Creating a High-Throughput Workflow for Automated Peptide Characterization using LC-MS

Sara Forsberg

Degree Project in Analytical Chemistry, 2023
Department of Chemistry
Lund University
Sweden

MSc, 30 hp

LUND UNIVERSITY

Creating a High-Throughput Workflow for Automated Peptide Characterization using LC-MS

Sara Forsberg

LUND
UNIVERSITY

Degree Project in Analytical Chemistry

2023

MSc, 30 hp

Supervisors

Tomas Leek & Dr. Lena von Sydow
R&I Chemistry, AstraZeneca

Margareta Sandahl
Senior lecturer, Center for Analysis and Synthesis, Lund University

Examiner

Peter Spégel
Senior lecturer, Center for Analysis and Synthesis, Lund University

Lund University

Department of Chemistry

Center for Analysis and Synthesis

P.O. Box 124

SE-221 00 Lund, Sweden

Popular Science Summary

Creating a Workflow for Characterizing Peptides

At the early stage of a drug discovery project, several hundreds of compounds are of interest to be profiled for the upcoming steps. At this stage it is crucial to have analysis techniques that can characterize the compounds with high quality. In this case, it means to confirm the compound's identity and relative purity, to be sure the results are reliable. When considering analysis of a large number of compounds in a short time, it is called high-throughput analysis.

This degree project was performed at the pharmaceutical company AstraZeneca in Gothenburg. The application is in line with AstraZeneca's research interest. The overall aim was to develop an analytical workflow that could be applied for a high-throughput analysis of a class of molecules called peptides. This was achieved by using the analysis technique liquid chromatography connected to mass spectrometry (LC-MS). From liquid chromatography, estimations of relative purity could be obtained and from mass spectrometry confirmation of the peptides' identity. The reason for a high-throughput workflow is to save solvent and energy during the analysis, save time for the user (shorter runs), and it has more user-friendly manual processing for non-experts to perform the characterization. One key success factor was to obtain a shorter run time, but to still get high-quality results.

During the experimental work, 20 peptides of different sizes and chemical properties were investigated. Initially, they needed to be dissolved to be able to run them on the LC-MS instrument. Then, experiments were performed on the LC-MS. A workflow was developed and then it was optimized. Different analysis parameters were investigated, for example the gradient of mobile phases and flow rate. Additionally, different software for processing of raw data from the workflow were considered. This to find the most suitable one to be used in the workflow that would make it easier to automatically perform processing and reporting of data.

As result, it was seen that a shorter method in runtime of instrument could be applied on a range of peptides and gave a good separation of the peptides and impurities. This was obtained by using a CSH column of 50 mm, flow rate of 0.5 mL/min and a gradient time of 3.5 min with the interval of either 10-60% or 5-55%. To know if this workflow will be suitable for a peptide of interest, limits was set up based on the peptide's lipophilicity, which means tendency to be in an organic phase instead of an aqueous phase. These limits had to be drawn depending on the peptide's ion class (base, neutral or zwitterion). As for processing of raw data the application Intact Mass was chosen due to having good opportunities of estimating a combined absorbance and mass spectral purity with high quality. Also, because it can sort out complex spectra with many masses from different peptides.

In conclusion, a more high-throughput friendly workflow for peptide characterization has been developed that can be applied on a range of peptides. Properties of peptides that describes lipophilicity can be used to predict whether this workflow is suitable or not. The application Intact Mass was chosen for processing of raw data since it can estimate a combined relative purity and potential impurities.

Abstract

In the early stage of a drug discovery project, there is a need for efficient methods that can analyse peptides in short time. This includes methods that confirm the peptide's identity and estimates its relative purity in an efficient and reliable way. The aim was to create a workflow for peptide characterization for AstraZeneca's in-house peptides that was applicable for a high-throughput analysis.

The approach was to develop and optimise an LC-MS method based on 20 therapeutic peptides considering stationary phase, gradient of mobile phase and flow rate. Also, to evaluate different peptide descriptors to see if they could be used to predict the suitability of a peptide for this workflow and to consider the usefulness and different features of software for processing of raw data.

The result was that an LC-MS method with an acquisition time of 5 minutes was developed. The method comprised a CSH column (1.7 µm, 2.1 x 50 mm), 0.5 mL/minute flow rate and gradient time of 3.5 minutes (slope 14 %B/minutes). The descriptors ClogK_D and aqueous solubility were useful to predict if this method was applicable for the peptides in question. Three software for processing of raw data was considered and the software Waters Connect with its application Intact Mass was chosen. Intact Mass can deconvolve the neutral mass, yield a purity as a combined UV and mass spectral purity (UVxMS) and perform a simple impurity profiling.

In conclusion, an LC-MS method adapted for a high-throughput workflow was accomplished that succeeded to obtain adequate results regarding retention time and separation of potential impurities for a range of peptides. Furthermore, descriptors turned out to be useful for predicting suitability of this workflow and an appropriate software could be applied for processing of raw data.

Keywords: High-throughput analysis, Impurity profiling, UHPLC-HRMS, Peptide characterization, Peptide descriptors.

Acknowledgements

First, I want to thank the Physical and Analytical Chemistry (PAC) team at Medicinal Chemistry, Research and Early Development, Respiratory and Immunology (RI), BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden, for welcoming me and giving me this interesting project. Especially thanks to my supervisors Lena von Sydow and Tomas Leek in the PAC team for guiding me and giving me valuable advice, as well as Kathrin Stavenhagen, Manasses Jora and Marta Passamonti in the PAC team.

I would also like to extend a big thank you to Werngard Czechtizky, Executive Director and Head of Department for Medicinal Chemistry, Respiratory and Immunology, AstraZeneca, for giving me the opportunity to come to AstraZeneca. As well as a thank to other helpful persons at AstraZeneca including Emma Evertsson, Leonardo De Maria, and Wu Su. I also want to thank Bernhard Poschner at Waters Corporation for all the help.

Further, I want to thank my supervisor Margareta Sandahl and examiner Peter Spégel, Centre for Analysis and Synthesis at Lund University, for good advice and support. Lastly, I want to thank my family and friends for supporting me through this project.

Table of Contents

List of Abbreviations

AZ	AstraZeneca in Gothenburg
ESI+	ElectroSpray Ionisation in Positive mode
HPLC	High-Performance Liquid Chromatography
HRMS	High Resolution Mass Spectrometry
K_D	Partition ratio
LC-MS	Liquid Chromatography Mass Spectrometry
pI	Isoelectric point
QTOF	Quadrupole Time-Of-Flight
RP	Reversed-Phase
RT	Retention Time
S	Aqueous Solubility
TFA	TriFluoroacetic Acid
TIC	Total Ion Current
UHPLC	Ultrahigh-Performance Liquid Chromatography

1 Introduction and Theory

Most molecules of interest in drug discovery have until some decades ago belonged to either of two categories: small molecules and larger biological proteins [1]. However, recently new modalities, including therapeutic peptides and oligonucleotides, have attracted attention. In this report, the focus was on peptides. Therapeutic peptides consist of natural and/or synthetic amino acid residues connected through peptide bonds. The length of the peptide therapeutic drugs on the market varies from a few residues up to 50 residues, or even more [1-3]. Novel technologies and strategies have made peptide development more feasible. This has led to possibilities to modify amino acids in the peptide to improve their potency as drugs such as improving their half-life, stability, solubility, and oral bioavailability [2-4].

Peptides are used in a vast therapeutic area including oncology, metabolic, respiratory, and cardiovascular diseases as a few examples [2, 5]. Oxytocin, Desmopressin and Atosiban are examples of three successful peptide therapeutic drugs which all three affect the hormone levels in the human body in various ways [3]. Peptides can be used for respiratory diseases, such as asthma and cystic fibrosis, as well. Several potential peptide drugs are in development for this, for example SPLUNC1-derived peptides [5].

In the early stage of a drug development project several hundreds or even thousands of compounds can be of interest. It is then of importance to have an efficient workflow that will accurately confirm the identity of the compound and with a high-quality estimate its relative purity. This to make sure that the scientist is working with the right compound and that there are no interfering compounds such as related degradation products. The process of analysing a large number of compounds in a short time is referred to as high-throughput analysis and is desirable since it shortens instrument time, is more sustainable since a lower solvent consumption is needed, and it allows non-experts to use the workflow. In addition, processing of data and reporting of outcome can be more automated as well, which will save time in manual evaluation. To make a high-throughput workflow efficient there is a need for short acquisition and processing times and also to use a method well adapted for the peptides giving precise, reliable, and robust results.

This degree project was performed within the Physical and Analytical Chemistry team belonging to the Respiratory and Immunology department at AstraZeneca (AZ) in Gothenburg. The aim was to create a high-throughput workflow that can be applied for AZ's in-house peptides. The following research questions were considered:

- *Can a generic LC-MS method be created for a range of peptides?*
- *Can peptide descriptors be used to predict whether this workflow is suitable or not?*

- *Is there a software that can combine UV absorbance, mass spectral purity and impurity profiling?*

The approach to create this workflow for therapeutic peptides was to evaluate peptide descriptors, dissolving peptides, using UHPLC-HRMS for creating the method and considering different software for processing of raw data.

Peptides

1.1.1 Peptide Descriptors

In a high-throughput workflow for peptide characterization, it is desired to know if there already exist a suitable workflow for the peptide of interest. Calculated properties of peptides, so-called descriptors, can then be considered for the prediction.

Among the various descriptors, the partition ratio (K_D) is interesting since it describes the distribution of the analyte in an organic and aqueous phase. K_D or more commonly the logarithm, logK_D, can be calculated through various calculation programmes. Two common, substructure-based methods to calculate logK_D are ClogK_D and AlogK_D, both of which divides the molecule into smaller fragments or into single atoms which are then summed together. ClogK_D are fragment based and is more commonly used because of more correct predictions than the atom based AlogK_D method [6]. The partition ratio considers only the neutral form of the compound, to compensate for all charged states of the solute the distribution ratio (*D*) can be used instead which also can be calculated through various calculation programmes [7].

Another descriptor that describes solubility of a peptide is the aqueous solubility (*S*). It is the concentration of a compound that is solubilized in an aqueous phase during equilibrium. Thus, using this descriptor could give an alert of a very hydrophilic peptide that might not be suitable at all to run, instead elutes with the void volume. Furthermore, other useful descriptors could be the isoelectric point (*pI*) and the peptides' charge at a specific pH.

1.1.2 Peptide Characterization

Peptide characterization during drug development is required according to several regulation agencies such as the European Medicines Agency (EMA) due to the complex development procedure [8, 9]. The process includes several steps and types of analysis, there are both chemical analysis which aims more at confirmation of identity and purity, and biophysical approaches [4]. Peptide characterization for this application needs to confirm the peptide's identity and estimate its relative purity, hence, also being able to separate the peptide from structurally similar impurities. For separation, chromatographic

methods are the most predominant methods due to good separation abilities, reliability, robustness, and widely used as a routine method. High Resolution Mass spectrometry is usually applied for confirmation of identity since it determines the molecular weight by several decimals accuracy which gives a reliant identification [4, 8].

Analysis and Evaluation

1.1.3 Reversed-phase Ultrahigh-Performance Liquid Chromatography

The chromatographic mode reversed-phase (RP) is the dominant approach for peptide separation because of its robustness, flexibility and versatility. It utilizes differences in hydrophobicity for separation [4, 8]. Ultrahigh-performance liquid chromatography (UHPLC) is more advantageous in comparison to HPLC concerning resolution, throughput, and sensitivity, which allows for only a few minutes of analysis time compared to longer times required in HPLC [8]. This is suitable for complex samples, such as peptides, which have a diversity regarding structure and physicochemical properties [4]. As a stationary phase in RP peptide separations, the packing material is typically silica chemically modified with C18 derivatized silanes [4].

The choice of mobile phase composition is a crucial part for obtaining an adequate chromatographic performance of peptides. An aqueous mobile phase is appropriate with an ion-pairing reagent and a gradient of organic solvent [4]. Isocratic elution is impractical since the retention of peptides is strongly organic solvent dependent, thus, using a gradient of increasing organic solvent concentration provides a better option to succeed when having many peptides [4, 8]. Adding an ion-pairing reagent, such as trifluoroacetic acid (TFA), to the mobile phase lowers the pH and thereby protonating the weak acids and forms an ion-pairing complex with the weak bases. Therefore, secondary interactions of the free peptide with the stationary phase are mitigated [10]. This leads to an improved peak shape where a higher concentration of TFA is better [4, 10]. However, it is desired to keep the TFA concentration low since TFA decreases the MS signal. Thus, to be able to have a low TFA concentration and still contain a favourable peak shape, a column packing material with a positive charged surface between the C18 ligands can be used, for example the Charged Surface Hybrid (CSH) provided by Waters [10].

Furthermore, it is desired to perform the separation below pH 3 to prevent undesirable interactions between positively charged amino acids and free silanol groups on the surface of the stationary phase [4]. TFA is an acid which decreases the pH, in addition an acidic buffer can be added. One important parameter to consider for a high-throughput workflow is the column length where a faster analysis can be achieved with a shorter column length [9, 11]. Other important parameters are the gradient of the mobile phase, the stationary phase used, and the flow rate. A higher flow rate is desired to faster re-

condition the column and reduce the column void volume which allows for a shorter acquisition time [11], but a too high flow rate could worsen the separation if having a short acquisition time.

1.1.4 High Resolution Mass Spectrometry

Mass spectrometry (MS) is mostly used for identification purposes and has the major advantages of high sensitivity, potential for high-throughput analysis, and compatibility with chromatographic systems [4]. Thus, it can be used as a detection method for HPLC, and UHPLC, either as the only detection method or in combination with UV detection. The advantage of having multiple detection methods is that complementary information can be achieved, *e.g.*, MS can detect and separate impurities from co-eluting peaks and UV is preferrable for estimation of the relative purity [4, 12]. Relative purity estimated by MS is more substance dependent, if not calibrated correctly. For complex mixtures, such as therapeutic peptides which can have several charges and are close in mass, high resolution mass spectrometry (HRMS) is preferred [4]. Identification of the peptide occur by accurate mass analysis of the identified molecular weight. Impurity profiling means to detect and identify potential impurities in the sample. Impurities can be related peptide products from the synthesis (deleted, added, or modified amino acid, etc.) or a related degradation product (oxidation, deamidation, etc.) [12].

1.1.5 Integrating Software

Processing and reporting of the raw data is performed using a software. Different software has different features and can be applied in various ways depending on the purpose. There are several vendors for analytical LC-MS instruments, *e.g.*, Waters Corporation, Agilent Technologies, Bruker, Sciex and Thermo Scientific, which have their own hardware instruments and accompanying software. Some of the software can be used across vendors, whilst some are more vendor dependent.

The UHPLC-MS instruments used for this study were from Waters, so Waters software were considered for this report. Waters has developed several software for their LC-MS hardware instruments during the recent decades where MassLynx™ Mass Spectrometry Software came first, followed by Unifi Scientific Information System. Recently Waters Connect, with its application Intact Mass, was launched. These are three different software with different layouts and features. Raw data can be, if run with the right settings, exported and imported between these software and, therefore, all three of them can be considered for integration and evaluation.

2 Scope of the Project

The main purpose of this project was to develop an analytical workflow that could be applied on a range of AZ's in-house peptides. To be able to do this, several restrictions had to be drawn. For examples, a limited set of UHPLC parameters were considered, among those reducing the acquisition time. Hence not everything else was evaluated.

Also, the specific instrument to be used for this workflow in the future was used for the experimental part, and thus, the workflow was evolved based on this. Furthermore, only commercially available or already published therapeutic peptides were chosen since this report will be published externally. Regarding investigating peptide descriptors, no chemometrics to see potential correlations between the descriptors or inclusion of external data were performed, due to time constraints.

There exists multiple software that are compatible independently of vendor of the hardware instrument. Some that are open source and some that are proprietary (*e.g.*, Byonic™ from Protein Metrics, and General Protein/Mass Analysis for Windows (GPMAW) from Lighthouse data). Thus, these could have been investigated, but were decided to be out of scope.

3 Material and Methods

Chemicals

For dissolving peptides, tert-butanol and acetic acid from Sigma-Aldrich was used. The mobile phase consisted of acetonitrile and milli-Q water, obtained from a Milli-Q water instrument from Merck. Trifluoroacetic acid (TFA) was used as an ion-pairing reagent and dimethyl sulfoxide (DMSO) as blank samples. More information of chemicals can be found in Table 1. The peptides used in the experimental work are summarized in Table 2. In Appendix A the amino acid sequence, monoisotopic molecular weight, structure formula, etc. can be seen.

Table 1 *Chemicals used in the experimental part.*

Acetic acid	ReagentPlus®, ≥ 99%	Sigma-Aldrich
Acetonitrile	Optima® LC/MS grade	Fisher Scientific
DMSO	LC-MS grade, ≥ 99.7%	Thermo Scientific
tert-Butanol	Anhydrous, ≥ 99.5%	Sigma-Aldrich
Trifluoroacetic acid (TFA)	99%	Alfa Aesar

Table 2 *Peptides analysed for the experimental part.*

Argipressin	CAS 113-79-1	MedChemExpress
Atosiban	CAS 90779-69-4	In-house synthesized
Bacitracin	CAS 1405-87-4	Sigma-Aldrich
Bivalirudin	CAS 128270-60-0	MedChemExpress
Carbetocin	CAS 37025-55-1	MedChemExpress
Cyclic somatostatin	CAS 38916-34-6	MedChemExpress
Deslorelin	CAS 57773-65-6	MedChemExpress
Desmopressin	CAS 62288-83-9	MedChemExpress
(Deamino-Cys1,Val4,D-Arg8)-Vasopressin	CAS 43157-23-9	Bachem
Felypressin	CAS 56-59-7	MedChemExpress
GNRH	CAS 34973-08-5	In-house synthesized
Leuprolide acetate	CAS 74381-53-6	In-house synthesized
ONEG-peptide	---	In-house synthesized
Oxytocin	CAS 50-56-6	Key Organics
PLP-peptide	---	In-house synthesized
Ser-Asp-Gly-Arg-Gly	CAS 108608-63-5	Sigma-Aldrich
Teriparatide	CAS 53332-67-4	MedChemExpress
TP2-peptide	---	In-house synthesized
Triptorelin	CAS 57773-63-4	Bachem
Vancomycin hydrochloride	CAS 1404-93-9	Sigma-Aldrich

Instrumentation

Two analytical UHPLC-HRMS instruments were used. A Waters ACQUITY Synapt G2-Si UPLC system connected to both a PDA detector and a quadrupole time-of-flight (QTOF) mass spectrometer, and a Waters ACQUITY RDa UPLC system with an RDa TOF mass spectrometer as well as a TUV detector. Both are TOF instruments, but the RDa instrument has a more automated start-up, no collision cell, and is more robust and user friendly for non-experts.

Methods

The different parts of this project included software tools to obtain descriptor values, solubilization of peptides, the development and optimization of the analytical workflow, and integration of software.

3.1.1 Descriptor Values

The descriptor values were acquired using an AZ application where first a search tool of the desired properties had to be constructed. From this, calculated values of ClogK_D, AlogK_D, logD, aqueous solubility and pI for each peptide could be obtained. The search tool was designed to be used for future application where only an internal AZ number for each peptide need to be used as input. For the descriptor values of the peptides for this report, see Appendix B.

In addition, another application was used to obtain the charge of the peptide at a specific pH (see Appendix B). For this study, the pH was approximately 2 for the mobile phase, thus charge at pH 2 was chosen. The application used for this is called pIChemiSt, which is developed by AZ for peptides. This tool is an open source and described in Frolov, *et.al* [13]. An example of one of the peptides named Felypressin is seen in **Figure 1**.

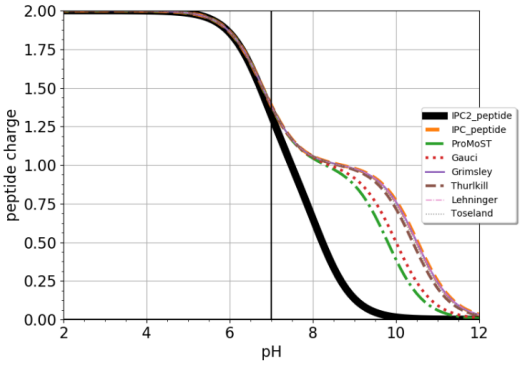


Figure 1 *Plot of peptide charge vs pH for Felypressin. Obtained from pIChemiSt.*

3.1.2 Dissolving of Peptides

The peptides used were either already in solution and stored in a freezer (- 20 °C), or in a solid powder from the purchaser also stored in a freezer (- 20 °C) and these needed to be solubilized. The solid

peptides were dissolved in accordance with the instructions given from Sigma-Aldrich[1] and the procedure was the following:

1. A small amount of solid was dissolved in approximately 1 mL of solvent in a glass vial aiming at a concentration of 1 mg/mL for a good value of absorbance. Glass vials were used to see how well the peptides had been dissolved.
2. UV sonication using a Bandelin SONOREX for about 10 min to make sure they got fully dissolved.

The solvent used was tert-butanol, but if the peptide had not been solubilized after the sonication step, then a small amount of milli-Q water or acetic acid was added. If a peptide had too high UV absorbance, it was diluted in a mixture of ACN/Milli-Q (1:1). For the high-throughput method later, the peptides for analyses will already be dissolved in a suitable solvent. Likewise, the concentration will be set in advance to a specific value to give an adequate absorbance.

3.1.3 Analysis Method Developing and Optimization

The peptides were analysed on either a Waters ACQUITY Synapt G2-Si UPLC instrument, with an ACQUITY UPLC CSH C18 column (1.7 µm, 2.1 x 100 mm), or they were analysed on a Waters ACQUITY RDa UPLC instrument, with an ACQUITY UPLC BEH C18 column (1.7 µm, 2.1 x 50 mm) or an ACQUITY UPLC BEH C18 column (1.7 µm, 2.1 x 50 mm). For more details, see Table 3.

Table 3 *LC-MS methods. For all methods the column temperature was 40°C, absorbance at 214 nm, and ESI+ mode and the mobile phases buffered with a mixture of formic acid and ammonium formate (pH 3), except for methods 2 and 3 where it is with only formic acid. The gradient start at 0.5 min for the RDa detector methods and at 1.0 min for the Synapt G2-Si method.*

Method	Instrument	Column	Mobile phase	Flow rate (mL/min)	Gradient (%B)	Gradient time (min)	Slope (%B)
1	Synapt G2-Si	CSH 100 mm	A: 100% MQ, 0.03% TFA B: 95% ACN, 0.03% TFA	0.5	10-60	7	7%/min
2	RDa detector	BEH 50 mm	A: 100% MQ + 0.1% FA B: ACN + 0.1% FA	0.4	10-60	3.5	14%/min
3	RDa detector	BEH 50 mm	A: 100% MQ + 0.1% FA B: ACN + 0.1% FA	0.4	10-60	2,4	21%/min
4	RDa detector	CSH 50 mm	A: 100% MQ, 0.03% TFA B: 95% ACN, 0.03% TFA	0.4	10-60	3.5	14%/min
5	RDa detector	CSH 50 mm	A: 100% MQ, 0.03% TFA B: 95% ACN, 0.03% TFA	0.5	10-60	3.5	14%/min
6	RDa detector	CSH 50 mm	A: 100% MQ, 0.03% TFA B: 95% ACN, 0.03% TFA	0.5	5-55	3.5	14%/min

[1]https://www.sigmaaldrich.com/SE/en/technical-documents/technical-article/research-and-disease-areas/cell-and-developmental-biology-research/solubility-guidelines

The methods were based on a generic method from preliminary studies (data not shown here), which was further optimized based on the results obtained during this study. On the RDa instrument a shorter column was used, to be able to shorten the acquisition times. The impact of the slope of the gradient, two different flow rates, and a method for more hydrophilic peptides were investigated.

3.1.4 Software Evaluation

The comparison of software was mostly theoretically considered, knowing what features they have, and to some extent practically, *e.g.,* ease of use and user experience. The focus was on the new application Waters Connect and its sub-application Intact Mass. For this, a method for processing raw data had to be constructed and furthermore optimized to fit our purposes. This included optimizing parameters such as minimum area and height for integration of peaks, based on either optical or total ion current (TIC) integration or both. This was evaluated in a team which consisted of both AZ and Waters colleagues.

3.1.5 Calibration of Instruments

To confirm a reliable outcome, the instruments needed to be calibrated. As for the RDa instrument, it has a build-in calibration but for the Synapt G2-Si a standard suitability test was run in the beginning to confirm it detected correctly. Additionally, on the Synapt G2-Si, purge and prime of the sample syringe was performed to reduce potential errors. For both instruments, every time a new sample set was to be run, two or three blanks were run to make sure the system was adequately conditioned and stable. The relative purity of the samples was considered, and not the absolute, so no calibration function for quantification was needed. The purity was defined as the relative absorbance at UV 214 nm.

4 Results and Discussions

Overall, an analytical high-throughput workflow was created with the aim of characterization of peptides. Initially, a generic chromatographic method was applied for the peptides which was further optimized considering the column, organic solvent gradient and flow rate. Integration and evaluation of software as well as the usefulness of peptide descriptors were also considered.

Confirmation of the peptide's identity for all samples was performed through comparing the experimentally detected accurate mass value, with the theoretical exact monoisotopic mass. This was performed on the Synapt G2-Si instrument and on the RDa instrument. The identity could be confirmed for all peptides except Deslorelin, which had been stored as a solution in the freezer and is likely to have been degraded. A new sample of Deslorelin from solid was prepared. This sample could be observed to have the expected exact mass.

Development and Optimization of the Workflow

The generic method tested (method 2 in **Table 3**) with a BEH column, gradient with a slope of 14%B/min and gradient time of 3.5 min, and flow rate of 0.400 mL/min was observed to retain most of the peptides. The retention times (RT) from this method are visualised in **Figure 2**. The peptide (Deamino-Cys1,Val4,D-Arg8)-Vasopressin is abbreviated as d-vasopressin. The UV chromatograms of the different peptides analysed with method 2 can be seen in Appendix C.

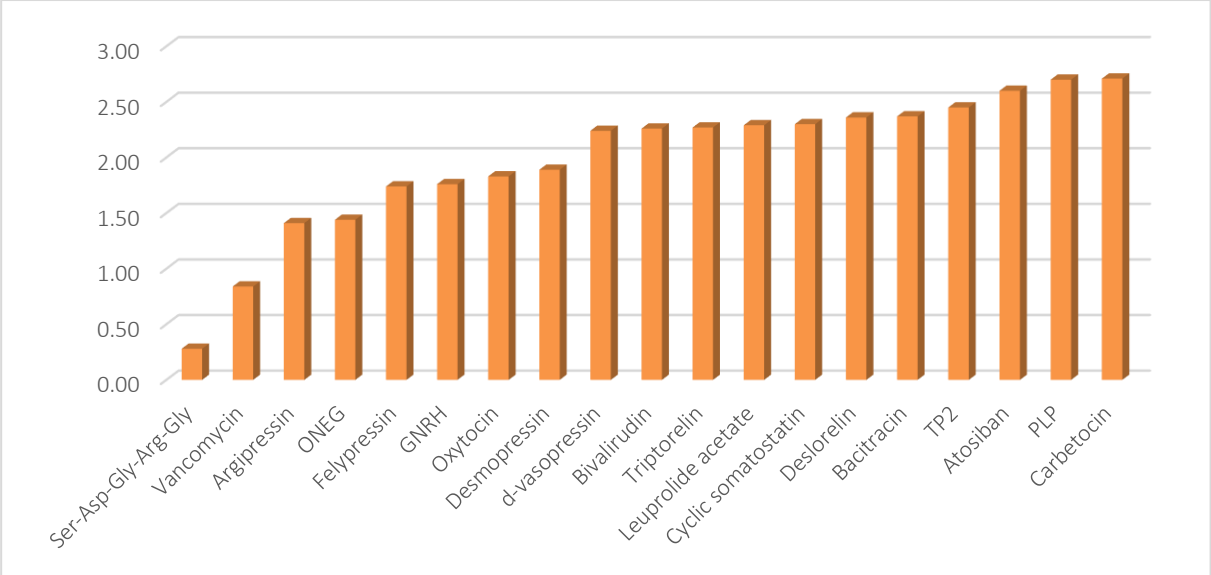


Figure 2 *Retention time of peptides in method 2 (BEH column, 50 mm, 14%B/min) ordered from shortest to longest retention time. All peptides were eluted between 1.0 min and 3.0 min, except for Vancomycin at 0.8 min and the pentapeptide Ser-Asp-Gly-Arg-Gly at 0.3 min.*

For an isocratic elution the retention factor (*k*) can be used to determine whether the retention time is within an appropriate range. For a gradient elution the gradient retention factor can be used instead

(k*) but was not included here. Instead, it was concluded that the interval of these peptides' retention time, except the pentapeptide's, can be assumed to be appropriate since the gradient in method 2 was set to be from 0.5 min to 4.0 min.

A shorter retention time (in reversed-phase) means that the compound is more hydrophilic, it is not retained as much on the C18 stationary phase. Then it can be concluded that the pentapeptide was very hydrophilic since it was not retained at all, it eluted with the void volume (around 0.3 min). Thus, only considering the RT, this method can be used for all the included peptides except the pentapeptide.

The second thing to evaluate, apart from appropriate retention times, was whether this method has succeeded to separate potential impurities. The risk with having a shorter gradient time with steeper gradient slope is that the separation of related product impurities might not succeed as well as with a longer gradient. In comparison with the outcome from the Synapt G2-Si instrument (**Figure 3**) with a 100 mm column and t_g = 7 min, a good separation of impurities was still achieved using the RDa instrument. Therefore, it can be concluded that a shorter column (50 mm) and shorter gradient time (t_g = 3.5 min) can still separate impurities to a large extent, enough for this application. An even higher degree of separation of impurities is not needed since the samples are expected to be pure.

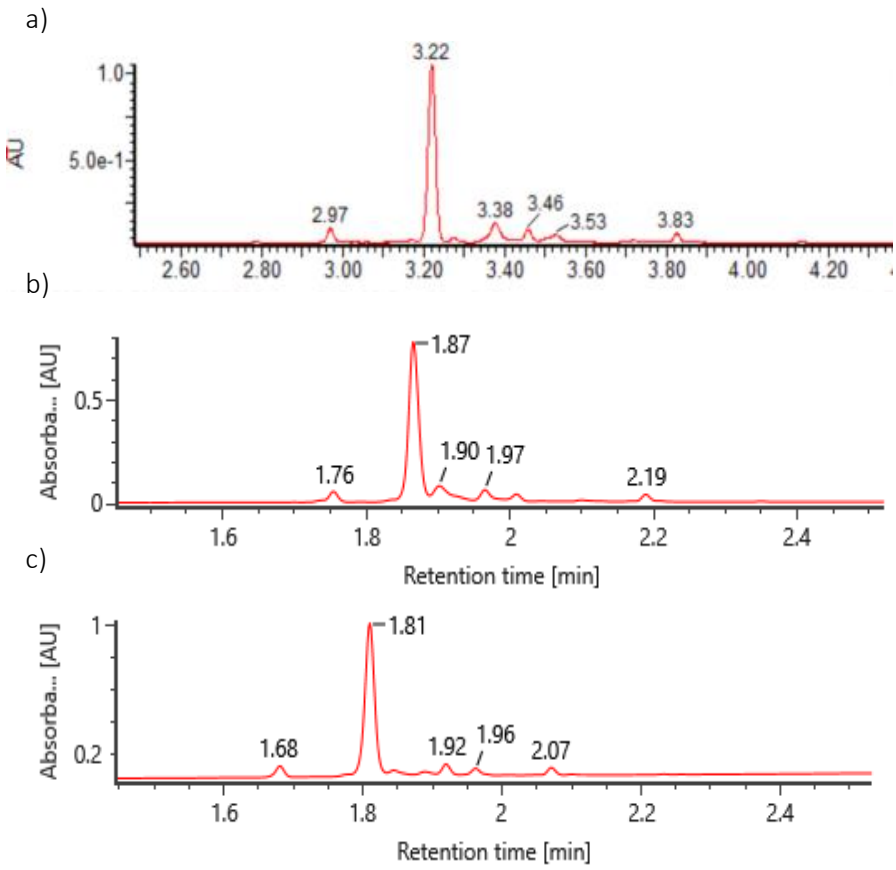


Figure 3 *Impurities of Desmopressin. a) Synapt G2-Si instrument, CSH column, 100 mm. b) RDa instrument, BEH column, 50 mm. c) RDa instrument, CSH column, 50 mm.*

4.1.1 *Gradient Elution*

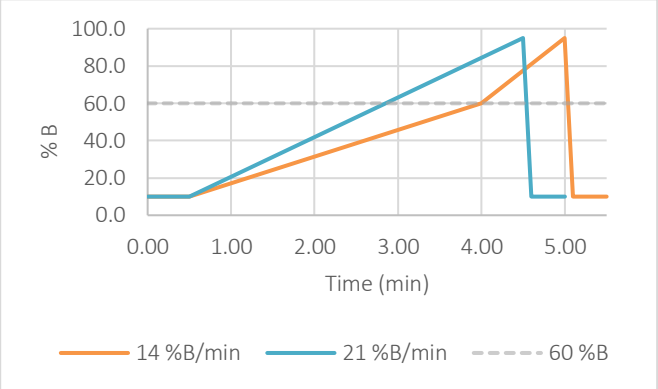


Figure 4 *Comparison of two slopes of 10-60 %B gradients with BEH column of 50 mm, flow rate 0.4 mL/min and mobile phase composition A: 100% MQ + 0.1% FA and B: ACN + 0.1% FA.*

Separations of peptides with LC is very dependent on the mobile phase composition. Therefore, the gradient used will likely be an important factor for how well the separation has succeeded. A comparison between two 10-60 %B gradient slopes was performed (**Figure 4**). One gradient slope of 14%B/min and the other at 21 %B/min, both starting at 0.5 min and then ending at 4.0 min and 2.9 min respectively, giving the latter gradient a steeper slope. They were performed on the RDa instrument with a BEH column (method 2 and 3). In the comparison all peptides except Teriparatide was used since data for this was missing for method 2 and 3.

The retention times using the higher slope of gradient (21%B/min) are shorter compared to the lower slope (14%B/min), see the blue and orange columns in **Figure 5**. Where the largest difference was seen for the longer retention times, which is reasonable since the longer time the earlier a certain %B value is reached (**Figure 4**). By comparing the RT of these two slopes of gradients, no clear conclusions can be drawn since both have a majority of peptides with adequate retention times. Thus, the separation of impurities must be considered as well.

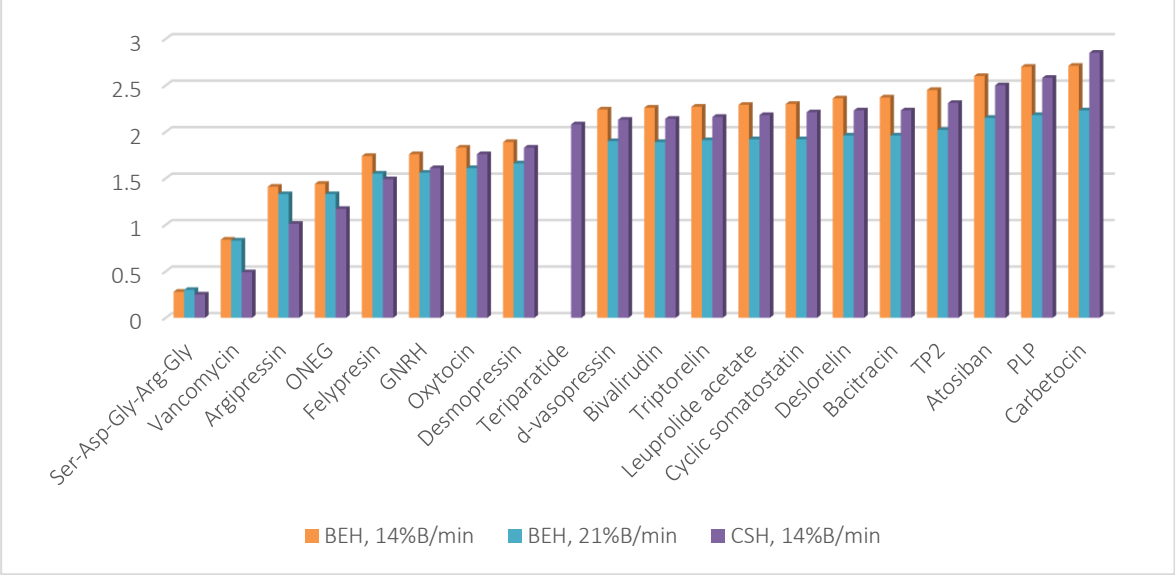


Figure 5 *Retention times of peptides for different methods. Flow rate 0.4 mL/min. For the BEH column the mobile phase composition was A: 100% MQ + 0.1% FA and B: ACN + 0.1% FA and for the CSH column A: 100% MQ, 0.03% TFA and B: 95% ACN, 0.03% TFA.*

In **Figure 6**, the peptides Desmopressin and Felypressin are seen. The rest of the UV chromatograms are found in Appendix D. The 14%B/min gradient gave a better separation of impurities, resulting in more peaks than the 21%B/min gradient. As for Desmopressin, a new peak to the right of the product peak was seen for the 14%B/min gradient, but not for the 21%B/min gradient. For Felypressin, the 21%B/min gradient has a separation resulting in two peaks where the shorter gradient slope gives four peaks, which indicates a better separation. For the other peptides analysed, the separation was the same or somewhat better (for example Atosiban and Leuprolide acetate have a better separation of impurities).

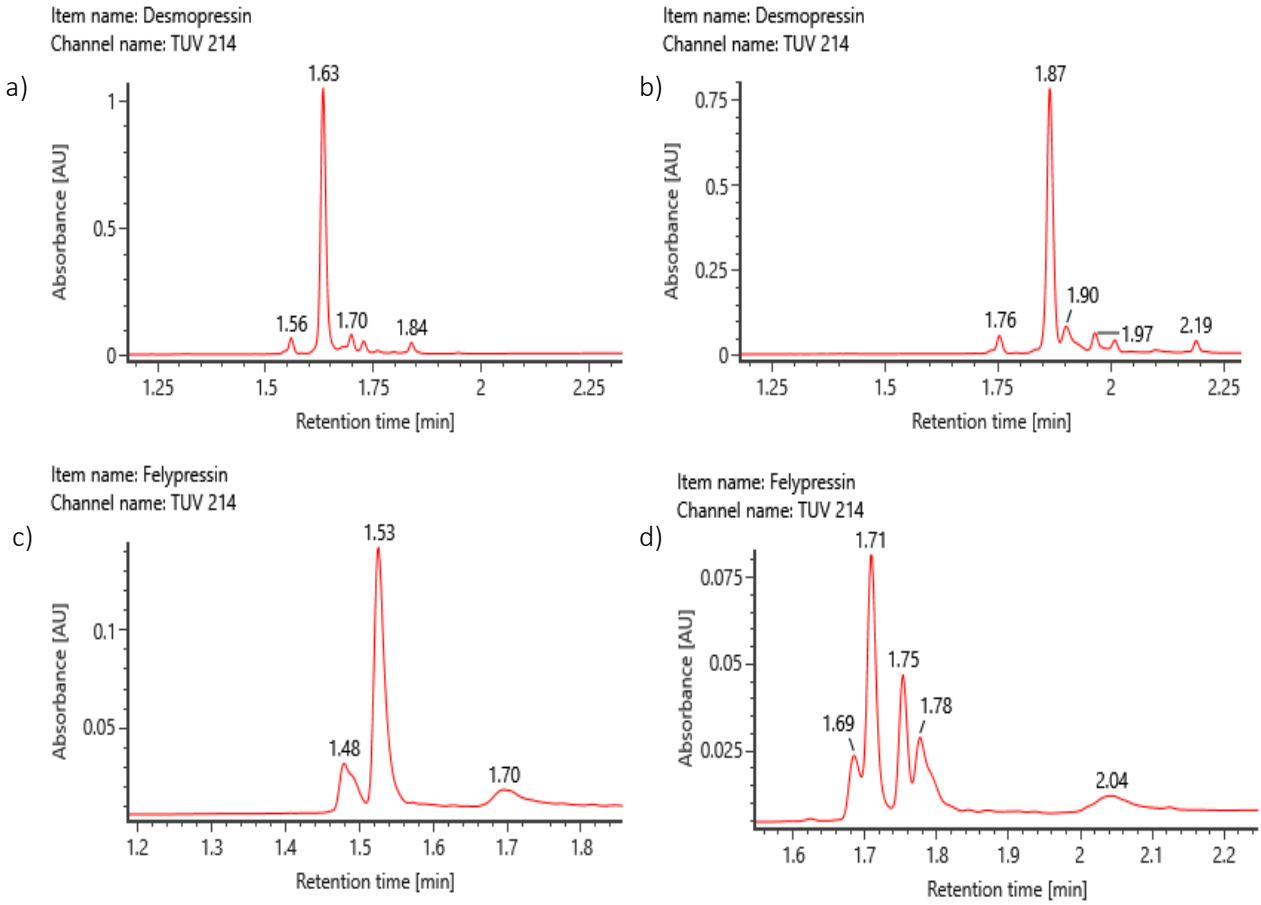


Figure 6 *Comparison of slopes of gradients with BEH column, mobile phase composition was A: 100% MQ + 0.1% FA and B: ACN + 0.1% FA. a) Desmopressin, 21%B/min. b) Desmopressin, 14%B/min. c) Felypressin, 21%B/min. d) Felypressin, 14%B/min.*

To evaluate the performance of a gradient separation, the peak capacity (*PC*) was considered the best according to Neue, *et.al.* [14]. Peak capacity is a measure of separation efficiency and can be calculated in several ways [11, 14]. A simplified version with the purpose to be able to compare methods, not determine absolute values, was used here (Equation 1).

$$PC = \frac{t_g}{w_{FWHM}} \qquad (1)$$

Where *PC* is the peak capacity, t_g the gradient time determined from the time the gradient starts until it ends, and w_{FWHM} is the full width at half maximum. The peak capacity for Desmopressin and Felypressin is shown in **Table 4** and the largest value of peak capacity was seen for the 14 %B/min gradient for both

peptides. Maximizing the peak capacity is important since that means that a higher number of peaks can be resolved. Thus, the 14%B/min gradient is the better in terms of peak capacity.

The advantage of a steeper slope of the gradient, which is equal to having a shorter gradient time, is that shorter retention times can be achieved, which was confirmed here. This is beneficial from a high-throughput perspective, since more samples can be run in a short time. However, the risk with having a shorter gradient time is that separation of potential impurities is not performed as well, as was illustrated in the above examples for Desmopressin and Felypressin. Hence, it was a trade-off between optimizing for high separation and an efficient method. Considering that a higher peak capacity was achieved with the 14%B/min gradient which has short enough acquisition time and that a visually better separation was observed, it was decided to continue with this gradient for the analytical LC method in the workflow.

Table 4 Summary of peak capacity for comparison of methods with BEH column of 50 mm, mobile phase composition A: 100% MQ + 0.1% FA and B: ACN + 0.1%. FA

	Gradients	
	21 %B/min	**14 %B/min**
Gradient time (min)	2,4	3,5
Desmopressin		
width at FWHM	0,028	0,031
PC	86	113
Felypressin		
width at FWHM	0,027	0,028
PC	89	125

4.1.2 Column

Two different C18 columns were evaluated, named BEH and CSH. They had the same dimensions (1.7 µm, 2.1 x 50 mm) and were both from Waters to be applied in an ACQUITY UPLC system. The comparison was performed between method 2 and 4 (Table 3), with a flow rate of 0.4 mL/min and gradient of 14 %B/min from 0.5 to 4.0 min. For the BEH column the mobile phase composition was A: 100% MQ + 0.1% FA and B: ACN + 0.1% FA and for the CSH column A: 100% MQ, 0.03% TFA and B: 95% ACN, 0.03% TFA.

In Figure 7, showing the peptides Desmopressin and Felypressin, a better separation was observed with the CSH column in combination with having TFA in the mobile phase. The peaks for these two peptide samples are more drawn out for the CSH analysis compared to the BEH analysis. This could be explained by the CSH column provides a better peak shape by decreasing secondary interactions by having the positive charge in between the C18 groups on the stationary phase surface. As well as TFA improves the peak shape.

As for the rest of the peptides (Appendix D), there was an improvement in separation of impurities with the CSH column for some of the peptides. For other peptides, no major difference in peak shape was observed. For the peptide TP2, a small peak has disappeared for the CSH column, which was present in the BEH column, and for Vancomycin the adjacent peaks have come closer together. Thus, this means that for most of the peptides an improvement was seen, a few of the peptides were unaffected of the column change, and for two of the peptides a slightly worse result was obtained. This correlates to some extent with the result obtained in the article *The importance of ion-pairing in peptide purification by*

reversed-phase liquid chromatography by the authors Åsberg, *et.al.* (2017), where two peptides were analysed and for one of them an improvement of peak shape with the CSH column was seen.

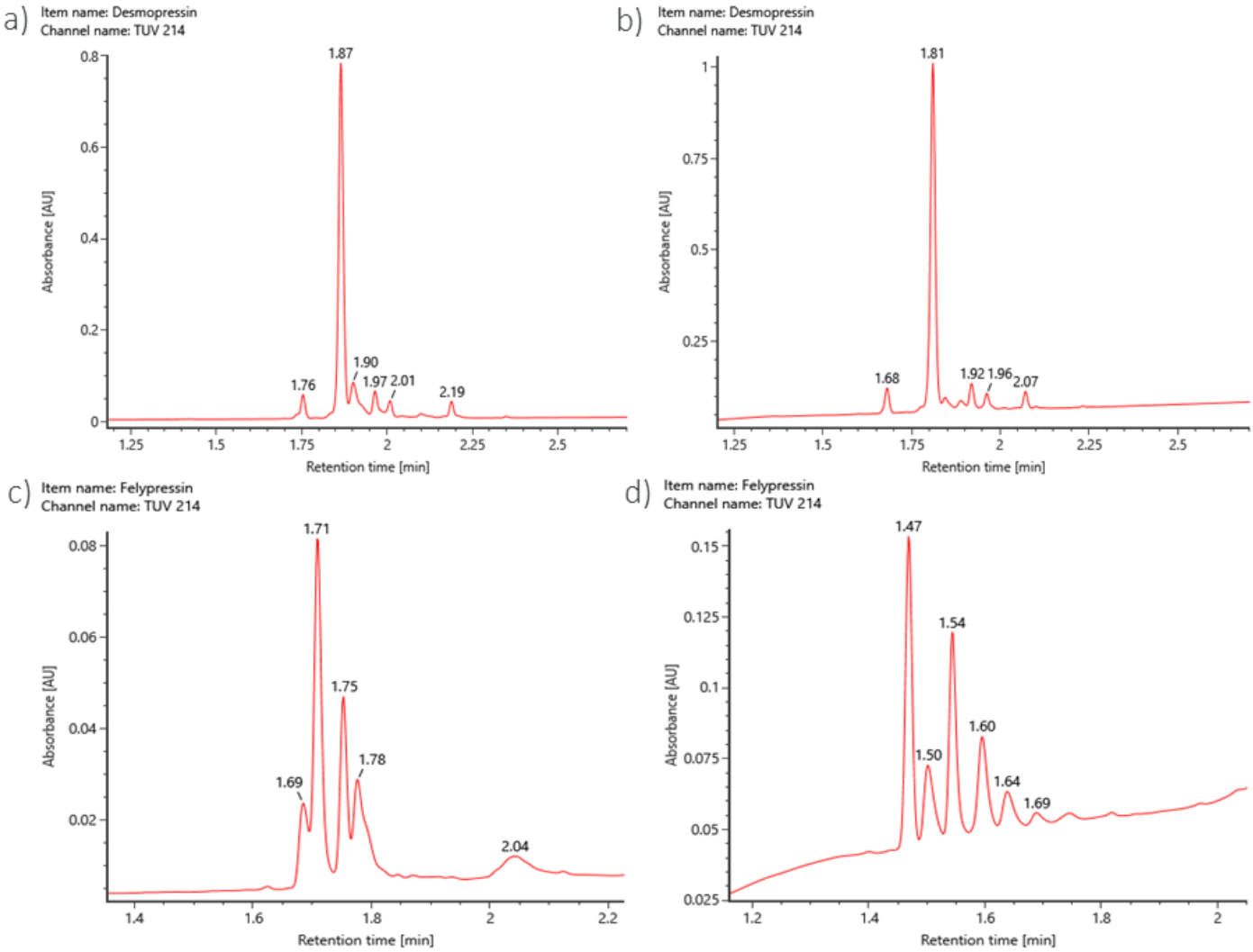


Figure 7 *Comparison of stationary phase, gradient 14 %B/min. For the BEH column the mobile phase composition was A: 100% MQ + 0.1% FA and B: ACN + 0.1% FA and for the CSH column A: 100% MQ, 0.03% TFA and B: 95% ACN, 0.03% TFA. a) Desmopressin, BEH column. b) Desmopressin, CSH column. c) Felypressin, BEH column. d) Felypressin, CSH column.*

The retention times for the peptides with the CSH column in combination with mobile phase containing TFA were slightly shorter or almost equal to the ones with the BEH column (**Figure 5**). Remarkably lower RT was achieved for Vancomycin, which elutes at the time of when the gradient starts. Thus, this peptide should be considered more carefully in the future, to make it retain better.

When evaluating the performance of the separation by the peak capacity (**Table 5**) it was seen that the peak capacity was the same for Desmopressin. For Felypressin a higher value was obtained for the CSH column.

To conclude, the CSH column in combination with TFA in the mobile phase was considered the better choice for this analytical workflow mainly since it theoretically has an advantage for peptide separations over the BEH column due to preventing secondary interactions and thus improving peak shape.

Table 5 *Peak capacity of Desmopressin and Felypressin for the BEH and CSH column. For the BEH column the mobile phase composition was A: 100% MQ + 0.1% FA and B: ACN + 0.1% FA and for the CSH column A: 100% MQ, 0.03% TFA and B: 95% ACN, 0.03% TFA.*

	Columns	
	BEH	CSH
Gradient time (min)	3,5	3,5
Desmopressin		
width at FWHM	0,031	0,031
PC	113	113
Felypressin		
width at FWHM	0,028	0,020
PC	125	175

4.1.3 Flow Rate

The conclusion for the analytical workflow was to use a 14%B/min gradient with a gradient time of 3.5 min together with the CSH column and TFA in the mobile phase. With these settings, a method was achieved that separated the peptide well from its impurities. One parameter left that was desired to increase was the flow rate. A higher flow rate re-conditions the column faster which minimises the risk of having residual compounds that will affect the next sample. Therefore, an increase from 0.4 mL/min to 0.5 mL/min was evaluated.

This was performed on the RDa instrument, with the 14%B/min gradient on the CSH 50 mm column (method 4 and 5, Table 3). Since the developed method was concluded since earlier to perform well on the peptides, only four peptides with impurities were chosen to be run to see potential changes of separation, the other were considered to perform well with the higher flow rate. This to save solvent consumption.

When comparing the peak capacity (Table 6) it was seen that increasing the flow rate increased the peak capacity for the peptides Desmopressin, Felypressin and Bacitracin. As for Vancomycin the peak capacity was reduced, but since Vancomycin was eluting earlier, at the void volume for the higher flow rate, this value may be incorrect.

Table 6 *Comparison of peak capacity for flow rate with CSH column of 50 mm, mobile phase composition was A: 100% MQ, 0.03% TFA and B: 95% ACN, 0.03% TFA.*

	Flow rate	
	0.4 mL/min	0.5 mL/min
Gradient time (min)	3,5	3,5
Desmopressin		
width at FWHM	0,031	0,027
PC	113	130
Vancomycin		
width at FWHM	0,023	0,024
PC	152	146
Felypressin		
width at FWHM	0,020	0,017
PC	175	206
Bacitracin		
width at FWHM	0,018	0,016
PC	194	219

The UV chromatograms for the peptides is seen in Figure 8 where a higher flow rate reduces the RT, which is beneficial for a high-throughput method. The separation was not affected significantly, the spectra were similar except that a higher flow rate lengthens the distance between the peaks (Desmopressin and Felypressin). Thus, selecting 0.5 mL/min as a flow rate was appropriate. Although, care need to be taken for quick-eluting peaks, such as Vancomycin, since this was eluting at the front with 0.5 mL/min.

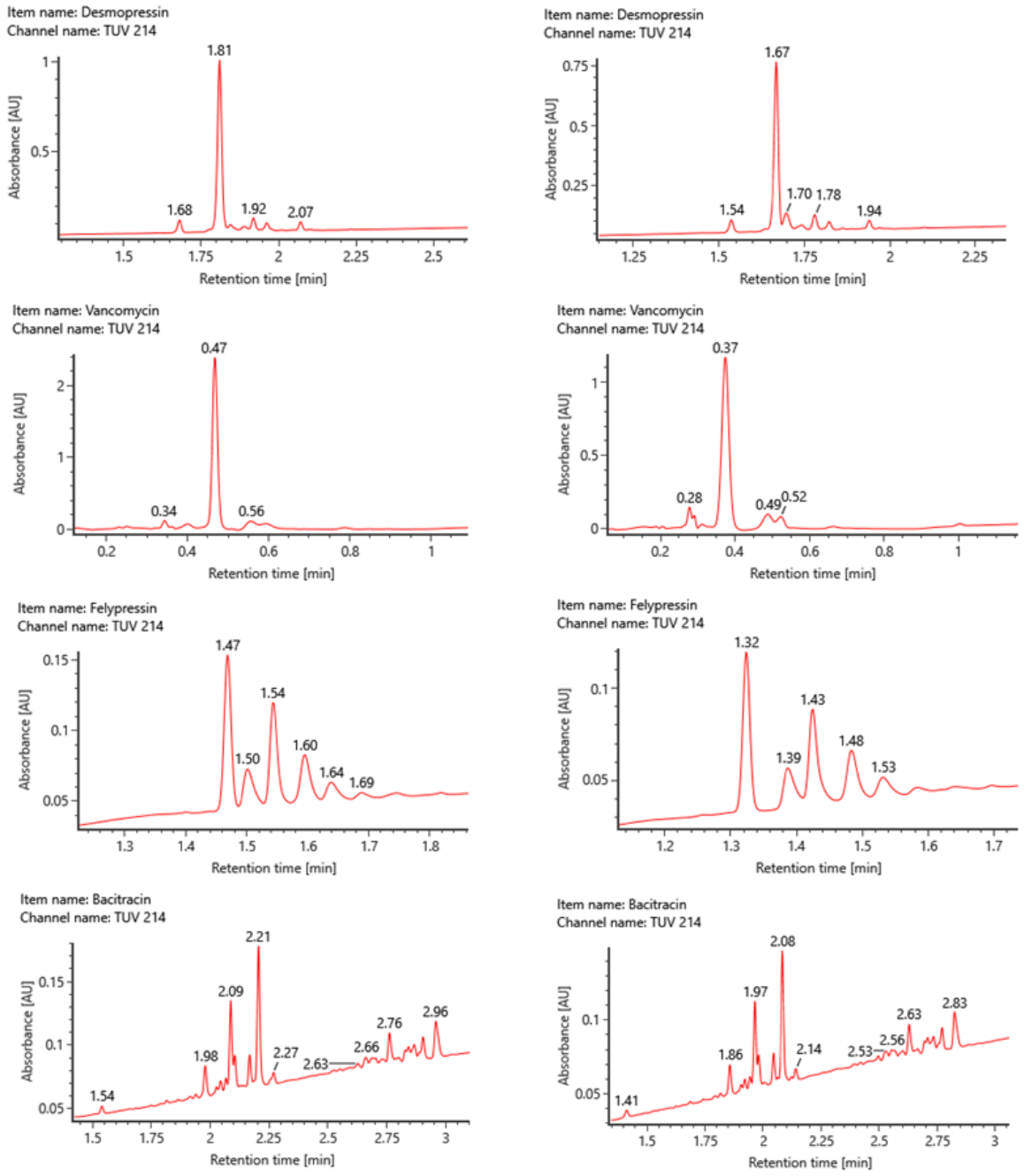


Figure 8 *Left column has a flow rate of 0.4 mL/min and right column a flow rate of 0.5 mL/min. Column: CSH, 50 mm. Mobile phase: A: 100% MQ, 0.03% TFA and B: 95% ACN, 0.03% TFA.*

4.1.4 Hydrophilic Compounds

Since peptides consist of amino acids, there is likely a chance of encounter more hydrophilic compounds than very hydrophobic compounds, for example Vancomycin and the pentapeptide. To create a more hydrophilic method, the starting concentration of the gradient was decreased to 5 %B instead of 10 %B (method 6 in **Table 3**). The risk of lowering the organic solvent concentration in the mobile phase is that less hydrophilic peptides can become less soluble and precipitate in the capillaries of the system.

The two peptides eluting earliest (the pentapeptide and Vancomycin) were included, as well as a third peptide (Desmopressin) for comparison of methods. They were run on the RDa instrument, with the CSH column and a flow rate of 0.5 mL/min. The UV chromatograms can be seen in Figure 9.

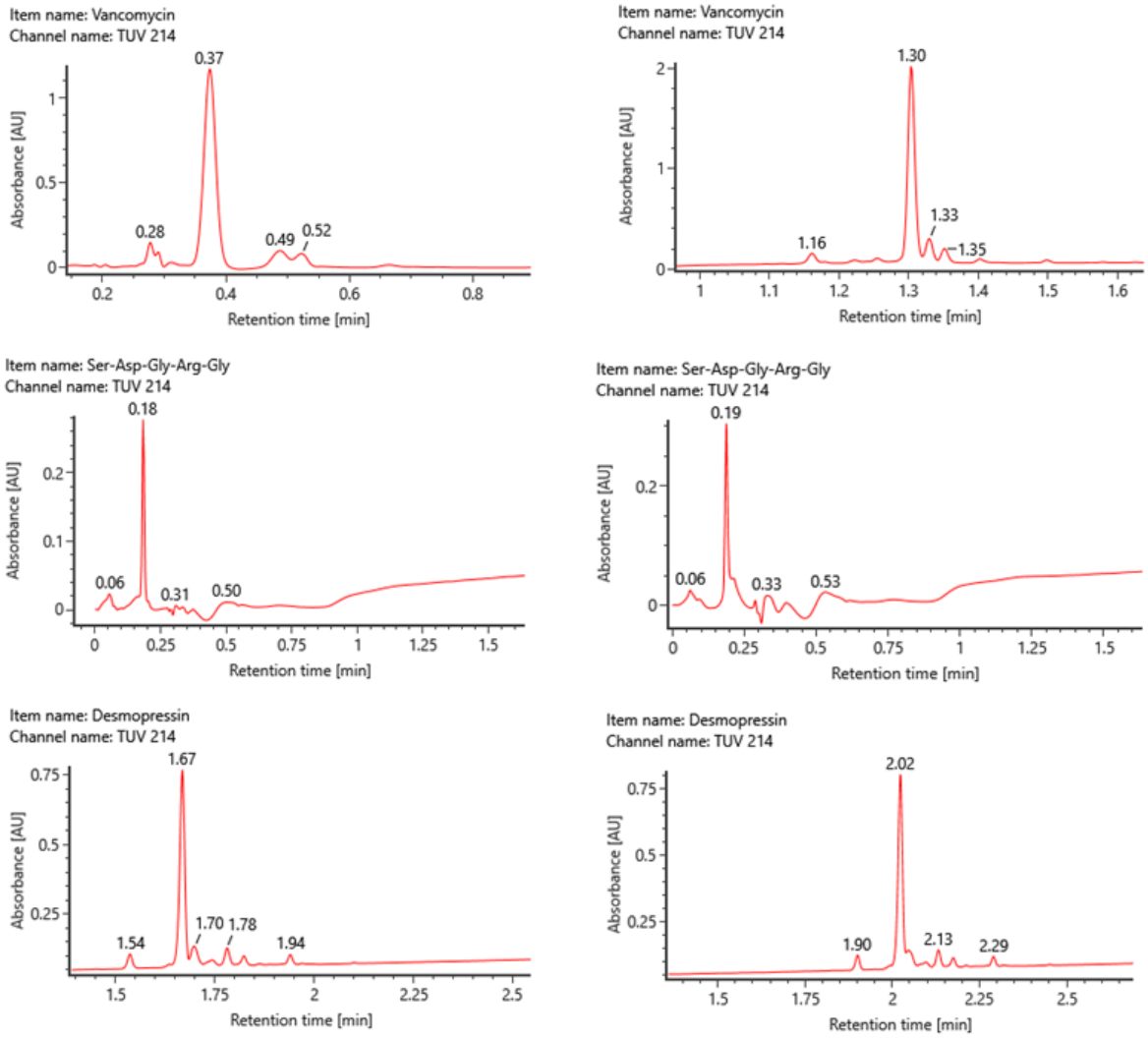


Figure 9 *Left column: gradient 10-60 %B. Right column: gradient 5-55 %B. Using a CSH column of 50 mm and mobile phase composition of A: 100% MQ, 0.03% TFA and B: 95% ACN, 0.03% TFA. Vancomycin and Desmopressin were eluted later and the pentapeptide was still not retained.*

The retention time was increased for Vancomycin, it did not elute with the void volume longer which was the aim. Desmopressin had also an increased retention time. As for the pentapeptide it was still not retained, probably due to the peptide being too hydrophilic. The peak capacity was better for Vancomycin, but for Desmopressin there was a small decrease (Table 7). As a conclusion, the 5-55 %B method is an appropriate solution for more hydrophilic peptides. However, this is dependent on their hydrophilicity, too

Table 7 *Comparison of peak capacity for the more hydrophilic method using a CSH column of 50 mm, mobile phase composition A: 100% MQ, 0.03% TFA and B: 95% ACN, 0.03% TFA.*

	Hydrophilic method	
	10-60 %B	5-55 %B
Gradient time (min)	3,5	3,5
Vancomycin		
width at FWHM	0,024	0,021
PC	146	167
Desmopressin		
width at FWHM	0,027	0,028
PC	130	125

hydrophilic compounds, such as the pentapeptide, cannot be retained but as for Vancomycin this method is an appropriate option.

Evaluation of Peptide Descriptors

Peptide descriptors were investigated for two reasons. 1) If they can be used to obtain an alert if a peptide is very hydrophilic or lipophilic, the latter can indicate that the peptide is not easily dissolved and thus can create stop in the system. 2) If they can be used to predict if the created 10-60 %B method is suitable to use, or if the 5-55 %B method should be used instead.

Several descriptors were investigated (Appendix E). The descriptors ClogK_D and aqueous solubility (S) described the retention time better than the other descriptors included. These two descriptors were plotted against each other (**Figure 10**), and limits for very hydrophilic or hydrophobic peptides could be set from this. The limit for a too hydrophilic peptide was based on the descriptor values of the pentapeptide since this was not retained at all for the created methods (both the ClogK_D and aqueous solubility limit). As for the ClogK_D limit for a too lipophilic peptide, it was set to be shortly above the value of the peptide with highest ClogK_D value (TP2, ClogK_D = 0.3).

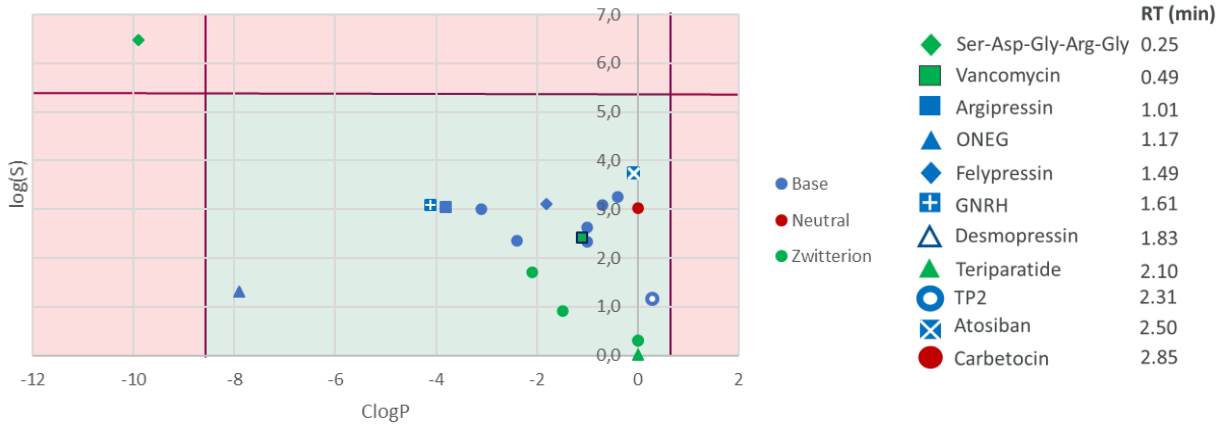


Figure 10 *Plot of two descriptors that describes lipophilicity. A few peptides are highlighted due to having interesting properties, see legend to the right, the other are visualised by a blue circle. The limits for obtaining an alert are marked in red. The retention times are from the method with a 14 %B/min gradient from 10-60 %B and a flow rate of 0.4 mL/min, CSH column of 50 mm, and mobile phase composition of A: 100% MQ, 0.03% TFA and B: 95% ACN, 0.03% TFA.*

Secondly, it was desired to be able to predict which of the two created gradient methods that is suitable to use for future analysis. All peptides except Vancomycin could be analysed with the 10-60 %B method, which needed to be analysed with the 5-55 %B method to not elute with the void volume. Dividing the peptides based on their ion class it was shown that limits for which method to use should be based on the peptide's ion class separately (**Figure 11**). Setting a special limit for the zwitterions and another for the bases can separate Vancomycin from the other retained peptides. For the limitations see Appendix F.

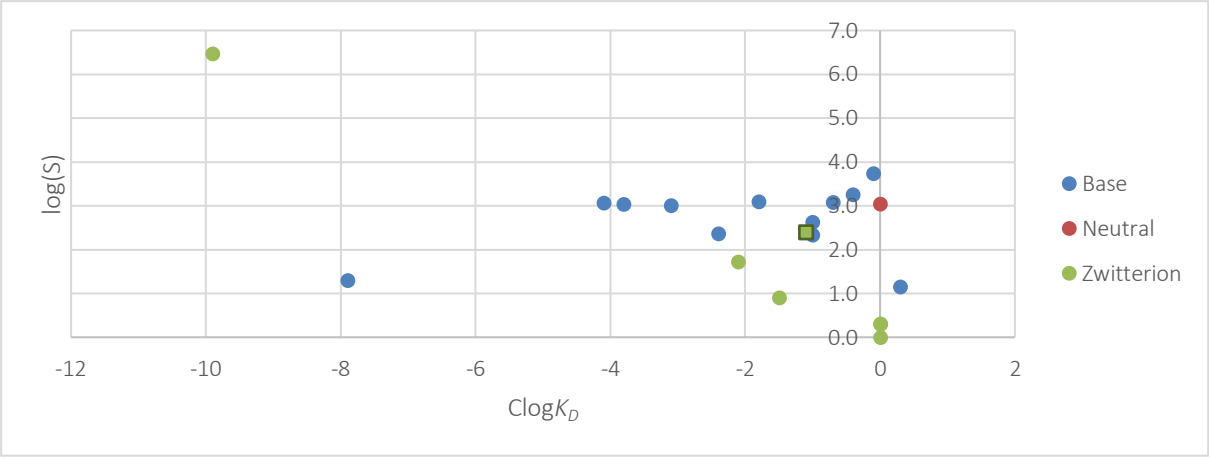


Figure 11 *Peptides were based according to their ion class. Vancomycin is visualised with a green box with a darker border.*

Integration and Evaluation of Software

The application Intact Mass provided in Waters Connect was chosen for processing of raw data acquired by the RDa instrument. This application can in a clear overview and with high accuracy give the analyst information about accurate confirmation of identity and the samples relative purity (**Figure 12**). The limits of pass and warning for identification of mass and relative purity is set by the user.

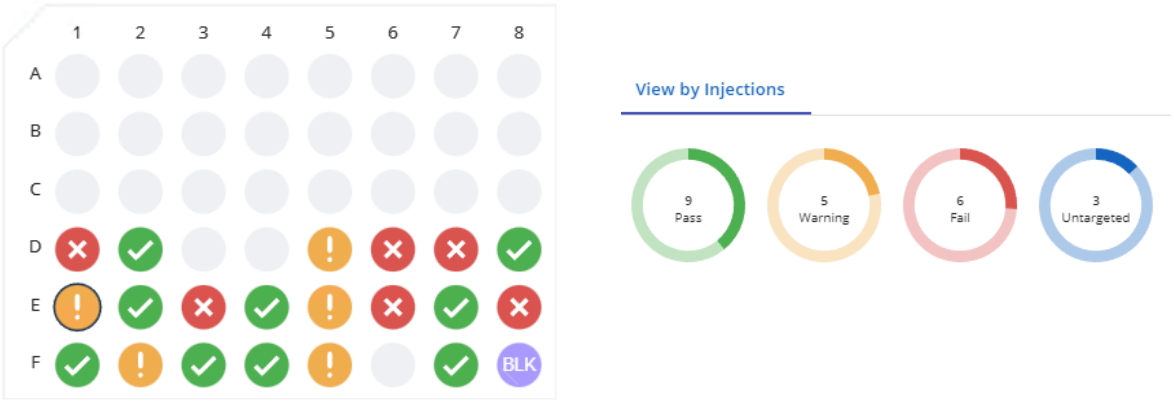


Figure 12 *An overview of the visualisation of the result from a plate with the Intact Mass application. The blank was run in a triplicate, the other samples once.*

It can be designed for a simpler overview, which is appropriate for analyses by non-expert users, of only reporting the most abundant deconvolved mass per peak (**Figure 13**). Thus, giving only the major impurities based on relative UV purity of the area of the peaks. As well as the processing can be designed to include more detailed data with several deconvolved masses per peak (**Figure 14**). Then a mass spectral purity for each peak is calculated and combined with the relative UV purity to give an UVxMS relative purity of the sample. The limits of masses to consider for deconvolution based on percentages of the intensity of the most abundant mass can be adjusted in the settings, hence, there is a possibility for a high-quality estimation of relative purity. However, this method also includes isotopes of the most

abundant mass in the relative mass spectral purity, which probably can be avoided by further optimising the setting parameters for the method.

Component		Observed TIC RT (mins)	Observed UV RT (mins)	Observed RT delta (mins)	Response	%	Observed neutral mass (Da)
Felypressin	●	1.49	1.47	0.02	10,712,667	100.00	1,039.434
	■	1.52	1.50	0.02	2,018,431	100.00	1,039.434
	▲	1.56	1.54	0.02	3,452,938	100.00	2,078.878
	◆	1.62	1.60	0.02	143,688	100.00	2,078.778
	○	1.66	1.64	0.02	343,269	100.00	4,157.774
	□	1.71	1.69	0.02	110,704	100.00	4,158.798

Figure 13 *Processing of raw data with Intact Mass. This method shows only one deconvoluted mass per peak, the most abundant one, thus the % value of mass spectral purity is 100 for all peaks. Each symbol represents one peak.*

Component		Observed TIC RT (mins)	Observed UV RT (mins)	Observed RT delta (mins)	Response	%	Observed neutral mass (Da)
Felypressin	●	1.49	1.47	0.02	10,712,667	83.97	1,039.434
	●	1.49	1.47	0.02	1,229,119	9.63	1,040.449
	●	1.49	1.47	0.02	816,609	6.40	1,041.422
	■	1.52	1.50	0.02	2,018,431	52.34	1,039.434
	■	1.52	1.50	0.02	1,345,380	34.89	2,078.883
	■	1.52	1.50	0.02	268,487	6.96	2,080.014
	■	1.52	1.50	0.02	224,170	5.81	2,079.760

Figure 14 *Processing with Intact Mass. This second method gives several deconvolved neutral masses per peak, including the different masses of isotopes. Thus, the mass spectral purity (%) is calculated considering different isotopes.*

A simple impurity profiling can be performed in Intact Mass by defining impurities, such as a deamidation or oxidation [12, 15]. The software then suggests potential impurities by calculating differences in accurate mass.

Comparison with the two other available software MassLynx and Unifi, Intact Mass has the advantage of giving a combined UV and mass spectral purity. In addition, Intact Mass can give the deconvoluted mass for the different impurities, unlike the other software evaluated. Relative UV purity can be obtained in the other two software as well, but not yet the mass spectral purity. Both MassLynx and Unifi can present data for an analytical plate using a colour scale, similar to Intact Mass.

In Intact Mass, a report can be created for a single sample, which includes the chromatograms as well, or as a summary for several peptides. The summary report can be edited to some extent, but not as much compared to Unifi and MassLynx. For example, no chromatograms can be included.

In the future, acquisition will be possible in Intact Mass but for this created workflow the acquisition will be in Unifi since this is the software currently compatible with the RDa instrument. To set up the analysis in Unifi a lot of copy and paste is required. It would be beneficial to reduce this to make it more automated.

The Created Workflow for Peptide Characterization

The aim with this degree project was to create an analytical workflow that has the capability of analysing a large number of peptides in an efficient way. The product that will be handed over is therefore the workflow illustrated in **Figure 15**. The application could for example be for a chemist, coming with a plate of peptides desired to be characterized with accurate mass and relative purity. The workflow for the chemist will then be to follow the procedure described in general terms here.

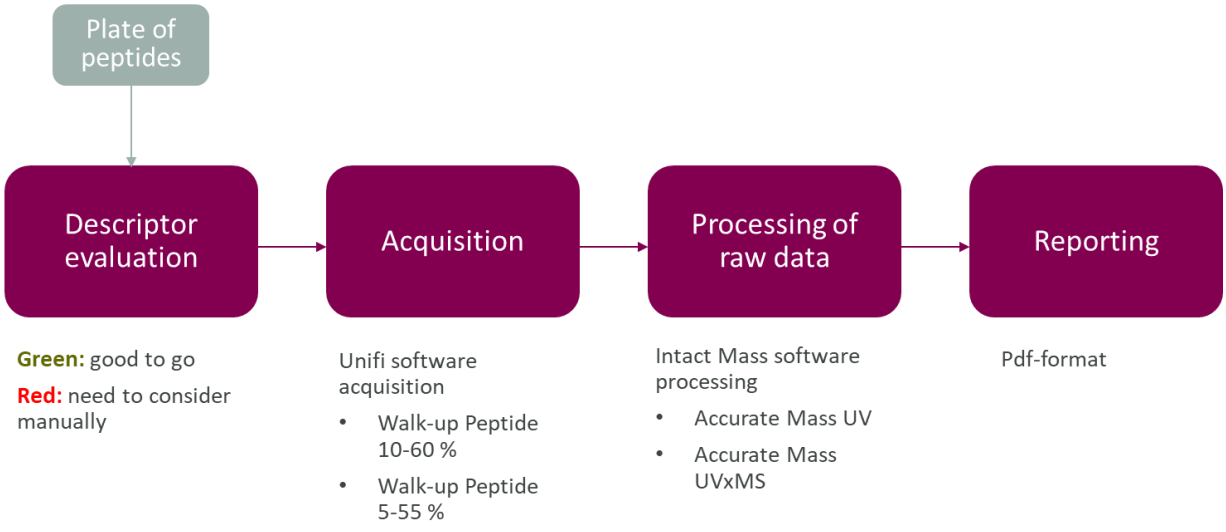


Figure 15 *An overview of the created analytical workflow for a high-throughput system for peptide characterization.*

Preparations of Peptides: The plate of peptides (or single sample of peptide) must meet the requirements to be run in this workflow. First, it needs to be dissolved appropriately. That means the solvent should be compatible with the LC-MS system and no residual solids can be present in the vial, the liquid needs to be clear. Second, the concentration of the peptide should be the specified value to obtain an appropriate absorbance in the TUV detector.

Descriptor evaluation: Before going to the lab, a search in the created tool for descriptor values needs to be conducted. The output is the different descriptors with a colour scale. If the descriptor value is shown as red, it means that the peptide could be either too hydrophilic or hydrophobic and manual

interpretation is needed to decide whether it should be analysed or not. If shown as green the descriptor values of that peptide are suitable for the workflow. If the ion class is a base or neutral, then the 10-60 %B method should be chosen. However, if the ion class is a zwitterion a second evaluation is needed to evaluate which of the two methods is most suitable.

Acquisition: Acquisition is performed on an RDa instrument. If all peptides are suitable for the generic method, then the method called *Walk-up Peptide 10-60 %* should be chosen. If the peptides are concluded to be hydrophilic from the search tool, then the method called *Walk-up Peptide 5-55 %* should be chosen for that peptide instead. The acquisition will be performed in Unifi since this is the software currently compatible with the system. Then the raw data needs to be exported from Unifi and imported to Intact Mass where it can be processed. In the near future, acquisition will be possible in Intact Mass as well.

Processing of Raw Data: The raw data is imported to Intact Mass where it will be processed. There are two methods that can be chosen for processing of raw data. The first, called *Accurate Mass UV*, estimates the relative purity based on UV area only, and gives back the deconvoluted molecular weight for each impurity. The second method estimates the relative UV purity based on area together with the mass spectral purity, as well as giving a combined result, UVxMS.

Reporting: In Intact Mass a summary report of all peptides is created. This can be designed to some extent after need and later converted to a pdf-document.

5 Conclusions

Two different LC methods were developed suitable for a high-throughput analysis that can be applied on a range of peptides. One LC-MS method that is more generic and one that is suitable for more hydrophilic compounds that are not retained with the generic method. The peptide descriptors ClogK_D and aqueous solubility can be used to predict whether this workflow is suitable or not for a peptide of interest. In the software Waters Connect with its application Intact Mass a processing method can be designed that provide a combined UV absorbance and mass spectral purity (UVxMS) and performs a simple impurity profiling. In overall, a workflow that is more applicable for a high-throughput analysis of peptides has been developed, however, work remains to make it more automated.

6 Future aspects

In this degree project a workflow for analytical characterization of peptides was created which was constructed based on 20 peptides. The next step would be to validate this method on a range of new peptides to see how well it performs outside the included data. There is also some remaining work needed for the processing methods in Intact Mass. It would also be desired to make the workflow more automated, especially to make it possible to also acquire data, not just evaluate, in Intact Mass, instead of Unifi. This would save a lot of time and reduce the number of copy pastes needed.

7 References

1. Craik, D.J., et al., *The future of peptide-based drugs.* Chem Biol Drug Des, 2013. **81**(1): p. 136-47.
2. Lau, J.L. and M.K. Dunn, *Therapeutic peptides: Historical perspectives, current development trends, and future directions.* Bioorg Med Chem, 2018. **26**(10): p. 2700-2707.
3. Muttenthaler, M., et al., *Trends in peptide drug discovery.* Nat Rev Drug Discov, 2021. **20**(4): p. 309-325.
4. D'Addio, S.M., et al., *New and Evolving Techniques for the Characterization of Peptide Therapeutics.* J Pharm Sci, 2016. **105**(10): p. 2989-3006.
5. Fellner, R.C., S.T. Terryah, and R. Tarran, *Inhaled protein/peptide-based therapies for respiratory disease.* Mol Cell Pediatr, 2016. **3**(1): p. 16.
6. Mannhold, R., et al., *Calculation of molecular lipophilicity: State-of-the-art and comparison of log P methods on more than 96,000 compounds.* J Pharm Sci, 2009. **98**(3): p. 861-93.
7. Visconti, A., et al., *Prediction and interpretation of the lipophilicity of small peptides.* J Comput Aided Mol Des, 2015. **29**(4): p. 361-70.
8. Fekete, S., J.L. Veuthey, and D. Guillarme, *New trends in reversed-phase liquid chromatographic separations of therapeutic peptides and proteins: theory and applications.* J Pharm Biomed Anal, 2012. **69**: p. 9-27.
9. Lam, A.K., et al., *Fast and high-throughput LC-MS characterization, and peptide mapping of engineered AAV capsids using LC-MS/MS.* Mol Ther Methods Clin Dev, 2022. **27**: p. 185-194.
10. Åsberg, D., et al., *The importance of ion-pairing in peptide purification by reversed-phase liquid chromatography.* J Chromatogr A, 2017. **1496**: p. 80-91.
11. Gilar, M., T.S. McDonald, and F. Gritti, *Impact of instrument and column parameters on high-throughput liquid chromatography performance.* J Chromatogr A, 2017. **1523**: p. 215-223.
12. Zeng, K., et al., *Liquid Chromatography-High Resolution Mass Spectrometry for Peptide Drug Quality Control.* AAPS J, 2015. **17**(3): p. 643-51.
13. Frolov, A.I., et al., *pIChemiSt horizontal line Free Tool for the Calculation of Isoelectric Points of Modified Peptides.* J Chem Inf Model, 2023. **63**(1): p. 187-196.
14. Neue, U.D., *Theory of peak capacity in gradient elution.* J Chromatogr A, 2005. **1079**(1-2): p. 153-61.
15. Qian Cutrone, J.J., et al., *Tiered analytics for purity assessment of macrocyclic peptides in drug discovery: Analytical consideration and method development.* J Pharm Biomed Anal, 2017. **138**: p. 166-174.

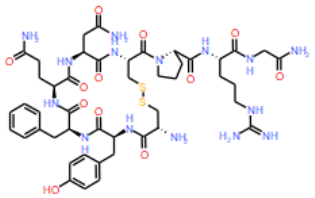
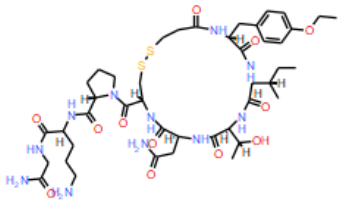
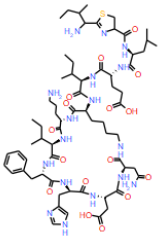
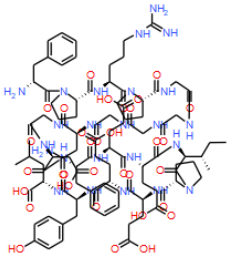
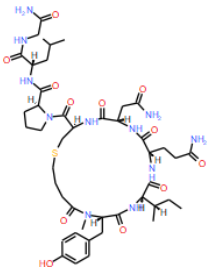
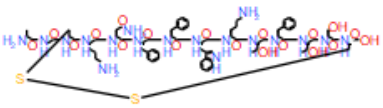
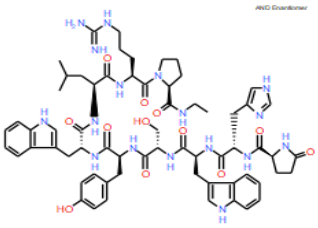
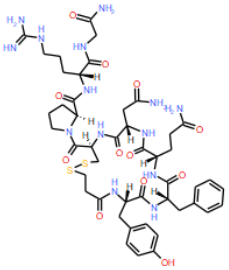
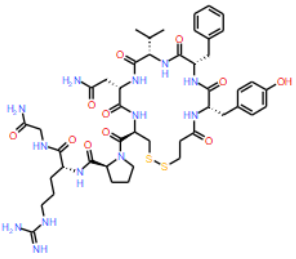
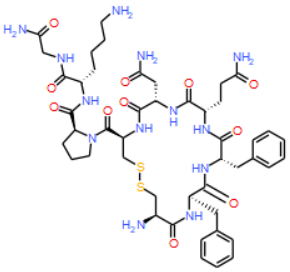
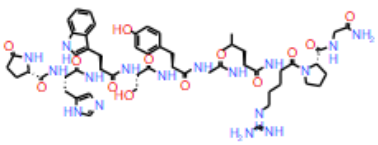
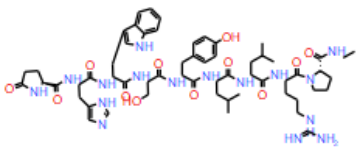
8 Appendix

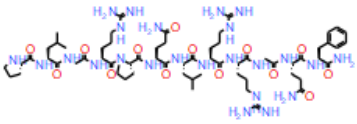
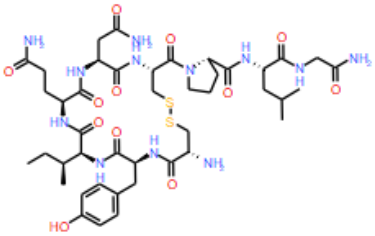
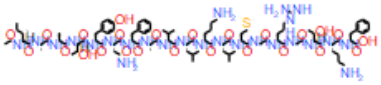
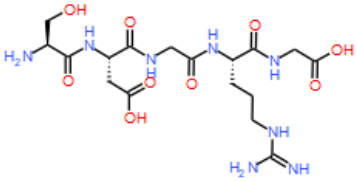
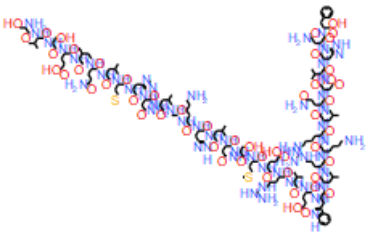
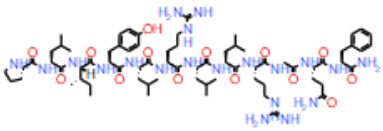
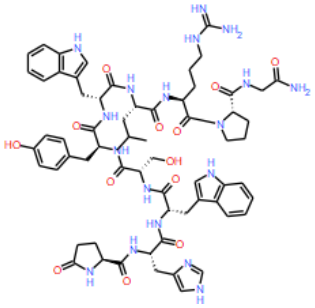
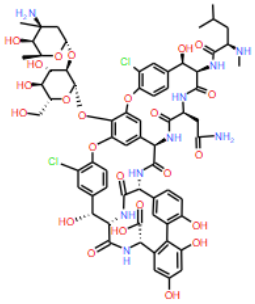
Appendix A

aa = amino acid residues

Compound Name	Formula	Monoisotopic Molecular Weight	Ion Class	Amino acid sequence	Nr of aa
Argipressin	C46 H65 N15 O12 S2	1083.4378	Base	CYFQNCPRG	9
Atosiban	C43 H67 N11 O12 S2	993.4412	Base	XITNCPXG	8
Bacitracin	C66 H103 N17 O16 S	1421.7489	Zwitterion	XXXKXXFXDX	10
Bivalirudin	C98 H138 N24 O33	2178.9858	Zwitterion	XPRPGGGGNGDFEEIPEEYL	20
Carbetocin	C45 H69 N11 O12 S	987.4848	Neutral	IQNCPLG	7
Cyclic somatostatin	C76 H104 N18 O19 S2	1636.7166	Zwitterion	AGCKNFFWKTFTSC	14
Deslorelin	C64 H83 N17 O12	1281.6407	Base	XHWSYWLRX	9
Desmopressin	C46 H64 N14 O12 S2	1068.4269	Base	YFQNCPXG	8
d-Vasopressin	C46 H65 N13 O11 S2	1039.4368	Base	YFVNCPXG	8
Felypressin	C46 H65 N13 O11 S2	1039.4368	Base	CXFQNCPKG	9
GNRH	C55 H75 N17 O13	1181.5730	Base	XHWSYGWLPG	10
Leuprolide acetate	C59 H84 N16 O12	1208.6454	Base	XHWSYXLRX	9
ONEG	C63 H106 N24 O14	1422.8320	Base	PLGRPQLRRGQF	12
Oxytocin	C43 H66 N12 O12 S2	1006.4364	Base	CYIQNCPLG	9
PLP	C100 H159 N25 O25 S	2142.1659	Zwitterion	XIAATYNFAVLKLMGRGTKF	20
Ser-Asp-Gly-Arg-Gly	C17 H30 N8 O9	490.2136	Zwitterion	SDGRG	5
Teriparatide	C181 H291 N55 O51 S2	4115.1308	Zwitterion	SVSEIQLMHNLGKHLNSMERVEWLRKKLQDVHNF	34
TP2	C72 H118 N20 O14	1486.9136	Base	PLIYLRLLRGQF	12
Triptorelin	C64 H82 N18 O13	1310.6308	Base	XHWSYXLRPG	10
Vancomycin	C66 H75 Cl2 N9 O24	1447.4302	Zwitterion	XGNGGGGX	8

Structural formula of the peptides.

Argipressin 	Atosiban 	Bacitracin 
Bivalirudin 	Carbetocin 	Cyclic somatostatin 
Deslorelin 	Desmopressin 	d-vasopressin 
Felypressin 	GNRH 	Leuprolide acetate 

ONEG 	Oxytocin 	PLP 
Ser-Asp-Gly-Arg-Gly 	Teriparatide 	TP2 
Triptorelin 	Vancomycin 	

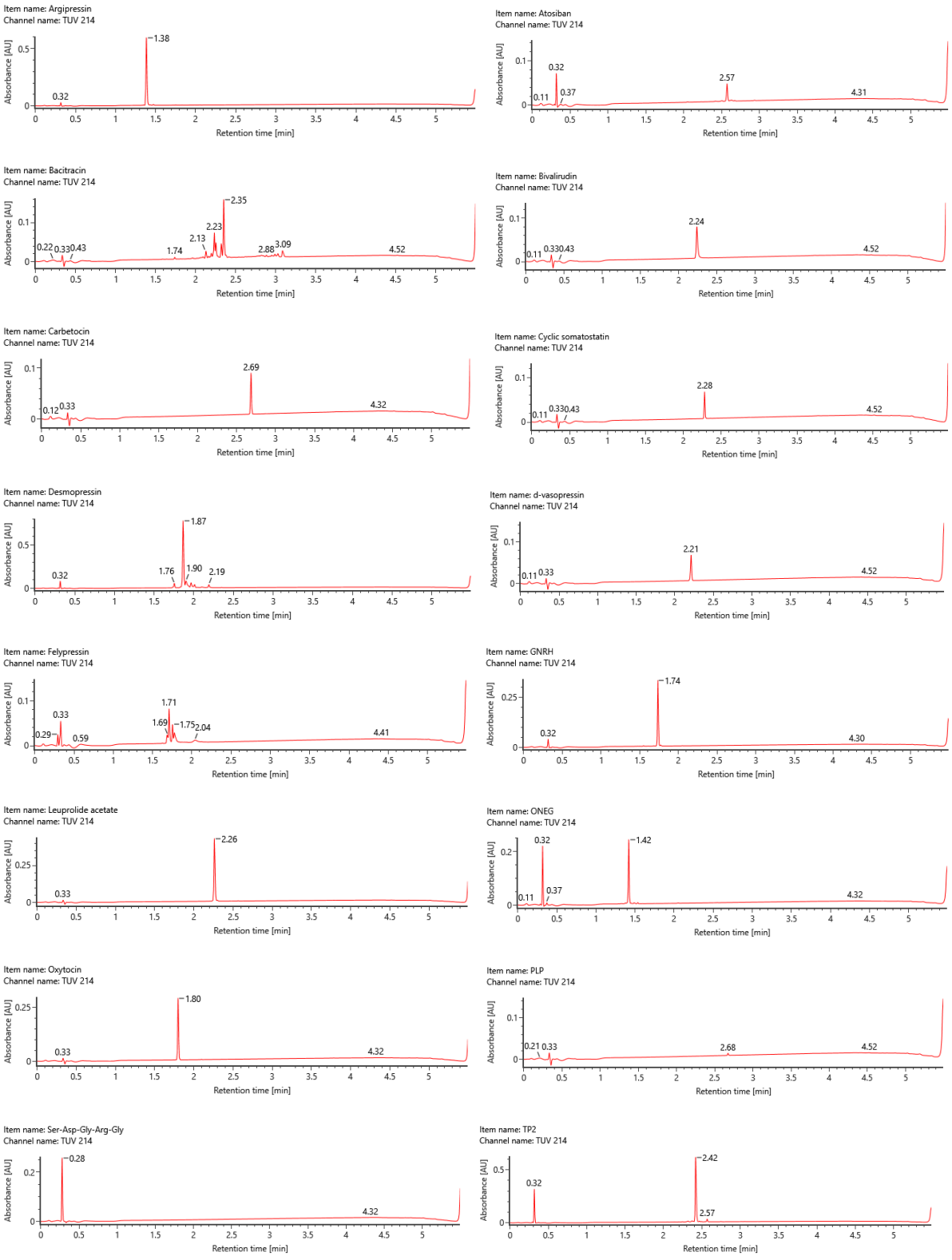
Appendix B

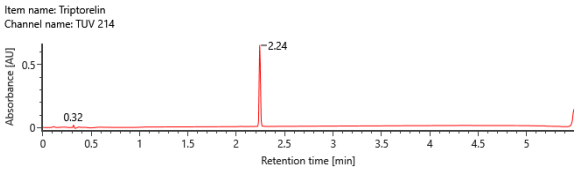
Descriptor values obtained from the search tool: aqueous solubility, ClogK_D, AlogK_D and AZ logD. The charge at pH 2 is obtained from the open-source tool pIChemiSt.

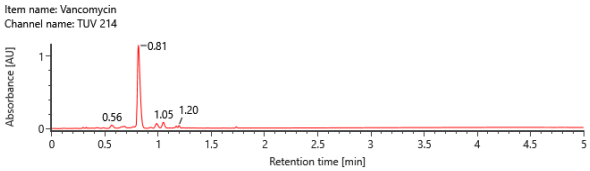
Compound Name	Aqueous Solubility pH 7.4 prediction (µM)	ClogK$_D$	AlogK$_D$	pI	AZ logD (pH 7.4)	Charge at pH 2
Argipressin	1094	-3.8	-5.5	11.1	-2.9	2
Atosiban	5459	-0.1	-4.3	11.5	-1.9	1
Bacitracin	52	-2.1	-2.6	8.1	-1.20	3
Bivalirudin	2	0.0	-9.3	3.7	-8.5	2.9
Carbetocin	1078	-0.0	-3.7	8.5	-0.2	0
Cyclic somatostatin	8	-1.5	-4.8	10.5	-2.9	2.8
Deslorelin	215	-1.0	-1.5	11.1	1.3	2
Desmopressin	1004	-3.1	-4.7	11.1	-2.2	1
d-Vasopressin	1785	-0.4	-2.8	11.1	-2.0	1
Felypressin	1254	-1.8	-5.1	11.4	-2.1	2
GNRH	1166	-4.1	-5.1	11.1	-1.0	1
Leuprolide acetate	425	-1.0	-1.9	11.1	0.7	2
ONEG	20	-7.9	-6.8	13.3	-1.9	4
Oxytocin	1201	-0.7	-4.3	8.4	-1.3	1
PLP	2	0.0	-4.9	10.6	-1.5	3
Ser-Asp-Gly-Arg-Gly	2932475	-9.9	-5.6	6.1	-6.4	2.8
Teriparatide	1	0.0	-19.2	7.5	-9.9	8.5
TP2	14	0.3	-0.0	12.2	2.7	3
Triptorelin	231	-2.4	-3.1	11.1	0.1	2
Vancomycin	250	-1.1	-0.5	7.6	-2.8	1.2

Appendix C

UV chromatograms of the peptides with the BEH column (50mm), mobile phase composition A: 100% MQ, 0.03% TFA and B: 95% ACN, 0.03% TFA, 14 %B/min, and flow rate 0.4 mL/min.

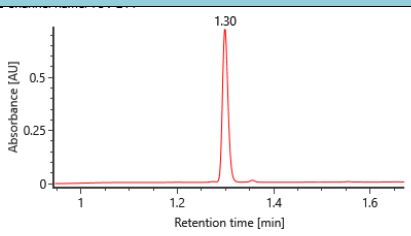
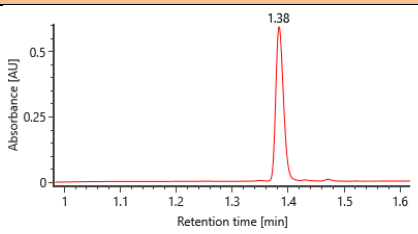
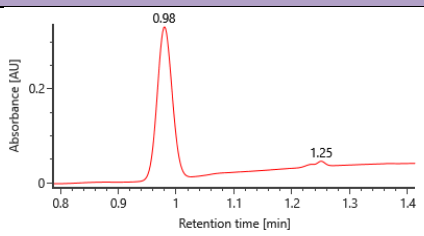
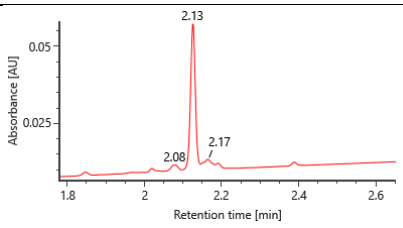
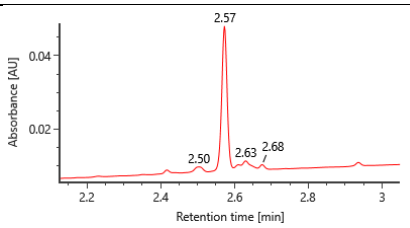
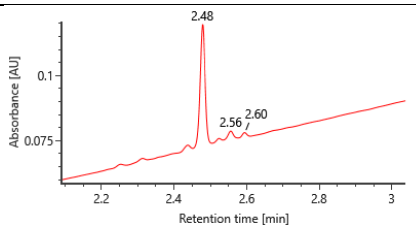
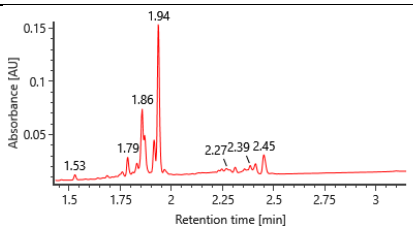
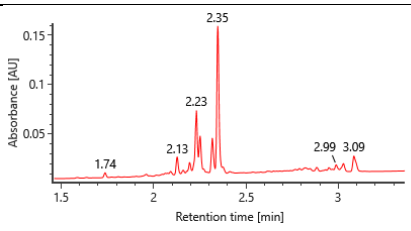
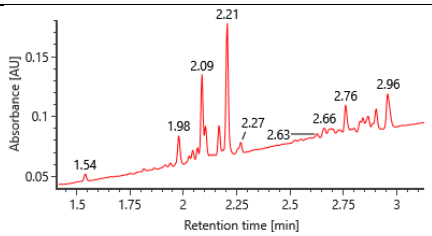
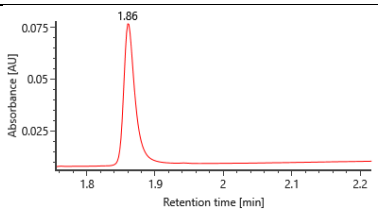
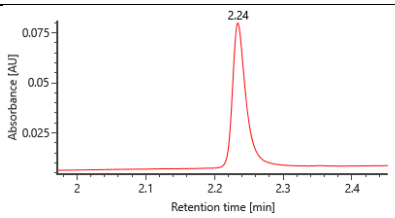
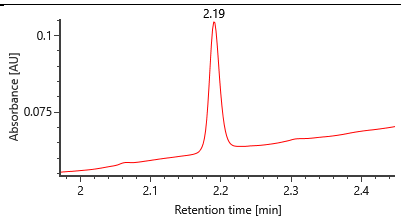
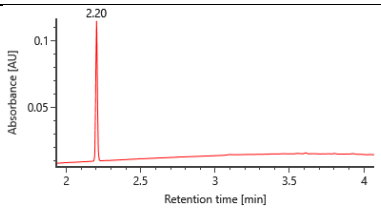
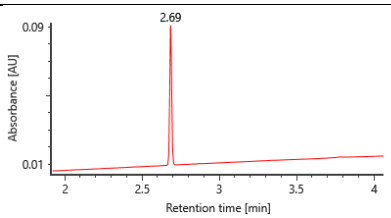
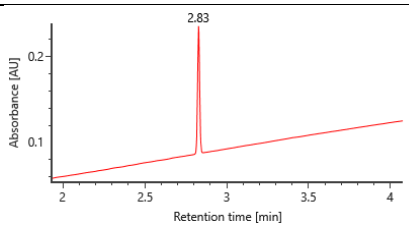


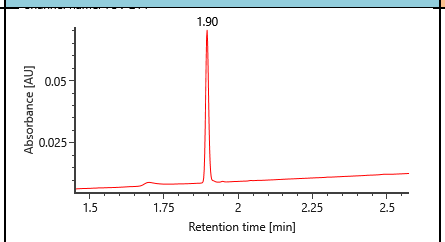
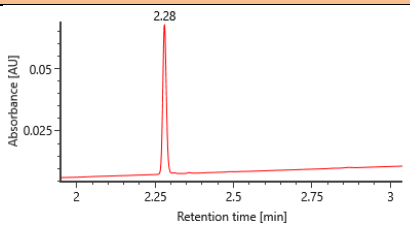
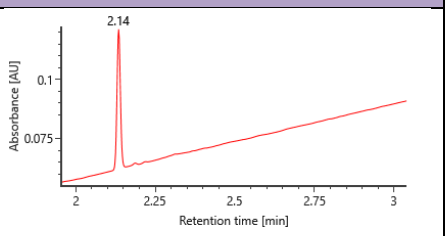
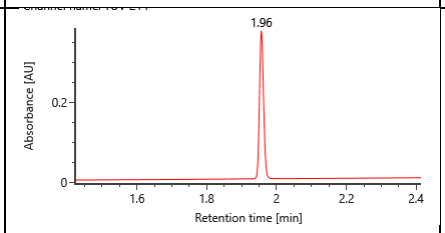
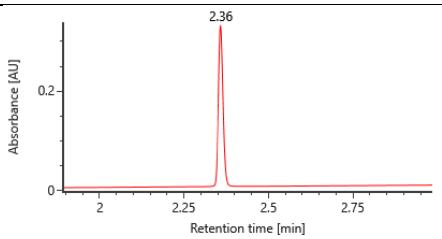
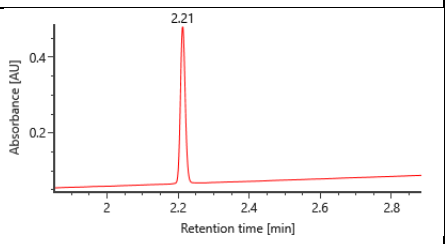
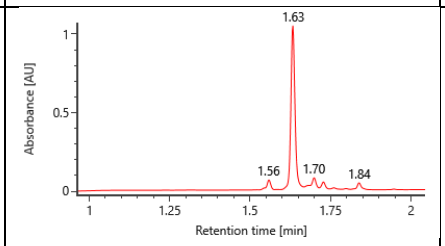
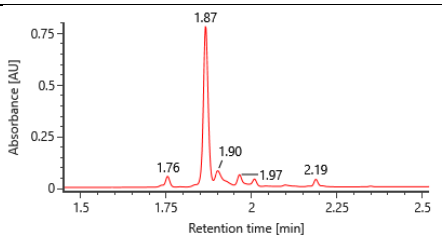
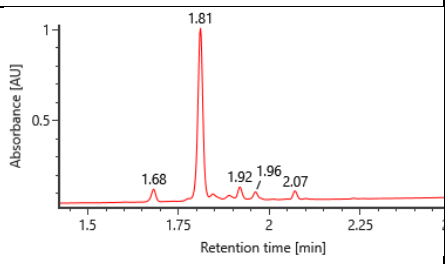
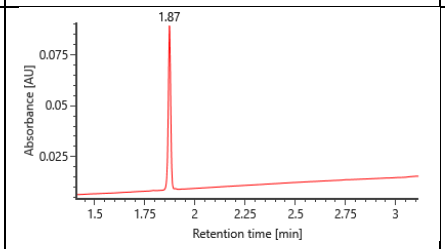
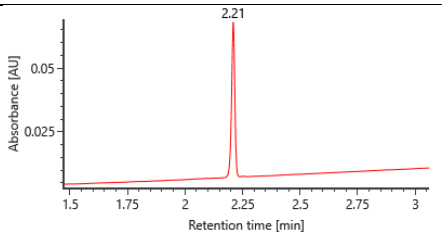
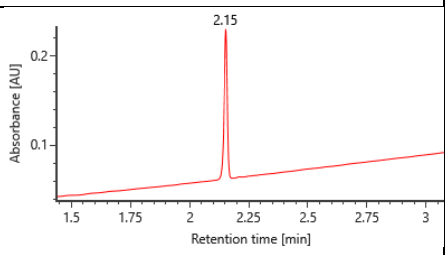
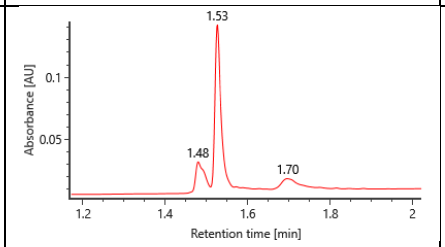
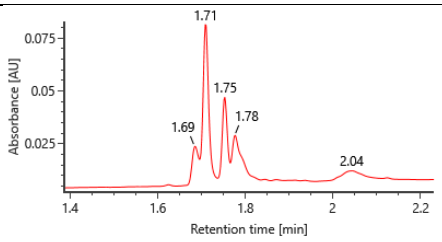
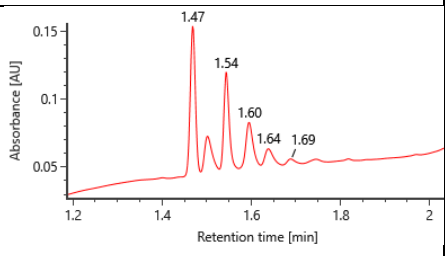
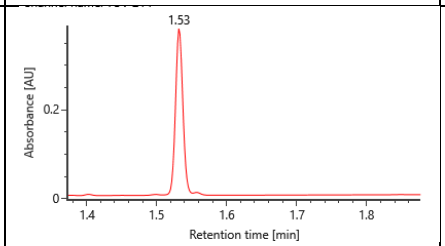
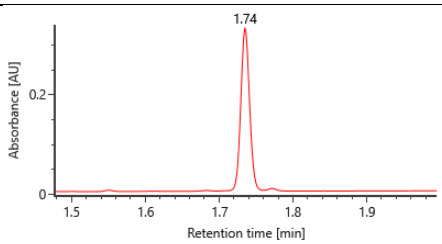
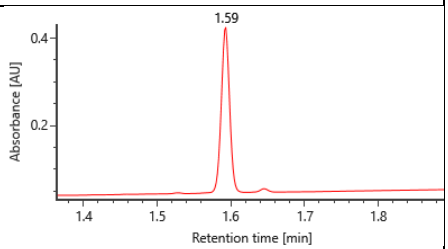

Item name: Triptorelin
Channel name: TUV 214
Absorbance [AU]
0.5
0
2.24
0.32
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
Retention time [min]

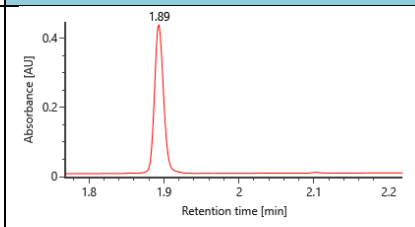
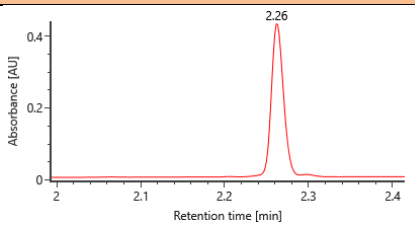
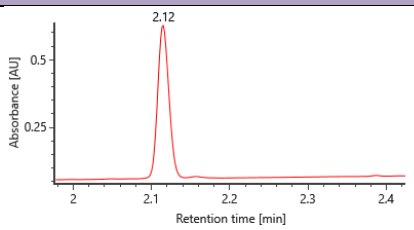
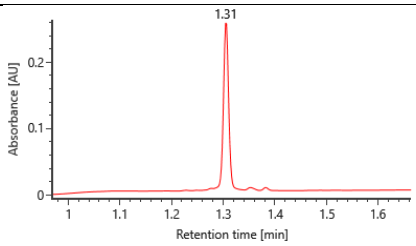
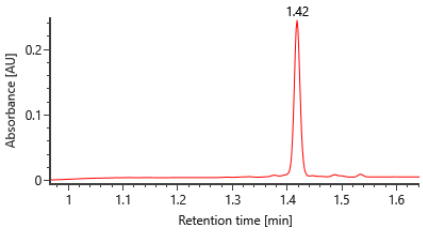
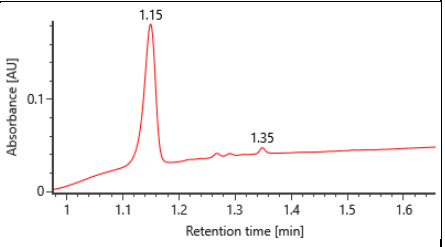
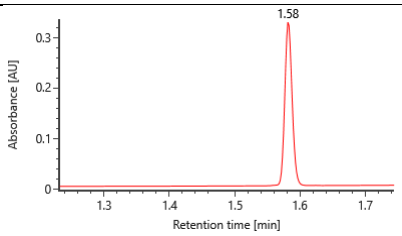
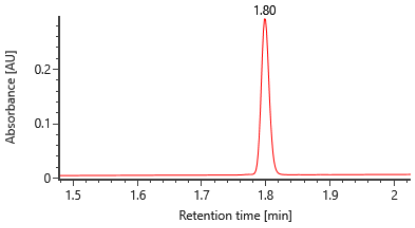
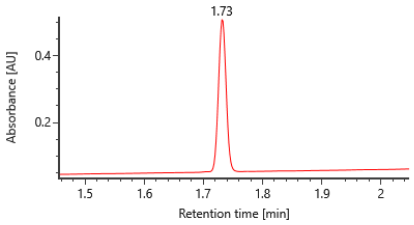
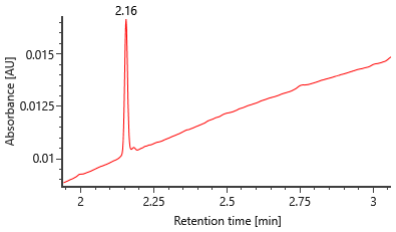
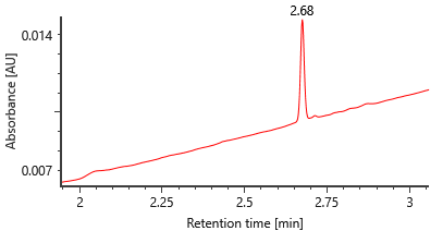
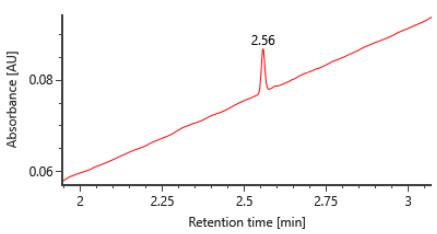
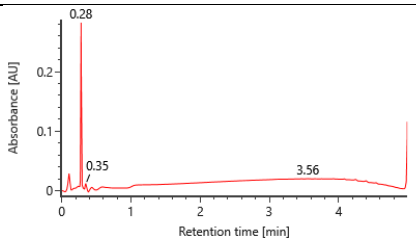
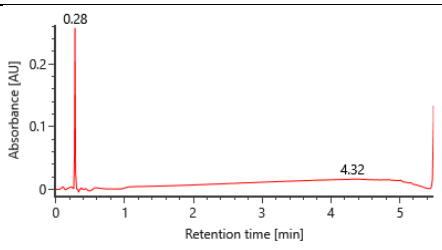
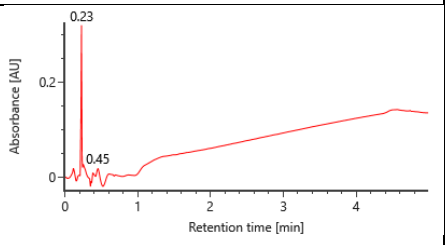
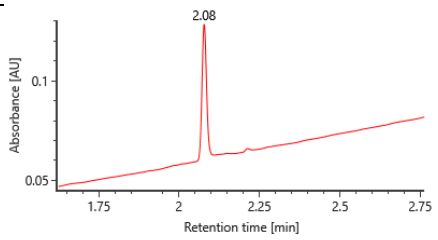

Item name: Vancomycin
Channel name: TUV 214
Absorbance [AU]
1
0
0.81
0.56
1.05
1.20
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
Retention time [min]

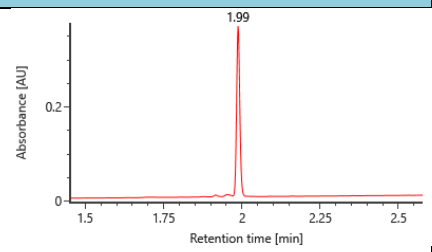
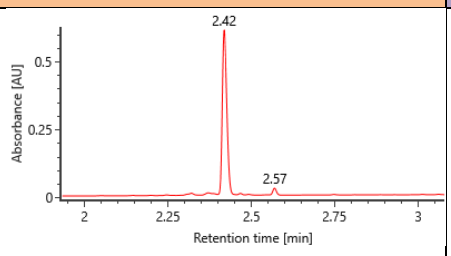
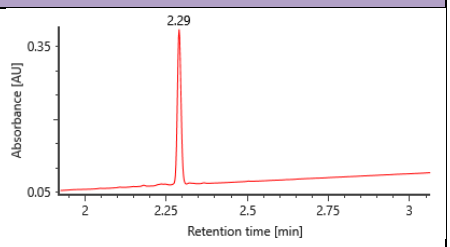
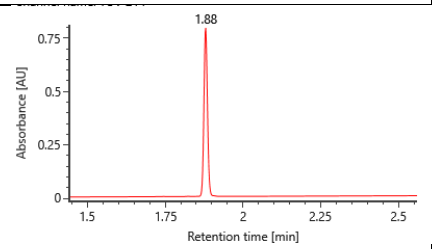
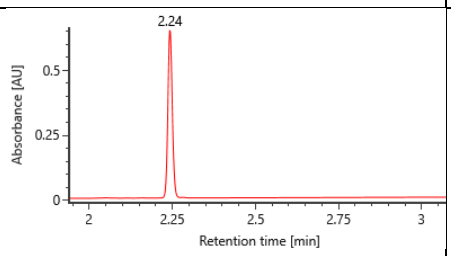
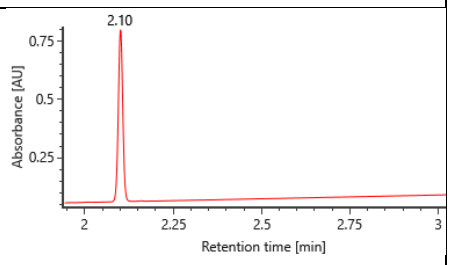
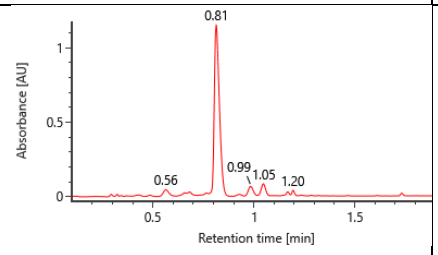
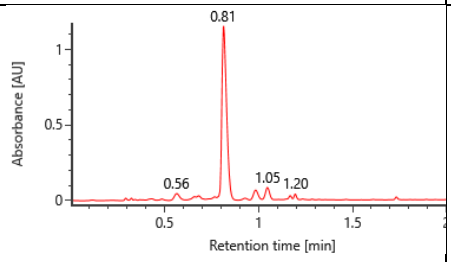
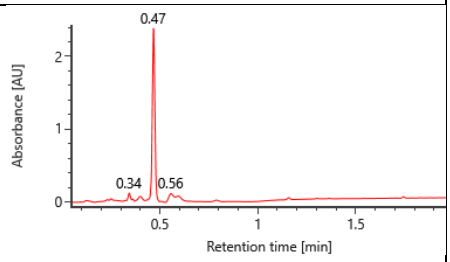
Appendix D

UV chromatograms for peptides comparing two gradients 14 %B/min and 21 %B/min with a BEH column, flow rate 0.4 mL/min (left and middle column) and comparing the two different stationary phases BEH and CSH, with gradient of 14 %B/min, flow rate 0.4 mL/min (middle and right column) on the RDa instrument. For the BEH column the mobile phase composition was A: 100% MQ + 0.1% FA and B: ACN + 0.1% FA and for the CSH column A: 100% MQ, 0.03% TFA and B: 95% ACN, 0.03% TFA.

	BEH column, 21 %B/min	BEH column, 14 %B/min	CSH column, 14 %B/min
Argipressin			
Atosiban			
Bacitracin			
Bivalirudin			
Carbetocin			

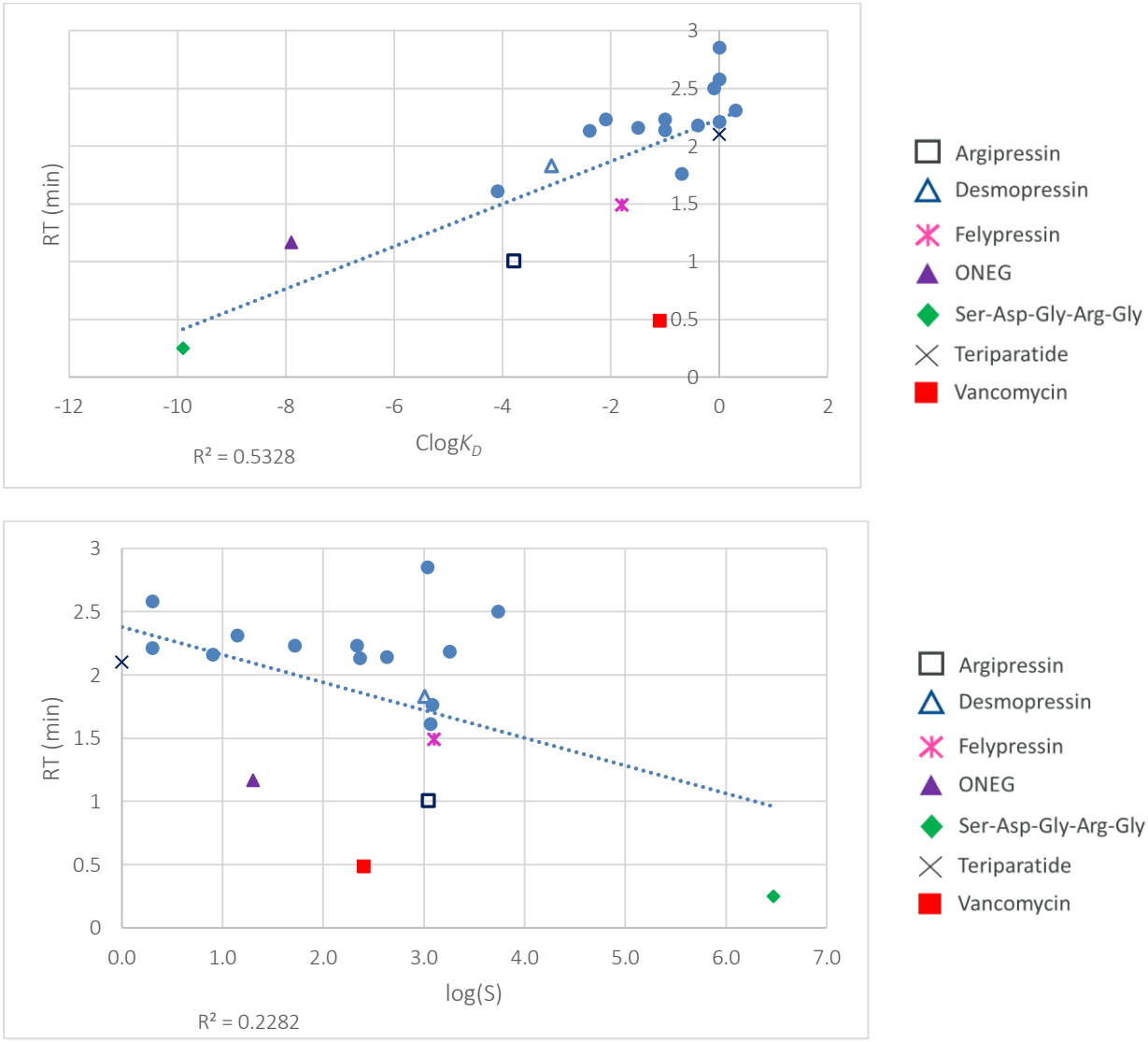
	BEH column, 21 %B/min	BEH column, 14 %B/min	CSH column, 14 %B/min
Cyclic somato-statin			
Deslorelin			
Desmo-pressin			
d-vasopressin			
Felypressin			
GNRH			

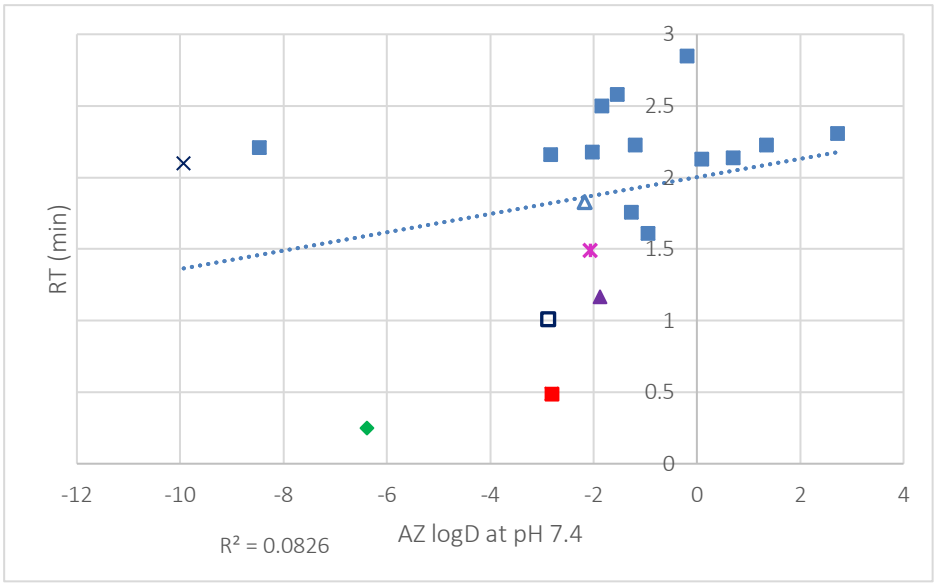
	BEH column, 21 %B/min	BEH column, 14 %B/min	CSH column, 14 %B/min
Leuprolide acetate			
ONEG			
Oxytocin			
PLP			
Ser-Asp-Gly-Arg-Gly			
Teripara-tide	No data	No data	

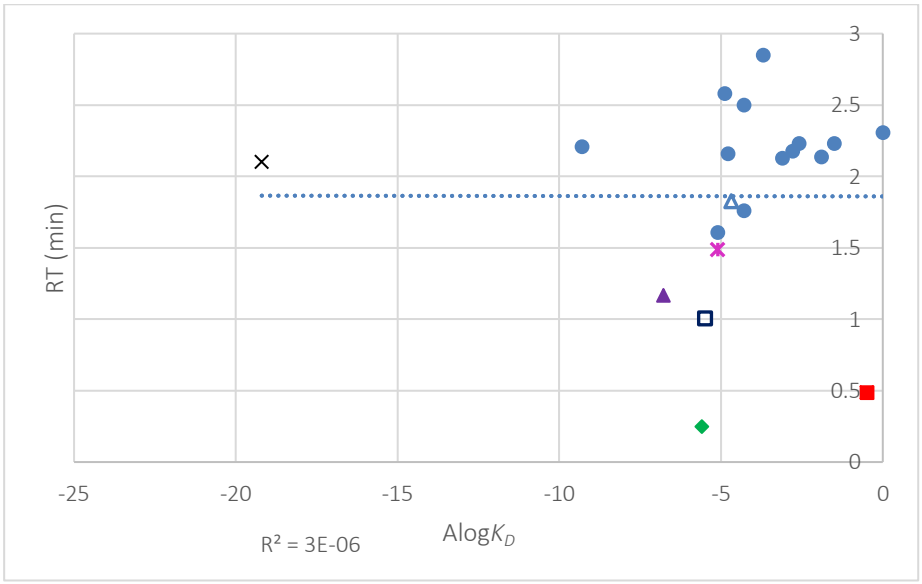
	BEH column, 21 %B/min	BEH column, 14 %B/min	CSH column, 14 %B/min
TP2			
Triptorelin			
Vanco-mycin			

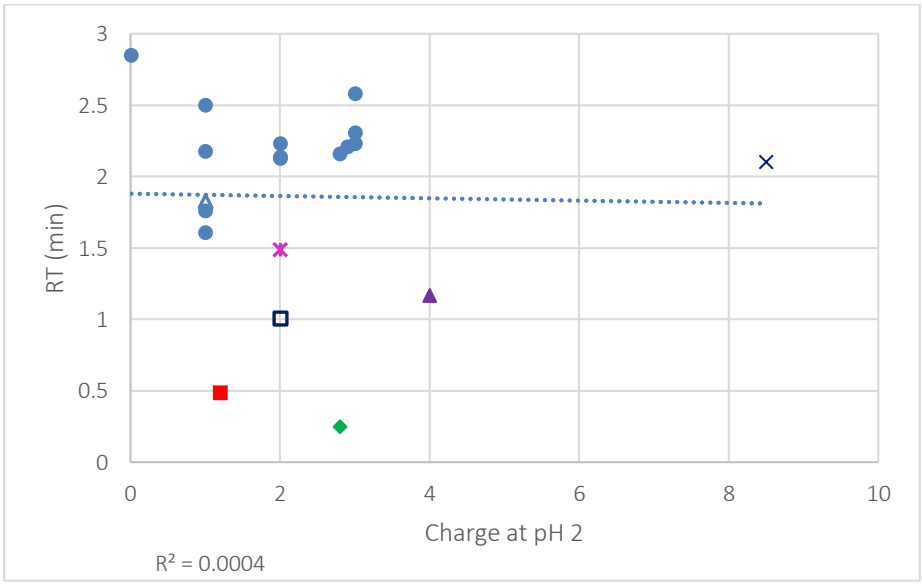
Appendix E

Descriptor evaluation. ClogK_D and aqueous solubility (S) were performing the best. A few peptides below are highlighted due to having interesting properties. The rest are visualised by a blue circle.




RT (min)
AZ logD at pH 7.4
R² = 0.0826
Argipressin
Desmopressin
Felypressin
ONEG
Ser-Asp-Gly-Arg-Gly
Teriparatide
Vancomycin


RT (min)
AlogK_D
R² = 3E-06
Argipressin
Desmopressin
Felypressin
ONEG
Ser-Asp-Gly-Arg-Gly
Teriparatide
Vancomycin


RT (min)
Charge at pH 2
R² = 0.0004
Argipressin
Desmopressin
Felypressin
ONEG
Ser-Asp-Gly-Arg-Gly
Teriparatide
Vancomycin

Appendix F

Limits for suitability of methods, predicted by peptide descriptors.

Green = 10-60 %B method

Orange = 5-55 %B method