UNIVERSITY OF
KWAZULU-NATAL

INYUVESI
YAKWAZULU-NATALI

# Sub/supercritical fluid chromatography purification of biologics


## Kamini Govender


## 2020

**A thesis presented for the degree of**
**Doctor of Philosophy (Pharmaceutical Chemistry)**


University of KwaZulu-Natal (Westville Campus)
College of Health Sciences
The discipline of Pharmaceutical Sciences
Catalysis and Peptide Research Unit

# Sub/supercritical fluid chromatography purification of biologics


## Kamini Govender


## 2020

A thesis submitted to the College of Health Science, University of KwaZulu-Natal, Westville, for the degree of Doctor of Philosophy in Pharmaceutical Chemistry.

This is to certify that the contents of the thesis are the original work of Miss Kamini Govender, carried out under our supervision at the Catalysis and Peptide Research Unit (CPRU), University of KwaZulu-Natal, Westville Campus, Durban, South Africa.

As the candidate's supervisors, we have approved this thesis for submission.

Supervisor:

Signed: ---------- Name: **Prof. T. Naicker** Date:----------

Supervisor:

Signed: ---------- Name: **Prof. H.G. Kruger** Date:----------

# Abstract

Peptide and protein drugs are highly versatile with numerous therapeutic properties such as anti-cancer, anti-diabetic, anti-hypertensive, and anti-microbial; which are therefore ideal candidates for the development of next-generation drugs. This is exemplified using these drugs for the treatment of diseases such as diabetes. Diabetes is one of the most prevalent non-communicable diseases worldwide. The rapid increase in the number of diabetic patients globally places a burden on current insulin manufacturers. The traditional reversed-phase high-performance liquid chromatography (RP-HPLC) purification methods of insulin and peptides are problematic, tedious with long run times of approximately 50 minutes, low yields employ harsh solvents such as acetonitrile, which has a negative impact on the environment. There is a need for a greener process for the purification of insulin and peptides. Sub/supercritical fluid chromatography (SFC) can provide the solution since it utilises greener mobile phases such as carbon dioxide ($CO_2$) and methanol, which can be recycled. However, there is a paucity of knowledge regarding the SFC purification of human insulin and peptides. Therefore, this research study aimed to provide an efficient, innovative approach for the biosynthesis of human insulin and the SFC purification of biosynthesised human insulin, as well as the extension into the SFC purification of peptides. The background of these topics is presented in Chapter One.

Chapter two (manuscript one) presents the development of a novel and more efficient method of human insulin biosynthesis in *Escherichia coli (E. coli)*. Several of the conventional steps were eliminated. The crude biosynthesised protein sequence was verified using protein sequencing, which had a 100% similarity to the human insulin sequence. The biological activity of the biosynthesised human insulin was tested *in vitro* using a 3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyl tetrazolium bromide (MTT) assay. The biosynthesis of human insulin was conducted on a laboratory-scale basis; future studies should investigate scaling up of this method.

Chapter three (manuscript two) was based on the SFC purification of the commercially available standard sample of human insulin and a crude biosynthesised sample of human insulin. The SFC purified standard the biosynthesised human insulin samples were detected and quantified using liquid chromatography-mass spectrometry (LC-MS) and protein sequencing techniques. SFC columns, i.e., silica, 2' ethyl pyridine, diol-HILIC, and the pentafluoro phenyl (PFP), were evaluated to determine the ideal column. The PFP column gave the best results since it displayed good peak shapes, resolution, retention factors, retention times, and the least relative standard deviation in comparison to the other columns. Therefore, the aforementioned column was selected for further analysis using the biosynthesised human insulin, whereby a column efficiency test was conducted on a semi-preparative scale, yielding 84% recovery. Subsequently, the biological activities of the SFC purified standard sample of human insulin and biosynthesised version were tested *in vitro* using a MTT assay. The results indicated that the biological activities of the standard and biosynthesised human insulin derivatives were retained subsequent to SFC purification. The biological activities were highly significant, with a p-value of $< 0.0001$.

From chapter three, band broadening and phase separation peaks were experienced during SFC purification of the commercially available standard sample of human insulin and biosynthesised human insulin. Therefore, in Chapter four (manuscript three), a SFC purification method was developed to purify peptides at an analytical scale. A tetrapeptide [insulin β chain peptide (15-18)], octapeptide [angiotensin II], nonapeptide [insulin β chain peptide (15-23)] were purified using four SFC columns, i.e., PFP, diol-HILIC, silica, and 2' ethyl pyridine. Subsequently, the 2' ethyl pyridine column was selected for further analysis based on the reproducibility, peak shapes, efficient separations and retention factors. The three peptides were monitored using LC-MS analysis. The successful peptide recoveries ranged from 80-102%.

Chapter five pertains to the summary and conclusion drawn from the study and reflects on possible future endeavours. The present study was successful in providing a more affordable and innovative approach for the biosynthesis of human insulin. The work also successfully developed a rapid, greener, and more efficient method of SFC purifying biosynthesised human insulin and peptides as opposed to conventional HPLC purification methods. As far as we are aware, this study is the first of its kind to purify biosynthesised human insulin and this combination of peptides using SFC purification techniques. Future research studies can focus on the SFC purification of larger protein molecules and consider the use of custom columns and or other modifiers for the improvement of the isolation of other highly sought after biologics within the pharmaceutical industry.

# Declaration 1-Plagiarism

I, Miss **Kamini Govender** declare that

1. The research reported in this thesis, except where otherwise indicated, and is my original research.
2. This thesis has not been submitted for any degree or examination at any other university.
3. This thesis does not contain other persons' data, pictures, graphs or other information unless specifically acknowledged as being sourced from other persons.
4. This thesis does not contain other persons' writing unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
    a. Their words have been re-written, but the general information attributed to them has been referenced
    b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks and referenced.
5. This thesis does not contain text, graphics or tables copied and pasted from the internet unless specifically acknowledged and the source being detailed in the thesis and in the reference sections.

Signature ……

Miss Kamini Govender

# Declaration 2-Publications

Detail of contribution to the publications that form part and/or include research presented in this dissertation (include publications in preparation, submitted, *in press* and published and give details of the contributions of each author to the experimental work and writing of each publication)

**List of publications:**

1) Kamini Govender[1], Tricia Naicker[1], Johnson Lin[2], Sooraj Baijnath[1], Anil Amichund Chuturgoon[3], Naeem Sheik Abdul[3], Taskeen Docrat[3], Hendrik Gerhardus Kruger[1*] and Thavendran Govender[4*]

**A novel and more efficient biosynthesis approach for human insulin production in *Escherichia coli (E. coli)*. *AMB Express*. Published, 2020. DOI: 10.1186/s13568-020-00969-w**

Kamini Govender biosynthesised human insulin using recombinant DNA technology, conducted, analysed all the experimental data and composed the entire manuscript including the supporting information. Taskeen Docrat, Dr Naeem Sheik Abdul and Prof. Anil Amichund Chuturgoon conducted the MTT assay. The rest of the authors are supervisors on the project and contributed to the conceptualisation of the idea and funding of the project.

2) Kamini Govender[1], Tricia Naicker[1], Sooraj Baijnath[1], Anil Amichund Chuturgoon[2], Naeem Sheik Abdul[2], Taskeen Docrat[2], Hendrik Gerhardus Kruger[1*] and Thavendran Govender[3*]

**Sub/supercritical fluid chromatography employing a water-rich modifier enables the purification of biosynthesized human insulin. Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences, Accepted, 2020. DOI: 10.1016/j.jchromb.2020.122126.**

Kamini Govender biosynthesised human insulin using recombinant DNA technology and developed a cost-effective method for the SFC purification of human insulin, analysed all the experimental data and composed the entire manuscript, including the supporting information. Taskeen Docrat, Dr Naeem Sheik Abdul and Prof. Anil Amichund Chuturgoon conducted the MTT assay. The rest of the authors are supervisors on the project and contributed to the conceptualisation of the idea, funding of the project, and provided scientific guidance.

3) Kamini Govender [1], Tricia Naicker[1], Sooraj Baijnath [1], Hendrik Gerhardus Kruger[1*] and Thavendran Govender[2*]

**The development of a sub/supercritical fluid chromatography-based purification method for peptides. Journal of Pharmaceutical and Biomedical Analysis, Submitted, 2020.**

Kamini Govender developed an efficient and cost-effective method for the SFC purification of peptides, analysed all the experimental data and composed the entire manuscript including the supporting information. The rest of the authors are supervisors on the project and contributed to the conceptualisation of the idea, funding of the project, and provided scientific guidance.

**Other publications which I contributed to:**

Satheesh Natarajan[1], Kamini Govender[1], Mokgokong Makwena[1], Sooraj Baijnath[1], Adeola Shobo[1], Per I Arvidsson[1,2], Thavendran Govender[1], Johnson Lin[3], Glenn Maguire[1,4], Tricia Naicker[1], Gert Kruger[1*] and Rae Silver[5*]

**Potential of brain mast cells for therapeutic application in the immune response to bacterial and viral infections. Submission in progress.**

Sateesh Natarajan and Kamini Govender, contributed to manuscript preparation, literature search, editing and writing of the manuscript. Mokgokong Makwena contributed to a literature search of the manuscript. Adeola Shobo contributed to a literature search and writing of the manuscript. The rest of the authors are supervisors on the project and contributed to the conceptualisation of the idea, funding of the project, and provided scientific guidance.

Signature ……

Miss Kamini Govender

# Dedication

This thesis is dedicated to my wonderful parents, Parma Varadarajalu Govender and Maeshwari Govender as well as my sister, Sovishnee Govender.

# Acknowledgements

I wish to express my sincere gratitude and appreciation to the following persons and institutions:

- I would like to thank God for giving me the strength, determination and perseverance throughout this study.

- To my parents, Parma and Maeshwari, thank you for your support and always inspiring me to be the best that I can be; I appreciate every sacrifice and endeavour that you have made throughout my studies. I could not have asked for better parents. Thank you for your love and support, which has allowed me to achieve my goals.

- I would like to thank my awesome sister, Sovishnee, for being a wonderful sister and for her encouragement throughout my studies. You are a true inspiration and my superhero.

- To my supervisors, Prof. T. Govender, Prof. H. G. Kruger, and Prof. T. Naicker, thank you for your guidance, support and excellent supervision throughout this study. Thank you for leading and inspiring the fourth revolution of industrial scientists in drug discovery.

- I would like to thank Prof. Tricia Naicker for her patience, assistance, and guidance. Thank you for your expert advice in the lab that enabled me to become a better Ph.D. student. I appreciate all the help you have given me throughout this study.

- I would also like to thank Prof. Gert Kruger for all of his time and expert opinion. He is dedicated and I appreciate all your help throughout this study. I truly respect and value your advice; I hope that one day I can have as many publications as you.

- I would like to thank Dr Sooraj Baijnath and Prof. Lin for their time, assistance, and guidance. I appreciate all the help given to me throughout this study.

- To Prof. Thavi Govender, I humbly thank you for conceptualizing this project. I sincerely thank you for allowing me to conduct this research for my Ph.D. project. Thank you for introducing me to the field of SFC purification. I appreciate all your help, guidance and assistance throughout my project.

- To all my research colleagues at the Catalysis and Peptide Research Unit (CPRU): Nakita, Arno, Lloyd, Kim, Shimaa, Sibusiso, Kamal, Spha, Neliswa, Mbo, Leon, Advaitaa, Sane and Yahya thank you for your support and friendship throughout this study. Especially for creating a wonderful work environment conducive to research.

- I would like to acknowledge the Central Analytical Facility (Stellenbosch University, South Africa) for conducting the protein sequencing.

- This work was made possible through financial support from the National Research Foundation (NRF) and the Technology Innovations Agency (TIA) of South Africa.

- This work was made possible through the University of KwaZulu-Natal (Westville) as I used the University's equipment and premises throughout this project.

# Abbreviations

| | |
|---|---|
| Å | Angstrom |
| aa | Amino acid |
| Anova | Analysis of variance |
| BPR | Back pressure regulator |
| C terminal | Carboxyl terminal |
| CHCA | Alpha-cyano-4-hydroxycinnamic acid |
| $Co^{2+}$ | Cobalt ion |
| $CO_2$ | Carbon dioxide |
| cRAP | Contaminant protein database |
| $Cu^{2+}$ | Copper ion |
| Da | Dalton |
| °C | Degrees Celsius |
| DCCA | Diabetes control and complications trial |
| DMSO | Dimethyl sulfoxide |
| DNA | Deoxyribonucleic acid |
| *E. coli* | *Escherichia coli* |
| FA | Formic acid |
| g | Grams |
| GST | Glutathione S-transferase |
| HCCA | Alpha-cyano-4-hydroxycinnamic acid |
| HepG2 | Hepatocellular carcinoma |
| HFIP | Hexafuoroisopropanol |
| HG | Hyperglycaemic |
| HILIC | Hydrophilic interaction liquid chromatography |
| *HNF1a* | Hepatic nuclear transcription factor |
| $^{125}I$ | Radioisotope of iodine |
| ID | Internal diameter |
| IDF | International diabetes federation |
| IEC | Ion exchange chromatography |
| IMAC | Immobilized metal affinity chromatography |
| IPF-1 | Insulin promotor factor-1 |
| IPTG | Isopropyl β-D thiogalactoside |
| IU | International units |
| k | Retention factor |
| kb | Kilo-base |

| | |
|---|---|
| kDa | Kilo-dalton |
| kV | Kilovolt |
| L | Litre |
| LB | Luria-bertani |
| LC-MS | Liquid chromatography-mass spectrometry |
| µL | Micro litre |
| µg | Microgram |
| µm | Micrometre |
| MALDI-TOF-MS | Matrix-assisted laser desorption ionisation-time of flight-mass spectrometry |
| mg | Milligram |
| mL | Millilitre |
| mm | Millimetre |
| mM | Millmolar |
| MS | Mass spectrometry |
| MS/MS | Tandem mass spectrometry |
| MTT | 3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyl tetrazolium bromide |
| m/z | mass to charge ratio |
| ng | Nanograms |
| NG | Normoglycaemic |
| Ni | Nickel |
| $Ni^{2+}$ | Nickel ion |
| NCD | Non-communicable disease |
| nm | Nanometre |
| $N_r$ | Not retained |
| N terminal | Amino terminal |
| NTA | Nitrilotriacetic acid |
| OD | Optical density |
| % | Percent |
| p-value | Calculated probability |
| PBS | Phosphate buffered saline |
| PCR | Polymerase chain reaction |
| PDA | Photodiode array |
| PFP | Pentafluoro phenyl |
| pI | Isoelectric point |
| RIA | Radioimmunoassay |
| RP-HPLC | Reversed phase-high performance liquid chromatography |
| rpm | Revolutions per minute |

| | |
|---|---|
| SDS-PAGE | Sodium dodecyl sulphate poly-acrylamide gel electrophoresis |
| SEC | Size exclusion chromatography |
| SFC | Sub/supercritical fluid chromatography |
| TEV | Tobacco Etch Virus |
| TFA | Trifluoroacetic acid |
| tRNA | Transfer ribonucleic acid |
| USA | United States of America |
| USP | United States Pharmacopeia |
| UV | Ultraviolet |
| WHO | World Health Organization |
| $Zn^{2+}$ | Zinc ion |

# List of Figures

**Figures from Chapter 1.**



**Figures from Chapter 2.**

**Figures from Chapter 3.**



**Figures from Chapter 4.**

# List of Tables

**Tables from Chapter 2**



**Tables from Chapter 3**



**Tables from Chapter 4**

# Table of contents

# Chapter 1

## Literature review

## 1.1 The prevalence of diabetes

Diabetes is a non-communicable disease (NCD) that occurs when the pancreas is incapable of producing enough insulin or when the body cannot effectively utilise the insulin it produced [1-3]. Diabetes is considered a global health problem because it is a prevalent non-communicable disease. The number of cases and prevalence of diabetes has steadily increased over the past few decades [1, 4].

In 2014 approximately 422 million adults were living with diabetes, in comparison to 108 million in 1980 [1, 2, 4]. In 2017 it was estimated that 451 million people between the ages of 18-99 years suffered from diabetes [5, 6]. In 2019 approximately 463 million adults suffered from diabetes globally, and 19 million are from Africa. It was estimated that in 2045, the number of people with diabetes would rise to 700 million [5-7]. Concerning the prevalence of diabetes in South Africa, in 2017, it was reported that 1.8 million people suffered from diabetes [7]. However, in 2019 the international diabetes federation (IDF) reported that approximately 4.5 million people were living with diabetes in South Africa, the rapid increase of people suffering from diabetes in South Africa has placed the country on the top 10 countries for an alarming increase in diabetes prevalence [7]. The drastic increase in the number of diabetic patients globally places a burden on current insulin manufacturers. As a result, they would be unable to meet the growing demand for affordable insulin because of limitations in production capacity and high production costs [8, 9]. This research study aims to biosynthesise human insulin using recombinant DNA technology and provide an innovative greener purification method using Sub/supercritical fluid chromatography (SFC).

The lack of access to affordable insulin eventually results in the premature death of diabetic patients [1, 10-12]. In 2012, there were approximately 3.7 million deaths related to blood glucose levels [1]. 2.2 Million deaths were associated with increased risk of cardiovascular as well as other diseases, and approximately 1.5 million deaths occurred amongst people that were diagnosed with diabetes [1]. These deaths are caused by the lack of access to affordable human insulin [1]. There is a need for access to a more affordable and greener process to manufacture insulin. Currently, there is no cure for diabetes; however, this metabolic disorder can be managed using insulin or oral medication [13]. This study will use recombinant DNA technology to biosynthesise human insulin that provides a greener and more affordable purification strategy. The following section provides more information regarding diabetes mellitus.

## 1.2 Diabetes mellitus

Diabetes mellitus is referred to as a metabolic disorder associated with complex aetiology [1, 2, 14]. This metabolic disorder is characterized by chronic hyperglycaemia associated with atypical fat, carbohydrate, and protein metabolism, which is as a result of defective insulin secretion. Diabetes causes severe long-term damage and, eventually, the failure of several organs [1, 2, 13, 15]. The long-term side effects of diabetes mellitus are retinopathy with potential blindness, neuropathy, and nephropathy [1, 2, 4, 13, 15]. People suffering from diabetes are also at a higher risk of cerebrovascular, cardiovascular, and peripheral vascular disease [1, 2, 11, 13, 15, 16].

The development of diabetes is triggered by numerous pathogenetic processes. As such, these processes destroy the beta cells of the pancreas resulting in insulin deficiency, whereas others result in the resistance to insulin action [15, 17]. The anomalies in fat, protein, and carbohydrate metabolisms result in an inability of insulin to target tissues due to the lack of insulin or insensitivity of insulin [15].

Diabetes can be diagnosed based on the amount of glucose found in blood plasma, which can be detected after a fasting plasma glucose test, two-hour plasma glucose (oral glucose tolerance test), or an A1C test. A fasting plasma glucose test can be defined when there is no caloric intake for approximately eight hours. The criterion for the diagnosis of diabetes is if the fasting plasma glucose is $\geq$ 7.0 mmol/L (126 mg/dL) [1, 18]. The glucose tolerance test is conducted with an anhydrous glucose load of 75 grams, and after a two-hour plasma oral glucose tolerance test, the range is $\geq$ 11.1 mmol/L (200mg/dL) [18]. An A1C test can also be performed using a standardized method to the Diabetes Control and Complications Trial (DCCT) with A1C criteria of $\geq$ 48 mmol/mol (6.5%) [18].

The clinical diagnosis of diabetes is associated with symptoms such as recurrent infections, weight loss, increased thirst, and urine volume. Severe symptoms include elevated levels of glycosuria, drowsiness eventually resulting in a coma [13, 15]. There are many forms of diabetes; these forms will be discussed next.

### 1.2.1 Type one diabetes

Type one is associated with processes of beta-cell destruction, which could eventually result in diabetes mellitus, whereby insulin would be required to prevent the onset of a coma and eventually resulting in death [15, 19-21]. It should be noted that type one diabetes is generally associated with the presence of islet cell antibodies, which result in the initiation of autoimmune responses, which eventually destroys the pancreatic beta cells [2, 15, 21].

Autoimmune diabetes mellitus was previously referred to as juvenile-onset diabetes, type one diabetes, or insulin-dependent diabetes. This form of diabetes is a result of autoimmune-mediated destruction of the beta cells located in the pancreas. The beta cells' destruction is variable, ranging from slow in certain individuals to rapid in others [22]. The highest occurrence of this type of diabetes takes place at the onset of adolescence or during childhood; however, the onset of diabetes can occur at any age [23]. This form of diabetes is associated with an individual's genetic predisposition to the autoimmune destruction of beta cells as well as environmental factors surrounding the individual, which is still poorly defined [15].

It must be noted that a few forms of one diabetes mellitus have no identifiable aetiology; this form of diabetes is classified as idiopathic diabetes. These patients have insulinopenia [15, 21] as well as a susceptibility to ketoacidosis [13]; however, there was no evidence of autoimmunity [24].

### 1.2.2. Type two diabetes

Type two diabetes is associated with disorders of insulin secretion and action [2, 5, 13, 14]. Type two diabetes mellitus is rapidly increasing and is common due to the increase in the prevalence of obesity and a sedentary lifestyle [25].

The majority of patients suffering from type two diabetes are generally obese. Obesity exacerbates insulin resistance [26]. Contrary to this, several people are not obese according to conventional weight criteria and could have a higher amount of body fat mainly situated in the central abdominal region [27]. The likelihood of developing type two diabetes increases with the following risk factors; age, obesity, and reduced physical activity [28].

### 1.2.3 Gestational diabetes

Gestational diabetes is characterized by carbohydrate intolerance; the onset or initial recognition is usually during pregnancy in certain individuals [15, 29-31]. Adult women, with a history of glucose intolerance, or high blood glucose levels are at a higher risk for developing gestational diabetes [15].

### 1.2.4 Genetic defects of beta-cell function and insulin action

Monogenic defects in the beta-cell function are associated with numerous forms of diabetes, usually characterized by an early onset of mild hyperglycaemia presenting in patients before the age of twenty-five years. Patients presenting with the above form of diabetes were previously referred to as maturity-onset diabetes of the young (MODY). These patients display reduced secretion of insulin with negligible insulin action [32, 33].

The abnormalities at three genetic loci were characterized, which were located on different chromosomes. The most common mutation is related to mutations located on chromosome twelve in the hepatic nuclear transcription factor known as *HNF1a* [34]. The second mutation is associated with the glucokinase gene located on chromosome 7p [35, 36]. Glucokinase plays an essential role in the conversion of glucose-6-phosphate, whereby the metabolism pathway, in turn, stimulates the secretion of insulin by the beta cells. Therefore, glucokinase functions as a glucose sensor for the beta cells. Due to the defective glucokinase gene, higher levels of glucose are required to mediate normal levels of insulin secretion.

The third type of mutation is associated with a mutation in *HNF4a* located on chromosome 20q [32]. The transcription factor *HNF4a* is associated with gene regulation and expression of *HNF1a*. The fourth form is associated with mutations in a transcription factor gene Insulin promotor factor-1 (*IPF*-1), in its homozygous form results in pancreatic agenesis [33]. Point mutations that occur in mitochondrial DNA were associated with deafness and diabetes mellitus [34]. A common mutation takes place in the transfer ribonucleic acid (tRNA) leucine gene at position 3243, resulting in an adenine to guanine substitution. Genetic defects that result in the failure to convert proinsulin to insulin are inherited in an autosomal dominant pattern with mild carbohydrate intolerance [37, 38].

Concerning genetic defects in insulin action, it must be noted that there are a few rare types of diabetes that occur from genetic anomalies of insulin action. These metabolic anomalies are linked with mutations of the insulin receptor can range from hyperinsulinemia to mild hyperglycaemia [37, 39]. This metabolic syndrome was referred to as Type A insulin resistance [39]. Rabson-Mendenhall and Leprechaunism syndrome are paediatric syndromes that have mutations in the insulin receptor gene, consequently leading to changes in the insulin receptor function and severe insulin resistance [37].

### 1.2.5 Diseases of the exocrine pancreas and endocrinopathies

Processes that injure the human pancreas can lead to diabetes mellitus; these processes include pancreatic carcinoma, pancreatitis, pancreatectomy, trauma, and infection [40]. The damage of the pancreas should be extensive for diabetes to occur, with cancer being an exception to this rule. For example, adenocarcinoma is located in the pancreas and was associated with diabetes [41]. Severe hemochromatosis can also damage beta cells, thereby impairing the secretion of insulin [42, 43].
Endocrinopathies can also be associated with diabetes. Numerous hormones, such as cortisol, glucagon, growth hormone, and epinephrine, inhibit insulin action. Diseases such as Cushing's syndrome, acromegaly, glucagonoma, and phaeochromocytoma are linked to excess secretion of these hormones, as a result, could cause diabetes [40, 44].

### 1.2.6 Drug-induced and infections that cause diabetes

Numerous drugs impede the secretion of insulin. Therapeutic drugs might not result in diabetes; however, these drugs can cause diabetes in patients with insulin resistance [45, 46]. Some toxins such as pentamidine [40, 47] and Vacor [40] can cause the destruction of pancreatic beta cells [40]. These drug reactions are rare. Numerous hormones and drugs can inhibit insulin action, such as glucocorticoids and nicotinic acid [45, 46].

Infections that cause diabetes can be associated with certain viruses and diseases which were linked with the destruction of beta-cells. Some patients with congenital rubella develop diabetes [47]. Cytomegalovirus, adenovirus, and mumps have also been identified in the induction of diabetes mellitus [48-50].

Diabetes mellitus is also associated with a variety of immunological diseases with differing aetiology presented in type one diabetes patients. An autoimmune disorder of the central nervous system that is associated with painful spasms and stiffness of the axial muscles is referred to as stiff-man syndrome [51]. People who have this disorder generally have high titres of GAD autoantibodies, and about approximately half will eventually develop diabetes.

### 1.2.7 Genetic and metabolic syndromes associated with diabetes

Numerous genetic syndromes such as Turner's syndrome, Down's syndrome, Wolfram's syndrome, and Klinefelter's syndrome are associated with an increased risk of diabetes mellitus [52]. Other manifestations associated with diabetes include hypogonadism, neural deafness, diabetes insipidus, and optic atrophy.

A person with diabetes suffers from irregular glucose tolerance and will generally also have one or more cardiovascular disease risk factors [16, 17, 53]. The symptoms of metabolic syndrome are present in patients up to ten years prior to the detection of glycaemic disorders [54]. This is a vital feature regarding the aetiology of cardiovascular disease risk and hyperglycaemia and the possible prevention of cardiovascular disease and mortality/morbidity in people with glucose intolerance [2, 54]. Metabolic syndrome patients with normal glucose tolerance are at high risk for future diabetes. Early detection and management of this syndrome will have an imperative influence on the prevention of cardiovascular disease and diabetes. Metabolic syndrome is associated with impaired glucose tolerance, diabetes mellitus, glucose intolerance, or insulin resistance coupled with insulin resistance, raised arterial pressure and diabetes [54]. Regardless of the prevention of diabetes and metabolic

syndrome, the affordability of readily available human insulin is imperative in the treatment of diabetes and its associated diseases.

### 1.3 The structure of insulin

In 1921, insulin was discovered by Charles Best and Frederick Banting [55]. Insulin plays an essential role in glucose homeostasis [56]. Therefore, it is instrumental in the treatment of diabetes. It is a 51 amino acid (aa) polypeptide. Insulin contains two polypeptide chains, the A chain has 21 amino acids. The B chain contains 30 amino acids. Insulin has three disulphide bonds. Two of the disulphide bonds interlink the A and B chains, whereas the third one is an intra A chain bond [56-61].

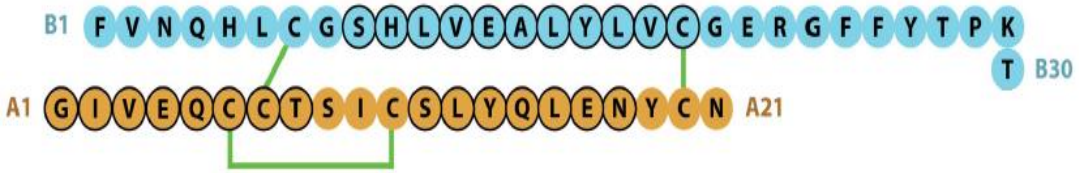


**Figure 1**: Insulin structure depicting polypeptide chains A and B, the A chain is depicted in brown, the B chain is depicted in blue, and the disulphide bonds are depicted in green, adapted from [59] (open access).

The insulin molecules net charge is produced from the ionisation of four tyrosine residues, two histidine residues, four glutamic acid residues, an arginine residue, a lysine residue, two α-amino, and two α-carboxyl groups. Insulin is negatively charged at neutral pH and has an isoelectric point (pI) of 5.3 in its denatured state [62]. Another intrinsic property of insulin is its innate ability to readily associate into higher-order states and dimers (Figure 2) [63]. The insulin molecule can associate into hexameric complexes in the presence of numerous divalent metal ions, for example, zinc [64]. X-ray crystallographic has identified the positions of six phenolic ligand binding sites on an insulin hexamer [65]. The ligand-binding stabilizes a conformational change that takes place at the N-terminal of the B-chain of insulin, shifting this conformational equilibrium of residues from B1 to B8 from a T-state to an R-state. This conformation transition is known as the T<−>R transition (Figure 2) [66].

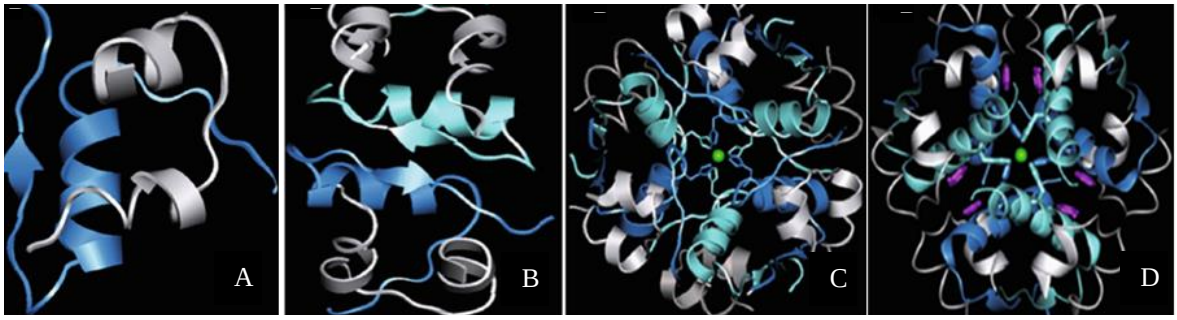


**Figure 2**: (A) An illustration of a T-state monomer of insulin with secondary and tertiary structures, B1-B8 region in an extended conformation. The A-chain is depicted in grey, and the B-chain is blue. (B) An illustration of a T-state dimer of insulin with secondary and tertiary structures. The A-chain is depicted in grey, and the B-chain is blue and cyan. (C**)** An illustration of secondary and tertiary structures of insulin in a T-state hexamer. The A-chain is depicted in grey, and the B-chain is blue and cyan with zinc depicted in green. (D) A representation of secondary and tertiary structures of insulin as an R-state hexamer in the presence of preservative. The A-chain is depicted in grey, and the B-chain is blue and cyan with zinc depicted in green, and the preservative is represented in magenta, picture adapted from [60].

### 1.4 Methods of verifying the biological activity of insulin

Human insulin is essentially in the regulation of glucose uptake into muscle and fat cells [58, 67]. Insulin activates the protein receptor (tyrosine kinase), ion transport systems, and activates the stimulation of glucose within seconds or minutes [58, 67].

There are numerous *in vitro* and *in vivo* tests that are used to determine the biological activity of human insulin. The bio-potency of biosynthesised insulin was tested *in vivo* using a United States Pharmacopeia (USP) rabbit hypoglycaemia assay [68-70]. The *in vitro* assays comprise of radioimmunoassay's (RIA) such as insulin receptor assays [67, 69-72]. In the radioimmunoassay, insulin can be radio-iodinated ($^{125}$I insulin), and the insulin receptor is purified from 3T3-L1 adipocytes [73]. The biological activity of insulin can also be assessed with insulin receptors derived from human lymphocytes and by measuring the lipid metabolism in rat hepatocytes and adipocytes [74]. Another *in vitro* test monitored the hexose uptake in 3T3-Ll adipocytes [73]. In this study, we determined the biological activity of biosynthesised human insulin using a 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide) MTT assay using the Hepatocellular carcinoma (HepG2) cells.

## 1.5 Polymerase chain reaction (PCR)

The polymerase chain reaction is a molecular technique used for mass amplification of DNA [75, 76]. PCR is an *in vitro* enzymatic technique which is used for the synthesis of DNA; two oligonucleotide primers hybridize the opposite strands of the target DNA [77, 78]. PCR was developed by the scientist Kary Mullis in 1984 [76, 79]. The PCR reaction requires three main components, such as a polymerase, template of DNA as well as the building blocks of DNA the nucleotide bases such as guanine, adenine, thymine, and cytosine [77].

It is an exponential amplification process that comprised of three steps: extension, annealing, and elongation [75, 77, 80, 81]. During the denaturation step, the DNA is exposed to high temperatures such as 90 °C to 97 °C and denatured. In the annealing step, the primers anneal to the DNA target strands. In the extension step, an extension of the DNA occurs, creating a complimentary copy strand of nucleic acid [75, 80].

PCR is a rapid confirmatory test that is highly specific and sensitive [75, 80]. PCR has a vast array of applications and is used in many fields of science, such as molecular biology and microbiology for DNA sequencing, cloning, and recombinant DNA technology [75, 81]. It's also used in forensic labs to amplify DNA from hair and blood samples, where a small amount of original DNA was available [75]. This study used PCR to clone human insulin and verify the positive transformants using colony PCR.

## 1.6 The biosynthesis of insulin using recombinant DNA technology

The first report of the synthesis of insulin by recombinant DNA technology was in 1978 by Goeddel and co-workers [82, 83]. In the initial biosynthesis of insulin, the A and B chains were combined and expressed separately in *E. coli*, yielding approximately 20 nanograms (ng) of insulin, which was detected by radioimmunoassay [70, 83]. This novel breakthrough in science started a major shift in insulin production, from the traditional animal pancreatic derived insulin to recombinant biosynthesis of insulin [70, 83, 84].

The use of recombinant technology is imperative for the production of recombinant proteins utilised for the development of protein drugs. Furthermore, recombinant protein production has an essential role in drug development. Recombinant proteins are produced in bacterial hosts, accounting for an estimated thirty percent of the current biopharmaceuticals in the market [85].

In 1978, the gene encoding for human insulin was cloned and expressed in *E. coli*. Thereafter in 1982, Genentech developed human insulin, which was licensed and marketed by Eli Lilly [8, 61, 70, 84]. This was the first licensed drug produced using recombinant DNA technology [8].

Initially, recombinant human insulin was produced by the separate expression of insulin A and B chains in *E. coli* [61, 68-70, 83]. After purification of these two chains were incubated under oxidizing conditions. This promoted the formation of intact insulin by disulphide bond formation resulting in mature insulin [69, 83]. Subsequently, an alternate expression method was developed in *E. coli* utilising the nucleotide sequence encoding for human proinsulin. Purification of proinsulin was followed by *in vitro* proteolytic excision of the connecting C peptide, resulted in mature insulin [61, 68, 70, 83]. This method was more efficient for the large scale synthesis of insulin as compared to the two chain combination approach and has been used commercially since 1986 [83].

### 1.7 The use of affinity tags in protein purification

Approximately 30% of therapeutic proteins are biosynthesised in *E. coli* using recombinant technology. There is comprehensive knowledge available regarding the fermentation profile of *E. coli*, expression vectors, strains, and protein folding [8, 86, 87]. *E.coli* has the following positive attributes such as its genetics are well characterized, displays rapid growth, and has a high production yield of proteins, easy to handle, and it is amenable to numerous genetic modifications thereby making it an ideal expression host [8, 86, 88, 89]. *E. coli* is an ideal bacterial expression host for large scale production of therapeutic proteins [87, 90-94]. The expression of a therapeutic protein requires a multidisciplinary approach [90]. There have been advances made with regards to protein refolding from inclusion bodies, molecular biology, and fermentation utilising ***E****. coli*. However, some bottlenecks that impact the cost-effectiveness of the biosynthesis of therapeutic proteins, such as the purification of therapeutic proteins from inclusion bodies, which result in low yields of the bioactive protein [90, 95].

Affinity tags are referred to as exogenous amino acid sequences that have a high affinity for a particular chemical or biological ligand. These tags are efficient tools for protein purification. The protein of interest and the tag are expressed as a single entity and subsequently purified. The tag has to be removed from the target protein as it can affect the physiochemical characteristics of the protein [96]. The cost incurred in the tag removal prevents large scale industrial applications [96, 97]. Due to advances in cutting edge technology in bioinformatics, proteomics, and genomics, the number of recombinant proteins is increasing rapidly [98]. Consequently, high throughput screening methods are required to rapidly identify the potential proteins with therapeutic or industrial enzymes [99].

Affinity tags have an essential role in protein purification. Affinity chromatography can achieve 90-99% purity. The use of affinity chromatography saves time due to the reduction in the number of chromatographic steps. As such, affinity tag technology will increase high-throughput purification, drug delivery methods, and large scale production plants [100]. There are a variety of affinity tags; this study employed histidine and glutathione *S*-transferase tags.

Immobilized metal affinity chromatography (IMAC) is extensively used to purify recombinant proteins with a small affinity tag comprising of poly-histidine residues. IMAC is based on the interaction with the specific amino acid side chains and a transition metal ion (such as a Zinc ion ($Zn^{2+}$), Copper ion ($Cu^{2+}$), Nickel ion ($Ni^{2+}$), Cobalt ion ($Co^{2+}$)) immobilized in a matrix [101]. The histidine amino acid exhibits the strongest interaction between the immobilized metal ions in the matrix. The electron donor groups located in the histidine imidazole ring form coordination bonds with the immobilized transition metal. Peptides that contain consecutive histidine residues are retained on the IMAC column matrix. After washing the matrix, the peptides which contain poly-histidine sequences are eluted by adding free imidazole or by adjusting the pH of the column buffer [100]. The poly-histidine affinity tag, also referred to as His six or His-tag, contains six consecutive histidine residues, however, can differ in size from two to approximately ten histidine residues [102].

Histidine tags are one of the most commonly used affinity tags. The purification is based on the use of chelated metal ions as affinity ligands. IMAC protein separation occurs based on the interaction between metal ions contained in an immobilized metal matrix and particular amino acid residues such as histidine [103]. The imidazole group of histidine has a high affinity for chelated metals [66]. The use of histidine tags at the carboxyl (C) or amino (N) terminals aids in the protein purification of the recombinant protein from the crude extract of the expression host in one step. There are a variety of strategies available for IMAC [103, 104], the most prevalent IMAC supports different chelating sepharose matrices or either nitrilotriacetic acid (NTA) as a ligand for immobilizing metals such as nickel (Ni–NTA). The binding specificity allows the purification of proteins under denaturing and native conditions [103]. The histidine tag is cleaved by a cleavage enzyme such as Tobacco Etch Virus (TEV) protease [105].

The pGex (Glutathione S-transferase) GST fusion protein system is used for the high-level purification of fusion proteins from bacterial and eukaryotic cell lysates [106]. In 1988 the single-step purification of polypeptides as fusions with GST was described [107]. These fusion proteins can be purified from the crude lysate by utilising affinity chromatography on immobilized glutathione matrices. Furthermore, the bound fusion proteins are eluted with reduced glutathione (10 mM) in non-denaturing conditions. The affinity tag can assist in the protection against intracellular protease cleavage and thereby, stabilize the cloned protein of interest [106].

The GST tag can be cleaved from the fusion protein while still attached to glutathione agarose, thereby providing a convenient method for the separation of the 26 kDa (kilo-dalton) GST tag from the protein of interest [102].

The removal of the affinity tag must be taken into consideration, and these proteases must be removed from the purified protein if the protein is intended for human use. Affinity tags can affect the functions and characteristics of the recombinant protein. Therefore the removal of the affinity tag from the protein may be achieved by the use of site-specific proteases [100]. Enterokinase is generally used for N-terminal fusions as it recognizes a five amino-acid polypeptide and cleaves the carboxyl end of lysine (D-D-D-D-K-X1) [100].

### 1.8 Traditional methods for insulin purification

Chromatography is an imperative separation method that allows for quantitative and qualitative analysis separation, identification, and purification of individual components of a mixture [108]. Biologics such as insulin are currently purified using reversed-phase high-performance liquid chromatography (RP-HPLC) [84].

The strategies of insulin biosynthesis in yeast and *E. coli* contain enzymatic and chemical reactions in the downstream processes. Thus, resulting in the formation of numerous derivatives associated with insulin, which is tedious to remove by conventional purification. Novo Nordisk A/S and other companies utilised RP-HPLC for protein purification [68]. The characteristics of a protein have a fundamental role in its purification, such as the total charge of a protein, hydrophobic groups displayed on the protein surface, the binding capacity of the protein to a stationary phase, shape, and size [108].

Samples can be separated by various forms of chromatography, such as affinity, ion exchange, and size exclusion chromatography (SEC) [109]. HPLC is a chromatographic separation technique that is extensively used for the quantitative separation and purification of steroids, amino acids, proteins, antibiotics, vitamins, hydrocarbons, nucleic acids, carbohydrates, polysaccharides and drugs [108, 109]. HPLC system was comprised of an injector, pumps, columns, detectors, and a data system. The sample is injected into the system by an injector, allowing the placement of the sample into the mobile phase, thereafter it is transported to the column. The pump transports the mobile phase throughout the system. The column comprises of a polymer or stainless-steel hardware that is filled with a specific separation packing material. Numerous packing materials comprise of polymeric such as microporous, mesoporous, macro-porous and nonporous or silica based such as bonded phases, or porous silica. The success of silica-based separation has led to numerous applications in reverse and

normal phase HPLC separation. After the sample is passed to the column, it is detected by a detector, the detectors used in HPLC can include light scattering, fluorescence, electrochemical, and ultraviolet (UV)-Vis absorption. Analytical systems can also be coupled to mass spectrometers [109]. A schematic representation of an HPLC is illustrated in Figure 3. HPLC is highly versatile and has analytical precision. There are numerous advantages over low-pressure liquid chromatography, such as better resolution, higher sensitivity, speed, the availability of a multitude of stationary phases, and easy sample recovery [109].

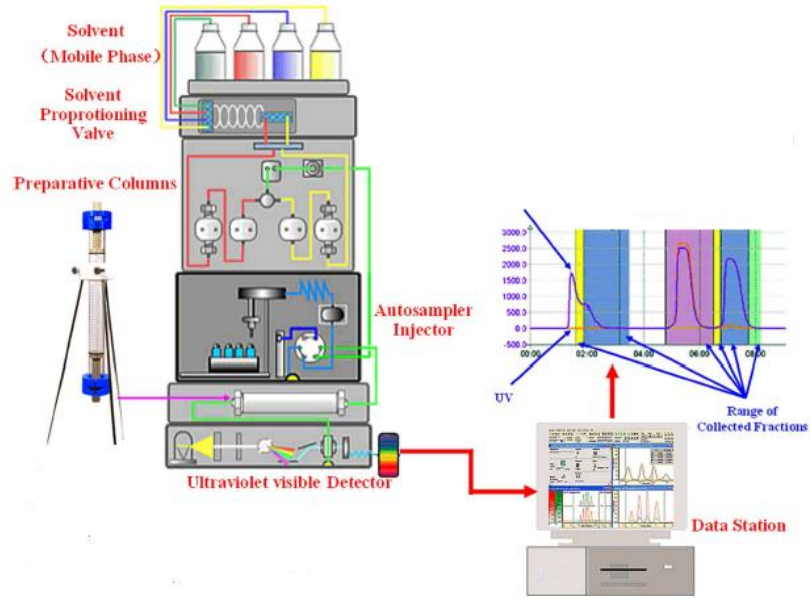


**Figure 3:** A schematic diagram of an HPLC system [110] (open access).

HPLC is the conventional method employed for the purification of biologics such as insulin [61, 68, 84]; however, this study will employ a greener form of SFC purification, which will be discussed next.

### 1.9 SFC purification

SFC is a purification technique that utilises a supercritical fluid as a mobile phase. This supercritical fluid is a condensed gas, such as carbon dioxide ($CO_2$) [111-114]. The aforementioned fluid is generally above the critical pressure and temperature [111-113]. $CO_2$ reaches its supercritical fluid state at a pressure of 73 bar and a temperature of 31°C [111-113]. A $CO_2$ phase diagram is illustrated in Figure 4.

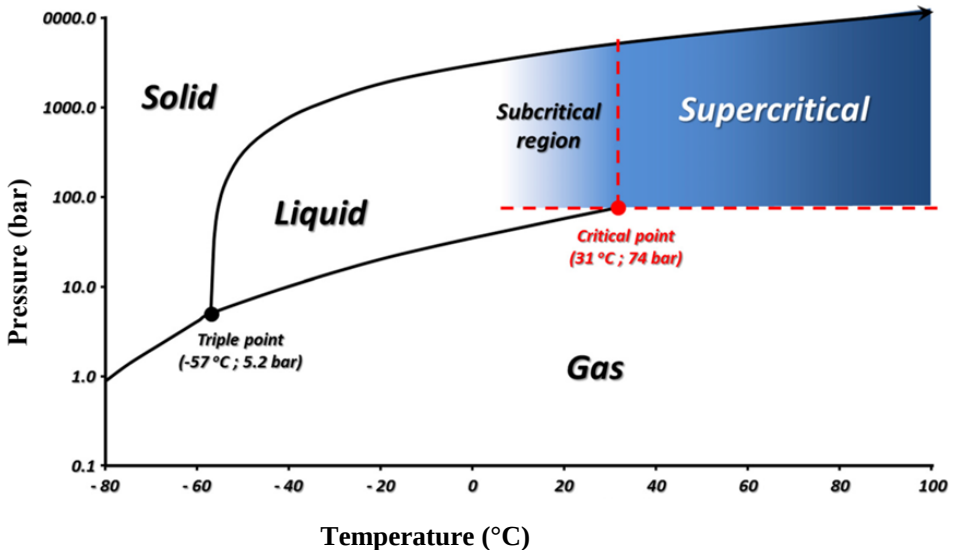


**Figure 4:** $CO_2$ phase diagram adapted from [115].

In an SFC system, the mobile phase is transported at a cold temperature such as 4°C to the pump and heated in the column oven above the supercritical temperature [116]. A back pressure regulator is used to maintain critical pressure throughout the SFC system [111, 117]. The sample is injected into the SFC system *via* an injector. The supercritical fluid carries the sample through a separation column. The sample is detected using a UV detector [111]. In order to improve the elution of organic substances, modifiers can be added [111]. Subsequently, the $CO_2$ is separated from the modifier and the eluted compounds *via* the expansion in the gas and liquid separators. The majority of the separated $CO_2$ leaves the system through an exhaust pipe and sent to a ventilation system. Thereafter, the fractions are collected; these fractions contain residual $CO_2$ and the separated compound [111, 113]. A schematic representation of the SFC is illustrated in Figure 5.

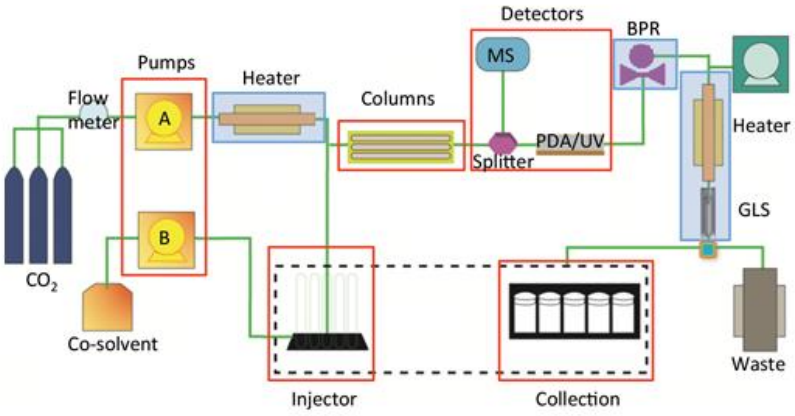


**Figure 5**: A schematic representation of a typical SFC flow path [118] (open access).

The SFC kinetics and thermodynamics provides a faster separation method than HPLC. SFC has the following advantages over HPLC, such as higher efficiency separation of drug substances and intermediates, short analysis time, faster method development, stacked injections, and column equilibration [112, 113, 117, 119-121]. There is a reduction of toxic and hazardous waste generated from SFC purification; as a result, it is a greener process [112, 113, 117]. The reduction in toxic solvent waste has health, safety, and cost benefits [112, 113]. The high-speed throughput of SFC purification is being used rapidly for semi-preparative and preparative as well as analytical purification of chiral compounds, which is extensively used in drug development [112, 113, 117]. Another advantage of SFC is the ease of recycling an unresolved fraction when the overlapping of chromatogram peaks occurs and poor peak resolution [117]. SFC can be utilised as an alternative purification method to HPLC for drug substances; as such, SFC is replacing HPLC in several pharmaceutical industries, such as purification, screening, and method development for chiral compounds such as enantiomers [112, 113, 117, 121]. HPLC uses toxic solvents such as acetonitrile. This study aims to use methanol as a green solvent to provide a novel method of purifying human insulin and other biologics. SFC is considered to be an environmentally friendly technology; it eliminates the use of hexane, heptane and acetonitrile, which is used in HPLC [113, 117]. As far as we are aware, this is one of the first studies to purify biosynthesised human insulin using SFC.

### 1.10 Purification of peptides

Peptides are used in the pharmaceutical industry and are highly versatile drugs as they can be employed for the treatment of type two diabetes [122]. Peptides display numerous advantages as drugs such as low toxicity, high biological activity, and high specificity [122, 123]. Peptides can be antimicrobial, antibacterial, anti-cancer, and anti-viral [122, 124, 125]. Peptides are used in various therapeutic fields, such as diabetes, arthritis, allergy, oncology, cardiovascular, and diagnostic research [123, 126].

There are numerous challenges associated with developing peptide drugs such as stability, solubility; however, the major hurdle associated with the peptide manufacturing process is peptide purification [123]. The downstream processes are very expensive due to yield losses experienced during HPLC purification, which is another major bottleneck experienced during the peptide manufacturing process [123, 125]. Therefore, extensive research should be conducted in the process of the development of peptide drugs to save time and costs [123]. Therefore, this study aims to purify peptides using SFC purification.

## 1.11 Matrix-assisted laser desorption ionisation-time of flight mass spectrometry

Matrix-assisted laser desorption ionisation-time of flight mass spectrometry (MALDI-TOF-MS) is an inexpensive method used for the rapid detection of fungal and bacterial organisms [127-137]. Mass spectrometry is an analytical method that is semi-quantitative and used for the elucidation of molecular structure or the composition of unidentified samples [127]. MALDI-TOF-MS is considered a soft ionisation technique because it ionizes biomolecules without causing fragmentation that is normally associated with traditional ionisation methods. The lack of fragmentation of small analytes allows for the identification of specific analytes against a backdrop of other signals; therefore, the presence of metabolites and small drugs can be identified from a range of other biomolecules [138].

Instruments that are used in mass spectrometry are comprised of three components an ionisation chamber, followed by a mass analyser and a detector (Figure. 6) [127]. In MALDI-TOF-MS a target plate is initially placed under vacuum, and a co-crystallized mixture of matrix and analyte is irradiated by a laser beam, resulting in localized heating of the analyte followed by desorption and ionisation, an ionic plume of ions are accelerated into the mass analyser by electric fields, thereafter the mass of the analyte is detected [138]. The matrix is used in MALDI to protect molecules from thermal degradation after irradiation by a pulsed laser, as the photons are absorbed by the matrix and thereafter transferred to the analyte in a relatively controlled manner [139] (Figure 6).

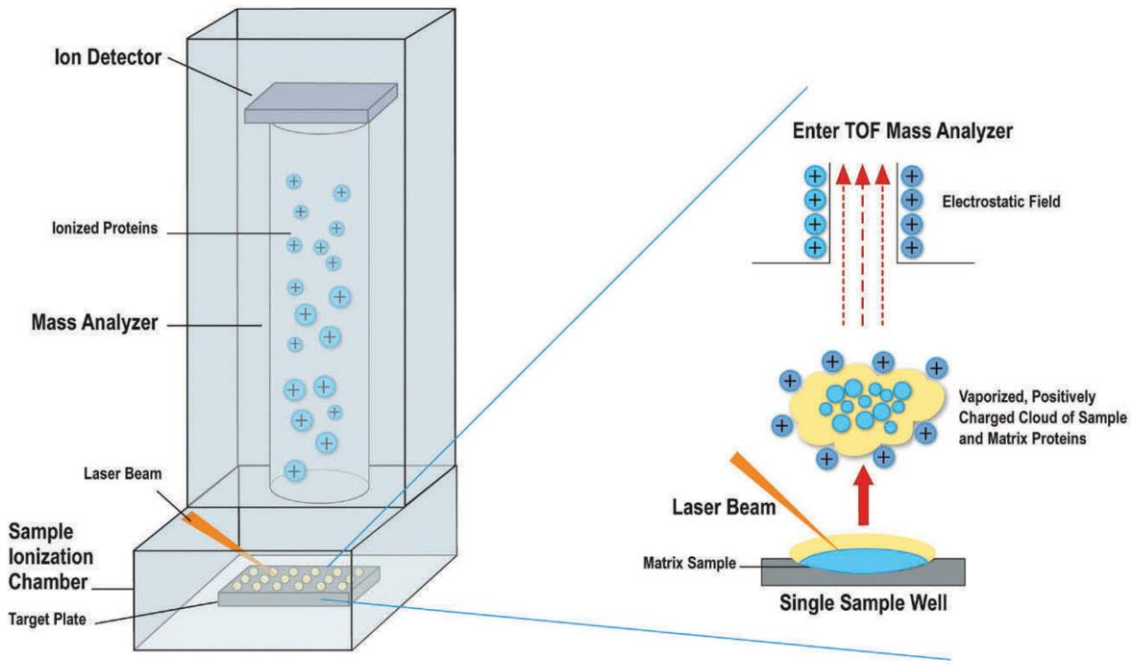


**Figure 6:** MALDI-TOF-MS sample ionisation process [140].

There are numerous advantages in the use of MALDI-TOF-MS, such as it does not require chromatographic separation, easy to use, fast detection of a variety of samples, and relatively simple sample preparation [129, 130, 141, 142]. A disadvantage of MALDI-TOF-MS is that it is less quantitative due to the efficiency of ionisation as a result of the heterogeneous distribution of the analyte and matrix co-crystallization [141, 142]. Other disadvantages include variability in signal intensity and poor reproducibility [142]. The shot to shot reproducibility is affected by the depth of the crystal layer, crystal sizes, and the distribution of the analyte throughout the crystallized matrix and analyte mixture [143]. This study employed MALDI-TOF-MS to detect biosynthesised human insulin.

### 1.12 Protein sequencing

Mass spectrometry has become an imperative tool for the determination and elucidation of amino acid sequences of peptides and proteins [144]. There are soft ionisation techniques such as MALDI [145] and electrospray ionisation [146], which are used to produce intact protein and peptide ions with minimal fragmentation located in the gas phase. There are more fragment ions detected in the above ionisation techniques compared to chemical ionisation and electron impact mass spectrometry [147]. These ions are rapidly detected with their mass to charge (m/z) ratios allowing for quick validation of peptides and proteins, which also allows for peptide mass fingerprinting [148-150]. Mass spectrometry is extensively used for the determination of peptides structure due to the different ionisation techniques available [147].

Peptide mass fingerprinting is based on the digestion of a protein or peptide by an enzyme such as trypsin; thereafter, the peptides are analysed using reverse-phase chromatography [144, 151]. The peptides are ionized and fragmented to form a signature from tandem mass spectrometry (MS/MS) spectra, which are used for rapid identification [151]. The peptides are identified by comparing the generated MS/MS spectra against the peptide sequences in peptide databases such as Mascot and SEQUEST [144, 152-154].

The peptide fragmentation pattern is dependent on numerous factors such as the charge of the ion, the size of the peptide, the amino acid composition, and the excitation method employed [144]. The peptide precursor ions dissociate under low collision energy parameters, which fragment the peptide backbone at the amide bonds [155-157]. There are three potential cleavage points in a peptide backbone A, B and C if there is a charge retained at the N-terminal of the peptide and if the charge is retained at the C terminal of a peptide then it is given the nomenclature of X, Y, and Z [147].

## 1.13 Aims and objectives of the study

1.13.1 To develop a novel and more efficient method for human insulin biosynthesis using recombinant DNA technology in *E. coli* (Chapter 2).

1.13.1.1 To verify human insulin using chromosomal DNA *via* PCR.

1.13.1.2 To detect and identify biosynthesised human insulin using MALDI-TOF-MS.

1.13.1.3 To conduct Liquid Chromatography-Mass Spectrometry (LC-MS) on the standard human insulin and biosynthesised human insulin samples.

1.13.1.4 To conduct protein sequencing of standard human insulin and biosynthesised human insulin.

1.13.1.5 To determine the biological activity of the standard human insulin and the crude biosynthesised human insulin using an *in vitro* MTT assay.

1.13.1.6 To conduct LC-MS of the standard human insulin and biosynthesised SFC purified human insulin samples.

1.13.2 Can human insulin be purified using SFC? (Chapter 3).

1.13.2.1 To purify human insulin using SFC.

1.13.2.2 To conduct protein sequencing of human insulin after SFC purification.

1.13.2.3 To determine the biological activities of standard human insulin and biosynthesised human insulin using an *in vitro* MTT assay.

1.13.3. Can crude peptides such as a tetrapeptide, octapeptide and a nonapeptide be purified using SFC? (Chapter 4).

1.13.3.1 To determine if a tetrapeptide, octapeptide, and a nonapeptide are soluble in methanol.

1.13.3.2 To develop an SFC purification method for a crude tetrapeptide, octapeptide and a nonapeptide.

1.13.3.3 To conduct LC-MS on the SFC purified tetrapeptide, octapeptide and nonapeptide.

## 1.14 Outline of thesis

This dissertation is divided into five chapters, including **Chapter one**.

**Chapter one** provides an overview of the current understanding of diabetes and the characteristics of insulin biosynthesis and purification.

**Chapter two** contains research data pertaining to the biosynthesis and optimisation of human insulin expression in *E. coli.*

**Chapter three** comprises of research data pertaining to the SFC purification of human insulin.

**Chapter four** pertains to research data regarding the SFC purification of peptides.

Lastly, **Chapter five** describes conclusions drawn from this study and reflects on future research endeavours.

## 1.15 References


1. WHO. *Global report on Diabetes, Available at http://www.who.int.* 2016.
2. Shahi, A., V. Prasad, S.S. Imam, and M. Abdul, *Pathophysiological Ramifications of Diabetic Condition: A Review.* Asian Journal of Biomedical and Pharmaceutical Sciences, 2018. **8**: p. 28-38.
3. Tuso, P., *Prediabetes and lifestyle modification: time to prevent a preventable disease*, in *The Permanente Journal*. 2014. p. 88.
4. Rao, G.H., *Management of Diabetes Epidemic: A global perspective.* EC Endocrinol Metab Res, 2018. **3**(2): p. 63-72.
5. Cho, N.H., J.E. Shaw, S. Karuranga, Y. Huang, J.D. da Rocha Fernandes, A.W. Ohlrogge, and B. Malanda, *IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045.* Diabetes research and clinical practice, 2018. **138**: p. 271-281.
6. Shoback, D. and D. Gardner, *Chapter 17.* Greenspan's Basic & Clinical Endocrinology (9th Ed.). New York: McGraw-Hill Medical, 2011.
7. Atlas, D., *International diabetes federation.* IDF Diabetes Atlas, 9th edn. Brussels, Belgium: International Diabetes Federation, 2019.
8. Baeshen, N.A., M.N. Baeshen, A. Sheikh, R.S. Bora, M.M.M. Ahmed, H.A. Ramadan, K.S. Saini, and E.M. Redwan, *Cell factories for insulin production.* Microbial Cell Factories, 2014. **13**(1): p. 1.
9. Nykiforuk, C.L., J.G. Boothe, E.W. Murray, R.G. Keon, H.J. Goren, N.A. Markley, and M.M. Moloney, *Transgenic expression and recovery of biologically active recombinant human insulin from Arabidopsis thaliana seeds.* Plant Biotechnology Journal, 2006. **4**(1): p. 77-85.
10. Association, A.D., *Economic costs of diabetes in the US in 2017.* Diabetes care, 2018. **41**(5): p. 917-928.
11. Gregg, E.W., Y.J. Cheng, S. Saydah, C. Cowie, S. Garfield, L. Geiss, and L. Barker, *Trends in death rates among US adults with and without diabetes between 1997 and 2006: findings from the National Health Interview Survey.* Diabetes care, 2012. **35**(6): p. 1252-1257.
12. Saydah, S.H., E.W. Gregg, H.S. Kahn, and M.K. Ali, *Mortality associated with less intense risk-factor control among adults with diabetes in the United States.* Primary care diabetes, 2018. **12**(1): p. 3-12.
13. Orieke, N., O. Asaolu, T. Fashanu, and O. Fasanmade, *A Coupled Insulin and Meal Effect Neuro-Fuzzy Model for The Prediction of Blood Glucose Level in Type 1 Diabetes Mellitus Patients.* Annals of Science and Technology, 2019.
14. Chaudhury, A., C. Duvoor, V.S. Reddy Dendi, S. Kraleti, A. Chada, R. Ravilla, A. Marco, N.S. Shekhawat, M.T. Montales, K. Kuriakose, A. Sasapu, A. Beebe, N. Patil, C.K. Musham, G.P.

Lohani, and W. Mirza, *Clinical Review of Antidiabetic Drugs: Implications for Type 2 Diabetes Mellitus Management.* Frontiers in Endocrinology, 2017. **8**(6).

15. Alberti, K.G.M.M. and P.f. Zimmet, *Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus. Provisional report of a WHO consultation.* Diabetic medicine, 1998. **15**(7): p. 539-553.
16. Franco, O.H., E.W. Steyerberg, F.B. Hu, J. Mackenbach, and W. Nusselder, *Associations of diabetes mellitus with total life expectancy and life expectancy with and without cardiovascular disease.* Archives of internal medicine, 2007. **167**(11): p. 1145-1151.
17. Zheng, Y., S.H. Ley, and F.B. Hu, *Global aetiology and epidemiology of type 2 diabetes mellitus and its complications.* Nature Reviews Endocrinology, 2018. **14**(2): p. 88.
18. Association, A.D., *2. Classification and diagnosis of diabetes: standards of medical care in diabetes—2018.* Diabetes care, 2018. **41**(Supplement 1): p. S13-S27.
19. Eiswirth, M., E. Clark, and M. Diamond, *Low carbohydrate diet and improved glycaemic control in a patient with type one diabetes.* Endocrinology, diabetes & metabolism case reports, 2018. **2018**(1).
20. Bending, D., P. Zaccone, and A. Cooke, *Inflammation and type one diabetes.* International immunology, 2012. **24**(6): p. 339-346.
21. Chowdhury, S., *Clinical assessment of obesity and insulin resistance in type 1 diabetes subjects seen at a center in Kolkata.* JAPI, 2009. **57**: p. 511.
22. Zimmet, P., T. Tuomi, I. Mackay, M. Rowley, W. Knowles, M. Cohen, and D. Lang, *Latent autoimmune diabetes mellitus in adults (LADA): the role of antibodies to glutamic acid decarboxylase in diagnosis and prediction of insulin dependency.* Diabetic medicine, 1994. **11**(3): p. 299-303.
23. Mølbak, A., B. Christau, B. Marner, K. Borch-Johnsen, and J. Nerup, *Incidence of insulin-dependent diabetes mellitus in age groups over 30 years in Denmark.* Diabetic medicine, 1994. **11**(7): p. 650-655.
24. McLarty, D., I. Athaide, G. Bottazzo, A. Swai, and K. Alberti, *Islet cell antibodies are not specifically associated with insulin-dependent diabetes in Tanzanian Africans.* Diabetes research and clinical practice, 1990. **9**(3): p. 219-224.
25. Tuomilehto, J., J. Lindström, J.G. Eriksson, T.T. Valle, H. Hämäläinen, P. Ilanne-Parikka, S. Keinänen-Kiukaanniemi, M. Laakso, A. Louheranta, and M. Rastas, *Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance.* New England Journal of Medicine, 2001. **344**(18): p. 1343-1350.
26. Campbell, P.J. and M.G. Carlson, *Impact of obesity on insulin action in NIDDM.* Diabetes, 1993. **42**(3): p. 405-410.

27. Kissebah, A.H., N. Vydelingum, R. Murray, D.J. Evans, R.K. KALKHOFF, and P.W. ADAMS, *Relation of body fat distribution to metabolic complications of obesity.* The Journal of Clinical Endocrinology and Metabolism, 1982. **54**(2): p. 254-260.
28. Zimmet, P.Z., *Kelly West Lecture 1991 challenges in diabetes epidemiology—from West to the rest.* Diabetes care, 1992. **15**(2): p. 232-252.
29. Flegal, K.M., M.D. Carroll, B.K. Kit, and C.L. Ogden, *Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999-2010.* Jama, 2012. **307**(5): p. 491-497.
30. Coustan, D.R., *Gestational diabetes mellitus.* Clinical chemistry, 2013. **59**(9): p. 1310-1321.
31. Barbour, L.A., *Unresolved controversies in gestational diabetes: implications on maternal and infant health.* Current Opinion in Endocrinology, Diabetes and Obesity, 2014. **21**(4): p. 264-270.
32. Byrne, M.M., J. Sturis, S. Menzel, K. Yamagata, S.S. Fajans, M.J. Dronsfield, S.C. Bain, A.T. Hattersley, G. Velho, and P. Froguel, *Altered insulin secretory responses to glucose in diabetic and nondiabetic subjects with mutations in the diabetes susceptibility gene MODY3 on chromosome 12.* Diabetes, 1996. **45**(11): p. 1503-1510.
33. Clement, K., M. Pueyo, M. Vaxillaire, B. Rakotoambinina, F. Thuillier, P. Passa, P. Froguel, J.-J. Robert, and G. Velho, *Assessment of insulin sensitivity in glucokinase-deficient subjects.* Diabetologia, 1996. **39**(1): p. 82-90.
34. Yamagata, K., N. Oda, P.J. Kaisaki, S. Menzel, H. Furuta, M. Vaxillaire, L. Southam, R.D. Cox, G.M. Lathrop, and V.V. Boriraj, *Mutations in the hepatocyte nuclear factor-1 alpha gene in maturity-onset diabetes of the young (MODY3).* 1996.
35. Froguel, P. and M. Vaxillaire, *Close linkage of glucokinase locus on chromosome 7p to early-onset non-insulin-dependent diabetes mellitus.* Nature, 1992. **356**(6365): p. 162.
36. Vionnet, N. and M. Stoffel, *Nonsense mutation in the glucokinase gene causes early-onset non-insulin-dependent diabetes mellitus.* Nature, 1992. **356**(6371): p. 721.
37. Gruppuso, P.A., P. Gorden, C.R. Kahn, M. Cornblath, W.P. Zeller, and R. Schwartz, *Familial hyperproinsulinemia due to a proposed defect in conversion of proinsulin to insulin.* New England Journal of Medicine, 1984. **311**(10): p. 629-634.
38. Robbins, D., S. Shoelson, A. Rubenstein, and H. Tager, *Familial hyperproinsulinemia. Two cohorts secreting indistinguishable type II intermediates of proinsulin conversion.* Journal of Clinical Investigation, 1984. **73**(3): p. 714.
39. Johns, D.R., *Mitochondrial DNA and disease.* New England Journal of Medicine, 1995. **333**(10): p. 638-644.
40. Association, A.D., *Diagnosis and classification of diabetes mellitus.* Diabetes care, 2010. **33**(Supplement 1): p. S62-S69.

41. Sanz, N., J. Karam, S. Horita, and G. Bell, *Prevalence of insulin-gene mutations in non-insulin-dependent diabetes mellitus.* The New England journal of medicine, 1986. **314**(20): p. 1322-1323.
42. Kahn, C.R., J.S. Flier, R.S. Bar, J.A. Archer, P. Gorden, M.M. Martin, and J. Roth, *The syndromes of insulin resistance and acanthosis nigricans: insulin-receptor disorders in man.* New England Journal of Medicine, 1976. **294**(14): p. 739-745.
43. Taylor, S.I., *Lilly Lecture: molecular mechanisms of insulin resistance: lessons from patients with mutations in the insulin-receptor gene.* Diabetes, 1992. **41**(11): p. 1473-1490.
44. Larsen, S., J. Hilsted, B. Tronier, and H. Worning, *Metabolic control and B cell function in patients with insulin-dependent diabetes mellitus secondary to chronic pancreatitis.* Metabolism, 1987. **36**(10): p. 964-967.
45. Phelps, G., P. Hall, I. Chapman, W. Braund, and M. Mackinnon, *Prevalence of genetic haemochromatosis among diabetic patients.* The lancet, 1989. **334**(8657): p. 233-234.
46. Yajnik, C., K. Shelgikar, S. Naik, S. Kanitkar, H. Orskov, K. Alberti, and T. Hockaday, *The ketosis-resistance in fibro-calculous-pancreatic-diabetes. 1. Clinical observations and endocrine-metabolic measurements during oral glucose tolerance test.* Diabetes research and clinical practice, 1992. **15**(2): p. 149-156.
47. Pandit, M.K., J. Burke, A.B. Gustafson, A. Minocha, and A.N. Peiris, *Drug-induced disorders of glucose tolerance.* Annals of internal medicine, 1993. **118**(7): p. 529-539.
48. O'Byrne, S. and J. Feely, *Effects of drugs on glucose tolerance in non-insulin-dependent diabetics (Part I).* Drugs, 1990. **40**(1): p. 6-18.
49. Gallanosa, A., D. Spyker, and R. Curnow, *Diabetes mellitus associated with autonomic and peripheral neuropathy after Vacor rodenticide poisoning: a review.* Clinical toxicology, 1981. **18**(4): p. 441-449.
50. Esposti, M.D., A. Ngo, and M.A. Myers, *Inhibition of mitochondrial complex I may account for IDDM induced by intoxication with the rodenticide Vacor.* Diabetes, 1996. **45**(11): p. 1531-1534.
51. King, M., D. Bidwell, A. Shaikh, A. Voller, and J. Banatvala, *Coxsackie-B-virus-specific IgM responses in children with insulin-dependent (juvenile-onset; type I) diabetes mellitus.* The lancet, 1983. **321**(8339): p. 1397-1399.
52. Solimena, M. and P. De Camilli, *Autoimmunity to glutamic acid decarboxylase (GAD) in stiffman syndrome and insulin-dependent diabetes mellitus.* Trends in neurosciences, 1991. **14**(10): p. 452-457.
53. Tripathy, B.B. and K.C. Samal, *Overview and consensus statement on diabetes in tropical areas.* Diabetes/Metabolism Research and Reviews, 1997. **13**(1): p. 63-76.
54. Barrett, T.G., S.E. Bundey, and A.F. Macleod, *Neurodegeneration and diabetes: UK nationwide study of Wolfram (DIDMOAD) syndrome.* The lancet, 1995. **346**(8988): p. 1458-1463.

55. Banting, F. and C. Best, *The Internal Secretion of the Pancreas.* Journal of Laboratory and Clinical Medicine, 1922. **7**(5): p. 465-480.
56. Ahmad, B., *Review: Pharmacology of insulin.* Br J Diabetes Vasc Dis, 2004. **4**(1): p. 10-14.
57. Vajo, Z., J. Fawcett, and W.C. Duckworth, *Recombinant DNA technology in the treatment of diabetes: insulin analogs.* Endocrine reviews, 2001. **22**(5): p. 706-717.
58. Malik, A., A. Al-Senaidy, E. Skrzypczak-Jankun, and J. Jankun, *A study of the anti-diabetic agents of camel milk.* International journal of molecular medicine, 2012. **30**(3): p. 585-592.
59. Lawrence, M., *Landmarks in insulin research.* Frontiers in Endocrinology, 2011. **2**: p. 76.
60. Beals, J.M., M.R. DeFelippis, P.M. Kovach, and J.A. Jackson, *Insulin*, in *Pharmaceutical Biotechnology*. 2013, Springer. p. 255-275.
61. Walsh, G., *Therapeutic insulins and their large-scale manufacture.* Applied Microbiology and Biotechnology, 2005. **67**(2): p. 151-159.
62. Kaarsholm, N.C., S. Havelund, and P. Hougaard, *Ionization behavior of native and mutant insulins: pK perturbation of B13-Glu in aggregated species.* Archives of biochemistry and biophysics, 1990. **283**(2): p. 496-502.
63. Pekar, A.H. and B.H. Frank, *Conformation of proinsulin. Comparison of insulin and proinsulin self-association at neutral pH.* Biochemistry, 1972. **11**(22): p. 4013-4016.
64. Goldman, J. and F.H. Carpenter, *Zinc binding, circular dichroism, and equilibrium sedimentation studies on insulin (bovine) and several of its derivatives.* Biochemistry, 1974. **13**(22): p. 4566-4574.
65. Derewenda, U., Z. Derewenda, E. Dodson, G. Dodson, C. Reynolds, G. Smith, C. Sparks, and D. Swenson, *Phenol stabilizes more helix in a new symmetrical zinc insulin hexamer.* Nature, 1989. **338**(6216): p. 594-596.
66. Blader, M.L. and M.F. Dunn, *Insulin hexamers: new conformations and applications.* Trends in Biochemical Sciences, 1991. **16**: p. 341-345.
67. Drejer, K., *The bioactivity of insulin analogues from in vitro receptor binding to in vivo glucose uptake.* Diabetes/metabolism reviews, 1992. **8**(3): p. 259-285.
68. Kroef, E.P., R.A. Owens, E.L. Campbell, R.D. Johnson, and H.I. Marks, *Production scale purification of biosynthetic human insulin by reversed-phase high-performance liquid chromatography.* Journal of Chromatography A, 1989. **461**: p. 45-61.
69. Chance, R.E., E.P. Kroeff, J.A. Hoffmann, and B.H. Frank, *Chemical, physical, and biologic properties of biosynthetic human insulin.* Diabetes care, 1981. **4**(2): p. 147-154.
70. Chance, R.E. and B.H. Frank, *Research, development, production, and safety of biosynthetic human insulin.* Diabetes care, 1993. **16**(Supplement 3): p. 133-142.
71. Heding, L.G., *Determination of total serum insulin (IRI) in insulin-treated diabetic patients.* Diabetologia, 1972. **8**(4): p. 260-266.

72. Snel, L., U. Damgaard, and I. Mollerup, *HPLC quantification of rDNA polypeptides like insulin precursors produced in yeast.* Chromatographia, 1987. **24**(1): p. 329-332.
73. Kohanski, R. and M.D. Lane, *Homogeneous functional insulin receptor from 3T3-L1 adipocytes. Purification using N alpha B1-(biotinyl-epsilon-aminocaproyl) insulin and avidin-sepharose.* Journal of Biological Chemistry, 1985. **260**(8): p. 5014-5025.
74. Heinemann, L. and B. Richter, *Clinical pharmacology of human insulin.* Diabetes care, 1993. **16**(Supplement 3): p. 90-100.
75. Joshi, M. and J. Deshpande, *Polymerase chain reaction: methods, principles and application.* International Journal of Biomedical Research, 2010. **2**(1): p. 81-97.
76. Mullis, K. and F. Faloona, *Methods Enzymol. 155,335-350.* Mullis335155Methods Enzymol, 1987.
77. Erlich, H.A., *PCR technology*. 1989: Springer.
78. Pohl, G. and I.-M. Shih, *Principle and applications of digital PCR.* Expert review of molecular diagnostics, 2004. **4**(1): p. 41-47.
79. Bartlett, J.M. and D. Stirling, *A short history of the polymerase chain reaction*, in *PCR protocols*. 2003, Springer. p. 3-6.
80. Yu, M., Y. Cao, and Y. Ji. *The principle and application of new PCR Technologies*. in *IOP Conference Series: Earth and Environmental Science*. 2017. IOP Publishing.
81. Innis, M.A., D.H. Gelfand, J.J. Sninsky, and T.J. White, *PCR protocols: a guide to methods and applications*. 2012: Academic press.
82. Goeddel, D.V., D.G. Kleid, F. Bolivar, H.L. Heyneker, D.G. Yansura, R. Crea, T. Hirose, A. Kraszewski, K. Itakura, and A.D. Riggs, *Expression in Escherichia coli of chemically synthesized genes for human insulin.* Proceedings of the National Academy of Sciences, 1979. **76**(1): p. 106-110.
83. Chance, R.E., N.B. Glazer, and K.L. Wishner, *Insulin lispro (humalog)*, in *Biopharmaceuticals, an Industrial Perspective*. 1999, Springer. p. 149-171.
84. Mollerup, I., S.W. Jensen, P. Larsen, O. Schou, L. Snel, and M.C. Flickinger, *Insulin Purification*, in *Encyclopedia of Industrial Biotechnology*. 2009, John Wiley & Sons, Inc.
85. Overton, T.W., *Recombinant protein production in bacterial hosts.* Drug discovery today, 2014. **19**(5): p. 590-601.
86. Huang, C.-J., H. Lin, and X. Yang, *Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements.* Journal of industrial microbiology & biotechnology, 2012. **39**(3): p. 383-399.
87. Baeshen, M.N., A.M. Al-Hejin, R.S. Bora, M. Ahmed, H. Ramadan, K.S. Saini, N.A. Baeshen, and E.M. Redwan, *Production of biopharmaceuticals in E. coli: current scenario and future perspectives.* J Microbiol Biotechnol, 2015. **25**(7): p. 953-962.

88. Swartz, J.R., *Advances in Escherichia coli production of therapeutic proteins.* Current Opinion in Biotechnology, 2001. **12**(2): p. 195-201.
89. Al-Hejin, A.M., R.S. Bora, and M.M.M. Ahmed, *Plasmids for Optimizing Expression of Recombinant Proteins in E. coli*, in *Plasmid*. 2019, IntechOpen.
90. Panda, A.K., *Bioprocessing of therapeutic proteins from the inclusion bodies of Escherichia coli*, in *Biotechnology in India II*. 2003, Springer. p. 43-93.
91. Kamionka, M., *Engineering of therapeutic proteins production in Escherichia coli.* Current pharmaceutical biotechnology, 2011. **12**(2): p. 268-274.
92. Smith, S.M., K.L. Walker, A.S. Jones, C.J. Smith, and C. Robinson, *Characterization of a novel method for the production of single-span membrane proteins in Escherichia coli.* Biotechnology and Bioengineering, 2019. **116**(4): p. 722-733.
93. Selas Castiñeiras, T., S.G. Williams, A.G. Hitchcock, and D.C. Smith, *E. coli strain engineering for the production of advanced biopharmaceutical products.* FEMS microbiology letters, 2018. **365**(15): p. fny162.
94. Ovaa, H., *Unnatural amino acid incorporation in E. coli: current and future applications in the design of therapeutic proteins.* Frontiers in chemistry, 2014. **2**: p. 15.
95. Kane, J.F. and D.L. Hartley, *Formation of recombinant protein inclusion bodies in Escherichia coli.* Trends in Biotechnology, 1988. **6**(5): p. 95-101.
96. Norouzi, R., Z. Hojati, and Z. Badr, *Overview of the recombinant proteins purification by affinity tags and tags exploit systems.* Journal of Fundamental and Applied Sciences, 2016. **8**(3): p. 90-104.
97. Young, C.L., Z.T. Britton, and A.S. Robinson, *Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications.* Biotechnology journal, 2012. **7**(5): p. 620-634.
98. Arnau, J., C. Lauritzen, G.E. Petersen, and J. Pedersen, *Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins.* Protein Expression and Purification, 2006. **48**(1): p. 1-13.
99. Korf, U., T. Kohl, H. van der Zandt, R. Zahn, S. Schleeger, B. Ueberle, S. Wandschneider, S. Bechtel, M. Schnölzer, and H. Ottleben, *Large-scale protein expression for proteome research.* Proteomics, 2005. **5**(14): p. 3571-3580.
100. Terpe, K., *Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems.* Applied Microbiology and Biotechnology, 2003. **60**(5): p. 523-533.
101. Porath, J., J. Carlsson, I. Olsson, and G. Belfrage, *Metal chelate affinity chromatography, a new approach to protein fractionation.* Nature, 1975. **258**: p. 598-599.
102. Kimple, M.E., A.L. Brill, and R.L. Pasker, *Overview of affinity tags for protein purification.* Current protocols in protein science, 2013: p. 9.9. 1-9.9. 23.

103. Porath, J., *Immobilized metal ion affinity chromatography.* Protein Expression and Purification, 1992. **3**(4): p. 263-281.
104. Abdullah, N. and H. Chase, *Removal of poly-histidine fusion tags from recombinant proteins purified by expanded bed adsorption.* Biotechnology and Bioengineering, 2005. **92**(4): p. 501-513.
105. Gaberc-Porekar, V. and V. Menart, *Potential for using histidine tags in purification of proteins at large scale*, in *Chemical Engineering & Technology: Industrial Chemistry-Plant Equipment-Process Engineering-Biotechnology*. 2005. p. 1306-1314.
106. Frangioni, J.V. and B.G. Neel, *Solubilization and purification of enzymatically active glutathione S-transferase (pGEX) fusion proteins.* Analytical biochemistry, 1993. **210**(1): p. 179-187.
107. Smith, D.B. and K.S. Johnson, *Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase.* Gene, 1988. **67**(1): p. 31-40.
108. Coskun, O., *Separation techniques: chromatography.* Northern clinics of Istanbul, 2016. **3**(2): p. 156.
109. Reuhs, B.L., *High-performance liquid chromatography*, in *Food analysis*. 2017, Springer. p. 213-226.
110. Lu, X., Z. Zheng, S. Miao, H. Li, Z. Guo, Y. Zhang, Y. Zheng, B. Zheng, and J. Xiao, *Separation of Oligosaccharides from Lotus Seeds via Medium-pressure Liquid Chromatography Coupled with ELSD and DAD.* Scientific reports, 2017. **7**: p. 44174.
111. Berger, T.A., *Supercritical fluid chromatography*. Primer, publication number 5991-5509EN. 2015, USA: Agilent technologies. 186.
112. Miller, L., *Evaluation of non-traditional modifiers for analytical and preparative enantioseparations using supercritical fluid chromatography.* Journal of Chromatography A, 2012. **1256**: p. 261-266.
113. Miller, L., *Use of dichloromethane for preparative supercritical fluid chromatographic enantioseparations.* Journal of Chromatography A, 2014. **1363**: p. 323-330.
114. Noireau, A., E. Lemasson, F. Mauge, A.-M. Petit, S. Bertin, P. Hennig, É. Lesellier, and C. West, *Purification of drug degradation products supported by analytical and preparative supercritical fluid chromatography.* Journal of pharmaceutical and biomedical analysis, 2019. **170**: p. 40-47.
115. Nováková, L., A.G.-G. Perrenoud, I. Francois, C. West, E. Lesellier, and D. Guillarme, *Modern analytical supercritical fluid chromatography using columns packed with sub-2 μm particles: a tutorial.* Analytica chimica acta, 2014. **824**: p. 18-35.
116. Tarafder, A., *Theories for Preparative SFC*, in *Supercritical Fluid Chromatography*. 2017, Elsevier. p. 245-274.
117. Speybrouck, D. and E. Lipka, *Preparative supercritical fluid chromatography: A powerful tool for chiral separations.* Journal of Chromatography A, 2016. **1467**: p. 33-55.

118. Waters. *Beginner's Guide to Preparative SFC*. 2019; Available from: https://www.waters.com/waters/en_GB/Beginner%27s-Guide-to-Preparative-SFC/nav.htm?locale=en_GB&cid=134932615.
119. Płotka, J.M., M. Biziuk, C. Morrison, and J. Namieśnik, *Pharmaceutical and forensic drug applications of chiral supercritical fluid chromatography*. TrAC Trends in Analytical Chemistry, 2014. **56**: p. 74-89.
120. Nogle, L.M., C.W. Mann, W.L. Watts Jr, and Y. Zhang, *Preparative separation and identification of derivatized β-methylphenylalanine enantiomers by chiral SFC, HPLC and NMR for development of new peptide ligand mimetics in drug discovery.* Journal of pharmaceutical and biomedical analysis, 2006. **40**(4): p. 901-909.
121. Miller, L. and M. Potter, *Preparative chromatographic resolution of racemates using HPLC and SFC in a pharmaceutical discovery environment.* Journal of Chromatography B, 2008. **875**(1): p. 230-236.
122. Albericio, F. and H.G. Kruger, *Therapeutic peptides.* Future medicinal chemistry, 2012. **4**(12): p. 1527-1531.
123. Ayoub, M. and D. Scheidegger, *Peptide drugs, overcoming the challenges, a growing business.* Chimica oggi, 2006. **24**(4): p. 46.
124. Da Costa, J.P., M. Cova, R. Ferreira, and R. Vitorino, *Antimicrobial peptides: an alternative for innovative medicines?* Applied Microbiology and Biotechnology, 2015. **99**(5): p. 2023-2040.
125. Eckert, R., *Road to clinical efficacy: challenges and novel strategies for antimicrobial peptide development.* Future microbiology, 2011. **6**(6): p. 635-651.
126. Sun, L., *Peptide-based drug development.* Mod Chem Appl, 2013. **1**(1): p. 1-2.
127. Theel, E.S., *Matrix-assisted laser desorption ionization-time of flight mass spectrometry for the identification of bacterial and fungal isolates.* Clinical Microbiology Newsletter, 2013. **35**(19): p. 155-161.
128. Oviano, M. and G. Bou, *Matrix-assisted laser desorption ionization–time of flight mass spectrometry for the rapid detection of antimicrobial resistance mechanisms and beyond.* Clinical microbiology reviews, 2018. **32**(1): p. e00037-18.
129. Holland, R., J. Wilkes, F. Rafii, J. Sutherland, C. Persons, K. Voorhees, and J. Lay Jr, *Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry.* Rapid Communications in Mass Spectrometry, 1996. **10**(10): p. 1227-1232.
130. Seng, P., M. Drancourt, F. Gouriet, B. La Scola, P.-E. Fournier, J.M. Rolain, and D. Raoult, *Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry.* Clinical Infectious Diseases, 2009. **49**(4): p. 543-551.

131. Demirev, P.A., Y.-P. Ho, V. Ryzhov, and C. Fenselau, *Microorganism identification by mass spectrometry and protein database searches.* Analytical chemistry, 1999. **71**(14): p. 2732-2738.

132. Claydon, M.A., S.N. Davey, V. Edwards-Jones, and D.B. Gordon, *The rapid identification of intact microorganisms using mass spectrometry.* Nature biotechnology, 1996. **14**(11): p. 1584.

133. Edwards-Jones, V., M.A. Claydon, D.J. Evason, J. Walker, A. Fox, and D. Gordon, *Rapid discrimination between methicillin-sensitive and methicillin-resistant Staphylococcus aureus by intact cell mass spectrometry.* Journal of Medical Microbiology, 2000. **49**(3): p. 295-300.

134. Steensels, D., J. Verhaegen, and K. Lagrou, *Matrix-assisted laser desorption ionization-time of flight mass spectrometry for the identification of bacteria and yeasts in a clinical microbiological laboratory: a review.* Acta Clinica Belgica, 2011. **66**(4): p. 267-273.

135. De Respinis, S., V. Monnin, V. Girard, M. Welker, M. Arsac, B. Cellière, G. Durand, P.P. Bosshard, C. Farina, and M. Passera, *Matrix-assisted laser desorption ionization–time of flight (MALDI-TOF) mass spectrometry using the Vitek MS system for rapid and accurate identification of dermatophytes on solid cultures.* Journal of clinical microbiology, 2014. **52**(12): p. 4286-4292.

136. Patel, R., *Matrix-assisted laser desorption ionization–time of flight mass spectrometry in clinical microbiology.* Clinical Infectious Diseases, 2013. **57**(4): p. 564-572.

137. Christensen, J.J., R. Dargis, M. Hammer, U.S. Justesen, X.C. Nielsen, M. Kemp, and D.M.-T.M.S. Group, *Matrix-assisted laser desorption ionization–time of flight mass spectrometry analysis of Gram-positive, catalase-negative cocci not belonging to the Streptococcus or Enterococcus genus and benefits of database extension.* Journal of clinical microbiology, 2012. **50**(5): p. 1787-1791.

138. Turker, S.D., W.B. Dunn, and J. Wilkie, *MALDI-MS of drugs: Profiling, imaging, and steps towards quantitative analysis.* Applied Spectroscopy Reviews, 2017. **52**(1): p. 73-99.

139. Allwood, D., R. Dreyfus, I. Perera, and P. Dyer, *Optical absorption of matrix compounds for laser-induced desorption and ionization (MALDI).* Applied surface science, 1997. **109**: p. 154-157.

140. Luedke, K. *MALDI-TOF MS For the Diagnosis of Infectious Diseases*. 2015; Available from: https://news.mayocliniclabs.com/2015/04/13/maldi-tof-ms-for-the-diagnosis-of-infectious-diseases-2/.

141. Ho, H.P., P. Rathod, M. Louis, C.K. Tada, S. Rahaman, K.J. Mark, J. Leng, D. Dana, S. Kumar, and M. Lichterfeld, *Studies on quantitative phosphopeptide analysis by matrix-assisted laser desorption/ionization mass spectrometry without label, chromatography or calibration curves.* Rapid Communications in Mass Spectrometry, 2014. **28**(24): p. 2681-2689.

142. Wu, Z., M. Khan, S. Mao, L. Lin, and J.-M. Lin, *Combination of nano-material enrichment and dead-end filtration for uniform and rapid sample preparation in matrix-assisted laser desorption/ionization mass spectrometry*. Talanta, 2018. **181**: p. 217-223.
143. Parker, L., A. Engel-Hall, K. Drew, G. Steinhardt, D.L. Helseth Jr, D. Jabon, T. McMurry, D.S. Angulo, and S.J. Kron, *Investigating quantitation of phosphorylation using MALDI-TOF mass spectrometry*. Journal of mass spectrometry, 2008. **43**(4): p. 518-527.
144. Paizs, B. and S. Suhai, *Fragmentation pathways of protonated peptides.* Mass spectrometry reviews, 2005. **24**(4): p. 508-548.
145. Karas, M. and F. Hillenkamp, *Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons.* Analytical chemistry, 1988. **60**(20): p. 2299-2301.
146. Fenn, J.B., M. Mann, C.K. Meng, S.F. Wong, and C.M. Whitehouse, *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.
147. Roepstorff, P. and J. Fohlman, *Proposal for a common nomenclature for sequence ions in mass spectra of peptides.* Biomedical mass spectrometry, 1984. **11**(11): p. 601-601.
148. Mann, M., P. Højrup, and P. Roepstorff, *Use of mass spectrometric molecular weight information to identify proteins in sequence databases.* Biological mass spectrometry, 1993. **22**(6): p. 338-345.
149. Yates, J.R., S. Speicher, P.R. Griffin, and T. Hunkapiller, *Peptide mass maps: a highly informative approach to protein identification.* Analytical biochemistry, 1993. **214**(2): p. 397-408.
150. Henzel, W.J., T.M. Billeci, J.T. Stults, S.C. Wong, C. Grimley, and C. Watanabe, *Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases.* Proceedings of the National Academy of Sciences, 1993. **90**(11): p. 5011-5015.
151. Keller, A., A.I. Nesvizhskii, E. Kolker, and R. Aebersold, *Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search.* Analytical chemistry, 2002. **74**(20): p. 5383-5392.
152. Fenyö, D., *Identifying the proteome: software tools.* Current Opinion in Biotechnology, 2000. **11**(4): p. 391-395.
153. Perkins, D.N., D.J. Pappin, D.M. Creasy, and J.S. Cottrell, *Probability-based protein identification by searching sequence databases using mass spectrometry data.* ELECTROPHORESIS: An International Journal, 1999. **20**(18): p. 3551-3567.
154. Eng, J.K., A.L. McCormack, and J.R. Yates, *An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database.* Journal of the American Society for Mass Spectrometry, 1994. **5**(11): p. 976-989.

155. Hunt, D.F., J.R. Yates, J. Shabanowitz, S. Winston, and C.R. Hauer, *Protein sequencing by tandem mass spectrometry*. Proceedings of the National Academy of Sciences, 1986. **83**(17): p. 6233-6237.
156. Biemann, K., *Contributions of mass spectrometry to peptide and protein structure.* Biomedical & environmental mass spectrometry, 1988. **16**(1-12): p. 99-111.
157. Papayannopoulos, I.A., *The interpretation of collision-induced dissociation tandem mass spectra of peptides.* Mass spectrometry reviews, 1995. **14**(1): p. 49-73.

# Chapter 2

## A novel and efficient biosynthesis approach of human insulin production in *Escherichia coli* (*E. coli*).

**A novel and more efficient biosynthesis approach for human insulin production in *Escherichia coli (E. coli)*.**

Kamini Govender[1], Tricia Naicker[1], Johnson Lin[2], Sooraj Baijnath[1], Anil Amichund Chuturgoon[3], Naeem Sheik Abdul[3], Taskeen Docrat[3], Hendrik Gerhardus Kruger[1*] and Thavendran Govender[4*]

[1]Catalysis and Peptide Research Unit, School of Health Sciences, University of KwaZulu-Natal, Durban, South Africa

[2]School of Life Sciences, University of KwaZulu-Natal, Durban, South Africa

[3]School of Laboratory Medicine and Medical Sciences, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa

[4] Department of Chemistry, University of Zululand, Private Bag X1001, KwaDlangezwa 3886, South Africa

*Corresponding authors: kruger@ukzn.ac.za, Orchid ID: 0000-0003-0606-2053 and

govendert@unizulu.ac.za

Catalysis and Peptide Research Unit (CPRU)

E-block, 6th Floor, Room E1-06-016

University of KwaZulu-Natal, Westville Campus, South Africa

Contact number: +27312601845

**Abstract**

Insulin has captured researchers' attention worldwide. There is a rapid global rise in the number of diabetic patients, which increases the demand for insulin. Current methods of insulin production are expensive and time-consuming. A PCR-based strategy was employed for the cloning and verification of human insulin. The human insulin protein was then overexpressed in *E. coli* on a laboratory scale. Thereafter, the optimisation of human insulin expression was conducted. The yield of human insulin produced was approximately 520.92 (mg/L), located in the intracellular fraction. Human insulin was detected using MALDI-TOF-MS and LC-MS methods. The crude biosynthesised protein sequence was verified using protein sequencing, which had a 100% similarity to the human insulin sequence. The biological activity of human insulin was tested *in vitro* using a MTT assay, which revealed that the crude biosynthesised human insulin displayed a similar degree of efficacy to the standard human insulin. This study eliminated the use of affinity tags since an untagged pET21b expression vector was employed. Tedious protein renaturation, inclusion body recovery steps, and the expensive enzymatic cleavage of the C-peptide of insulin were eliminated, thereby making this method of biosynthesising human insulin a novel and more efficient method.

**Keywords**:
biosynthesis of human insulin; diabetes; *E. coli*.

# 1 Introduction

In 2018, it was estimated that 405.6 million people suffered from Type 2 diabetes, and this number is projected to increase to approximately 510.8 million by the year 2030. Based on these estimates, the global usage of insulin is estimated to rise from 516.1 million vials (1000 IU) to 633.7 million vials in 2030 (Basu et al., 2019). The increase in diabetic patients, coupled with the development of oral and inhalation methods of insulin delivery, requires considerable amounts of insulin. The current producers of insulin would not be able to cope with the rapid demand of affordable insulin as a result of high production costs and production capacity limitations (Baeshen et al., 2014).

Insulin has an essential role in glucose homeostasis (Ahmad 2004). It is produced by the beta cells of the pancreas and is one of the main anabolic hormones in the human body (Voet and Voet, 2011). It is synthesised by beta cells in the pancreas and has a fundamental role in fat and carbohydrate metabolism (Ahmad, 2004). Human insulin is a 51 amino acid (aa) polypeptide. It contains two polypeptide chains. The A chain has 21 amino acids and the B chain contains 30 amino acids. Insulin has three disulphide bonds. Two of the disulphide bonds interlink the A and B chains, whereas the third one is an intra A chain bond (Ahmad, 2004; Vajo et al., 2001). Insulin was first discovered in 1921 by Charles Best and Frederick Banting (Banting & Best, 1922). Prior to 1982, insulin was extracted from the pancreas of animals such as bovines and porcine (Beals et al., 2013; Mollerup et al., 2009).

The biotechnological industry can, possibly, develop suitable innovations to curb the shortage in the supply of human insulin. Novel innovations in methods of insulin production and purification have the potential to significantly advance the pharmaceutical biotechnology industry. The use of recombinant deoxyribonucleic acid (DNA) technology allowed for a direct method of biosynthesising human insulin, which did not require animal-derived pancreatic tissue (Walsh, 2005). In 1982, Novo Nordisk derived semi-synthetic human insulin, whereby porcine insulin was converted into human insulin enzymatically. Human insulin was biosynthesised by Eli Lilly using recombinant DNA technology in 1982. Two production methods were reported by Eli Lilly (Mollerup et al., 2009). The first method consisted of cloning the A and B chains of insulin separately in *E. coli*; thereafter, these two chains were isolated, purified, and chemically attached. Subsequently, the final reversed phase-high performance liquid chromatography (RP-HPLC) purification steps were conducted (Chance et al., 1981; Frank & Chance, 1983; Mollerup et al., 2009). The second method of cloning recombinant insulin in *E. coli* consisted of the expression of proinsulin under a tryptophan promoter that had a methionine bond to proinsulin. This study cloned the A and B chains of human insulin together in *E. coli*. In 1988 Novo Nordisk launched the biosynthesis of human insulin using recombinant DNA

technology (Mollerup et al., 2009). The proinsulin biosynthesis strategy of human insulin is the preferred method.

However, this procedure results in the formation of inclusion bodies and requires the use of enzymes such as carboxypeptidase B to remove the C-peptide (Redwan et al., 2007; Zieliński et al., 2019) and the use of affinity tags (Redwan et al., 2007). The inclusion body formation requires additional steps to be taken, such as the renaturation of the protein (Redwan et al., 2007; Zieliński et al., 2019). This study aimed to create a novel and efficient method for the biosynthesis of human insulin employing a polymerase chain reaction (PCR)-based strategy using the pET21b expression vector in *E. coli* for the optimisation of human insulin expression.

## 2 Materials and methods

### 2.1 Bacterial strains

The bacterial strains employed in this study were obtained from the School of Life Sciences, in the Discipline of Microbiology (University of KwaZulu-Natal, Westville Campus, South Africa). The pCMV6-XL5 plasmid integrated with the gene encoding for human proinsulin was procured from ORI Gene (United States of America (USA). Commercial human insulin protein was obtained from Sigma-Aldrich Inc (Germany) as a control in this study. All restriction enzymes were purchased from Thermo-Scientific, USA.

### 2.2 Cloning of human insulin in *E. coli*

An amplification of human insulin was conducted using a PCR master mix according to the manufacturer's instructions (Thermo-Scientific, USA). The pCMV6-XL5 plasmid was used as a DNA template with the following primers: (PGEX-*Bam*HI-F: 5'-**GGA TCC** ATG GCC CTG TGG ATG CG–3' and PGEX-*Xho*I-R: 5'–**CTC GAG** CTA GTT GCA GTA GTT CTC C–3). The PCR conditions were as follows: a denaturation at 95°C for 30 seconds, followed by an annealing step at 61.2°C for 30 seconds; thereafter an elongation step at 72°C for one minute, and lastly another elongation at 72°C for two minutes. The PCR comprised 30 cycles. The PCR amplicons were subsequently visualised on a 1.5% agarose gel (Helling et al., 1974) and viewed using a Syngene G BOX gel documentation system (Vacutec, South Africa). The respective band was cut off using a scalpel under a UV trans-illuminator (UVP incorporated, USA). PCR products were then purified using the Zymoclean™ gel DNA recovery kit (ZYMO Research, USA), following the manufacturer's instructions.

The pET21b vector DNA was isolated using a Zyppy™kit (ZYMO Research, USA), as per the manufacturer's instructions. Thereafter, the PCR amplicons and the pET21b vector were digested with *Bam*HI and *Xho*I, according to the manufacturer's instructions (Thermo-Scientific, USA). Thereafter, the restricted PCR products were inserted into the pET21b vector using T4 DNA ligase (Thermo-Scientific, USA), according to the manufacturer's instructions. The pET21b vector human proinsulin (designated as pET21b-hPin) was transformed in chemically competent *E. coli* BL21 (DE3) cells plasmid using the calcium chloride heat shock method (Inoue et al., 1990).

### 2.3 The expression and isolation of the protein

The transformed *E. coli* BL21 (DE3) (Thermo-Scientific, USA) was verified using colony PCR. The positive clones were incubated in 10 mL of Luria-Bertani (LB) medium (1% bacto tryptone, 0.5% yeast extract, 1% NaCl) containing 50 µg/mL ampicillin and 34 µg/mL chloramphenicol (Merck, Germany) overnight at 37°C at 180 revs per minute (rpm). One mL of an overnight culture was used to inoculate 100 mL of LB broth using the above method (Maseko et al., 2016; Volontè et al., 2011). The expression of human proinsulin was induced by the addition of 0.1 to 1 mM Isopropyl β-D thiogalactoside (IPTG) at an early exponential phase ($OD_{600}$ 0.4-0.7).

The over-expressed protein was recovered following to a similar method adapted from Maseko et al. (2016). After overnight induction at 16°C, the cultures were transferred to 50 mL Cell Star® centrifuge tubes (Greiner Bio-One, Austria) and spun down at 8000 rpm for 15 minutes at 4°C. The pellet was resuspended in 50 mM Tris-HCl (pH = 8). The crude protein extracts were sonicated for five to ten minutes on ice until the cells were homogenised with a Sonic Ruptor 400 Ultrasonic homogeniser (Omni International, United States of America). Thereafter, the samples were centrifuged at 8000 rpm for approximately 15 minutes at 4°C, and the supernatant was transferred to a clean centrifuge tube and stored at 4°C. The proinsulin was converted to mature human insulin as a result of the autocatalytic cleavage of the C-peptide (Gooch, 2011). Thereafter, the crude protein was concentrated using 3 kDa (kilo-Dalton) Amicon®ultra-size exclusion centrifugal filters (Merck, Germany), according to the manufacturer's instructions.

### 2.4 Detection of human insulin by matrix-assisted laser desorption ionisation-time of flight-mass spectrometry (MALDI-TOF-MS)

The auto flex III smart beam MALDI-TOF-MS (Bruker Daltonics, Germany) was used for the detection of the standard human insulin and biosynthesised human insulin samples. The MALDI-TOF spectra were obtained from linear mode, where by ions charged at a voltage of 20 kilovolts (kV) within a molecular mass detection range of 1500 to 7000 Daltons (Da) were employed.

The instrument contained an ultraviolet nitrogen laser at 337 nanometres (nm). FlexControl version 3.4 build 119 software was used for the data acquisition and FlexAnalysis of the MALDI-TOF spectra. The target that was employed in the study was a ground steel target plate (Bruker Daltonics, Germany). The protein samples were diluted and spotted on the ground steel plate according to the manufacturer's instructions (Bruker Daltonics, Germany). The matrix utilised in this study was alpha-cyano-4-hydroxycinnamic acid (CHCA or HCCA) (Bruker Daltonics, Germany). A standard curve was generated whereby standard human insulin was diluted to the required concentration using the HCCA matrix. The standard curve comprised of 0 ng/mL, 10 ng/mL, 50 ng/mL, 100 ng/mL, and 1000 ng/mL concentrations (Refer to Figure S4 in the supplementary section.).

### 2.5 Liquid chromatography-mass spectrometry (LC-MS)

The sample containing human insulin was further confirmed using the LC-MS-2020 (Shimadzu, Japan) with a YMC-Triart C18 column (150 mm x 4.6 mm internal diameter, the pore size of 120 Å and a particle size of 5 µm) (YMC, Japan), coupled to a NM32LA nitrogen generator (Peak Scientific Instruments, United Kingdom). The MS spectra were obtained using the positive mode with a mass range of 200 to 1500 m/z. Mobile phase A was Millipore water (Millipore, USA) and mobile phase B was acetonitrile (Merck, Germany). Both these mobile phases contained 0.1% (v/v) formic acid as an ion-pairing agent. Flow rates of 1.0 mL/minute were used with a total run time of 25 minutes. The gradient profile was from 5% to 95% acetonitrile in seven minutes; and at 95% acetonitrile for a further 18 minutes. The column temperature was maintained at 40°C. The nebuliser was set at 1.5 bar with a desolvation gas temperature of 250°C and dry gas flow rates of 10 L/minute.

### 2.6 Protein sequencing of human insulin

In-solution digestion of the protein samples were conducted at the Central Analytical Facility (Stellenbosch University, South Africa) on an Ultimate 3000 RSLC (Thermo Fisher Scientific, USA), coupling with a mass spectrometry (fusion mass spectrometer) (Thermo Fisher Scientific, USA) and an ionisation source (Nanospray Flex). The files generated from the mass spectrometer were imported onto the Proteome Discoverer v1.4. The procedure was conducted according to the manufacturer's instructions (Thermo Fisher Scientific, USA), and analysed using an Amanda algorithm and Sequest. Thereafter, a concatenated database interrogation was conducted using the Uniprot P09211, concatenated with the contaminant protein database (cRAP). The peptides were validated using Target-Decoy PSM validator mode. The results were imported according to the manufacturer's instructions using Scaffold 1.4.4 software (Proteome Software Inc, USA), and the identified peptides were validated with the peptide, Protein Prophet, as well as X!Tandem algorithms from the Scaffold 1.4.4 software (Proteome Software Inc, USA).

### 2.7 Biological activity

The biological activity of the standard human insulin and the crude biosynthesised human insulin was conducted *in vitro* using a 3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyl tetrazolium bromide (MTT) assay. Hepatocellular carcinoma (HepG2) (Highveld Biologicals, South Africa) cell line was used. The HepG2 cells were cultured according to a method similar to that reported by Abdul *et al*. (2016). The cell culture reagents were obtained from Whitehead Scientific (South Africa). The standard human insulin and crude biosynthesised human insulin stocks were made in Eagle's minimum essential medium (EMEM) (catalogue number: 12-136F, Lonza, Switzerland) up to a concentration of 1 mg/mL. In the MTT assay, three treatment concentrations were used of the standard human insulin and crude biosynthesised human insulin: low (50 µg/mL), medium (100 µg/mL), and high (150 µg/mL).

The HepG2 cells were seeded into a 96-well plate ($2x10^4$ cells/well, 24 hours). Thereafter, the culture medium was removed, and cells were washed with PBS. The HepG2 cells were then treated with standard human insulin and crude biosynthesised human insulin hyperglycaemic (25 mM) media (Chen et al. 2006) for 15 minutes. The treatments were removed by washing the cells in PBS, and 120 µl MTT (5 mg/ml, PBS) was added to each well. This was followed by an incubation of four hours at 37°C. The formazan crystals were solubilised by the addition of 100 µl/well DMSO. The optical densities (OD) were measured spectrophotometrically (BioTek uQuant, USA) at 570 nm with a reference wavelength of 690 nm. All experiments were conducted in triplicate.

The following formula was used to calculate cell viability (%):

$$\%\ Cell\ viability = \frac{OD\ of\ treated\ cells}{OD\ of\ control\ cells} \times 100$$

### 2.8 Statistical analysis

Two-way analysis of variance (ANOVA) was conducted. Analyses were performed using the statistical software package Graph pad in stat version 8.1.0 (325) 64 bit for Windows (Graph pad software, San Diego California).

# 3 Results

## 3.1 PCR amplification and purification of human proinsulin replicons

PCR amplification of human proinsulin was conducted using PCR primers PGEX-*Bam*HI-F and PGEX-*Xho*I-R. The mass amplified and purified product corresponded to the expected amplicon size of 345 bp (Figure S1a, lane 2). The human proinsulin replicons were inserted into pET21b. The pET21b-hPin plasmid (Figure 1) was then transformed successfully into *E. coli* BL21 (DE3) cells, which were confirmed by colony PCR (345 bp; Figure S1b).

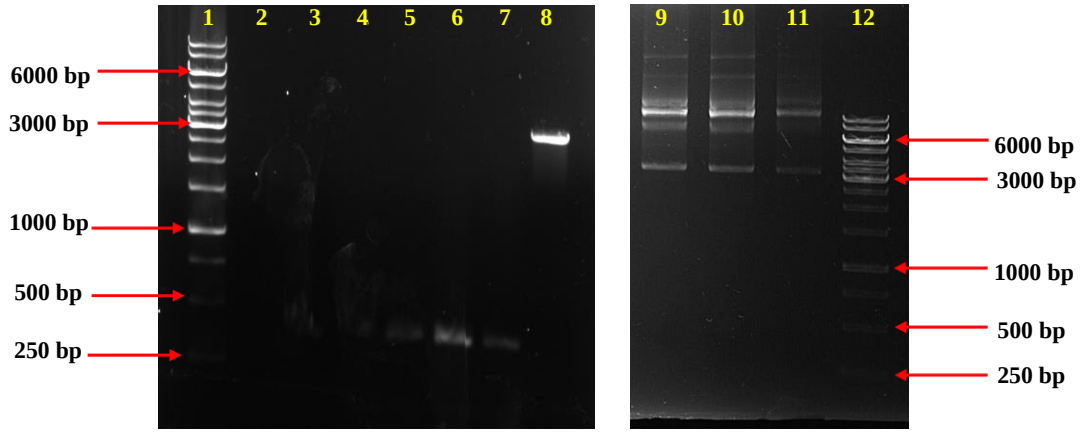


**Figure 1:** Images displaying the PCR amplicons of proinsulin DNA, purified vector DNA, and the integrated pET21b-hPin vector with the inserted proinsulin gene. Lanes 1 and 12 contain one kb molecular weight marker. Lanes 2 to 7 contain the PCR amplicons of proinsulin flanked with *Bam*HI and *Xho*I restricted ends; Lane 8 contains the purified pET21b miniprep product; Lanes 9 to 11 contain the integrated pET21b-hPin vector miniprep products.

## 3.2 Detection and optimisation of expression of human insulin using MALDI-TOF-MS

The expression of human insulin was conducted by expressing the protein at varying IPTG concentrations, such as 0.1 mM, 0.5 mM, and 1 mM. The standard human insulin and the crude biosynthesised human insulin were detected using MALDI-TOF-MS, as the expected size of human insulin was obtained (approximately 5.8 kDa). Figure 2 shows an example of the MALDI-TOF spectrum, illustrating the supernatant sample of crude biosynthesised human insulin, which was induced at 1 mM IPTG. The results were confirmed using commercial human insulin as the sample using MALDI-TOF (Figure S2) and LC-MS (Figure S3). The LC-MS spectra detected the $[M+4H]^{4+}$, $[M+5H]^{5+}$, $[M+6H]^{6+,}$ and $M+7H]^{7+}$charged states of the human insulin protein in all IPTG induced samples (Figure S3).

Human insulin was detected in all the intracellular and inclusion body fractions under all IPTG induction conditions. Table 1 shows that 0.1 mM IPTG induced the optimum yield for recovering human insulin using the *E. coli* BL21 DE3-pET21b-hPin expression system, according to the standard curve using the commercially available standard human insulin (Figure S4). Most of the induced human insulin was in the intracellular (soluble) fraction, with 520.92 mg/L of human insulin received in the intracellular fraction for 0.1 mM IPTG induction; compared to 393.81 mg/L and 210.83 mg/L under 0.5 mM and 1.0 mM IPTG, respectively.

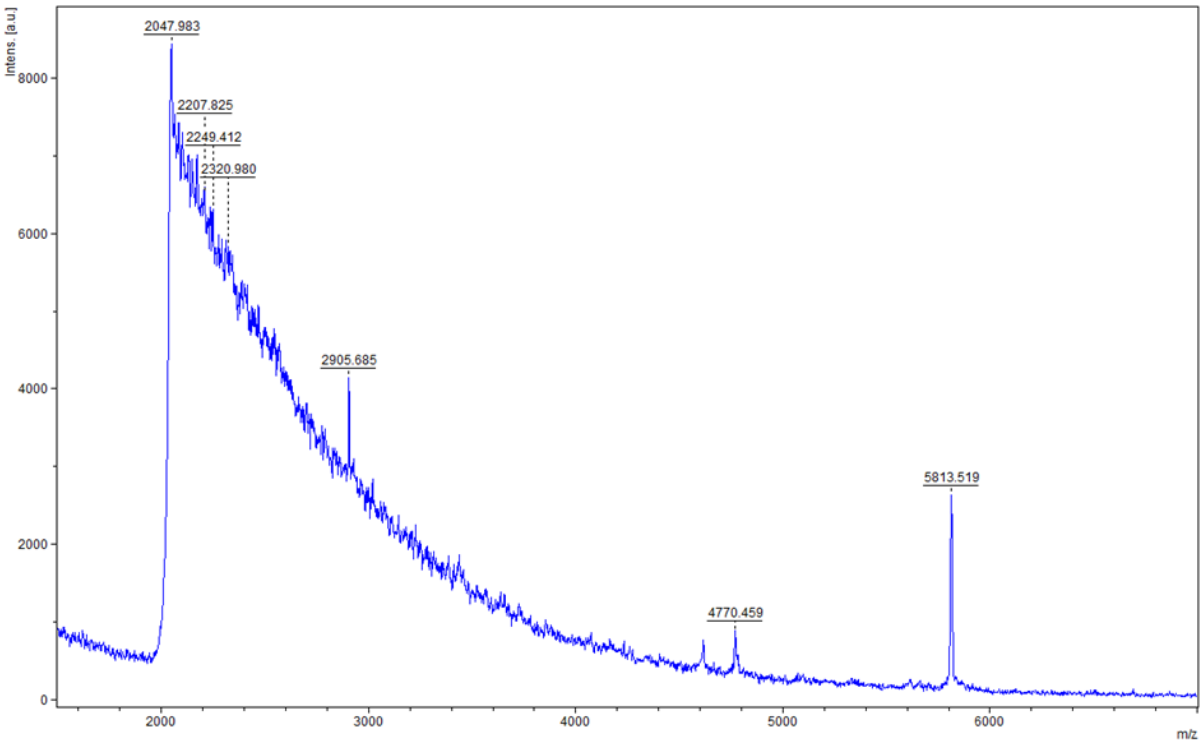


**Figure 2:** A MALDI-TOF spectrum illustrating the supernatant sample of crude biosynthesised human insulin, which was induced at 1 mM IPTG.

**Table 1**: The concentrations of human insulin obtained for the IPTG induction calculated based on commercial human insulin.

| IPTG induced sample | Concentration (mg/L) (Intracellular) | Concentration (mg/L) (Inclusion body) |
|---|---|---|
| 1 mM | 210.83 | 65.45 |
| 0.5 mM | 393.81 | 64.20 |
| 0.1 mM | 520.92 | 107.82 |

### 3.3 Protein sequencing of induced biosynthesised human insulin

The crude biosynthesised human insulin protein was sequenced. The standard human insulin was set as a positive control. The protein sequencing results confirmed that the biosynthesised protein was human insulin (Figure 3; Table S1). After a concatenated database search, the fragmented biosynthesised human insulin peptide sequence was found to have a 100% similarity match to that of the human insulin sequence from the protein database.

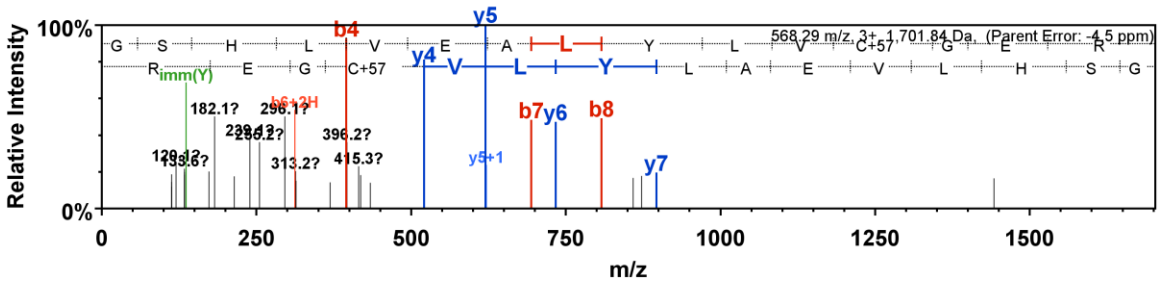


**Figure 3:** A peptide spectrum illustrating the protein sequence of crude biosynthesised human insulin, which was a 100% match to the human insulin sequence derived from the Scaffold 1.4.4 software.

### 3.4 Biological activity of human insulin

The biological activity was tested using a MTT assay *in vitro*, using the HepG2 cell line under hyperglycaemic conditions (Figure 4). In the MTT assay, three concentrations were used of the standard human insulin and crude biosynthesised human insulin: low (50 µg/mL), medium (100 µg/mL) and high (150 µg/mL). The results revealed that, for the low treatment the crude biosynthesised human insulin displayed higher cell viability than the standard human insulin; whereas the medium and high concentrations exhibited similar cell viability to that of the standard human insulin. The results were statistically significant with a calculated probability (p-value) of < 0,0001.

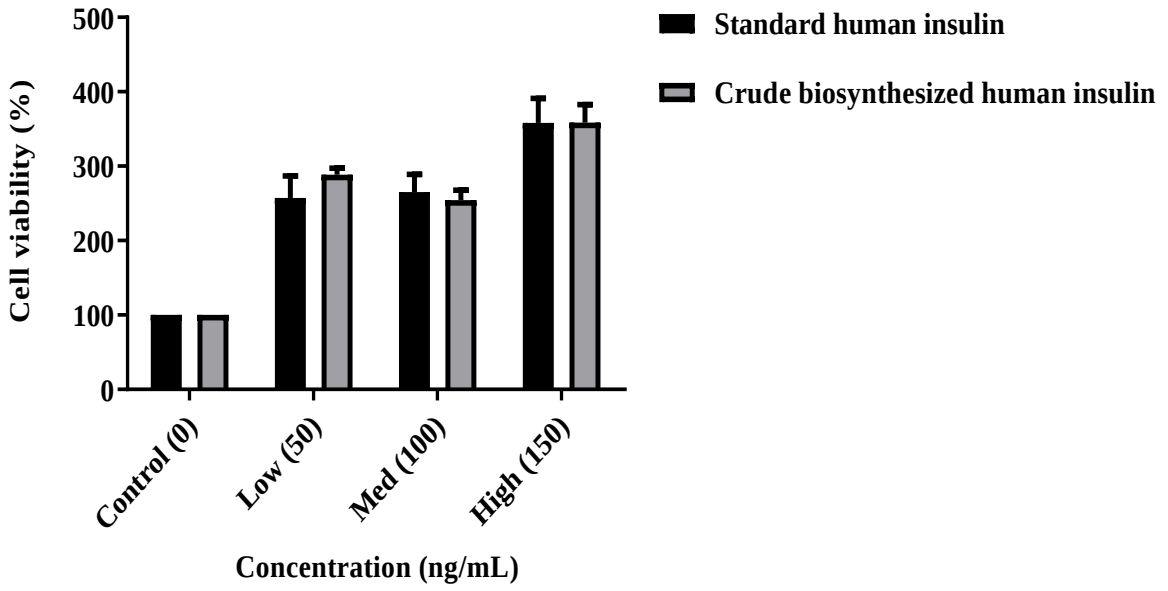


**Figure 4:** A graph illustrating the cell viability of HepG2 cells, under hyperglycaemic conditions for standard and crude biosynthesised human insulin at low (50 ng/mL), medium (100 ng/mL) and high (150 ng/mL) MTT concentrations (p-value < 0,0001).

## 4 Discussion and conclusion

This study developed a novel and efficient biosynthetic laboratory-scale method of human insulin production using an innovative recombinant cloning strategy, pET21b-hPin vector; and an *E. coli* BL21 (DE3) expression system was employed using proinsulin (Figure 2). Insulin can be biosynthesised using two methods. The first method is when the A and B chains of insulin are cloned separately, isolated, and subsequently purified as S-sulfonate derivatives. The chains are thereafter combined to form insulin, which is then purified (Kroef et al., 1989). The second approach involves the use of proinsulin, which is the preferred method of biosynthesising insulin, since a single fermentation is conducted, as opposed to in the first method of insulin production (Kroef et al., 1989). In terms of costs associated with affinity tag removal on a large scale, the current project was based on the biosynthesis and analysis of the untagged form of human insulin (Norouzi et al., 2016; Young et al., 2012).

This study used a novel PCR-based strategy in the cloning of human insulin. The human insulin gene was amplified and confirmed using PCR (Figure S1). Subsequently, the pET21b-hPin plasmid was transformed into an expression host BL21 (DE3). The production of biosynthesised human insulin was validated using MALDI-TOF-MS, LC-MS, and protein sequencing (Figure 3). Commercial human insulin was used as a positive control.

Mention is made in the literature of the human insulin protein being expressed at 1 mM IPTG (Redwan et al., 2007). In this study, the human insulin protein was initially induced with 1 mM IPTG at 37°C. Thereafter, the optimisation of the expression of human insulin was conducted, where the IPTG concentration was varied. The samples were detected using MALDI-TOF-MS (Figure 2) and LC-MS (Figure S3). A standard curve derived from various sample concentrations of standard human insulin was generated using MALDI-TOF-MS to extrapolate the respective protein concentrations from the biosynthesised human insulin supernatant and pellet fractions. The concentrations of the protein found in the intracellular fractions (soluble fractions) were 520.92 mg/L, 393.81 mg/L and 210.83 mg/L; and in the pellets (inclusion body/insoluble fractions) they were 107.82 mg/L, 64.20 mg/L and 65.45 mg/L for the respective IPTG concentrations of 0.1 mM, 0.5 mM and 1 mM (Table 1). The optimum condition for the induction of human insulin was 0.1 mM IPTG.

This study optimised human insulin protein expression, where the human insulin protein was expressed in the soluble (intracellular) fraction resulting in the elimination of the use of chaotropic agents. The use of chaotropic agents, such as urea and guanidine hydrochloride, can result in complete secondary structure denaturation (Singh et al., 2015). The use of guanidine hydrochloride and urea to solubilise protein results in a low yield of the bioactive therapeutic protein (Upadhyay et al., 2016). The proinsulin was produced in inclusion bodies at a 10% level in *E. coli* with a yield of 1.3 mg/L of insulin (Redwan et al., 2007). Solubilisation and refolding can often lead to poor recovery of the protein of the desired bioactive protein (Singh et al., 2015). This study had a yield of approximately 520.92 (mg/L) of human insulin located in the intracellular (soluble) fraction, which is higher than in previous reports (Redwan et al., 2007).

Researchers have found that protein expression conducted at a high inducer concentration and temperature resulted in the expression of the biosynthesised protein at an increased translational rate and can eventually result in the formation of inclusion bodies (Carrió and Villaverde, 2005). The formation of inclusion bodies in bacterial expression hosts such as *E. coli* presents challenges in the recovery process of bioactive proteins (Upadhyay et al., 2016). The extraction of recombinant biosynthesised proteins from inclusion bodies usually results in low yields of the bioactive protein, and this process is laborious (Singh et al., 2015). Inclusion bodies require downstream processing such as isolation from the cell, solubilisation, purification, and refolding of the protein (Redwan et al., 2007; Singh et al., 2015; Zieliński et al., 2019). In this study, the protein was expressed mainly in the soluble fraction; therefore, the inclusion body recovery and renaturation of the protein steps were eliminated, which decreased the expenses associated with these steps. Therefore, the method of human insulin biosynthesis employed in this study is a more efficient process than the current methods (Zieliński et al., 2019).

The proinsulin cloning strategy uses the nucleotide sequence encoding for human proinsulin, which consisted of A, B, and C chains (Chance et al., 1999). In the above strategy, proinsulin was cloned, and the C-peptide was cleaved using *Achromabacter lyticus* protease, carboxypeptidase B, and trypsin (Baeshen et al., 2014; Mollerup et al., 2009; Morihara et al., 1980). *Achromobacter lyticus* protease (Morihara et al. 1980) is a lysine-specific enzyme which can enzymatically convert proinsulin to insulin; it has an advantage over trypsin digestion as it avoids non-specific cleavage after the arginine at position B22 (Mollerup et al., 2009). The C-peptide bond was enzymatically cleaved by the utilisation of cyanogen bromide; this was followed by protein folding and the formation of disulphide bonds (Frank and Chance, 1983). The enzymes that are normally used in the C-peptide cleavage, such as carboxypeptidase B, are costly (Redwan et al., 2007). This ground-breaking study eliminated the expensive digestion of the C-peptide, which occurred as a result of autocatalytic cleavage (Gooch, 2011).

The human insulin that was biosynthesised in this study was detected and verified using MALDI-TOF-MS, LC-MS, and protein sequencing. Standard human insulin served as a positive control for the protein sequencing of human insulin with a 90% similarity to the human insulin sequence from the protein database (Figure S5). The crude biosynthesised human insulin protein produced in this study was verified using protein sequencing. After a concatenated peptide database search, the crude biosynthesised human insulin was unequivocally proven to be human insulin with a 100% similarity (Figure 3). Therefore, this study successfully biosynthesised and optimised the expression of human insulin in *E. coli*.

This study employed an *in vitro* MTT assay to determine the biological activity. The MTT assay is used to measure the viability of metabolically active cells based on the cells' ability to produce reducing equivalents (Carrió and Villaverde, 2005; Morihara et al., 1980). The cell uptake of MTT by a protein facilitated mechanism or by endocytosis leads to the reduction of MTT. This yields a purple formazan product, which is impermeable to cell membranes and results in its accumulation within living cells (Hansen and Bross, 2010; Maioli et al., 2009; Mosmann, 1983). The solubilisation of the cells results in the release of the purple product, which is detected colorimetrically (Maioli et al., 2009). The cells' respiratory chain (Slater et al., 1963) and other electron transport systems (Liu et al., 1997) cause the reduction of MTT. The reduction of MTT by living cells is an indication of mitochondrial activity, which therefore serves as a measure of cell viability (Maioli et al., 2009). Insulin is involved in the regulation of glucose uptake into muscle and fat cells (Drejer, 1992). Insulin works within seconds to activate the tyrosine kinase receptor, ion transport systems, and stimulation of glucose (Drejer, 1992). Therefore, the glucose utilisation by the HepG2 cells increases cellular respiration and mitochondrial activity, increasing cell viability. The biological activity was tested in this study using a MTT assay *in vitro,* using a HepG2 cell line with normoglycaemic (5.5 mM) and

hyperglycaemic (25 mM) conditions (Chen et al., 2006). The MTT assay exhibited increased cell viability in the hyperglycaemic treatments. There were three treatment concentrations of standard human insulin and crude biosynthesised human insulin: low (50 ng/mL), medium (100 ng/mL), and high (150 ng/mL). The analysis was conducted on standard human insulin and biosynthesised human insulin. The standard human insulin served as a positive control. The cell viability of the crude biosynthesised human insulin showed (288,42% ± 9,04%) low, (254,57% ± 12,96%) medium and (358,70% ± 24,17%) high concentrations, in comparison to standard human insulin which had (257,37% ± 29,65%) low, (265,23% ± 23,61%) medium and (358,07% ± 33,10%) high concentrations. In the low concentration, the crude biosynthesised human insulin displayed higher cell viability than the standard human insulin; whereas the medium and high concentrations exhibited similar cell viability to that of the standard human insulin. As such, the efficacy was determined to be the same as the standard human insulin. Therefore, we can conclude that the novel method employed in this study is an efficient and effective method of human insulin production.

The biological activity test conducted in this study revealed that the crude biosynthesised human insulin displays similar efficacy to that of the commercially available standard human insulin. This study successfully provided a novel method for the biosynthesis of human insulin since it employed the untagged pET21b expression vector in *E. coli*. The elimination of affinity tags has implications with regards to protein purification, as well as the reduction in downstream processing costs, such as glutathione S-transferase and histidine tag removal. In addition, this method also eliminated the tedious protein enzymatic cleavage of the C-peptide, the recovery of human insulin from inclusion bodies, and human insulin renaturation steps; thereby making this process less labour-intensive. This study was conducted on a laboratory-scale; however, it has the potential to be up-scaled industrially. Future studies can focus on finding greener and more affordable purification methods for human insulin.

## 5 Acknowledgements

This study was made possible through financial support from the University of KwaZulu-Natal, National Research Foundation (NRF) grant number: 105216, the Technology Innovation Agency (TIA) of South Africa.

## 6 Conflict of interest

The authors declare no conflict of interest.

## 7 Author's contribution

Kamini Govender biosynthesised human insulin using recombinant DNA technology, analysed all the experimental data and composed the entire manuscript, including the supporting information. Taskeen Docrat, Dr Naeem Sheik Abdul, and Prof. Anil Amichund Chuturgoon conducted the MTT assay. The rest of the authors are co-supervisors on the project and contributed to the conceptualisation of the idea, funding of the project, and scientific guidance.

## 8 References


Abdul NS, Nagiah S, Chuturgoon AA (2016) Fusaric acid induces mitochondrial stress in human hepatocellular carcinoma (HepG2) cells. Toxicon 119:336-344

Ahmad B (2004) Review: Pharmacology of insulin. Br J Diabetes Vasc Dis 4(1):10-14

Baeshen NA, Baeshen MN, Sheikh A, Bora RS, Ahmed MMM, Ramadan HA, Saini KS, Redwan EM (2014) Cell factories for insulin production. Microb Cell Fact 13(1):1

Banting F, Best C (1922) The Internal Secretion of the Pancreas. J Lab Clin Med 7(5):465-480

Basu S, Yudkin JS, Kehlenbrink S, Davies JI, Wild SH, Lipska KJ, Sussman JB, Beran D (2019) Estimation of global insulin use for type 2 diabetes, 2018–30: a microsimulation analysis. Lancet Diabetes Endocrinol 7(1):25-33

Beals JM, DeFelippis MR, Kovach PM, Jackson JA (2013) Insulin Pharmaceutical biotechnology. Springer, pp 255-275

Carrió MM, Villaverde A (2005) Localization of chaperones DnaK and GroEL in bacterial inclusion bodies. J bacteriol 187(10):3599-3601

Chance RE, Glazer NB, Wishner KL (1999) Insulin lispro (humalog) Biopharmaceuticals, an Industrial Perspective. Springer, pp 149-171

Chance RE, Kroeff EP, Hoffmann JA, Frank BH (1981) Chemical, physical, and biologic properties of biosynthetic human insulin. Diabetes care 4(2):147-154

Chen Q, Xia Y, Qiu Z (2006) Effect of ecdysterone on glucose metabolism *in vitro*. Publisher. Accessed 23 April 2019 2019

Drejer K (1992) The bioactivity of insulin analogues from *in vitro* receptor binding to *in vivo* glucose uptake. Diabetes/metabolism reviews 8(3):259-285

Frank B, Chance R (1983) Two routes for producing human insulin utilizing recombinant DNA technology. MMW, Munchener medizinische Wochenschrift:S14

Gooch JW (2011) Autocatalytic. Encyclopedic Dictionary of Polymers:876-876 doi:https://doi.org/10.1007/978-1-4419-6247-8_13197

Hansen J, Bross P (2010) A cellular viability assay to monitor drug toxicity Protein Misfolding and Cellular Stress in Disease and Aging. Springer, pp 303-311

Helling RB, Goodman HM, Boyer HW (1974) Analysis of endonuclease R· *Eco*RI fragments of DNA from lambdoid bacteriophages and other viruses by agarose-gel electrophoresis. J Virol 14(5):1235-1244

Inoue H, Nojima H, Okayama H (1990) High efficiency transformation of *Escherichia coli* with plasmids. Gene 96(1):23-28

Kroef EP, Owens RA, Campbell EL, Johnson RD, Marks HI (1989) Production scale purification of biosynthetic human insulin by reversed-phase high-performance liquid chromatography. J Chromatogr A 461:45-61

Liu Y, Peterson DA, Kimura H, Schubert D (1997) Mechanism of cellular 3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyltetrazolium bromide (MTT) reduction. J Neurochem 69(2):581-593

Maioli E, Torricelli C, Fortino V, Carlucci F, Tommassini V, Pacini A (2009) Critical appraisal of the MTT assay in the presence of rottlerin and uncouplers. Biological procedures online 11(1):227

Maseko SB, Natarajan S, Sharma V, Bhattacharyya N, Govender T, Sayed Y, Maguire GEM, Lin J, Kruger HG (2016) Purification and characterization of naturally occurring HIV-1 (South African subtype C) protease mutants from inclusion bodies. Protein Expr Purif 122:90-96 doi:https://doi.org/10.1016/j.pep.2016.02.013

Mollerup I, Jensen SW, Larsen P, Schou O, Snel L, Flickinger MC (2009) Insulin Purification Encyclopedia of Industrial Biotechnology. John Wiley & Sons, Inc.

Morihara K, Oka T, Tsuzuki H, Tochino Y, Kanaya T (1980) Achromobacter protease I-catalyzed conversion of porcine insulin into human insulin. Biochem Biophys Res Commun 92(2):396-402

Mosmann T (1983) Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays. J Immunol Methods 65(1-2):55-63

Norouzi R, Hojati Z, Badr Z (2016) Overview of the recombinant proteins purification by affinity tags and tags exploit systems. J Fundament Appl Sci 8(3):90-104

Redwan ERM, Matar SM, El-Aziz GA, Serour EA (2007) Synthesis of the human insulin gene: protein expression, scaling up and bioactivity. Prep Biochem Biotechnol 38(1):24-39

Singh A, Upadhyay V, Upadhyay AK, Singh SM, Panda AK (2015) Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. Microb Cell Fact 14(1):41

Slater T, Sawyer B, Sträuli U (1963) Studies on succinate-tetrazolium reductase systems: III. Points of coupling of four different tetrazolium salts III. Points of coupling of four different tetrazolium salts. Biochim Biophys Acta 77:383-393

Upadhyay V, Singh A, Jha D, Singh A, Panda AK (2016) Recovery of bioactive protein from bacterial inclusion bodies using trifluoroethanol as solubilization agent. Microb Cell Fact 15(1):100

Vajo Z, Fawcett J, Duckworth WC (2001) Recombinant DNA technology in the treatment of diabetes: insulin analogs. Endocr Rev 22(5):706-717

Voet D, Voet JG (2011) Biochemistry, 4-th Edition. NewYork: John Wiley & Sons Inc:492-496

Volontè F, Piubelli L, Pollegioni L (2011) Optimizing HIV-1 protease production in *Escherichia coli* as fusion protein. Microb Cell Fact 10(1):53

Walsh G (2005) Therapeutic insulins and their large-scale manufacture. Appl Microbiol Biotechnol 67(2):151-159

Young CL, Britton ZT, Robinson AS (2012) Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications. Biotechnol J 7(5):620-634

Zieliński M, Romanik-Chruścielewska A, Mikiewicz D, Łukasiewicz N, Sokołowska I, Antosik J, Sobolewska-Ruta A, Bierczyńska-Krzysik A, Zaleski P, Płucienniczak A (2019) Expression and purification of recombinant human insulin from *E. coli* 20 strain. Protein Expr Purif 157:63-69

# Chapter 3

## Sub/supercritical fluid chromatography employing a water-rich modifier enables the purification of biosynthesized human insulin.

**Sub/supercritical fluid chromatography employing a water-rich modifier enables the purification of biosynthesised human insulin.**

Kamini Govender[1], Tricia Naicker[1], Sooraj Baijnath[1], Anil Amichund Chuturgoon[2], Naeem Sheik Abdul[2], Taskeen Docrat[2], Hendrik Gerhardus Kruger[1*] and Thavendran Govender[3*]

[1]Catalysis and Peptide Research Unit, School of Health Sciences, University of KwaZulu-Natal, Durban, South Africa

[2]School of Laboratory Medicine and Medical Sciences, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa

[3]Department of Chemistry, University of Zululand, Private Bag X1001, KwaDlangezwa 3886, South Africa

*Corresponding authors: kruger@ukzn.ac.za or govendert@unizulu.ac.za, contact number: +27359026209

**Abstract**

There is a paucity of knowledge surrounding the SFC purification of human insulin. The current conventional method of insulin purification involves traditional RP-HPLC that utilises copious amounts of toxic solvents. In this study, we envisaged the development of an environmentally friendly SFC method for biosynthesised human insulin purification. Various commercially available SFC columns derived with silica, 2'ethyl pyridine, diol-HILIC, and the PFP functionalities were evaluated to determine the optimal stationary phase for purification. The PFP column gave the best results with respect to efficiencies of this important biologic that yielded average recoveries of 84%. LC-MS was used to initially detect and quantify the SFC purified standard sample of insulin (purchased) as well as the biosynthesised version. Protein sequencing was employed to verify the amino acid sequencing of the insulins; as such, the standard had a 90% probability to human insulin from the database, whereas the biosynthesised version had a 96% probability. The biological activities of both versions of the SFC purified proteins were assessed *in vitro* using a MTT assay. The results indicated that the biological activities of both samples were retained post-SFC purification. This study successfully proposes a greener and more efficient method for the purification of insulin derivatives.

**Keywords**: human insulin; sub/supercritical fluid chromatography purification.

## 1 Introduction

Human insulin is generally purified *via* a three-step approach comprising of ion exchange, size exclusion, and reversed-phase high-performance liquid chromatography (RP-HPLC) chromatography [1, 2]. The large scale pharmaceutical purification of biosynthesised insulin derived from recombinant deoxyribonucleic acid (DNA) technology is conventionally purified using preparative RP-HPLC methods [2]. This approach usually results in yields of approximately ≥ 75% on $C_{18}$ [3] and $C_8$ [2] stationary phase columns. In conventional analytical separations conducted *via* RP-HPLC, acetonitrile is employed as the mobile phase with flow rates of 1.0 mL/minute, and as much as 50 mL of waste is generated per run [2, 4]. In 2019, approximately 463 million people were diagnosed with Type 2 diabetes. If all of these patients are treated with insulin twice a day with the basal insulin treatment method of 0.1 units/kg/day, approximately $1.25 \times 10^8$ litres of acetonitrile would be generated in a year [2, 5-7]. Acetonitrile is the most extensively used organic solvent in RP-HPLC; however, the waste generation has a negative impact on the environment [8-10]. Apart from having a poor environmental profile, it's waste disposal also presents challenges, and it is also expensive [11, 12]. Therefore, there is a need for "greener" separation and purification processes that produce less waste and utilise more environmentally friendly solvents [4, 10, 13, 14].

SFC (sub/supercritical fluid chromatography) has been extensively used as analytical, semi-preparative, and preparative tools in the pharmaceutical industry over conventional HPLC for the separation of chiral compounds "[15-20]" and "peptides" [18, 21, 22]. SFC is gaining significant momentum in the pharmaceutical industry as it is a robust and powerful method that is well established for the rapid separation of enantiomers, with a two-fold reduction in time relative to conventional chiral phase HPLC [23-28]. It has numerous advantages such as better reproducibility, faster equilibration, and shorter run times, thereby increasing chromatographic productivity while reducing solvent consumption and waste [29-32]. SFC is considered to be environmentally friendly as the carbon dioxide ($CO_2$) used as the major mobile phase and can be efficiently recycled [30, 33].

Schiavone, *et al*. reported the SFC purification of commercially available bovine insulin as well as other proteins such as ubiquitin, cytochrome C, and apomyoglobin [34]. They employed techniques such as size exclusion chromatography coupled with hydrogen deuterium exchange (SEC-HDX) technology to monitor the after-effects of the SFC purification on the aforementioned proteins. The results indicated that the insulin could withstand the SFC conditions as the higher order structure was retained, but the same was not observed for ubiquitin, cytochrome C, and apomyoglobin, which could not re-fold to their original conformations post-SFC.

A major bottleneck in the drug development process is the purification of complex polar mixtures; therefore, new SFC techniques are required to increase respective analytes throughput, enabling efficient separations which expand the SFC applications to more polar biologic analytes [22, 35]. Studies have indicated that the addition of small amounts of water to the $CO_2$/methanol modifier has shown to enhance SFC chromatography by improving analytes solubility, peak shapes, allows for the separation of more polar analytes, a decrease in run times and enables a greener SFC purification process as less organic solvent is consumed [22, 34-37]. In our study, we demonstrated the effect of combining water with methanol, trifluoroacetic acid (TFA), and $CO_2$ modifier to enable the SFC purification of biosynthesised human insulin.

We aimed to develop a method for the purification of human insulin utilising SFC as opposed to the conventional established RP-HPLC approach [38]. Herein, an innovative insulin purification method is reported utilising SFC technology to purify crude biosynthesised human insulin at a semi-preparative scale. The standard (purchased) and the biosynthesised samples of insulin were quantified and characterised using liquid chromatography-mass spectrometry (LC-MS).

In this study also employed protein sequencing to validate that the purchased standard and biosynthesised samples were indeed human insulin derivatives. Organic solvents such as alcohols (which were used as modifiers) can cause the complete unfolding of a protein (denaturation), resulting in the formation of precipitates or aggregates [34, 39, 40]. Furthermore, the biological activities of the proteins after purification were evaluated to determine the effects of the SFC conditions on their structure employing an *in vitro* 3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyl tetrazolium bromide (MTT) assay. The post evaluation was important since the protein has a tendency to form insoluble precipitates when exposed to heat or hydrophobic surfaces. Therefore, this study will investigate the SFC purification of human insulin analogues and, thereafter, assess their biological activities.

## 2 Materials and methods

### 2.1 Chemicals and reagents

The standard sample of human insulin (catalogue number: I2643) was purchased from Sigma-Aldrich Inc. (Germany). The biosynthesised sample of the insulin employed in this study was genetically engineered *via* recombinant DNA technology utilising *E. coli* [41]. RP-HPLC grade acetonitrile, methanol, formic acid, and TFA were purchased from (Merck, Germany). Ultrapure water was obtained from a Millipore Milli-Q Gradient (Millipore, USA) water purification system. The hepatocellular carcinoma (HepG2) cell line was procured from Highveld Biologicals (South Africa), and the cell culture reagents were obtained from Whitehead Scientific (South Africa).

All other reagents were purchased from Merck (Germany) unless stated otherwise. The following SFC columns were employed in this study for the SFC purification of the biosynthesised human insulin: silica (250 mm × 4.6 mm; particle size 5 µm, Supelco™, USA), 2' ethyl pyridine (250 mm × 4.6 mm; particle size 5 µm, Princeton chromatography Inc., USA), diol-Hydrophilic interaction liquid chromatography (HILIC) (250 mm × 4.6 mm; particle size 5 µm, YMC, Japan) and pentafluoro phenyl (PFP) (250 mm × 4.6 mm; particle size 5 µm, YMC, Japan).

### 2.2 SFC purification

SFC purifications were conducted on a Sepiatec Prep SFC 30 system (Sepiatec, Germany). The SFC instrument comprised of a high-pressure pump, a back-pressure regulator, an injector with an injection loop, column oven and an ultraviolet (UV) detector. The UV detector was set at a wavelength of 220 nanometres (nm). The column oven temperature was maintained at 40°C, and injection volumes of 50 µL were used. The $CO_2$ pump head was cooled with a chiller, which was set at 5°C (PolyScience, USA). Various isocratic and gradient conditions were investigated. The mobile phases used were technical grade $CO_2$ (Afrox, South Africa) and a methanol modifier (Merck, Germany) that contained 5% (v/v) water and 0.2% (v/v) TFA. The gradient profile was 5-60%, whereby the modifier was initially held at 5% for one minute and thereafter increased to 60% over 6 minutes and maintained for a further 5 minutes. Flow rates of 4 mL/minute were optimal for all columns, with a total run time of approximately 13 minutes. Isocratic SFC conditions were conducted with a 10% modifier for approximately 10 minutes. After fraction collection, the standard and the biosynthesised samples were spun down using a rotary evaporator (Heidolph, Germany) and carefully transferred to microcentrifuge tubes (ISOLAB Laborgerate GmbH, Germany). The samples were freeze-dried overnight using a FreeZone 4.5 litre freeze drier (Labconco, USA).

### 2.3 LC-MS monitoring

A LC-MS-2020 (Shimadzu, Japan) was used to determine the purities of both the crude and purified samples. Injection volumes of 10 µL of unpurified standard (1 mg/mL) and crude biosynthesised human insulin (0.52 mg/mL) were used. However, post-SFC, in the LC-MS analysis, the injection volumes were increased to 50 µL of the samples. The LC-MS was coupled to a NM32LA nitrogen generator (Peak Scientific Instruments, United Kingdom). There were two mobile phases employed in the LC-MS analyses; mobile phase A consisted of Millipore water (Millipore, USA) and mobile phase B comprised of acetonitrile (Merck, Germany). The mobile phases above both contained 0.1% (v/v) formic acid, which served as an ion-pairing agent, with 1.0 mL/minute flow rates and run times of 25 minutes.

A 5% to 95% acetonitrile gradient was conducted; it was equilibrated for approximately 7 minutes at 5% acetonitrile, and thereafter ramped up to 95% over 18 minutes. A YMC-Triart $C_{18}$ (YMC, Japan) stainless steel column was employed for the LC-MS analysis, which contained the dimensions of 4.6 mm internal diameter (ID), a pore size of 120 Å with a particle size of 5 µm and a length of 150 mm. The temperature of the column was maintained at 40°C. The MS spectra were obtained using a mass range of 200 to 1500 m/z in a positive mode. The desolvation gas temperature was 250°C. The dry gas had a 10 L/minute flow rate, and the nebuliser was set at 1.5 bar.

### 2.4 Protein sequencing of insulin

The standard (purchased) and the biosynthesised samples of insulin were analysed at the Central Analytical Facility (Stellenbosch University, South Africa) [42] using LC-MS/MS conducted on an Ultimate 3000 RSLC (Thermo Fisher Scientific, USA) coupled to a Fusion mass spectrometer (Thermo Fisher Scientific, USA), *via* nano ionisation source (Nanospray Flex). The mass spectrometer files were subsequently imported onto the Proteome Discoverer v1.4 according to the manufacturer's instructions (Thermo Fisher Scientific, USA) and analysed using Sequest and an Amanda algorithm. The Uniprot P09211 linked to the contaminant protein database (cRAP) was used whereby a concatenated database interrogation was conducted. The peptides were validated using Target-Decoy PSM validator mode. The peptides were validated and identified using Protein Prophet and X! Tandem algorithms from the Scaffold 1.4.4 software (Proteome Software Inc, USA) according to the manufacturer's instructions.

### 2.5 Biological activity of insulin

A MTT assay was used to determine the biological activities of the standard and the biosynthesised insulin samples *in vitro*. HepG2 cell lines were employed in this study and were grown according to a similar method reported by Abdul *et al*., 2016 [43]. The standard and the biosynthesised insulin stock solutions were made in Eagle's minimum essential medium (EMEM) (catalogue number: 12-136F, Lonza, Switzerland) up to a concentration of 1 mg/mL. Three concentrations of the standard and biosynthesised samples of the insulin [low (50 µg/mL), medium (100 µg/mL), and high (150 µg/mL)] were used in the MTT assays.

The cells were grown on a 96-well plate for 24 hours ($2x10^4$ cells/well) followed by removal of the culture medium cells washed with phosphate-buffered saline (PBS). The HepG2 cells were subsequently treated with either the standard or biosynthesised samples of insulin prior to SFC purifications in normoglycaemic (5.5 mM) and hyperglycaemic (25 mM) media [44] for 15 minutes.

The above process was also conducted after SFC purifications for both the standard and the biosynthesised samples of the insulin. Thereafter, the cells were washed in PBS and 120 µl of MTT (5 mg/ml, PBS) and incubated for 4 hours at 37°C. Formazan crystals were solubilised *via* the addition of 100 µl of dimethyl sulfoxide (DMSO) per well, and optical densities (OD) were measured spectrophotometrically (BioTek uQuant, USA) at a wavelength of 570 nm. The experiments were conducted in triplicate.

The following formula was used to calculate cell viability (%):

$$\%\ \text{Cell viability} = \frac{\text{OD of treated cells}}{\text{OD of control cells}} \times 100$$

### 2.6 Statistical analyses

A two-way analysis of variance (ANOVA) was employed in this study. Analyses were conducted using the statistical software package Graph pad in stat version 8.1.0 (325) 64 bit for Windows (Graph pad software, San Diego, California).

## 3 Results and discussion

### 3.1 SFC results: column optimisation

Currently, there is limited information regarding SFC purification of insulin. However, as already mentioned, there is one recent report in which commercially available bovine insulin was purified using SFC on a preparative scale, and they reported the use of a 2'picolylamine column, with a mixture of acetonitrile and methanol as the modifier [34]. In our study, standard and crude biosynthesised human insulin derivatives were purified using SFC on a semi-preparative scale. The use of acetonitrile was eliminated due to its toxicity, therefore, making this a "greener" purification process [45]. LC-MS analysis, protein sequencing, and an *in vitro* MTT assays were employed in order to monitor both the standard and crude biosynthesised insulin derivatives post-SFC purification.

The characteristics of a protein play a fundamental role in its purification process and include its total charge, number of exposed hydrophobic groups, its binding capacity to the stationary phase as well as its shape, and size [46]. The selection of the correct stationary phase is imperative in the method development stages of SFC purifications. Proteins can interact with the chromatographic stationary phases in a variety of manners, whereby the retention times are affected by the protein's molecular composition [47].

Studies have indicated that to separate and purify strongly polar molecules, the addition of water can improve the molecule's solubility, peak shapes, chromatographic selectivity and speed [22, 34-37, 48, 49]. In this study, we implemented the addition of 5% (v/v) water to methanol with 0.2% (v/v) TFA and technical grade $CO_2$ to enable the SFC purification of human insulin. Initially the standard sample of insulin (1 mg/mL) was dissolved in 50% (v/v) water and 50% (v/v) methanol with 0.2 % (v/v) formic acid. The initial conditions for the evaluation of all of the columns employed a gradient method of 5-60% of the modifier over 13 minutes and the silica, diol-HILIC, 2' ethyl pyridine, and PFP columns were used for the initial screening of conditions. With the silica column, four peaks were observed with relatively long retention times, poor peak resolution, and high baseline noise leading us to believe that the protein was not eluted from the column (Figure 1-A). The 2' ethyl pyridine column displayed similar properties to that of the silica column (Figure 1-B). The diol-HILIC column had two peaks with good UV intensities (Figure 1-C), and the LC-MS analysis of both peaks (upon separation) gave the same retention times and masses when exposed to RP-HPLC (Figures S1 and S2 in the supporting information). This observation can be attributed to the protein separation of secondary structures that occurred between the protein and the packing material due to adsorption effects [50, 51]. Therefore, the PFP column was selected for further purification and optimisations of the biosynthesised sample of the insulin, as it displayed relatively better properties such as peak shapes, resolution, retention factors, retention times and the least relative standard deviation in comparison to the other columns (refer to Figure 1-D, Tables 1 and S1). This column is known for displaying ion-exchange, dipole-dipole, and π-π interactions with a variety of analytes as well as being highly stable [52, 53]. Subsequently, we optimised the method to even work well under isocratic conditions with a 10% methanol modifier that contains 5% (v/v) water and 0.2% (v/v) TFA (Figure S3). Isocratic conditions are preferred for stacked injections, as this is one of the main advantages of SFC protocols [24, 26].

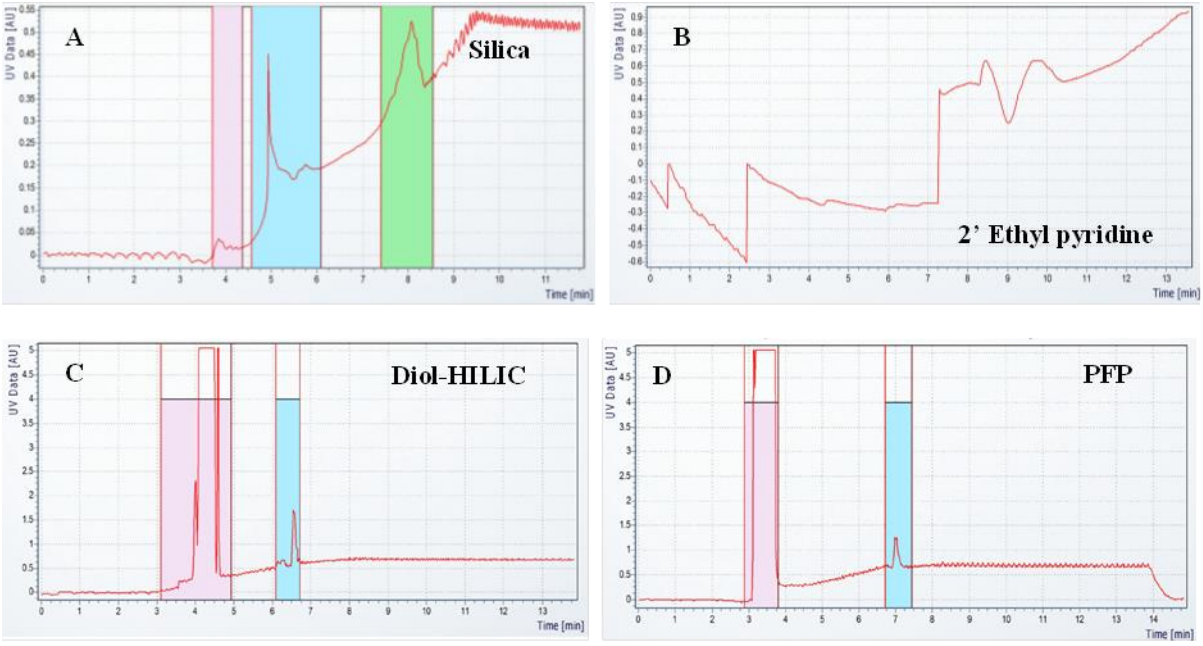


**Figure 1:** Represents SFC chromatograms of the standard sample of insulin (1 mg/mL) on analytical size columns (250 mm × 4.6 mm): (A) silica column (B) 2'ethyl pyridine (C) diol-HILIC (D) PFP. All columns were run using a gradient mode with a modifier range of 5-60%.

**Table 1**: Retention factors (k) of the standard sample of insulin (1 mg/mL).

| Column | k1 | k2 | k3 | k4 | Assessment of column |
|---|---|---|---|---|---|
| silica | 4.1 | 6.0 | 6.9 | 10 | Long retention times, poor peak resolution and baseline noise |
| 2' ethyl pyridine | 10.6 | 12.0 | - | - | Long retention times, baseline noise, poor peak shapes and peak resolution |
| diol-HILIC | 4.7 | 7.9 | - | - | Poor peak shape and peak splitting |
| PFP | 3.9 | 8.4 | - | - | No peak splitting, good peak shape, and resolution |

In an attempt to eliminate the phase separation peak and band broadening anomalies experienced during the SFC purification, various combinations of acids such as formic acid, acetic acid, and organic solvents such as DMSO as well as hexafluoro isopropanol (HFIP) were combined with methanol (Figure S4) while optimisation with the standard insulin sample. The use of HFIP and methanol was successful in the elimination of the phase separation peak for the standard sample and was the best condition in comparison to the former conditions (Figure S4).

### 3.2 PFP column efficiency test of the biosynthesised sample of insulin

Although a mixture of HFIP and methanol was used to dissolve the standard sample, Tris-HCl was required for the biosynthesised samples, since precipitation occurred under the former conditions and was probably due to the solubilities of some other proteins. This resulted in inaccurate total protein concentration determinations. This is a common effect as alcohol tends to denature some proteins [40]. Fortunately, Tris-HCl worked well for both the solubilisation of the standard and the biosynthesised samples while also the best condition to prevent the occurrence of phase separation peaks (Figures S3 and S5). With the optimised conditions in hand, the purification of the biosynthesised sample of insulin was attempted. PFP column efficiency tests were conducted using the biosynthesised sample, which yielded an average recovery of 84% (Figure 2-B). The weights of the freeze-dried samples were recorded; a PFP column efficiency test was determined according to the following equation:

$$\text{Column efficiency} = \frac{\text{Freeze} - \text{dried weight of biosynthesised sample after SFC purification}}{\text{Freeze} - \text{dried weight of biosynthesised sample before SFC purification}} \times 100$$

$$= \frac{17.8\ \text{mg}}{21.1\ \text{mg}} \times 100$$

= 84.36%

### 3.3 LC-MS mass spectra and protein sequencing of insulin

The standard, as well as the biosynthesised samples of insulin, were detected and confirmed using LC-MS analysis. The retention time for the standard sample of insulin was 12.5 minutes (Figures S6). The crude biosynthesised sample had numerous peaks before SFC purification which was indicative of the presence of numerous other proteins with a purity of approximately 19% (Figure 2-A); however, after SFC purification, there was one peak present at a retention time of 12.9 minutes with a purity of approximately 85% (LC-MS) therefore indicating the SFC purification was successful (Figure 3). The difference in retention times was attributed to the differences in sequences of insulin derivatives. The observed masses were validated with LC-MS, whereby the commercially available standard sample of insulin was 968.87 m/z (Figure S6), and the biosynthesised sample had a mass of 968.08 m/z. (Figure 3). The standard sample of insulin had a 90% probability to the human insulin from the database used for the protein sequencing (Figure 4, Tables S2 and S3), whereas the biosynthesised version had a 96% probability (Figure 5, Tables S4 and S5).

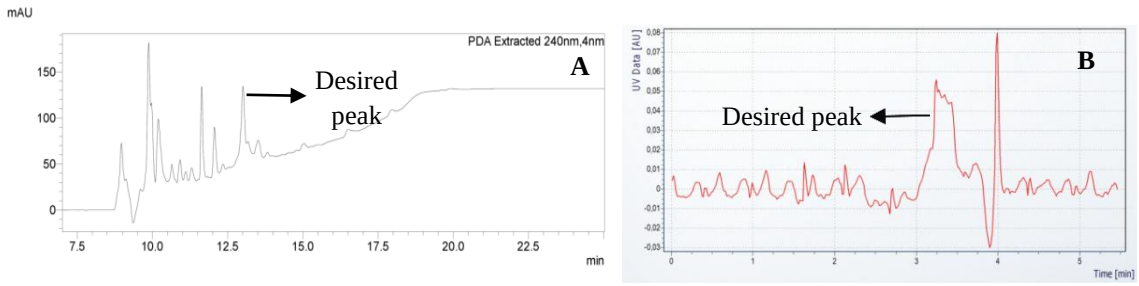


**Figure 2:** A Represents a LC-MS spectrum obtained from a YMC-Triart $C_{18}$ (dimensions: 150 mm × 4.6 mm (length x internal diameter), pore size: 120 Å, and a particle size of 5 µm) of the crude biosynthesised sample of insulin and B represents SFC chromatograms of the crude biosynthesised sample of insulin, which was separated on a PFP based column with the following characteristics; dimensions: 250 mm × 4.6 mm (length x internal diameter), pore size: 120 Å and a particle size of 5 µm.

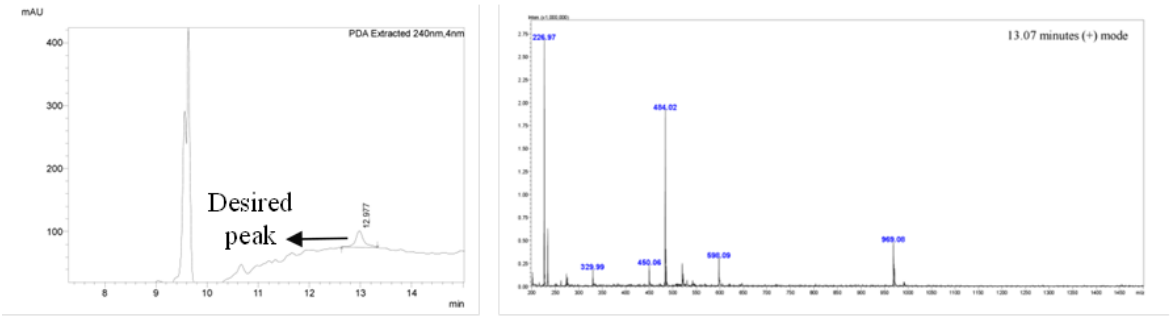


**Figure 3**: Represents a LC-MS spectrum obtained from a YMC-Triart $C_{18}$ (dimensions: 150 mm × 4.6 mm (length x internal diameter), pore size: 120 Å, and a particle size of 5 µm) of SFC purified biosynthesised sample of insulin.

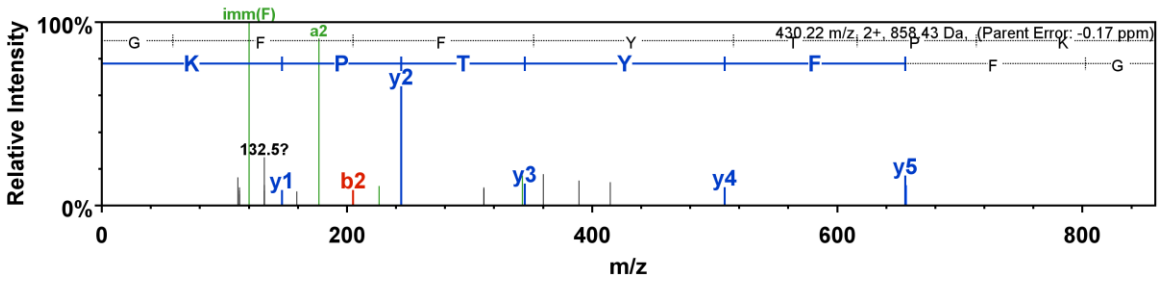


**Figure 4:** A peptide spectrum illustrating the standard sample of insulin, which had a 90% peptide sequence probability to human insulin derived from the Scaffold 1.4.4 software (Central Analytical Facility, Stellenbosch University, South Africa) [42].

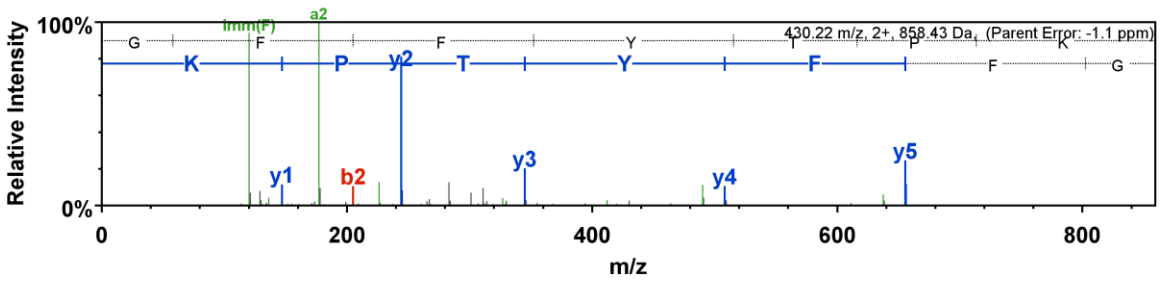


**Figure 5:** A peptide spectrum illustrating the biosynthesised sample of insulin after SFC purification with a 96% probability in peptide sequence to human insulin derived from the Scaffold 1.4.4 software (Central Analytical Facility, Stellenbosch University, South Africa) [42].

In this study, the standard sample of insulin was employed as a control for the optimisation of the SFC conditions of a biosynthesised insulin analogue. Insulin is an unstable entity as it can undergo various modifications, such as chemical reactions with surrounding molecules and proteins that are prone to denaturation during handling in the pharmaceutic industry [54]. Denaturation can also be caused by organic solvents, heating, freezing, pH extremes, and denaturants [55, 56]. Proteins undergo conformational changes with high temperature [57]. Proteins will also precipitate in high concentrations of organic solvents, such as alcohol [58]. Proteins are labile under high temperatures, and this can compromise biological activities and tertiary structures [58]. In this study, insulin was exposed to increased temperature (40°C) as well as alcohol, such as methanol under the SFC conditions (up to 260 bars of pressure). Therefore, the biological activity of the biosynthesised sample of insulin was tested *in vitro* using a MTT assay to determine the effect of SFC purification. The standard sample of insulin was utilised as a positive control in the MTT cell viability assay (Figure 6).

### 3.4 Determination of biological activities of insulin *via* a MTT assay

The MTT assay measures metabolically active cell viabilities, which is based on the cells' ability to produce reducing equivalents [59, 60]. The metabolically active cells uptake MTT by endocytosis or *via* a protein facilitated mechanism and results in the formation of a purple formazan product, which accumulates within living cells [60-62]. The metabolism of the dye releases the purple formazan product, which is quantified *via* colorimetric analysis [61]. MTT is reduced in metabolically active cells by the cells' electron transport systems [63] and respiratory chain [64].

The HepG2 cell line was grown under normoglycaemic (5.5 mM) and hyperglycaemic (25 mM) treatment conditions [44]. The MTT assay was employed to measure the cell viabilities of metabolically active cells based on their ability to produce reducing equivalents [65, 66]. Insulin is associated with the regulation of glucose uptake into muscle and fat cells and works within seconds to activate the tyrosine kinase receptor and the stimulation of glucose [67].

The glucose utilisation by the HepG2 cells increases cellular respiration and mitochondrial activities, increasing cell viabilities. The MTT assay results correlated with the former observation and showed an increase in cell viabilities for the hyperglycaemic treatments. Three treatment concentrations; low (50 ng/mL), medium (100 ng/mL) and high (150 ng/mL) were evaluated. The analyses were conducted using crude biosynthesised samples, and SFC purified biosynthesised samples of insulin. The cell viabilities of the SFC purified standard samples for the low, medium, and high concentrations were 219.79% ± 72.40%, 258.76% ± 25.83%, 356.35% ± 14.28% respectively. The cell viabilities of the SFC purified biosynthesised analogue were 260.15% ± 3.02%, 358.13% ± 17.00%, 350.24% ± 44.16% for the low, medium and high concentrations respectively (refer to Figure 6 and Figure 7). Both SFC insulin analogues exhibited similar cell viabilities to that of post-SFC purifications. As such, the MTT assay results revealed that the biological activities were retained after SFC purifications in all three concentrations. Future studies can investigate the use of ion mobility-mass spectrometry for analysis of samples post SFC [68, 69].

This study proves that SFC purification could be employed for the purification of insulin derivatives and provides a suitable alternative to RP-HPLC. In this study, the PFP column efficiency test yielded a recovery of 84% compared to traditional RP-HPLC methods of ≥ 75% [2]. The SFC run time was decreased to approximately 10 minutes (isocratic conditions, Figure S3) as compared to the traditional RP-HPLC of 50 minutes, which was conducted at an analytical scale [2]. This drastic reduction in time would decrease the amount of solvents being used in the overall purification process of insulin; therefore, this study reduced the carbon footprint of biosynthesised insulin.

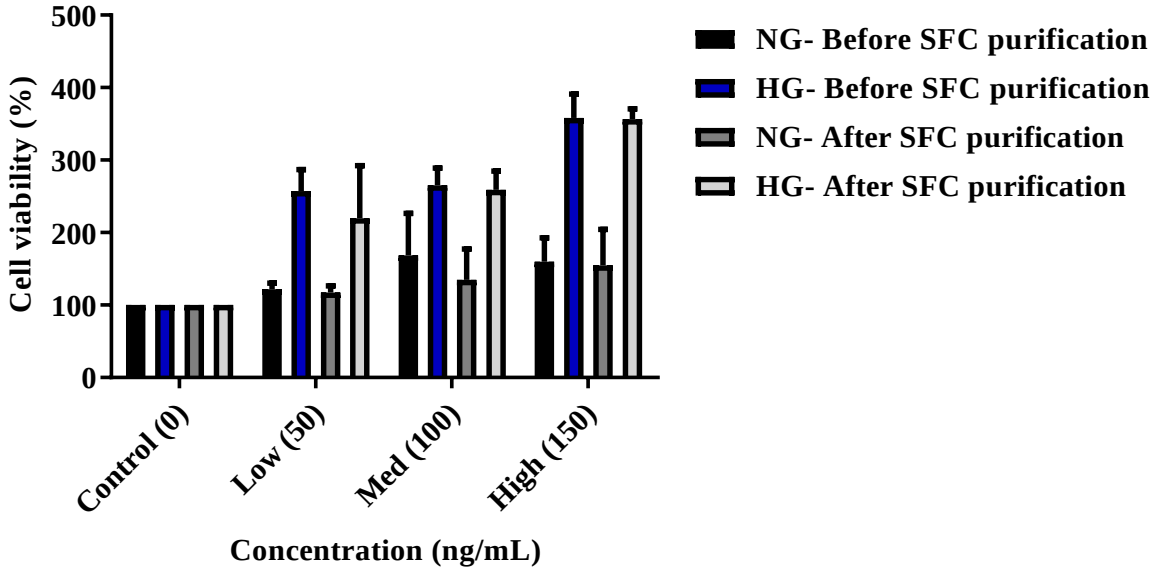


**Figure 6:** A graph displaying the cell viabilities of HepG2 cells, under normoglycaemic (NG) and hyperglycaemic (HG) conditions for low (50 ng/mL), medium (100 ng/mL) and high (150 ng/mL) MTT concentrations treated with standard insulin sample before SFC purification and after SFC purification (p-value < 0.0001).

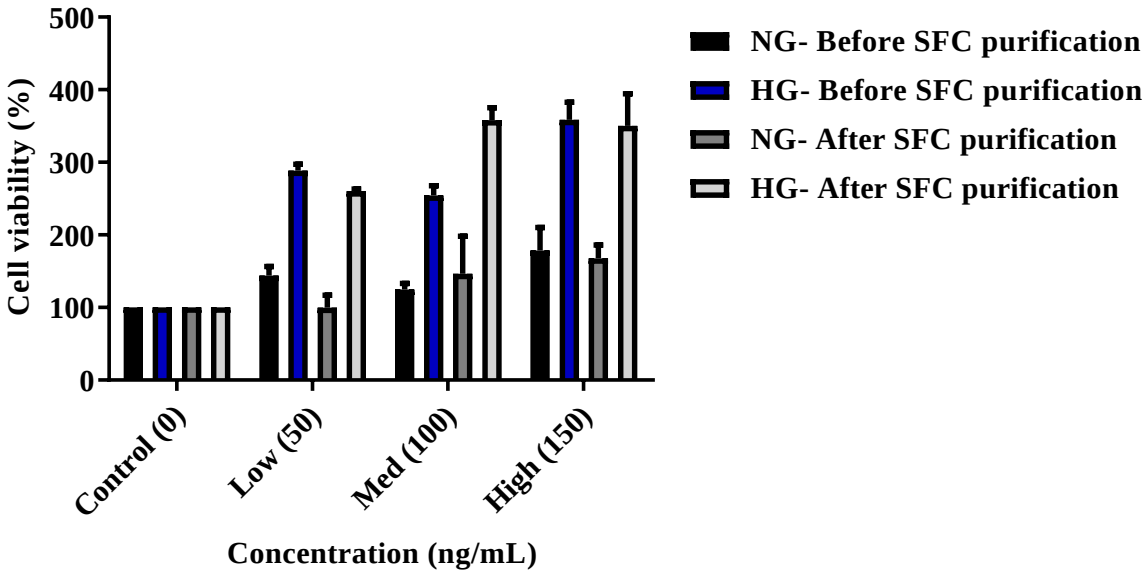


**Figure 7:** A graph displaying the cell viabilities of HepG2 cells, under normoglycaemic (NG) and hyperglycaemic (HG) conditions under low (50 ng/mL), medium (100 ng/mL) and high (150 ng/mL) MTT concentrations treated with biosynthesised insulin sample before SFC purification and after SFC purification (p-value < 0.0001).

## 4 Conclusion

In 2019 the number of adults with Type 2 diabetes was approximately 463 million; it is projected to increase to approximately 700 million by 2045 [8]. The rapid rise in patients with diabetes globally creates a severe demand for affordable insulin. This study successfully shows that SFC technology can be employed as a rapid, cost-effective, and greener method of purifying biosynthesised insulin as opposed to using conventional HPLC. Several SFC columns were assessed, and the PFP column gave the best results since it displayed good peak shapes, resolution, retention factors, retention times, and the least relative standard deviation in comparison to the other columns. Initially, the conditions for the purification were optimised using a standard sample. Thereafter, column efficiency tests were conducted at an analytical scale, yielding 84% recovery of the biosynthesised sample. The biological activity of the biosynthesised insulin analogue was determined using a MTT cytotoxicity assay after SFC purification with excellent retention of activity. The biological activities were highly significant, with a p-value of < 0.0001. Since the recovery from the SFC column was 84%, future endeavours for improvement of this method can implement the use of custom columns. In addition, the modifier effects, band broadening, and phase separation peaks are areas of development for industrial innovators. Future research can also investigate the use of different modifiers to determine its effect on the solubility of various proteins. This work serves as a proof-of-concept starting point and forms one of only two reports on the purification of biologically important proteins using cutting edge SFC technology.

## 5 Acknowledgements

This study was made possible through financial support from the University of KwaZulu-Natal, National Research Foundation (NRF), and Technology Innovation Agency (TIA).

## 6 Conflict of interest

The authors declare no conflict of interest.

## 7 Authors contribution

Kamini Govender conducted the experimental work, which formed the basis of her Ph.D. study, Taskeen Docrat, Dr. Naeem Sheik Abdul, and Prof. Anil Chuturgoon conducted the MTT assays. The rest of the authors are co-supervisors on the project and contributed to the conceptualisation of the idea, funding of the project, and scientific guidance.

## 8 References

1. Mollerup, I., S.W. Jensen, P. Larsen, O. Schou, L. Snel, and M.C. Flickinger, *Insulin Purification*, in *Encyclopedia of Industrial Biotechnology*. 2009, John Wiley & Sons, Inc.
2. Kroef, E.P., R.A. Owens, E.L. Campbell, R.D. Johnson, and H.I. Marks, *Production scale purification of biosynthetic human insulin by reversed-phase high-performance liquid chromatography.* Journal of Chromatography A, 1989. **461**: p. 45-61.
3. Parman, A. and J. Rideout, *Purification of porcine proinsulin by high-performance liquid chromatography.* Journal of Chromatography A, 1983. **256**: p. 283-291.
4. Welch, C.J., N. Wu, M. Biba, R. Hartman, T. Brkovic, X. Gong, R. Helmy, W. Schafer, J. Cuff, and Z. Pirzada, *Greening analytical chromatography.* TrAC Trends in Analytical Chemistry, 2010. **29**(7): p. 667-680.
5. Association, A.D., *8. Pharmacologic Approaches to Glycemic Treatment: Standards of Medical Care in Diabetes—2018*. Diabetes care, 2018. **41**(Supplement 1): p. S73-S85.
6. Chamberlain, M. and L. Stimmler, *The renal handling of insulin.* The Journal of clinical investigation, 1967. **46**(6): p. 911-919.
7. Atlas, D., *International diabetes federation.* IDF Diabetes Atlas, 9th edn. Brussels, Belgium: International Diabetes Federation, 2019.
8. Welch, C.J., T. Brkovic, W. Schafer, and X. Gong. *Performance to burn? Re-evaluating the choice of acetonitrile as the platform solvent for analytical HPLC*. Green Chemistry 2009 [cited 11 8]; 1232-1238].
9. McConvey, I.F., D. Woods, M. Lewis, Q. Gan, and P. Nancarrow, *The importance of acetonitrile in the pharmaceutical industry and opportunities for its recovery from waste.* Organic Process Research & Development, 2012. **16**(4): p. 612-624.
10. Sheldon, R.A., *The greening of solvents: Towards sustainable organic synthesis.* Current Opinion in Green and Sustainable Chemistry, 2019. **18**: p. 13-19.
11. Gilomen, K., H. Stauffer, and V. Meyer, *Detoxification of acetonitrile—water wastes from liquid chromatography.* Chromatographia, 1995. **41**(5-6): p. 488-491.
12. Miller, L.M., J.D. Pinkston, and L.T. Taylor, *Modern Supercritical Fluid Chromatography: Carbon Dioxide Containing Mobile Phases*. 2019: John Wiley & Sons.

13. Li, J., J. Albrecht, A. Borovika, and M.D. Eastgate, *Evolving green chemistry metrics into predictive tools for decision making and benchmarking analytics.* ACS Sustainable Chemistry & Engineering, 2017. **6**(1): p. 1121-1132.
14. Welch, C.J., T. Nowak, L.A. Joyce, and E.L. Regalado, *Cocktail chromatography: enabling the migration of HPLC to nonlaboratory environments.* ACS Sustainable Chemistry & Engineering, 2015. **3**(5): p. 1000-1009.
15. Patel, M., F. Riley, M. Ashraf-Khorassani, and L. Taylor, *Supercritical fluid chromatographic resolution of water soluble isomeric carboxyl/amine terminated peptides facilitated via mobile phase water and ion pair formation.* Journal of Chromatography A, 2012. **1233**: p. 85-90.
16. Barhate, C.L., M.F. Wahab, D. Tognarelli, T.A. Berger, and D.W. Armstrong, *Instrumental idiosyncrasies affecting the performance of ultrafast chiral and achiral sub/supercritical fluid chromatography.* Analytical chemistry, 2016. **88**(17): p. 8664-8672.
17. Lesellier, E. and C. West, *The many faces of packed column supercritical fluid chromatography – A critical review.* Journal of Chromatography A, 2015. **1382**(Supplement C): p. 2-46.
18. Bolaños, B., M. Greig, M. Ventura, W. Farrell, C.M. Aurigemma, H. Li, T.L. Quenzer, K. Tivel, J.M. Bylund, and P. Tran, *SFC/MS in drug discovery at Pfizer, La Jolla.* International Journal of Mass Spectrometry, 2004. **238**(2): p. 85-97.
19. Biba, M. and J. Liu, *A Perspective on the Application of Preparative Supercritical Fluid Chromatography Using Achiral Stationary Phases in Pharmaceutical Drug Discovery and Development.* American Pharmaceutical Review, 2016.
20. Płotka, J.M., M. Biziuk, C. Morrison, and J. Namieśnik, *Pharmaceutical and forensic drug applications of chiral supercritical fluid chromatography*. TrAC Trends in Analytical Chemistry, 2014. **56**: p. 74-89.
21. Ventura, M., *Advantageous use of SFC for separation of crude therapeutic peptides and peptide libraries.* Journal of pharmaceutical and biomedical analysis, 2020: p. 113227.
22. Liu, J., A.A. Makarov, R. Bennett, I.A. Haidar Ahmad, J. DaSilva, M. Reibarkh, I. Mangion, B.F. Mann, and E.L. Regalado, *Chaotropic Effects in Sub/Supercritical Fluid Chromatography via Ammonium Hydroxide in Water-Rich Modifiers: Enabling Separation of Peptides and Highly Polar Pharmaceuticals at the Preparative Scale.* Analytical chemistry, 2019. **91**(21): p. 13907-13915.
23. Miller, L. and M. Potter, *Preparative chromatographic resolution of racemates using HPLC and SFC in a pharmaceutical discovery environment.* Journal of Chromatography B, 2008. **875**(1): p. 230-236.
24. Nogle, L.M., C.W. Mann, W.L. Watts Jr, and Y. Zhang, *Preparative separation and identification of derivatized β-methylphenylalanine enantiomers by chiral SFC, HPLC and NMR for development of new peptide ligand mimetics in drug discovery.* Journal of pharmaceutical and biomedical analysis, 2006. **40**(4): p. 901-909.
25. Speybrouck, D. and E. Lipka, *Preparative supercritical fluid chromatography: A powerful tool for chiral separations.* Journal of Chromatography A, 2016. **1467**: p. 33-55.
26. Miller, L., *Evaluation of non-traditional modifiers for analytical and preparative enantioseparations using supercritical fluid chromatography.* Journal of Chromatography A, 2012. **1256**: p. 261-266.
27. Miller, L., *Use of dichloromethane for preparative supercritical fluid chromatographic enantioseparations.* Journal of Chromatography A, 2014. **1363**: p. 323-330.
28. Barhate, C.L., L.A. Joyce, A.A. Makarov, K. Zawatzky, F. Bernardoni, W.A. Schafer, D.W. Armstrong, C.J. Welch, and E.L. Regalado, *Ultrafast chiral separations for high throughput enantiopurity analysis.* Chemical Communications, 2017. **53**(3): p. 509-512.
29. Zheng, J., J. Pinkston, P. Zoutendam, and L. Taylor, *Feasibility of supercritical fluid chromatography/mass spectrometry of polypeptides with up to 40-mers.* Analytical chemistry, 2006. **78**(5): p. 1535-1545.
30. Berger, T.A., *Supercritical fluid chromatography*. Primer, publication number 5991-5509EN. 2015, USA: Agilent technologies. 186.
31. Taylor, L.T., *Supercritical fluid chromatography.* Analytical chemistry, 2010. **82**(12): p. 4925-4935.

32. DaSilva, J.O., D. Lehnherr, J. Liu, R. Bennett, I.A. Haidar Ahmad, M. Hicks, B.F. Mann, D.A. DiRocco, and E.L. Regalado, *Generic Enhanced Sub/Supercritical Fluid Chromatography: The Blueprint for Highly Productive and Sustainable Separation of Primary Hindered Amines*. ACS Sustainable Chemistry & Engineering, 2020.
33. Rajendran, A., *Design of preparative-supercritical fluid chromatography*. Journal of Chromatography A, 2012. **1250**: p. 227-249.
34. Schiavone, N.M., R. Bennett, M.B. Hicks, G.F. Pirrone, E.L. Regalado, I. Mangion, and A.A. Makarov, *Evaluation of global conformational changes in peptides and proteins following purification by supercritical fluid chromatography*. Journal of Chromatography B, 2019. **1110**: p. 94-100.
35. Roy, D., M.F. Wahab, T.A. Berger, and D.W. Armstrong, *Ramifications and Insights on the Role of Water in Chiral Sub/Supercritical Fluid Chromatography*. Analytical chemistry, 2019. **91**(22): p. 14672-14680.
36. Liu, J., E.L. Regalado, I. Mergelsberg, and C.J. Welch, *Extending the range of supercritical fluid chromatography by use of water-rich modifiers*. Organic & Biomolecular Chemistry, 2013. **11**(30): p. 4925-4929.
37. Taylor, L.T., *Packed column supercritical fluid chromatography of hydrophilic analytes via water-rich modifiers.* Journal of Chromatography A, 2012. **1250**: p. 196-204.
38. Reuhs, B.L., *High-performance liquid chromatography*, in *Food analysis*. 2017, Springer. p. 213-226.
39. Brange, J., L. Andersen, E.D. Laursen, G. Meyn, and E. Rasmussen, *Toward understanding insulin fibrillation.* Journal of pharmaceutical sciences, 1997. **86**(5): p. 517-525.
40. Hirota, N., Y. Goto, and K. Mizuno, *Cooperative α-helix formation of β-lactoglobulin and melittin induced by hexafluoroisopropanol*. Protein science, 1997. **6**(2): p. 416-421.
41. Govender, K., T. Naicker, J. Lin, S. Baijnath, A.A. Chuturgoon, N.S. Abdul, T. Docrat, H.G. Kruger, and T. Govender, *A novel and more efficient biosynthesis approach for human insulin production in Escherichia coli (E. coli).* AMB Express, 2020. **10**(1): p. 43.
42. ; Available from: http://www.sun.ac.za/english/faculty/science/CAF.
43. Abdul, N.S., S. Nagiah, and A.A. Chuturgoon, *Fusaric acid induces mitochondrial stress in human hepatocellular carcinoma (HepG2) cells.* Toxicon, 2016. **119**: p. 336-344.
44. Chen, Q., Y. Xia, and Z. Qiu. *Effect of ecdysterone on glucose metabolism in vitro*. Life sciences 2006 [cited 2019 23 April 2019]; 1108-1113].
45. Sheldon, R.A., *The greening of solvents: Towards sustainable organic synthesis.* Current Opinion in Green and Sustainable Chemistry, 2018.
46. Coskun, O., *Separation techniques: chromatography*. Northern clinics of Istanbul, 2016. **3**(2): p. 156.
47. Aguilar, M., *HPLC of peptides and proteins: basic theory and methodology*. Methods in molecular biology (Clifton, NJ), 2004. **251**: p. 3.
48. West, C., J. Melin, H. Ansouri, and M.M. Metogo, *Unravelling the effects of mobile phase additives in supercritical fluid chromatography. Part I: polarity and acidity of the mobile phase*. Journal of Chromatography A, 2017. **1492**: p. 136-143.
49. Bennett, R., M. Biba, J. Liu, I.A.H. Ahmad, M.B. Hicks, and E.L. Regalado, *Enhanced fluidity liquid chromatography: A guide to scaling up from analytical to preparative separations.* Journal of Chromatography A, 2019. **1595**: p. 190-198.
50. Mori, S. and M. Kato, *High-performance aqueous size-exclusion chromatography using diol-bonded porous glass packing materials. Retention behavior of some proteins.* Journal of Liquid Chromatography, 1987. **10**(14): p. 3113-3126.
51. Pesek, J.J., M.T. Matyska, and L. Mauskar, *Separation of proteins and peptides by capillary electrochromatography in diol-and octadecyl-modified etched capillaries.* Journal of Chromatography A, 1997. **763**(1-2): p. 307-314.
52. Danielson, N.D., L.G. Beaver, and J. Wangsa, *Fluoropolymers and fluorocarbon bonded phases as column packings for liquid chromatography*. Journal of Chromatography A, 1991. **544**: p. 187-199.

53. West, C., E. Lemasson, S. Khater, and E. Lesellier, *An attempt to estimate ionic interactions with phenyl and pentafluorophenyl stationary phases in supercritical fluid chromatography.* Journal of Chromatography A, 2015. **1412**: p. 126-138.
54. Wang, W., *Instability, stabilization, and formulation of liquid protein pharmaceuticals.* International journal of pharmaceutics, 1999. **185**(2): p. 129-188.
55. Manning, M.C., K. Patel, and R.T. Borchardt, *Stability of protein pharmaceuticals.* Pharmaceutical research, 1989. **6**(11): p. 903-918.
56. Wang, W., *Protein aggregation and its inhibition in biopharmaceutics.* International journal of pharmaceutics, 2005. **289**(1-2): p. 1-30.
57. Roumeliotis, P. and K. Unger, *Assessment and optimization of system parameters in size exclusion separation of proteins on diol-modified silica columns.* Journal of Chromatography A, 1981. **218**: p. 535-546.
58. Rubinstein, M., *Preparative high-performance liquid partition chromatography of proteins*, in *Analytical biochemistry*. 1979. p. 1-7.
59. Nikzad, S., M. Baradaran-Ghahfarokhi, and P. Nasri, *Dose-response modeling using MTT assay: a short review*. Life Sci J, 2014. **11**(9s): p. 432-7.
60. Hansen, J. and P. Bross, *A cellular viability assay to monitor drug toxicity*, in *Protein Misfolding and Cellular Stress in Disease and Aging*. 2010, Springer. p. 303-311.
61. Maioli, E., C. Torricelli, V. Fortino, F. Carlucci, V. Tommassini, and A. Pacini, *Critical appraisal of the MTT assay in the presence of rottlerin and uncouplers.* Biological procedures online, 2009. **11**(1): p. 227.
62. Mosmann, T., *Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays.* Journal of immunological methods, 1983. **65**(1-2): p. 55-63.
63. Liu, Y., D.A. Peterson, H. Kimura, and D. Schubert, *Mechanism of cellular 3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyltetrazolium bromide (MTT) reduction.* Journal of neurochemistry, 1997. **69**(2): p. 581-593.
64. Slater, T., B. Sawyer, and U. Sträuli, *Studies on succinate-tetrazolium reductase systems: III. Points of coupling of four different tetrazolium salts III. Points of coupling of four different tetrazolium salts.* Biochimica Et Biophysica Acta-Molecular and Cell Biology of Lipids, 1963. **77**: p. 383-393.
65. Carrió, M.M. and A. Villaverde, *Localization of chaperones DnaK and GroEL in bacterial inclusion bodies.* Journal of bacteriology, 2005. **187**(10): p. 3599-3601.
66. Morihara, K., T. Oka, H. Tsuzuki, Y. Tochino, and T. Kanaya, *Achromobacter protease I-catalyzed conversion of porcine insulin into human insulin.* Biochemical and Biophysical Research Communications, 1980. **92**(2): p. 396-402.
67. Drejer, K., *The bioactivity of insulin analogues from in vitro receptor binding to in vivo glucose uptake.* Diabetes/metabolism reviews, 1992. **8**(3): p. 259-285.
68. Lanucara, F., S.W. Holman, C.J. Gray, and C.E. Eyers, *The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics.* Nature Chemistry, 2014. **6**(4): p. 281.
69. Ben-Nissan, G. and M. Sharon, *The application of ion-mobility mass spectrometry for structure/function investigation of protein complexes.* Current Opinion in Chemical Biology, 2018. **42**: p. 25-33.

# Chapter 4

**The development of a sub/supercritical fluid chromatography-based purification method for peptides.**

**The development of a sub/supercritical fluid chromatography-based purification method for peptides.**

Kamini Govender[1], Tricia Naicker[1], Sooraj Baijnath[1], Hendrik Gerhardus Kruger[1*], and Thavendran Govender[2*]

[1]Catalysis and Peptide Research Unit, School of Health Sciences, University of KwaZulu-Natal, Durban, South Africa

[2] Department of Chemistry, University of Zululand, Private Bag X1001, KwaDlangezwa 3886, South Africa

*Corresponding authors: kruger@ukzn.ac.za and govendert@unizulu.ac.za

Catalysis and Peptide Research Unit (CPRU)

E-block, 6th Floor, Room E1-06-016

University of KwaZulu-Natal, Westville Campus, South Africa

Contact number: +27312601845

**Abstract**

Peptide drugs are essential components of the pharmaceutical industry with a multiplicity of therapeutic properties, such as being anti-hypertensive, anti-microbial, anti-diabetic, and anti-cancer. These molecules are similar in physiological structure and function to the body's endogenous signalling molecules and are therefore ideal candidates for the development of the next generation of drugs. However, the purification of these peptides can be problematic due to poor solubility and stability, which often results in low peptide yields. Peptides are traditionally purified *via* RP-HPLC methods, which are tedious and employ harsh solvents that generate harmful waste to the environment. There is a growing need for more cost-effective and sustainable purification methods of these biologics. SFC can provide a greener peptide purification approach with more environmentally friendly mobile phases such as $CO_2$ and methanol, which can easily be recycled with minimal environmental impact. Currently, there is limited knowledge regarding the SFC purification of peptides. Herein, this study investigated SFC methods to purify a tetrapeptide (LYLV), octapeptide (DRVYIHPF), and nonapeptide (LYLVCGERG) on commercially available columns at an analytical scale. The 2' ethyl pyridine column proved to be optimal based on its reproducibility, peak shapes, efficient separations, and retention factors with peptide recoveries ranging from 80-102%. The run times were reduced to 13 minutes, as opposed to the traditional RP-HPLC methods of 50 minutes, thus making this SFC method an efficient, greener, and more cost-effective approach for the purification of these peptides.

**Keywords**: Sub/supercritical fluid chromatography (SFC); PFP column; diol-HILIC column; 2' ethyl pyridine column; tetrapeptide [insulin β chain peptide (15-18)]; octapeptide [angiotensin II]; nonapeptide [insulin β chain peptide (15-23)].

## 1 Introduction

Peptides are a privileged class of pharmaceutical biologics that display distinct therapeutic and biochemical properties [1]. Advances in genomics, transcriptomics, proteomics, and protein engineering have resulted in the synthesis of peptides with similar molecular structures to that of endogenous signalling compounds in the body, such as neurotransmitters [2-5]. Peptides are extremely versatile, with several medicinal benefits ranging from anti-viral, anti-diabetic, anti-cancer, and anti-microbial properties [6-8]. The physiochemical characteristics of peptides are that of being relatively small in size while displaying structural diversities, low toxicities, high biological activities, high tissue affinities, and specificities [9-11]. Peptides are postulated to be the next-generation of highly prescribed drugs, as exemplified by their already routine use in several major diseases [6, 11-15].

There are several problems associated with the mass production of peptides, such as the high costs concerning the synthesis and especially the purification [12]. Currently, the purification of peptides are effected using a range of chromatographic methods, including reversed-phase high-performance liquid chromatography (RP-HPLC) [16-20], size exclusion chromatography (SEC) [16, 19-21], and ion-exchange chromatography (IEC) [16, 19, 20, 22]. These methods are normally lengthy and costly [19, 23-26]. This is a major bottleneck that is commonly experienced during the production of peptide therapeutics [8, 12]. Conventional analytical separations generate approximately 60 mL's of waste per run that is potentially hazardous to the environment and increases operating costs [20, 25-28]. Another major problem is associated with the solubilities and stabilities of these compounds during the purification processes [12]. Consequently, there is a need for more greener and environmentally friendly processes [26, 29-32].

Purification *via* sub/supercritical fluid chromatography (SFC) of such compounds may provide a possible solution, with a considerable gain in the momentum of this technique in the pharmaceutical industry in recent times [16, 33-36]. Essentially, SFC can reduce the amount of organic solvents being used during the purification processes since it employs carbon dioxide ($CO_2$) as the major mobile phase, thereby resulting in less energy being consumed during the evaporation and fractionation processes [37]. The separation of drugs on the SFC has proven to have higher efficiencies, faster run times, and stacked injections, which allows for higher chromatographic productivities and the reduction of solvent waste [16, 33, 34, 38-44]. However, there are also disadvantages associated with SFC purification, for example, the mismatch of modifiers utilised to dissolve the analyte and the mobile phase employed, which can result in broad peak profiles [33, 45]. Currently, there is limited information regarding the SFC purification of biomolecules, such as proteins and peptides. However, there was a recent study by Schiavone *et al*. in which they reported the successful preparative scale

SFC purification of bradykinin (peptide) and commercially available proteins such as bovine insulin, ubiquitin, and cytochrome C [16]. The aforementioned study employed a 2'picolylamine column with $CO_2$ and a mixture of acetonitrile and methanol together with 5% $H_2O$ and 0.2% trifluoroacetic acid (TFA) additives as a modifier [16]. However, the potential drawbacks of the aforementioned study were that the higher-order structures of ubiquitin, cytochrome C, and apomyoglobin were not retained after SFC purification. Schiavone *et al*. concluded that the bradykinin and the bovine insulin can be purified using this method since insulin's higher-order structures were retained post-SFC [16]. Another study conducted by Tognatelli *et al*. employed SFC on an analytical scale using a 2'ethyl pyridine column whereby they successfully purified a mixture of peptides such as G3502, V8376, methionine encephalin acetate, angiotensin II acetate, and leucine encephalin acetate and utilised $CO_2$ and methanol as a modifier with 0.1% TFA [25].

A major bottleneck in the drug development process is the purification of polar mixtures; therefore, it is imperative that new SFC techniques are developed, which expand SFC purification to more polar analytes [33, 35, 46]. Studies have shown that the addition of small amounts of water to the methanol/ $CO_2$ modifier has enhanced SFC chromatography by improving analytes solubility and aids in the separation of more polar analytes [16, 33, 35, 46-48]. Therefore, in our study, we demonstrated the effect of combining water with methanol, TFA, and $CO_2$ modifier to enable the SFC purification of crude peptides.

Our study aimed to build on this developing field and employed peptides associated with diabetes and cardiovascular diseases by also evaluating SFC technology of some commercially available columns at an analytical scale. The following analytes were selected as candidates since these peptides are commonly used in the treatment of non-communicable diseases, a tetrapeptide [insulin β chain peptide (15-18)], octapeptide [angiotensin II], and a nonapeptide [insulin β chain peptide (15-23)][49]. These crude peptides were assessed on the pentafluoro phenyl (PFP), diol-hydrophilic interaction liquid chromatography (HILIC), and the 2' ethyl pyridine SFC columns employing technical grade $CO_2$ and methanol, which contained 0.2% TFA as well as 5% water as the modifier. Thereafter, post-SFC the aforementioned peptides were quantified using liquid chromatography-mass spectrometry (LC-MS).

## 2 Materials and methods

### 2.1 Chemicals and reagents

Methanol, acetonitrile, TFA, and formic acid were purchased from (Merck, Germany). Ultrapure water was obtained from a Millipore Milli-Q Gradient (Millipore, USA) water purification system. All other solvents and reagents were acquired from Merck (Germany) unless stated otherwise. The crude peptides (Table 1) were purchased from AnaSpec, Incorporated (USA).

### 2.2 LC-MS monitoring

The current study employed LC-MS-2020 (Shimadzu, Japan) for the monitoring of the peptides. The instrument consisted of a binary pump, an autosampler, a mass spectrometer, and a NM32LA nitrogen generator (Peak Scientific Instruments, United Kingdom). The two mobile phases utilised in the LC-MS analyses were mobile phase A: Millipore water (Millipore, USA) and mobile phase B: acetonitrile (Merck, Germany). These mobile phases both contained formic acid at a concentration of 0.1% (v/v) as the ion-pairing agent (Merck, Germany). Injection volumes of 10 µL for the crude tetrapeptide (1 mg/mL), octapeptide (2 mg/mL) and the nonapeptide (1.5 mg/mL) were used. The tetrapeptide was analysed using LC-MS method A with a 5% to 95% gradient profile whereby the acetonitrile was ramped up from 5-95% over 18 minutes. The octapeptide and the nonapeptide were analysed using a LC-MS method B, whereby the acetonitrile gradient was changed to 5-70% over 22 minutes. The flow rates for both methods were 1.0 mL/minute on a stainless steel YMC-triart $C_{18}$ column (YMC, Japan) with the following dimensions; particle size of 5 µm, a 4.6 mm internal diameter, 150 mm in length and pore size of 120 Å was used. The nebuliser setting was 1.5 bar, and the desolvation gas temperature was set at 250°C. The MS spectra were conducted in positive mode with a mass range from 200 to 1500 m/z.

**Table 1**: Molecular weight and sequences of crude peptides used in this study

| Peptide | Peptide length | Molecular weight (Daltons) | Peptide sequence |
|---|---|---|---|
| Insulin β chain peptide (15-18) | Tetrapeptide (4) | 505.7 ± 0.2% | LYLV |
| Angiotensin II | Octapeptide (8) | 1045.2 ± 0.2% | DRVYIHPF |
| Insulin β chain peptide (15-23) | Nonapeptide (9) | 1008.3 ± 0.2% | LYLVCGERG |

### 2.3 Columns

**Table 2**: Physical dimensions of SFC columns employed in this study

| Column packing material | Dimensions (L x I.D.) mm | Pore Size (Å) | Particle Size (µm) | Supplier |
|---|---|---|---|---|
| PFP | 250 × 4.6 | 120 | 5 | YMC (Japan) |
| diol-HILIC | 250 × 4.6 | 120 | 5 | YMC (Japan) |
| 2’ ethyl pyridine | 250 × 4.6 | 100 | 5 | Princeton chromatography Inc. (USA) |
| 2’ ethyl pyridine | 250 × 10 | 300 | 5 | Princeton chromatography Inc. (USA) |

### 2.4 Solubility of crude peptides

The crude insulin β chain peptide (15-18) (tetrapeptide) and angiotensin II (octapeptide) were dissolved in neat methanol to a concentration of 1 mg/mL and 2 mg/mL respectively. However, the crude insulin β chain peptide (15-23) nonapeptide needed approximately 96% (v/v) methanol, 4% (v/v) water and 0.16% (v/v) TFA for complete solvation at a concentration of 1.5 mg/mL.

### 2.5 SFC instrumentation

The SFC system (Prep SFC 30, Sepiatec, Germany) was composed of an injector loop, column oven (maintained at a temperature of 40°C), a back pressure regulator (BPR) (set at 150 bar), a $CO_2$ pump with the pump head maintained at 5°C (digital chiller from Polyscience, USA), a high-pressure pump for the modifier, and ultraviolet (UV) wavelength detector (set at 220 nanometres (nm)). 50 µL injections were conducted for the tetrapeptide (1 mg/mL), octapeptide (2 mg/mL) and the nonapeptide (1.5 mg/mL). The mobile phases employed in this study were technical grade $CO_2$ (Afrox, South Africa) and methanol, which contained 0.2% (v/v) TFA and 5% (v/v) water (Merck, Germany). A 5-60% gradient was used, whereby the methanol modifier was initially held for 1 minute, ramped up to 60% over 6 minutes, and held for a further 5 minutes. The total run time was approximately 13 minutes. Flow rates of 8 mL/minute were utilised for the 2’ ethyl pyridine (10mm internal diameter (I.D.) column); whereas 4 mL/minute flow rates were utilised for the smaller diameter columns (4mm I.D.). The following parameters were employed to evaluate the SFC column performances, such as reproducibility, peak shapes, efficient SFC peptide separations, and retention factors. After fraction collection, the purified peptide samples were spun down using a rotary evaporator (Heidolph, Germany) and subsequently stored at 4°C for further analysis. The SFC purifications of the three aforementioned peptides were tested in triplicate (n=3), and standard deviations and relative standard deviations were conducted.

### 2.6 Peptide recoveries

Multiple injections (50 µL) were run on the Sepiatec Prep SFC 30 system (Sepiatec, Germany). After fraction collection, the SFC purified tetrapeptide, octapeptide, and the nonapeptide was carefully placed in the rotary evaporator (Heidolph, Germany) to reduce the amount of methanol and subsequently freeze-dried overnight using a FreeZone 4.5 litre freeze drier (Labconco, USA). The weights of the freeze-dried peptide samples were recorded; the peptide recoveries were determined according to the following equation:

$$\text{Peptide recovery } = \frac{\text{Peptide weight after SFC purification}}{\text{Peptide weight before SFC purification}} \times 100$$

## 3 Results and discussion

The separation of peptides and proteins is dependent on the differential adsorption of the respective solutes and their affinities for the stationary phases, with the degree of its binding affinities being dependent on the solute's structures. When a molecule has a higher binding affinity for a particular stationary phase, it is retained longer on the column as compared to a molecule with a lower binding affinity [19]. The correct selection of an ideal stationary phase is imperative for the separation of these analytes. The analyses and purifications of highly polar compounds such as nucleic acids and amino acids have been widely avoided, due to poor peak shapes and long retention times, when using traditional mobile phases and stationary phases using SFC [16]. Studies have indicated that this technique can be utilised to separate and purify strongly polar molecules through the addition of water to the modifier phase to improve the molecule's solubility, peak shapes and to aid in a reduction of stationary phase interactions [16, 33, 35, 46-48, 50, 51].

This study implemented the use of 5% water with 0.2% TFA to methanol/$CO_2$, which was employed as the modifier. Various SFC columns were evaluated to determine the optimal stationary phases for the purification of the crude, tetrapeptide [insulin β chain peptide (15-18)], octapeptide [angiotensin II], and nonapeptide [insulin β chain peptide (15-23). The peptides were detected and monitored using LC-MS analysis before and after the SFC purification.

### 3.1 The LC-MS analyses of the crude tetrapeptide, octapeptide, and nonapeptide

Both the crude tetrapeptide and the octapeptide dissolved well in neat methanol; however, the crude octapeptide required more time under sonication and required the addition of 4% water for complete dissolution. The crude tetrapeptide (Figure S1), octapeptide (Figure S4), and nonapeptide (Figure S7) were analysed *via* LC-MS before the subjection to SFC purification to confirm their masses and purity. The crude tetrapeptide displayed two major peaks (Figure S1); one with a mass of 506.55 m/z which correlated to the tetrapeptide (68.36%) (Figures S1 and S2) and the other with a mass of 612.54 m/z (Figures S1 and S3) that corresponded to an impurity (31.65%) which was as a result of the rink amide resin linker by-product [52]. The LC-MS spectra for the crude octapeptide showed one major peak (93%) (Figures S4 and S5) with a mass of 1045.66 m/z and several smaller peaks near the closest one having a mass of 605.24 m/z corresponding to a double coupling of tyrosine (7%) (Figures S4 and S6). The crude nonapeptide (92.59%) had a mass of 1008.64 m/z (Figures S7 and S8) with a shoulder peak mass of 1064.75 m/z (Figures S7 and S9) corresponding to the impurity of a tert-butyloxycarbonyl protecting group (7.41%). With this information in hand, the peptides were tested on various SFC columns and conditions for optimal purification.

### 3.2 The SFC purification of peptides

The crude tetrapeptide, octapeptide, and the nonapeptide were evaluated using the following SFC columns; PFP, diol-HILIC, and 2'ethyl pyridine. During the SFC purification of these peptides, there were UV wavelength detector noise observed on the SFC chromatograms (Figures 1, 3 and 4), there are three possible sources of this noise, i.e., the flow cell which is electronic, mechanical and thermal noise which leads to a reduction in sensitivity [53]. The pressure fluctuations refractive indices changes due to the BPR of the SFC [54], and the opening and closing of the BPR valve can result in the formation of pressure pulses within the system, which were observed in Figures 1, 3, and 4.

#### 3.2.1 SFC column optimisation of the tetrapeptide

The factors that influence the selection of the optimum separation and resolution of peptides are the number and the overall distribution of charged groups of a particular peptide, which would affect the hydrophobicity, ionisation, and polarisability of the peptide [55]. The tetrapeptide was tested on the aforementioned SFC columns (refer to Figure 1 A-D and Table 3). The peak shapes were relatively good on the PFP, and the diol HILIC; however, the LC-MS results indicated that the tetrapeptide contained a mixture (tetrapeptide and the same impurity from the crude sample) post-SFC purification, therefore, indicating that there may be only partial purification occurring (Figure S10).

The tetrapeptide was also separated on the 2’ ethyl pyridine 100 Å column yielding the best retention factor (5.51) from the three stated columns with good peak shapes, minimal baseline noise, and no peak splitting (refer to table 3). Thereafter, the 2’ethyl pyridine column 300 Å (Figure 1-C) was selected for further analysis to evaluate the effect of pore size of the stationary phase on the purification of the tetrapeptide, refer to Figure 2 and Table 3, post-SFC and was confirmed with LC-MS analysis (Figures S11 and S12). As can be seen from both the SFC and LC-MS chromatograms, the purification was much more efficient with the macro-porous stationary phase.

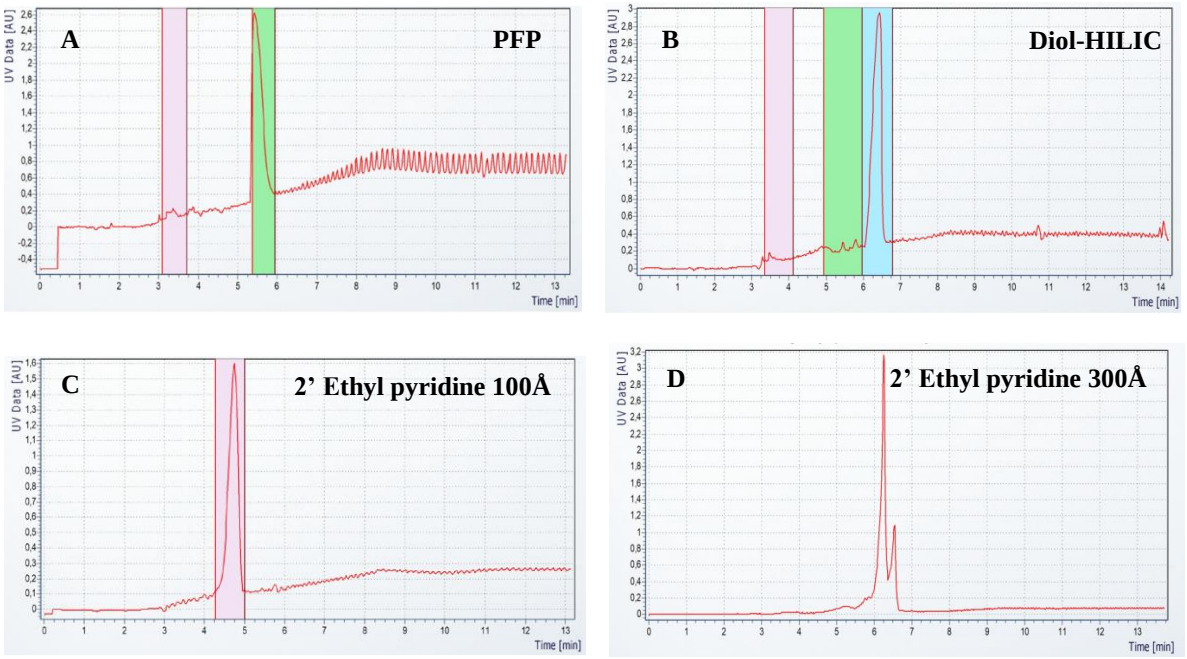


**Figure 1:** Represents 50 µL injections of SFC chromatograms of tetra-peptide (1 mg/mL) on the following columns: **A**: PFP (250 mm × 4.6 mm), **B**: diol-HILIC (250 mm × 4.6 mm), **C**: 2’ ethyl pyridine (250 mm × 4.6 mm), **D**: 2’ ethyl pyridine (250 mm × 10 mm). All columns were tested on a gradient mode with a modifier range of 5-60%.

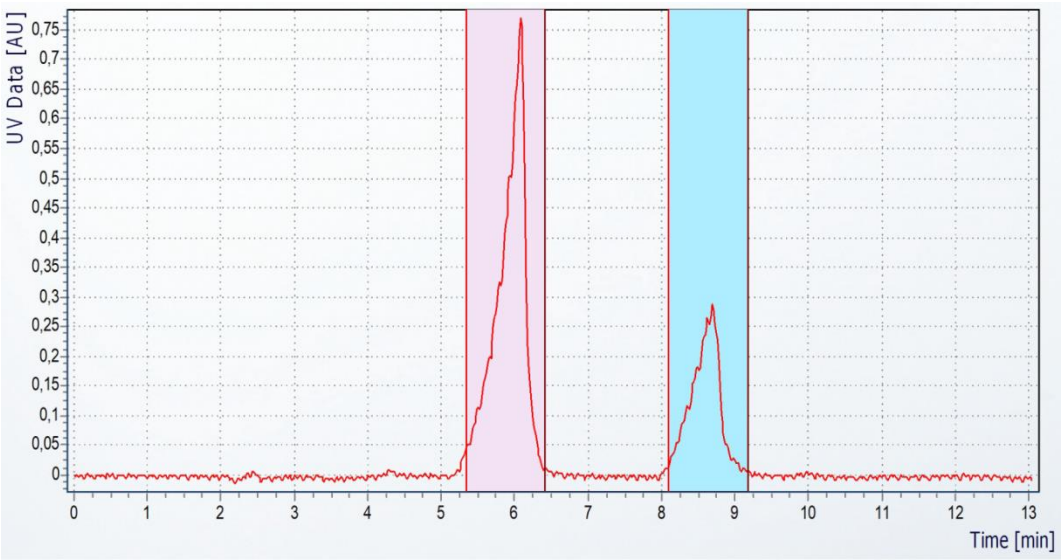


**Figure 2:** Represents a 50 µL injection of a SFC chromatogram of the tetrapeptide (1 mg/mL) tested on a 2' ethyl pyridine column-based column with the following characteristics; dimensions: 250 mm× 10 mm, pore size: 300 Å and a particle size of 5 µm. It was conducted on an isocratic mode with a modifier range of 20%.

### 3.2.2 SFC column optimisation of the octapeptide

The crude octapeptide was assessed on the aforementioned SFC columns; (refer to Figure 3 A-D and Table 4). The octapeptide was unretained on the PFP column (Figure 3-A). The PFP column is comprised of a pentafluoro phenyl group bonded phase, which contained fluorine atoms that have a partial negative charge, resulting in significant ionic, π-π, and dipole-dipole interactions [56, 57]. This could be attributed to the fact that peptides and proteins comprise of multiple amino acids, which consist of basic and acidic side chains, which give the protein a significant amount of charge [58]. The diol-HILIC column displayed poor peak shapes and noise on the baselines (Figure 3-B). This column is based on layered organic/inorganic hybrid silica and synthesised with a dihydroxy propyl group. The composition of the diol-HILIC column is silica molecules joined to oxygen atoms, resulting in siloxane bond formations. The remaining unreacted hydroxyl moieties are known as silanols [59]. These silanol groups are acidic; as a result, the diol-HILIC column can exhibit cation exchange properties [58]. With the diol-HILIC column, secondary interactions can occur between the peptides and the stationary phase packing material; this phenomenon is referred to as adsorption effects, which contributed to the results obtained in our study [60-64]. The LC-MS results post-SFC indicated the presence of a mixture which comprised of the octapeptide and the same impurities from the crude sample (Figure S13).

Initially, the 2’ ethyl pyridine 100 Å column was employed for the octapeptide, which displayed good peak shapes, minimal noise and no peak splitting and this phase was deemed suitable for further optimisations (Figure 3-C). We therefore proceeded with evaluating the macro-porous version of this column. The 2’ ethyl pyridine 300 Å separated the octapeptide with better peak shapes and retention factors of 2.63 and therefore warranted further use (Refer to Figures 3-D, S14, S15 and Table 4).

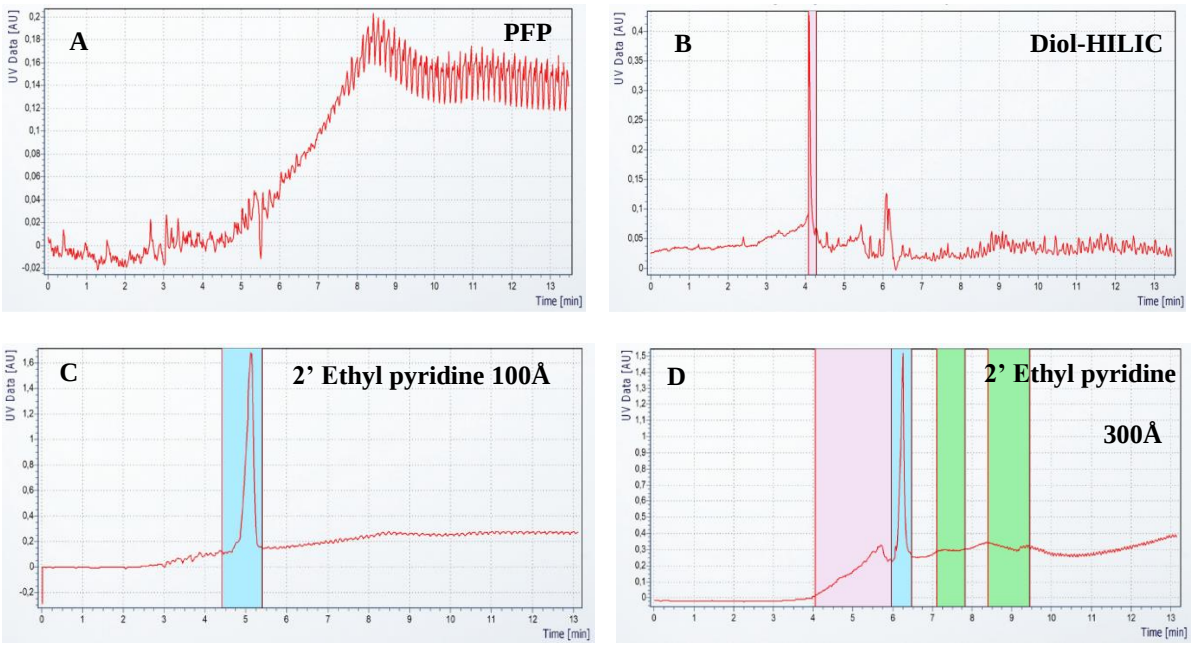


**Figure 3:** Represents 50 µL injections of SFC chromatograms of the octapeptide (2 mg/mL) on the following columns: **A**: PFP (250 mm × 4.6 mm), **B**: diol-HILIC (250 mm × 4.6 mm), **C**: 2’ ethyl pyridine (250 mm × 4.6 mm), and **D**: 2’ ethyl pyridine (250 mm × 10 mm). All columns were tested on a gradient mode with a modifier range of 5-60%.

### 3.2.3 SFC column optimisation of the nonapeptide

The crude nonapeptide underwent a similar evaluation on the aforementioned SFC columns (refer to Figures 4 A-D, S16-19). The PFP and diol-HILIC columns' separations were inadequate with peak splitting, poor peak resolution, and partial separations (refer to Figures 4 and Table 5). Amongst the microporous columns, the 2’ethyl pyridine column 100 Å (Figure 4-C) displayed the best retention factors and peak shapes with minimal noise (refer to table 5). Therefore, we proceeded to assess the effectiveness of the macro-porous 2’ ethyl pyridine 300 Å column (Figure 4-D), which displayed even better peak shapes and peak resolution with a retention factor of 2.92, than its microporous counterpart.

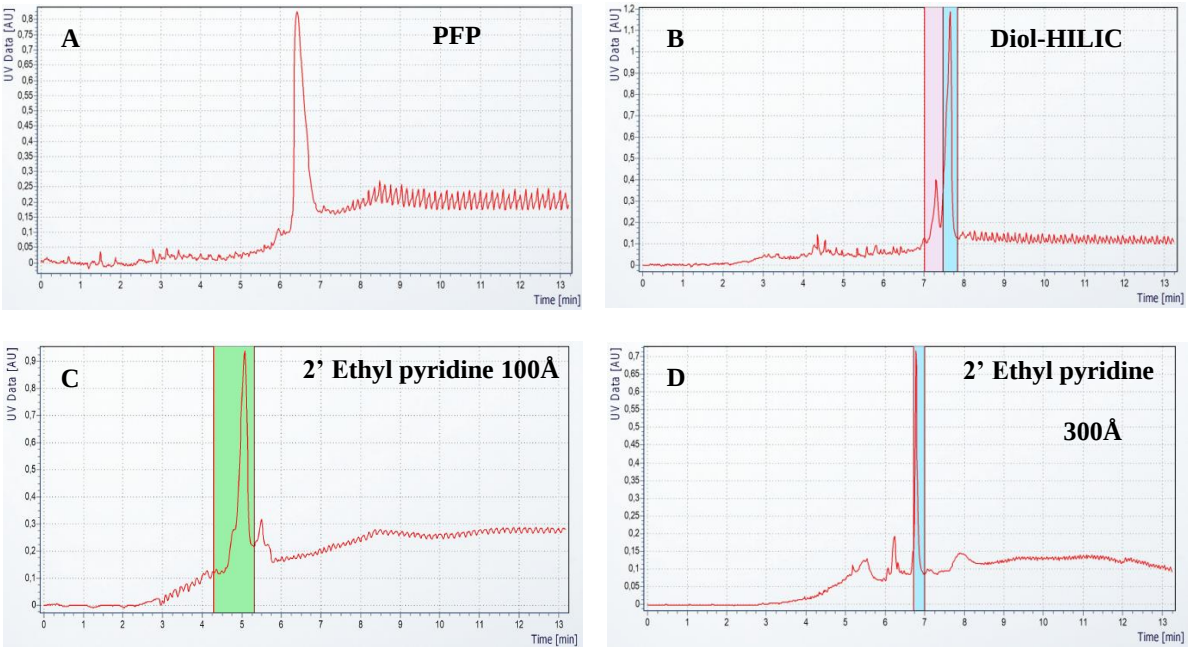


**Figure 4:** Represents 50 µL injections of SFC chromatograms of nonapeptide (1.5 mg/mL) on the following columns: **A**: PFP (250 mm × 4.6 mm), **B**: diol-HILIC (250 mm × 4.6 mm), **C**: 2' ethyl pyridine (250 mm × 4.6 mm), and **D**: 2' ethyl pyridine (250 mm × 10 mm). All columns were conducted on a gradient mode with a modifier range of 5-60%.

**Table 3**: Retention factors (k) of the tetrapeptide (1 mg/mL) for the 2' ethyl pyridine, diol-HILIC, and PFP columns.

| Column | k1 | k2 | SFC conditions | Assessment of column |
|---|---|---|---|---|
| PFP | 3.45 | 6.53 | 5-60% gradient | Good peak shapes, partial purification, the purified sample contained a mixture. |
| diol-HILIC | 3.79 | 6.53 | 5-60% gradient | Good peak shapes, however, the purified sample contained a mixture. |
| 2' ethyl pyridine 100 Å | 5.51 | - | 5-60% gradient | Good peak shapes, minimal baseline noise, and no peak splitting. |
| 2' ethyl pyridine 300 Å | 2.49 | 2.78 | 5-60% gradient | Poor peak shapes and peak resolution, however, the peptide separation occurred. |
| 2' ethyl pyridine 300 Å | 2.49 | 3.83 | 20% isocratic | Good peak shapes and resolution. |

**Table 4**: Retention factors (k) of the octapeptide (2 mg/mL) for the 2’ ethyl pyridine, diol-HILIC, and PFP columns conducted at 5-60% gradient conditions.

| Column | k1 | k2 | Assessment of column |
|---|---|---|---|
| PFP | $N_r$ | $N_r$ | Octapeptide was not retained on the column. |
| diol-HILIC | 4.66 | 7.22 | Noise in the baseline, poor peak shapes, and peak splitting. Partial purification as a mixture was present after SFC purification. |
| 2’ ethyl pyridine 100 Å | 5.85 | - | Good peak shapes, minimal baseline noise, and no peak splitting |
| 2’ ethyl pyridine 300 Å | 2.34 | 2.63 | Good peak shapes. Separation of the octapeptide was achieved on this column. |

$N_r$ denotes a compound that was not retained on the column.

**Table** :5 Retention factors (k) of the nonapeptide (1.5 mg/mL) for the 2’ ethyl pyridine, diol-HILIC, and PFP columns conducted at 5-60% gradient conditions.

| Column | k1 | k2 | k3 | k4 | Assessment of column |
|---|---|---|---|---|---|
| PFP | 7.90 | - | - | - | Good peak shape, however, there was only one main peak, and there was noise in the baseline. |
| diol-HILIC | 9.11 | 9.62 | - | - | Peak splitting and poor peak resolution. |
| 2’ ethyl pyridine 100 Å | 5.51 | 6.53 | - | - | Good peak shapes, poor peak resolution, minimal noise, and good retention factors. |
| 2’ ethyl pyridine 300 Å | 2.20 | 2.63 | 2.92 | 3.65 | Separation of the nonapeptide occurred with good peak shapes and peak resolution. Best retention factors for the peak of interest of 2.92. |

### 3.3 SFC recoveries conducted on the 2’ ethyl pyridine column

The 2’ ethyl pyridine column was previously shown to be successful in the SFC separation of peptides such as leucine encephalin acetate as well as methionine encephalin acetate, which was reported by Tognatelli *et al*. [25] and proprietary macrocyclic peptides by Ventura *et al*.[36].

In our study, the tetrapeptide, octapeptide, and the nonapeptide confirmed this (Figures 1-C, 3-C, and 4-C). It is not uncommon that macromolecules such as peptides require wider pore columns to ensure the internal surface is accessible to the analyte (pore diameters usually ≥ 300 Å) [65]. Macro-porous columns range from approximately 100 to 4000 Å [66]. These columns also allow for better scale-up to be conducted as they enable the use of higher flow rates and throughputs, as well as the ability to process larger sample volumes, thereby increasing the efficiency of peptide separations. The separation of the tetrapeptide [insulin β chain peptide (15-18)], octapeptide [angiotensin II], nonapeptide [insulin β chain peptide (15-23)] were reproducible on the 2' ethyl pyridine column with relative standard deviation values of 0-0.02% (see Table 6). The standard deviation and relative standard deviation indicated good precision of the method [67]. Therefore, the 2'ethyl pyridine column (300 Å) was selected for further evaluation whereby scale-up peptide recovery tests were conducted for the three crude peptides (refer to Table 7, Figures S20, S21, and S22). The nonapeptide yielded the best recovery of 102%, followed by octapeptide with 100%, and the tetrapeptide with 80%. The possible reason for the difference in the tetrapeptide recovery from the other two peptides is that the tetrapeptide contained more impurities (31.65%) as opposed to approximately 7% impurities for both the octapeptide and the nonapeptide. These results correlated to the impurities observed in LC-MS results (section 3.1).

**Table 6**: Reproducibility of retention times (n=3) for the separation of a tetrapeptide (1 mg/mL) [insulin β chain peptide (15-18)], octapeptide (2 mg/mL) [angiotensin II], nonapeptide (1.5 mg/mL) [insulin β chain peptide (15-23)]. Conducted on a 2' ethyl pyridine column (250 mm × 10 mm). Conducted on a gradient mode with a modifier range of 5-60%.

| Name of peptide | Average (minutes) | Standard deviation (minutes) | Relative standard deviation (%) |
|---|---|---|---|
| Tetrapeptide | 5.83 | 0,14 | 0.02 |
| Octapeptide | 6.25 | 0 | 0 |
| Nonapeptide | 6.75 | 0 | 0 |

**Table 7**: Represents the recoveries of the tetrapeptide, octapeptide, and the nonapeptide post-SFC purification. Peptide recoveries were conducted on the SFC using 5 mg samples (whereby n=1).

| Peptide | Peptide length | Weight before SFC (mg) | Weight after SFC (mg) | Recovery (%) |
|---|---|---|---|---|
| **Insulin β chain peptide (15-18)** | **Tetrapeptide (4)** | **5** | **4** | **80** |
| **Angiotensin II** | **Octapeptide (8)** | **5** | **5** | **100** |
| **Insulin β chain peptide (15-23)** | **Nonapeptide (9)** | **5** | **5.1** | **102** |

Interestingly, this study had an overall run time of 13 minutes generating 29.24 mLs of waste per run as opposed to the conventional analytical RP-HPLC methods of 50 minutes generating approximately 60 mLs of waste per run [25]. This reduction in run time analysis has implications regarding the generation of less waste [68] and therefore decreases the impact on the environment.

## 4 Conclusion

The SFC purification of the tetrapeptide (LYLV), octapeptide (DRVYIHPF), and nonapeptide (LYLVCGERG) on three different phases of SFC columns (PFP, diol-HILIC, and the 2' ethyl pyridine) were assessed on an analytical scale. The purity of the three peptides were detected and monitored using LC-MS analyses. The 2'ethyl pyridine stationary phase proved to be optimal for these peptides as it displayed good peak shapes and resolution. With this stationary phase, the SFC purification methods were reproducible and precise for all three peptides, with observed low relative standard deviations ranging from 0-0.02%. The recoveries ranged from 80-102%. It should be noted that traditional methods of peptide purification are approximately 50 minutes, while the SFC approach decreased the run times to approximately 13 minutes. There was a considerable decrease in the amount of solvents employed, making it a greener process. Since there is limited information in current literature regarding SFC purification of peptides and proteins, the present study provided an imperative method development strategy. The present study included the use of water and TFA in the $CO_2$/methanol modifier as well as eliminating the use of toxic acetonitrile and therefore serves as an important proof-of-concept starting point for other peptides. Future studies can focus on the use of SFC purifications on a larger preparative scale, investigate higher molecular weight peptides/proteins, vary the mobile phases, and consider the use of customised columns.

## 5 Acknowledgments

This study was made possible through financial support from the University of KwaZulu-Natal, National Research Foundation (NRF) of South Africa and Technology Innovation Agency (TIA) of South Africa.

## 6 Conflict of interest

The authors declare no conflict of interest.

## 7 Authors contribution

Kamini Govender conducted the experimental work, analysed all the experimental data and composed manuscript including the supporting information, which formed the basis of her Ph.D. study, the rest of the authors are co-supervisors on the project and contributed to the conceptualisation of the idea, funding of the project and scientific guidance.

## 8 References


1. Lau, J.L. and M.K. Dunn, *Therapeutic peptides: Historical perspectives, current development trends, and future directions.* Bioorganic and Medicinal Chemistry, 2018. **26**(10): p. 2700-2707.
2. Thompson, J.W., *Opioid peptides.* British Medical Journal (Clinical Research Ed.), 1984. **288**(6413): p. 259.
3. Otsuka, M. and S. Konishi, *Substance P—the first peptide neurotransmitter?* Trends in neurosciences, 1983. **6**: p. 317-320.
4. Di, L., *Strategic approaches to optimizing peptide ADME properties.* The AAPS journal, 2015. **17**(1): p. 134-143.
5. Sood, A. and R. Panchagnula, *Peroral route: an opportunity for protein and peptide drug delivery.* Chemical Reviews, 2001. **101**(11): p. 3275-3304.
6. Albericio, F. and H.G. Kruger, *Therapeutic peptides.* Future medicinal chemistry, 2012. **4**(12): p. 1527-1531.
7. Da Costa, J.P., M. Cova, R. Ferreira, and R. Vitorino, *Antimicrobial peptides: an alternative for innovative medicines?* Applied Microbiology and Biotechnology, 2015. **99**(5): p. 2023-2040.
8. Eckert, R., *Road to clinical efficacy: challenges and novel strategies for antimicrobial peptide development.* Future microbiology, 2011. **6**(6): p. 635-651.
9. Zompra, A.A., A.S. Galanis, O. Werbitzky, and F. Albericio, *Manufacturing peptides as active pharmaceutical ingredients.* Future medicinal chemistry, 2009. **1**(2): p. 361-377.
10. Agyei, D., C.M. Ongkudon, C.Y. Wei, A.S. Chan, and M.K. Danquah, *Bioprocess challenges to the isolation and purification of bioactive peptides.* Food and Bioproducts Processing, 2016. **98**: p. 244-256.
11. Daliri, E., D. Oh, and B. Lee, *Bioactive peptides.* Foods, 2017. **6**(5): p. 32.

12. Ayoub, M. and D. Scheidegger, *Peptide drugs, overcoming the challenges, a growing business.* Chimica oggi, 2006. **24**(4): p. 46.
13. Sun, L., *Peptide-based drug development.* Mod Chem Appl, 2013. **1**(1): p. 1-2.
14. Kumar, A., Y.E. Jad, A. El-Faham, G. Beatriz, and F. Albericio, *Green solid-phase peptide synthesis 4. γ-Valerolactone and N-formylmorpholine as green solvents for solid phase peptide synthesis.* Tetrahedron Letters, 2017. **58**(30): p. 2986-2988.
15. Hruby, V.J., *Design of cyclic peptides with biological activities from biologically active peptides: the case of peptide modulators of melanocortin receptors.* Peptide Science, 2016. **106**(6): p. 884-888.
16. Schiavone, N.M., R. Bennett, M.B. Hicks, G.F. Pirrone, E.L. Regalado, I. Mangion, and A.A. Makarov, *Evaluation of global conformational changes in peptides and proteins following purification by supercritical fluid chromatography.* Journal of Chromatography B, 2019. **1110**: p. 94-100.
17. Conlon, J.M., *Purification of naturally occurring peptides by reversed-phase HPLC.* Nature protocols, 2007. **2**(1): p. 191.
18. Piot, J., Q. Zhao, D. Guillochon, P. Dhulster, G. Ricart, and D. Thomas, *Semi-preparative purification and characterization of peptides from complex haemoglobin hydrolysate by HPLC-mass spectrometry.* Chromatographia, 1990. **30**(3-4): p. 205-210.
19. Aguilar, M., *HPLC of peptides and proteins: basic theory and methodology.* Methods in molecular biology (Clifton, NJ), 2004. **251**: p. 3.
20. Ausubel Fred, M., B. Roger, E. Kingston Robert, and D. Moore David, *Current Protocols in Molecular Biology.* Current Protocols in Molecular Biology. John Wiley & Sons, 2003.
21. Lemieux, L., J.-M. Piot, D. Guillochon, and J. Amiot, *Study of the efficiency of a mobile phase used in size-exclusion HPLC for the separation of peptides from a casein hydrolysate according to their hydrodynamic volume.* Chromatographia, 1991. **32**(11-12): p. 499-504.
22. Shukla, K.K., S.B. Badgujar, P.B. Bhanushali, and S.G. Sabharwal, *Simplified purification approach of urinary neutrophil gelatinase-associated lipocalin by tangential flow filtration and ion exchange chromatography.* Journal of Chromatography B, 2017. **1051**: p. 68-74.
23. Eschelbach, J.W. and J.W. Jorgenson, *Improved protein recovery in reversed-phase liquid chromatography by the use of ultrahigh pressures.* Analytical chemistry, 2006. **78**(5): p. 1697-1706.
24. Roumeliotis, P. and K. Unger, *Assessment and optimization of system parameters in size exclusion separation of proteins on diol-modified silica columns.* Journal of Chromatography A, 1981. **218**: p. 535-546.
25. Tognarelli, D., A. Tsukamoto, J. Caldwell, and W. Caldwell, *Rapid peptide separation by supercritical fluid chromatography.* Bioanalysis, 2010. **2**(1): p. 5-7.

26. Welch, C.J., T. Brkovic, W. Schafer, and X. Gong. *Performance to burn? Re-evaluating the choice of acetonitrile as the platform solvent for analytical HPLC*. Green Chemistry 2009 [cited 11 8]; 1232-1238].
27. Issaq, H.J., K.C. Chan, G.M. Janini, T.P. Conrads, and T.D. Veenstra, *Multidimensional separation of peptides for effective proteomic analysis.* Journal of Chromatography B, 2005. **817**(1): p. 35-47.
28. McConvey, I.F., D. Woods, M. Lewis, Q. Gan, and P. Nancarrow, *The importance of acetonitrile in the pharmaceutical industry and opportunities for its recovery from waste.* Organic Process Research & Development, 2012. **16**(4): p. 612-624.
29. Sheldon, R.A., *The greening of solvents: Towards sustainable organic synthesis.* Current Opinion in Green and Sustainable Chemistry, 2019. **18**: p. 13-19.
30. Li, J., J. Albrecht, A. Borovika, and M.D. Eastgate, *Evolving green chemistry metrics into predictive tools for decision making and benchmarking analytics.* ACS Sustainable Chemistry & Engineering, 2017. **6**(1): p. 1121-1132.
31. Welch, C.J., T. Nowak, L.A. Joyce, and E.L. Regalado, *Cocktail chromatography: enabling the migration of HPLC to nonlaboratory environments.* ACS Sustainable Chemistry & Engineering, 2015. **3**(5): p. 1000-1009.
32. Miller, L.M., J.D. Pinkston, and L.T. Taylor, *Modern Supercritical Fluid Chromatography: Carbon Dioxide Containing Mobile Phases*. 2019: John Wiley & Sons.
33. Govender, K., T. Naicker, S. Baijnath, A.A. Chuturgoon, N.S. Abdul, T. Docrat, H.G. Kruger, and T. Govender, *Sub/supercritical fluid chromatography employing water-rich modifier enables the purification of biosynthesized human insulin.* Journal of Chromatography B, 2020, Accepted.
34. Barhate, C.L., L.A. Joyce, A.A. Makarov, K. Zawatzky, F. Bernardoni, W.A. Schafer, D.W. Armstrong, C.J. Welch, and E.L. Regalado, *Ultrafast chiral separations for high throughput enantiopurity analysis.* Chemical Communications, 2017. **53**(3): p. 509-512.
35. Liu, J., A.A. Makarov, R. Bennett, I.A. Haidar Ahmad, J. DaSilva, M. Reibarkh, I. Mangion, B.F. Mann, and E.L. Regalado, *Chaotropic Effects in Sub/Supercritical Fluid Chromatography via Ammonium Hydroxide in Water-Rich Modifiers: Enabling Separation of Peptides and Highly Polar Pharmaceuticals at the Preparative Scale.* Analytical chemistry, 2019. **91**(21): p. 13907-13915.
36. Ventura, M., *Advantageous use of SFC for separation of crude therapeutic peptides and peptide libraries.* Journal of pharmaceutical and biomedical analysis, 2020: p. 113227.
37. Saito, M., *History of supercritical fluid chromatography: instrumental development.* Journal of bioscience and bioengineering, 2013. **115**(6): p. 590-599.

38. Płotka, J.M., M. Biziuk, C. Morrison, and J. Namieśnik, *Pharmaceutical and forensic drug applications of chiral supercritical fluid chromatography.* TrAC Trends in Analytical Chemistry, 2014. **56**: p. 74-89.
39. Miller, L., *Evaluation of non-traditional modifiers for analytical and preparative enantioseparations using supercritical fluid chromatography.* Journal of Chromatography A, 2012. **1256**: p. 261-266.
40. Miller, L., *Use of dichloromethane for preparative supercritical fluid chromatographic enantioseparations.* Journal of Chromatography A, 2014. **1363**: p. 323-330.
41. Speybrouck, D. and E. Lipka, *Preparative supercritical fluid chromatography: A powerful tool for chiral separations.* Journal of Chromatography A, 2016. **1467**: p. 33-55.
42. Nogle, L.M., C.W. Mann, W.L. Watts Jr, and Y. Zhang, *Preparative separation and identification of derivatized β-methylphenylalanine enantiomers by chiral SFC, HPLC and NMR for development of new peptide ligand mimetics in drug discovery.* Journal of pharmaceutical and biomedical analysis, 2006. **40**(4): p. 901-909.
43. Miller, L. and M. Potter, *Preparative chromatographic resolution of racemates using HPLC and SFC in a pharmaceutical discovery environment.* Journal of Chromatography B, 2008. **875**(1): p. 230-236.
44. DaSilva, J.O., D. Lehnherr, J. Liu, R. Bennett, I.A. Haidar Ahmad, M. Hicks, B.F. Mann, D.A. DiRocco, and E.L. Regalado, *Generic Enhanced Sub/Supercritical Fluid Chromatography: The Blueprint for Highly Productive and Sustainable Separation of Primary Hindered Amines.* ACS Sustainable Chemistry & Engineering, 2020.
45. Tarafder, A., *Theories for Preparative SFC*, in *Supercritical Fluid Chromatography*. 2017, Elsevier. p. 245-274.
46. Roy, D., M.F. Wahab, T.A. Berger, and D.W. Armstrong, *Ramifications and Insights on the Role of Water in Chiral Sub/Supercritical Fluid Chromatography.* Analytical chemistry, 2019. **91**(22): p. 14672-14680.
47. Taylor, L.T., *Packed column supercritical fluid chromatography of hydrophilic analytes via water-rich modifiers.* Journal of Chromatography A, 2012. **1250**: p. 196-204.
48. Liu, J., E.L. Regalado, I. Mergelsberg, and C.J. Welch, *Extending the range of supercritical fluid chromatography by use of water-rich modifiers.* Organic & Biomolecular Chemistry, 2013. **11**(30): p. 4925-4929.
49. Rao, G.H., *Management of Diabetes Epidemic: A global perspective.* EC Endocrinol Metab Res, 2018. **3**(2): p. 63-72.
50. West, C., J. Melin, H. Ansouri, and M.M. Metogo, *Unravelling the effects of mobile phase additives in supercritical fluid chromatography. Part I: polarity and acidity of the mobile phase.* Journal of Chromatography A, 2017. **1492**: p. 136-143.

51. Bennett, R., M. Biba, J. Liu, I.A.H. Ahmad, M.B. Hicks, and E.L. Regalado, *Enhanced fluidity liquid chromatography: A guide to scaling up from analytical to preparative separations.* Journal of Chromatography A, 2019. **1595**: p. 190-198.
52. Stathopoulos, P., S. Papas, and V. Tsikaris, *C-terminal N-alkylated peptide amides resulting from the linker decomposition of the Rink amide resin. A new cleavage mixture prevents their formation.* Journal of peptide science: an official publication of the European Peptide Society, 2006. **12**(3): p. 227-232.
53. Helmy, R., M. Biba, J. Zang, B. Mao, K. Fogelman, V. Vlachos, P. Hosek, and C.J. Welch, *Improving sensitivity in chiral supercritical fluid chromatography for analysis of active pharmaceutical ingredients.* Chirality: The Pharmacological, Biological, and Chemical Consequences of Molecular Asymmetry, 2007. **19**(10): p. 787-792.
54. Berger, T.A. and B.K. Berger, *Minimizing UV noise in supercritical fluid chromatography. I. Improving back pressure regulator pressure noise.* Journal of Chromatography A, 2011. **1218**(16): p. 2320-2326.
55. Ausubel Fred, M., B. Roger, E. Kingston Robert, and D. Moore David, *Seidman JG, Smith John A., Struhl Kevin.* Current Protocols in Molecular Biology. John Wiley & Sons, 2003.
56. West, C., E. Lemasson, S. Khater, and E. Lesellier, *An attempt to estimate ionic interactions with phenyl and pentafluorophenyl stationary phases in supercritical fluid chromatography.* Journal of Chromatography A, 2015. **1412**: p. 126-138.
57. West, C. and E. Lemasson, *Unravelling the effects of mobile phase additives in supercritical fluid chromatography—Part II: Adsorption on the stationary phase.* Journal of Chromatography A, 2019.
58. Green, J.S. and J.W. Jorgenson, *Minimizing adsorption of proteins on fused silica in capillary zone electrophoresis by the addition of alkali metal salts to the buffers.* Journal of Chromatography A, 1989. **478**: p. 63-70.
59. Ghose, S., T.M. McNerney, and B. Hubbard, *Preparative protein purification on underivatized silica.* Biotechnology and Bioengineering, 2004. **87**(3): p. 413-423.
60. Mizutani, T. and A. Mizutani, *Estimation of adsorption of drugs and proteins on glass surfaces with controlled pore glass as a reference.* Journal of pharmaceutical sciences, 1978. **67**(8): p. 1102-1105.
61. Strazhesko, D.N., V.B. Strelko, V.N. Belyakov, and S.C. Rubanik, *Mechanism of cation exchange on silica gels.* Journal of Chromatography A, 1974. **102**: p. 191-195.
62. Schmidt, D.E., R.W. Giese, D. Conron, and B.L. Karger, *High performance liquid chromatography of proteins on a diol-bonded silica gel stationary phase.* Analytical chemistry, 1980. **52**(1): p. 177-182.

63. Pesek, J.J., M.T. Matyska, and L. Mauskar, *Separation of proteins and peptides by capillary electrochromatography in diol-and octadecyl-modified etched capillaries.* Journal of Chromatography A, 1997. **763**(1-2): p. 307-314.
64. Mori, S. and M. Kato, *High-performance aqueous size-exclusion chromatography using diol-bonded porous glass packing materials. Retention behavior of some proteins.* Journal of Liquid Chromatography, 1987. **10**(14): p. 3113-3126.
65. Heftmann, E., *Chromatography: Fundamentals and applications of chromatography and related differential migration methods-Part B: Applications*. 2004: Elsevier.
66. Unger, K.K., *Packings and stationary phases in chromatographic techniques.* Chromatographic Science Series, 1990. **47**.
67. Belouafa, S., F. Habti, S. Benhar, B. Belafkih, S. Tayane, S. Hamdouch, A. Bennamara, and A. Abourriche, *Statistical tools and approaches to validate analytical methods: methodology and practical examples.* International Journal of Metrology and Quality Engineering, 2017. **8**: p. 9.
68. González-Ruiz, V., A.I. Olives, and M.A. Martín, *Core-shell particles lead the way to renewing high-performance liquid chromatography.* TrAC Trends in Analytical Chemistry, 2015. **64**: p. 17-28.

# Chapter 5

## Summary and conclusion

# 5 Summary and conclusion

Protein and peptide biologics are imperative in the pharmaceutical industry as these molecules are multifaceted, having anti-cancer, anti-viral, anti-hypertensive, and anti-diabetic properties; these biologics can be utilised in the treatment of cancer, arthritis, diabetes, and cardiovascular diseases. The present study focused on human insulin and peptides associated with diabetes. Diabetes is a major non-communicable disease that occurs when the human body is incapable of effectively using insulin. The prevalence of diabetes is placing a substantial burden on our health care system and increases the demand for affordable insulin. This creates a massive burden on current insulin manufacturers. Consequently, there is a need for scientists to create more affordable methods of human insulin purification. The current RP-HPLC purification methods of insulin and peptides are problematic with low yields, and employ harsh mobile phases such as acetonitrile that are hazardous to the environment. There is a need for a greener process for the purification of these compounds. SFC can provide the solution to this problem since SFC purification is a green process since it utilises $CO_2$ and methanol as mobile phases, which can be recycled. However, there is limited information regarding the SFC purification of biosynthesised human insulin and peptides. Therefore, the aim of the present study endeavoured to develop a novel, more efficient approach to the biosynthesis of human insulin. Furthermore, this research study also aimed to purify biosynthesised human insulin and it also led to the purification of other peptides using SFC.

In chapter two of this thesis, a novel and more efficient method of human insulin biosynthesis was developed employing recombinant DNA technology such as polymerase chain reaction techniques, which were utilised for the cloning and verification of human insulin in *E. coli*. The present study quantified the biosynthesised human insulin using MALDI-TOF-MS as well as verified it using LC-MS and protein sequencing. Consequently, the protein sequencing results indicated that the biosynthesised human insulin's sequence was unequivocally proven to be human insulin since the protein sequencing was 100% similar to that of human insulin. The biological activity of the crude biosynthesised human insulin was evaluated *in vitro* utilising a MTT assay, and the data indicated that the crude biosynthesised human insulin displayed similar efficacy to that of the commercially available standard sample of human insulin. The biological activity results were highly significant, with a p-value value of $< 0.0001$. It should be noted that this study was successful in creating an innovative, novel, and more affordable strategy for the biosynthesis of human insulin in *E. coli*. Numerous downstream processing costs were eliminated, such as the use of affinity tags for purification, renaturation of the protein, enzymatic cleavage of the C peptide and tedious recovery steps associated with inclusion body formation. However, the current research study was conducted on a laboratory-scale basis; future studies can upscale the current work industrially.

In chapter three, a SFC purification method was developed whereby commercially available human insulin (standard) and crude biosynthesised human insulin were purified on the SFC. The following SFC columns were assessed using standard human insulin: silica, 2' ethyl pyridine, diol-HILIC, and the PFP, to determine the optimum for purification of the biosynthesised human insulin. The PFP column yielded the best results which displayed good peak shapes, resolution, retention times, retention factors, and the least relative standard deviation in comparison to the other columns. Since the aforementioned column was determined to be the best, a column efficiency test was conducted on a semi-preparative scale utilising the biosynthesised human insulin yielding an 84% recovery. Herein, the detection and quantification of the SFC purified biosynthesised human insulin was conducted using protein sequencing, and LC-MS analysis. To determine the effect of the SFC purification on the biosynthesised human insulin, a MTT assay was conducted; the results were highly significant with a p-value of $< 0.0001$. Interestingly, the *in vitro* MTT cytotoxicity results revealed that the biological activities of the biological activities of the standard and biosynthesised human insulin derivatives were indeed retained post-SFC purification.

In chapter three, numerous challenges were encountered during SFC purification of human insulin, such as the development of phase separation peaks, adsorption effects, band broadening, and the use of methanol with proteins on the SFC. Therefore, chapter four endeavoured to develop a SFC purification method for peptides at an analytical scale. A tetrapeptide [insulin β chain peptide (15-18)], octapeptide [angiotensin II], nonapeptide [insulin β chain peptide (15-23)] were purified using four SFC columns; the PFP, the diol-HILIC, and 2' ethyl pyridine. The respective peptides were monitored using LC-MS analysis. The 2' ethyl pyridine column was selected for further testing based on peak shapes, reproducibility, retention factors, and efficiency of SFC peptide separations of all three peptides. Thereafter, recovery tests for the three peptides were conducted on the 2' ethyl pyridine column, and the peptide recoveries were significant, ranging from 80-102%. The research results indicated that the present study was successful in the SFC purification of these peptides. As far as we are aware, this is one of the first studies to purify biosynthesised human insulin and this combination of peptides using SFC technology.

This innovative study can provide a suitable alternative to the traditional RP-HPLC method of purifying crude biosynthesised human insulin and peptides. The RP-HPLC purification method is time-consuming, with run times of approximately 50 minutes, and utilises hazardous solvents that are difficult to dispose of and have a negative impact on the environment. However, in the current research study, the SFC purification run times for the methanol soluble peptides was 13 minutes, and the biosynthesised human insulin was approximately 10 minutes. As such, the significant reduction in run time makes the SFC purification methods of human insulin and peptides less laborious, more efficient, and an overall greener process as opposed to the conventional RP-HPLC methods.

In conclusion, this innovative application of SFC can be employed as a key platform for the purification of peptides and proteins. Protein and peptide biologics are essential in the pharmaceutical industry as these biologics have a plethora of therapeutic properties.

Future research studies can use different modifiers, hydrophobic ion-pairing, or chemical modification of peptide and proteins to enhance solubility in non-aqueous solvents. Moreover, studies can also incorporate the use of custom-made columns to evaluate its effect on a range of proteins and peptides, thus paving the way forward to innovative next-generation isolation of novel and current biologics using SFC purification.

# Appendices

# Appendix 1 for Chapter 2

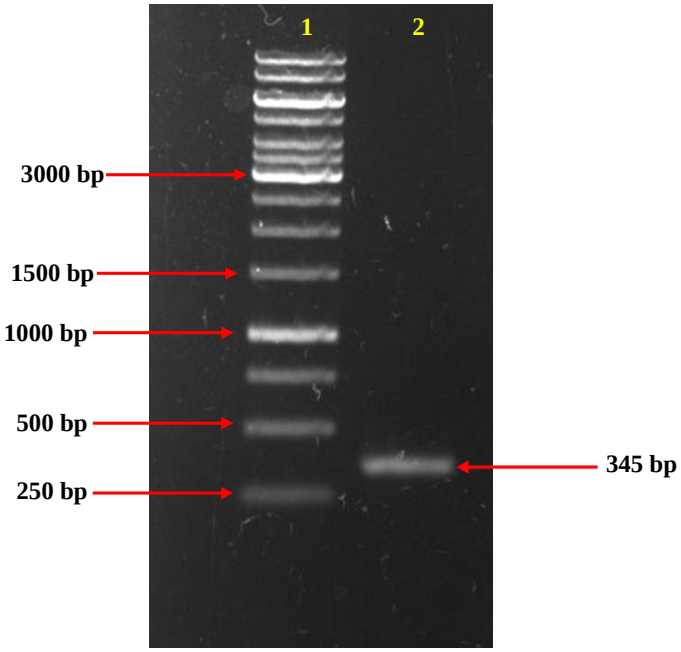


**Figure S1a:** An image depicting a purified PCR product of human proinsulin gene. Lane 1 contains the one kb molecular weight marker, and lane 2 contains the purified *Bam*HI and an *Xho*I ended PCR product of human proinsulin gene (345 bp).

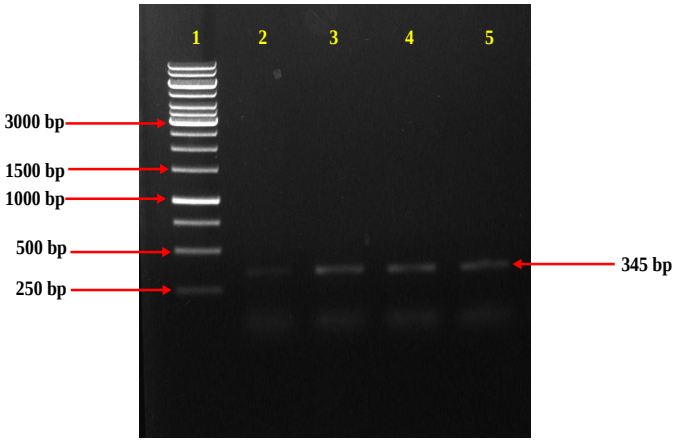


**Figure S1b:** An image displaying colony PCR products of positive transformants obtained from clones that contain the pET21b-hPin vector. Lane 1 contains the one kb molecular weight marker, and lanes 2-5 contain colony PCR amplicons (345 bp).

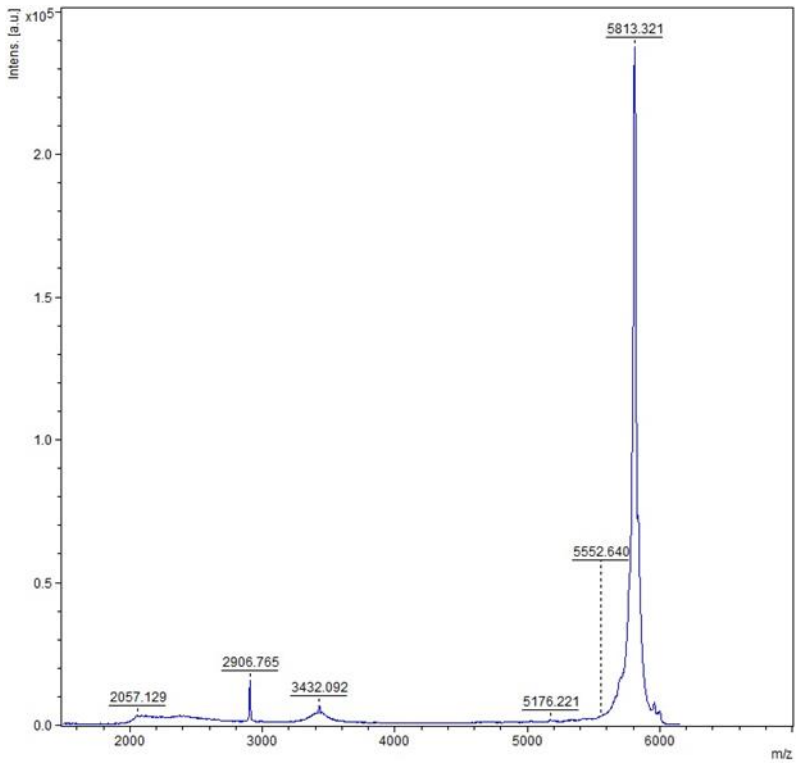


**Figure S2:** An image depicting a MALDI-TOF spectrum of standard human insulin as a positive control.

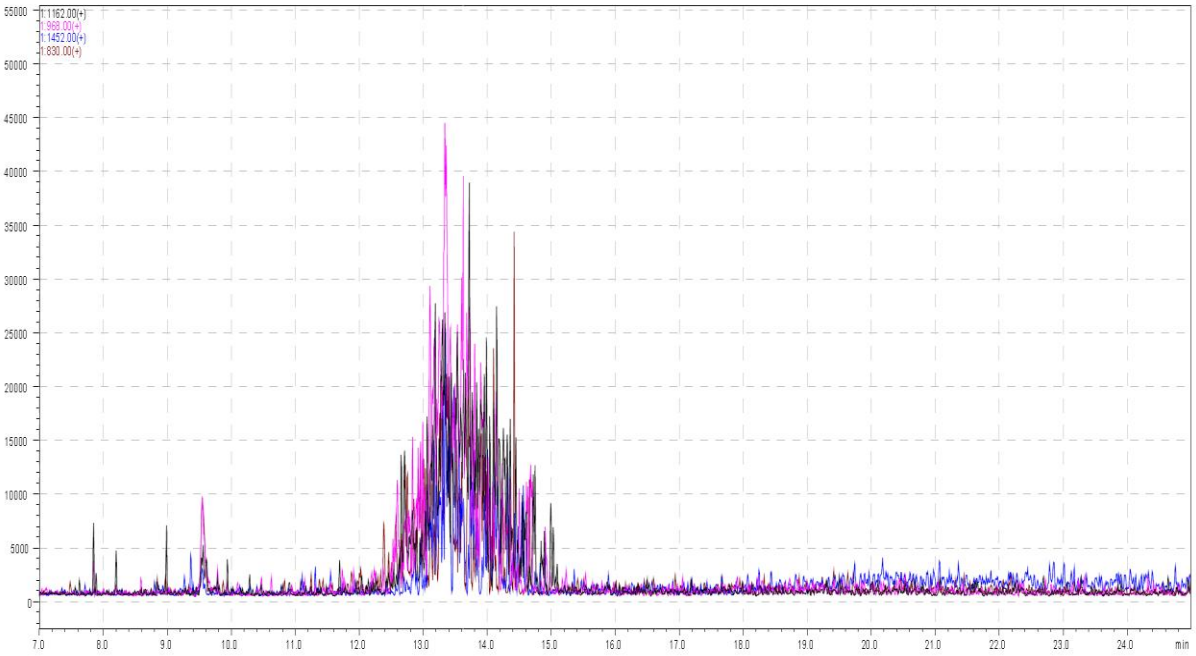


**Figure S3:** An LC-MS chromatogram illustrating 1 mM IPTG induced biosynthesised human insulin sample.

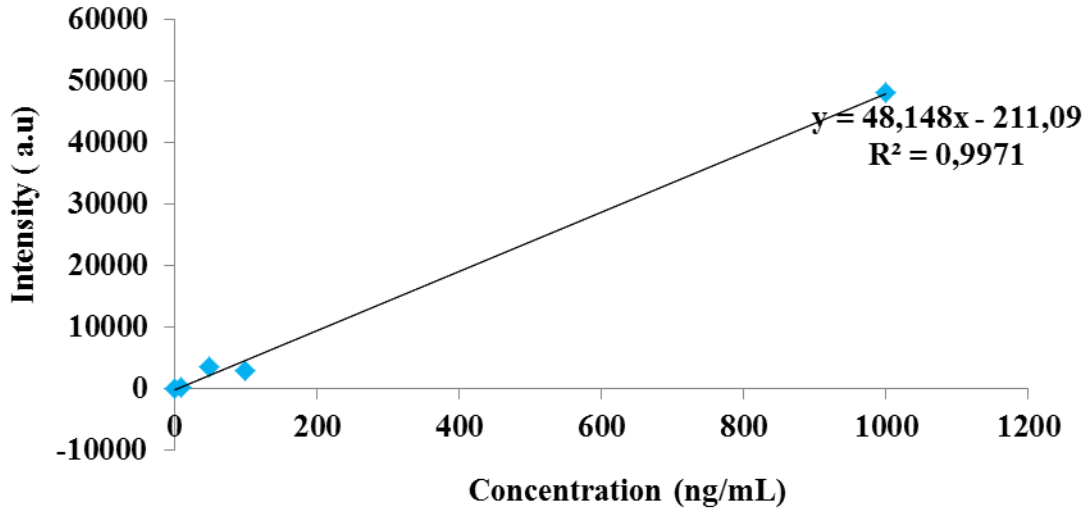


**Figure S4:** A MALDI-TOF standard curve of the insulin standard at 0 ng, 10 ng, 50 ng, 100 ng, and 1000 ng.

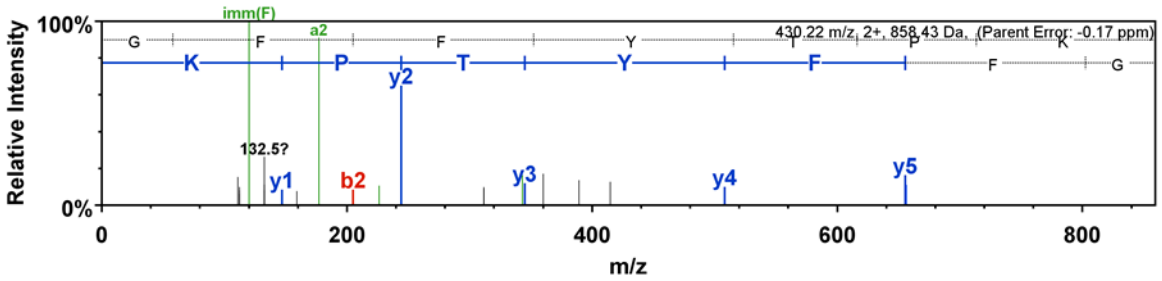


**Figure S5:** A peptide spectrum illustrating the protein sequence of standard human insulin, which was 90% similar to the human insulin sequence derived from Scaffold 1.4.4 software.

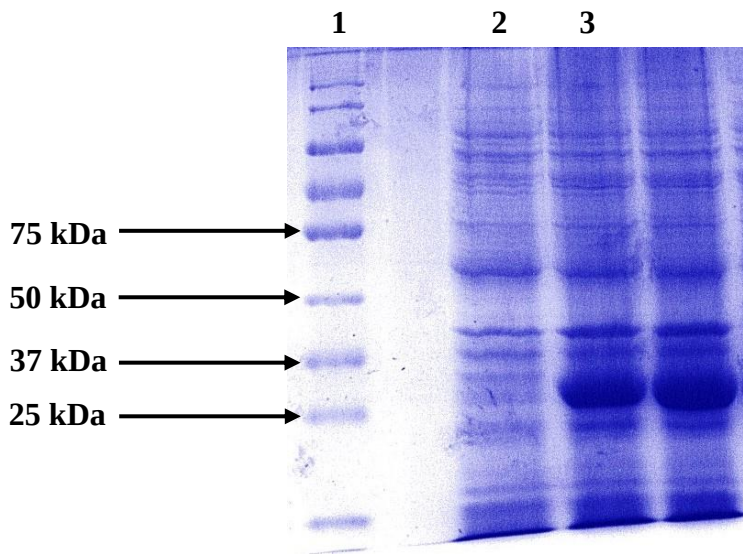


**Figure S6:** An SDS-page image displaying the over-expressed GST tagged-human insulin proteins in the wild type BL21 (DE3) strain. Lane 1 contains the PageRuler™ pre-stained protein ladder (Thermo-Scientific, USA). Lane 2 contains the un-induced wild type strain, and lane 3 contains the IPTG induced BL21 (DE3) transformant.

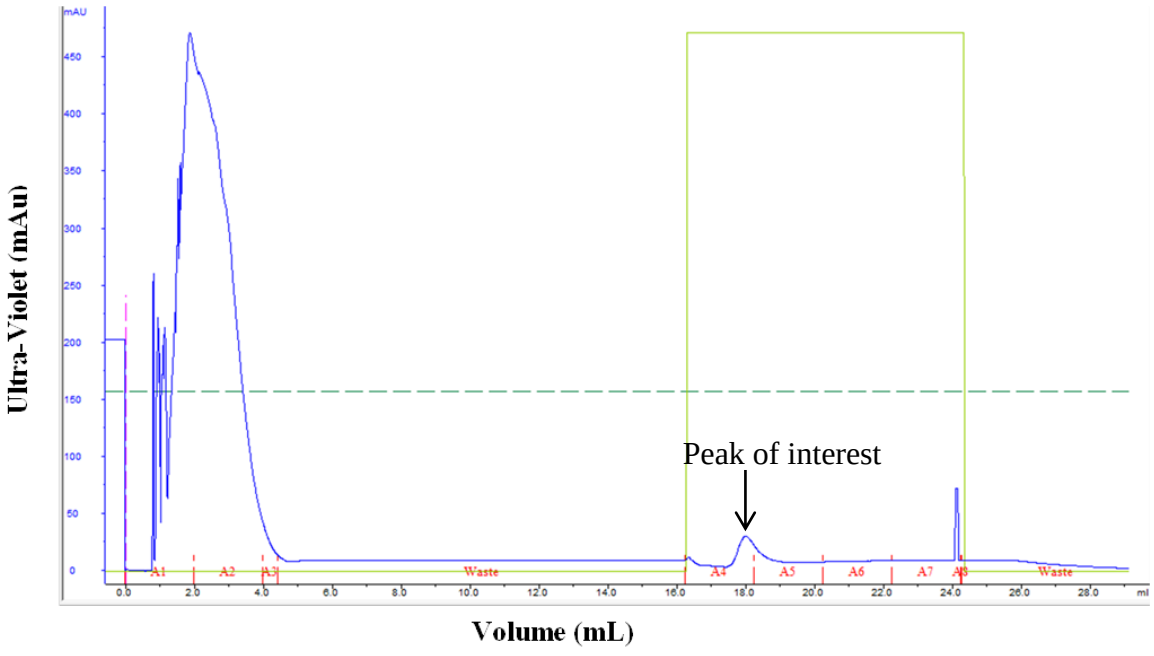


**Figure S7**: Illustrating an elution profile of a GST tagged human insulin obtained from the AKTA purifier 100-950.

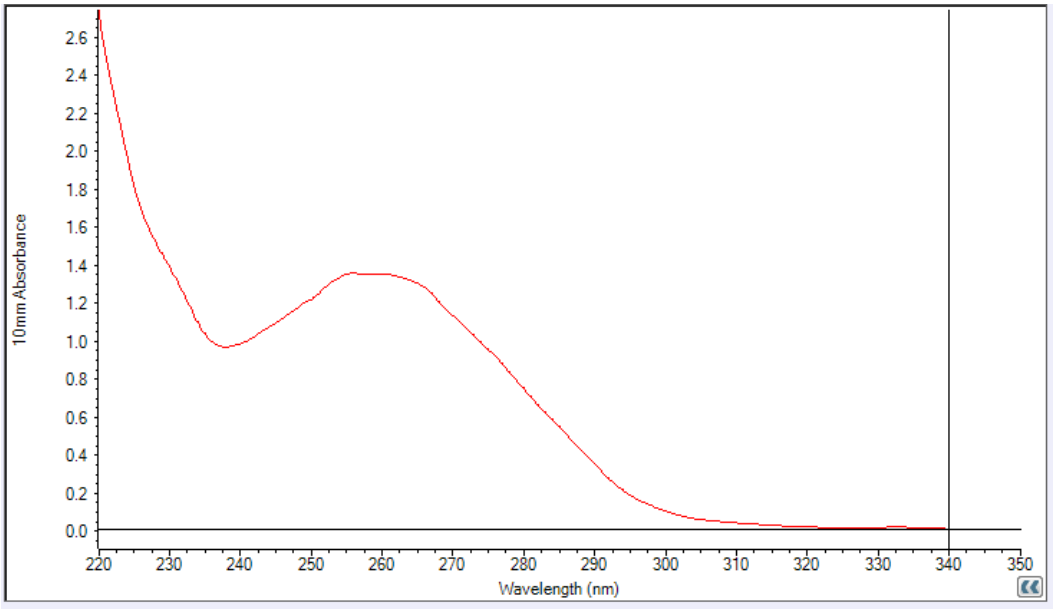


**Figure S8:** Biosynthesised human insulin protein UV spectrum (220-340 nm) obtained from NanoDrop™ 2000c.

**Table S1:** Peptide fragmentation table of crude biosynthesised human insulin, which yielded a 100% match to human insulin in the protein database.

| B | B Ions | B+2H | B-NH3 | B-H2O | AA | Y Ions | Y+2H | Y-NH3 | Y-H2O | Y |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 58.0 | 29.5 | | | G | 1,702.9 | 851.9 | 1,685.8 | 1,684.8 | 15 |
| 2 | 145.1 | 73.0 | | 127.1 | S | 1,645.8 | 823.4 | 1,628.8 | 1,627.8 | 14 |
| 3 | 282.1 | 141.6 | | 264.1 | H | 1,558.8 | 779.9 | 1,541.8 | 1,540.8 | 13 |
| 4 | 395.2 | 198.1 | | 377.2 | L | 1,421.7 | 711.4 | 1,404.7 | 1,403.7 | 12 |
| 5 | 494.3 | 247.6 | | 476.3 | V | 1,308.7 | 654.8 | 1,291.6 | 1,290.7 | 11 |
| 6 | 623.3 | 312.2 | | 605.3 | E | 1,209.6 | 605.3 | 1,192.6 | 1,191.6 | 10 |
| 7 | 694.4 | 347.7 | | 676.3 | A | 1,080.6 | 540.8 | 1,063.5 | 1,062.5 | 9 |
| 8 | 807.4 | 404.2 | | 789.4 | L | 1,009.5 | 505.3 | 992.5 | 991.5 | 8 |
| 9 | 970.5 | 485.8 | | 952.5 | Y | 896.4 | 448.7 | 879.4 | 878.4 | 7 |
| 10 | 1,083.6 | 542.3 | | 1,065.6 | L | 733.4 | 367.2 | 716.3 | 715.4 | 6 |
| 11 | 1,182.7 | 591.8 | | 1,164.6 | V | 620.3 | 310.6 | 603.3 | 602.3 | 5 |
| 12 | 1,342.7 | 671.8 | | 1,324.7 | C+57 | 521.2 | 261.1 | 504.2 | 503.2 | 4 |
| 13 | 1,399.7 | 700.4 | | 1,381.7 | G | 361.2 | 181.1 | 344.2 | 343.2 | 3 |
| 14 | 1,528.7 | 764.9 | | 1,510.7 | E | 304.2 | 152.6 | 287.1 | 286.2 | 2 |
| 15 | 1,702.9 | 851.9 | 1,685.8 | 1,684.8 | R | 175.1 | 88.1 | 158.1 | | 1 |

# Appendix 2 for Chapter 3

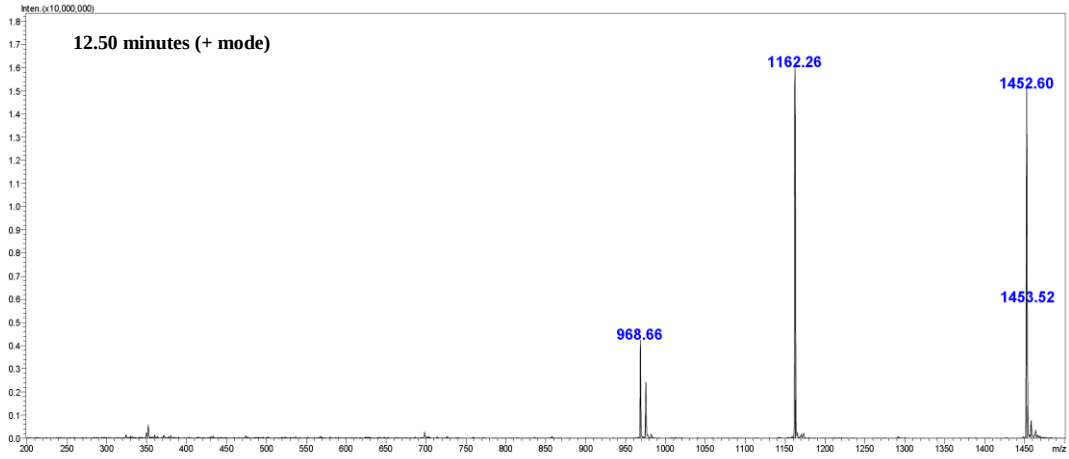


**Figure S1**: Represents a LC-MS spectrum obtained from a YMC-Triart $C_{18}$ (150 mm × 4.6 mm) of SFC purified standard sample of insulin (peak1).

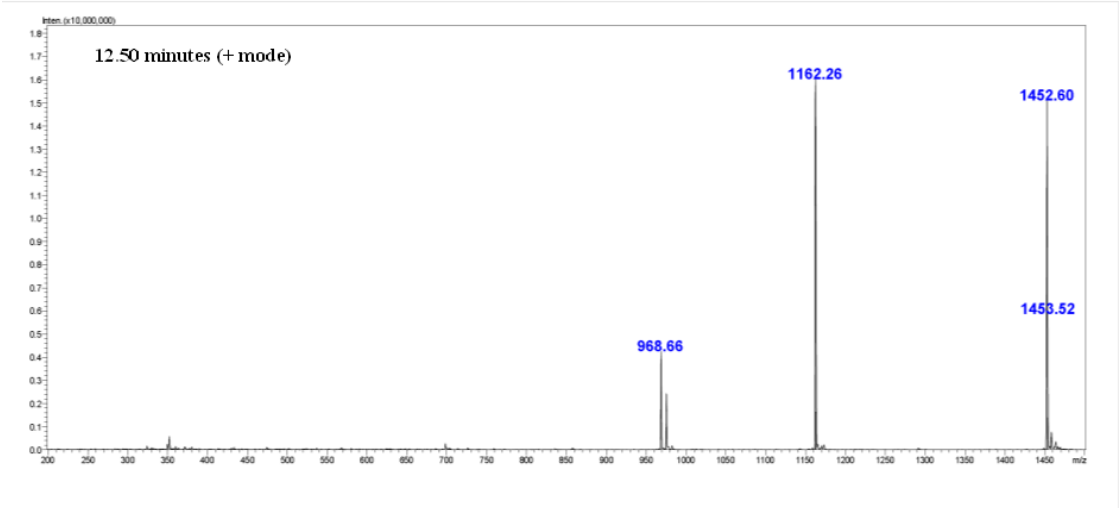


**Figure S2**: Represents a LC-MS spectrum obtained from a YMC-Triart $C_{18}$ (150 mm × 4.6 mm) of the SFC purified standard sample of insulin (peak2).

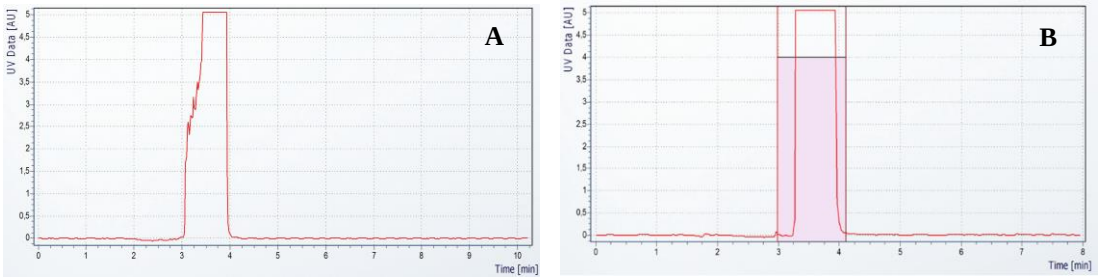


**Figure S3:** Represents SFC chromatograms of the A) standard sample of insulin (1 mg/mL) and B) biosynthesised sample of insulin (0.52 mg/mL), which were both dissolved in 50 mM Tris-HCl buffer and 0.2 % formic acid. A 50 µL injection was tested on a PFP based column (250 mm × 4.6 mm). The SFC separations were conducted using an isocratic mode with a 10% modifier.

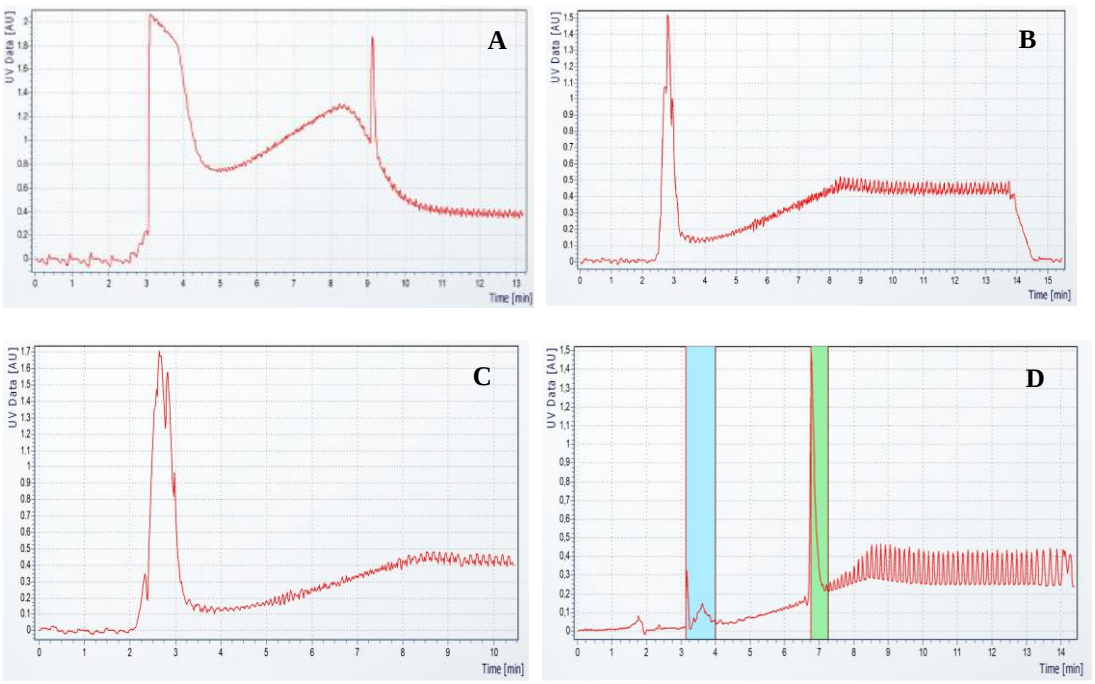


**Figure S4:** Represents SFC chromatograms of 50 µL injections of the following mixtures: A) DMSO and methanol (30:70), B) acetic acid and methanol (7:93), C) formic acid and methanol (6:94) and HFIP and methanol (50:50) separated on a PFP column (250 mm × 4.6 mm). The SFC separation was conducted using a gradient mode with a modifier range of 5-60%.

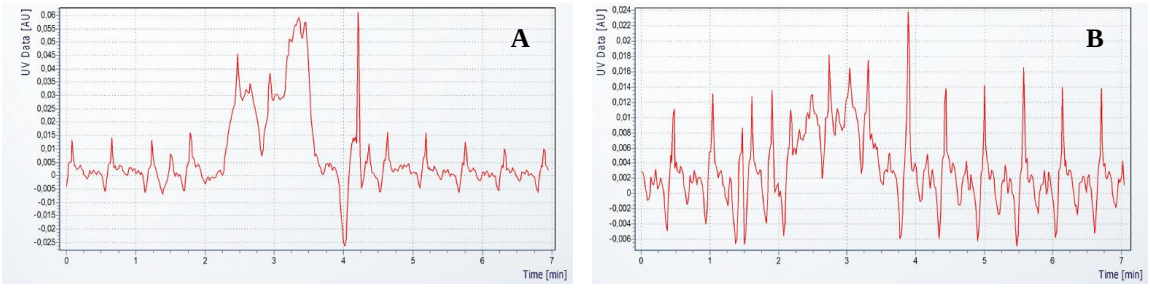


**Figure S5:** Represents SFC chromatograms of standard sample of insulin (0.01 mg/mL) dissolved in 50 mM Tris-HCl buffer and 0.2 % formic acid: A) 25 µL injection B) 10 µL injection, which were separated on a PFP (250 mm × 4.6 mm). The SFC separations were conducted using an isocratic mode with a 10% modifier.

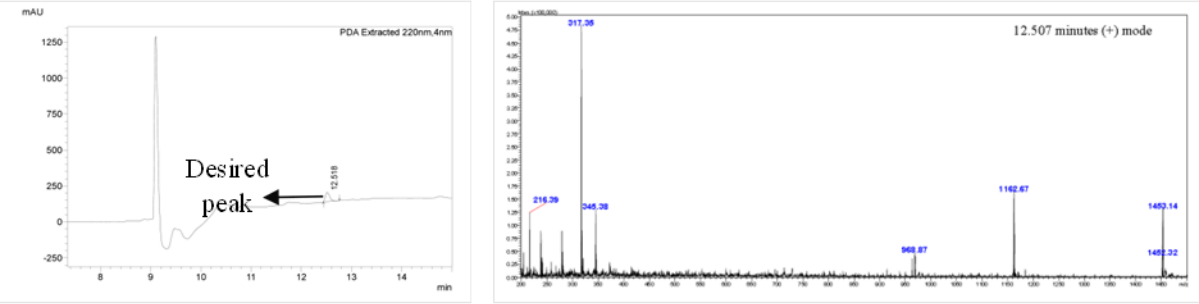


**Figure S6**: Represents a LC-MS spectrum obtained from a YMC-Triart $C_{18}$ (dimensions: 150 mm × 4.6 mm of the SFC purified standard sample of insulin.

**Table S1**: Reproducibility of retention times (n=3) for the separation of the standard sample of human insulin. A 5-60 % gradient was conducted.

| Column | Peaks | Average (minutes) | Standard deviation (minutes) | Relative standard deviation (%) | Assessment of column |
|---|---|---|---|---|---|
| GreenSep™ silica | 1 | 3,8 | 0.1 | 3.8 | Long retention times, poor peak resolution and baseline noise |
| | 2 | 5.1 | 0.1 | 1.5 | |
| | 3 | 5.7 | 0.1 | 2.6 | |
| | 4 | 7.9 | 0.1 | 1.8 | |
| 2'ethyl pyridine | 1 | 8.3 | 0.3 | 3.5 | Long retention times, baseline noise, poor peak shapes and peak resolution |
| | 2 | 9.5 | 0 | 0 | |
| diol-HILIC | 1 | 4,2 | 0,1 | 1,7 | Poor peak shape and peak splitting |
| | 2 | 6,5 | 0 | 0 | |
| PFP | 1 | 3,6 | 0 | 0 | No peak splitting, good peak shape, and resolution |
| | 2 | 6.9 | 0 | 0 | |

**Table S2**: Peptide fragmentation table of standard human insulin, which yielded a 90% probability to human insulin in the protein database.

| B | B Ions | B+2H | B-NH3 | B-H2O | AA | Y Ions | Y+2H | Y-NH3 | Y-H2O | Y |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 58.0 | | | | G | 859.4 | 430.2 | 842.4 | 841.4 | 7 |
| 2 | 205.1 | | | | F | 802.4 | 401.7 | 785.4 | 784.4 | 6 |
| 3 | 352.2 | | | | F | 655.3 | | 638.3 | 637.3 | 5 |
| 4 | 515.2 | | | 497.2 | Y | 508.3 | | 491.3 | 490.3 | 4 |
| 5 | 616.3 | | | 598.3 | T | 345.2 | | 328.2 | 327.2 | 3 |
| 6 | 713.3 | 357.2 | | 695.3 | P | 244.2 | | 227.1 | | 2 |
| 7 | 859.4 | 430.2 | 842.4 | 841.4 | K | 147.1 | | 130.1 | | 1 |

**Table S3**: Peptide table report of the standard human insulin, which yielded a 90% probability to human insulin in the protein database.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Valid | Weight | Sequenc | Prob | SEQUEST | SEQUEST | XI Tande | MS-Amar | NTT | Modifica | Observec | Actual M | Charge | Delta Da | Delta PPI | Retentio | Intensity | TIC | Start | Stop | # Other P | Other Prc | Spectrum ID | | | |
| 2 | TRUE | 1.0 | GFFYTPK | 90% | 1.550559 | 0.431837 | | | 2 | | 430.2211 | 858.4277 | 2 | 0.000038 | 0.04435 | 3,042.39 | | 1.94028E | 1 | 0 | 0 | | 20190129_KG_32-3442-13758_13758 | | | |
| 3 | TRUE | 1.0 | GFFYTPK | 90% | 1.516015 | 0.4590188 | | | 2 | | 430.2209 | 858.4272 | 2 | -0.000450 | -0.5238 | 3,072.68 | | 9413750 | 1 | 0 | 0 | | 20190129_KG_32-3547-13968_13968 | | | |
| 4 | TRUE | 1.0 | GFFYTPK | 90% | 1.411593 | 0.4248729 | | | 2 | | 430.2210 | 858.4275 | 2 | -0.000206 | -0.2397 | 3,103.04 | | 3646240 | 1 | 0 | 0 | | 20190129_KG_32-3641-14166_14166 | | | |
| 5 | TRUE | 1.0 | GFFYTPK | 90% | 1.183619 | 0.42504108 | | | 2 | | 430.2209 | 858.4273 | 2 | -0.000389 | -0.4528 | 3,133.22 | | 1700300 | 1 | 0 | 0 | | 20190129_KG_32-3749-14375_14375 | | | |
| 6 | TRUE | 1.0 | GFFYTPK | 90% | 1.104835 | 0.4752052 | | | 2 | | 430.2210 | 858.4275 | 2 | -0.000145 | -0.1687 | 3,193.83 | | 325915 | 1 | 0 | 0 | | 20190129_KG_32-3948-14781_14781 | | | |
| 7 | TRUE | 1.0 | GFFYTPK | 90% | 1.152466 | 0.3155001 | | | 2 | | 430.2216 | 858.4287 | 2 | 0.001076 | 1.252 | 3,163.45 | | 674771 | 1 | 0 | 0 | | 20190129_KG_32-3837-14568_14568 | | | |
| 8 | END OF FILE | | | | | | | | | | | | | | | | | | | | | | | | | |

**Table S4**: Peptide fragmentation table of crude biosynthesised human insulin, which yielded a 96% probability to human insulin in the protein database.

| B | B Ions | B+2H | B-NH3 | B-H2O | AA | Y Ions | Y+2H | Y-NH3 | Y-H2O | Y |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 58.0 | | | | G | 859.4 | 430.2 | 842.4 | 841.4 | 7 |
| 2 | 205.1 | | | | F | 802.4 | 401.7 | 785.4 | 784.4 | 6 |
| 3 | 352.2 | | | | F | 655.3 | | 638.3 | 637.3 | 5 |
| 4 | 515.2 | | | 497.2 | Y | 508.3 | | 491.3 | 490.3 | 4 |
| 5 | 616.3 | | | 598.3 | T | 345.2 | | 328.2 | 327.2 | 3 |
| 6 | 713.3 | 357.2 | | 695.3 | P | 244.2 | | 227.1 | | 2 |
| 7 | 859.4 | 430.2 | 842.4 | 841.4 | K | 147.1 | | 130.1 | | 1 |

**Table S5**: Peptide table report of crude biosynthesised human insulin, which yielded a 96% probability to human insulin in the protein database.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Valid | Weight | Sequenc | Probabil | SEQUEST | SEQUEST | XI Tande | MS-Amar | NTT | Modifica | Observec | Actual M | Charge | Delta Da | Delta PPI | Retentio | Intensity | TIC | Start | Stop | # Other P | Other Prc | Spectrum ID | | | |
| 2 | TRUE | 1.0 | GFFYTPK | 96% | 1.401384 | 0.45985177 | | | 2 | | 430.2206 | 858.4267 | 2 | -0.000938 | -1.092 | 2,820.26 | | 4880830 | 1 | 0 | 0 | | 20190129_KG_54-6622-15922_15922 | | | |
| 3 | TRUE | 1.0 | GFFYTPK | 96% | 1.405259 | 0.43911934 | | | 2 | | 430.2209 | 858.4273 | 2 | -0.000328 | -0.3818 | 2,850.35 | | 7796630 | 1 | 0 | 0 | | 20190129_KG_54-6709-16114_16114 | | | |
| 4 | TRUE | 1.0 | GFFYTPK | 96% | 1.405259 | 0.4178313 | | | 2 | | 430.2209 | 858.4273 | 2 | -0.000328 | -0.3818 | 2,850.35 | | 7796630 | 1 | 0 | 0 | | 20190129_KG_54-6709-16114_16114 | | | |
| 5 | TRUE | 1.0 | GFFYTPK | 96% | 1.405259 | 0.4178313 | | | 2 | | 430.2209 | 858.4273 | 2 | -0.000328 | -0.3818 | 2,850.35 | | 7796630 | 1 | 0 | 0 | | 20190129_KG_54-6709-16114_16114 | | | |
| 6 | TRUE | 1.0 | GFFYTPK | 96% | 1.401384 | 0.40328243 | | | 2 | | 430.2206 | 858.4267 | 2 | -0.000938 | -1.092 | 2,820.26 | | 4880830 | 1 | 0 | 0 | | 20190129_KG_54-6622-15922_15922 | | | |
| 7 | TRUE | 1.0 | GFFYTPK | 96% | 1.401384 | 0.40328243 | | | 2 | | 430.2206 | 858.4267 | 2 | -0.000938 | -1.092 | 2,820.26 | | 4880830 | 1 | 0 | 0 | | 20190129_KG_54-6622-15922_15922 | | | |
| 8 | TRUE | 1.0 | GFFYTPK | 96% | 1.401384 | 0.40328243 | | | 2 | | 430.2206 | 858.4267 | 2 | -0.000938 | -1.092 | 2,820.26 | | 4880830 | 1 | 0 | 0 | | 20190129_KG_54-6622-15922_15922 | | | |
| 9 | TRUE | 1.0 | GFFYTPK | 96% | 1.405259 | 0.34777182 | | | 2 | | 430.2209 | 858.4273 | 2 | -0.000328 | -0.3818 | 2,850.35 | | 7796630 | 1 | 0 | 0 | | 20190129_KG_54-6709-16114_16114 | | | |

# Appendix 3 for Chapter 4

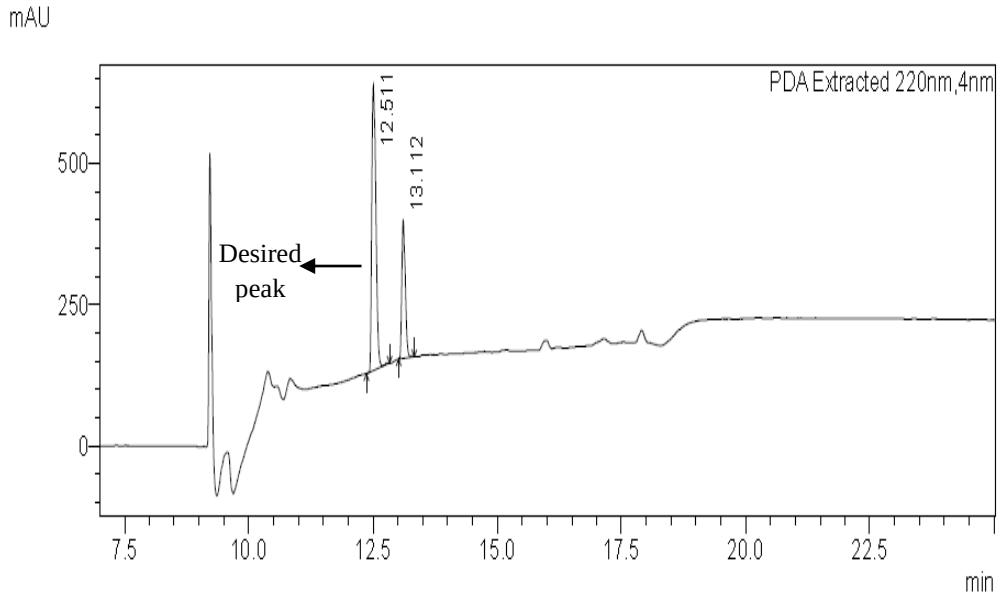


**Figure S1:** Represents a LC-MS spectrum obtained of a crude tetrapeptide separated on a YMC-Triart $C_{18}$ column (dimensions: 150 mm × 4.6 mm, pore size: 120 Å and a particle size of 5 µm).

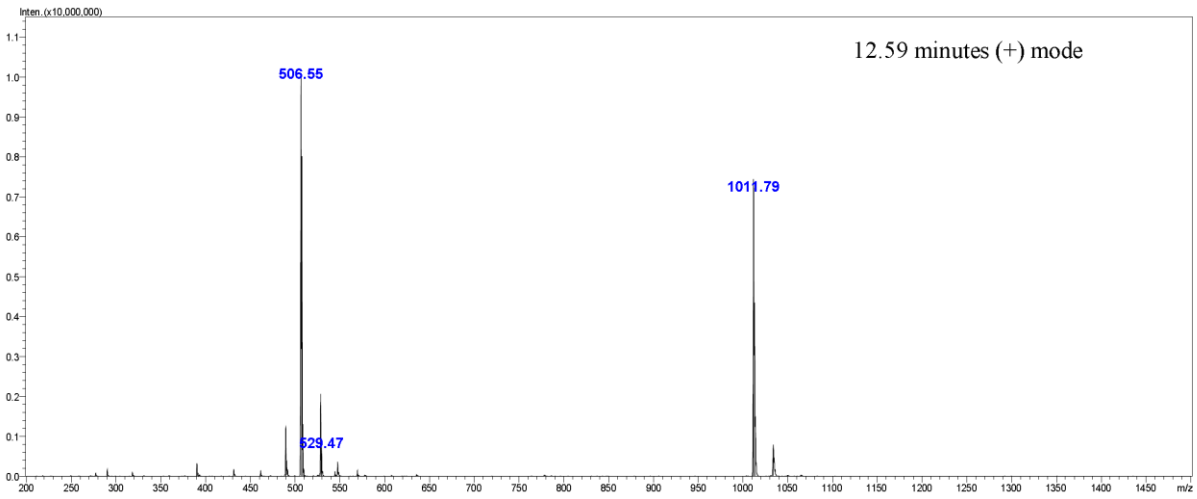


**Figure S2**: Represents a LC-MS spectrum of a crude tetrapeptide separated on a YMC-Triart $C_{18}$ column (dimensions: 150 mm × 4.6 mm, pore size: 120 Å and a particle size of 5 µm).

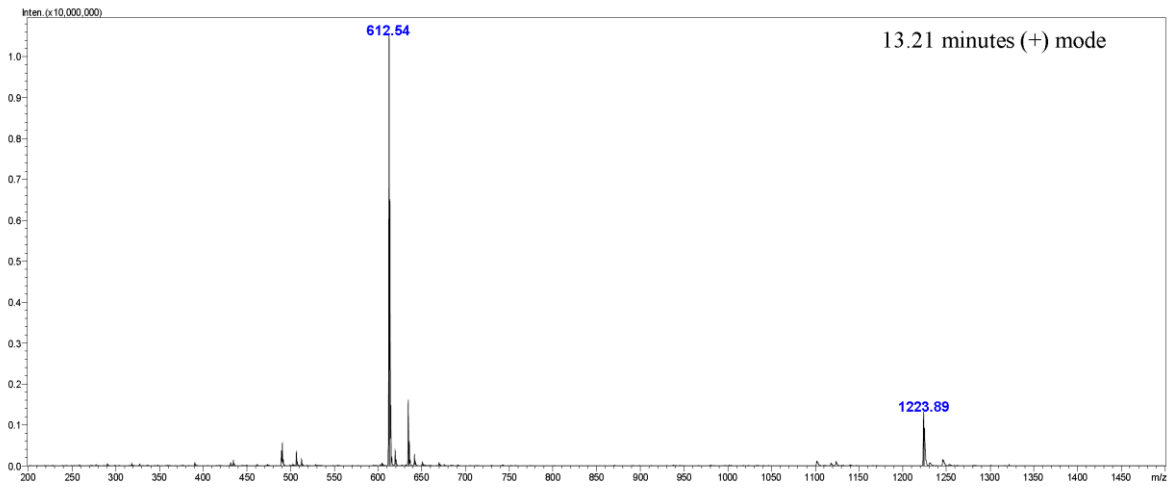


**Figure S3**: Represents a LC-MS spectrum of a crude tetrapeptide illustrating the mass of tetrapeptide with an impurity separated on a YMC-Triart $C_{18}$ column (dimensions: 150 mm × 4.6 mm, pore size: 120 Å and a particle size of 5 µm).

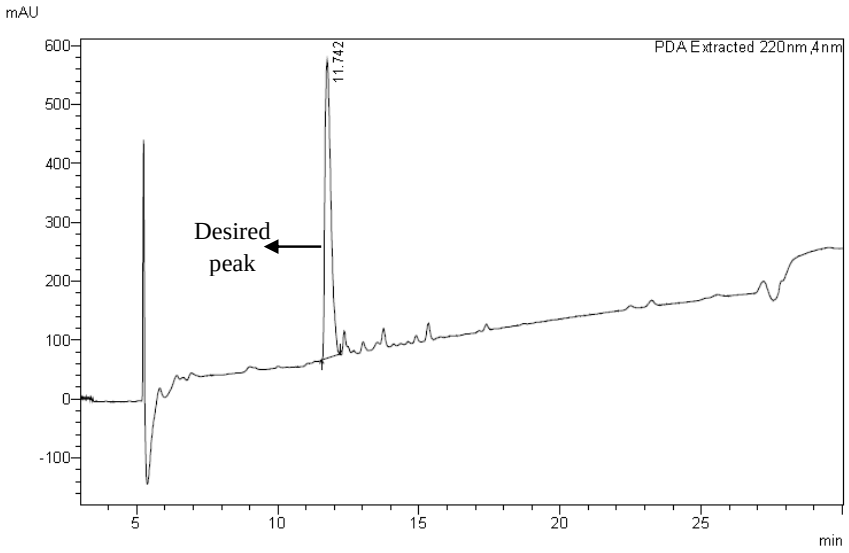


**Figure S4**: Represents a LC-MS spectrum of a crude octapeptide separated on a YMC-Triart $C_{18}$ column (dimensions: 150 mm × 4.6 mm, pore size: 120 Å and a particle size of 5 µm).

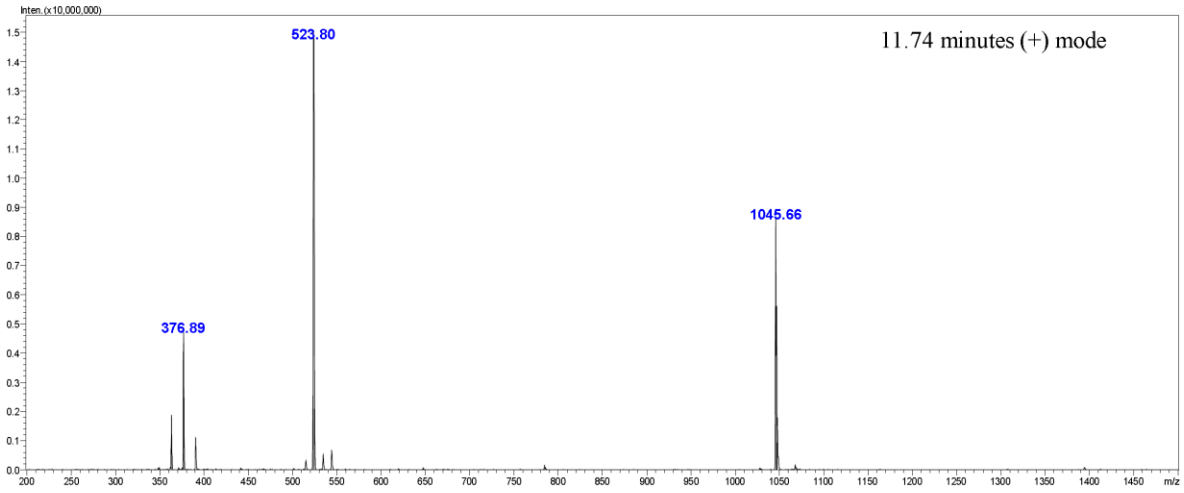


**Figure S5**: Represents a LC-MS spectrum of a crude octapeptide separated on a YMC-Triart $C_{18}$ column (dimensions: 150 mm × 4.6 mm, pore size: 120 Å and a particle size of 5 µm).

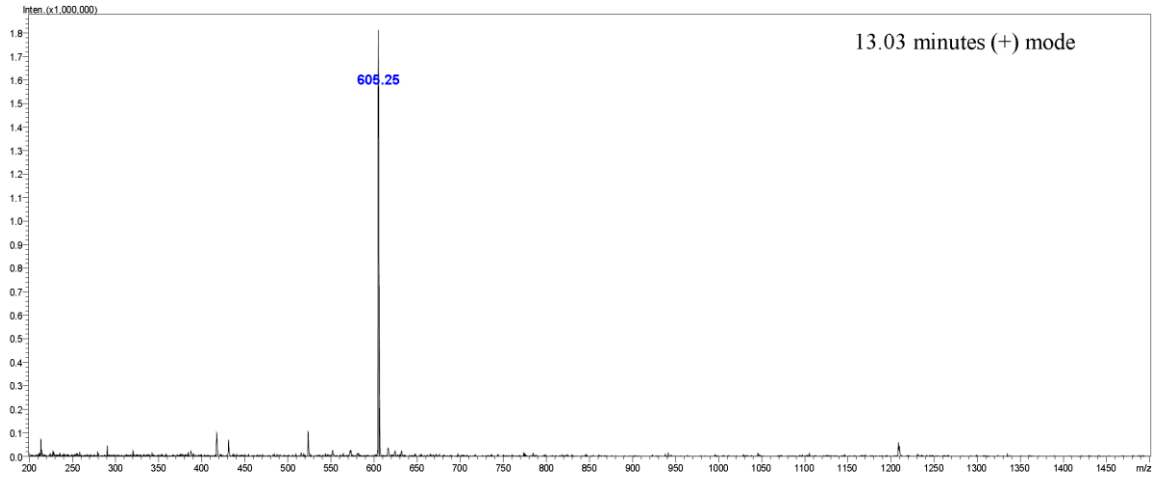


**Figure S6**: Represents a LC-MS spectrum of a crude octapeptide illustrating the mass of octapeptide with an impurity separated on a YMC-Triart $C_{18}$ column (dimensions: 150 mm × 4.6 mm, pore size: 120 Å and a particle size of 5 µm).

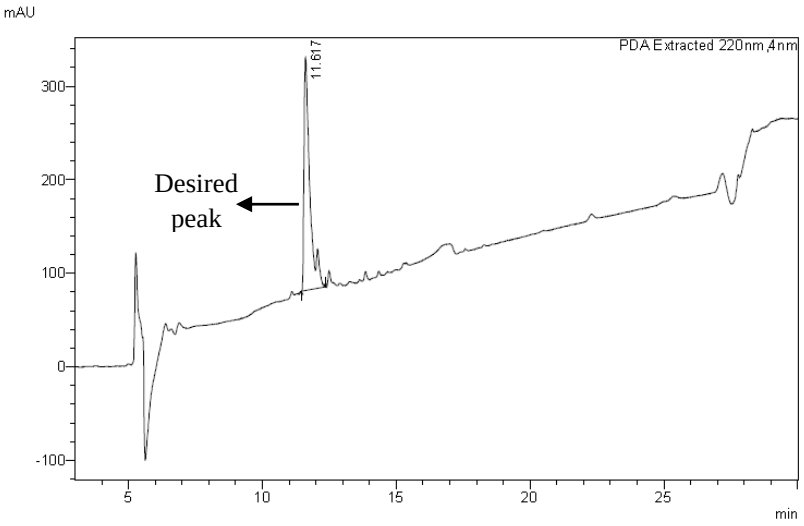


**Figure S7**: Represents a LC-MS spectrum of a crude nonapeptide separated on a YMC-Triart $C_{18}$ column (dimensions: 150 mm × 4.6 mm, pore size: 120 Å and a particle size of 5 µm).

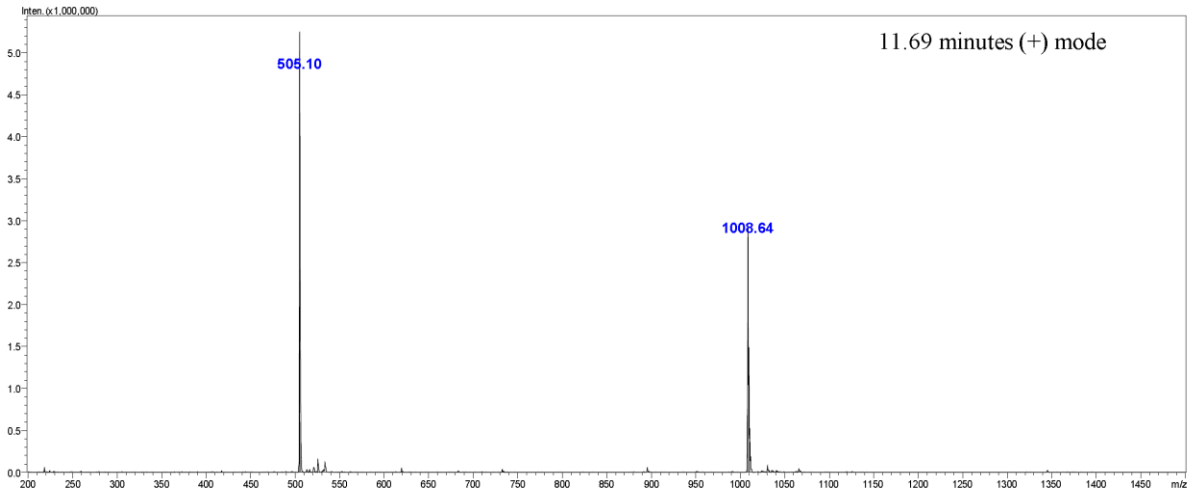


**Figure S8**: Represents a LC-MS spectrum of a crude nonapeptide separated on a YMC-Triart $C_{18}$ column (dimensions: 150 mm × 4.6 mm, pore size: 120 Å and a particle size of 5 µm).

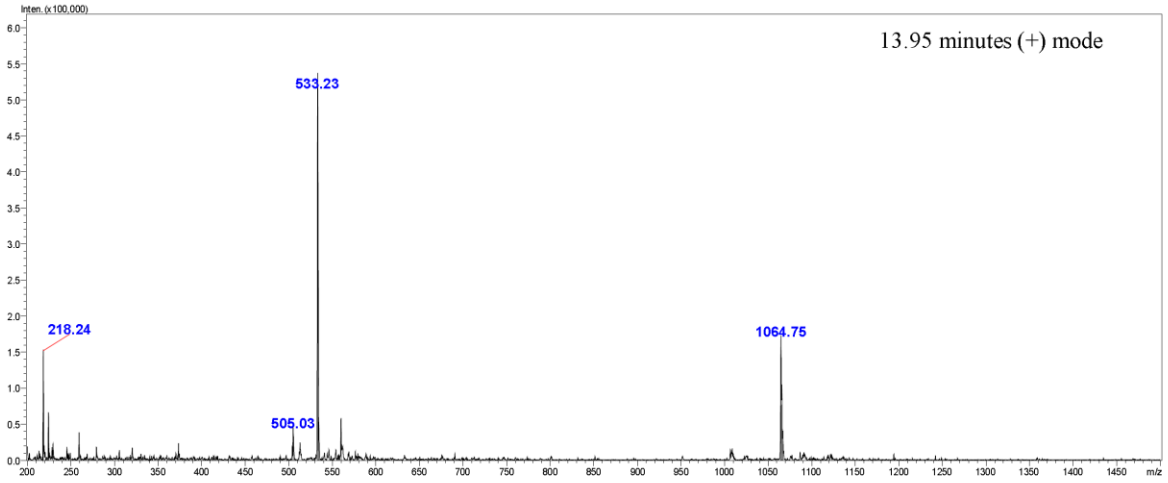


**Figure S9**: Represents a LC-MS spectrum of a crude nonapeptide illustrating the mass of nonapeptide with an impurity separated on a YMC-Triart $C_{18}$ column (dimensions: 150 mm × 4.6 mm, pore size: 120 Å and a particle size of 5 µm).

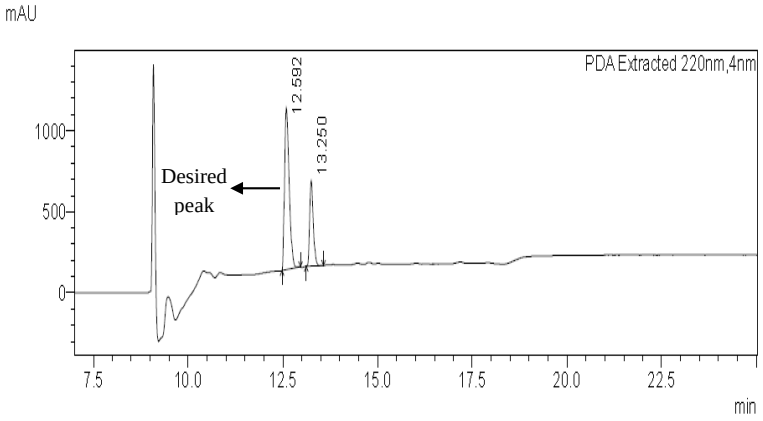


**Figure S10:** Represents a LC-MS spectrum separated on a YMC-Triart $C_{18}$ (dimensions: 150 mm × 4.6 mm, pore size: 120 Å and a particle size of 5 µm) of crude tetrapeptide after SFC purification using the PFP column.

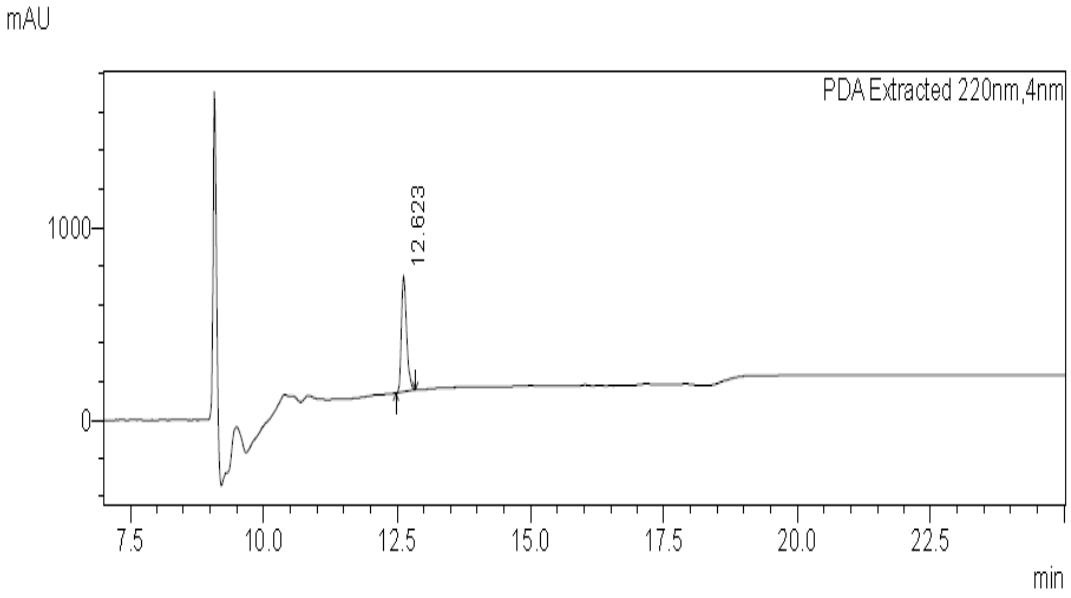


**Figure S11:** Represents a LC-MS spectrum separated on a YMC-Triart $C_{18}$ column (dimensions: 150 mm × 4.6 mm, pore size: 120 Å and a particle size of 5 µm) of a SFC purified tetrapeptide separated on the 2' ethyl pyridine column.

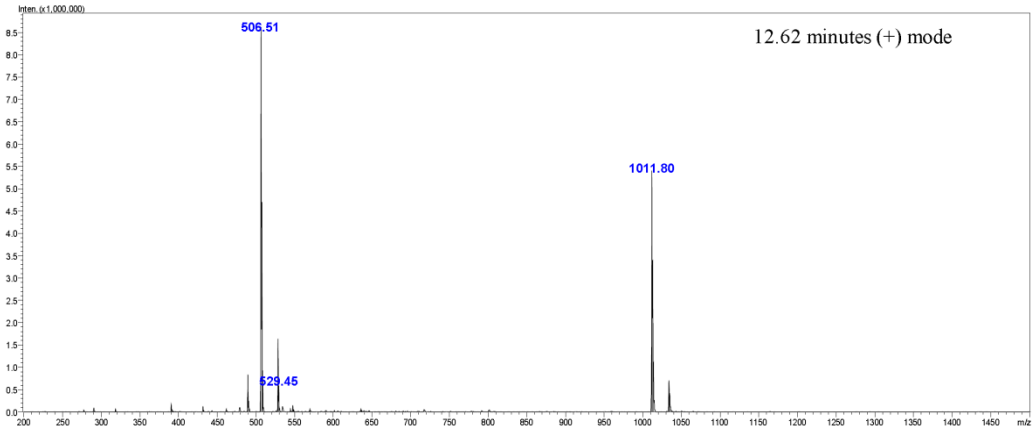


**Figure S12:** Represents a LC-MS spectrum separated on a YMC-Triart $C_{18}$ column (dimensions: 150 mm × 4.6 mm, pore size: 120 Å and a particle size of 5 µm) of a SFC purified tetrapeptide using the 2' ethyl pyridine column.

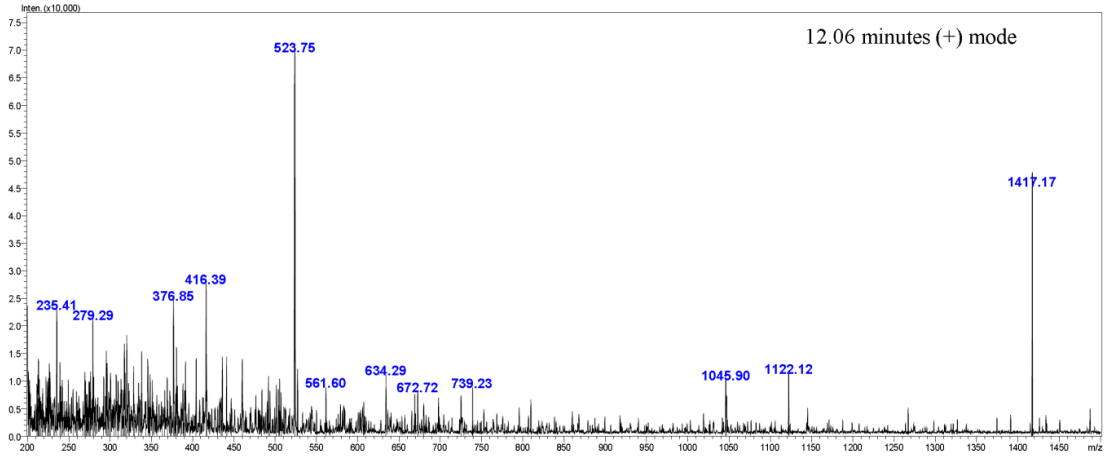


**Figure S13:** Represents a LC-MS spectrum separated on a YMC-Triart $C_{18}$ (dimensions: 150 mm × 4.6 mm, pore size: 120 Å and a particle size of 5 µm) of crude octapeptide after SFC purification using the diol-HILIC column.

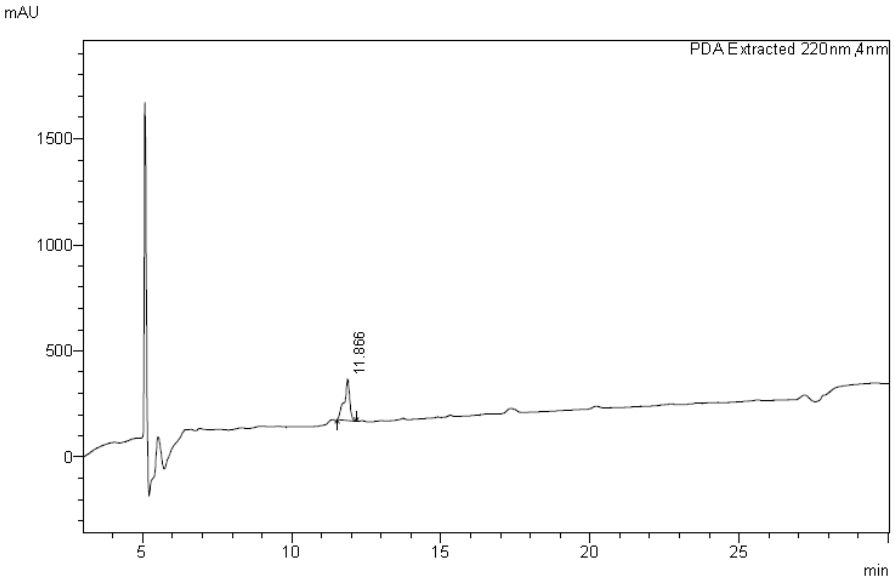


**Figure S14:** Represents a LC-MS spectrum separated on a YMC-Triart $C_{18}$ column (dimensions: 150 mm × 4.6 mm, pore size: 120 Å and a particle size of 5 µm) of a SFC purified octapeptide separated on the 2' ethyl pyridine column.

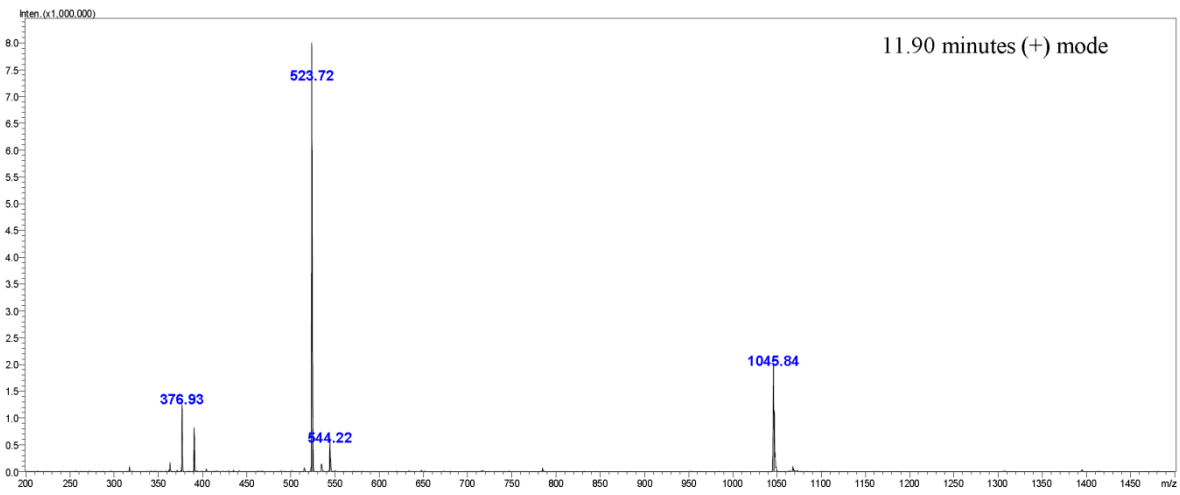


**Figure S15:** Represents a LC-MS spectrum separated on a YMC-Triart $C_{18}$ column (dimensions: 150 mm × 4.6 mm, pore size: 120 Å and a particle size of 5 µm) of a SFC purified octapeptide separated on the 2' ethyl pyridine column.

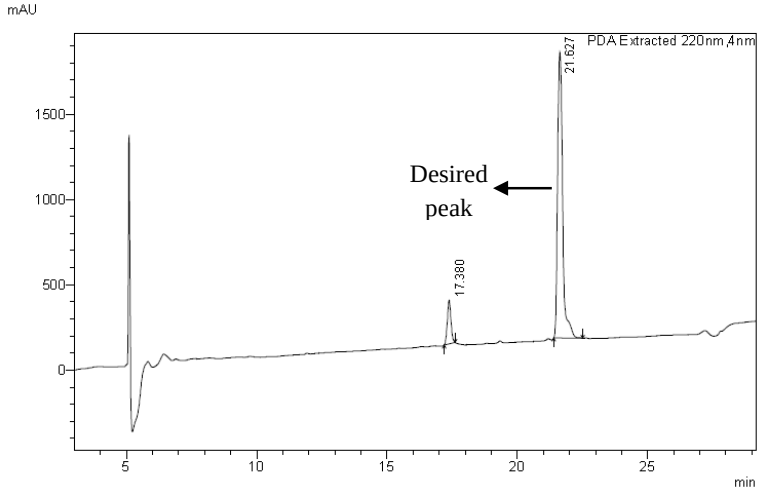


**Figure S16:** Represents a LC-MS spectrum separated on a YMC-Triart $C_{18}$ (dimensions: 150 mm × 4.6 mm, pore size: 120 Å and a particle size of 5 µm) of crude nonapeptide after SFC purification using the diol-HILIC column.

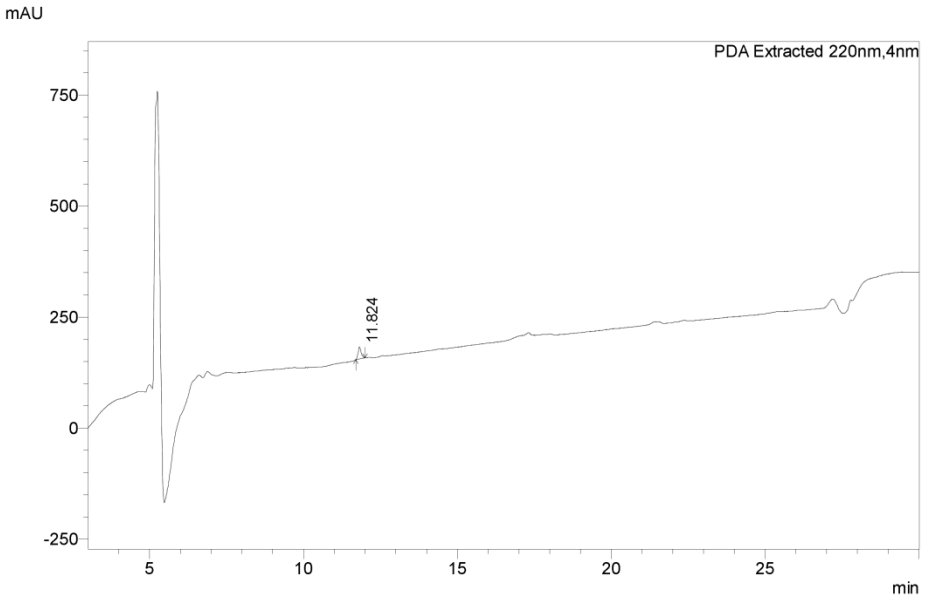


**Figure S17:** Represents a LC-MS spectrum obtained utilizing a YMC-Triart $C_{18}$ column (dimensions: 150 mm × 4.6 mm, pore size: 120 Å and a particle size of 5 μm) of a SFC purified nonapeptide using the 2' ethyl pyridine column.

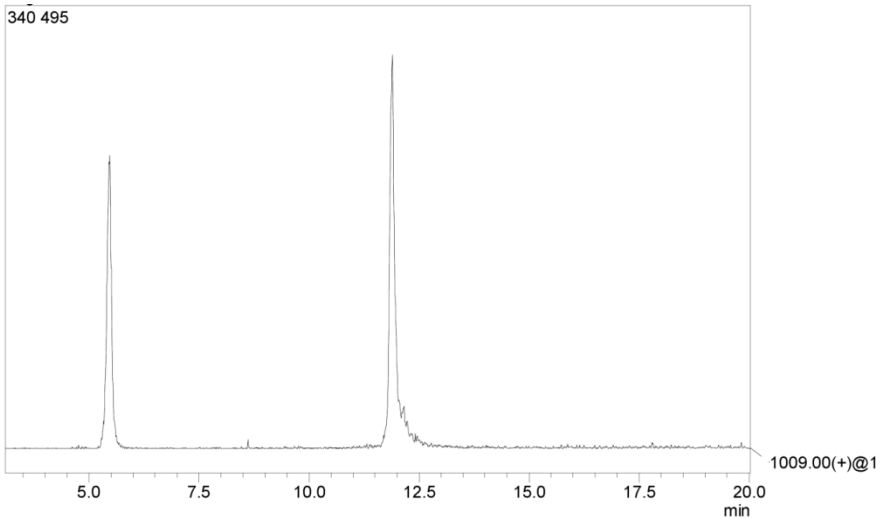


**Figure S18:** Represents a LC-MS spectrum obtained utilizing a YMC-Triart $C_{18}$ column (dimensions: 150 mm × 4.6 mm, pore size: 120 Å and a particle size of 5 μm) of a SFC purified nonapeptide using the 2' ethyl pyridine column.

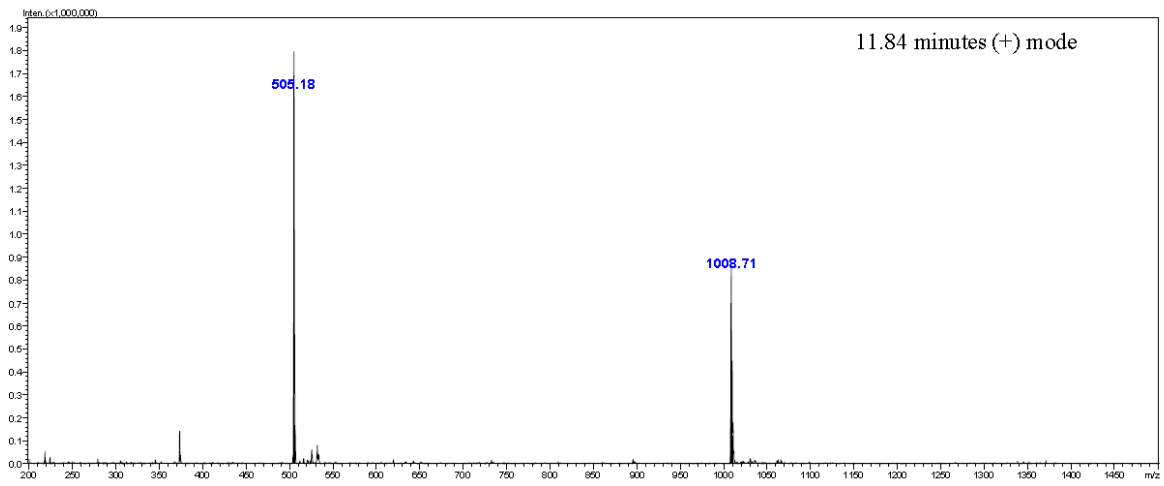


**Figure S19:** Represents a LC-MS spectrum obtained utilizing a YMC-Triart $C_{18}$ column (dimensions: 150 mm × 4.6 mm, pore size: 120 Å and a particle size of 5 µm) of SFC purified nonapeptide using the 2' ethyl pyridine column.

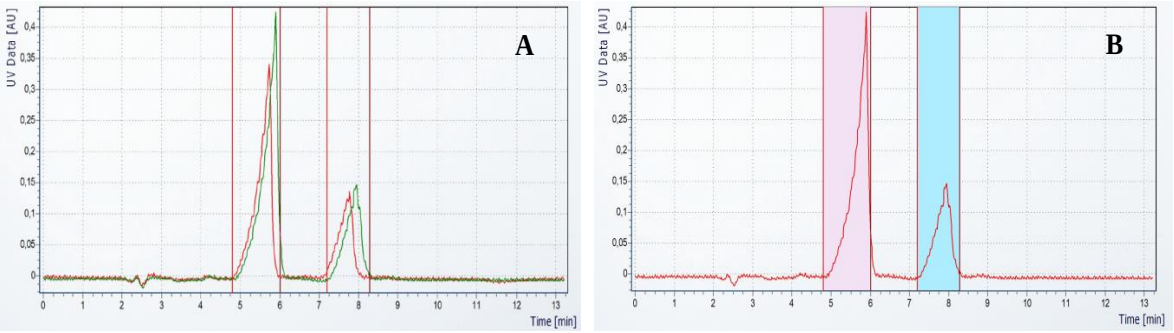


**Figure S20**: A and B illustrates SFC chromatograms of 50 µL injections of the tetrapeptide which was separated on the 2' ethyl pyridine column with the following characteristics; dimensions: 250 mm × 10 mm, pore size 300 Å and particle size of 5 µm, utilizing 20% isocratic conditions.

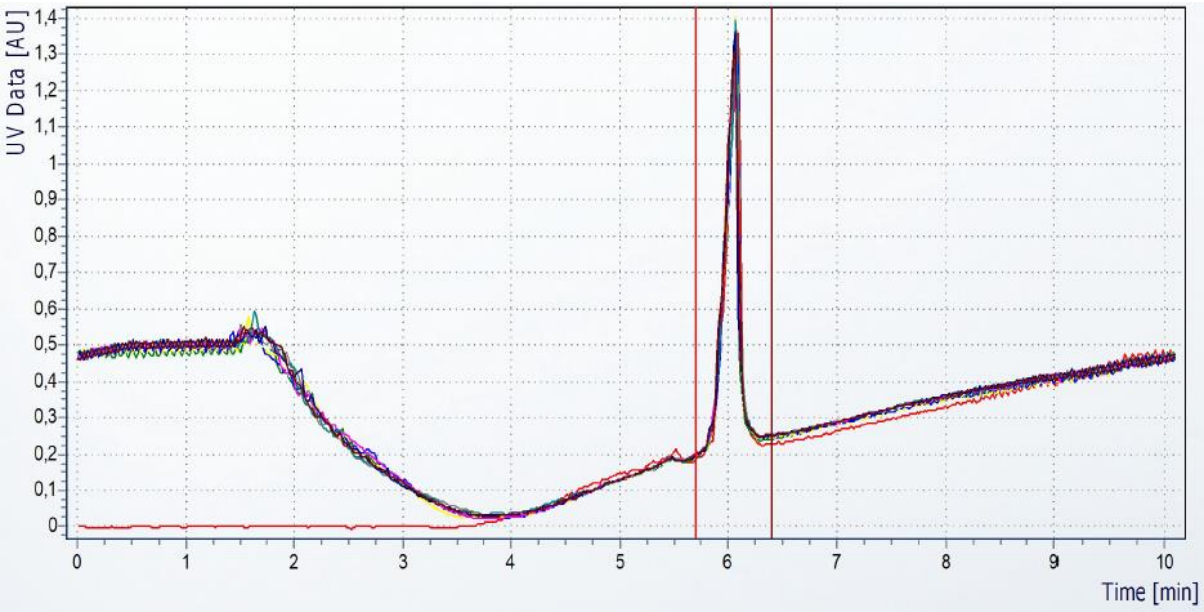


**Figure S21**: Illustrating a SFC chromatogram of multiple injections of 50 µL of the octapeptide which was separated on the 2' ethyl pyridine column with the following characteristics; dimensions: 250 mm × 10 mm, pore size 300 Å and particle size of 5 µm, utilizing a 5-60% gradient.

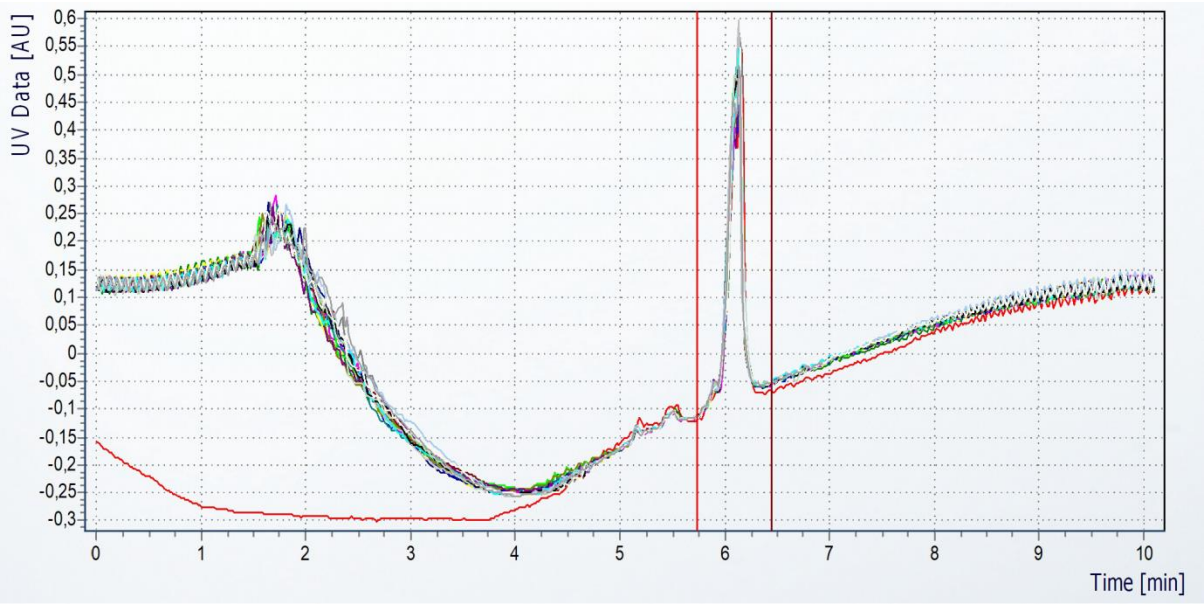


**Figure S22**: Illustrating a SFC chromatogram of multiple injections of 50 µL of the nonapeptide which was separated on the 2' ethyl pyridine column with the following characteristics; dimensions: 250 mm × 10 mm, pore size 300 Å and particle size of 5 µm, utilizing a 5-60% gradient.

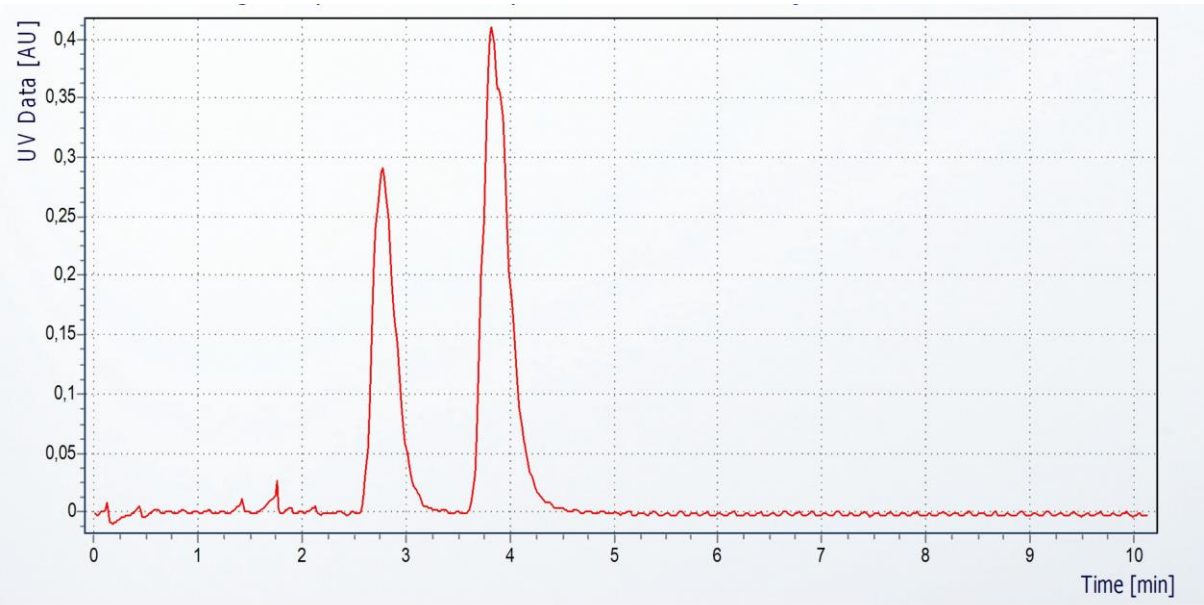


**Figure S23**: Represents an SFC chromatogram of a mixture of 2' Bromo-6-nitro toluene (0.5 mg/mL) and N-benzyl benzamide (1.25 mg/mL) of tested on a 2' ethyl pyridine column with the following dimensions: 250 mm × 10 mm, pore size: 300 Å and a particle size of 5 µm.

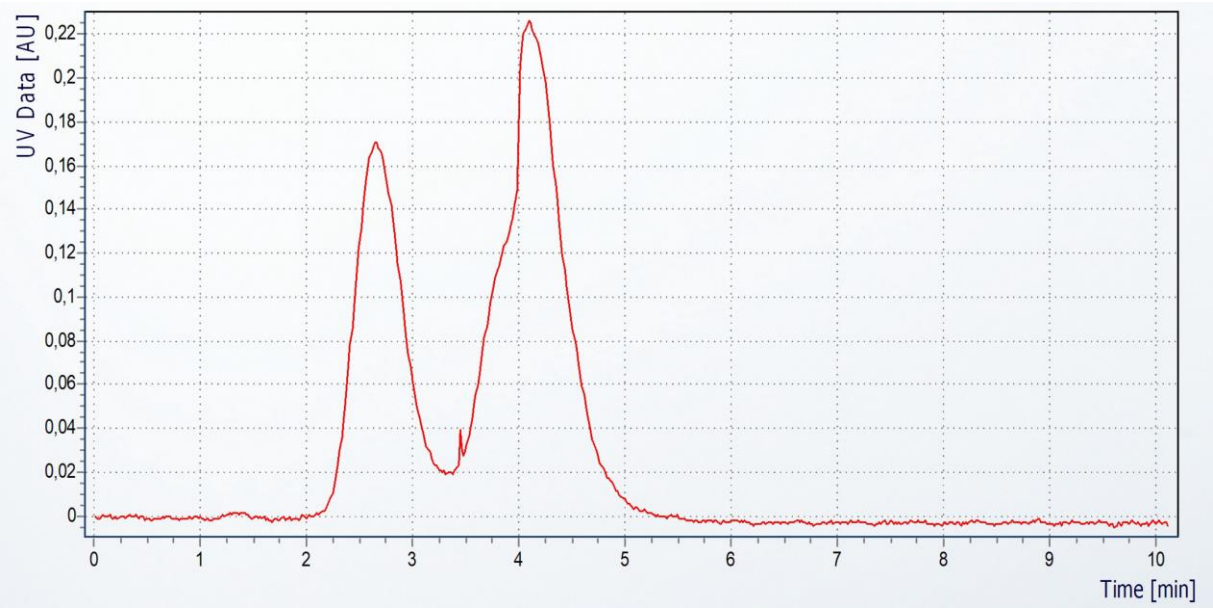


**Figure S24**: Represents an SFC chromatogram of a mixture of 2' Bromo-6-nitro toluene (0.5 mg/mL) and N-benzyl benzamide (1.25 mg/mL) of tested on a 2' ethyl pyridine column with the following dimensions: 250 mm × 4.6 mm, pore size: 100 Å and a particle size of 5 µm.

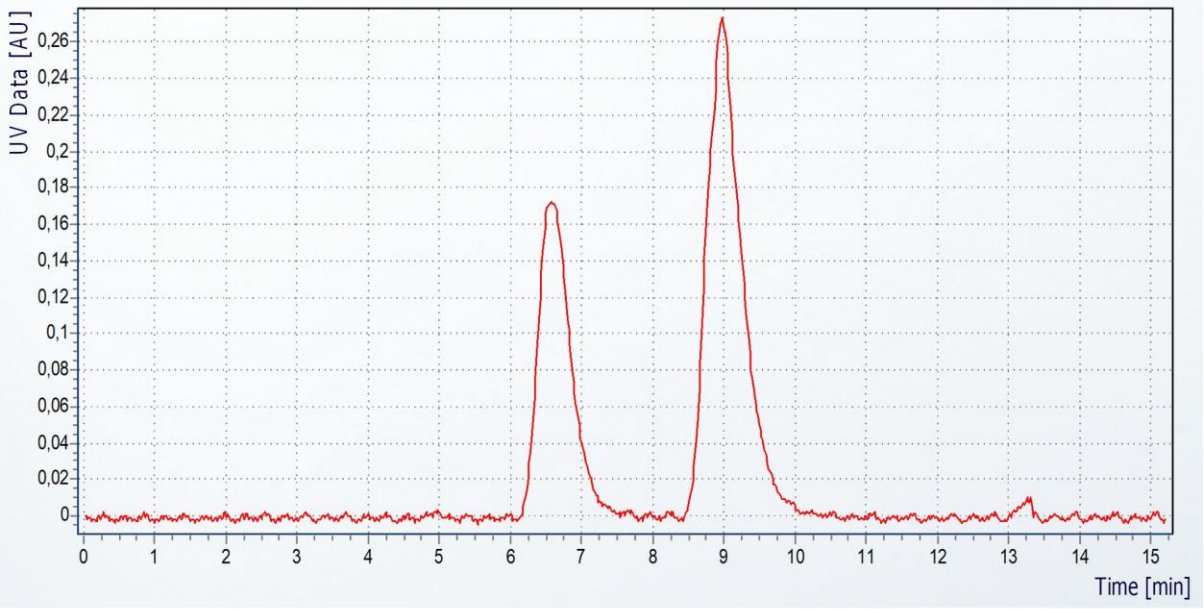


**Figure S25**: Represents an SFC chromatogram of a mixture of 2' Bromo-6-nitro toluene (0.5 mg/mL) and N-benzyl benzamide (1.25 mg/mL) were tested on a diol-HILIC column with the following dimensions: 250 mm × 4.6 mm, pore size: 120 Å and a particle size of 5 µm.

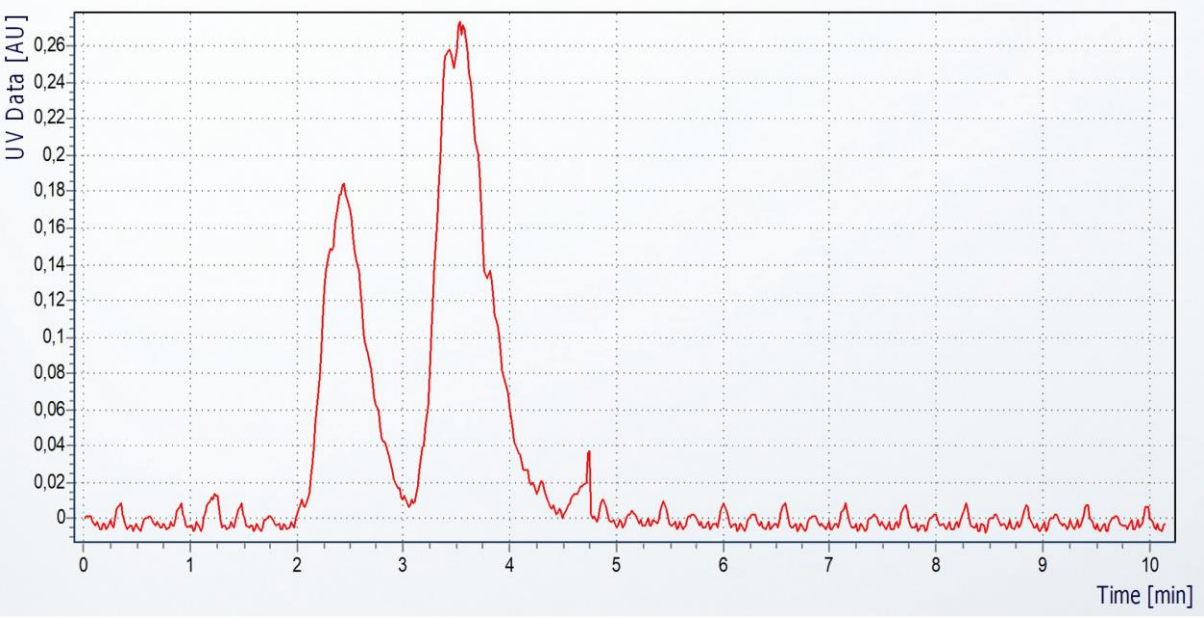


**Figure S26**: Represents an SFC chromatogram of a mixture of 2' Bromo-6-nitro toluene (0.5 mg/mL) and N-benzyl benzamide (1.25 mg/mL) were tested on a PFP column with the following dimensions: 250 mm × 4.6 mm, pore size: 120 Å and a particle size of 5 µm.