Author:
Dodd, Sarah B

Title:
Design and Synthesis of Novel SRPK1 Inhibitors in the Treatment of Breast Cancer

School of Chemistry

Design and Synthesis of Novel SRPK1 Inhibitors in the Treatment of Breast Cancer

Sarah Dodd

This thesis is submitted in fulfilment of the requirements for the Master of Research programme at the University of Bristol

September 2018

Dr. Chris Arthur

Dr. Paul Gates

Acknowledgements

I would like to thank Dr Chris Arthur, my supervisor for his help, support and advice during the year and for allowing me to conduct and learn valuable skills from such a fascinating study.

I will also like to thank the N317 Lab group especially Tom Burnham for his help and advice in using new techniques and equipment.

Also, I will like to express my thanks to the J. Adams group for donating the plasmids used in this research.

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: Sarah Dodd DATE: September 10, 2018

Table of Contents

1. Abstract

The use of computational methods in drug design is valuable in reducing costs and time. It allows researchers to optimise lead compounds and to predict and measure their binding affinity. This helps streamline the drug discovery process. This thesis discusses the use of computational methods to discover novel compounds for the inhibition of the SRPK1 protein. The protein plays a significant role in angiogenesis through the regulation of anti-angiogenic VEGF isoform formation. Many studies have shown that inhibition of SRPK1 has a correlation to reduced angiogenesis. Inhibitors of the protein can be useful in the treatment of a number of conditions like neovascular eye disease and cancer.

Four novel compounds were discovered through pharmacophore modelling and docking studies which was used to measure the strength of the interaction between compound and target. Though not all the compounds could be synthesised, four other compounds underwent synthesis. Unfortunately, *in vitro* toxicity and binding studies could not be conducted. The *in vitro* testing however was conducted on SRPIN340 after it was successfully synthesised. The toxicity tests for SRPIN340 showed toxicity around a concentration of 25 µM.

2. Introduction

Angiogenesis is the formation and maintenance of new blood vessels from pre-existing vasculature.[2] This is not to be confused with vasculogenesis- the *de novo* creation of blood vessels.[3] Angiogenesis is a constant and vital process in the body due to its role in growth and development, wound healing and disease. It is also one of the first and most critical processes to begin in embryo formation and growth.[4] Blood vessels are important in the distribution and diffusion exchange of oxygen, nutrients, antibodies and hormones. Therefore, changes in activity or regulation of angiogenesis can prove detrimental to the body and may bolster the activity of certain diseases and conditions. Diseases which include cancer, rheumatoid arthritis and wet age-related macular degeneration (AMD). Due to this, angiogenesis has become a target in the control and treatment in the above diseases and other ischaemic illnesses.[4, 5]

2.1 The Angiogenic Process

Angiogenesis occurs as a series of events, but these events can occur in two different ways. Both ways occur *in utero* and throughout life from the infanthood to adulthood and they are: sprouting and intussusceptive angiogenesis.[2]

Intussusceptive angiogenesis is the formation of blood vessels by the growth and extension of the capillary wall into the lumen to split one vessel into two new ones.[6, 7] On the other hand, sprouting angiogenesis which is more pertinent to this research, is characterised by the formation of `sprouts' from endothelial cells (EC) to create new blood vessels and this happens in multiple stages.

Firstly, in the case of a hypoxic environment, oxygen sensing mechanisms begin EC activation using angiogenic factors such as VEGF (Vascular Endothelial Growth Factor). Then, the capillary wall is degraded by extracellular proteinases and from this, the degraded part of the begins to serve as a branch point for the new blood vessel.

ECs begin to proliferate to make sprouts which then migrate into the extracellular membrane (ECM) towards the area with a high concentration of angiogenic stimuli. Tip cells on the sprouts have processes called filopodia which secrete high levels of proteolytic enzymes that help create a route in the ECM for the sprout.[8] They also have heavy concentrations of VEGF receptors (VEGFR2) which help them direct the sprouts to where they can sense high levels of the angiogenic stimuli.[9] When enough of the tip cells reach their destination, there is a reorganisation of the ECs to form tubules.[10]

2.2 VEGF

As mentioned before, the ECs proliferate and move towards the VEGF concentration gradient. VEGF is a polypeptide growth factor recognised to play a significant role in inducing angiogenesis.[9] The VEGF family mainly consists of six members with each having specialised roles. They are: VEGF-A (the first to be identified), VEGF-B, VEGF-C, VEGF- D, VEGF- E (from the orf parapox virus) and placenta growth factor (PIGF).[11] VEGF-A is perhaps the most important in terms of the angiogenic pathway and coupled with exon splicing of the VEGF-A gene, molecular variants of this same protein can be made with only slight differences in the total amino acid number. In humans, the main five variants are VEGF-A_{121}, VEGF-A_{145}, VEGF-A_{165}, VEGF-A_{189} and VEGF-A_{206}. VEGF-A_{165} is one of the most pro-angiogenic isoforms whilst the VEGF-A_{165b} isoform is anti-angiogenic [12]therefore highlighting the changes splicing can have on the function of a protein. The splicing process that cells use to determine which of these isoforms to generate are specially regulated by the proteins SPRK1 (Serine/arginine-rich kinase 1) and SRSF1 (Serine/arginine-rich splicing factor).

2.3 SRPK1

SRPKs are a group of enzymes that specifically phosphorylate the serine residues in serine/arginine proteins (SR proteins). SR proteins are an essential class of splicing factors which in turn is essential in pre- mRNA processing. Phosphorylation of these proteins by SRPKs is necessary for them to conduct both constitutive and alternative splicing (Figure 1). [1]Splicing is the process of removing introns from pre-mRNA and consequently joining the exons to form mature RNA.[13] This reaction is catalysed by spliceosomes. Constitutive splicing is the process where the exons are spliced in the same way and order, always with a spliceosome. On the other hand, alternative splicing allows for variations in the expression of genes.[1, 14] Different sites for splicing are selected with placing different exons in different orders, therefore, allowing for a wider variety of isoforms.

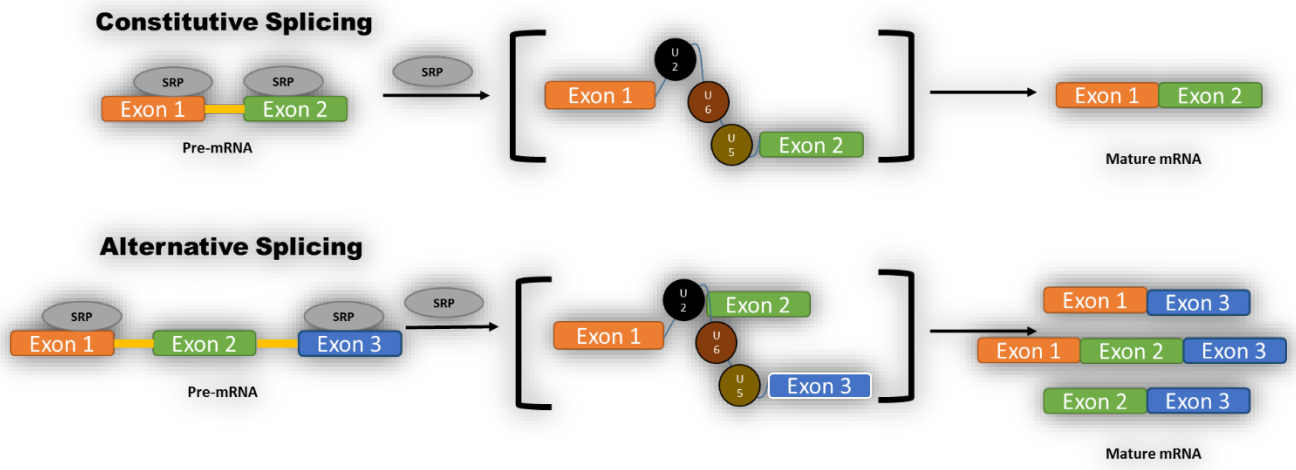
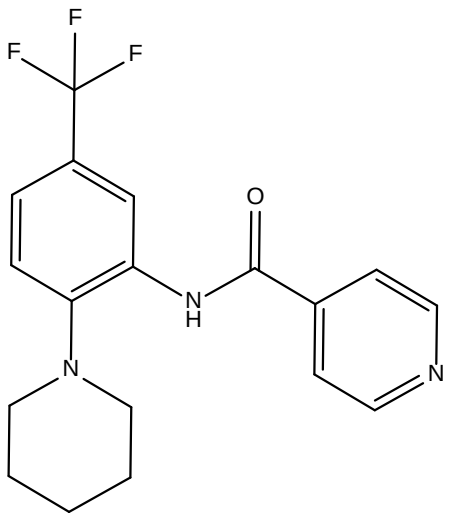


Figure 1: Constituted and Alternative splicing by serine/arginine proteins (SRP) with the possible mature mRNA products for each mechanism. Adapted from Wang et al. Intron domains characterised by the U circles.

SRPK1 is key in the up-regulation of the alternative splicing of VEGF-A to create pro-angiogenic isoforms through the phosphorylation of the SRSF1.[15] SRPK1 is usually anchored in the cytoplasm but translocated to the nucleus. When in the nucleus, an accumulation of SR proteins in the cytoplasm is induced subsequently altering downstream splicing events. Due to this important role, it is imperative that SRPK2 is tightly regulated. SRPK1 has been seen to be upregulated in cancers such as breast[16], colon, pancreatic and lung cancers.[17] Where SRPK1 expression has been reduced, studies showed an increase in pro-apoptotic proteins sensitised breast cancer cells to apoptosis.[18] Another study showed a correlation between SRPK1 expression and the preferential metastasis of breast cancer cells into the brain and the lungs[19]. To target the inhibition of SRPK1 can, therefore, be an avenue where angiogenesis can be reduced.

Over the years some compounds have been found to have an inhibitory effect on the SPRK1 protein. With some being small molecule inhibitors that fit into the active site cleft of kinase core of the protein. One of these small molecule inhibitors is SRPIN340 (Figure 2) which was determined by Fukuhara *et al*.[20] It has shown a Ki of 0.89 µM and IC50 of 0.14 µM in mice.



SRPIN340 (1)

Figure 2: Molecular structure of SRPIN340

Three components of the SRPIN340 molecule make it good at binding to the SRPK1 protein. The X-ray structure of the protein reveals a hydrophobic pocket created by a kinase hinge region next to a helix in the SRPK1 insertion domain packs which is shown in Figure 3 and Figure 4.[21] When this pocket is analysed when bound to SRPIN340, the interactions between molecule and protein can be seen. The CF_3 group orients the molecule towards the hydrophobic pocket allowing the molecule to enter the binding site. This is due to its small size but very electronegative nature. The piperidine and isonicotinamide groups serve as hydrogen bond acceptors which interact with the altered DFG and catalytic lysine sequences in the protein. These interactions can be seen in Figure 4.[22]

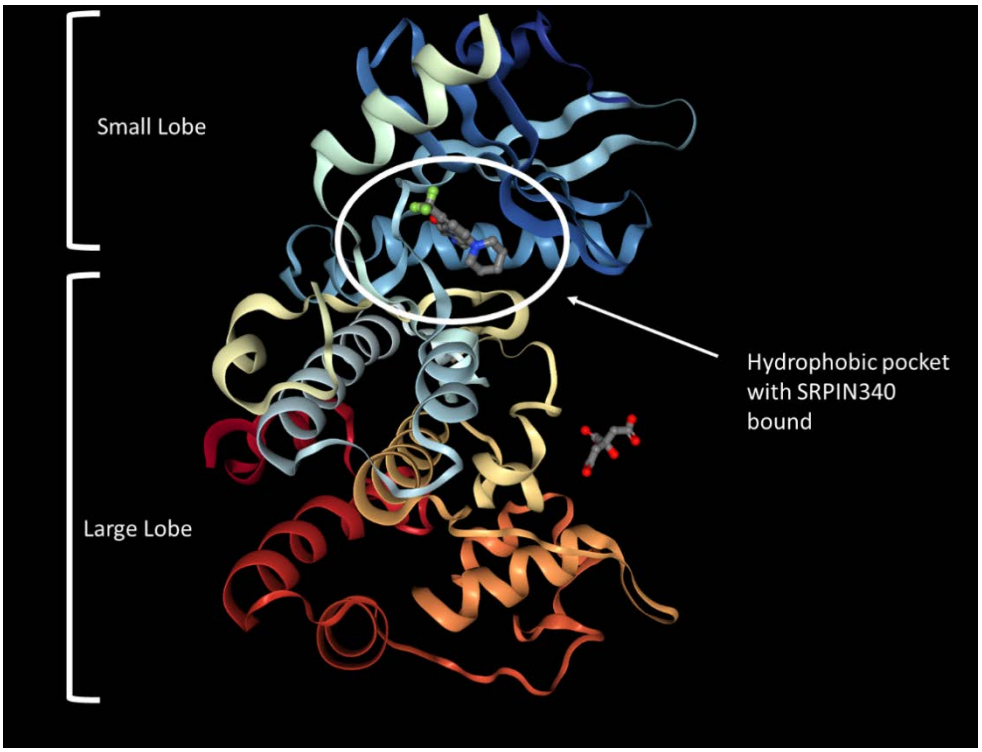


Figure 2: X- Ray structure of SRPK1 bound to SRPIN340 showing the large and small lobes.

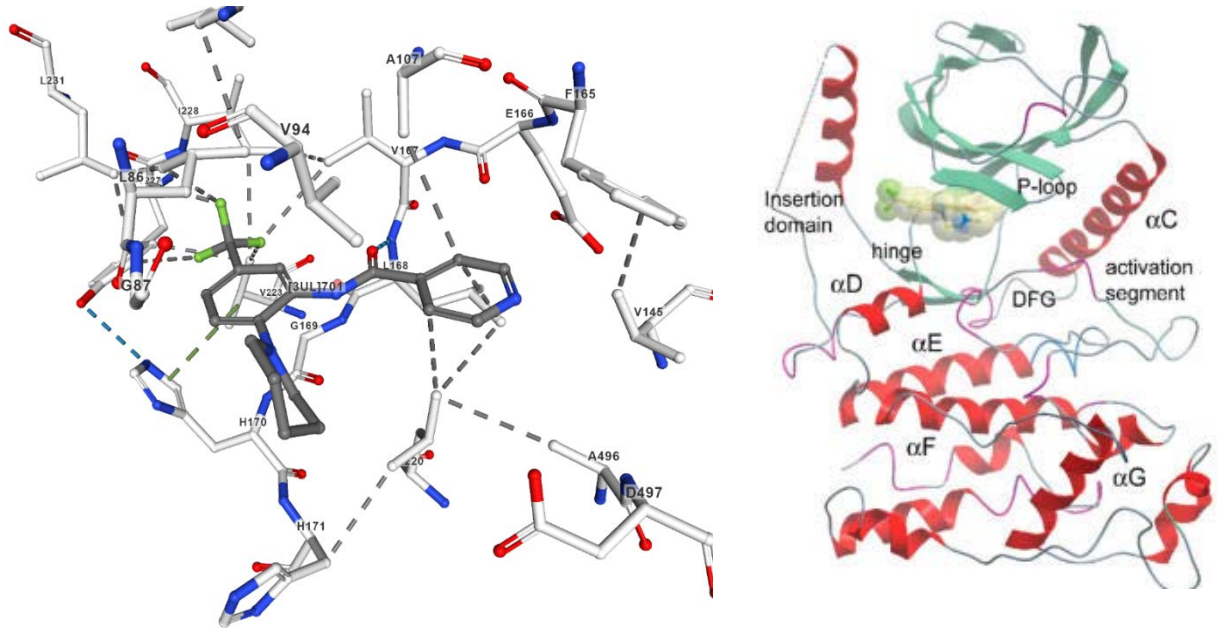


Figure 4: Left portion shows the molecular interactions between SPRIN340 and SRPK1 residues. SPRIN340 is coloured in dark grey and protein residues are in white. The right shows more detailed regions of the SRPK1 protein. Left portion is from the RSCB PDB[21] and the right portion is adapted from Batson et. al. [23]

Though SRPIN340 shows great promise, it's not ideal as an antiangiogenic drug. Firstly, it lacks selectivity in its inhibition of kinases. Not only has SRPIN340 been shown to inhibit SRPK1, but it has also shown inhibition is SRPK2[20] and this could lead to higher toxicities and unpredictable effects. Secondly, SPRIN340 is almost insoluble in water and best dissolved in DMSO with an estimated logP of 3.589. This makes administration of the drug difficult. For oral ingestions, solubility in water is imperative as drugs with low water solubility often require higher doses to have a therapeutic effect and the use of higher doses of a drug can cause problems when it comes to toxicity.[24, 25]

However, SRPIN340 is a small molecule inhibitor. This is an advantage because some of the current treatments of neovascular eye conditions and some cancers employ the use of monoclonal antibodies. The use of these antibodies usually requires a subcutaneous injection and sometimes directly to the afflicted area. This can result in low patient compliance as doing it in a pain-free way can be challenging. Also, there are risks of infections and other adverse effects on the injection site.

2.4 Angiogenesis in Disease

Mammalian cells can only be at maximum 100 to 200 µm from the nearest blood vessel to access and receive the oxygen and nutrients they require.[26] When the number of cells increase, more blood vessels are needed to provide the constant stream of oxygen and nutrients needed and this is when angiogenesis is necessary[3]. Without angiogenesis and a limited amount of blood vessels, the size of the group of cells is restricted. In a healthy mammal, the process of angiogenesis is highly regulated by pro- and anti-angiogenic proteins. These molecules are exploited in angiogenesis- dependent diseases.

2.4.1 Macular Degeneration

Age-related macular degeneration is one of the leading causes of blindness especially amongst the elderly in the developed world.[27] This condition is caused by the damage to the macula of the retina[28] as seen in Figure 5. Though early on there are little to no symptoms, sufferers begin to experience a slow worsening of vision in the affected eye. In most cases, there is not complete loss of vision, but the reduced vision makes it difficult for sufferers to perform daily tasks such as driving and reading.

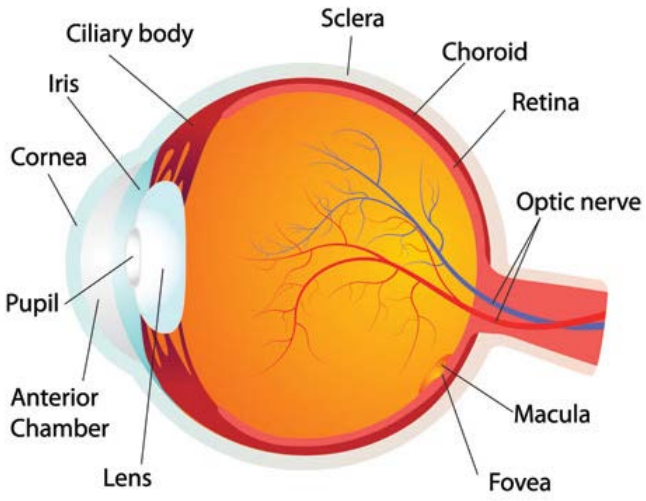


Figure 5: A drawn anatomy of the eye showing the macula- the part of the eye most affected by AMD.[29]

Some of the most common symptoms of macular degeneration are distorted vision (where straight lines may appear wavy), blurred vision and a loss in contrast sensitivity.[30] As mentioned before, macular degeneration rarely leads to complete vision loss. This is because the macula only takes about 2% of the area of the retina, therefore, leaving the remaining 98% of the peripheral field.[31] However, even though such a small portion is taken by the macular almost half of the processing power of the visual cortex goes to information gained from the macula.

The three main stages to the development of AMD are the early, intermediate and late stages. In AMD, there is an accumulation of deposits named drusen (deposits usually made from proteins and lipids) in the macular and it is the size of this buildup, that is used to characterise the stages of AMD.[32]

Early AMD is characterised with medium-sized drusen, around the width of a human hair with intermediate AMD having a larger amount of drusen. In the late stages, enough damage has occurred to the retina that loss of vision has begun.[32] This late-stage can be divided into two sub-categories: wet and dry AMD.

Dry AMD accounts for about 80 to 90% of all cases and tends to have a slow progression.[33] In this category, large amounts of drusen have accumulated as well as an irreversible loss of retinal cells. In some cases, there is a progression to the wet type. Wet AMD, also known as neovascular AMD, is caused by the abnormal increase in the growth of blood vessels in the choriocapillaris[34] as seen in figure 6. These abnormal blood vessels tend to be fragile and can easily break leading to the release of proteins and blood in the eye which goes on to cause damage that is irreversible to the photoreceptors resulting in a rapid loss in vision.[34]

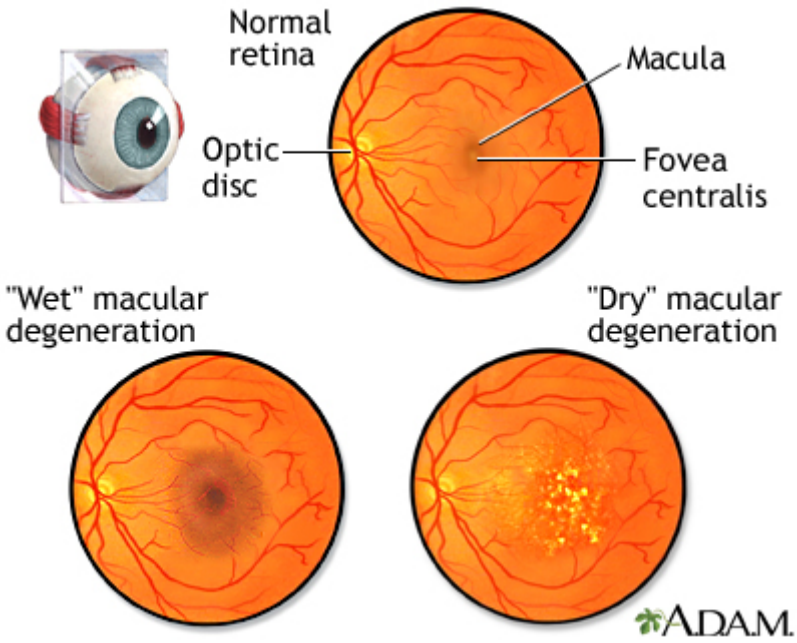


Figure 6: Frontal section of the eye showing the difference in the macula in Wet and Dry AMD.[35]

Dry AMD has no medical treatment available on the other hand wet AMD has a number of treatments. Laser coagulation in conjunction with medication has been known to stop and sometimes even reverse angiogenesis in the condition. Bevacizumab and ranibizumab (structures in Figure 9) have been shown to inhibit angiogenesis with similar efficacy to Laser coagulation. Other approved anti-angiogenic drugs used include aflibercept and pegaptanib.

2.4.2 Angiogenesis and Cancer

The proliferation and metastatic spread of cancer cells is dependent on the increased growth of vasculature for the required stream of nutrients, metabolites and oxygen. It has been shown that pre-cancerous cells or tissue become cancerous after acquiring the ability to switch on angiogenesis.[36] This ability can be induced through different mechanisms such as hypoxia-induced tumour angiogenesis.[37]

Metastasis is made efficient when tumour cells infiltrate blood vessels, travel around the circulatory system and relocate to a different part of the body such as an organ, bone or tissue and proliferate.[38, 39] This highlights the vital role angiogenesis plays in the growth and metastatic nature of tumour cells. Further studies have even gone to show that when cancer cells are starved of blood circulation, they only grew to be around 2 mm in diameter, unlike cells which were in environments where angiogenesis could be switched on. The cells grew to a much bigger size.[40]

2.4.2.1 Tumour Angiogenesis

To gain the nutritional needs required to grow and expand, cancer cells use co-opt existing vasculature and then go on to form vasculature of their own down the line as seen Figure 7.[41]

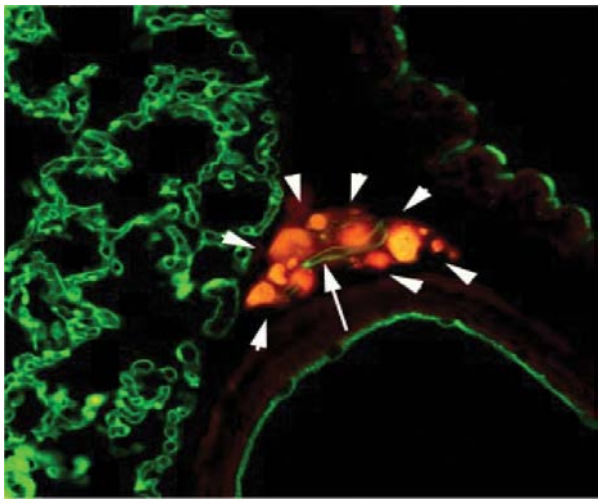


Figure 7: Tumour angiogenesis in the lung. Cancer cells are in red, vasculature labelled with anti-PECAM antibodies in green.

Oxygen has been known to be crucial in the regulation of angiogenesis. In blood vessels, both the smooth muscle (SMC) and endothelial cells (ECs) have mechanisms specifically for the sensing of oxygen. Such mechanisms include the use of heme-oxygenases and oxygen-sensitive NADPH oxidases.[42] Another type of oxygen sensors interacts with the hypoxia-inducible transcription factor (HIF) family which can initiate the expression of many genes vital to cell survival and angiogenesis. These features are attributed to normal blood vessel structures.[43] When co-opting, the tumour continues to expand and take over more of the adjacent vasculature but through this takeover, the morphology of the vessels become altered. This alteration goes on to affect its physiology and its therapeutic response.[44]

Unlike normal blood vessels, tumour- associated vessels are not hierarchic, or efficient in the delivery of nutrients to a tumour. This lack of efficiency is due to the inefficient coating of smooth muscle on the vessel. They also are randomly structured with highly irregular patterns of branching[45] (Figure 8).[41] Another alteration in morphology in tumours are that the ECs are hyperactive by releasing proteins such as endothelial growth factor (EGF), VEGF and oestrogen. All these proteins encourage additional endothelial cell growth.

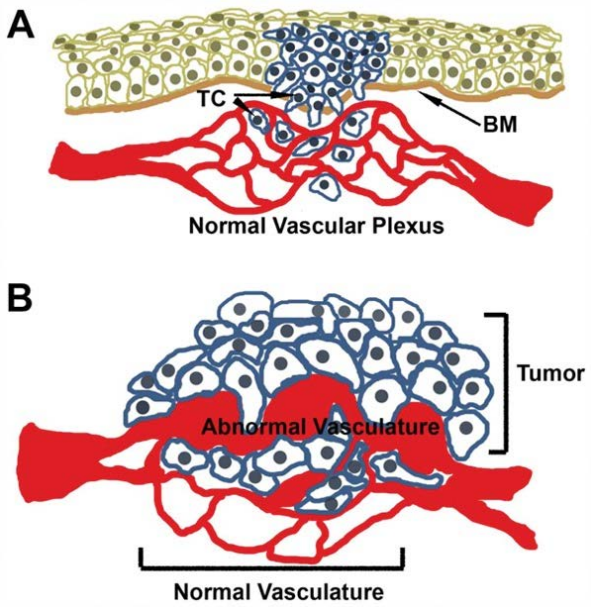


Figure 8: Vasculature co-opting by tumour cells (TC). (A): TC migrating across the membrane and associating with normal nearby vessels. (B) Co-opted vessel modification after a period of time by the TCs.[41]

As stated previously, tumour vasculature is inefficient in the provision of nutrients and oxygen to a tumour. This causes hypoxia and acidosis in the cancer cells and if left too long, necrosis and apoptosis are induced in the tumour[46] however, hypoxia coupled with angiogenesis mutations in the cell leads to the activation of hypoxia-induced angiogenesis[37, 47] which starts with upregulation of the VEGF mRNA and protein.[48]

2.4.2.2 Current Anti-angiogenesis Cancer Treatments

There are many different cancer treatments that target different pathways and angiogenesis inhibitors. By stopping or reducing the production of the new blood vessels, the tumour growth can be halted through starvation. These compounds can then be used in conjunction with other blood vessels in cancer treatments.

Angiogenesis inhibitors can be placed in two classes: direct and indirect inhibitors. The direct inhibitors target the ECs.[49] They prevent the proliferation and migration of ECs due to the cells responding to inducers like VEGF and Platelet-derived growth factor (PDFG)[50]. Examples of such inhibitors are angiostatin, canstatin, tumstatin and arrestin but these examples are fragment peptides of larger proteins that are released through proteolysis.[51]

Indirect inhibitors target tumours and tumour- associated cells. They also inhibit the expression or function of pro-angiogenic proteins like VEGF and EGF.[52] A lot of anti-angiogenesis approved drugs fall in this category. Some of these drugs on the market are; thalidomide, gefitinib, bevacizumab, sunitinib and axitinib which are illustrated in figure 9.

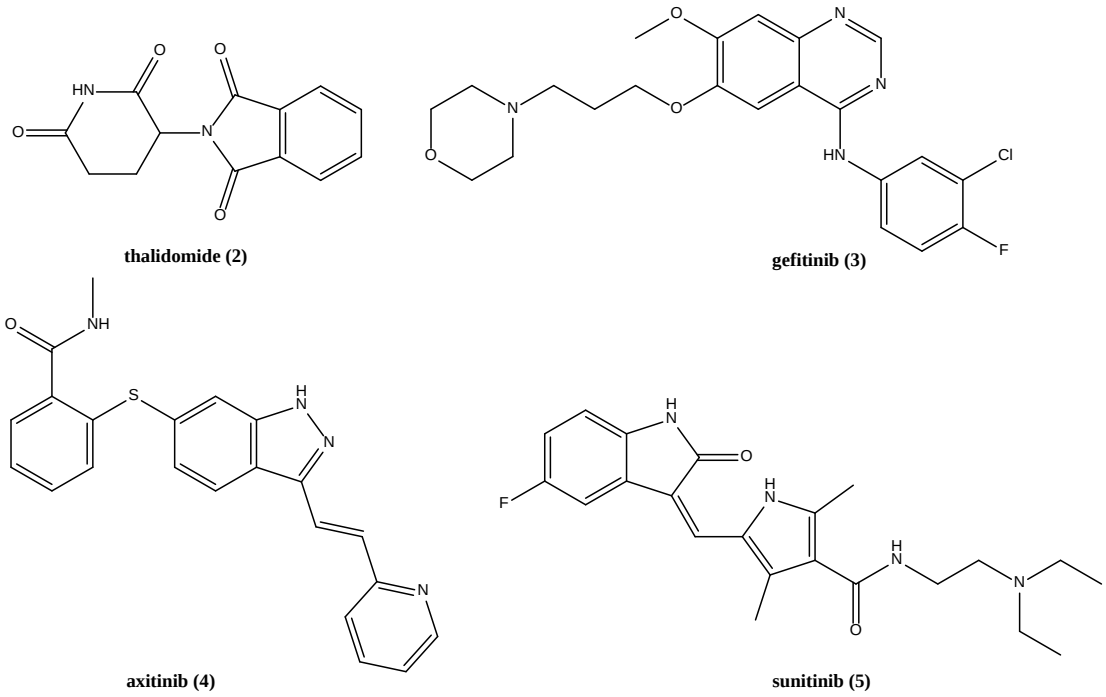


Figure 9: Structures of some of the current anti-angiogenic drugs used in the treatment of cancer.

Once used as an anti-emetic drug, thalidomide has been used in treating myeloma and cancers that produce VEGF and similar angiogenic cytokines like placental growth factor (PLGF).[53] It inhibits the phosphorylation of Akt which is significant in the pathways of many pro-angiogenic factors. Due to its side effects of the drug such as thromboembolism and neuropathy, analogues have been made like pomalidomide. Thalidomide as a therapy on its own has not been shown to have much significant anti-angiogenic effect yet, has proven effective against many different myeloma[54] when coupled with dexamethasone which is a corticosteroid commonly used in the treatment of rheumatic problems.

Bevacizumab (Avastin) is a humanized monoclonal anti-VEGF- A antibody. It is used to treat lung, renal, ovarian and glioblastomas though it is not effective in breast cancers. It binds to extracellularly to VEGF-A to prevent binding to VEGFRs found on the surfaces of ECs.[55] Some other monoclonal antibodies used are ramucirumab and IMC-18F1 with others like cetuximab and volociximab being under consideration.

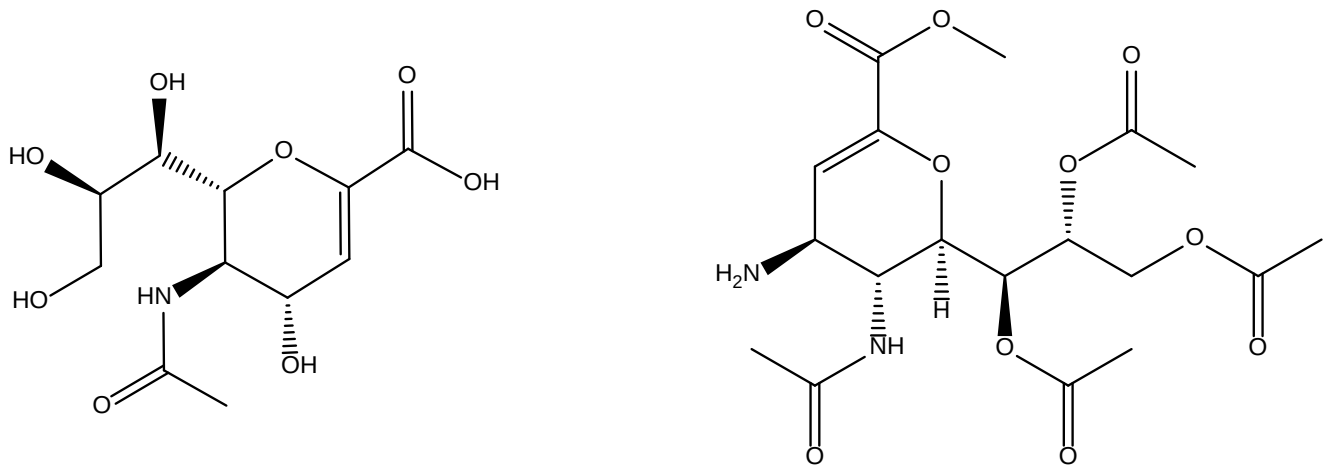
Another class of anti-angiogenesis drug are the receptor tyrosine kinase inhibitor small molecules (RTKIs). Examples of these are sunitinib and axitinib seen in figure 9 with sunitinib being primarily used in the treatment of kidney and renal cell carcinoma. They inhibit the enzymatic activity of the tyrosine kinase fragment found in multiple receptors involved in angiogenesis pathways.

2.5 Computational Methods in Drug Discovery

The discovery and development of drugs is often a very long, costly and complex process. A lot of the time and money spent in the process is usually due to synthesising and testing leads. Over the years, the use of computational methods in the drug discovery pipeline has become increasingly more important. Computer-aided drug design (CADD) has become a way to reduce costs and time because of its lead optimisation abilities. It has become especially more significant with its ability to create *in silico* models of molecules and isolating desired attributes such as binding affinity and specificity of the proposed molecule.

Many heavily used drugs on the market today were discovered through computational methods and these drugs are found in therapies for HIV/AIDS and other viral diseases and cancer. In influenza, one of the most notable examples is the drug Zanamivir (illustrated in Figure 10). The viral target used was the neuraminidase enzyme. By studying and exploiting its structure through x-ray crystallography, the molecule was designed specifically to fit the binding site of the target.

Neuraminidase is an enzyme important to the influenza virus. It cleaves the terminal sialic acid residues on the surface of host cells and the virus itself. It also helps in the release of the newly formed virus from the surface of infected cells.[56, 57] Inhibitors to this enzyme tend to be sialic acid analogues and they work through competitive inhibition by binding to the active site. This leaves no room for the enzyme to bind to the sialic acid on the surface of the host cell. When the protein neuraminidase was crystallised, this allowed a 3D structure of the active site to be explored. From this, inhibitors could be created[58]. A lead compound 2-deoxy-2,3-didehydro-N-acetyl neuraminic acid (Neu5Ac2en) was found to be a small molecule inhibitor and this was used as a scaffold to find derivatives[59]. Using the structure of the active site, modifications of Neu5Ac2en could be tested to see its potency. With improvements in X-ray crystallography, important residues in the active sites were identified which resulted in knowing what substitutions to make. An example was finding that swapping the C4 hydroxyl group with a more basic group. After even more improvements with X-ray crystallography, it was discovered that an even larger basic group could be accommodated. So, a guanidyl group was used as the basic group to make 4-deoxy-4-guanidino-Neu5Ac2en. This was then named Zanamivir. The differences in Zanamivir compared to Neu5Ac2en can be seen in Figure 10.



Neu5Ac2en (6) **Zanamivir (7)**

Figure 10: Inhibitors Neu5Ac2n and the Zanamivir

2.5.1 Important Processes

One of the most important processes used in computational techniques in drug discovery is molecular recognition. In this process, *in silico* calculations of the interaction between a small molecule and a protein can be made. Through this, the behaviour of these interactions can be seen on an atomic level of detail which would be more difficult to do experimentally. Seeing this behaviour is done by linking structural and energetic properties using physics-based energy functions such as molecular mechanics. Interactions on a molecular level can be studied using docking studies.

Molecular docking is a method used in predicting the most suitable orientation of a ligand when complexed to an active site as seen in Figure 11. The primary aim is to find an optimised conformation for both ligand and receptor to a point where the free energy of the complex is minimised. Docking studies are particularly important in not just finding optimum orientation but to also predicting the binding affinity of the ligand.

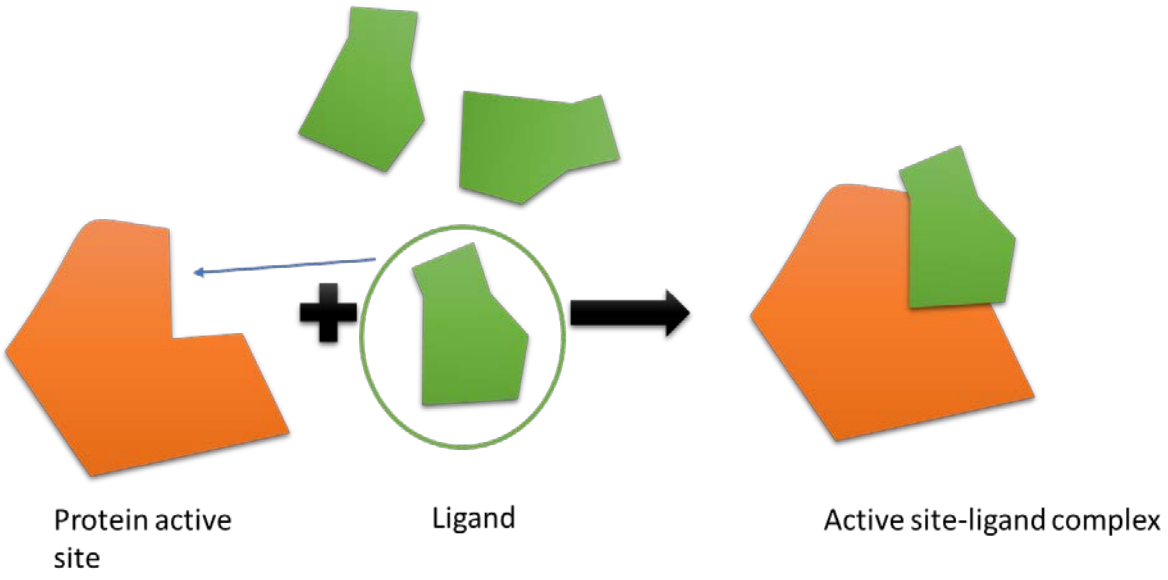


Figure 11: Schematic illustration of the docking of a molecule/ligand (green) to the active site of a target protein to form a complex. Docking studies would highlight the ligand in the optimum orientation and conformation (circled).

Different forces affect how the ligand binds to the active site of the target. Some include electrostatic forces which are charges that reside in both ligand and active site. Steric forces are influenced by entropy which is where the complex naturally seeks to find the conformation with the least free energy. Solvent-related forces come from structural changes of the solvent and occur when things like ions and proteins bring about structural changes to the solvent. Such forces include hydrogen bond and hydrophobic interactions.

Docking studies occur in a number of stages. Firstly, a three-dimensional structure of the receptor should be built. An example of a source for such structures is the PDB and if not available online, X-ray crystallography of the protein can be used. When the structure is obtained, it is essential for the active sites within the receptor to be identified. Some proteins may have multiple active sites but the one of pertinence should be located. The next stage is the ligand selection and preparation. Ligands can be created by the researcher or obtained from databases like PubChem and ZINC. Once this library has been collected, docking can begin. Using software like AUTODOCK and GRAMM, the ligand is docked unto the active and site and the molecular interactions are monitored. A function score or binding affinity is given and from this, molecules in the library can be ranked. Due to this, docking can be used to screen out non-active molecules or those with particularly low binding affinities. It can also be used in optimisation to find substitutions that can be made to the lead compounds to make more selective and potent analogues.

Another important process used through computation methods is virtual screening. This is the process of taking a large library of available small compounds and evaluating their bioactivity computationally. Different software such as DOCK, KNIME and Pharmit can use a scoring method (either pre-set or customised) to rank the molecules in the library to highlight the most bioactive. Software like Pharmit use pharmacophore studies to score and rank the molecules.

It was once thought that particular functional groups in a molecule were responsible for the molecules biological effect, but the term pharmacophore was coined by Scheuler who defines it as "a molecular framework that carries the essential features responsible for the drug's biological activity.".[60] However, a pharmacophore is more an abstract concept and should not be reduced to just functional groups like guanidines or to scaffolds like prostaglandins. The use of pharmacophores has become very important in CADD. Here, molecular patterns are labelled according to their features such as being anionic, aromatic or cationic. Through this, different molecules can be compared using these pharmacophore labels rather than using actual atoms or functional groups.

With these patterns, a pharmacophore model can be built. Most models have spheres to highlight each feature or label with the size of each sphere representing the room for deviation from the exact position. This deviation from the exact position can be used as a method for ranking the molecules in a library and this property is called the root mean square deviation (RMSD). By coupling the RMSD score with docking scores, it possible to screen through libraries to find molecules with a high binding affinity that do not deviate too much from the original position of the lead compound.

2.5.2 Structures in Drug Discovery

When using structure-targeted methods for finding new lead compounds, it is imperative that a correct and accurate structure is used- especially in mechanisms concerning ligand binding, enzymatic functions and protein-protein interactions. Techniques such as x-ray crystallography serve as a powerful method to determine structures of target proteins although they provide static structures and the process of making the proteins into crystals may be at times very difficult. In this research, the X-ray structure for SRPK1 plays a significant role when conducting pharmacophore studies.

3. Project Aims

The process of angiogenesis is essential to survival and when the process is exploited through upregulation or suppression, it can a very detrimental effect on life. Cancers and neovascular diseases are examples where the effect of the upregulation of angiogenesis can be seen. Chemotherapies can be made to target and inhibit pro-angiogenic factors like SRPK1. A current molecule known for inhibiting is SRPK1 is SRPIN340. Whilst this shows promise, SRPIN340 lacks hydrophilicity which makes it hard for oral administration. It also is not very potent meaning a higher dose are needed for a therapeutic effect which can consequently bring about toxicities.

This project aims to create novel compounds that can serve as a more potent and hydrophilic substitute for SRPIN340. This will be done through pharmacophore and docking studies using made libraries from ZINC and Swiss Bioisostere. From this process, a few molecules will be chosen for synthesis. After synthesis, the molecules will be complexed with SRPK1 to establish inhibition. The SRPK1 used will be induced and harvested from transformed *E. coli* BL21(DE3) cells. Toxicity studies of both SRPIN340 and the newly synthesised molecules will be done using two strains of breast cancer cells, so comparisons can be made.

4. Experimental

4.1 Protein Expression and Purification

The materials used to make the media, buffers and other compounds were bought from Sigma-Aldrich and Fisher Scientific. The BL21(DE3) *E. coli* were bought from Novagen whilst the recombinant plasmids were donated by the J. Adams research group. All bacterial growth and protein expression procedures were carried out in sterile conditions with the use of a Bioquell PLC Microflow biological safety cabinet. The protocols used in cell culturing and protein expression and purification were adapted from Aubol *et. al*[61]

Medium Preparation

In this process, Luria Broth (LB) was made from NaCl (10 g/L), yeast extract (5 g/L) and Tryptone (20 g/L) with the pH adjusted to 7.5. Conical flasks were each filled with about 200 mL of media and sealed using a bung and aluminium foil. The flasks were then autoclaved at 121°C, 15 psi for 20 minutes for sterilisation.

Cell Transformation and Culturing

The recombinant plasmids encoding the SRPK1 gene were transformed into the *E. coli* BL21(DE3) cells through heat shock and grown on LB agar plate containing the antibiotic carbenicillin at 37°C overnight. Single colonies were taken from the plate and added to 10 mL of LB media containing carbenicillin (100 mg/mL) which were in 50mL falcon tubes. The colonies were then grown in agitated conditions at 220 rpm and 37°C. They were left to grow until the media became opaque in colour. 3mL of this was added to every 200 mL portion of media in conical flasks with added 200 µg carbenicillin (100 mg/mL). The flasks were then put into an incubator to stir at 220 rpm and 37°C until the OD_{500} was about 0.7

Protein Expression

1 mL of a 100 mM stock solution of isopropyl B-D thiogalactopyranoside was added to each of the conical flasks which induced protein expression. The flasks were then left to be shaken at the same speed but with a temperature of 16^{o}C overnight. The cultures were then collected and centrifuged for 10 minutes at 1400 rpm.

Cell Harvesting and Lysis

The suspended cell pellet was then thawed and centrifuged for 15 minutes at 8000 rpm and the pellet was collected and suspended in lysis buffer (20 mM MES, 50 mM NaCl, 20 glycerol, pH 6.5, 1mM PMSF). This was then sonicated for 20 minutes on a 5s on: 5s off cycle whilst being kept on ice, the lysate solution was then centrifuged for 10 minutes at 14000 rpm. The supernatant was then collected for purification

Protein Purification

1L each of a wash buffer (20 mM MES, 50 mM NaCl, pH 7.5, 5% glycerol, 5 mM imidazole) and an elution buffer (20 mM MES, 50 mM NaCl, pH 7.5, 5% glycerol, 1 M imidazole) were made, filtered and degassed. The collected supernatant was then applied to a 1mL Ni^{2+} Sepharose (HisTrap excel) column attached to an AKTA FPLC. The fractions with the desired protein were collected and then mixed with some storage buffer (50 mM Tris- HCl, pH 7.5, 150 mM NaCl, 0.25 mM Dithiothreitol (DTT), 0.25 mM Ethylenediaminetetraacetic acid (EDTA), 0.1 mM Phenylmethylsulfonyl fluoride (PMSF), 25% glycerol).

SDS PAGE- Identification

20 µL of the collected protein fraction and 20 µL of loading buffer were mixed in 0.5 mL Eppendorf tubes. The mixture was then boiled at 90^{o}C for ten minutes. This was then loaded onto a polyacrylamide gel which was run at 120V and 400 mA for around 100 minutes in an anode and cathode buffer.

The gel was then stained with Coomassie blue staining solution for 30 minutes and then destained by heating it in water in a microwave until most of the dye was removed.

All buffers and solutions required for the SDS Page are as follows:

- Loading buffer: 100 mL (9.46 g trizma base, 1 mL mercaptoethanol, 1 g SDS, 10 mL glycerol, 0.01 g bromophenol blue, pH 6.8)
- Anode buffer: 500 mL (12.11 g trizma base, pH 8.9)
- Cathode buffer: 500 mL (6.055 g trizma base, 6.96 g tricine, 0.5 g SDS, pH 8.25)
- Staining solution: 500 mL (1.25 g Coomassie blue, 225 mL ethanol, 50 mL glacial acetic acid)
- Gel buffer: 500 mL (181.7 g trizma base, 1.5 g SDS, pH 8.45)
- Gel page made with:

Table 1: Table showing components for SDS page gels

Component	**Stacking gel (4%)**	**Separating gel (10%)**
Gel Buffer (mL)	2.5	3.33
40% acrylamide (mL)	1	2.5
Distilled water (mL)	3.25	2.8
Glycerol (mL)	-	1.3
TEMED (µL)	25	25
10% Ammonium persulfate (AMPS) (µL)	25	25

4.2 SRPIN340 Toxicity Testing – Breast Cancer

4.2.1 Initial Cell Culturing

MC47 and MDA breast cancer strains donated by breast cancer patients were collected and washed with phosphate-buffered saline (PBS), pH 7.4 in a cell culture flask. The cells were then washed with trypsin and left in a CO_2 incubator for about 5 minutes to detach the cells from the flask walls. A microscope was used to confirm cell detachment. The cells were washed again with PBS to inhibit the function of the trypsin. The cells were then removed from the flasks and centrifuged at 200 x g for five minutes. The supernatant was discarded, and the cells were kept. 10 µL of each strain was extracted and stained with 10µL trypan blue cell stain and then counted in an automated cell counter. To get more accurate results, the cell suspension was diluted so an estimated 5000 cells were present in each well as shown in Table 2.

Table 2: Live cell counts, and calculations used for the determination of cell concentration

Cell Strain	Number of live cells (10^6/ mL)	Concentration required (10^4/ mL)	Number of cells required (µL)	Amount of medium required (mL)
MCF7	5.4	5	140	14.86
MDA	1.7	5	440	14.56

The MDA strain was mixed with Dulbecco's modified Eagle medium (DMEM) whilst the MC47 strain was mixed with minimal essential medium (MEM). A 12x8 well plate was chosen for each strain and each well was filled with 100 µL of cells. These well plates were then left in a CO_2 incubator overnight.

Both cell strain media were discarded and substituted with a 100 µL of 5% foetal bovine solution (FBS) and then both strains were incubated in CO_2. Eleven solutions of differing SRPIN340 concentrations were prepared through serial dilutions with the starting amount being 50 µM of SRPIN340 dissolved in 0.1% dimethyl sulfoxide (DMSO) in water. The published IC_{50} value of SRPIN340 is 0.96 µM, therefore, all concentrations needed to be higher.[62] The solutions were then added to the wells are illustrated in Figure 12. The plates were left to incubate for three days. After the incubation, half of the wells were stained with calcein acetoxymethyl (calcein-AM) and the other half with Alamar Blue. These dyes were chosen as they fluoresce when they come into contact with live cells. A microplate reader (CLARIOstar) was used to measure the fluorescence to come with estimated concentrations of live cells. The results can be seen in Figure 27.

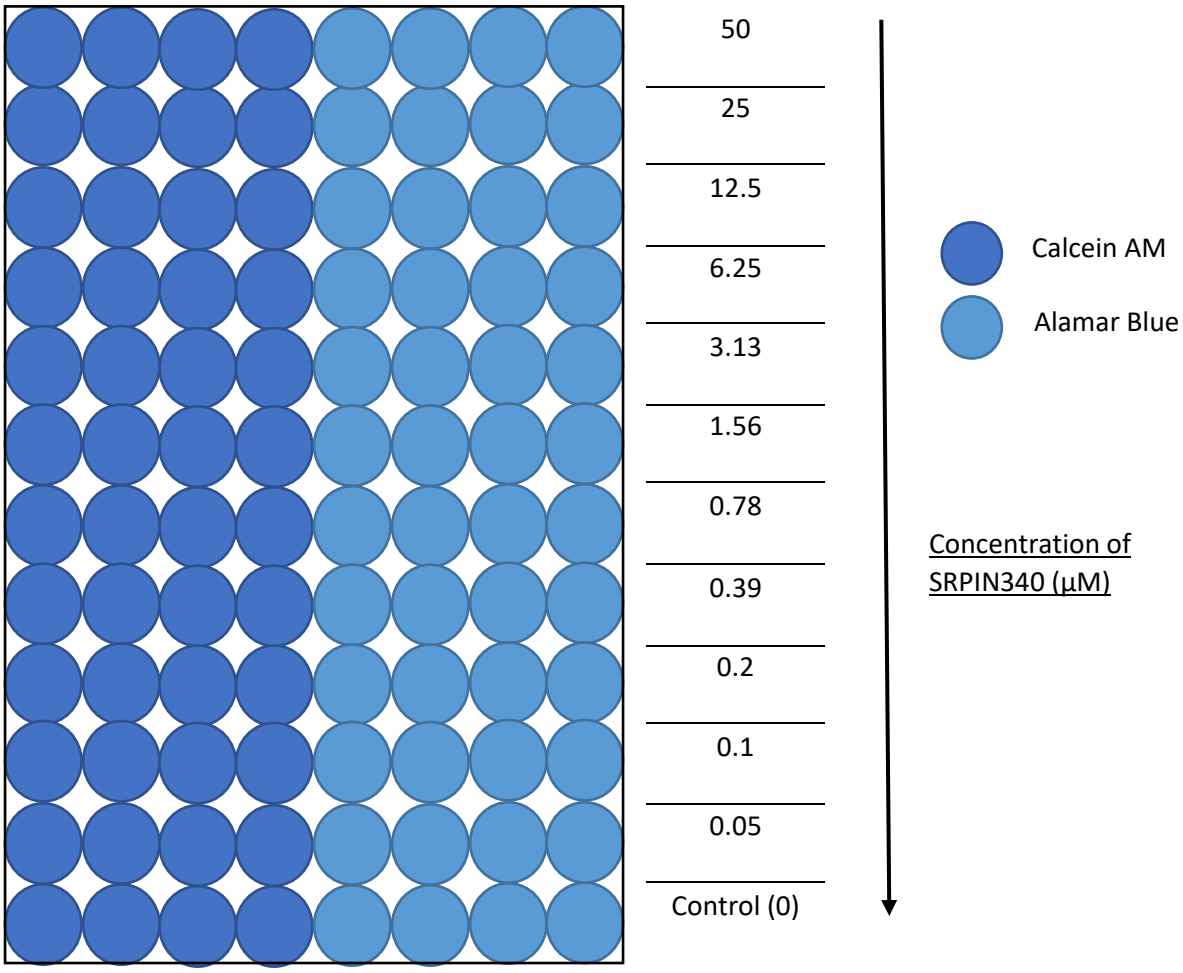


Figure 12: Illustration of cell wells, cell concentrations and cell stains

4.3 Synthesis

4.3.1 General Reaction Mechanism

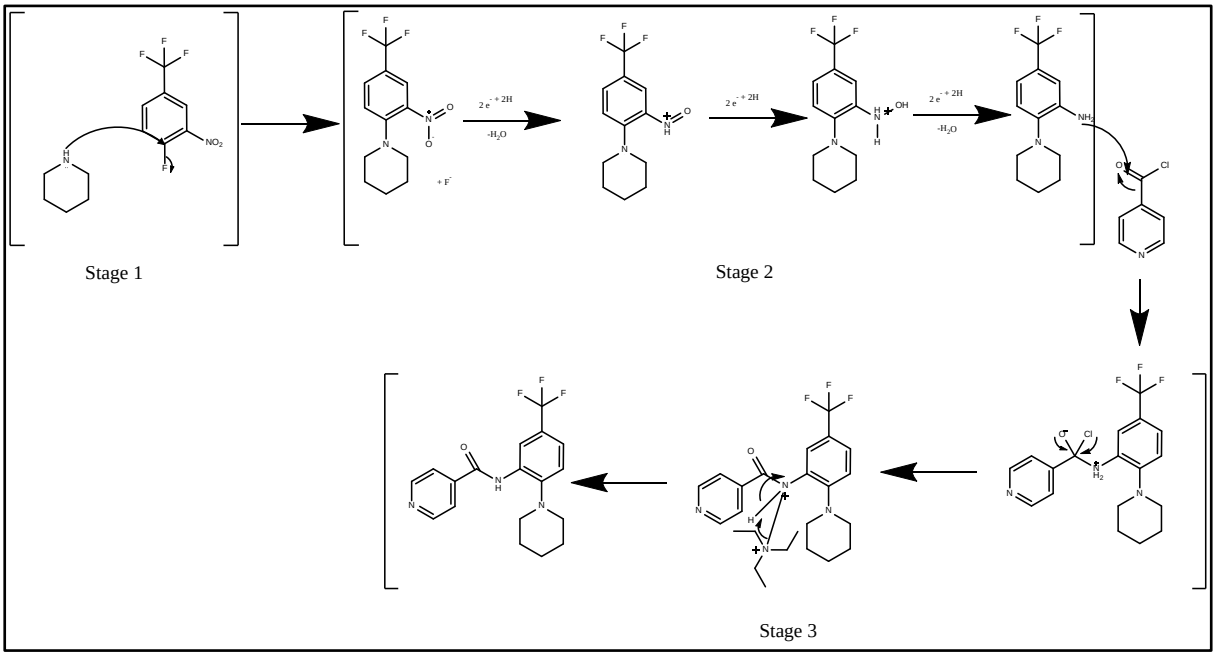


Figure 13: Reaction mechanism for the synthesis of SRPIN340

Figure 13 outlines the reaction mechanism of the synthesis of SRPK1 and the consequent synthesis of the chosen novel compounds. Stage one consists of the nucleophilic substitution of the fluoride ion in the 1- fluoro-2-nitro-4-(trifluoromethyl)benzene molecule with either piperidine or the other nucleophiles needed for the synthesis of the selected compounds. Stage 2 is the reduction of the nitro group to an amine group which is necessary for stage 3. The final stage will be the nucleophilic attack of the acyl chloride by the amine with triethylamine being present to mop up and residual protons. Since all the reactants for each synthesis fall in the same molecular groups, the mechanisms should be fairly similar.

4.3.2 Synthesis of SRPIN340:

The reaction protocol used was copied from the procedure stated by Fukuhara *et al.* in the supplementary information from the original paper.[20] The scheme has been illustrated in Figure 14.

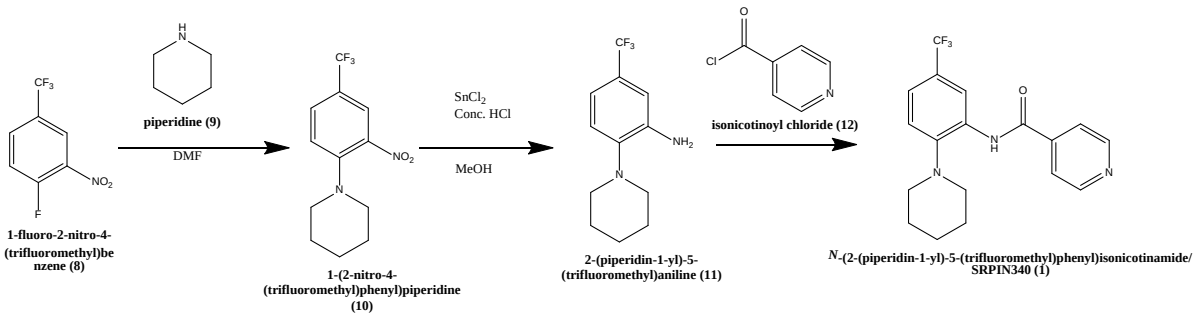


Figure 14: Reaction scheme for the synthesis of SRPIN340

- **8-10**: **9** (440 µL, 4.4 mmol) and **8** (570 µL, 2.6 mmol) were combined and dissolved in DMF (2 mL) at room temperature and stirred for an hour. Water was added to the mixture. The solution then underwent extraction with diethyl ether with the combined organic layers being washed with brine, dried with magnesium sulphate then being filtered and dried *in vacuo*. Compound **10** (1-[2-nitro-4- trifluoromethyl)phenyl] piperidine) is an orange solid (1 g, 3.64 mmol, 90% yield). 1H NMR (400 MHz, Chloroform-d) δ 8.03 (dd, J = 2.3, 0.9 Hz, 1H), 7.65 – 7.56 (m, 1H), 7.13 (d, J = 8.8 Hz, 1H), 3.12 (dd, J = 6.4, 4.2 Hz, 5H), 1.77 – 1.59 (m, 5H).

- **10-11**: Anhydrous tin dichloride (3.5 g, 9.3 mmol), **10** (680 mg, 2.4 mmol), concentrated HCl (2.8 mL, 17 mmol) and methanol (20 mL) at 0 ^{o}C were mixed in a round bottom flask. The mixture was allowed to reach room temperature and was stirred overnight. The acid was quenched with saturated sodium bicarbonate solution and the product was extracted with ethyl acetate. The combined collected organic layers were washed with brine and dried with magnesium sulphate, filtered and dried *in vacuo*.

Compound **11** {1-[2-amino-4-(trifluoromethyl) phenyl] piperidine} is a cream coloured solid (0.5 g, 2.04 mmol, 77% yield). ^{1}H NMR ($CDCl_3$, 400 MHz) δ ^{1}H NMR (400 MHz, Chloroform-*d*) δ 7.04 – 6.90 (m, 4H), 4.06 (s, 2H), 2.85 (d, *J* = 6.1 Hz, 4H), 1.71 (q, *J* = 5.6 Hz, 4H), 1.26 (s, 1H).

- **11-1**: Triethylamine (0.6 mL, 3 mmol), **12** (1.8 g, 2.04 mmol) and **11** (500 mg, 1 mmol) were dissolved in methanol (5 mL) at 0^{o}C and left to stir for 2 hours. Water was added, and the product was extracted thrice with ethyl acetate. The combined organic layers were washed with brine, dried with magnesium sulphate, filtered then dried *in vacuo*. The product was then purified using a silica coloumn. Compound **11**: N-[2-(1-piperidinyl)-5-(trifluoromethyl)phenyl] isonicotinamide (SRPIN340) is a white solid (0.62 g, 1.78 mmol, 87.32% yield)). ^{1}H NMR (400 MHz, Chloroform-*d*) δ 9.53 (s, 1H), 8.90 – 8.83 (m, 3H), 7.80 – 7.73 (m, 2H), 7.39 (dd, *J* = 8.3, 2.0 Hz, 1H), 7.28 (d, *J* = 16.8 Hz, 2H), 1.78 (p, *J* = 5.5 Hz, 5H), 1.71 – 1.64 (m, 3H).

4.3.3 Synthesis of 4-fluoro-3-methyl-N-(2-(2-methyl-3-oxopiperazin-1-yl)-5-(trifluoromethyl)phenyl) benzamide

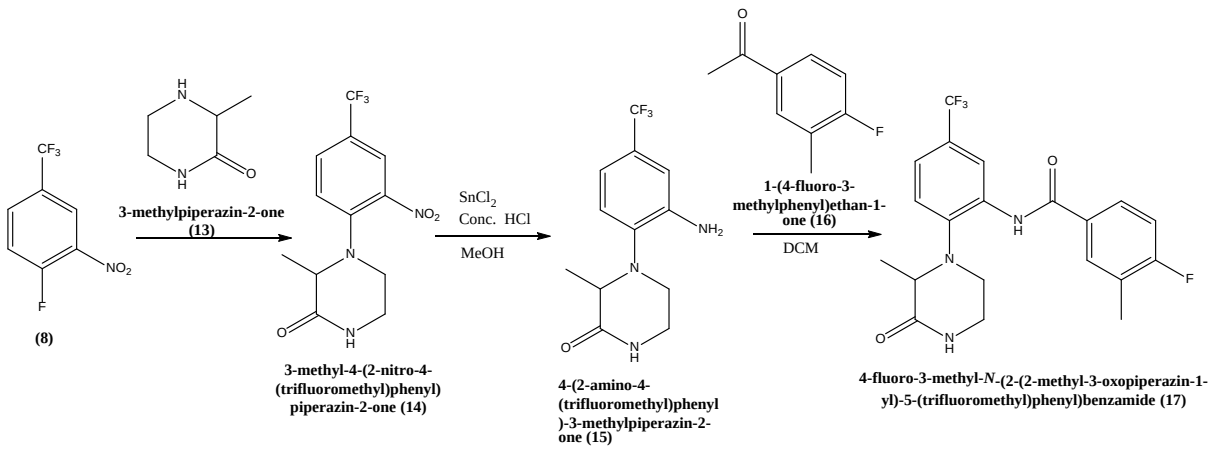


Figure 15: Reaction scheme of the Synthesis of 4-fluoro-3-methyl-N-(2-(2-methyl-3-oxopiperazin-1-yl)-5-(trifluoromethyl)phenyl) benzamide

- **8-14**: **13** (0.2 g, 1.8 mmol) and **8** (0.19 g, 0.9 mmol) were dissolved in DMF (2 mL) at room temperature and stirred for an hour. Water was added to the solution and the product was extracted with diethyl ether with the combined organic layers being washed with brine, dried with magnesium sulphate then being filtered and dried *in vacuo*. Compound **14** [3-methyl-4-(2-nitro-4-(trifluoromethyl) phenyl) piperazin-2-one] is a yellow solid (0.14 g, 0.461 mmol, 51.8% yield) 1H NMR (400 MHz, Chloroform-d) δ 8.31 (s, 1H), 8.02 – 7.79 (m, 3H), 7.28 (d, J = 18.6 Hz, 3H), 6.40 (s, 1H), 4.15 (q, J = 7.2, 6.3 Hz, 2H), 3.56 – 3.32 (m, 3H), 3.16 – 2.96 (m, 2H), 1.46 – 1.26 (m, 5H).

- **14-15**: Compound **14** (0.14 g, 0.5 mmol) anhydrous tin dichloride (2.3 g, 12.16 mmol) concentrated HCl (2 mL) and methanol (10 mL) at 0^{o}C were mixed in a round bottom flask. The mixture was allowed to reach room temperature and heated at reflux for 30 minutes. The acid was quenched with saturated sodium bicarbonate solution and the product was extracted with ethyl acetate. The combined collected organic layers were washed with brine and dried with magnesium sulphate, filtered and dried *in vacuo*. Compound **15** [4-(2-amino-4-(trifluoromethyl) phenyl)-3-methylpiperazin-2-one] is a cream coloured solid (0.14g, 0.29 mmol, 57.14%) 1H NMR (400 MHz, Chloroform-d) δ 7.26 (s, 1H), 7.04 (s, 0H), 6.97 (s, 1H), 4.11 (q, J = 7.2 Hz, 1H), 3.65 (s, 0H), 3.47 (q, J = 7.0 Hz, 34H), 1.42 (s, 10H).

  15-17: Triethylamine (0.6 mL, 3 mmol) **15** (0.06g, 0.22 mmol) and **16** (0.06 g, 0.4 mmol) were dissolved in methanol (5 mL) at 0^{o}C and left to stir for 2 hours. Water was added, and the product was extracted thrice with ethyl acetate. The combined organic layers were washed with brine, dried with magnesium sulphate, filtered then dried *in vacuo*. The product was then purified using a silica gel chromatography (1:1 ethyl acetate: hexane). Compound **17**: 4-fluoro-3-methyl-N-(2-(2-methyl-3-oxopiperazin-1-yl)-5-(trifluoromethyl)phenyl) benzamide appeared to be a white solid but NMR and Mass spectra indicated this was not made.

4.3.4 Synthesis of 2,3-difluoro-4-methyl-N-(2-(2-methyl-3-oxopiperazin-1-yl)-5-(trifluoromethyl)phenyl) benzamide

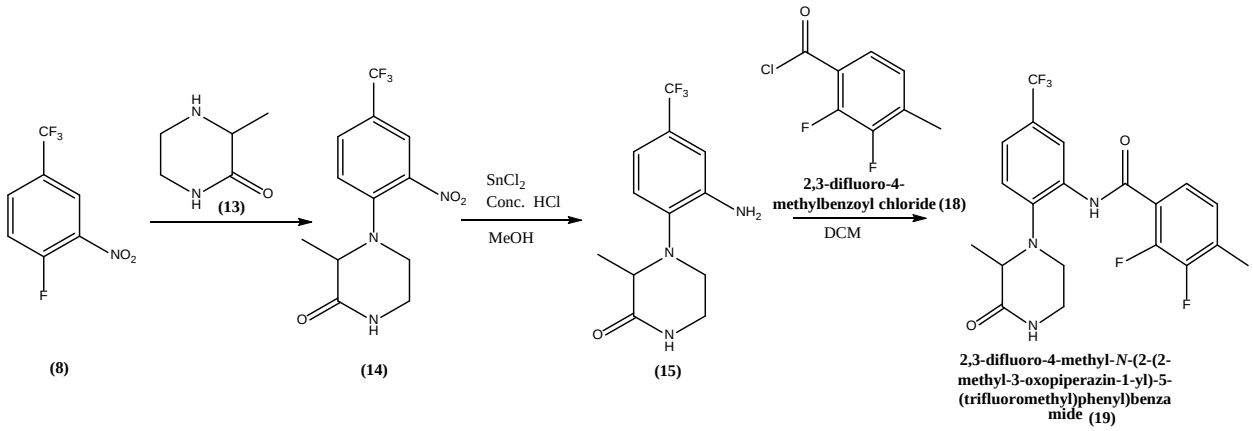


Figure 16: Reaction scheme of the synthesis of 2,3-difluoro-4-methyl-N-(2-(2-methyl-3-oxopiperazin-1-yl)-5-(trifluoromethyl)phenyl) benzamide

- **15-19**: Half of the amount of **15** created in the synthesis of **17** was used in this process. Triethylamine (0.6 mL, 3 mmol), **18** (0.07g, 0.4 mmol) and **15** (0.06g, 0.22 mmol) were dissolved in methanol (5 mL) at 0^{o}C and left to stir for 2 hours. Water was added, and the product was extracted thrice with ethyl acetate. The combined organic layers were washed with brine, dried with magnesium sulphate, filtered then dried *in vacuo*. The product was then purified using a silica gel chromatography (1:1 ethyl acetate: hexane). Compound **19**: 2,3-difluoro-4-methyl-N-(2-(2-methyl-3-oxopiperazin-1-yl)-5-(trifluoromethyl)phenyl) benzamide) is a white solid but NMR and mass spectra indicated that this was not made.

4.3.5 Synthesis of 2,3-difluoro-N-(2-(hexahydropyrrolo[1,2-a] pyrazin-2(1H)-yl)-5-(trifluoromethyl) phenyl)-4-methylbenzamide

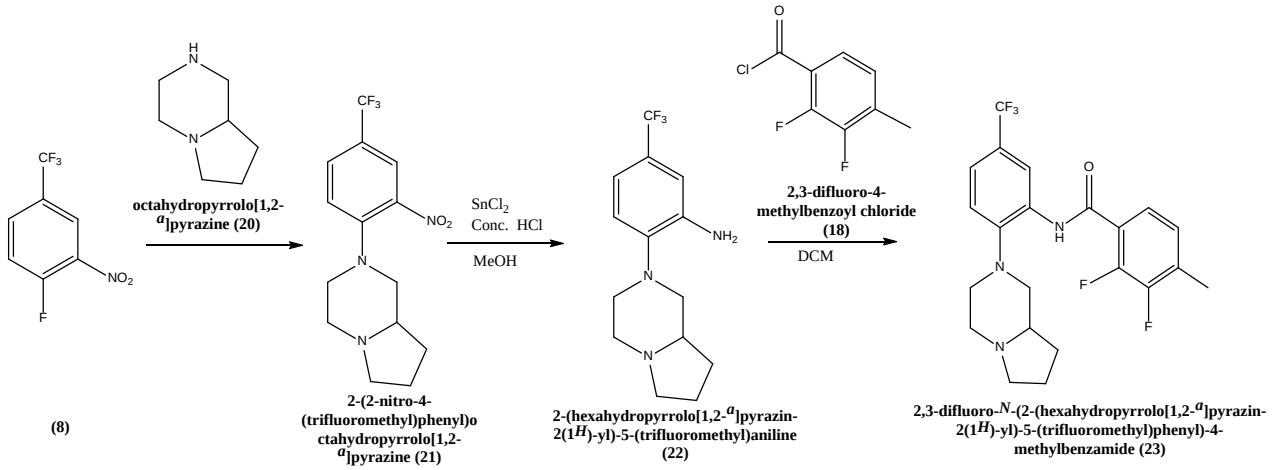


Figure 17: Reaction scheme of the synthesis Synthesis of 2,3-difluoro-N-(2-(hexahydropyrrolo[1,2-a] pyrazin-2(1H)-yl)-5-(trifluoromethyl) phenyl)-4-methylbenzamide

- **8-21**: **20** (0.15 g, 1.18 mmol) and **8** (0.12g, 0.57 mmol) were dissolved in DMF (2 mL) at room temperature and stirred for an hour. Water was added to the solution and the product was extracted with diethyl ether with the combineCF3d organic layers being washed with brine, dried with magnesium sulphate then being filtered and dried *in vacuo*. **21**: 2-(2-nitro-4-(trifluoromethyl) phenyl) octahydropyrrolo [1,2-a] pyrazine is a dark orange solid (0.13 g, 0.41 mmol, 35% yield) 1H NMR (400 MHz, Chloroform-d) δ 7.99 (s, 1H), 7.63 (dd, J = 8.8, 2.2 Hz, 1H), 7.17 (d, J = 8.8 Hz, 1H), 3.95 (s, 0H), 3.41 – 3.04 (m, 4H), 2.89 (d, J = 30.7 Hz, 4H), 2.48 (td, J = 11.0, 3.3 Hz, 1H), 2.29 (q, J = 8.8 Hz, 1H), 2.14 (d, J = 0.8 Hz, 1H), 1.96 – 1.69 (m, 2H), 1.54 – 1.39 (m, 1H).

- **21-22**: **21** (0.13 g, 0.41 mmol, anhydrous tin dichloride (2.3g, 12.16 mmol), concentrated HCl (2 mL) and methanol (10 mL) at 0^{o}C were mixed in a round bottom flask. The mixture was allowed to reach room temperature and heated at reflux for 30 minutes. The acid was quenched with saturated sodium bicarbonate solution and the product was extracted with ethyl acetate. The combined collected organic layers were washed with brine and dried with magnesium sulphate, filtered and dried *in vacuo*.

  Compound **22**: 2-(hexahydropyrrolo[1,2-a] pyrazin-2(1H)-yl)-5-(trifluoromethyl)aniline is a light yellow coloured solid (0.08, 0.28 mmol, 72% yield) 1H NMR (400 MHz, Chloroform-d) δ 7.16 – 6.85 (m, 1H), 4.23 – 3.91 (m, 1H), 3.46 (q, J = 7.0 Hz, 0H), 3.23 (dt, J = 10.9, 2.2 Hz, 0H), 3.16 – 3.05 (m, 1H), 2.85 (td, J = 11.5, 11.0, 3.0 Hz, 0H), 2.54 (dd, J = 10.9, 9.8 Hz, 0H), 2.40 (td, J = 11.7, 11.1, 3.1 Hz, 0H), 2.31 – 2.12 (m, 1H), 2.02 (s, 0H), 1.93 – 1.68 (m, 1H), 1.59 – 1.35 (m, 0H), 1.21 (dt, J = 19.7, 7.1 Hz, 1H).

- **22-23**: Triethylamine (0.6 mL, 3 mmol)**, 18** (0.05 g, 0.3 mmol) and **22** (0.04, 0.14 mmol) were dissolved in DCM (5 mL) at 0^{o}C and left to stir for 2 hours. Water was added, and the product was extracted thrice with ethyl acetate. The combined organic layers were washed with brine, dried with magnesium sulphate, filtered then dried *in vacuo*. The product was then purified using a silica gel chromatography (1:1 ethyl acetate:hexane). Compound **23**: 2,3-difluoro-N-(2-(hexahydropyrrolo[1,2-a] pyrazin-2(1H)-yl)-5-(trifluoromethyl) phenyl)-4-methylbenzamide is a brown/ orange solid (0.04 g, 0.09 mmol, 66.66% yield). 1H NMR (400 MHz, Chloroform-d) δ 7.56 (ddd, J = 8.4, 6.5, 1.9 Hz, 2H), 7.01 – 6.85 (m, 2H), 4.37 (q, J = 7.1 Hz, 3H), 3.43 (d, J = 13.8 Hz, 0H), 3.30 – 3.15 (m, 0H), 3.10 (q, J = 7.3 Hz, 2H), 2.41 – 2.22 (m, 5H), 1.38 (td, J = 7.2, 3.6 Hz, 7H), 1.30 – 1.17 (m, 1H).

4.3.6 Synthesis of 4-fluoro-N-(2-(hexahydropyrrolo[1,2-a] pyrazin-2(1H)-yl)-5-(trifluoromethyl) phenyl)-3-methylbenzamide

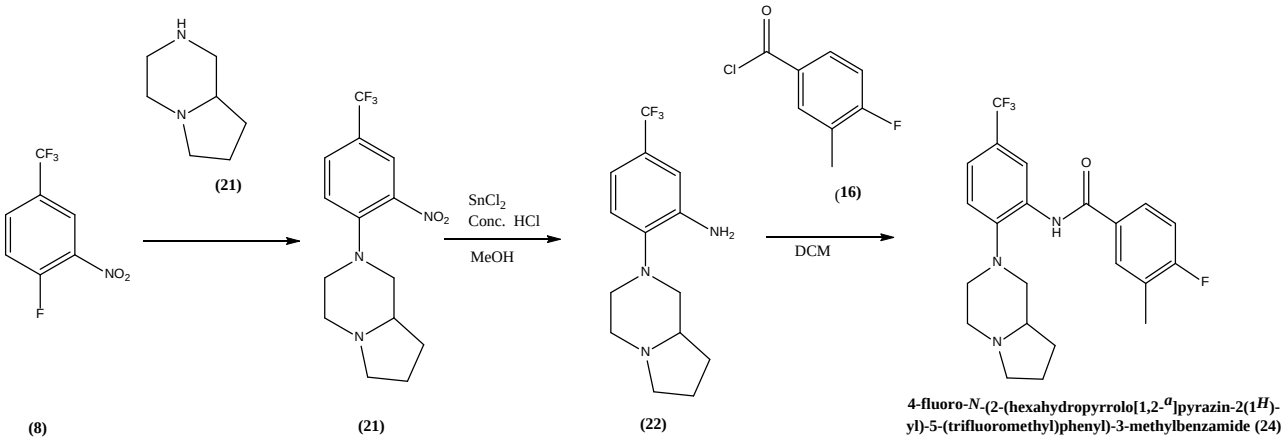


Figure 18: Reaction scheme of the synthesis of 4-fluoro-N-(2-(hexahydropyrrolo[1,2-a] pyrazin-2(1H)-yl)-5-(trifluoromethyl) phenyl)-3-methylbenzamide

- **22-24**: Taking half of the amount of **22** created in the synthesis of **23** was used in this process. Triethylamine (0.6 mL, 3 mmol), **16** (0.05 g,0.3 mmol) and **22** (0.04 g, 0.14 mmol) were dissolved in methanol (5 mL) at 0^{o}C and left to stir for 2 hours. Water was added and the product was extracted thrice with ethyl acetate. The combined organic layers were washed with brine, dried with magnesium sulphate, filtered then dried *in vacuo*. The product was then purified using a silica gel chromatography (1:1 ethyl acetate: hexane). Compound **24**: 4-fluoro-N-(2-(hexahydropyrrolo[1,2-a]pyrazin-2(1H)-yl)-5-(trifluoromethyl)phenyl)-3-methylbenzamide is a brown solid (0.021, 0.05 mmol, 42% yield) 1H NMR (400 MHz, Chloroform-d) δ 10.31 (s, 1H, NH), 7.95 – 7.83 (m, 2H, 2CH), 7.33 (d, J = 1.4 Hz, 1H, CH), 7.13 (q, J = 8.7, 7.1 Hz, 1H, CH), 7.07 – 6.90 (m, 2H, 2CH), 3.52 – 3.20 (m, 6H, 3CH_2), 3.16 (qd, J = 7.3, 1.2 Hz, 2H, CH_2), 2.34 (s, 2H, CH_2), 2.27 (d, J = 1.8 Hz, 2H, CH_3, CH), 2.13 – 1.91 (m, 4H, 2CH_2), 1.23 (td, J = 7.1, 1.1 Hz, 2H, CH_2).

5. Results and Discussion

5.1 Protein Harvesting and Characterisation

For this research, the SRPK1 protein needed to be harvested. This was to test experimentally how well the novel compounds bound to the protein. The procedure for how the harvesting was achieved is outlined in the previous chapter. After the SRPK1 protein was harvested from the *E. coli*, it was run through an SDS page and also through an LC/MS to characterise the protein. The protein from the plasmid used was only a partial version of the original protein. The plasmid that was donated from the Adams group was a truncated version. In the full-length SRPK1 protein, there is the generic kinase core as well as a long spacer region found between the two lobes of the core. It was found by Ngo *et al*. that the purification of the full-length protein caused its degradation as there was a cleave that occurred in the spacer region.[63] When this space region was removed, Ngo *et al.* found that purification was then possible without compromising its kinase activity. It was the newly truncated version (SRPK1ΔNS3) that was donated in the plasmid. The comparison between the full length and the truncated is illustrated in figure 19.

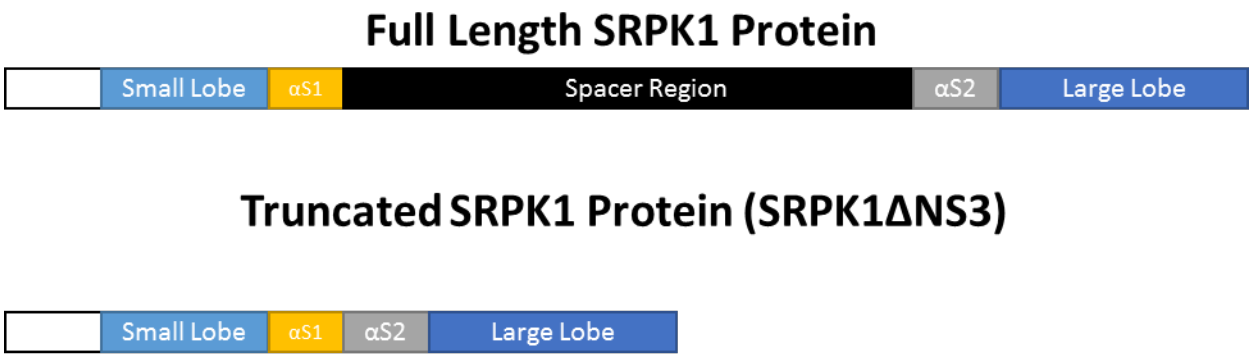


Figure 19: Illustrated comparison of the full length and truncated SRPK1 protein

Unfortunately, the papers by the Adams group featuring this truncated protein did not mention its weight. The work by Ngo came with a different weight usually around the 40 kDa weight. This was particularly interesting because when the SDS page and mass spectrometry was conducted, a protein weight of 35 kDa was deduced. The SDS page can be seen in figure 20 and the mass spectrum data in figure 21.

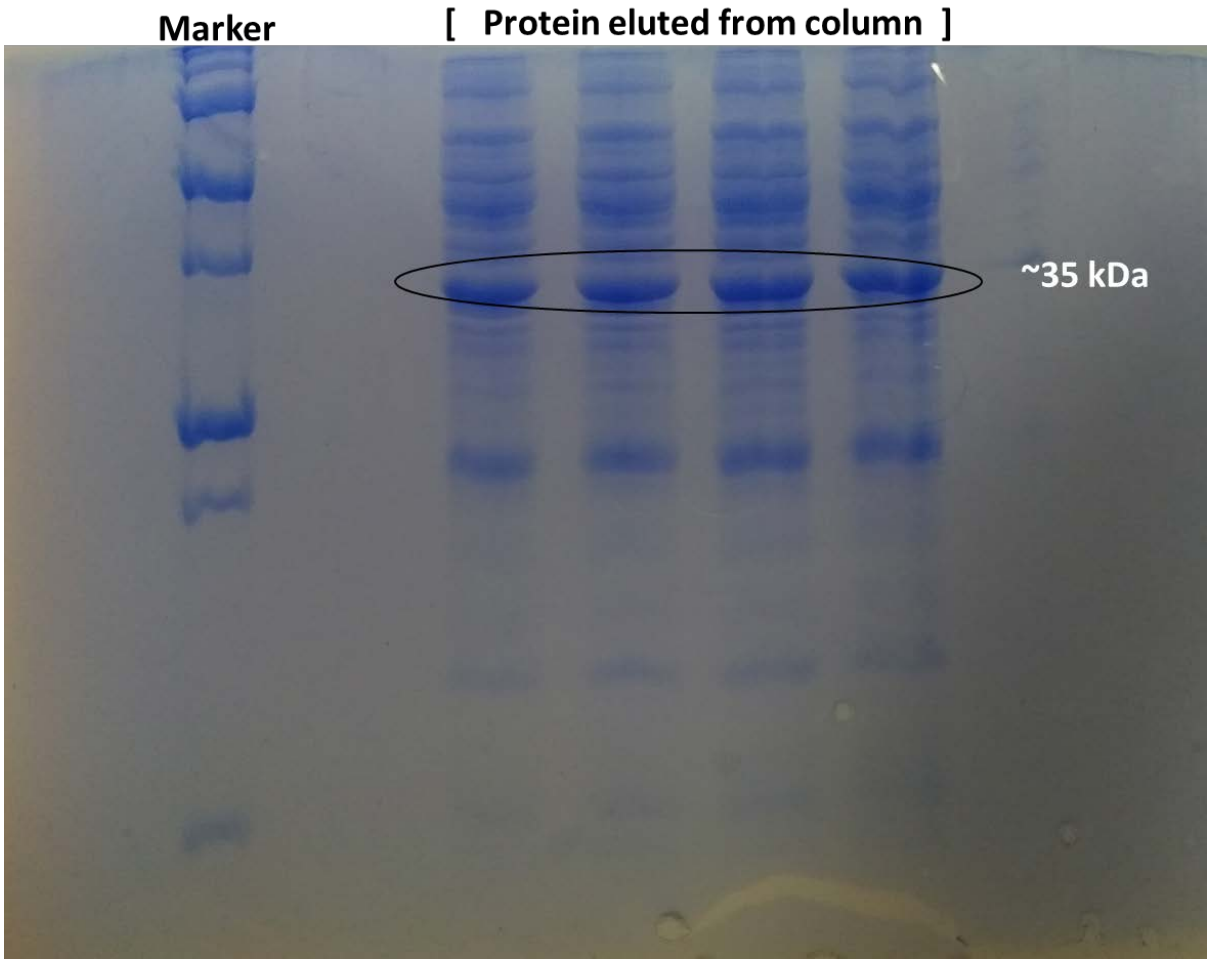


Figure 20: Run SDS page of eluted protein.

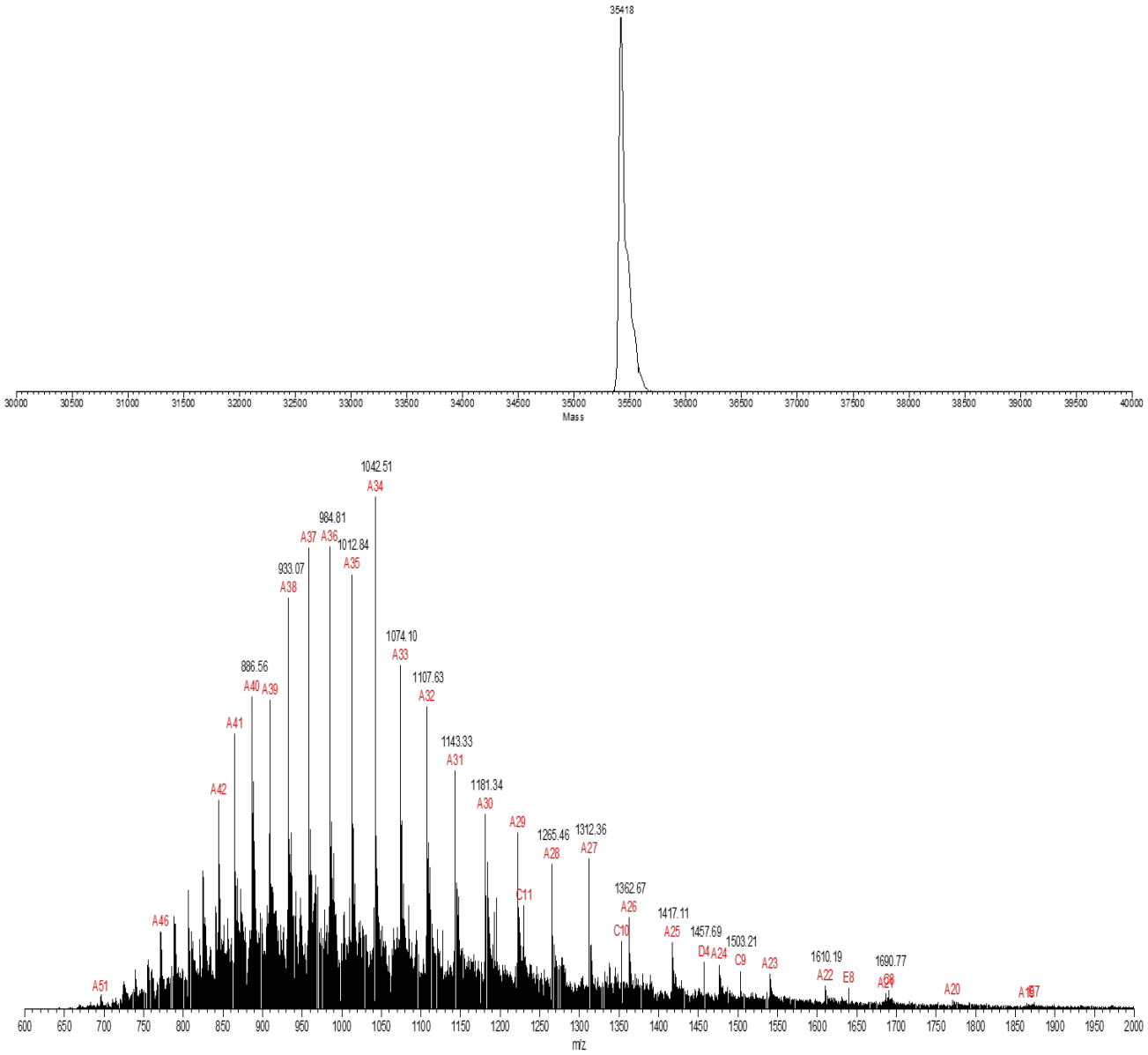


Figure 21: Mass spectra following protein purification. As seen in the top, seems to be clear peak showing mass of 35418 Da. This could ASF2

Further reading and analyses show that the protein being expressed with the weight of 35 kDa might be Alternative Splicing Factor 2 (ASF2). It has been confirmed to have a weight of around 33 kDa.[64] It is a protein necessary in most alternative splicing and acts in mRNA translation and nuclear export post-slicing.[65] The expression of this protein is not a bad result. Though the preferred protein would have been SRPK1, work that was planned would have tested the practical interactions between the synthesised molecules and both SRPK1 and ASF2 since both proteins often work in tandem.

As to why little to no levels of SRPK1 was produced, this could be perhaps be explained in the FPLC purification step. Since ASF2 is the smaller protein, it is more like to elute at a faster time compared to SRPK1 if assuming both proteins have similar polarities. If this process was to be repeated, a longer run time to look for a second peak which might indicate the presence of a second protein. A second run would have to be done perhaps using size exclusion chromatography since the weight is known. Also, as seen in the SDS page, the collected protein was not pure. This could have been due to there being other histidine-tagged or histidine-rich proteins from a previous use trapped in the column even after it had been washed.

5.2 Computational Methods

Computational methods were used to find potential new lead compounds which could be more potent and more effective than SRPK1. The structure of SRPIN340 was used as a scaffold and the molecule was split into three fragments as seen in figure 22. Different libraries were created to find substitutions for each fragment then these libraries were combined and put through the software KNIME which created a new library made from different combinations of the fragments. Outlined below is the process of discovery.

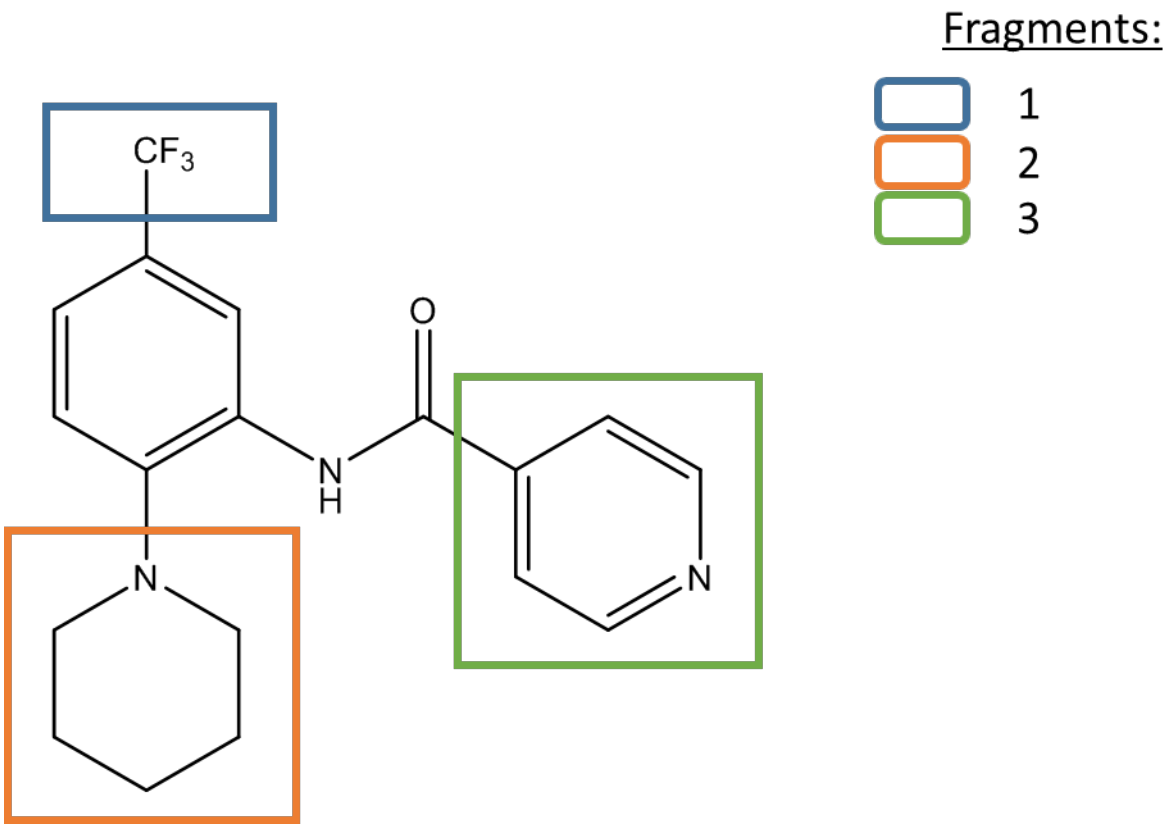


Figure 22: **The different fragments of SRPIN340 used to create the libraries**

5.2.1 Compound Library Formation

As mentioned before, different libraries were created from different parts of the SRPIN340 molecule. The alternatives used to make these libraries were sourced from different places. Fragment one was sourced from the SwissBioisostere database and can be seen in Figure 23.[66] Swiss Bioisostere is a database of molecular replacements. It provides information on molecular replacements and how they perform in biochemical assays to give an idea on what bioisosteric modifications can be made to a lead compound to optimise it. This software was chosen because suggested replacements for the fragments came from published data. Also, information on the biochemical assays helped give a better idea of the best replacements.

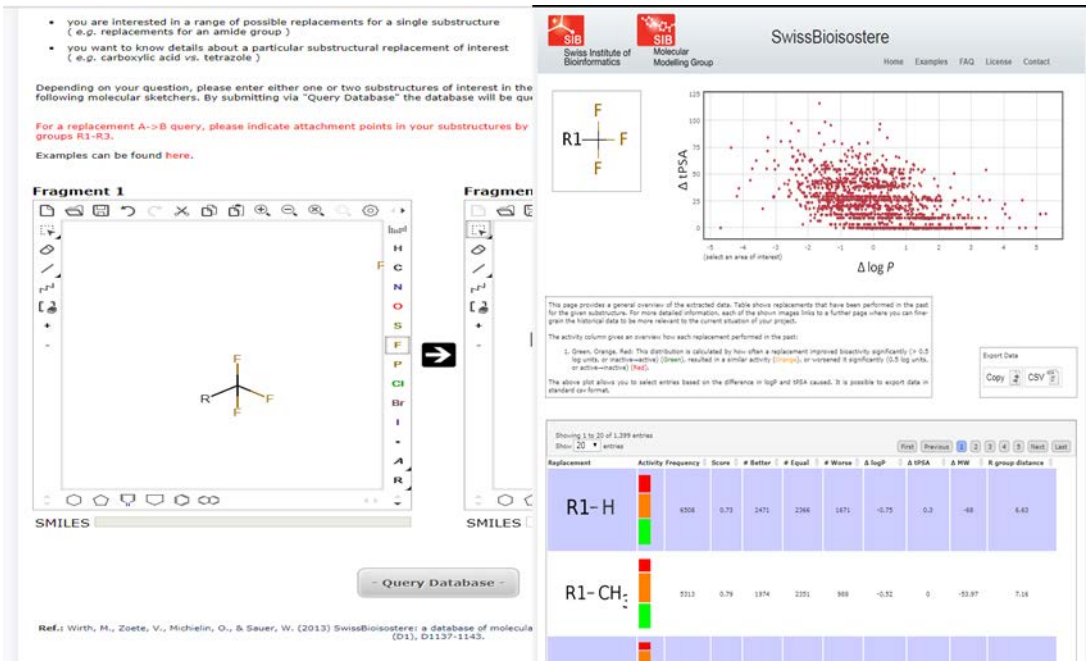


Figure 23: Picture showing the use of SwissBioisostere

The list was generated into a CSV file. The lists for substitutes for fragment two and three were sourced from ZINC15 which is a database of commercially available compounds.[67]

After gaining the libraries, they were put through KNIME software in three different entry ports as pictured in figure 24. Through these ports, bioisosteric transformations were made and then different combinations from the three lists were then made. The nodes used in the KNIME workflow were programmed using Anaconda python programming.

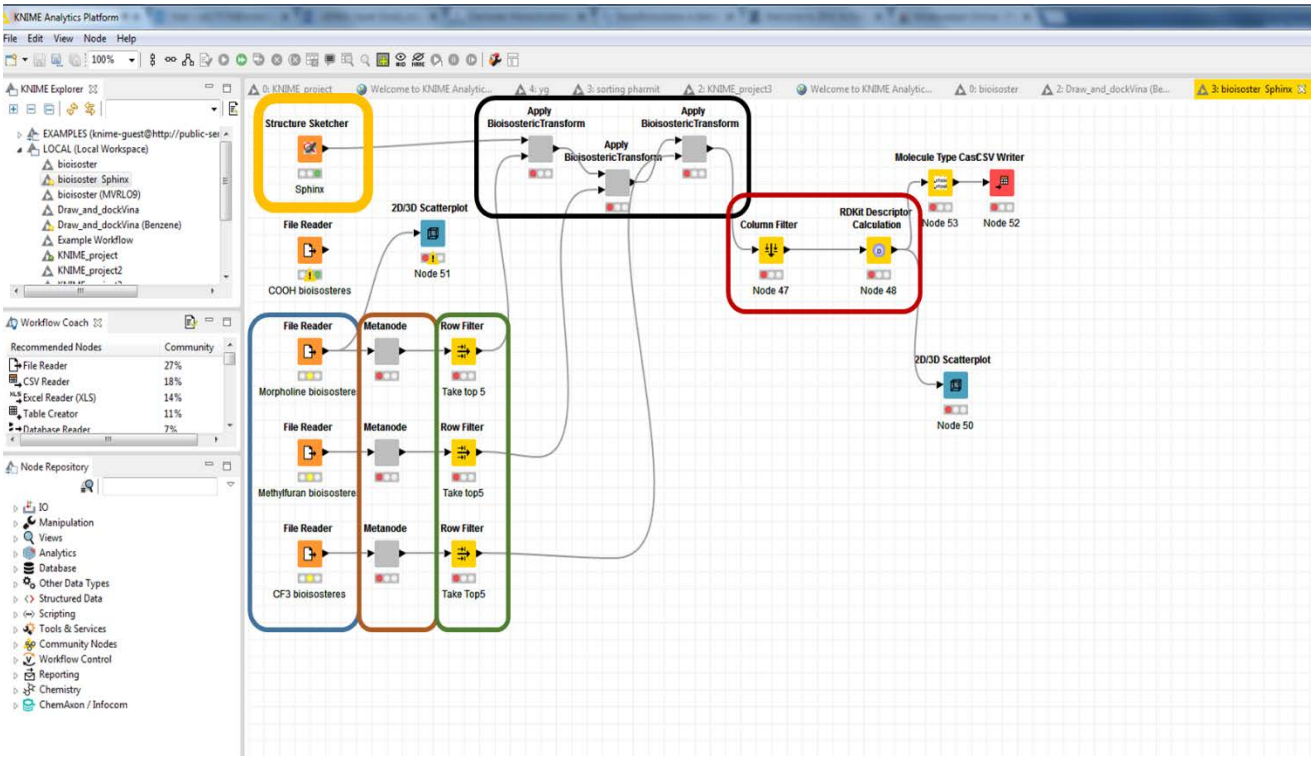


Figure 24: KNIME workflow used in the creation of analogues through combinations of the three libraries.

The nodes in the blue rectangle pictured in figure 24 were the ports through which the libraries were introduced to the workflow. After all 3 nodes had the libraries uploaded to them, the structure sketcher in the yellow square was used to upload the molecular structure for SRPIN340. This enabled the workflow to have a scaffold to work with. The nodes in brown sorted out the libraries to ensure they were in the correct format needed to apply the transformations. The libraries already came with each bioisosteric replacement having a score so the filters in green took the 100 molecules with the highest score in each library. The nodes in black (meta nodes) were each programmed to add the bioisosteric changes to the scaffold through different combinations. Each library was added in sequenced steps and after all combinations were completed, the new library was passed through the nodes in red which allowed only the product made to be visible. The RDKit Descriptor node gave properties to the newly formed products. It gave information such as LogP and molecular weight.

The final list of all the combinations had around 250 thousand molecules. With such a large number of molecules, it was imperative to deduce which were most likely to have an inhibitory effect on the target protein's binding site. To determine this, a 3D pharmacophore study was needed and the software Pharmit was used. An X-ray crystal structure of SRPK1 as shown in Figure 25 was found from the Protein Data Bank (PDB) and was derived by Morooka *et al*.[21] The crystal structure with SRPIN340 bound to the receptor had the label 4WUA with the ligand 3UL and it is with this, the pharmacophore was built.

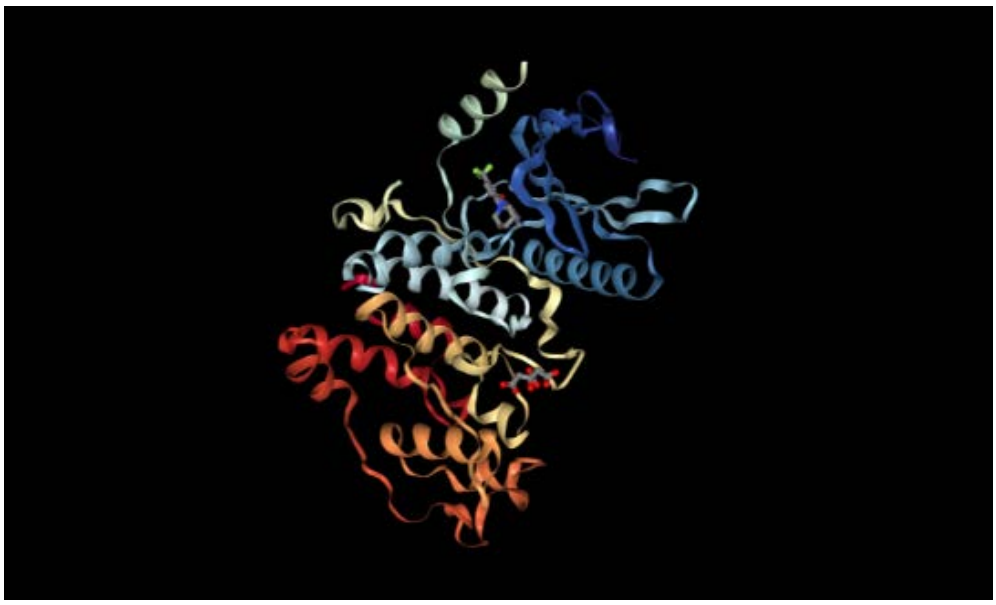


Figure 25: X-ray structure of SRPK1

The library list generated from before by KNIME was uploaded to the Pharmit software and the pharmacophore model searched through the generated list. The attributes that determined the pharmacophore's biological activity were chose as shown in the yellow rectangle in figure 26. There were over 25000 hits which were ranked in order of Root Mean Squared Deviation (RMSD). RMSD is the variation in the position of atoms in the pharmacophore when compared to the original

conformation and positions of the original pharmacophore. Therefore, the smaller the RMSD, the higher the similarity of a molecule to the compound in structure.

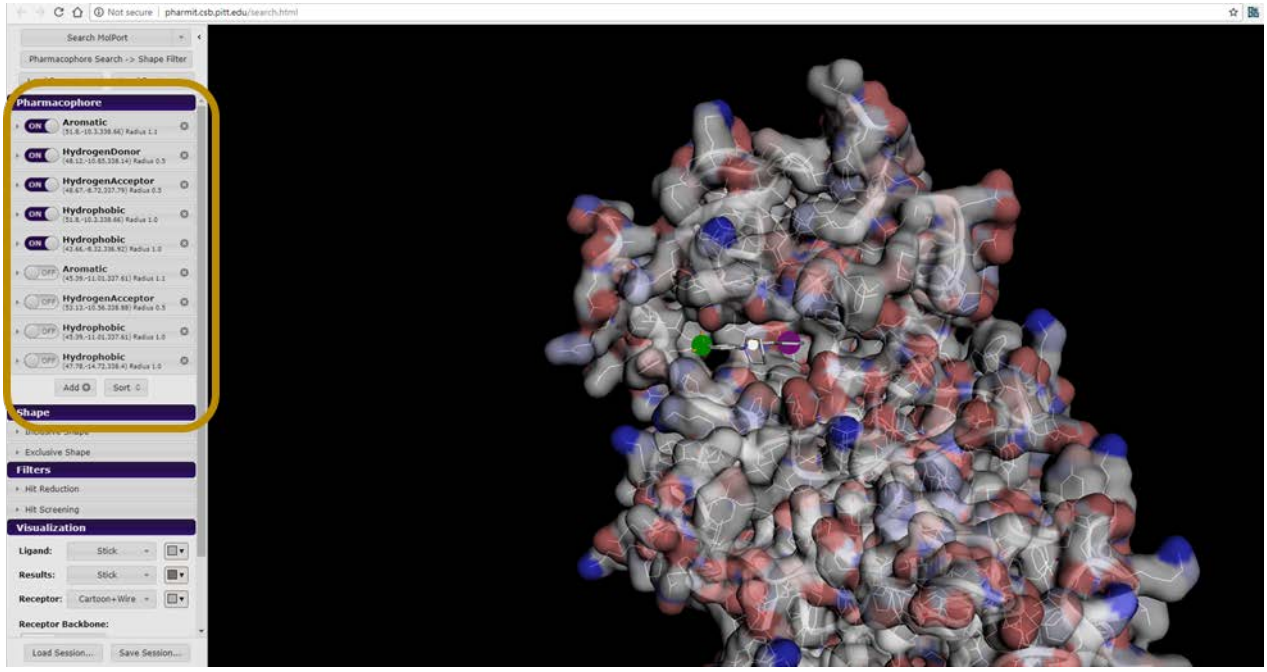


Figure 26: Pharmacophore modelling to score molecules in the library. Parameters set for the pharmacophore can be seen highlighted in the yellow rectangle.

Though there was a high number of molecules, it was noticed that a lot of them were repeated just in slightly different conformations so, in another effort to lower the numbers, a restriction was placed to only consider the best conformation of each molecule. This still generated a large list. The list was downloaded and put through another KNIME workflow.

The new workflow in KNIME had nodes to help filter out molecules using a set of parameters. These parameters were to find molecules that followed Lipinski's rule of 5[68] with a few modifications, had a reasonable binding affinity and to ensure that synthesis was possible. The parameters in detail were Lipinski's rule of five which outline certain properties most molecules need to have to be potentially effective as drugs.

Applying this to the workflow, molecular weight was one of the most important properties given. Lipinski's rule mentions that potential drugs should have a molecular weight less than 500 g/mol and therefore all molecules with a weight of over 450 were excluded. It must be emphasised how important hydrophilicity is as this determines whether a drug can be easily absorbed by the gastrointestinal tract. One of the major problems with SRPIN340 was its inability to be easily dissolved in water. Using Lipinski's rule, all molecules with a LogP of higher than 5 were excluded to ensure solubility was possible in water. Also, molecules with more than 5 hydrogen bond donor and 10 hydrogen bond acceptors were also excluded.

Outside of Lipinski's rule, one of the other filters used included affinity scores. All molecules with a score higher than -11 were excluded. A lot of the molecules scored between -11 and -12. In comparison, SRPIN340 had a score of -9.7, therefore, a score lower than that was needed for a more potent inhibitor. Another filter used focused on the fragments themselves. It was noticed when going through the list that some of the fragments repeated themselves as expected but a lot of them were not necessarily easy to synthesise. Some were searched for in databases including Sigma Aldrich, Molbase, Reaxys and Scifinder. These fragments were either not easy to make and required unavailable environments or they were simply not available in the catalogues for purchase. It was therefore not reasonable to choose the molecules that had these fragments and so such molecules were filtered out.

After all the filtering was done, lead compounds could not be chosen simply on their affinity scores or a single factor, therefore, the Pareto Principle was used. The Pareto principle states that most things are not distributed evenly. This means some things contribute more to the whole than others. An example would be in a work situation where 20% of the workers do the vast majority results whilst the remaining 80% do not contribute as much. It helps in decision making especially when figuring out where to place the highest priority instead of casting a wide net. Here, it was

determined that some properties contribute more to the molecules biological activity in comparison to others. The two most important factors chosen were affinity scores and LogP.

With this, a scatter plot was created. Due to the importance of both factors, a compromise had to be made and it meant not choosing a molecule with a low hydrophobicity whilst sacrificing binding affinity and vice versa. When this reasoning was used, molecules that fell near the curve as shown in figure 27 suited both factors. From this, 4 molecules were chosen, and the needed fragments were bought and synthesised as described in the synthesis.

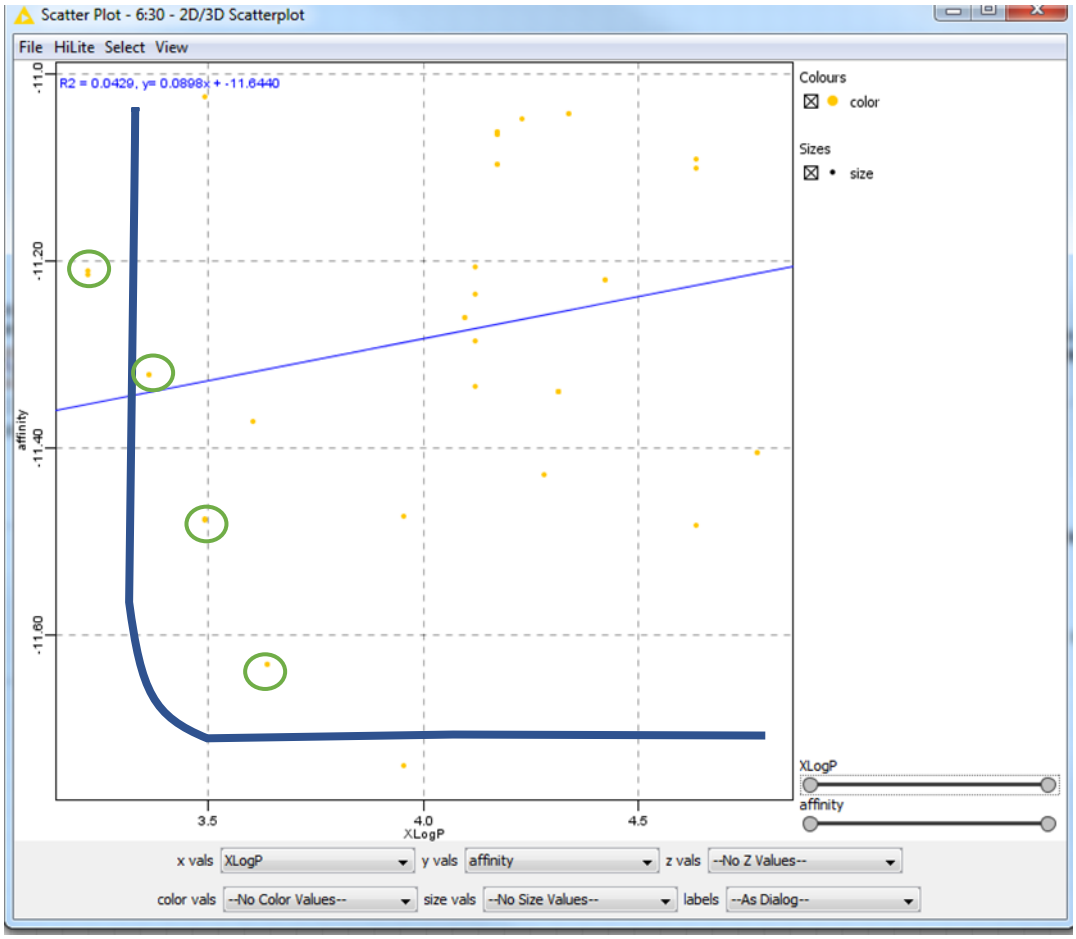
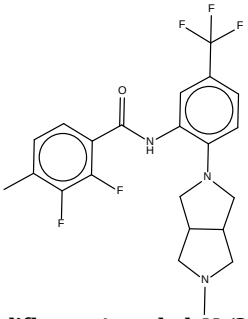


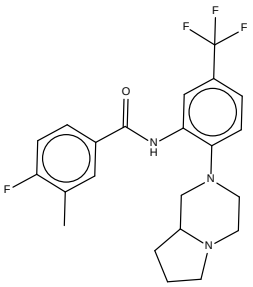
Figure 27: Scatter plot from KNIME created from the final compound list. Chosen molecules are highlighted in green circles.

5.3 Novel Compound Selection

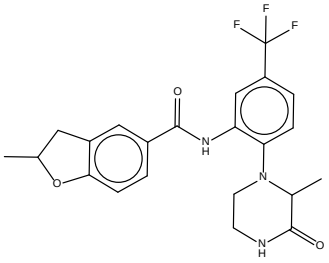
After the computational studies as outlined in the previous chapter was completed. Four molecules were chosen out of the thousands created. This selection was made based on LogP, molecular weight, binding affinity and if synthesis was feasible.



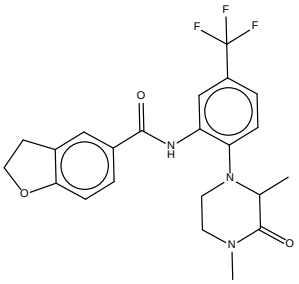
2,3-difluoro-4-methyl-*N*-(2-(5-methylhexahydropyrrolo[3,4-*c*]pyrrol-2(1*H*)-yl)-5-(trifluoromethyl)phenyl)benzamide (25)



4-fluoro-*N*-(2-(hexahydropyrrolo[1,2-*a*]pyrazin-2(1*H*)-yl)-5-(trifluoromethyl)phenyl)-3-methylbenzamide (24)



2-methyl-*N*-(2-(2-methyl-3-oxopiperazin-1-yl)-5-(trifluoromethyl)phenyl)-2,3-dihydrobenzofuran-5-carboxamide (26)



***N*-(2-(2,4-dimethyl-3-oxopiperazin-1-yl)-5-(trifluoromethyl)phenyl)-2,3-dihydrobenzofuran-5-carboxamide (27)**

Figure 28: The four compounds which were chosen to synthesise

The particulars of each molecule can be seen below in Table 3

Table 3: Table showing the selected compounds and the property values when selected

Compound	Molecular Weight	LogP	Binding Affinity	RMSD
SRPIN340	385.12	4.43	-9.7	0
2,3-difluoro-4-methyl-N-(2-(5-methylhexahydropyrrolo[3,4-c]pyrrol-2(1H)-yl)-5-(trifluoromethyl)phenyl)benzamide (26)	439.16	4.69	-11.81	1.05
4-fluoro-N-(2-(hexahydropyrrolo[1,2-a]pyrazin-2(1H)-yl)-5-(trifluoromethyl)phenyl)-3-methylbenzamide (25)	421.18	4.82	-11.84	1.16
2-methyl-N-(2-(2-methyl-3-oxopiperazin-1-yl)-5-(trifluoromethyl)phenyl)-2,3-dihydrobenzofuran-5-carboxamide (27)	433.43	3.956	-11.74	0.89
N-(2-(2,4-dimethyl-3-oxopiperazin-1-yl)-5-(trifluoromethyl)phenyl)-2,3-dihydrobenzofuran-5-carboxamide (28)	433.43	3.638	-11.63	1.14

As shown, there were stronger binding affinities in all the molecules chosen compared to SRPIN340. All the components needed to synthesise the molecules were then ordered. Unfortunately, not all the fragments arrived. Only two of fragments 2 and two from fragment 3 ever arrived and they did not correspond with each other. It was decided that the synthesis will carry on and synthesis molecules with the fragments available whilst mixing and matching the fragments. In the end, 4 new molecules could be potentially made. These are illustrated in Figure 29.

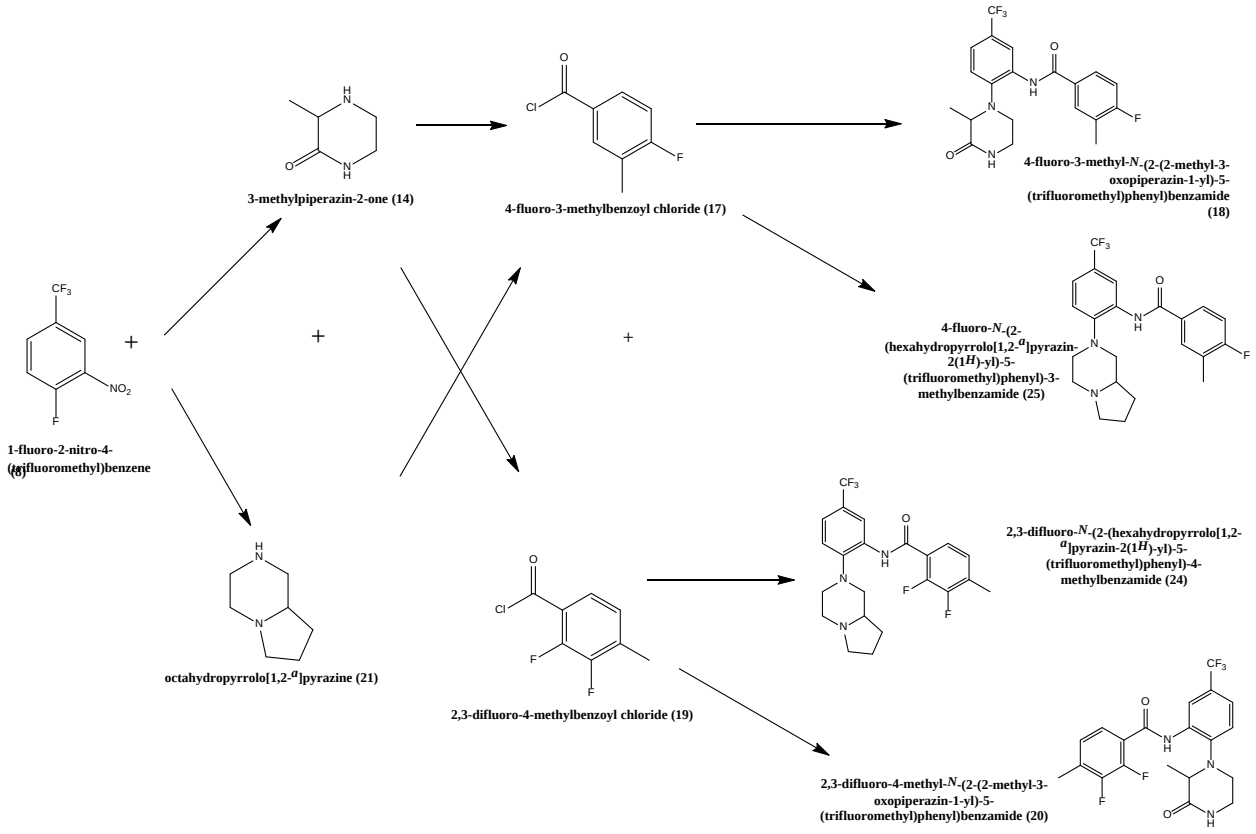


Figure 29: Illustration of the combinations of fragments

The synthesis of each molecule is detailed in the previous chapter including analytical data. Copies of analytical data can be found in the appendix. The molecules containing the 3-methylpiperazin-2-one fragment proved difficult to synthesise in every stage of the reaction. All analytical data showed unexpected information and the generic reaction had to be modified many times to find the molecule needed. Some tweaks included using heat and increasing the amount of $SnCl_2$ used as well as using more concentrated acid in the reduction step.

After the synthesis, unfortunately, there was not enough time to test the molecules on the SRPK1 protein to test for its activity using experimental methods. Also, *in vivo* testing could not be done on the mammalian cells.

5.4 *In vitro* toxicity testing of SRPIN340 on mammalian cells.

Two cell strains from metastatic tumours found in breast cancer patients were used for the toxicity testing. Breast cancer cells were chosen due to studies showing that the overexpression of SRPK1 lessens apoptosis in breast cancer cells.[69] The results for the toxicity testing of SRPIN340 on the two breast cancer cell strains can be seen in figure 30.

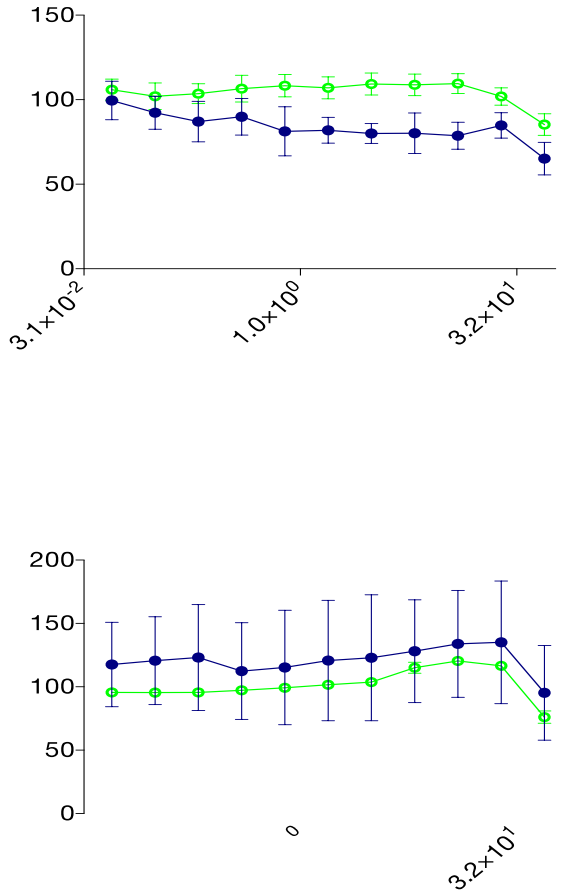


Figure 30: Graph showing SRPIN340 toxicity.

For the MCF7 strain, SRPIN340 toxicity could be seen in concentrations above 25 µM in cells stained with Alamar blue but with the cells stained with calcein AB, toxicity could be seen in concentrations above 10 µM. In the MDA strain, toxicity showed in concentrations above 25 µM and it was the same for stains.

5.5. Synthesis

The synthesis of SRPIN340 was successful and was produced with a high yield of 87.32%. But this did not replicate in the production of the novel compounds. Compounds 17, 19 were not made and the rest of the novel compounds were produced but in very low titres. The particular reason this could have happened is not ascertained but there are possible theories.

Firstly, in stage 1, the nucleophiles used were piperazines are less basic than piperidine. With this reduced basicity, the piperazines are less nucleophilic and will be more difficult to do the substation. This can be seen in drops in yields. SPRIN240 had a yield of 90% in stage one compared to the rest of the molecules with a maximum yield of 51.8% in the first stage. In a paper by Magano et al,[70] sodium tert-butoxide, a palladium catalyst and toluene with the entire reaction being done in reflux showed success so this might be an avenue to take in future.

In stage 2, an instant colour change was visible when the reaction was started. The solution went from brightly coloured yellow/orange to light yellow then to clear, so a reaction was evident. Yet when the extraction was done, the was little to no product produced at all. In acidic conditions like that needed for the reduction, the piperazines can become reactive. The keto group on the 3-methylpiperazin-2-one especially reacts in acidic conditions since it protonates it become a hydroxy group. This is evident in the product being made. The octahydropyrollo[1,2-α]pyrazine with fairly good yields of around 70% since this acid reaction did not affect it. In the future, blocking the keto group might increase the yield produced.

6. Conclusions

The known kinase inhibitor SRPin340 was successfully synthesised with its presence being determined with NMR spectroscopy. Toxicity studies were conducted using two strains of mammalian breast cancer cells. The studies showed the concentrations of around 25 µM caused toxicity.

SRPK1 Protein was also successfully purified from recombinant *E. coli* cells. The protein was then purified using nickel-affinity FPLC and the presence and weight of the protein were determined through an SDS PAGE. The results from the page showed the protein to have a weight of around 35 kDa. This was also confirmed by LC-MS. This weight did not match previous work using the same plasmid which gave a weight of 45 kDa. Even when the expression was repeated, the same 35 kDa weight was the result. This protein purified is believed to be the ASF2 protein.

Using SRPIN340 as a scaffold, three libraries were sourced from ZINC15 and SwissBioiostere with each library representing a fragment of the scaffold. These libraries were put through KNIME which created different combinations of the fragments to create new analogues of SRPIN340 to form a new library.

A 3D pharmacophore study was conducting using SRPIN340 complexed to SRPK1. The crystal structure for the complex was generated by Morooka *et al*. and was used in the Pharmit software. The forces that were felt to be vital in the biological activity of the pharmacophore was highlighted and with the information, the software screened through the new library to find the most likely biologically active. It also ranked the molecules according to their affinity scores. The list was downloaded to create another library.

A new KNIME workflow was created to help filter molecules that were too heavy, too hydrophobic, had low affinity scored and had fragments that were not easy to purchase or synthesise. Using Pareto's principle in a scatter plot, four molecules were chosen that had both low LogP values and high-affinity scores. The four fragments from the four molecules were then purchased for synthesis.

Not all the fragments arrived so with the ones available, different combinations of the fragments were synthesised to create new molecules. Two were very easy to synthesise but the other two did not create the final product intended. This was due to unpredicted interactions that certain functional groups had with catalysts and the reagents. Unfortunately, experimental binding studies could not be conducted as well as toxicity studies. This was due to time constraints.

7. Further work

In this section, improvements that can be made to this study will be discussed as well as what further work can be done.

Firstly, as it was not known why the protein harvested from the *E. coli* cells was smaller than anticipated. To figure out why, sequencing of the plasmid can be done to find the gene sequence, so a more accurate weight can be deduced. PCR can also be done on the plasmid to ensure the gene is still in the plasmid. From this, if the problem lies with the plasmid, new plasmids can be gotten or made. If it lies with the technique in protein expression, the experiment can be redone after evaluating. Size exclusion and longer runs could be done to ensure the purification of SPRK1 in the future also.

One of the biggest problems faced was the synthesis of molecules **18** and **25**. There were problems in every stage of the reaction and some stages had to be done multiple times with different conditions and working in small quantities did not allow for easy product retrieval from separations. In the future, trying with larger amounts might help in recuperating more product. Also during reactions, blocking and protecting reactive functional groups from making unwanted interactions could increase the possibilities of the product being made.

As mention in previous chapters, a lot of the fragments ordered did not arrive which made it difficult to create the molecules chosen from the computational studies. For future work, getting hold of those fragments and synthesising the molecules would be greatly advantageous to compare to the computational predictions.

Due to time constraints, toxicity and binding studies could not be conducted on the synthesised molecules. With SRPIN340 as a control, comparisons of potency and binding strength between the molecules and SRPIN340 can be made. This is essential to argue that the new molecules would be better.

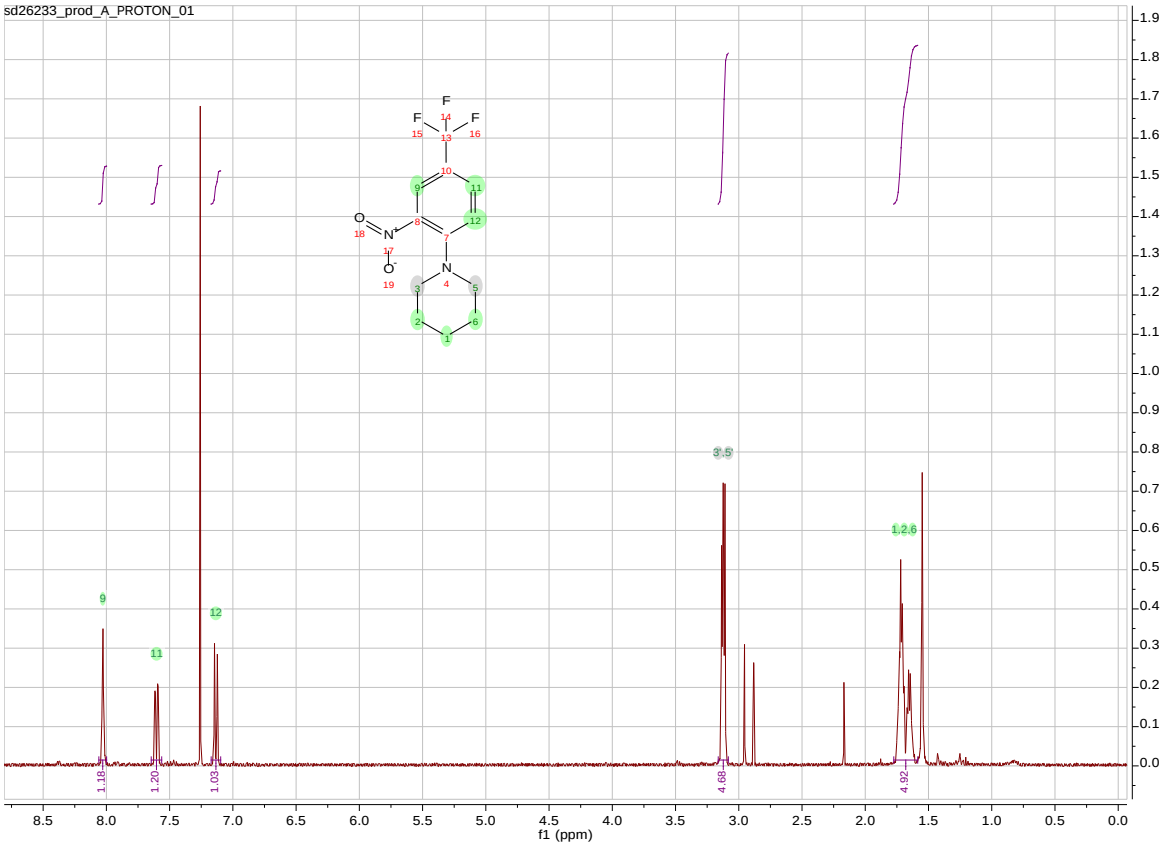
8. References

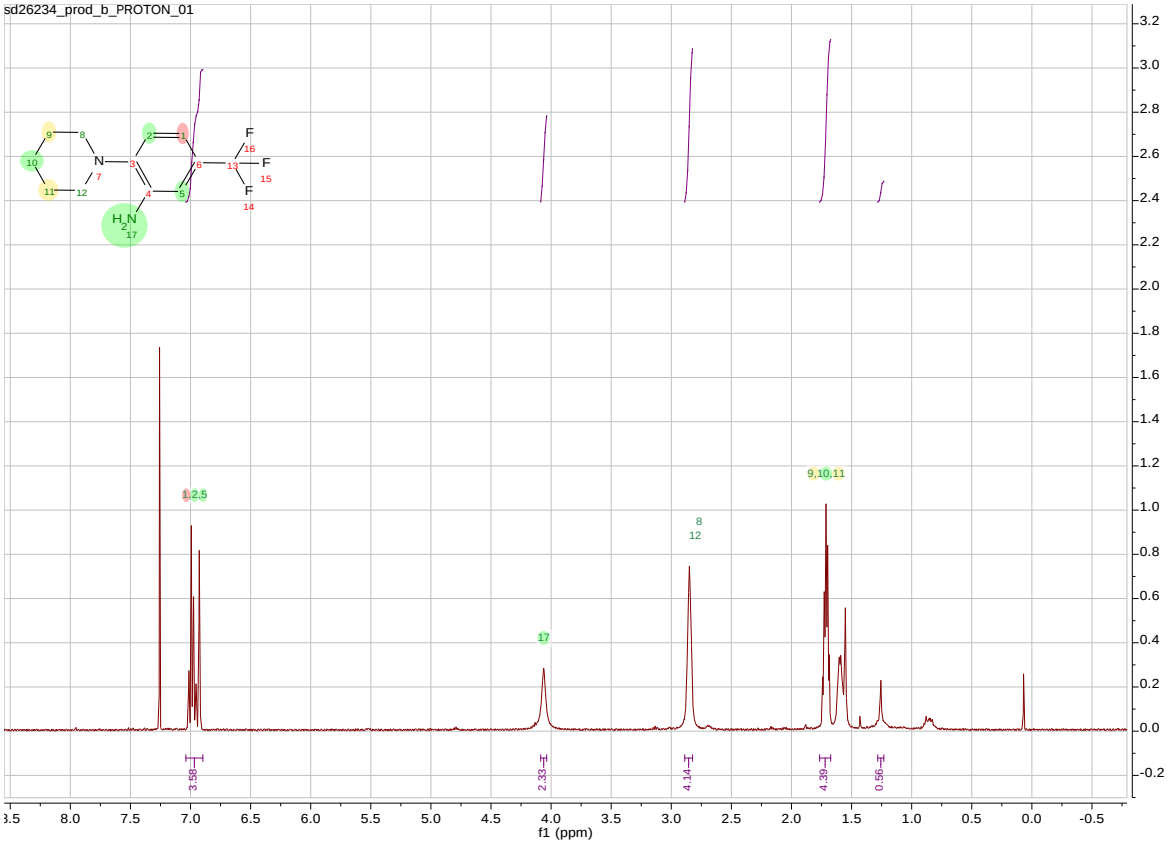
1. Y. A. N. Wang, J. Liu, B. O. Huang, Y.-M. Xu, J. Li, L.-F. Huang, J. I. N. Lin, J. Zhang, Q.-H. Min, W.-M. Yang and X.-Z. Wang, *Biomedical Reports*, 2015, **3**, 152-158.
2. T. H. Adair and J.-P. Montani, *Colloquium Series on Integrated Systems Physiology: From Molecule to Function*, 2010, **2**, 1-84.
3. W. Risau and I. Flamme, *Annu Rev Cell Dev Biol.* , 1995, **11**, 73-91.
4. P. Carmeliet, *Nature Medicine*, 2003, **9**, 653.
5. P. Carmeliet and R. K. Jain, *Nature*, 2000, **407**, 249.
6. P. H. Burri and M. R. Tarek, *The Anatomical Record*, 1990, **228**, 35-45.
7. S. J. Mentzer and M. A. Konerding, *Angiogenesis*, 2014, **17**, 499-509.
8. V. W. van Hinsbergh and P. Koolwijk, *Cardiovacular Research*, 2008, **78**, 203-212.
9. H. Gerhardt, *Organogenesis*, 2008, **4**, 241-246.
10. S. Chien, *American Journal of Physiology, Heart and Circulatory Physiology*, 2007, **292**, 209-224.
11. D. I. R. Holmes and I. Zachary, *Genome Biology*, 2005, **6**, 209-209.
12. C. Maes, K. Carmeliet P Fau - Moermans, I. Moermans K Fau - Stockmans, N. Stockmans I Fau - Smets, D. Smets N Fau - Collen, R. Collen D Fau - Bouillon, G. Bouillon R Fau - Carmeliet and G. Carmeliet, *Mechanisms of Development*, 2002, **111**, 61-73.
13. F. B. Perler, E. O. Davis, G. E. Dean, F. S. Gimble, W. E. Jack, N. Neff, C. J. Noren, J. Thorner and M. Belfort, *Nucleic Acids Research*, 1994, **22**, 1125-1127.
14. V. O. Wickramasinghe, M. Gonzàlez-Porta, D. Perera, A. R. Bartolozzi, C. R. Sibley, M. Hallegger, J. Ule, J. C. Marioni and A. R. Venkitaraman, *Genome Biology*, 2015, **16**, 201.
15. D. G. Nowak, E. M. Amin, E. S. Rennel, C. Hoareau-Aveilla, M. Gammons, G. Damodoran, M. Hagiwara, S. J. Harper, J. Woolard, M. R. Ladomery and D. O. Bates, *The Journal of Biological Chemistry*, 2010, **285**, 5532-5540.
16. N. Bullock and S. Oltean, *The Journal of Pathology*, 2017, **241**, 437-440.
17. X. Xu, Y. Wei, S. Wang, M. Luo and H. Zeng, *Oncotarget*, 2017, **8**, 61944-61957.
18. G. M. Hayes, L. J. Carrigan Pe Fau - Miller and L. J. Miller, *Cancer Research*, 2007, **67**, 2072-2080.
19. S. E. van Roosmalen W Fau - Le Devedec, O. Le Devedec Se Fau - Golani, M. Golani O Fau - Smid, I. Smid M Fau - Pulyakhina, A. M. Pulyakhina I Fau - Timmermans, M. P. Timmermans Am Fau - Look, D. Look Mp Fau - Zi, C. Zi D Fau - Pont, M. Pont C Fau - de Graauw, S. de Graauw M Fau - Naffar-Abu-Amara, C. Naffar-Abu-Amara S Fau - Kirsanova, G. Kirsanova C Fau - Rustici, P. A. C. t. Rustici G Fau - Hoen, J. W. M. Hoen Pa Fau - Martens, J. A. Martens Jw Fau - Foekens, B. Foekens Ja Fau - Geiger, B. Geiger B Fau - van de Water and B. van de Water, *J Clin Invest*, 2015, **125**, 1648-1664.
20. T. Fukuhara, T. Hosoya, S. Shimizu, K. Sumi, T. Oshiro, Y. Yoshinaka, M. Suzuki, N. Yamamoto, L. A. Herzenberg, L. A. Herzenberg and M. Hagiwara, *Proceedings of the National Academy of Sciences*, 2006, **103**, 11329.
21. S. Morooka, M. Hoshina, I. Kii, T. Okabe, H. Kojima, N. Inoue, Y. Okuno, M. Denawa, S. Yoshida, J. Fukuhara, K. Ninomiya, T. Ikura, T. Furuya, T. Nagano, K. Noda, S. Ishida, T. Hosoya, N. Ito, N. Yoshimura and M. Hagiwara, *Molecular Pharmacology*, 2015, **88**, 316-325.
22. T. Ikura, M. Hoshina, T. Hosoya, M. Hagiwara and N. Ito, *Journal*, 2016, DOI: 10.2210/pdb4WUA/pdb.

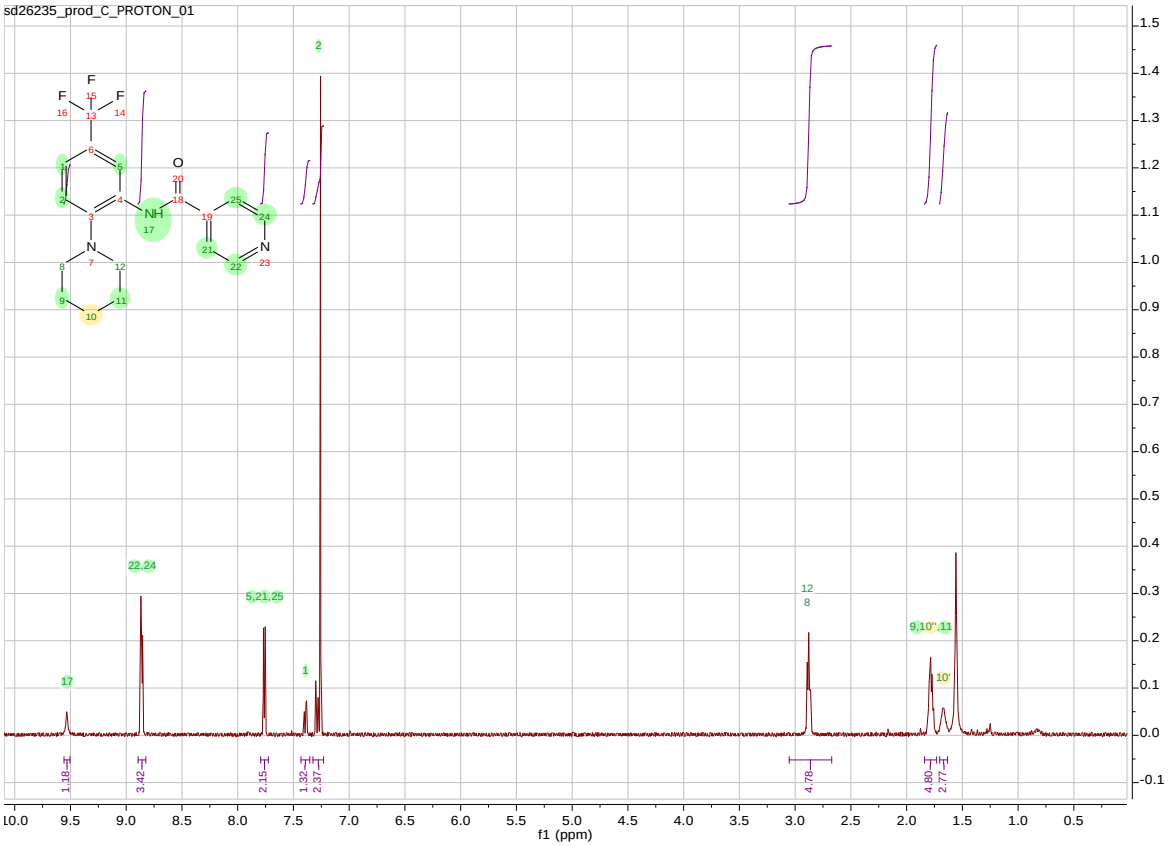
23. J. Batson, H. D. Toop, C. Redondo, R. Babaei-Jadidi, A. Chaikuad, S. F. Wearmouth, B. Gibbons, C. Allen, C. Tallant, J. Zhang, C. Du, J. C. Hancox, T. Hawtrey, J. Da Rocha, R. Griffith, S. Knapp, D. O. Bates and J. C. Morris, *ACS Chemical Biology*, 2017, **12**, 825-832.
24. V. R. Vemula, V. Lagishetty and S. Lingala, *International Journal of Pharmaceutical Sciences Review and Research*, 2010, **5**, 41-51.
25. E. H. Kerns and L. Di, *Drug-Like Properties: Concept, Structure Design and Methods, From ADME to Toxicity Optimization*, 2008.
26. M. Lovett, K. Lee, A. Edwards and D. L. Kaplan, *Tissue Engineering. Part B, Reviews*, 2009, **15**, 353-370.
27. K. M. Gehrs, D. H. Anderson, L. V. Johnson and G. S. Hageman, *Annals of medicine*, 2006, **38**, 450-471.
28. T. Ayoub and N. Patel, *Journal of the Royal Society of Medicine*, 2009, **102**, 56-61.
29. RNIB, Understand your eye anatomy, (accessed 20/08, 2018).
30. H.-d. Zou, X. Zhang, X. Xu, F.-h. Wang and S.-j. Zhang, *Zhonghua Yan Ke Za Zhi*, 2005, **41**, 15-19.
31. M. Flax, *Archives of Ophthalmology*, 2007, **125**, 858-859.
32. N. E. Institute, Facts About Age-Related Macular Degeneration, https://nei.nih.gov/health/maculardegen/armd_facts, (accessed 20/08, 2018).
33. T. Iroku-Malize and S. Kirsch, *FP Essentials*, 2016, **445**, 24-28.
34. D. Yorston, *Community Eye Health*, 2006, **19**, 4-5.
35. F. W. Lusby, Macular degeneration, (accessed 20/08, 2018).
36. P. M. Gullino, *JNCI: Journal of the National Cancer Institute*, 1978, **61**, 639-643.
37. D. Liao and R. S. Johnson, *Cancer Metastasis Reviews*, 2007, **26**, 281-290.
38. N. Nishida N Fau - Nishida, H. Yano H Fau - Yano, T. Nishida T Fau - Nishida, T. Kamura T Fau - Kamura and M. Kojiro M Fau - Kojiro, *Vascular Health Rish Managament*, 2006, **2**, 213-219.
39. J. Folkman, *N Engl J Med*, 1971, **285**, 1182-1186.
40. L. Muthukkaruppan Vr Fau - Kubai, R. Kubai L Fau - Auerbach and R. Auerbach, *Journal of the National Cancer Institute*, 1982, **69**, 699-708.
41. S. Ziyad and M. L. Iruela-Arispe, *Genes & Cancer*, 2011, **2**, 1085-1096.
42. J. P. Ward, *Biochimica et Biophysica Acta*, 2008, **1777**, 1-14.
43. G. L. Semenza, *Nature Reviews: Cancer*, 2003, **3**, 721-732.
44. F. Hillen F Fau - Hillen and A. W. Griffioen Aw Fau - Griffioen, *Cancer Metastasis Reviews*, 2007, **26**, 489-502.
45. J. A. Nagy Ja Fau - Nagy, S.-H. Chang Sh Fau - Chang, S.-C. Shih Sc Fau - Shih, A. M. Dvorak Am Fau - Dvorak and H. F. Dvorak Hf Fau - Dvorak, *Seminars in Thrombosis and Hemostasis*, 2010, **36**, 321-331.
46. J. M. Brown, *Methods in Enzymology*, 2007, **435**, 297-321.
47. B. L. Krock Bl Fau - Krock, N. Skuli N Fau - Skuli and M. C. Simon Mc Fau - Simon, *Genes and Cancer*, 2011, **2**, 1117-1133.
48. K. H. Plate, H. A. Breier G Fau - Weich, W. Weich Ha Fau - Risau and W. Risau, *Nature*, 1992, **259**, 845-848.
49. A. E. El-Kenawi and A. B. El-Remessy, *British Journal of Pharmacology*, 2013, **170**, 712-729.
50. A. Abdollahi, C. Hahnfeldt P Fau - Maercker, H.-J. Maercker C Fau - Grone, J. Grone Hj Fau - Debus, W. Debus J Fau - Ansorge, J. Ansorge W Fau - Folkman, L. Folkman J Fau - Hlatky, P. E. Hlatky L Fau - Huber and P. E. Huber, *Molecular Cell*, 2004, **13**, 649-663.
51. M. Rajabi M Fau - Rajabi and S. A. Mousa Sa Fau - Mousa, *Biomedicines*, 2017, **5**, 34.
52. F. Ciardiello, R. Caputo R Fau - Bianco, V. Bianco R Fau - Damiano, G. Damiano V Fau - Fontanini, S. Fontanini G Fau - Cuccato, S. Cuccato S Fau - De Placido, A. R. De Placido S Fau - Bianco, G. Bianco Ar Fau - Tortora and G. Tortora, *Clinical Cancer Research*, 2001, **7**, 1459-1465.
53. G. Neufeld and O. Kessler, *Cancer Metastasis Reviews*, 2006, **25**, 373-385.

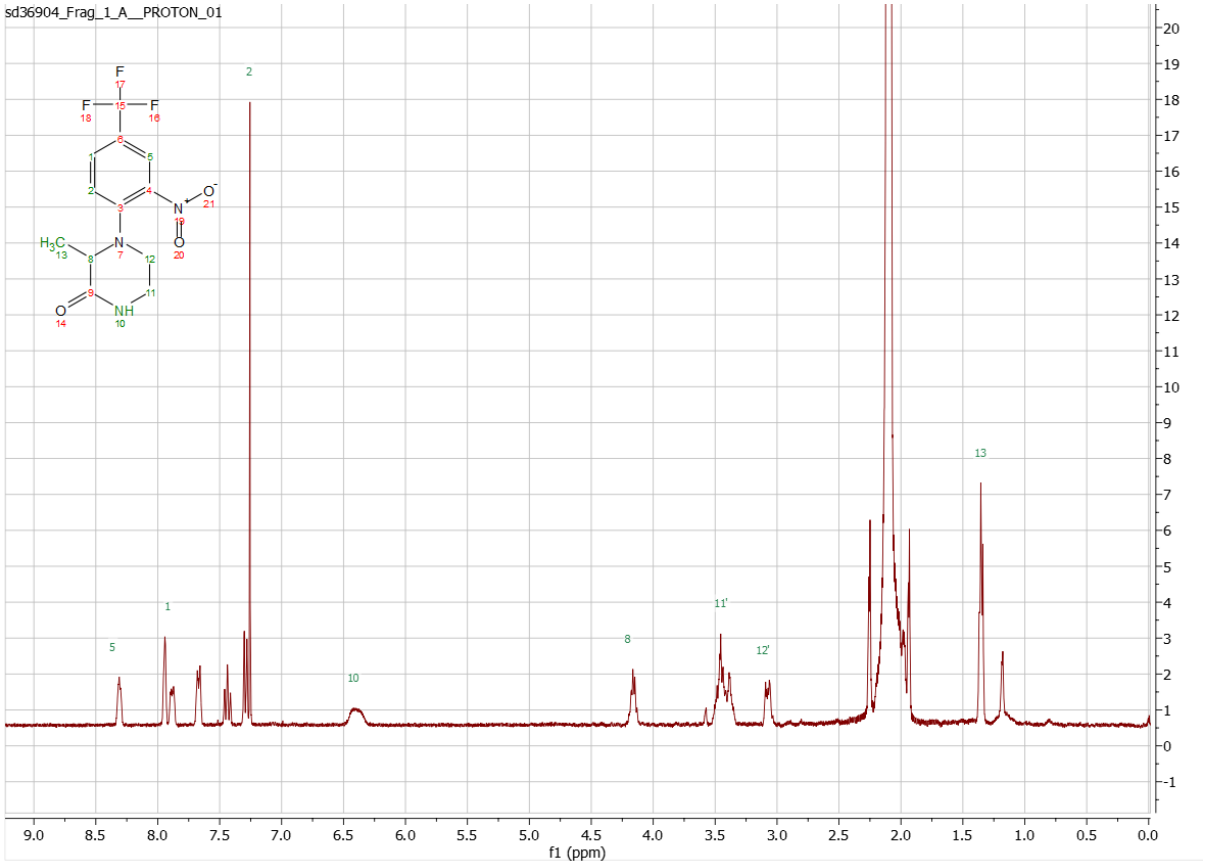
54. S. V. Rajkumar and T. E. Witzig, *Cancer Treatment Reviews*, 200, **26**, 351-362.
55. L. M. Ellis, *Seminars in Oncology*, 2006, **33**, 1-7.
56. J. L. McKimm-Breschkin, *Influenza and Other Respiratory Viruses*, 2013, **1**, 25-36.
57. A. Moscona, *New England Journal of Medicine*, 2005, **353**, 1363-1373.
58. M. Elliott, *Philosophical Transactions of the Royal Society of London. Series B*, 2001, **356**, 1885-1893.
59. M. von Itzstein, *Nature Reviews Drug Discovery*, 2007, **6**, 967.
60. F. W. Schueler, *Chemobiodynamics and Drug Design*, McGraw-Hill Book Co., Inc., New York, 1961.
61. B. E. Aubol, S. Chakrabarti, J. Ngo, J. Shaffer, B. Nolen, X.-D. Fu, G. Ghosh and J. A. Adams, *Proceedings of the National Academy of Sciences*, 2003, **100**, 12601.
62. M. V. Gammons, D. Fedorov O Fau - Ivison, C. Ivison D Fau - Du, T. Du C Fau - Clark, C. Clark T Fau - Hopkins, M. Hopkins C Fau - Hagiwara, A. D. Hagiwara M Fau - Dick, R. Dick Ad Fau - Cox, S. J. Cox R Fau - Harper, J. C. Harper Sj Fau - Hancox, S. Hancox JC Fau - Knapp, D. O. Knapp S Fau - Bates and D. O. Bates, *Investigative Ophthalmology and Visual Science*, 2013, **54**, 6052-6062.
63. J. C. Ngo, K. Gullingsrud J Fau - Giang, M. J. Giang K Fau - Yeh, X.-D. Yeh Mj Fau - Fu, J. A. Fu Xd Fau - Adams, J. A. Adams Ja Fau - McCammon, G. McCammon Ja Fau - Ghosh and G. Ghosh, *Structure*, 2007, **15**, 123-133.
64. A. R. Krainer, D. Conway Gc Fau - Kozak and D. Kozak.
65. G. Michlewski, J. F. Sanford Jr Fau - Caceres and J. F. Caceres.
66. M. Wirth, V. Zoete, O. Michielin and W. H. B. Sauer, *Nucleic Acids Research*, 2013, **41**, D1137-D1143.
67. Irwin, Sterling, Mysinger, Bolstad and Coleman, *J. Chem. Inf. Model*, 2012, DOI: DOI: 10.1021/ci3001277.
68. C. A. Lipinski, B. W. Lombardo F Fau - Dominy, P. J. Dominy Bw Fau - Feeney and P. J. Feeney, *Advanced Drug Delivery Reviews*, 2001, **46**, 3-26.
69. J. C. Lin, C. Y. Lin, W. Y. Tarn and F. Y. Li, *RNA*, 2014, **20**, 1621-1631.
70. J. Magano, A. Akin, M. Chen, K. Giza, J. Moon and J. Saenz, *Synthetic Communications - SYN COMMUN*, 2008, **38**, 3631-3639.

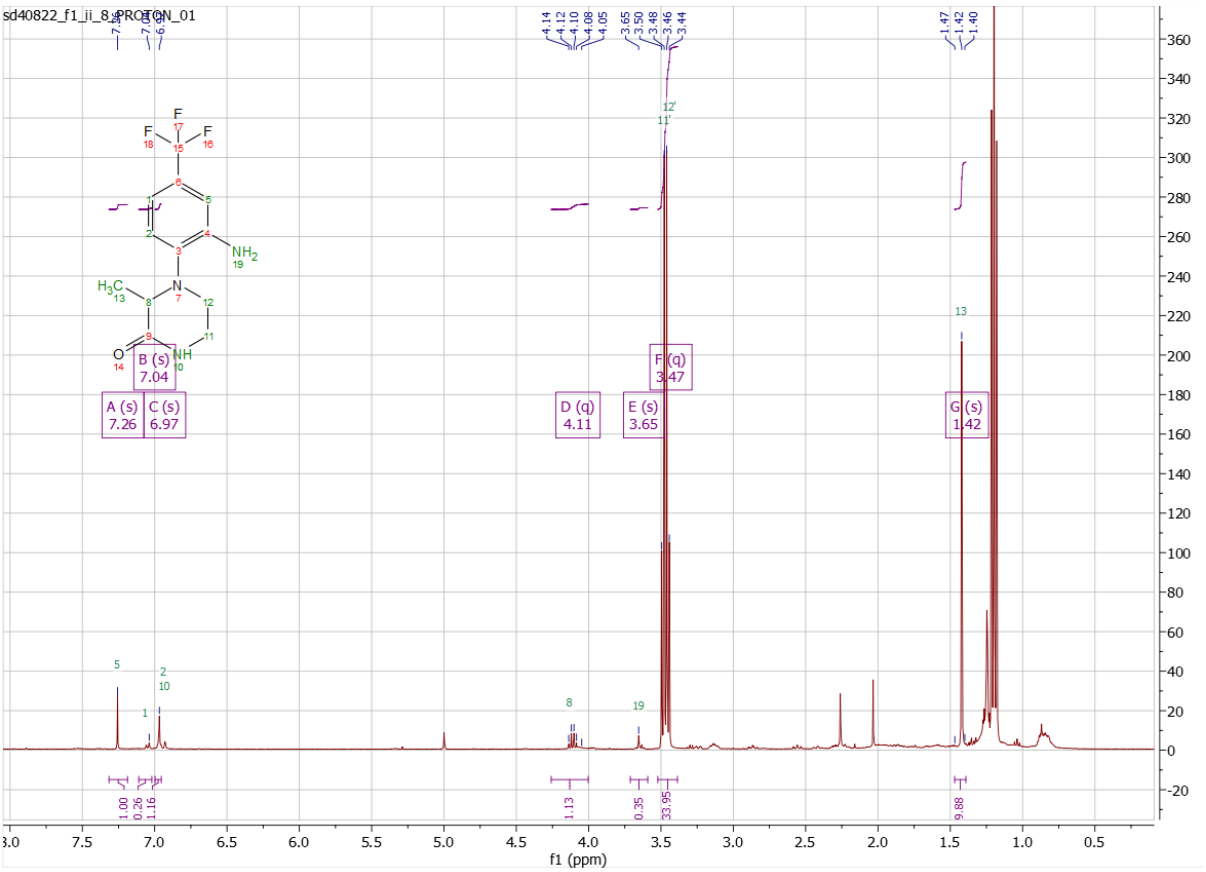
9. Appendix

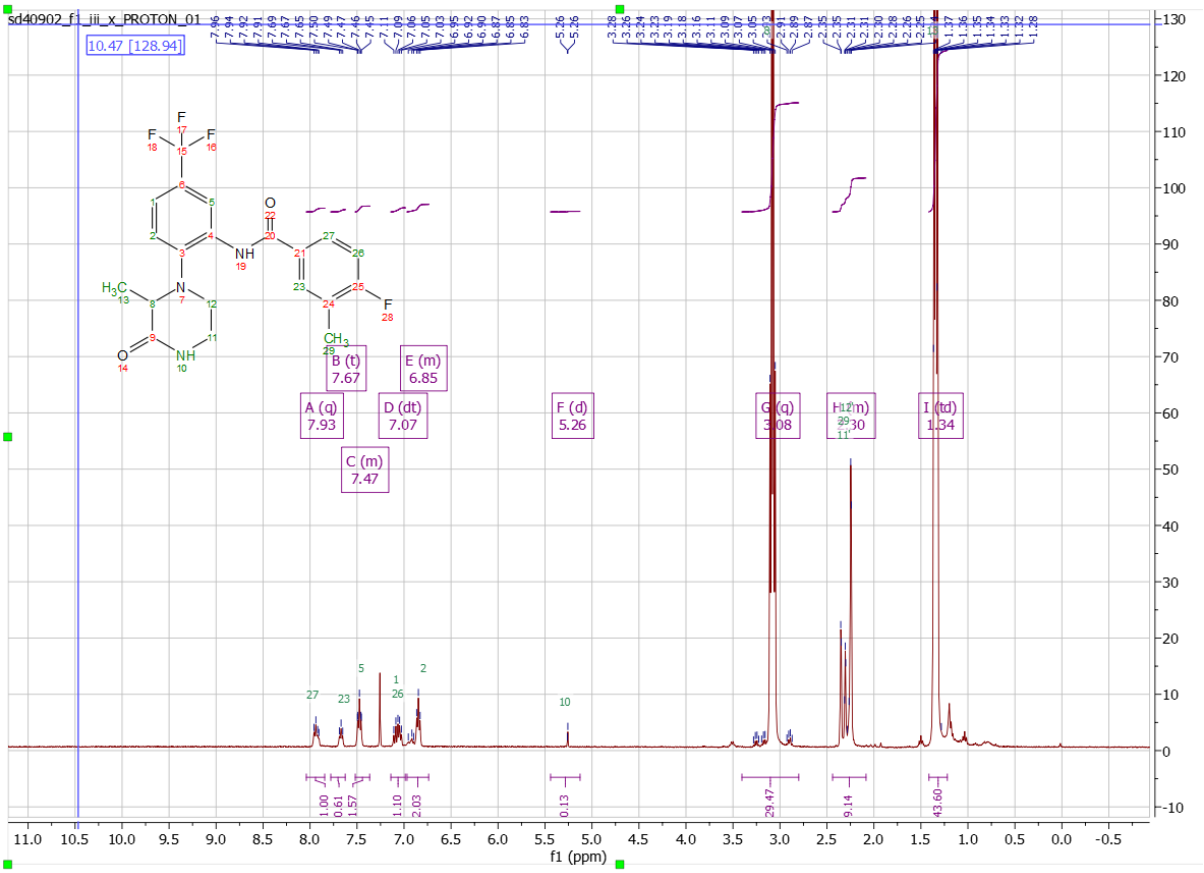

sd26233_prod_A_PROTON_01
f1 (ppm)
9
11
12
3,5
1,2,6
1.18
1.20
1.03
4.68
4.92

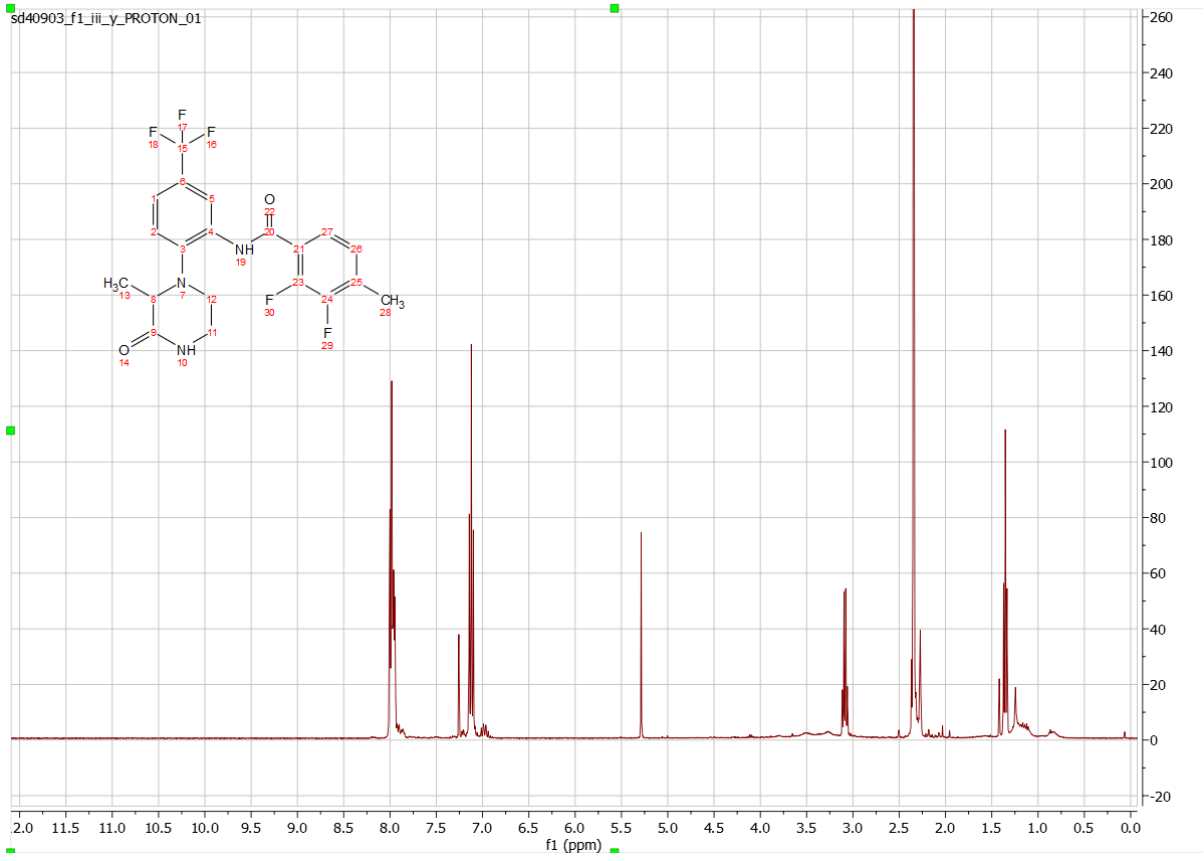

sd26234_prod_b_PROTON_01
1,2,5
8
12
17
9,10,11
3.98
2.33
4.14
4.39
0.56
f1 (ppm)

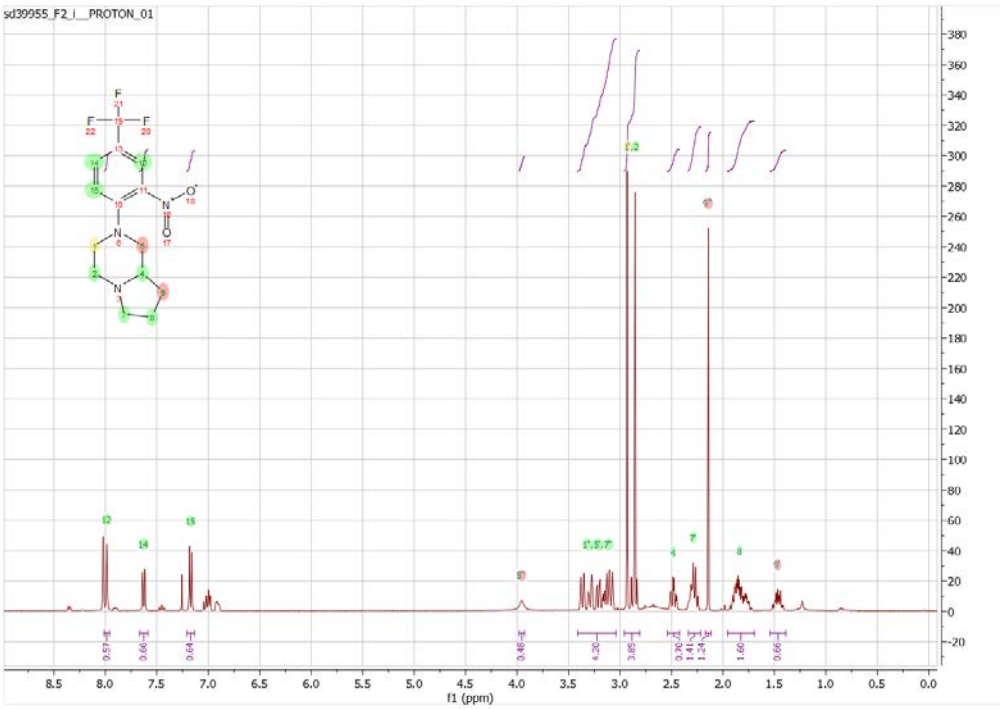

sd26235_prod_C_PROTON_01
2
22,24
5,21,25
17
1
12
8
9,10',11
10'
1.18
3.42
2.15
1.32
2.37
4.78
4.80
2.77
f1 (ppm)

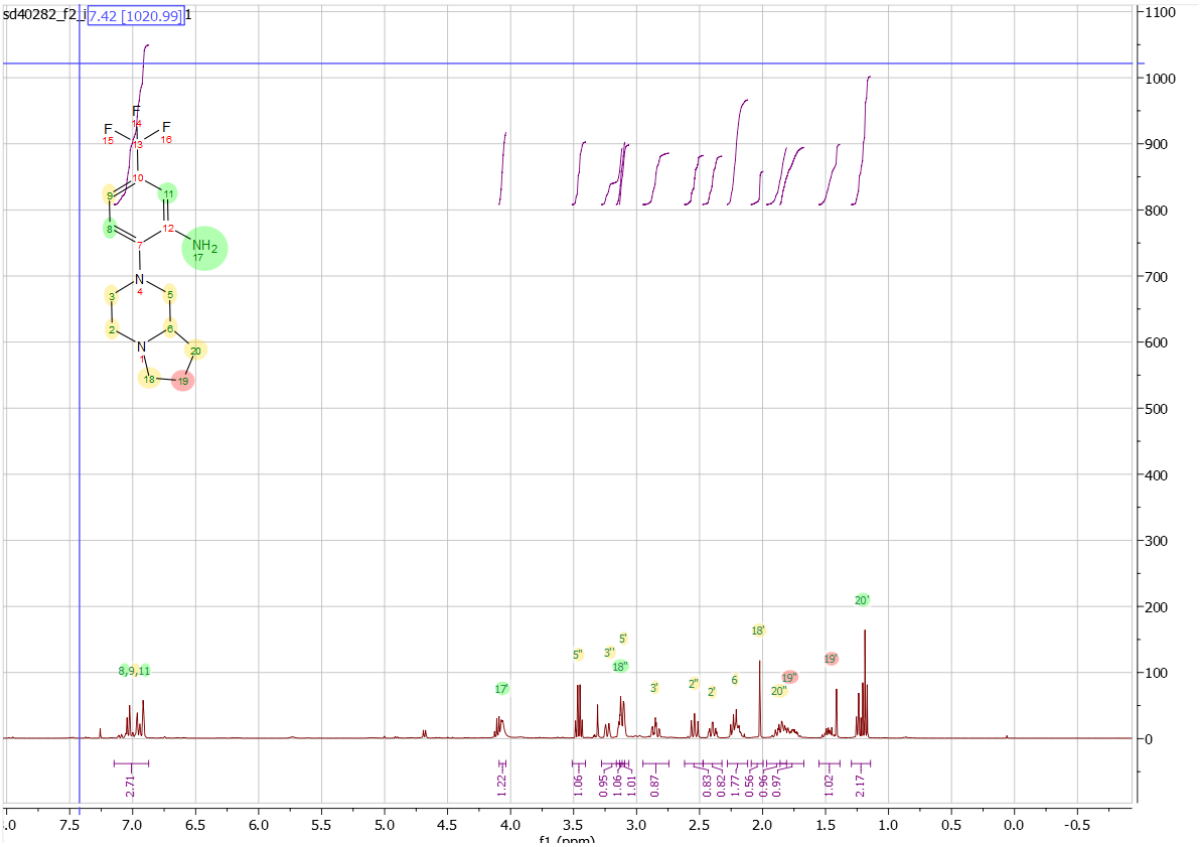

sd36904_Frag_1_A__PROTON_01
f1 (ppm)

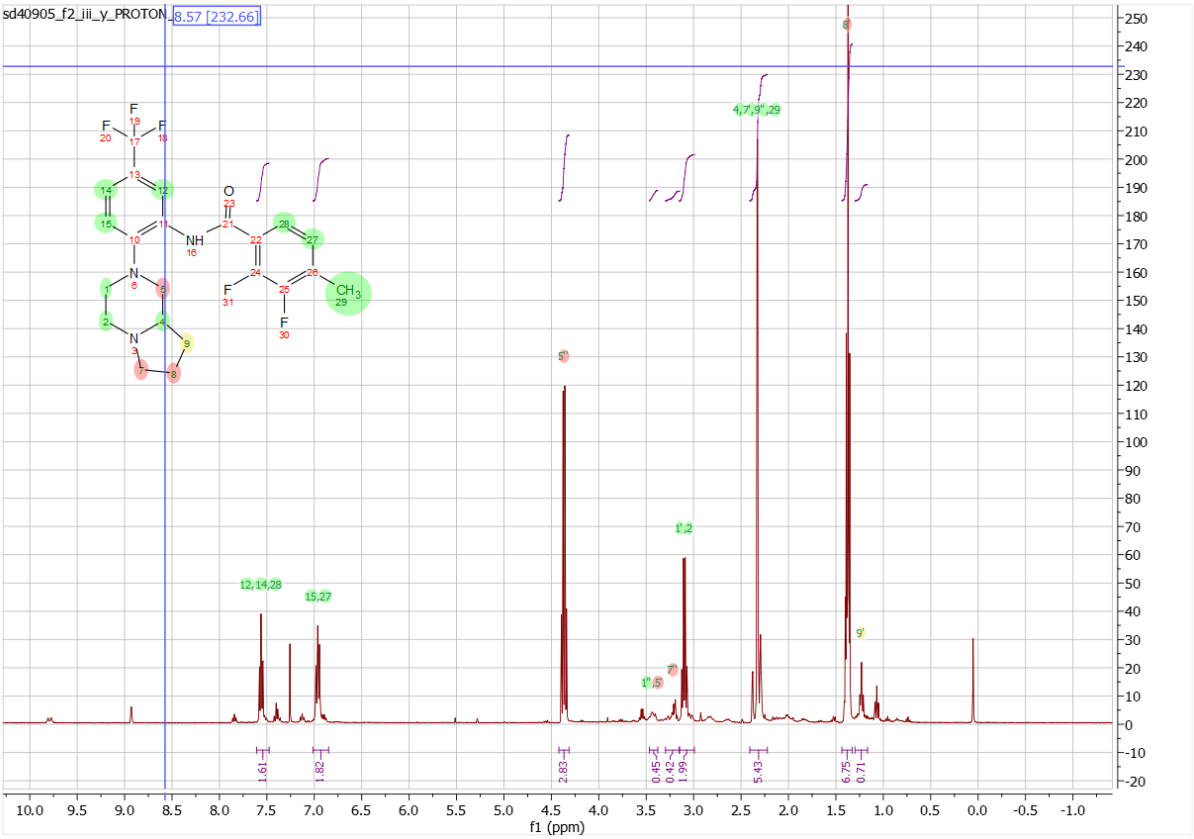

sd40822_f1_ii_8_PROTON_01
B (s) 7.04
A (s) 7.26
C (s) 6.97
F (q) 3.47
D (q) 4.11
E (s) 3.65
G (s) 1.42
f1 (ppm)

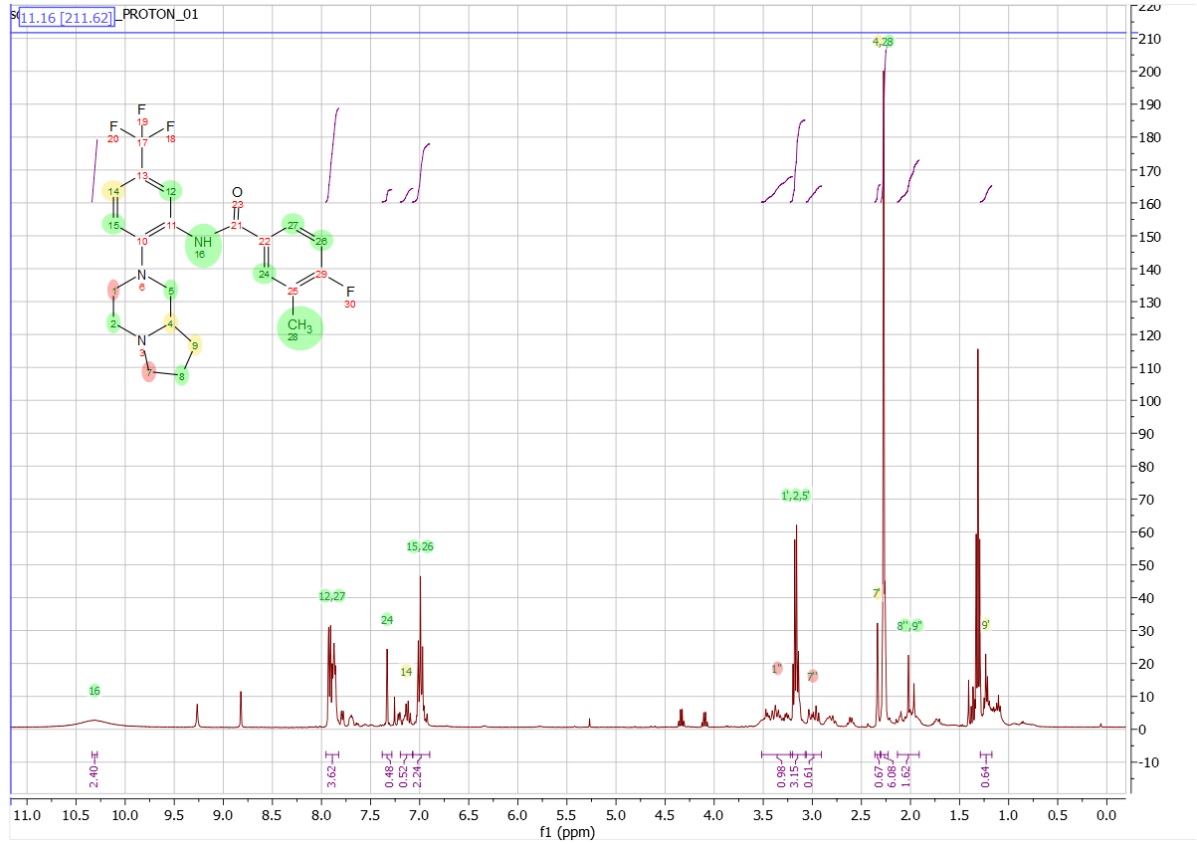

sd40902_f1_jii_x_PROTON_01
10.47 [128.94]
B (t)
7.67
E (m)
6.85
A (q)
7.93
D (dt)
7.07
C (m)
7.47
F (d)
5.26
G (q)
3.08
H (m)
1.30
I (td)
1.34
f1 (ppm)


sd40903_f1_jii_y_PROTON_01
f1 (ppm)


sd39955_F2_i__PROTON_01
f1 (ppm)


sd40282_f2_i
f1 (ppm)


sd40905_f2_iii_y_PROTON
8.57 [232.66]
f1 (ppm)


PROTON_01
11.16 [211.62]
f1 (ppm)