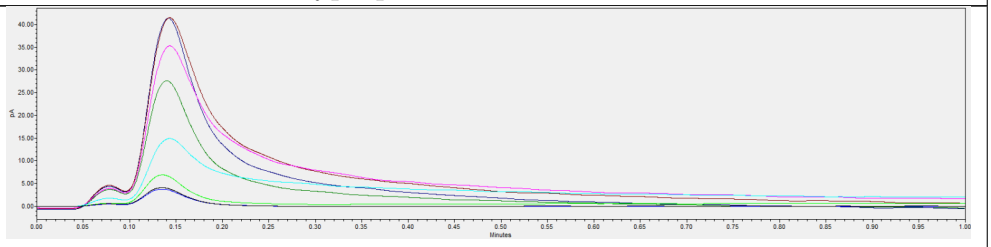
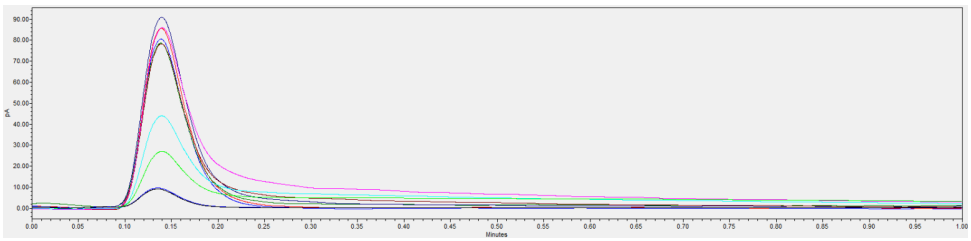
L-Tryptophan	
IPA_5_muMeOH	
IPA_15_muACN	

Table 19: *L-Thyroxine* chromatograms for methods with R^2<0.8.

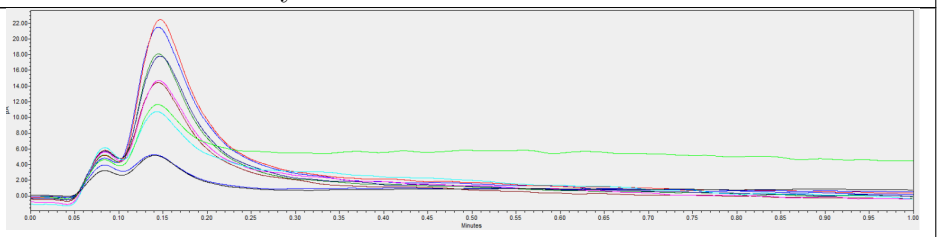
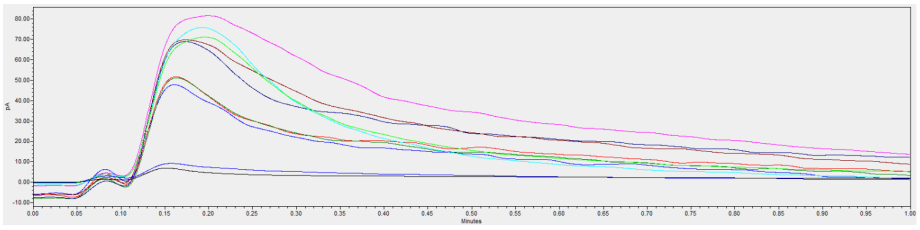
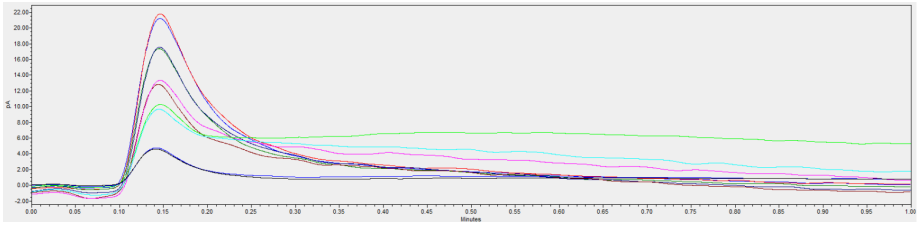
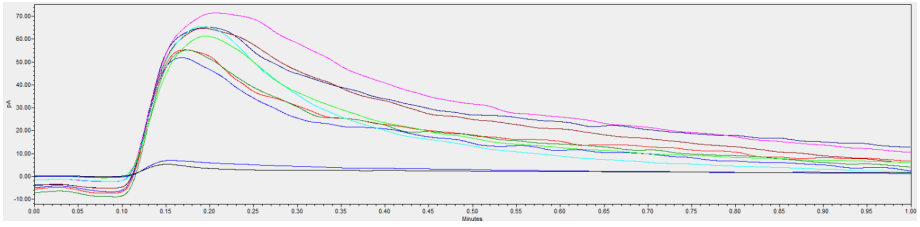
L-Thyroxine	
MeOH_15_muMeOH	
MeOH_25_muMeOH	
MeOH_15_muACN	
MeOH_25_muACN	

Table 20: *Metronidazole* chromatogram for a method with R^2<0.8.

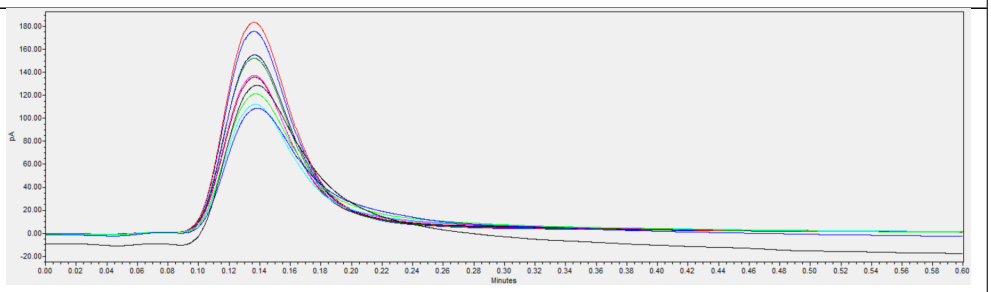
Metronidazole	
IPA_5_muMeOH	

Table 21: *Amoxicillin* chromatogram for a method with R^2<0.8.

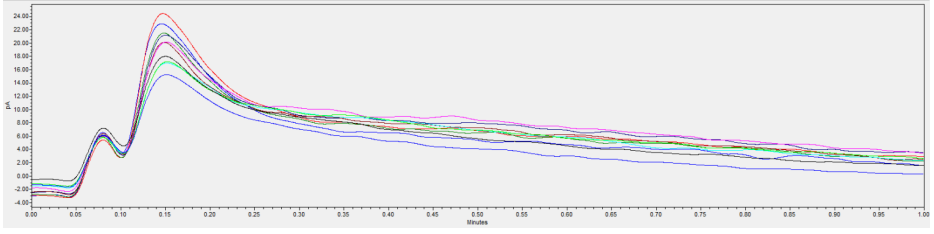
Amoxicillin	
IPA_15_muMeOH	

Table 22: *Penicillin G K* chromatogram for a method with R^2<0.8.

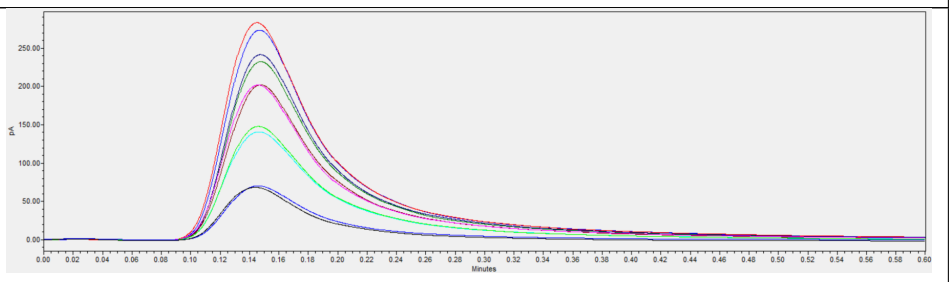
Penicillin G K	
MeOH_40_muACN	

Table 23: *Acyclovir* chromatograms for methods with R^2<0.8.

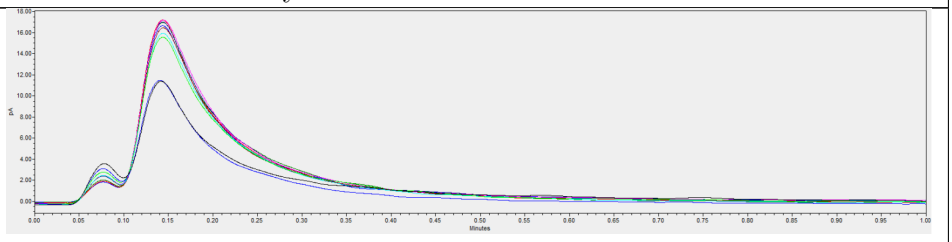
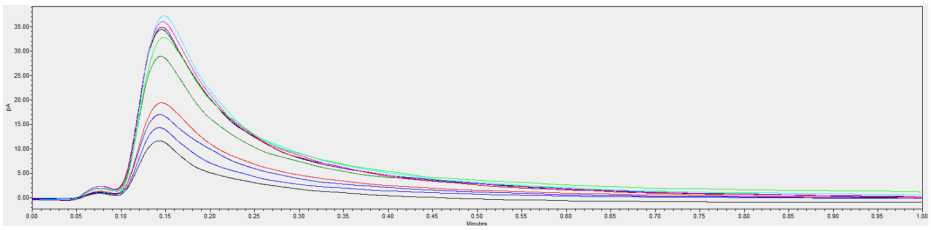
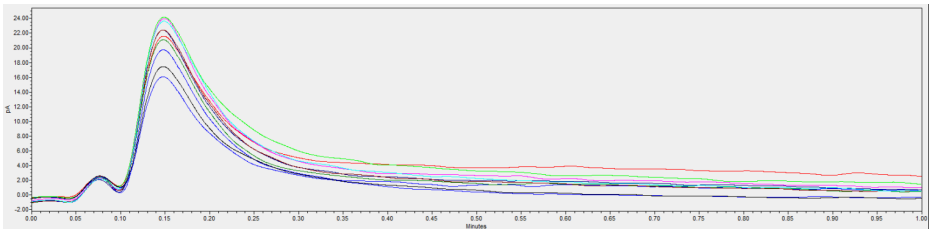
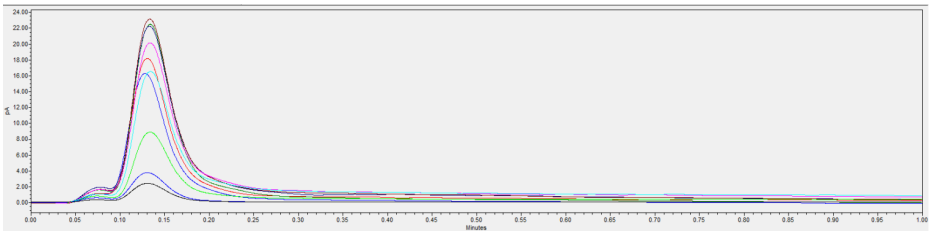
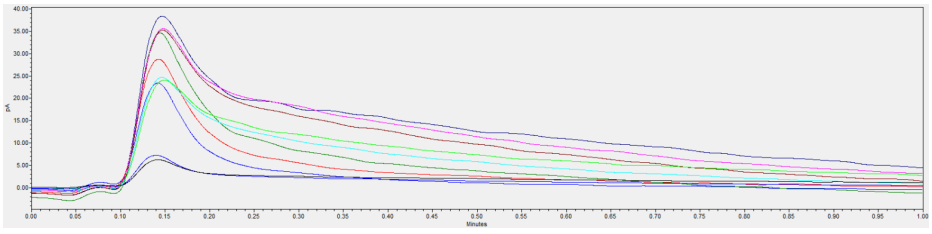
Acyclovir	
MeOH_5_muMeOH	
IPA_5_muMeOH	
IPA_15_muMeOH	
IPA_5_muACN	
IPA_15_muACN	

Table 24: *Cytidine* chromatograms for methods with R^2<0.8.

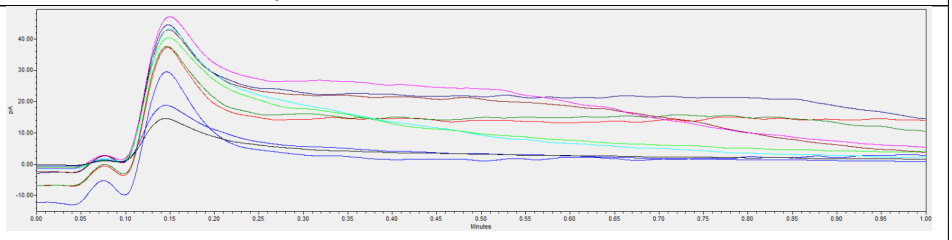
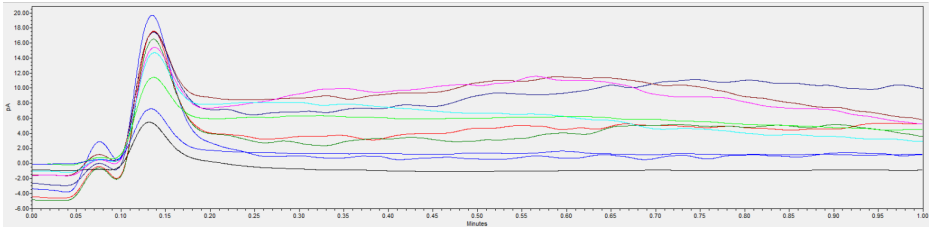
Cytidine	
MeOH_5_muMeOH	
MeOH_5_muACN	

Table 25: *DOPE* chromatograms for methods with R^2<0.8.

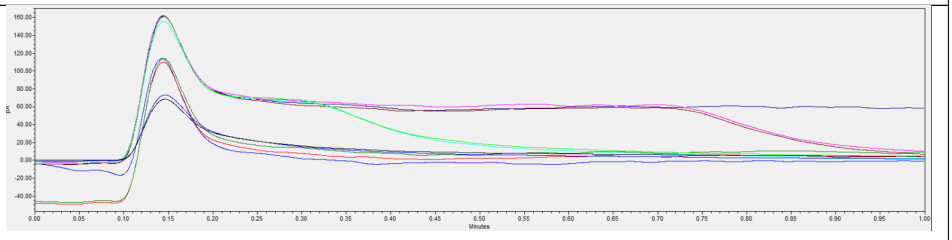
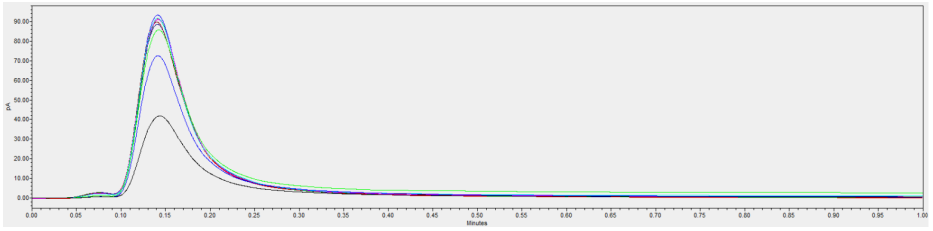
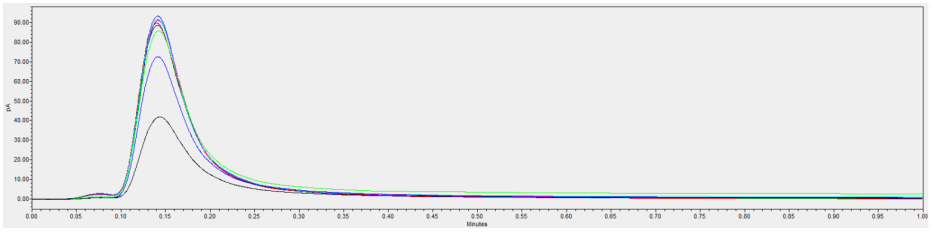
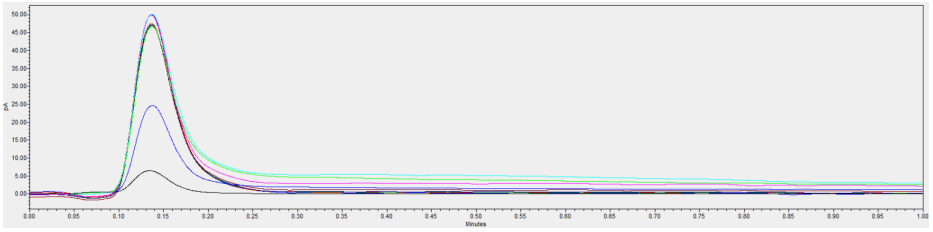
DOPE	
MeOH_5_muMeOH	
IPA_5_muMeOH	
MeOH_5_muACN	
IPA_5_muACN	

Table 26: *DSPC* chromatograms for methods with R^2<0.8.

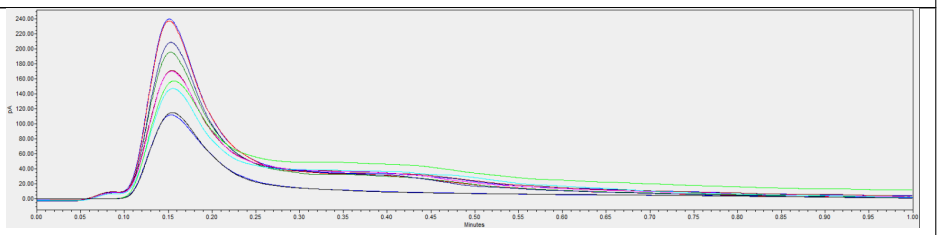
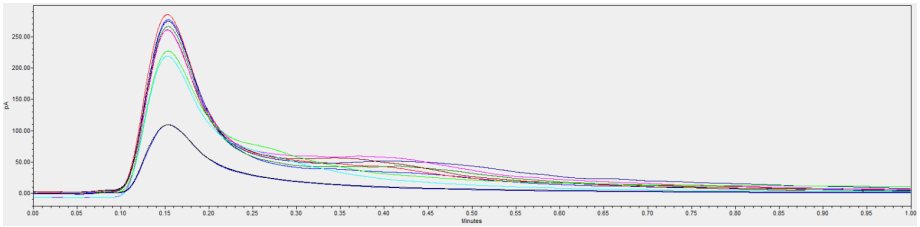
DSPC	
MeOH_25_muMeOH	
IPA_25_muACN	

Table 27: *Nicotinamide* chromatogram for a method with R^2<0.8.

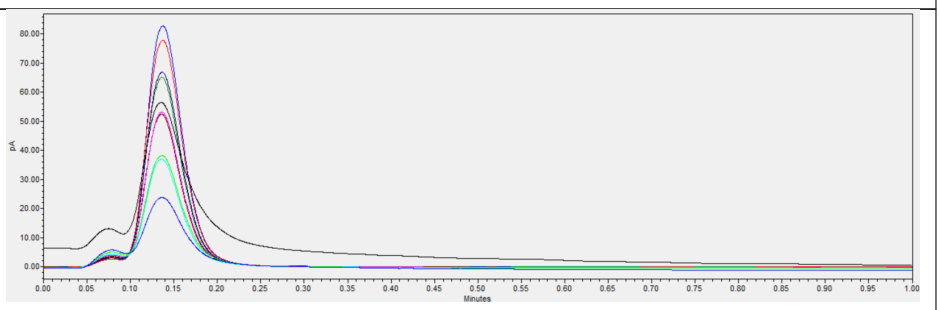
Nicotinamide	
MeOH_5_muMeOH	

Table 28: *Nicotinic acid* chromatograms for methods with R^2<0.8.

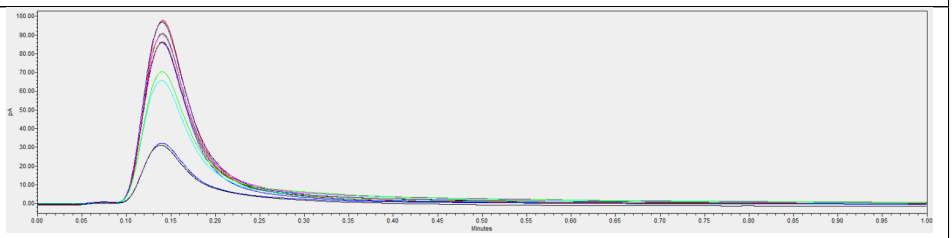
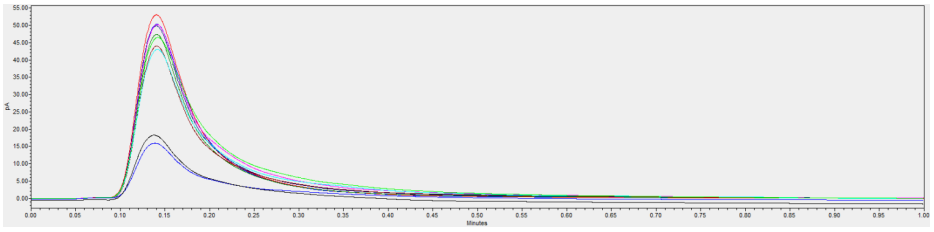
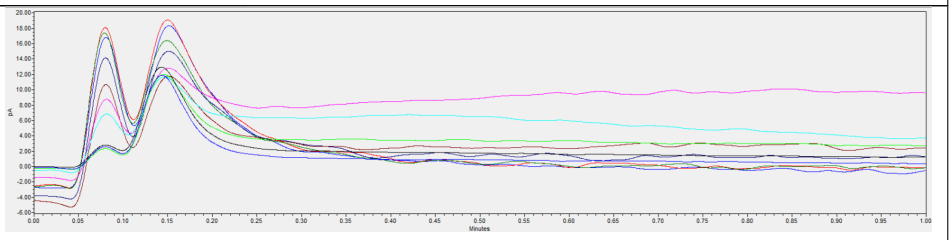
Nicotinic acid	
IPA_5_muMeOH	
MeOH_5_muACN	

Table 29: *Ascorbic acid* chromatogram for a method with R^2<0.8.

Ascorbic acid	
IPA_15_muMeOH	

F Adjusted values of concentrations for methods resulting in R^2<0.8

Table 30: Listed compounds, analysed concentrations and methods with R^2<0.8. Adjusted values give concentration range within which linear behaviour can be observed. '*' means that there is one data point that has been considered as an outlier.

Compound	Method	Concentration	R^2	Adjusted concentration	Adjusted R^2
L-Isoleucine	IPA_5_muMeOH	c_1-c_5	0.745	c_1-c_4	0.893
L-Leucine	MeOH_5_muMeOH	c_1-c_4	0.558	c_1-c_3	0.983
	IPA_15_muMeOH	c_1-c_5	0.734	c_1-c_4	0.936
	MeOH_5_muACN	c_1-c_4	0.484	c_1-c_3	0.993
	IPA_15_muACN	c_1-c_5	0.227	c_1-c_4	0.855
L-Glutamic acid	IPA_40_muACN	c_1-c_5	0.01	c_1-c_3	0.869
L-Methionine	MeOH_5_muMeOH	c_1-c_5	0.068	c_1-c_2	1
	IPA_15_muACN	c_1-c_5	0.321	c_1-c_3	0.996
L-Tryptophan	IPA_5_muMeOH	c_1-c_4	0.663	c_1-c_3	0.919
	IPA_15_muACN	c_1-c_5	0.513	c_1-c_4	0.959
L-Thyroxine	MeOH_15_muMeOH	c_1-c_5	0.635	c_1-c_2	1
	MeOH_25_muMeOH	c_1-c_5	0.291	c_1-c_3	0.968
	MeOH_15_muACN	c_1-c_5	0.373	c_1-c_2	1
	MeOH_25_muACN	c_1-c_5	0.44	c_1-c_3	0.974
Metronidazole	IPA_5_muMeOH	c_1-c_5	0.567	c_2-c_4	0.927
Amoxicillin	IPA_15_muMeOH	c_1-c_5	0.768	c_1-c_4	0.865
Penicillin G K	MeOH_40_muACN	c_1-c_5	0.582	c_1-c_3,c_5*	0.965
Acyclovir	MeOH_5_muMeOH	c_1-c_5	0.319	c_1-c_3	0.827
	IPA_5_muMeOH	c_1-c_5	0.03	c_1-c_2	1
	IPA_15_muMeOH	c_1-c_5	0.14	c_1-c_2	1
	IPA_5_muACN	c_1-c_5	0.483	c_1-c_4	0.832
	IPA_15_muACN	c_1-c_5	0.093	c_1-c_4	0.805
Cytidine	MeOH_5_muMeOH	c_1-c_5	0.259	c_1-c_3	1
	MeOH_5_muACN	c_1-c_5	0.23	c_1-c_3	0.992
DOPE	MeOH_5_muMeOH	c_1-c_5	0.008	c_1-c_3	0.999
	IPA_5_muMeOH	c_1-c_5	0.45	c_1-c_2	1
	MeOH_5_muACN	c_1-c_5	0.061	c_1-c_2	1
	IPA_5_muACN	c_1-c_5	0.058	c_1-c_2	1
DSPC	MeOH_25_muMeOH	c_1-c_5	0.639	c_1-c_2	1
	IPA_25_muACN	c_1-c_5	0.641	c_1-c_4	0.804
Nicotinamide	MeOH_5_muMeOH	c_1-c_5	0.76	c_2-c_5	1
Nicotinic acid	IPA_5_muMeOH	c_1-c_5	0.604	c_1-c_4	0.833
	MeOH_5_muACN	c_1-c_5	0.428	c_1-c_2	1
Ascorbic acid	IPA_15_muMeOH	c_1-c_5	0.54	c_1-c_3	0.999

G Detailed missing data

Table 31: Table listing missing parts of data.

Compound	Method	Missing concentration
L-Leucine	MeOH_5_muMeOH	c_5
	IPA_5_muMeOH	c_4, c_5
	MeOH_5_muACN	c_5
	IPA_5_muACN	c_4, c_5
L-Glutamic acid	IPA_5_muMeOH	c_4, c_5
	IPA_15_muMeOH	c_4, c_5
	IPA_25_muMeOH	c_4, c_5
	IPA_5_muACN	c_4, c_5
	IPA_15_muACN	c_4, c_5
	IPA_25_muACN	c_4, c_5
L-Methionine	IPA_5_muMeOH	c_5
	IPA_5_muACN	c_5
L-Tryptophan	MeOH_5_muMeOH	c_3-c_5
	IPA_5_muMeOH	c_5
	MeOH_5_muACN	c_3-c_5
	IPA_5_muACN	c_5
L-Thyroxine	MeOH_5_muMeOH	c_5
	IPA_5_muMeOH	c_5
	IPA_15_muMeOH	c_5
	MeOH_5_muACN	c_5
	IPA_5_muACN	c_5
	IPA_15_muACN	c_5
Acyclovir	MeOH_5_muACN	c_5
Thiamine hydrochloride	IPA_5_muMeOH	c_5
	IPA_5_muACN	c_5

H Trends observed from MODDE response contour plots

Table with observed trends from response contour plots obtained with MODDE.

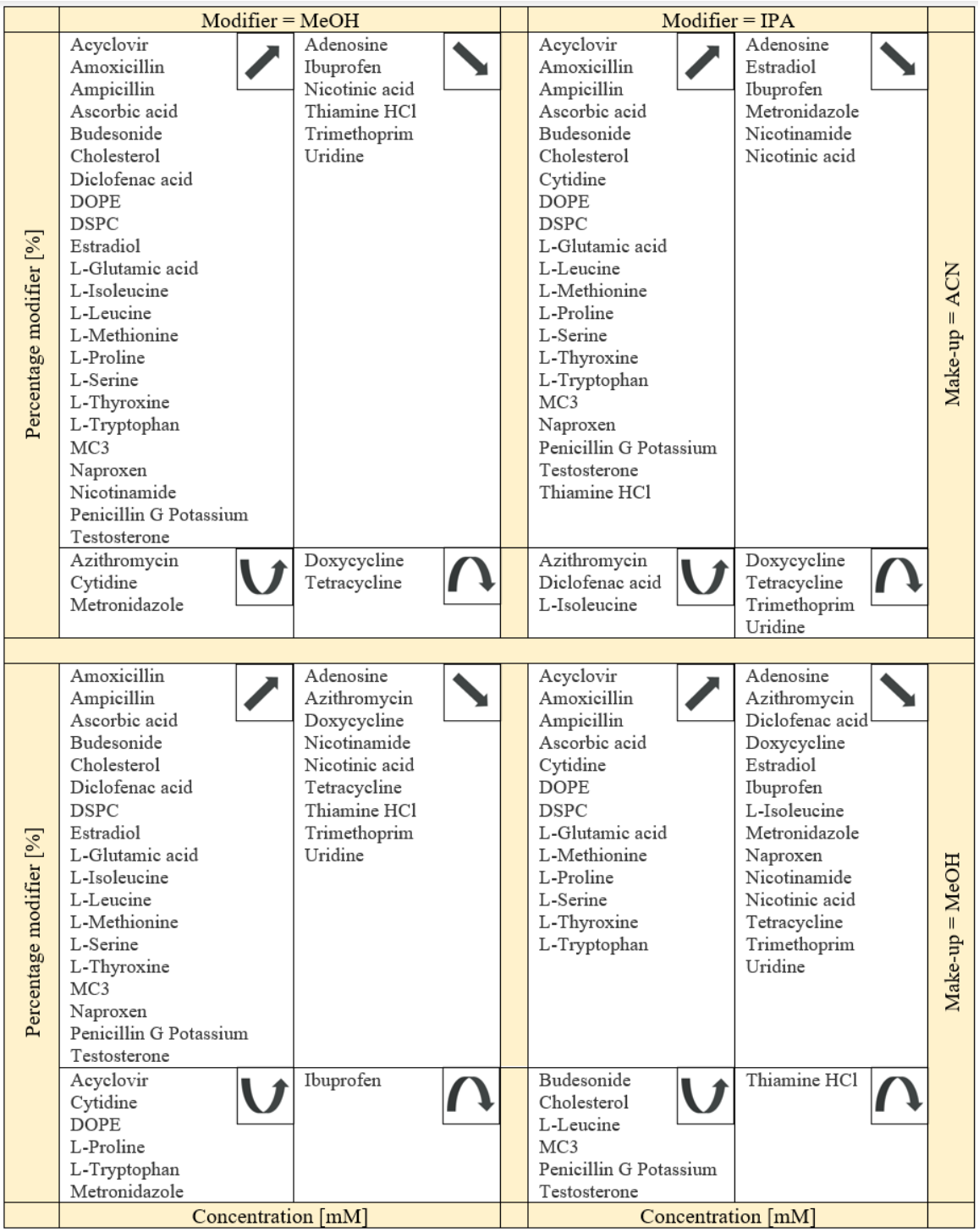


Figure 21: Observed behaviour of response contours of area for 34 analysed compounds.

I 4D response contour of area plots

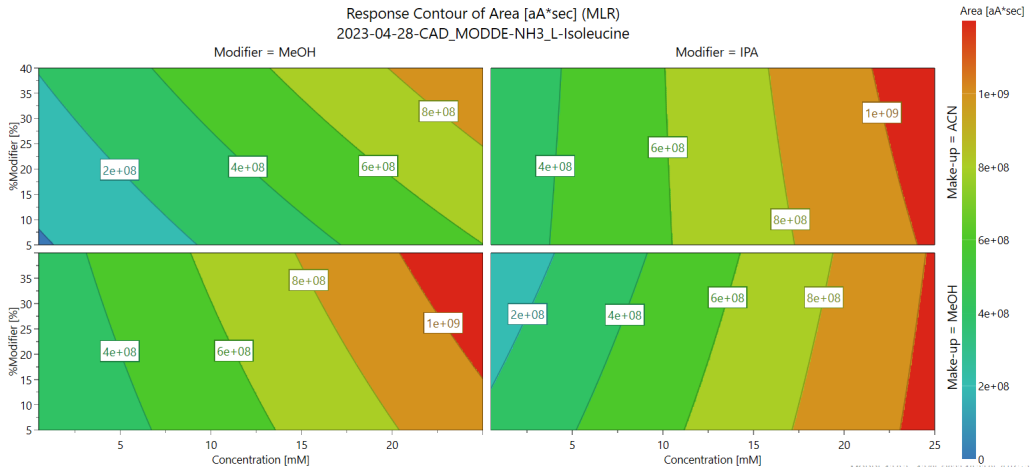


Figure 22: 4D response contour of area plot for *L-isoleucine* showing mixed behaviour of combined analysis parameters.

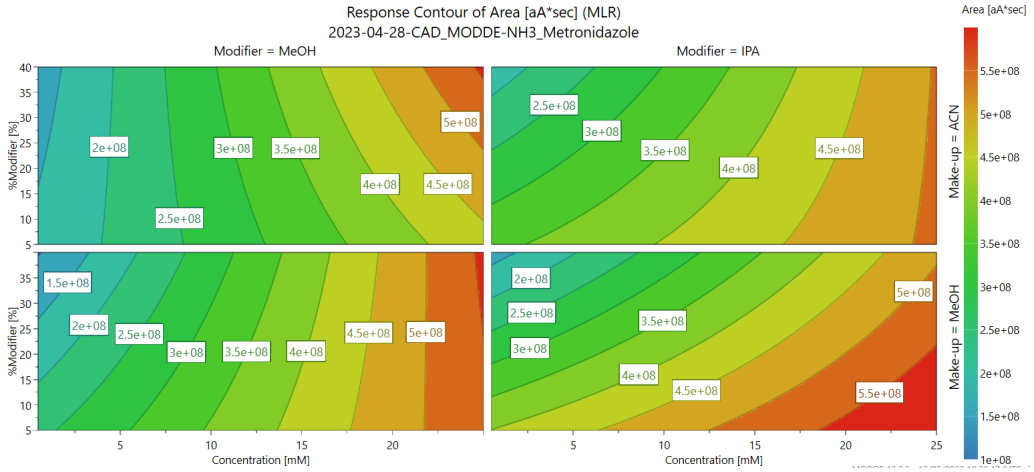


Figure 23: 4D response contour of area plot for *metronidazole* showing mixed behaviour of combined analysis parameters.

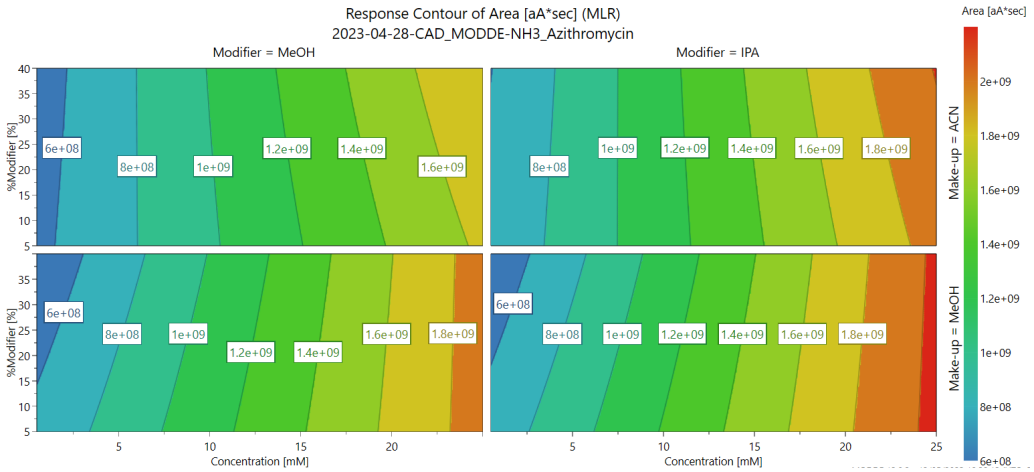


Figure 24: 4D response contour of area plot for *azithromycin* showing mixed behaviour of combined analysis parameters.

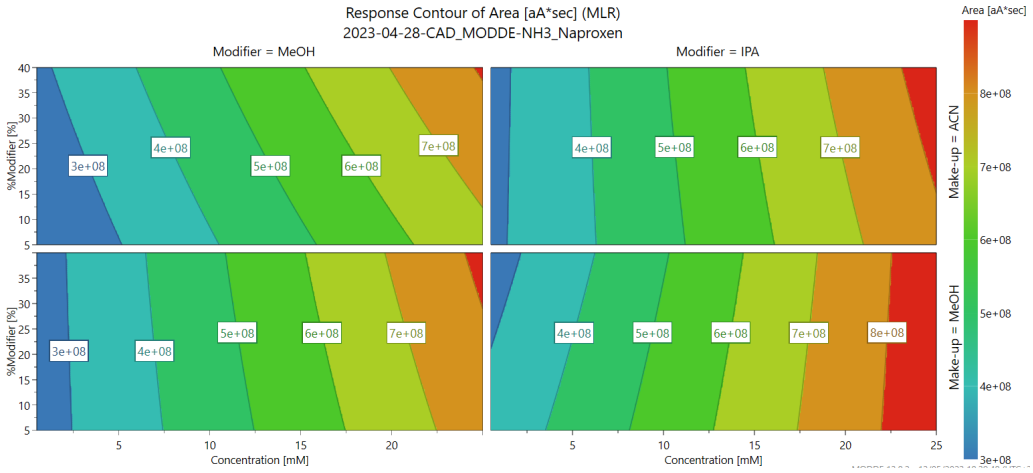


Figure 25: 4D response contour of area plot for *naproxen* showing mixed behaviour of combined analysis parameters.

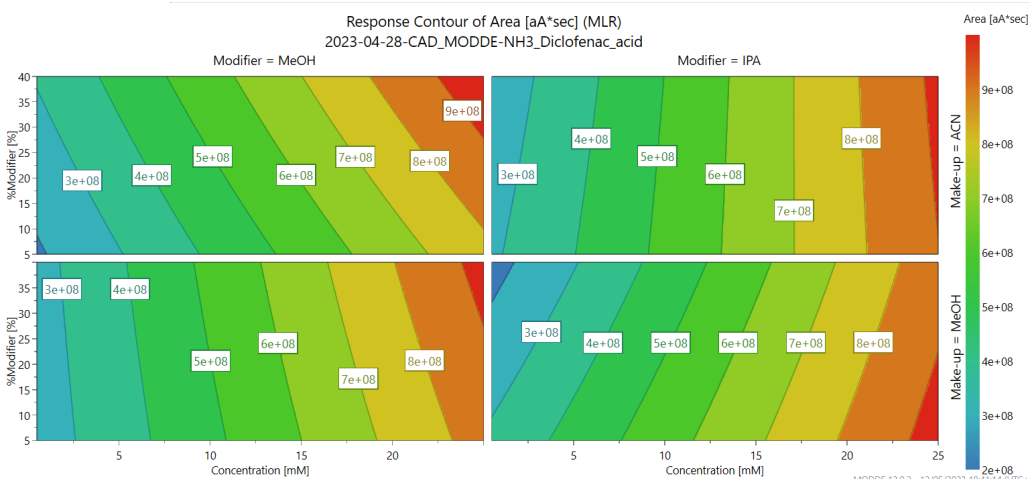


Figure 26: 4D response contour of area plot for *diclofenac acid* showing mixed behaviour of combined analysis parameters.

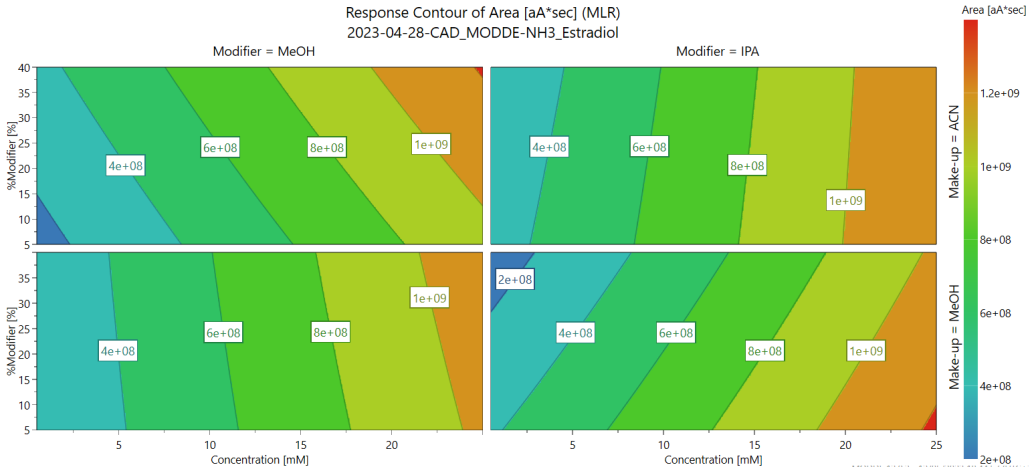


Figure 27: 4D response contour of area plot for *estradiol* showing mixed behaviour of combined analysis parameters.

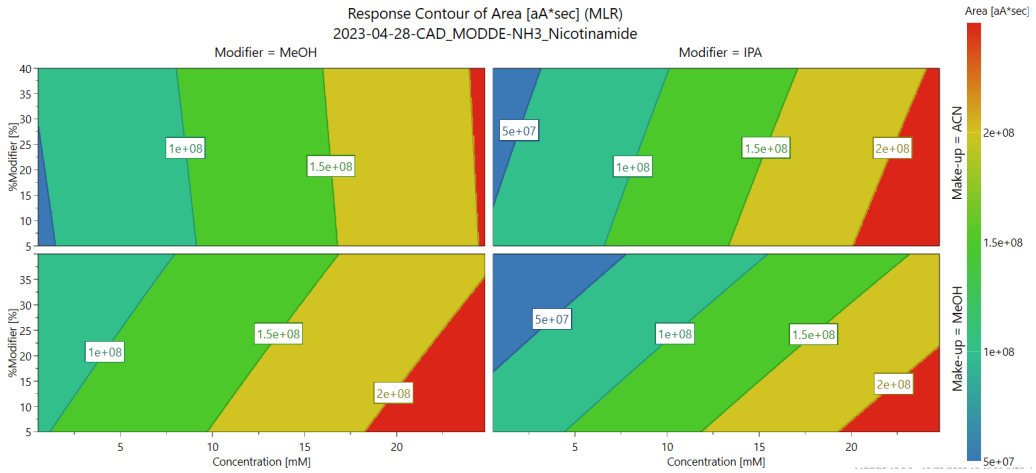


Figure 28: 4D response contour of area plot for *nicotinamide* showing mixed behaviour of combined analysis parameters.

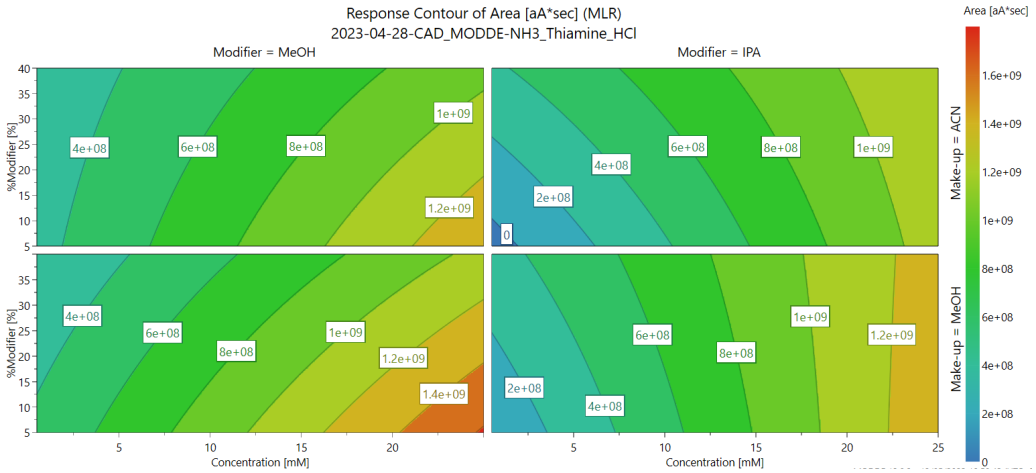


Figure 29: 4D response contour of area plot for *thiamine HCl* showing mixed behaviour of combined analysis parameters.