# Design, Synthesis and Biological Evaluation of Potential Inhibitors of S100P, a Protein implicated in Pancreatic Cancer

**Ramatoulie Camara**

Submitted to the University of Hertfordshire in partial fulfilment of the requirements of the degree of Doctor of Philosophy

**October 2015**

*To my Dad, Malang A. Camara, your memory still lives on*

## Acknowledgements

I would like to express my gratitude to my supervisory team: Dr Stewart B. Kirton, Dr Sharon Rossiter and Dr David Y. S. Chau. Thank you for all your cheerful optimism, support and guidance throughout my PhD journey. To the Worldwide Cancer Research (formerly known as the Association for International Cancer Research, AICR) and the University of the Hertfordshire, thank you for funding this project. Thank you to our collaborators Dr Tanja Crnogorac-Jurcevic and her team at Queen Mary, University of London, Mr Nasir Mahmoud in particular who carried out the biological screening presented in Chapter 3 of this thesis. I would like to extend my thanks to Ms Arshjot Grewal and Ms Aleksandra Kulikowska for their help with the synthesis of some of the analogues presented in Chapter 4. My gratitude goes to the staff at the Department of Pharmacy particularly Prof. Mire Zloh for his help when I wanted to run some molecular dynamic simulations. I would like to thank the technical staff especially Mr Virendra Shah, Dr Mark Scott and Mr James Stanley for all their help throughout my PhD studies. To the technical staff at CP Snow namely Mr Lee Rixon, Mrs Diana Francis, Mrs Sue Rawlins and their team, and everyone out there who made working in those labs a pleasure, I say thank you. To my fellow residents of LC101 both past, Miss Elizabeth Osoba, Dr Satyajit Shetage, and Dr Francesco Casserta, and present, Ms Soumaya Hakim, Ms Mubinah Beebeejaun, Ms Ilaria Passarini and Dr Ewelina Hoffman, thank you guys for all the fun memories we shared in the office! Miss Osoba and Dr Shetage, thank you for always being available any time to answer my never-ending thesis-related questions! To all my colleagues, past and present, especially my Gunner sister Ms Michelle Botha, and Mr Hassan Farrah, thank you all for your encouragement throughout this journey. To Emeritus Prof. Andrew Hutt, I would like to say a big thank you for taking the time to go through my Chapters, and giving me valuable and constructive feedback. I deeply appreciate this. To all my friends, near and far, thank you for being there for me. To my family, I thank you for your patience, love and support throughout my education. Finally, I would like to extend my deepest gratitude to my husband Anthony Macey, thank you for everything you have done for me, for being there when I needed a shoulder to lean on, for looking after me and making my writing-up period as stress-free as possible, thank you. S100P stole your wife but she has found her way back!

## Abstract

Pancreatic cancer is relatively uncommon. Despite its relative scarcity, it is the fourth-ranked cancer killer in the Western world with less than a 5% 5-year survival rate. The high mortality rate is due to the asymptomatic nature of the disease and the advanced stage at which it is usually diagnosed. S100P is a calcium-binding protein that has been shown to be highly expressed in the early stages of pancreatic cancer and has been proposed as a potential therapeutic target via the blocking of its interaction with its receptor RAGE, the receptor for advanced glycation end-products.

In this thesis, computational techniques were employed on the NMR ensemble of S100P (PDB Accession code 1OZO) to identify potential inhibitors of the S100P-RAGE interaction in the hope of identifying a series of novel leads that could be developed into clinical candidates for the treatment of pancreatic cancer. *In silico* studies identified putative binding sites at the S100P dimeric interface capable of accommodating cromolyn, an anti-allergy drug shown to bind to the protein both *in vitro* and *in vivo*. Virtual screening of >1 million lead-like compounds using 3D pharmacophore models derived from the predicted binding interactions between S100P and cromolyn, identified 9,408 "hits". These were hierarchically clustered according to similarities between chemical structures into 299 clusters and 77 singletons.

Biological screening of 17 of the "hits" identified from virtual screening stuidies, 4 of which were synthesised in-house, against pancreatic cancer cell lines identified five compounds that demonstrated an equal or greater capacity to reduce BxPC-3 S100P-expressing pancreatic cells' metastatic potential *in vitro* relative to cromolyn. Compound **24** in particular, showed significant ($p$<0.05) inhibition of invasion of these cells at a concentration of 100 μM that was comparable to cromolyn at the same concentration. This compound, structurally distinct from cromolyn, was successfully synthesised, purified and characterised in-house alongside 39 of its analogues. Biological screening of compound **24** and four of its analogues for anti-proliferative activity against BxPC-3 and Panc-1 pancreatic

cancer cell lines showed all five compounds significantly ($p$<0.0001) inhibiting proliferation in both cell lines at a concentration of 1 μM relative to the non-treated control. Hence, structurally distinct compounds that show promising inhibitory activity on the metastasis and proliferation of pancreatic cancer cells have been identified using a structure-based drug design methodology. These compounds, with further optimisation, could provide good starting points as therapeutic lead candidates for the treatment of pancreatic cancer.

## List of Figures

## List of tables

## Abbreviations

| | |
|---|---|
| 3D | three-dimensional |
| $Ac_2O$ | acetic anhydride |
| AcOH | acetic acid |
| ANOVA | analysis of variance |
| Ar | aromatic |
| Ar-C | aromatic carbon |
| Ar-H | aromatic hydrogen |
| ASR | age-standardised rates |
| br | broad |
| CA19-9 | carbohydrate antigen 19.9 |
| CADD | computer-aided drug design |
| CAM | chick chorioallantioc membrane |
| CDF | Cancer Drugs Fund |
| Cl | chlorine |
| $CO_2Et$ | propionate |
| $CO_2H$ | carboxylic acid |
| CT | computerised tomography |
| d | doublet |
| DCM | dichloromethane |
| dd | doublet of doublets |
| DMAP | 4-dimethylaminopyridine |
| DMEM | Dulbecco's Modified Eagle's Medium |
| DMSO | dimethyl sulfoxide |
| EGF | epidermal growth factor |
| ERCP | endoscopic retrograde cholangiopancreatography |
| ESI | electrospray ionisation |
| $Et_2O$ | diethyl ether |
| $Et_3N$ | triethylamine |
| EtOH | ethanol |
| EUS | endoluminal ultrasonography |
| FATCAT | Flexible structure Alignment by Changing Aligned Fragment Pairs with Twists |
| FBS | foetal bovine serum |
| FGI | functional group interconversion |
| FNAB | Endoscopic ultrasonography-guided fine-needle aspirational biopsy |
| AcOH | acetic acid |
| GBVI/WSA | Generalized-Born Volume Integral/Weighted Surface area |
| GI | gastrointestinal |
| $H_2O$ | water |
| HBA | hydrogen bond acceptor |
| HCl | hydrochloride |
| HTS | high-throughput screening |
| IARC | International Agency for Research on Cancer |
| IPMN | intraductal papillary neoplasm |
| IR | infrared |
| $K_2CO_3$ | potassium carbonate |

| | |
|---|---|
| LC-MS | liquid chromatography-mass spectrometry |
| LDH | lactate dehydrogenase |
| LibMCS | Library Maximum Common Substructure |
| LYVE1 | lymphatic vessel endothelial hyaluronan receptor 1 |
| M.p. | melting point |
| *m/z* | mass/charge ratio |
| MBD | MOE molecular database |
| MCN | mucinous cystic neoplasms |
| MCS | maximum common substructure |
| $Me_2N$ | dimethylamine |
| MeO | methoxy |
| MeOH | methanol |
| $MgSO_4$ | magnesium sulphate |
| MMFF94x | Merck Molecular Force Field 94 |
| MOE | Molecular Operating Environment |
| MRI | medical resonance imaging |
| MTS | 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium |
| MW | molecular weight |
| $NaBH_4$ | sodium borohydride |
| $NaHCO_3$ | sodium hydrogen carbonate |
| NaOAc | sodium acetate |
| NaOH | sodium hydroxide |
| ND | not determined |
| NHS | National Health Service |
| NMR | Nuclear Magnetic Spectroscopy |
| $NO_2$ | nitro |
| OD | optical density |
| ONS | Office for National Statistics |
| PanIN | pancreatic intraepithelial neoplasia |
| PDAC | pancreatic ductal adenocarcinoma |
| PDB | protein data bank |
| PET | positron emission tomography |
| pKa | acid dissociation constant |
| PLB | propensity for ligand binding |
| PPPD | pylorus preserving pancreaticoduodenectomy |
| PSP | protein–site–protein |
| q | quadruplet |
| qq | quartet of quartets |
| QSAR | quantitative structure activity relationship |
| RAGE | receptor for advanced glycation end-products |
| RASMs | readily available starting materials |
| REG1A | Regenerating Islet-Derived 1 Alpha |
| $R_f$ | retention factor |
| RMSD | root mean square deviation |
| RNAi | RNA interference |
| RPMI | Roswell Park Memorial Institute |
| s | singlet |

| | |
|---|---|
| SAR | structure-activity relationship |
| SBDD | structure-based drug design |
| SEM | standard error of mean |
| SIB | Swiss Institute of Bioinformatics |
| SMILES | Simplified Molecular-Input Line-Entry System |
| t | triplet |
| *t*-Bu | *tert*-butyl |
| TFF1 | Trefoil factor 1 |
| TGF | transforming growth factor |
| THF | tetrahydrofuran |
| TLC | thin-layer chromatography |
| TM | target molecule |
| TMS | tetramethylsilane |
| TNM | tumour lymph nodes metastasis |
| tPSA | total polar surface area |
| WHO | World Health Organisation |

# Table of contents

# CHAPTER 1: General Introduction

# 1. INTRODUCTION

Cancer is a collective term used to describe a group of diseases characterised by uncontrolled cell growth and division, with the potential to migrate to and invade distant tissues [1]. The World Health Organisation's (WHO) International Agency for Research on Cancer (IARC) estimated about 14 million new cancer cases and 8.2 million cancer-related deaths worldwide in 2012 [2]. In Europe, 3.45 million new cancer cases (excluding non-melanoma skin cancer) were estimated in 2012 with 1.75 million deaths occurring from the disease in the same year [3]. In the US the estimate for new cancer cases in 2014 was 1,665,540 with 585,720 deaths in the same period [4]. Data relating to cancer survival and mortality rates tend be 2-3 years behind the current year due to the time involved in data collection, analysis, compilation and dissemination [4].

The cost involved in cancer care – the term cancer used henceforth refers collectively to all diseases as described in [1] – and treatment is a huge financial burden for healthcare systems globally [5-8]. In the US, cancer care is projected to cost $173 billion in 2020, an increase of 39% from the estimated cost of $124.54 billion in 2010 [9] and according to the UK's Department of Health, cancer care costs the National Health Service (NHS) £5 billion annually [10]. Loss in productivity due to premature cancer-related deaths, or from family members taking time off work to care for sick relatives, also has an indirect but significant economic impact [11-13].

Many cancers have an unknown aetiology. However mutations in genes that play key roles in tumour suppression such as *p53* [14-17], *BRCA1* and *BRCA2* [18-

20], and/or over-expression of proto-oncogenes – normal genes that have the potential to differentiate into cancer-promoting genes i.e. oncogenes – have been linked with the disease [21-24]. These mutations can be hereditary or result from exposure to external factors such as smoking in lung cancer [25-28].

The search for a cure for almost all types of cancer has been ongoing for more than five decades with billions spent on research [29, 30]. Yet, despite the advances in scientific and medical knowledge, technology, and a deep and better understanding of the human body, a cure is yet to be found. Whilst there has been progress in chemotherapy since the use of nitrogen mustard compounds as chemotherapeutic agents in the 1940s [31], reducing toxicity of many anti-cancer drugs still remains a huge challenge [32-34]. Since cancer drugs are designed to target rapidly dividing cancer cells, healthy cells with similar characteristics are often susceptible to these drugs. Temporary alopecia (hair loss) is a common side effect of chemotherapy because many anti-cancer drugs impair the mitotic process of actively dividing hair follicles [35].

Despite the lack of a cure, early screening [36, 37], better screening methods such as computed tomography (CT, [38]), mammography [39], positron emission tomography (PET) scanning [40], medical resonance imaging (MRI, [41]), chemotherapy, radiotherapy, and changes in lifestyle behaviours have resulted in an improvement in incidence rates for some cancers. For example, a decline in smoking among men has seen a fall in lung cancer incidence for this group in Europe since the sixties [42, 43]. Lung cancer accounts for the top cause of cancer deaths in the developed world, followed by colon, breast, prostate and pancreatic cancers [4, 44-

46]. However, prognoses for lung, breast, prostate and colon cancers have improved considerably over time when compared to prognoses for pancreatic cancer [4, 47, 48]. Pancreatic cancer, as will be discussed later, remains one of the most lethal cancers [49-51].

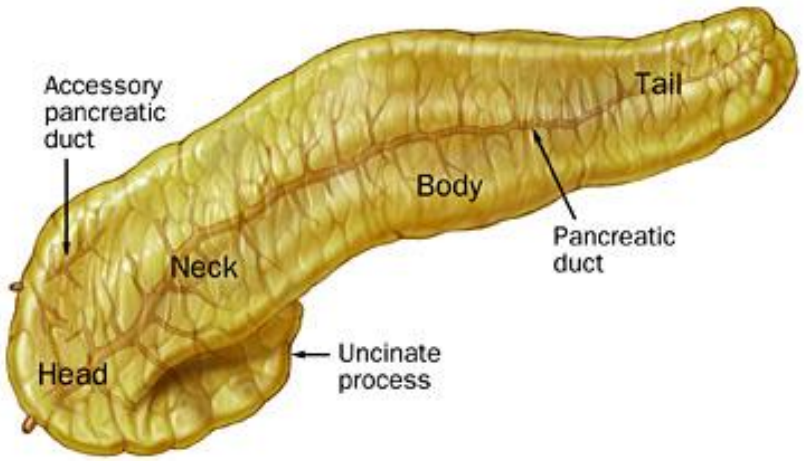
In this thesis, work will be presented relating to studies on one of the proposed clinical markers for pancreatic cancer, S100P, as a therapeutic target. This protein has been reported to be useful in the early detection of the disease but little has been done to target it as a point of therapy. Using computational, synthetic and biological methods, this thesis will show a medicinal chemistry approach to identifying a lead compound against S100P using a rational structure-based drug discovery. It is hoped that the results of this project will contribute significantly towards the search for a cure for pancreatic cancer and explore S100P as a therapeutic target.

## 1.1 The pancreas

The pancreas (Figure 1.1A) is a digestive organ and an endocrine gland situated in the peritoneal cavity [52]. It is covered by the first part of the small intestine and surrounded in front by the stomach, liver, left kidney, spleen and aorta (Figure 1.1B). Its location in the body makes it a rather inaccessible organ [53, 54]. The pancreas is divided into four parts: the head, neck, body and tail. Histologically, almost 80% of the pancreas is made up of exocrine acinar cells and their associated ducts, while the Islet of Langerhans, characterised by their distinctive clusters within the exocrine cells, make up the rest of the this dual-functioning organ [55].

As a digestive organ, the pancreas produces enzymes and fluids through its acinar cells and ducts [52]. These are then secreted into the duodenum via the ducts where they mix with the chyme from the stomach to help with digestion. The endocrine function of the pancreas is carried out by cells in the Islets of Langerhans. These synthesise and secrete the hormones insulin, glucagon, and somatostatin, which are involved in glucose metabolism [55].

**A**




**B**

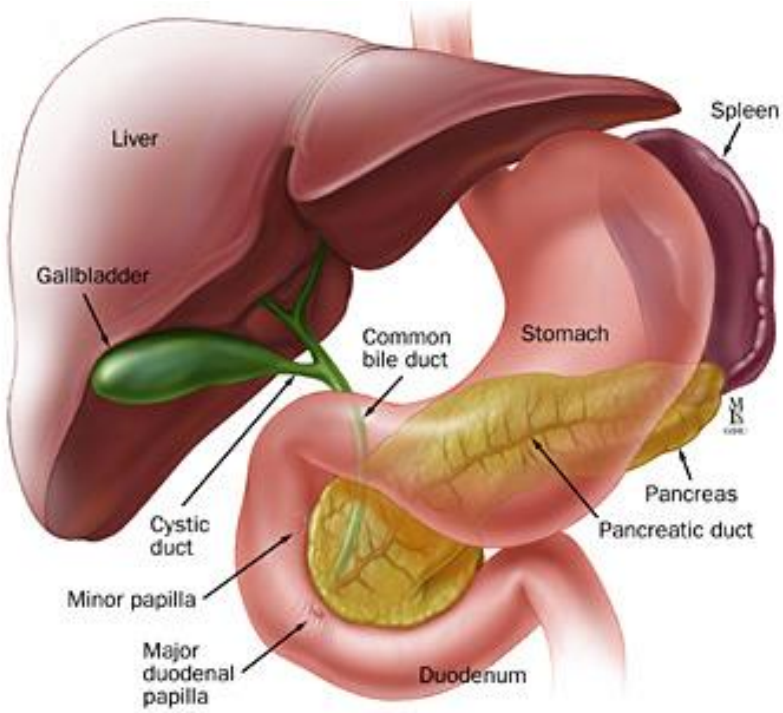



**Figure 1.1**: The pancreas. A) The pancreas is divided into four main parts: the head, neck, body and tail. B) Location of the pancreas in the duodenal cavity relative to organs nearby. (Images taken from http://goo.gl/ktQ6rW. Last accessed 10th October, 2015).

### 1.1.1 Diseases

Both the endocrine and the exocrine components of the pancreas can be affected by disease or tumours [52, 55]. Either component can be affected by disease independent of the other. For example, diseases that affect exocrine cells of the pancreas may not generally affect the endocrine cells producing hormones. The main diseases that can afflict the pancreas are pancreatitis (acute or chronic), pancreatic endocrine tumours (PETs) and pancreatic ductal adenocarcinomas (PDACs) [54, 56].

#### 1.1.1.1 **Pancreatitis**

The term pancreatitis means inflammation of the pancreas [52, 54]. This can occur when digestive enzymes produced by the acinar cells are released in the pancreas instead of their intended destination (duodenum) and they, in turn, digest the pancreas [52]. In a 1983 meeting under the auspices of the Pancreatic Society of Great Britain and Ireland, a group of international doctors specialising in pancreatitis offered the following definition to aid in the classification of pancreatitis:-

*''a continuing inflammatory disease of the pancreas, characterised by irreversible morphological change, and typically causing pain and/or permanent loss of function''* [57].

In acute pancreatitis, the pancreas becomes inflamed over a short period of time resulting in severe abdominal pain and distention, nausea and vomiting [58]. Gallstones and alcohol consumption have been reported to be the main cause of acute pancreatitis although in some cases the cause seems to be idiopathic [59]. While this condition affects many people and has been reported to be on the rise,

many of those affected only require brief hospital admission and few suffer fatality [60].

Chronic pancreatitis has a more dismal prognosis with a higher mortality rate compared to acute pancreatitis [61]. Although alcohol consumption has been reported to be the major contributing factor to chronic pancreatitis (in Western countries), some cases have been attributed to genetic mutations, whilst, others have an idiopathic aetiology [62]. Diminished pancreatic function has been reported to be a major sign of chronic pancreatitis [63] although other diagnostic tools such as endoscopic ultrasonography and invasive techniques are needed to confirm the presence of the disease. Acute pancreatitis acquired from severe alcohol abuse has also been linked to the subsequent development of chronic pancreatitis [64]. Like acute pancreatitis, pain is a major feature of chronic pancreatitis and takes centre stage in the management of the disease in addition to other measures such as use of painkillers e.g. opiates [65] and abstinence from alcohol (if the latter is factor) [62].

#### 1.1.1.2 **Pancreatic cancer**

More than 95% of pancreatic cancers are of exocrine origin [52, 56]. However, despite the exocrine pancreas being mainly composed of acinar cells, these are believed to make up less than 1% of the total pancreatic cancers with the vast majority arising from the pancreatic ducts (also known as adenocarcinomas) [66]. Pancreatic ductal adenocarcinoma (PDAC) is the most common form, accounting for more than 85% of all pancreatic cancer cases [67]. Other types of cancers affecting the exocrine pancreas include acinar cell carcinoma, adenosquamous carcinoma, giant cell tumour, intraductal papillary-mucinous neoplasm (IPMN), mucinous

cystadenocarcinoma, pancreatoblastoma, serous cystadenocarcinoma and solid and pseudopapillary tumours [68].

Advances in molecular biology techniques have led to a better understanding of the cellular changes that occur leading to the development of pancreatic cancer [69]. It is believed that invasive pancreatic ductal adenocarcinoma (PDAC) – any reference to pancreatic cancer henceforth refers to PDAC – begins from non-invasive precursor lesions such as mucinous cystic neoplasms (MCNs), intraductal papillary neoplasms (IPMNs) and pancreatic intraepithelial neoplasia (PanIN) [70, 71]. The first two lesions, MCNs and IPMNs, are non-inflammatory cystic lesions and can be distinguished from each other by the presence of an ovarian-type stroma in the former which is absent in the latter [72]. PanIN, which is subdivided into four groups: PanIN-1A, PanIN-1B, PainIN-2, and PanIN-3, is <5 mm in diameter and is the most common neoplastic lesions present in invasive adenocarcinoma [73, 74]. IPMNs in comparison to PanIN, are larger, >5 mm in diameter, and less common in invasive cancer [74, 75]. The distinct cellular and molecular changes that occur in PanINs (Figure 1.2) have been suggested as a model to follow the progression of pancreatic cancer from precursor lesions to infiltrating neoplasm [69].

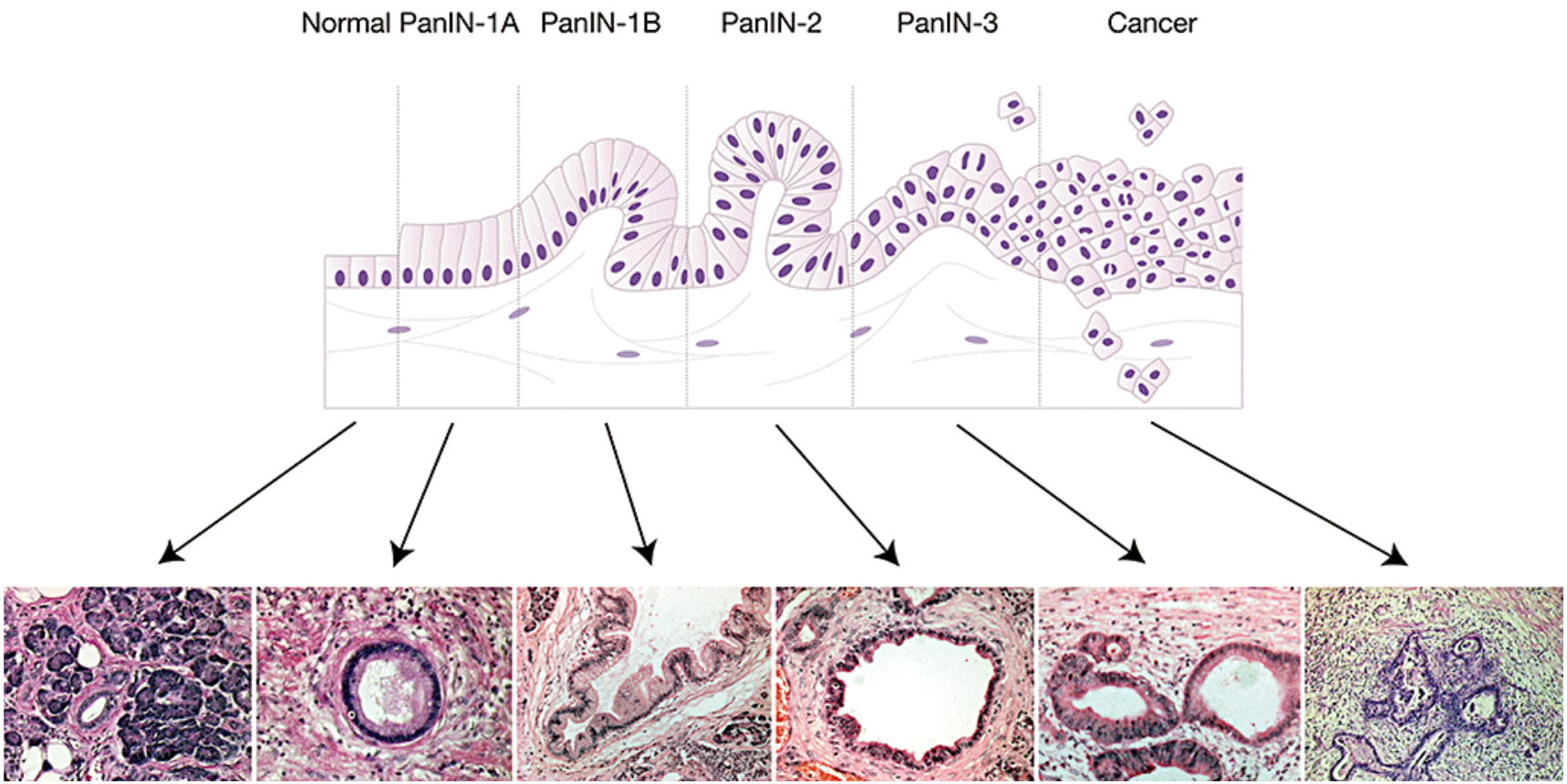



**Figure 1.2**: Progression model for pancreatic cancer. The different stages of cellular changes that take place during the development of PanIN into infiltrating cancer. (Image reproduced from Chang *et al*. [76] with permission from John Wiley and Sons, 2015).

Unlike exocrine pancreatic adenocarcinomas, cancers of the endocrine pancreas are less common, occurring more frequently in the young and have a much better prognosis [66]. Pancreatic endocrine tumours arise from cells of the Islet of Langerhans and are usually associated with a lack of, or excessive, secretion of hormone from cancer cells [77]. Although the molecular origins of pancreatic endocrine tumours are poorly understood, it is believed that genes that are mainly associated with exocrine pancreatic cancer are not involved in pancreatic endocrine tumours [78].

#### 1.1.1.2.1 Pancreatic ductal adenocarcinoma (PDAC) incidence and mortality rates

Pancreatic ductal adenocarcinoma (PDAC) is the fourth most common cause of cancer deaths in the United States and Europe behind lung, colon and breast cancers [3, 4, 66, 79, 80]. Unlike lung, colon, breast and prostate cancers which have high incidence rates, PDAC has a modest incidence rate and does not feature in the top five common cancers [3, 4]. However, the mortality rate from this cancer is as high as its incidence rate, highlighting its lethality [4]. In the USA, 46,420 new cases were estimated in 2014, with 39,590 deaths in the same year [4]. This was an increase from the number estimated in 2012 when more than 43,000 new cases of pancreatic cancer were predicted with an estimated 37,000 deaths in the same year [81]. In Europe, the estimated incidence for 2012 was 103,770, with 104,460 deaths (Table 1.1, [3]). According to the Office for National Statistics (ONS) for the UK, approximately 7,300 people were newly diagnosed with pancreatic cancer in England in 2011, with around 7,000 dying from the disease in the same period [82].

**Table 1.1**: Estimated numbers of new cancer cases and deaths from cancer (thousands), age-standardised rates (ASRs) (per 100,000) by sex and cancer site in Europe in 2012. (Table reprinted from Ferlay *et al*. [3] with permission from Elsevier Ltd, 2015).

| | **Incidence** | | | | | | **Mortality** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Both sexes | | Male | | Female | | Both sexes | | Male | | Female | |
| | Cases | ASR (E) | Cases | ASR (E) | Cases | ASR (E) | Deaths | ASR (E) | Deaths | ASR (E) | Deaths | ASR (E) |
| Oral cavity and pharynx | 99.6 | 11 | 73.9 | 18.2 | 25.8 | 4.9 | 43.7 | 4.7 | 34.2 | 8.4 | 9.4 | 1.6 |
| Oesophagus | 45.9 | 4.7 | 35.1 | 8.4 | 10.8 | 1.8 | 39.5 | 3.9 | 30.3 | 7.1 | 902 | 1.4 |
| Stomach | 139.6 | 13.7 | 84.2 | 19.5 | 55.4 | 9.3 | 107.3 | 10.3 | 63.6 | 14.6 | 43.7 | 7 |
| Colon and rectum | 446.8 | 43.5 | 241.6 | 55.7 | 205.2 | 34.7 | 214.7 | 19.5 | 113.2 | 25.2 | 101.5 | 15.4 |
| Liver | 63.4 | 6.2 | 42.8 | 10 | 20.6 | 3.3 | 62.1 | 5.9 | 39.9 | 9.1 | 22.2 | 3.4 |
| Gallbladder | 29.7 | 2.7 | 11.9 | 2.7 | 17.9 | 2.8 | 20.9 | 1.9 | 7.9 | 1.8 | 13 | 2 |
| **Pancreas** | **103.8** | **10.1** | **51.9** | **12.1** | **51.8** | **8.3** | **104.5** | **9.9** | **52.6** | **12.2** | **51.9** | **8.1** |
| Larynx | 39.9 | 4.4 | 36 | 8.8 | 3.9 | 0.8 | 19.8 | 2.1 | 18.1 | 4.3 | 1.7 | 0.3 |
| Lung | 409.9 | 41.9 | 290.7 | 68.3 | 119.2 | 21.6 | 353.5 | 35.2 | 254.4 | 59.1 | 99 | 17.2 |
| Melanoma of skin | 100.3 | 11.1 | 47.2 | 11.4 | 53.1 | 11 | 22.2 | 2.3 | 12.1 | 2.8 | 10.1 | 1.8 |
| Breast | | | | | 463.8 | 94.2 | | | | | 131.2 | 23.1 |
| Cervix uteri | | | | | 58.3 | 13.4 | | | | | 24.4 | 4.9 |
| Corpus uteri | | | | | 98.9 | 19.3 | | | | | 23.7 | 3.9 |
| Ovary | | | | | 65.5 | 13.1 | | | | | 42.7 | 7.6 |
| Prostate | | | 416.7 | 96 | | | | | 92.2 | 19.3 | | |
| Testis | | | 21.5 | 5.8 | | | | | 1.6 | 0.4 | | |
| Kidney | 115.2 | 12.1 | 71.7 | 17.2 | 43.4 | 8.1 | 49 | 4.7 | 31.3 | 7.2 | 17.7 | 2.8 |
| Bladder | 151.2 | 14.4 | 118.3 | 26.9 | 32.9 | 5.3 | 52.4 | 4.5 | 39.5 | 8.5 | 12.9 | 1.8 |
| Brain, nervous system | 57.1 | 6.6 | 30.7 | 7.8 | 26.4 | 5.6 | 45 | 4.9 | 24.6 | 6 | 20.4 | 4 |
| Thyroid | 52.9 | 6.3 | 12.3 | 3.1 | 40.7 | 9.3 | 6.3 | 0.6 | 2.1 | 0.5 | 4.3 | 0.7 |
| Hodgkin lymphoma | 17.6 | 2.3 | 9.3 | 2.5 | 8.3 | 2.1 | 4.6 | 0.5 | 2.6 | 0.6 | 2 | 0.4 |
| Non-Hodgkin lymphoma | 93.4 | 9.8 | 49.5 | 11.9 | 43.9 | 8 | 37.9 | 3.5 | 20.3 | 4.6 | 17.5 | 2.7 |
| Multiple myeloma | 38.9 | 3.8 | 20.5 | 4.7 | 18.4 | 3.1 | 24.3 | 2.2 | 12.2 | 2.7 | 12.1 | 1.8 |
| Leukaemia | 82.3 | 8.8 | 46.4 | 11.3 | 35.9 | 6.9 | 53.8 | 5.1 | 29.5 | 6.7 | 24.3 | 3.9 |
| All sites but non-melanoma skin cancers | 3439.6 | 355.7 | 1829.1 | 429.9 | 1610.5 | 306.3 | 1754.6 | 168 | 975.9 | 222.6 | 778.6 | 128.8 |

Of the top five common causes of cancer death, pancreatic cancer has the worst head-to-head prognosis, with a less than 3% 5-year survival rate ([4], Figure 1.3). This is mainly due to the asymptomatic nature of the disease and the late stage at the time of diagnosis [83]. Risk factors for developing the disease include age, smoking, alcohol consumption, chronic pancreatitis [84, 85] and genetic predisposition/defects [68].

#### 1.1.1.2.2 Diagnosis and staging

Some of the tools used to currently diagnose PDAC in the UK include ultrasound scanning, computerised tomography (CT) scanning, magnetic resonance imaging (MRI) scanning, endoluminal ultrasonography (EUS), laparoscopy, endoscopic retrograde cholangiopancreatography (ERCP), and biopsy [86]. Endoscopic ultrasonography-guided fine-needle aspirational biopsy (FNAB) is also used to aid in diagnosis and staging of PDAC [87-89]. These methods are used in combination to help increase the accuracy of the diagnosis. However, due to the asymptomatic nature of the disease and a lack of adequate indicative biomarkers, these techniques are not normally sensitive enough to detect the disease in its early stages [86]. Diagnosis often only occurs once the cancer has metastasised to nearby structures such as the bile duct, mesenteric and coeliac nerves, and duodenum [90]. Cancer of the head of the pancreas may sometimes lead to the obstruction of the bile duct leading to painless jaundice as one of the symptoms [91]. The similarities between PDAC and chronic pancreatitis [92] can make diagnosing pancreatic cancer a challenging task, with the latter being a high risk for the development of the cancer [85, 93, 94]. This can sometimes lead to initial misdiagnosis of the cancer as chronic pancreatitis and cause a delay to timely intervention and treatment.

Once the cancer is detected, it is “staged” based on the size or the tumour (T), the involvement of lymph nodes (N), and metastasis (M) of the cancer to nearby tissues or organs ([95], Figure 1.4, Table 1.2). In addition to the TNM classification, there is also a stage grouping that can include a combination of the TNM stages at different levels (Table 1.3). The accurate staging of the cancer is crucial in order to offer the best possible outcome for treatment, hence the ability and sensitivity of the diagnostic tool(s) used to detect the disease for accurate staging is vital.

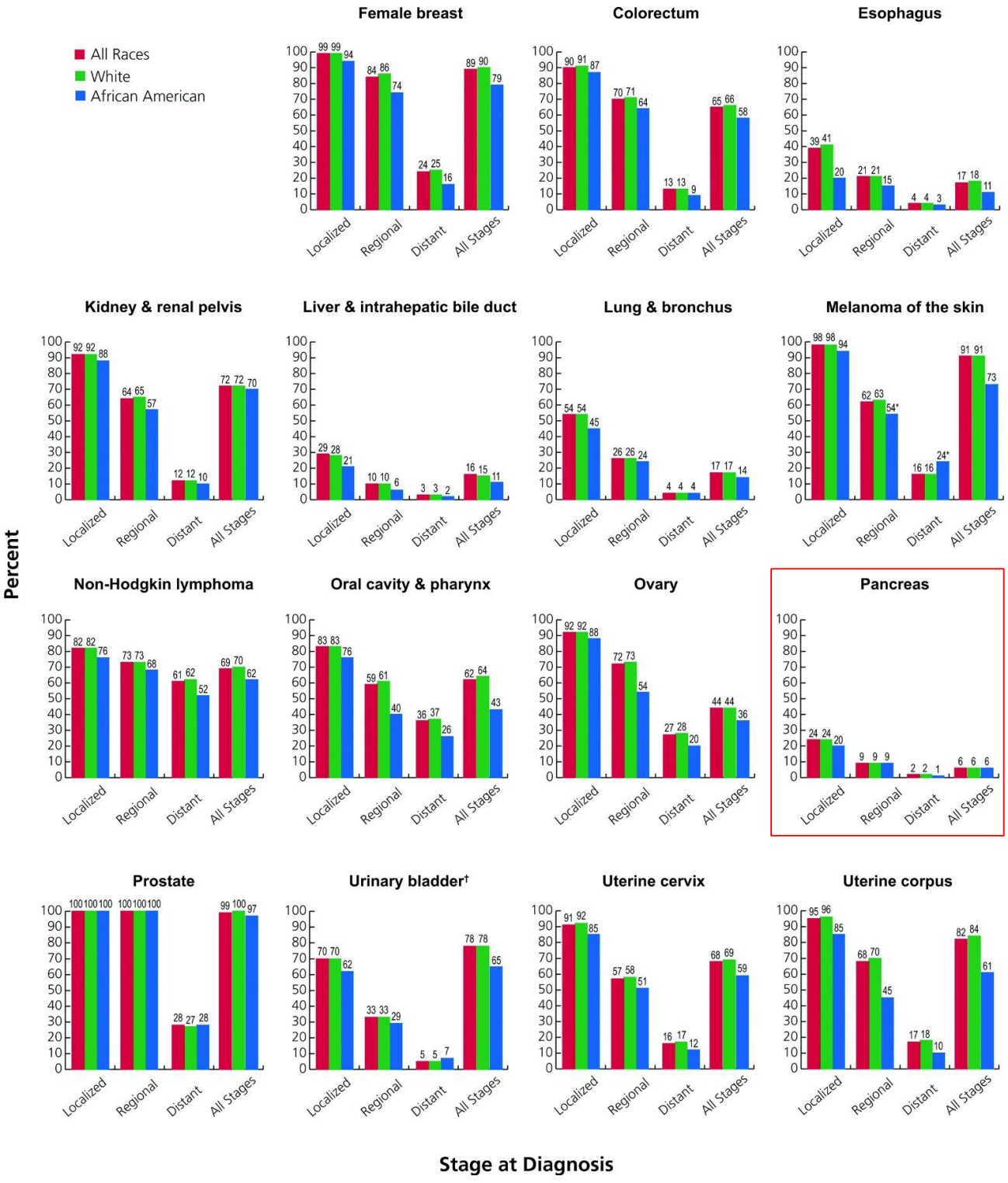



**Figure 1.3**: Five-year survival rates for selected cancers by race and stage at diagnosis, US, 2003-2009. Pancreatic cancer (enclosed) has the worst survival rate of all cancers shown. (Image reproduced from Siegel *et al*. [4] with permission from John Wiley and Sons, 2015).

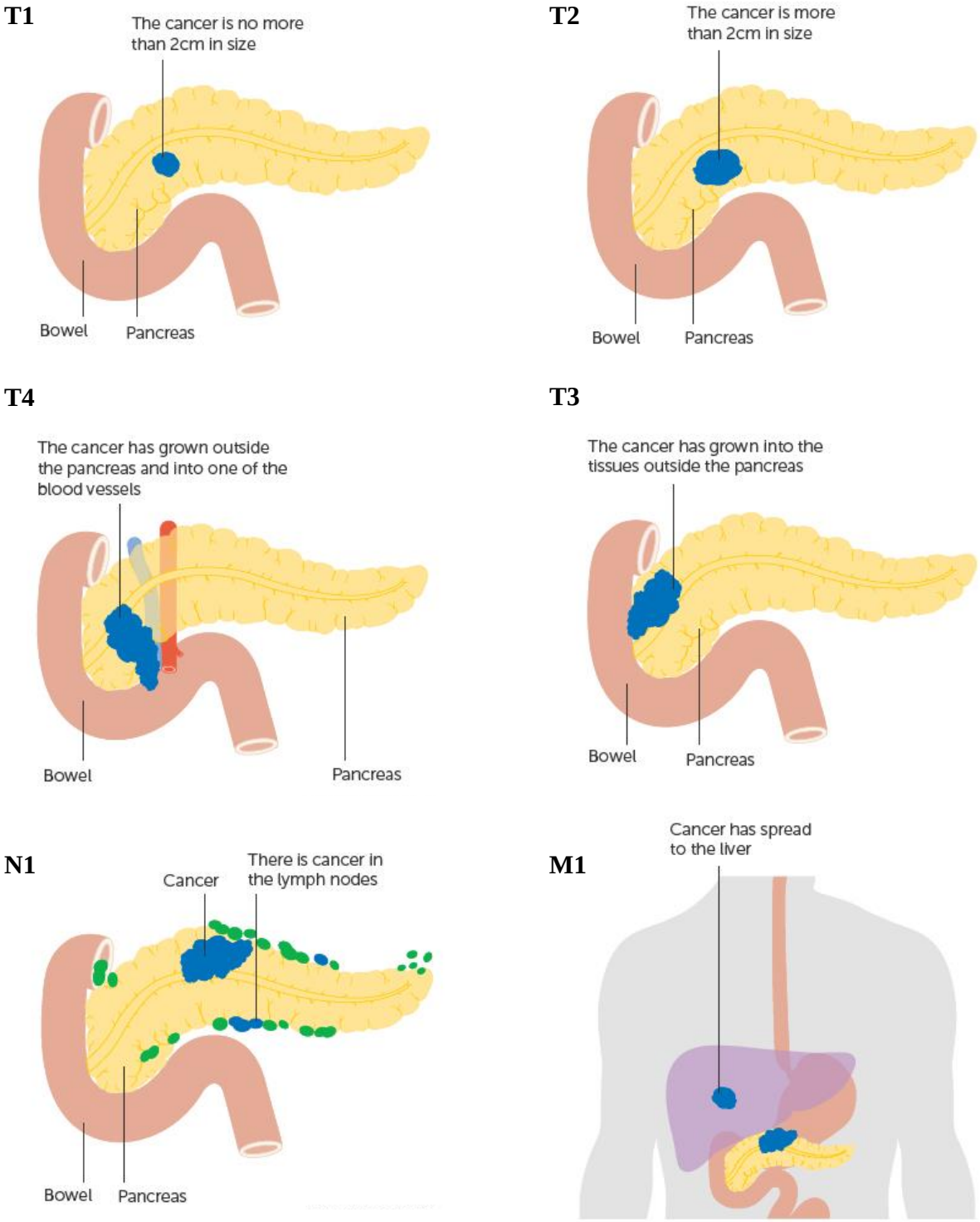



**Figure 1.4**: TNM staging of pancreatic cancer. The staging is used to describe the size of the primary tumour (T), presence of lymph nodes with cancer (N), and metastasis of the cancer to other tissues (M). (Images taken from http://goo.gl/CPkI6f. Last accessed 11th October, 2015).

**Table 1.2**: Breakdown of the TNM (Tumour, lymph Nodes, Metastasis) staging of pancreatic cancer. (Table adapted from Evans *et al*. [95]).

| | **Primary tumour (T)** | **Lymph nodes (N)** | **Distant metastasis (M)** |
|---|---|---|---|
| **Tis** | Carcinoma *in situ* (includes PanINs) | | |
| **X** | Primary tumour cannot be assessed | Regional lymph nodes cannot be assessed | Distant metastasis cannot be assessed |
| **0** | No evidence of primary tumour | No regional lymph node metastasis | No distant metastasis |
| **1** | Tumour limited in pancreas, 2 cm or less in diameter | Regional lymph node metastasis | Distant metastasis |
| **2** | Tumour limited to pancreas, more than 2 cm in diameter | | |
| **3** | Tumour extends beyond the pancreas but without involvement of celiac axis or superior mesenteric artery | | |
| **4** | Tumour involves the celiac artery (unresctable primary tumour) | | |

**Table 1.3**: Stage grouping of pancreatic cancer in combination with the TNM staging. (Table adapted from Evans *et al*. [95]).

| | Primary tumour (T) | Lymph nodes (N) | Distant metastasis (M) |
|---|---|---|---|
| **Stage 0** | Tis | N0 | M0 |
| **Stage IA** | T1 | N0 | M0 |
| **Stage IB** | T2 | N0 | M0 |
| **Stage IIA** | T3 | N0 | M0 |
| **Stage IIB** | T1 | N1 | M0 |
| | T2 | N1 | M0 |
| | T3 | N1 | M0 |
| **Stage III** | T4 | Any N | M0 |
| **Stage IV** | Any T | Any N | M1 |

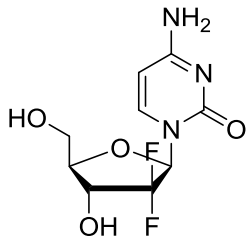
#### 1.1.1.2.3 Treatment

The type of treatment offered to combat PDAC usually depends on the stage at which the cancer is detected. Once the cancer has been conclusively confirmed, initial treatment is then based on whether the cancer is localised within the pancreas (0, IA and IB), locally invasive (IIA, IIB and III) or has metastasised to distant organs (IV) [91]. Surgery is offered as the first line of treatment to patients where the cancer has not metastasised [91, 96]. However because pancreatic cancer is usually quite advanced at the time of diagnosis only 5-25% of patients presenting with tumours are amenable to resection [79]. The three main types of operations offered to patients are the pylorus preserving pancreaticoduodenectomy (PPPD) where the head of the pancreas, duodenum, part of the bile duct and gall bladder are removed [97, 98]; the Whipple procedure – which is similar to PPPD but also involves removal of the lower part of the stomach [99]; and total pancreatectomy – a procedure where the whole pancreas is removed in addition to the duodenum, part of the stomach, part of the bile duct, gall bladder, the spleen, and surrounding lymph nodes [100]. Radiation therapy (radiotherapy) is often offered concomitantly with surgery for better efficacy [101].

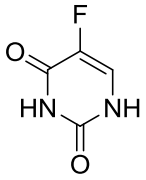
For patients in whom the cancer is at an advanced stage, surgery is often too late and not the route of therapy offered [102]. For these people, their only hope lies with chemotherapy and radiotherapy. In the UK, current chemotherapeutic interventions for pancreatic cancer include treatment with gemcitabine [96, 102, 103], an anticancer agent which is sometimes used alongside 5-fluorouracil (5-FU, [104], Figure 1.5). Nab-paclitaxel, the protein-bound form of paclitaxel (Figure 1.5) given in combination with gemcitabine [105], is the only drug on the Cancer Drugs

Fund (CDF) list in the UK for untreated metastatic pancreatic cancer [106]. This fund was established as an additional resource of funding to provide access to cancer drugs otherwise not routinely available on the National Health Service (NHS, [106]). However, this drug is being withdrawn from the CDF list because it is deemed to be economically unviable due to the perceived diminished benefits it offers to patients in comparison to its high cost [106], underlining the urgent need for drugs with better efficacy. Other treatments currently under clinical trial phases include the use of gemcitabine in combination with cytotoxic compounds other than 5-FU [107-109].

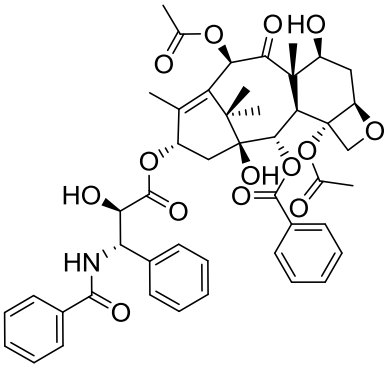
Chemotherapeutic intervention with these medicines only delays the return of the cancer and offers patients a modest extension of lifespan [110], with gemcitabine and 5-FU reportedly conferring only an 18% and 2% increased survival rate respectively at 12 months in patients with advanced pancreatic cancer [104]. The survival rate for nab-paclitaxel in combination with gemcitabine was found to be 35% after 12 months compared to 22% for the gemcitabine-only treatment group in Phase III clinical trials [111]. PDAC has a high recurrence rate that contributes to the high mortality rate. Survival rates after resection are low [112], and less than 5% of people diagnosed with the cancer survive beyond five years [67, 113].




**Gemcitabine (1)**




**5-Fluorouracil (2)**




**Paclitaxel (3)**

**Figure 1.5**: Chemotherapeutic agents used to treat pancreatic cancer.

1.1.1.2.4 Current research

Recently a lot of research has been carried out to develop a molecular [114-116] and genetic profile of pancreatic adenocarcinoma [117, 118]. Implicated genes include *KRAS, TP53, INK4A*, as well as abnormal or genetic mutations of the epidermal growth factor (EGF) and transforming growth factor (TGF)-β signalling pathways [69, 70, 119]. Carbohydrate antigen 19-9 (CA19-9) is a tumour marker present in pancreatic and biliary tract cancers [120] and has been used as a marker for pancreatic cancer in the advanced stages of the disease [121]. Its use as a clinical marker is however not encouraged, as it can give unreliable results given that it is also highly expressed in pancreatitis, cirrhosis [120] and other diseases involving bile ducts [120, 122]. It is for these reasons that, in 2006, the American Society of Clinical Oncology (ASCO) advised against its use for this purpose [123]. Their recommendation is to use the marker to monitor levels in people receiving chemotherapy.

Consequently, the need arises to find a marker(s) that can identify pancreatic cancer in its early stage to enable adequate intervention for better prognosis. A lot of research into PDAC is geared towards such an outcome and recently, three proteins, LYVE1, REG1A and TFF1, present in urine were reported to be able to differentiate between patients with, and without pancreatic cancer with better accuracy compared to CA19-9 [124]. The findings of this study offers an inexpensive, non-invasive method for early diagnosis of the disease, which is very promising. However, it remains to be seen how successful the results translate clinically. S100P is another small protein that has been shown to be highly expressed in early development of

pancreatic cancer, and has the potential to be used as a clinical marker for the disease [114, 121, 125, 126].

## 1.2 S100 proteins

S100 proteins are so named due to their 100% solubility in ammonium sulphate and were first identified in bovine brain cells approximately 50 years ago [127]. Since then 24 human isoforms have been identified [128-130].

S100 proteins are EF-hand type $Ca^{2+}$-binding proteins involved in signal transduction, regulation of protein phosphorylation, enzyme activities, transcription factors, $Ca^{2+}$ homeostasis, cell cycle progression and cell motility [128, 131]. They form the largest subgroup of the EF-hand $Ca^{2+}$-binding superfamily of proteins but unlike other $Ca^{2+}$ binding proteins such as calmodulin, troponin C, myosin light chains and parvalbumin, S100 proteins only act as modulators and are not involved in altering key cellular functions [132].

The term EF was first coined by Kretsinger and Nokolds [133] when they were describing the carp muscle calcium-binding protein parvalbumin. EF was used to describe two of the six α-helices of the main chain of the parvalbumin protein: helices E and F and their relationship to the bound calcium ion (Figure 1.6). In EF-hand $Ca^{2+}$-binding proteins, the EF region can be visualised as a clenched right hand with the forefinger and thumb extended at an approximate right angle to each other (Figure 1.6B). The extended thumb represents the direction of the COOH-terminus of helix F while the pointed forefinger is directed towards the $NH_2$-terminus of helix E. The clenched fingers represent the EF loop where the calcium ion is bound [133].

The EF-hand has a helix-loop-helix conformation and each hand is able to bind two calcium ions per protein [128].

Besides the canonical EF-hand on the C-terminus (with 12 amino acids) which is common to all EF-hand proteins, S100 proteins also exhibit another "pseudo" EF-hand present on the N-terminus (with 14 amino acids, Figure 1.6C) [134]. In S100 proteins, the N-terminal domain is made up of helices I and II ($H_I$ and $H_{II}$) joined by loop 1 ($L_1$) while the C-terminal domain is formed by helices III and IV linked by loop 3 ($H_{III}$-$L_3$-$H_{IV}$) (Figure 1.6A). This means that each S100 protein with the exception of S100A10 [135] can bind two $Ca^{2+}$ ions. However, since S100 proteins mainly exist as homodimers or in some cases heterodimers with other family members [136], each dimer is able to bind four calcium ions. Calcium binding in the different loops is not identical (Figure 1.6C); the C-terminus has a higher affinity for binding calcium ions than the N-terminus [137-140]. However, interaction with target proteins is not always calcium dependent and studies have shown that some S100 proteins are able to bind to their target with or without bound calcium ions [141, 142]. Binding of calcium ions is believed to induce a conformational change in the protein which results in exposure of a hydrophobic binding domain that helps with target recognition and binding [140]. In addition to binding calcium ions, some S100 proteins also bind other metal ions such as $Mg^{2+}$, $Zn^{2+}$ or $Cu^{2+}$ [143, 144] and sometimes binding of these metals increase the affinity for calcium binding [138].

Most S100 proteins are believed to exert their function by interacting with RAGE, the receptor for advanced glycation end-products [145, 146]. RAGE is an

immunoglobulin protein present in large quantities during inflammation and damage to epithelial cells [147]. Interaction of RAGE with its ligands which include S100 proteins has been implicated in Alzheimer's disease [148], diabetes [149], inflammation, colon [150], pancreatic [151], lung [152] and breast cancers [153].

Many S100 proteins have been isolated and their 3D holo structure [154-159] and/or apo form resolved experimentally [158, 160-164]. In addition, the three-dimensional structures of some family members bound to their target proteins have also been elucidated [165-168]. S100 proteins are found only in vertebrates and function both intracellularly and extracellularly [169]. Expression of S100 proteins is tissue specific [170] and they can be present in both normal and diseased cells [171]. It is their elevated presence in diseased cells that has been the focal point of intense research in recent years [118, 172-177].

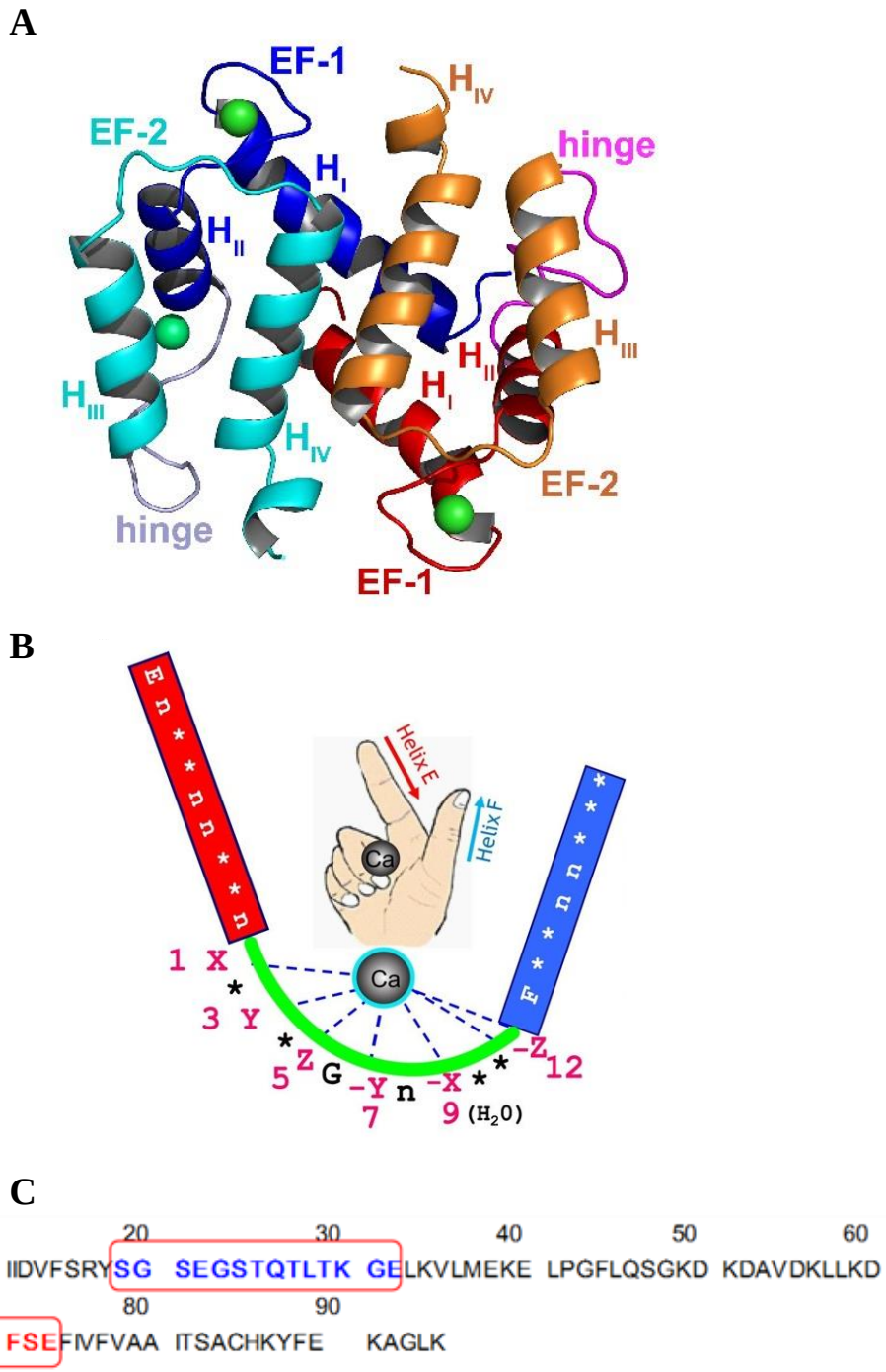



**Figure 1.6**: Calcium-binding S100 protein. A) Cartoon model showing the helix-loop-helix domains with bound $Ca^{2+}$ (image taken from http://goo.gl/aepDse. Last accessed 11th October, 2015). B) Illustration of $Ca^{2+}$ at the EF-hand using the thumb and index fingers (image taken from Zhou *et al*. [178] with permission from Elsevier Limited, 2015). C) Calcium-binding affinity of S100P (highlighted in blue and red), a member of the S100 family of proteins.

### 1.2.1 S100P and pancreatic cancer

S100P (Figure 1.7) is a 95 amino acids member of the S100 protein family first identified and isolated from human placenta in 1992 [179, 180]. Like other members of the S100 family of proteins, S100P consists of two EF-hands linked together by a flexible linker [159]. Studies have shown that the protein exists as a homodimer, with the monomers held together at the dimeric interface by non-covalent interactions [141]. The two EF-hands (EF-hand 1 and 2) lie in positions 12–47 and 49–84 respectively. The loop (position 62–73) linking the canonical EF-hand 2 on the C-terminus has a much higher affinity for binding calcium ions than loop 1 (position 19–32) linking the "pseudo" EF-hand 1 on the N-terminus [139, 141].

S100P has been reported to be present in both normal and malignant tissues [171] and it is its expression in the latter which is the subject of much research as it is overexpressed in many cancers [121, 172, 181, 182]. Its role in normal tissues is not clear, but a study by Sato and Hitomi [183] showed that it is present during the differentiation of human oesophageal epithelial cells implying that it may play a role in this process. In tumour cells, over-expression is believed to lead to tumour malignancy and proliferation [170, 177], increased resistance to therapy [181], decreased survival rates in cancer patients [184], immortalisation of breast cancer cells [174] and decreased sensitivity to treatment with chemotherapeutic agents [185, 186]. These actions are believed to be mediated through the protein's interaction with its receptor RAGE, the receptor for advanced glycation end-products [150, 151, 187, 188].

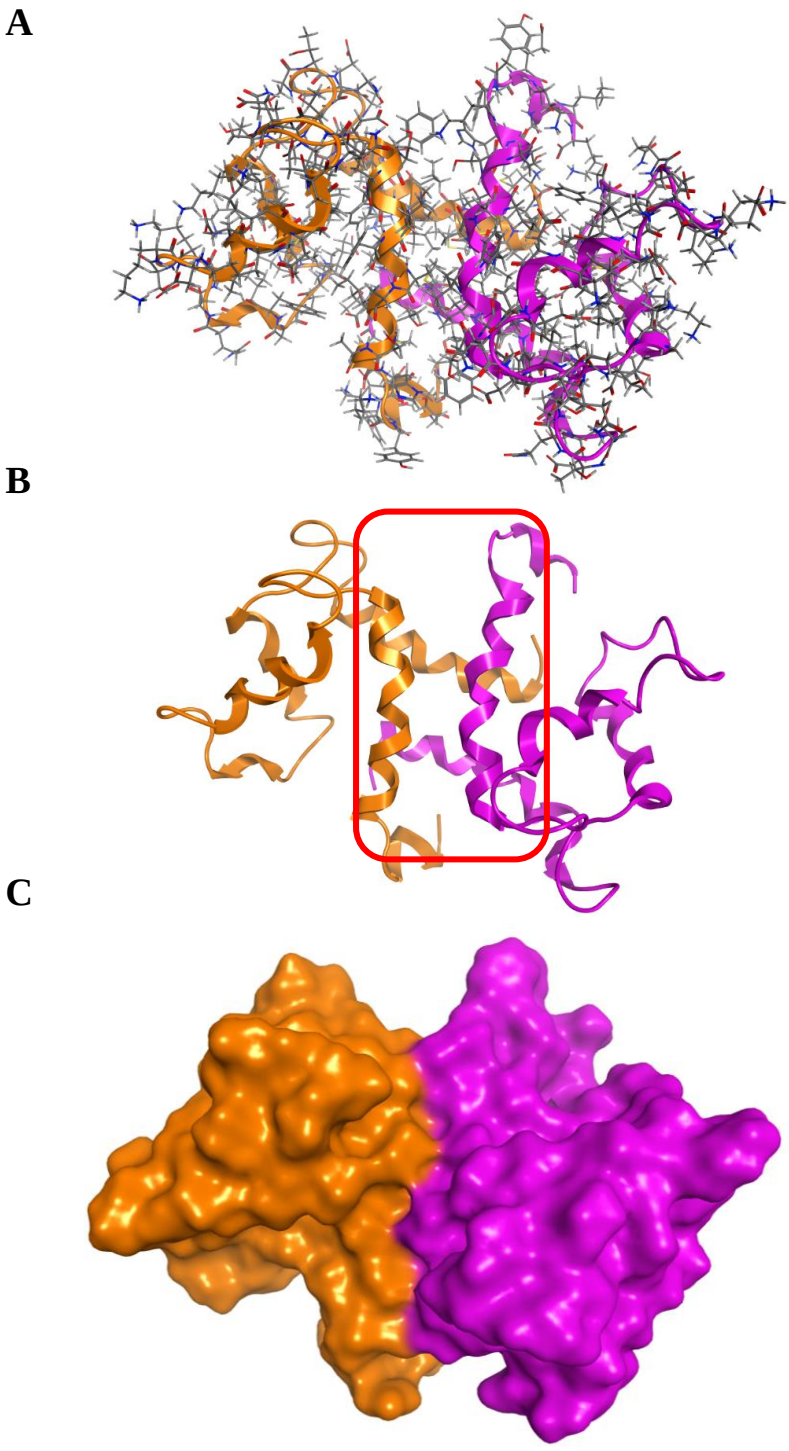


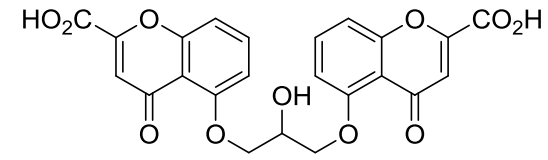

**Figure 1.7**: S100P protein homodimer. A) Ball and stick and ribbon depiction of the protein. B) Ribbon model of the protein showing the dimeric interface (enclosed in red). C) Molecular surface. (Images generated in MOE [189]).

There has been considerable evidence implicating S100P over-expression in pancreatic cancer [118, 151] and it has been proposed as a novel marker in identifying this cancer [121, 125, 126, 190-193]. A study of the salivary transcriptome for early markers of pancreatic cancer identified S100P as one of the over-expressed proteins in patients with pancreatic cancer [194]. This is promising as it will provide an easy and non-invasive procedure for early detection of the cancer in the clinic. The protein has also been shown to be highly sensitive and specific in discriminating between chronic pancreatitis and pancreatic cancer [121, 194].

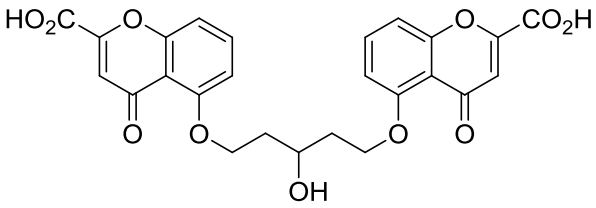
### 1.3 S100P as a therapeutic target and project aims

Although much has been reported on S100P and its expression in many tumours, there is little research published targeting the protein as a possible route of therapy for pancreatic cancer at the start of this project in 2011. In 2006 a study by Arumugam *et al*. [195] showed that S100P interacts with cromolyn (Figure 1.8), an anti-allergy drug that is used in the treatment of asthma [196], and that this interaction inhibits cell proliferation and tumour growth by blocking the formation of the S100P-RAGE adduct in pancreatic cancer cells. They showed that cromolyn's inhibitory effect was more significant in combination with gemcitabine than on its own [195]. An analogue of cromolyn, C5OH (Figure 1.8) has been reported to be more potent than cromolyn in inhibiting S100P-RAGE interaction both *in vitro* and *in vivo* [197]. However, cromolyn is a very water-soluble compound and is poorly absorbed by the gastrointestinal (GI) tract with less than 1% oral bioavailability [198, 199]. This renders it highly unsuitable as a drug candidate for cancer therapy. Consequently, there is a gap in the market to exploit this area in order to combat pancreatic cancer.

This project seeks to identify potential novel chemotherapeutic agents that will act as inhibitors of S100P-RAGE adduct formation. This will be achieved using computational and synthetic chemistry techniques to rationally design novel inhibitors based on an understanding of cromolyn-S100P interaction. Identified compounds will be subsequently screened for activity against pancreatic cancer cell lines. This project will use a medicinal chemistry approach incorporating computer-aided drug design (CADD) techniques (Chapter 2), synthetic chemistry (Chapters 3 and 4), and biological screening (Chapters 3 and 4) techniques – these form the early stages of a drug discovery process (Figure 1.9) – to identify the potential lead candidates.




**Cromolyn (4)**




**C5OH (5)**

**Figure 1.8**: Cromolyn and its analogue C5OH.

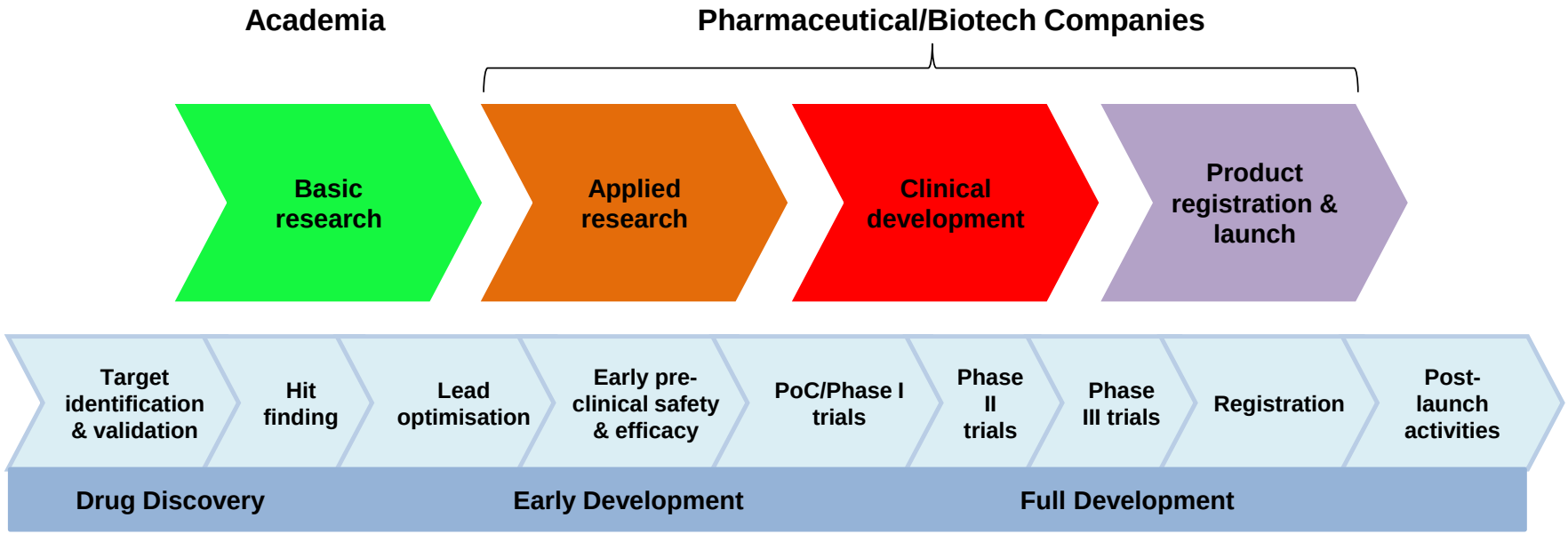



**Figure 1.9**: Drug discovery stages. Academia is mainly involved in (but not limited to) basic research involving target identification and validation, hit finding and lead optimisation. (Image adapted from Bevan [200]).

## 1.4 Overview of drug-target/receptor interaction

A drug can be viewed of as a compound that interacts with a biological system to elicit a physiological response [201]. Drugs are thus a diverse group of compounds and can vary broadly from caffeine in a cup of coffee to keep the mind alert, morphine to numb pain and snake venom that has the potential to kill [201]. In order for a drug to exert its effect, toxic or otherwise, it has to interact with a target or receptor. A biological target is an entity within a living system whose activity is modified by a drug or external stimulus [202]. This interaction between drug and target occurs via complementarity between the two and is held together by weak forces such as hydrogen bonding and hydrophobic interactions [201].

Medically, drugs are developed to fill a therapeutic void e.g. where there is no known cure for a disease, or to improve the efficacy of existing drugs [203]. In order to do so, a target is first identified and validated. A drug target could be a protein, receptor, an enzyme, ion channel, nuclear hormone receptor, membrane transport protein or nucleic acid [202]. One or more of these targets can be implicated in disease states, thus it is not unusual to have different drugs designed for different targets in the same disease. For instance, in cancer therapy, alkylating agents such as cisplatin target DNA [204] while topoisomerase inhibitors like Topotecan (Hycamtin, GSK) are designed to target the topoisomerase enzymes that separate DNA strands [205].

Target validation – defining the target's role or involvement in disease [206], is part of the first stages in a drug discovery process after the target's discovery (Figure 1.9). Many approaches, *in vitro*, *in vivo*, and *in silico*, are employed to

validate a target's role in disease. In sense reversal, an *in vitro* technique, the target gene's expression is reduced and the corresponding physiological response and effect on disease is studied [207]. Examples of such an approach include antisense technology, gene knockouts, use of ribozymes and RNA interference (RNAi) [208-211]. In contrast, proteomic validation techniques which include the use of antibodies [212], aptamers [213], peptides [214], and microarrays, can be used to modify a target protein's activity without affecting its expression [206, 215].

*In vivo* target validation techniques include the use of animal models e.g. mice and zebrafish to study the effect of gene silencing in disease states [216-218] whereas with *in silico* methods, computers are used to predict potential drug targets, e.g. from the human genome [219], or to predict interactions between a potential drug molecule and a model of a target macromolecule via computer simulation [206]. This simulation is carried out with or without prior knowledge of the structure of the biological target [220]. Examples of *in silico* drug validation techniques include *in silico* gene expression analysis [221], and homology modelling which involves building a 3D model of a target protein using information from known, homologous proteins as templates [222, 223].

Once a target is validated, it then becomes druggable. A druggable target is one that has the potential to interact with drug-like chemical compounds to produce a pharmacological effect [224]. S100P has been validated and identified as a druggable target [125, 150, 175, 177, 181, 185, 195, 225-228]. It has thus passed the first stages required of a drug discovery project i.e. target identification and validation. This project is mainly involved in the next stage of drug discovery i.e. "hit" finding, and

will achieve this by using *in silico* software to first identify putative pockets on the protein, then predict the interaction of cromolyn with these pockets via molecular simulations and docking. The predicted interactions will be used to design pharmacophore constraints as filters to identify "hits" from virtual libraries. All these steps are part of what is known as computer-aided drug design (CADD), and will be achieved using the Molecular Operating Environment (MOE) drug discovery software suite [189].

## 1.5 Computer-Aided Drug Design (CADD)

The use of computing power to solve chemical problems – computational chemistry, or to study biological systems – computational biology, or to organise and analyse biological and genomic data – bioinformatics, demonstrates the important role that computers play in modern-day drug discovery and development.

Computer-aided drug design (CADD) involves the use of computer power to expedite the time and resources spent at the discovery stage by identifying compounds with favourable properties and potential to be successful drug candidates [229]. To facilitate this, specialised software and algorithms have been developed with the ability to carry out calculations to identify molecular targets, predict binding interactions between target(s) and small molecules, and estimate binding affinity [230]. Structural databases such as the Protein Data Bank (PDB) containing 3D experimental data of macromolecules [231], the ZINC database holding millions of compounds for virtual screening [232], and the Cambridge Crystallography Data centre which contains crystal structure data of small molecules [233], and private in-house databases provide useful sources for starting points in most CADD projects.

The Swiss Institute of Bioinformatics (SIB) Click2Drug website (www.click2drug.org) has detailed listings of software, databases and web servers used in CADD.

The role of CADD following target identification and validation is dependent on two main factors: the availability of 3D experimental structure of the target, and/or known drugs that are active on the target. If there are no experimentally determined 3D structures of the target, a ligand-based or pharmacophore search of known compounds is adopted ([234], Figure 1.10). If the experimental 3D structure of the target macromolecule is available, a structure-based drug design (SBDD) approach is often taken to identify lead compounds [234].

In ligand-based drug design studies, structure/activity data from compounds that bind to a target with a known affinity is used to build 3D quantitative structure activity relationship (QSAR) pharmacophore models to search databases for structurally related compounds on the premise that compounds with similar structures/pharmacophores will share similar properties [235-238]. A pharmacophore is an abstract 3D representation of the geometric, steric and electronic properties of a compound that are necessary for target recognition and binding [239].

Receptor- or structure-based drug design (SBDD), which includes docking and *de novo* design, is employed where the 3D structure of a therapeutic target is known or available. It involves docking of a ligand(s) into the binding site of the target and predicting the quality of fit between ligand and target which is then scored and ranked according to the best fit [240-242]. A SBDD approach is used in this thesis as there is 3D experimental data available on S100P in the PDB database.

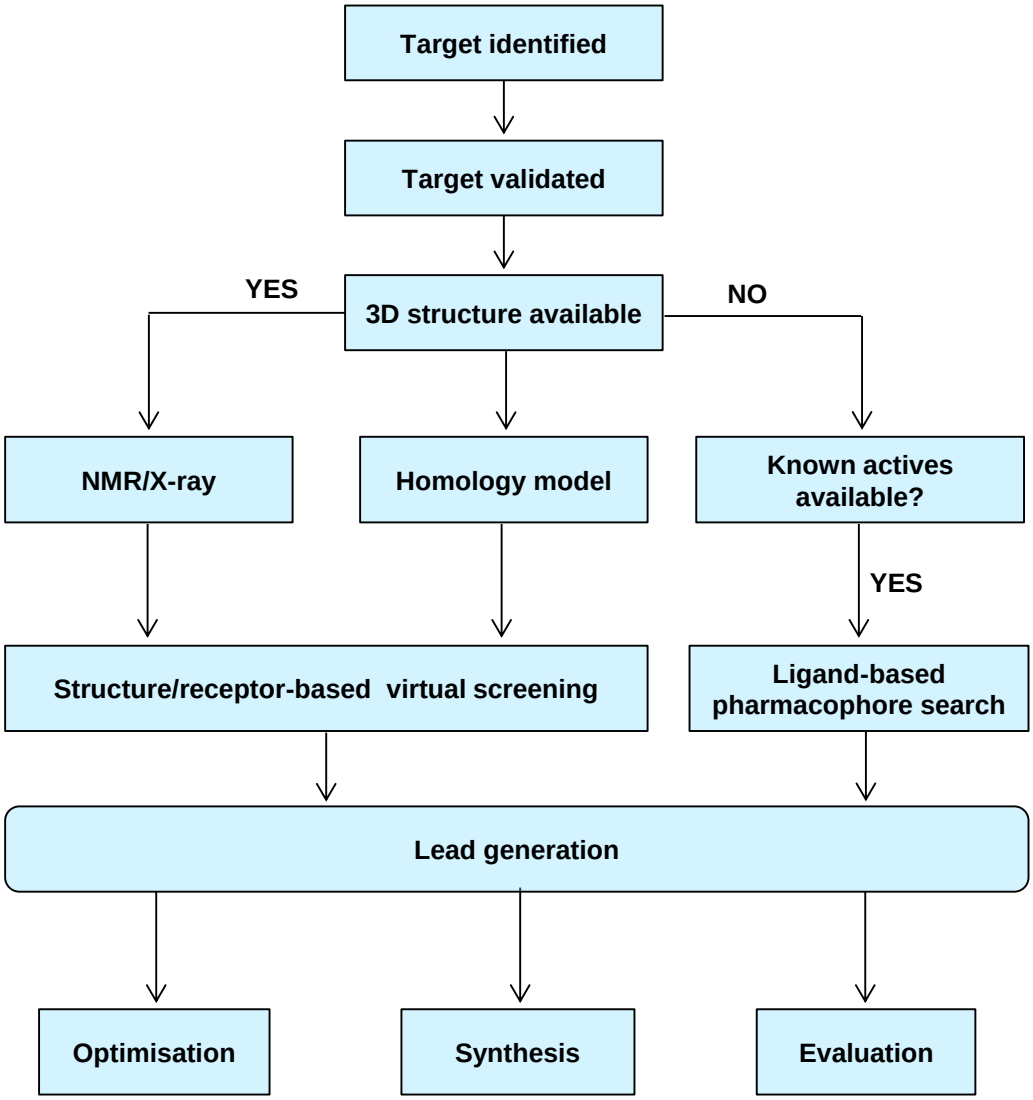



**Figure 1.10**: Simplified pathways of computer-aided drug design. (Image adapted from Veselovsky and Ivanov [234]).

### 1.5.1 Docking

Docking is used in SBDD for two main purposes: to predict the best pose of a ligand at the binding site of a target macromolecule and to estimate the tightness of fit between ligand and target [243, 244]. Docking plays a key role in SBDD where the information from the macromolecule–ligand complex could form the basis of whether a ligand is progressed through to the lead-optimisation stage or discarded. The strength of any docking algorithm lies in its ability to predict the binding affinity between the macromolecule and ligand complex [245]. This ability, known computationally as the scoring function, is a set of computed mathematical constraints designed to predict the experimental free energy of binding ($\Delta G$) between the ligand and macromolecule after docking [246]. Whereas docking will try and mimic as closely as possible the native conformation of the protein–ligand complex in the experimental structure, it is the scoring function that decides the quality of this fit [247].

For many docking algorithms, there is a high success rate in reproducing the conformational pose of the X-ray crystal structure but predicting the binding affinity is not usually as successful [248, 249]. Most scoring functions are designed based on experimental data and the accuracy of their results reflect that of the observed experimental data [250]. Any errors contained in the experimental data could therefore influence the results of the scoring function. Despite this limitation, docking algorithms, in most cases, are still capable of discriminating between potential actives and non-binding compounds [251].

Several docking algorithms are available, either free of charge or commercially via drug discovery software suites. The most commonly used docking programs, according to the number of citations in published articles between 1990-2013 [230], are AutoDock [252], GOLD [253], and Glide [254]. Others include FlexX [255], Surflex-Dock [256], Fitted, AutoDock Vina [257], and MOE [258].

### 1.5.2 Structure-based Drug Design (SBDD) on S100P

In this thesis, SBDD techniques were employed on the publically available 3D experimental structures of S100P to identify lead candidates with potential to bind to the protein (Chapter 2). This was achieved using pocket-detection algorithms, free and commercial, to identify putative binding pockets on the protein. The Molecular Operating Environment (MOE) drug discovery software suite [189] was used for docking, pharmacophore design and virtual screening studies once putative sites have been identified. ChemAxon's JChem software package (www.chemaxon.com) was used to cluster identified "hits" from virtual screening studies.

#### 1.5.2.1 Algorithms used to identify putative binding pockets on S100P

Knowledge of cavities on a target macromolecule is essential before embarking on structure-based drug design studies. These are the sites at which the designed small molecules will interact with the target. Most of the 3D experimental structures of macromolecules available in the PDB database are resolved in complex with ligands, showing the binding pocket(s). However, for those macromolecules that are not resolved in complex with ligands, as is the case in S100P, it becomes imperative to predict these sites prior to carrying out any SBDD studies.

The four pocket-detection algorithms used in this work, Fpocket [259], Site Finder [260], Pocket-Finder and Q-SiteFinder [261], to identify putative binding pockets on S100P use either geometry- (Fpocket, Site Finder, Pocket-Finder) or energy-based (Q-SiteFinder) approaches. Geometry-based methods use the shape of the protein to predict potential binding sites [262] while energy-based methods employ small chemical probes to study energetically favourable interactions with the protein surface [263].

Fpocket's search method uses alpha spheres and Voronoi tessellation to identify clefts and cavities on a protein surface [259]. An alpha sphere is one in contact with four other atoms on its boundary and contains no internal atom [259]. Voronoi tessellation involves the division of space into regions called Voronoi cells around a given set of centroid points [264]. In Fpocket, a cluster of alpha spheres is subjected to Voronoi tessellation, pruned to remove clusters with unfavourable characteristics such as polar groups, and then ranked according to their ability to bind to small molecules [259].

MOE's Site Finder uses the same alpha sphere method as Fpocket to detect potential binding pockets. Site Finder works by identifying regions of tight atomic packing and filtering out inaccessible regions and exposed hydrophilic spheres [260]. Spheres are then classified as hydrophobic or hydrophilic based on their hydrogen bonding character. Hydrophobic spheres with the most hydrophobic contacts with the receptor are ranked as the top predicted site [265].

Q-SiteFinder, an energy-based pocket-detection algorithm, identifies potential binding pockets by calculating the van der Waals interaction between a

methyl probe and the protein [266]. These probes are then clustered and ranked according to their probe energy [261]. Pocket-finder, a geometry-based algorithm developed by Laurie *et al*. [266] to compare against Q-SiteFinder, implements the POCKET algorithm [267] when searching for binding pockets. Potential sites are identified when a probe of radius 3 Å passes through a three-dimensional grid and undergoes a protein–site–protein (PSP) event, this occurs when there is an interaction between the probe sphere and an atom followed by an alternate period of no interaction then interaction [261]. Identified pockets are ranked by the number of probes in the sites instead of probe energy used in Q-SiteFinder.

#### 1.5.2.2 **Virtual screening of lead-like databases**

Virtual screening plays a crucial role in the rational approach to drug design and discovery by speeding up part of the lead identification process, where, with appropriate computing power, millions of compounds within virtual libraries can be screened for novel bioactive hits within hours [268]. Virtual screening does not replace the long established high-throughput screening (HTS) used in lead identification in the pharmaceutical industry; it is seen as a complementary tool to help streamline the process of lead identification [269-271]. This means that thousands of compounds are filtered out via virtual screening with those that have been predicted to be active carried through to HTS thus eliminating/decreasing cost and waste.

HTS is a highly automated process that uses robotics and sophisticated data analysis methods to screen a huge collection of molecules for biological actives in assays [272, 273]. The cost involved in putting together compound libraries and

equipment for HTS, which analyses millions of individual compounds in a biochemical assay, can be prohibitive [274]. Its routine application is therefore limited to the pharmaceutical industry and very few in academia who can afford the financial resources and human expertise to run it.

Virtual libraries for virtual screening could either be accessed from in-house databases [275, 276], commercial vendors, or from ZINC [232], a free database of commercially-available compounds for virtual screening. These databases contain compounds that are known as leads or lead-like compounds, a term that is used to differentiate them from drugs. Lead-like compounds have properties that make them amenable to optimisation before they could become drugs [277]. Some leads are marketed drugs whose structure have been modified to yield novel drugs whilst others, "pure leads", have no prior therapeutic value [277]. Leads also fulfil Lipinski's rule of five, which states that for a drug to be orally active it should have a molecular weight of less than 500, a lipophilicity (logP) value below 5, less than five hydrogen bond donors and less than 10 hydrogen bond acceptors [278]. Unlike drugs, leads tend to exhibit weak affinities in the high micromolar to millimolar range to their target [279].

The MOE software that was used in this study comes with lead-like databases that contain commercially available leads from specialist vendors. This database, and the ZINC lead-like database, were the virtual libraries screened for "hits" in this study.

#### 1.5.2.3 **Compound clustering**

Given a group of heterogeneous compounds, clustering is an analysis technique that finds similarity between individual compounds in order to put them together in a cluster [280]. The premise is that compounds in a given cluster will have more in common with each other than with others in a different cluster.

Clustering was applied in this study following virtual screening studies in order to identify a diverse collection of compounds that will be selected for validation studies. ChemAxon's Library Maximum Common Substructure (LibMCS) application (Version 5.8.2, 2013, ChemAxon, www.chemaxon.com), which uses maximum common substructure subgraph [281, 282] to identify common fragments in compounds in a hierarchical manner, was employed. Compounds sharing a user-defined maximum common substructure are hierarchically grouped together until there are no more shared fragments (Figure 1.11). Singletons are outlier compounds that do not share common subgraphs with any of the clusters.

The above SBDD techniques will form the basis for the next Chapter which presents a detailed methodology and results on computational studies on S100P.

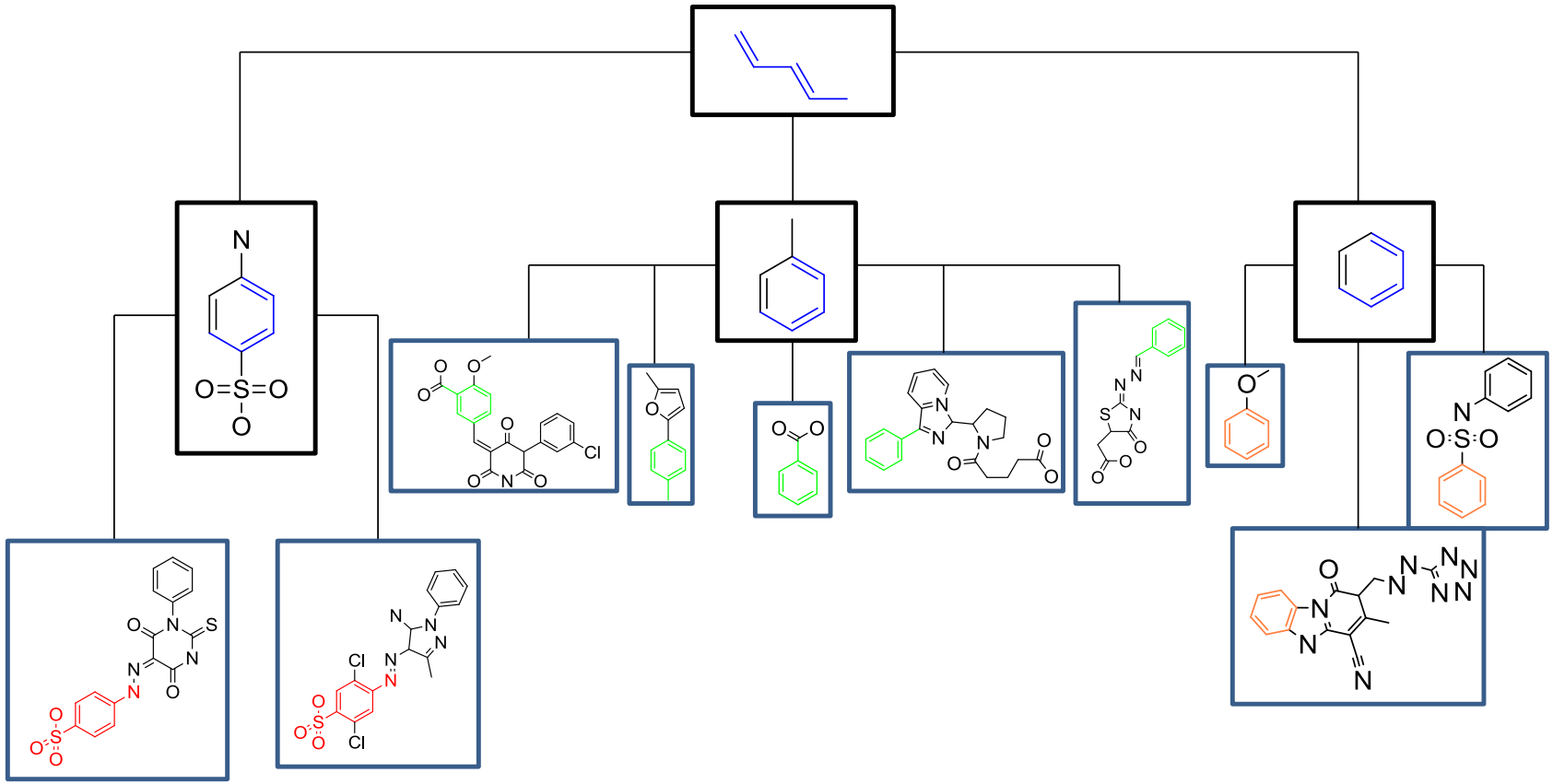


**Figure 1.11**: Substructure hierarchy system used by ChemAxon's Library MCS for clustering compounds. Initial structures are shown at the bottom of the hierarchy, the next level contains maximum common substructure of initial structures, and above that on the first level is the fragment common to all compounds. (Image drawn by Author, 2015).

# CHAPTER 2: *In Silico* Design of Potential S100P Inhibitors

# 2. PREFACE

Despite published data on the S100P's interaction with cromolyn [195], metal ions [141] and small peptides [138] at the outset of this project, there was limited published information regarding the mechanism by which these molecules might interact with the protein. The majority of the published work used biochemical methods such as column affinity chromatography [195, 283, 284], standard assays [172, 177, 181, 227] and microarray techniques [285-287] to ascertain the existence of affinity, rather than to probe mechanisms that gave rise to it.

As such, the following chapter details the computational tools used to identify and investigate potential binding sites on S100P and cromolyn, a ligand shown to bind to the protein [195]. The resulting binding interactions between the protein and ligand formed the basis for pharmacophore models used to screen lead-like databases, to identify novel "hit" molecules with potential to bind to S100P and inhibit the S100P RAGE interaction.

## 2.1 INTRODUCTION

Since the discovery of S100P in 1992 [179, 180], there is growing evidence of the crucial role it plays in many cancers. In breast cancer, S100P has been reported to immortalise human epithelial cells *in vitro* [174], make the cancer cells more proliferative [288], as well as confer poor survival rates to patients [289]. In prostate cancer, S100P over-expression is shown to positively correlate with tumour growth, metastasis and protection of host cells against camptothecin-induced apoptosis [181]. In lung cancer, increased expression of the protein was found to induce metastasis in non-small cell lung cancer patients, leading to increased angiogenesis [290]. A decrease in survival rate in lung cancer patients was also observed to correlate with increased S100P expression [184]. In colorectal cancer, S100P is reported to induce cancer cell growth, proliferation and metastasis via its binding to RAGE – the receptor for advanced glycation end-products [150]. In pancreatic cancer, it has been shown to be highly present in the early stages of the disease [291], and promotes growth, metastasis and invasion of cancer cells via its interaction with RAGE [151, 177]. Of all these cancers, pancreatic cancer has the worst prognosis with no specific early markers to identify the disease in its early stage. It is this urgency to find specific markers for the disease that S100P has been proposed as a potential clinical marker for the cancer [121, 125] which makes it an ideal therapeutic target against the cancer [191, 192, 287].

It is not known what causes the over-expression of S100P in pancreatic cancer [192] but hypomethylation of the gene has been found in many pancreatic cancer cell lines as well as in primary tumours [287]. Whilst the protein has been associated with

many targets, including ezrin [283, 292] and S100P binding protein [118, 293], it is its interaction with RAGE that is believed to be instrumental in promoting cancer cell growth, migration and invasion [150, 151, 187, 225]. Since interaction with RAGE occurs via the dimeric protein [294], any attempts at designing potential inhibitors against this interaction must therefore be carried out on the protein dimer, particularly on the dimeric interface. Dimer interfaces in proteins have been targeted in drug discovery studies to prevent dimerization [295, 296] or interaction with target [297].

## 2.2 METHODS

### 2.2.1 Identification, manipulation and analysis of experimental (3D) S100P structures

A search of the RCSB Protein Data Bank (PDB) website ([231], www.rsb.org, accessed in October 2011) using the keyword "S100P" generated results that contained information on two structures of the protein deposited in the database: an X-ray crystal structure (PDB ID: 1J55, [159]) and an NMR ensemble (PDB ID: 1OZO; [298]).

The X-ray crystal structure of S100P was extracted in its monomeric form (Figure 2.1A) and saved as a PDB file. The biologically functional homodimer was created using the UCSF Chimera package [299]. In the program, the protein was downloaded from the Protein Data Bank using its PDB Accession Code 1J55. Under Favourites, Model Panel, Biological Unit was selected for the program to display the biological oligomer using the REMARK 350 BIOMT matrices in the PDB file header. The generated dimer was saved as a PDB file.

Conversely, the NMR ensemble was extracted from the PDB website as one huge PDB file containing 16 models (Figure 2.1B). Individual models were separated by copying the header and the 95 amino acid (aa) sequences of both chains A and B for each model onto a WordPad document that were then saved as a Rich Text File using the file extension .pdb.

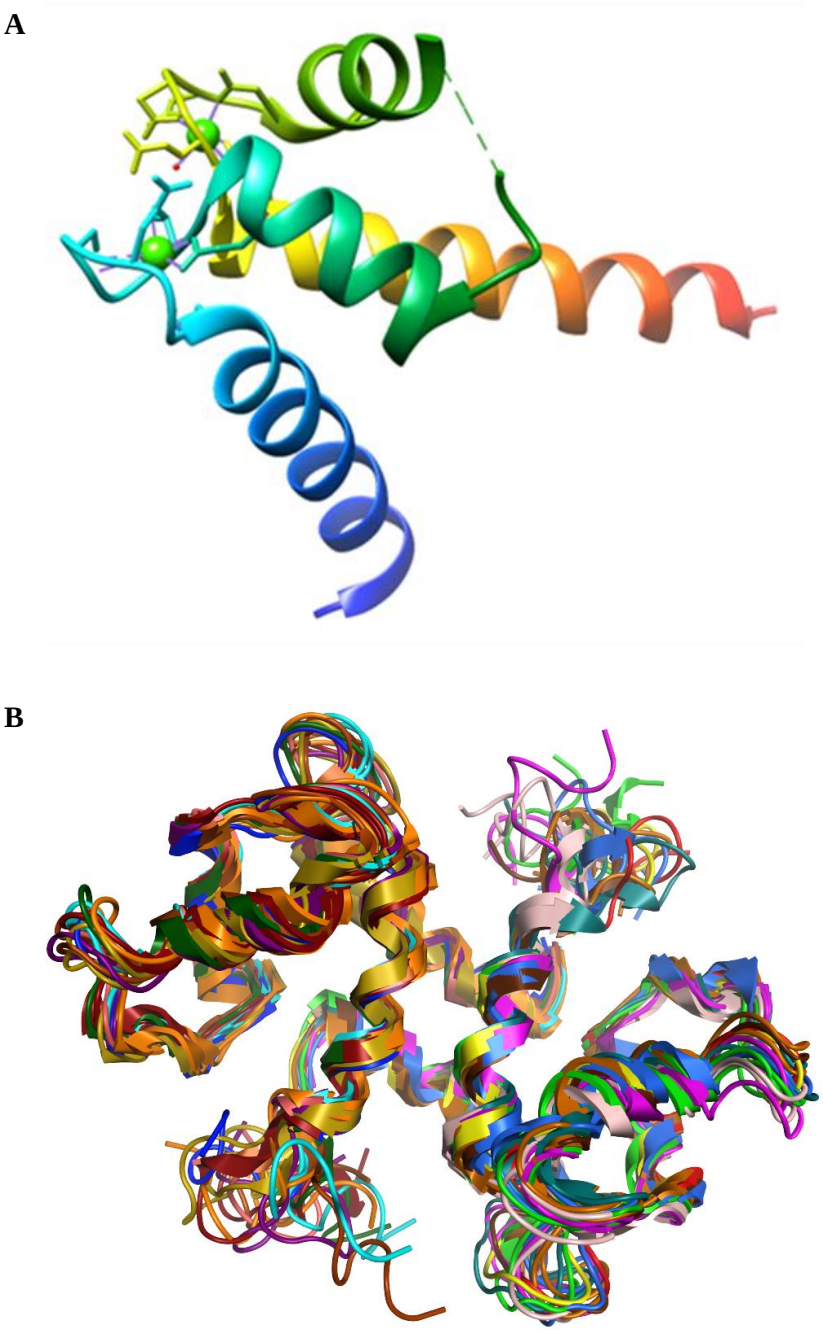



**Figure 2.1:** 3D experimental structures of S100P. A) X-ray crystal structure of S100P (PDB ID: 1J55; [159]). Calcium ions are shown as green spheres and missing residues 45-51 are depicted by a broken line. B) NMR ensemble of S100P (PDB ID 1OZO; [298]). The sixteen conformers of the ensemble are overlaid in different colours. (Images generated in MOE).

### 2.2.2 Identification of putative binding pockets at the dimeric interface of S100P

Four different pocket-detecting algorithms – Fpocket (http://bioserv.rpbs.univ-paris-diderot.fr/cgi-bin/fpocket [259]), Pocket-Finder (http://www.modelling.leeds.ac.uk/pocketfinder/ [261]), Q-SiteFinder (http://www.modelling.leeds.ac.uk/qsitefinder/ [261]), and Site Finder [260] – were employed to identify potential binding sites on the dimeric interface of the S100P protein. Where the algorithm was used online, individual S100P PDB files were uploaded to the respective servers and once potential pockets were identified, the resulting PDB files with cavity information were saved.

With the exception of Site Finder, an application within the Molecular Operating Environment (MOE2010.09) drug discovery software suite, all pocket-detection algorithms were available to use online for free (all were accessed in October, 2011). When Site Finder was used, the protein was prepared using the Protonation 3D wizard, which assigned ionisation states to the macromolecule, added protons where necessary and checked for tautomers and rotamers [300]. Default settings for the calculation of binding pockets were retained as follows: probe radius of 1.4 Å for a hypothetical hydrophilic hydrogen bonding atom (such as N or O), probe radius of 1.8 Å for a hypothetical hydrophobic atom (such as C), rejection of a hydrophilic sphere that has no hydrophobic alpha sphere within 3 Å. Once prepared, potential pockets identified by Site Finder on the protein were rendered with dummy atoms for subsequent docking studies.

### 2.2.3 Docking

Docking studies were carried out using the Dock application within MOE2009.10 to investigate potential interactions between cromolyn and S100P. Prior to docking, the ligand was built using the Builder feature in MOE and energetically minimised using the default Merck Molecular Force Field 94 (MMFF94x) force field settings to optimise its 3D coordinates. Both ligand and protein were prepared for docking using the Protonate 3D tool. The binding site(s) to dock the ligand into were restricted to those identified at the dimeric interface of the protein by the pocket-detecting algorithms previously described.

With the exception of sites identified by Site Finder, the dimeric binding cavities were recreated during docking by selecting the amino acid residues identified as belonging to the site by the cavity detection algorithm and restricting docking to a cut-off radius of 4.5 Å from these selected residues. The dimeric interfaces for Site Finder were rendered in MOE with dummy atoms using an identical distance restraint.

Once both protein and ligand were prepared and the binding site defined, the ligand was docked into the binding site using the default Triangle Matcher docking placement method. The resulting poses were ranked with the London scoring function. Fifty poses were retained for energy minimisation in the pocket before they were rescored using the Generalized-Born Volume Integral/Weighted Surface area (GBVI/WSA) scoring function [301]. Output poses are ranked according to the GBVI/WSA binding free energy (*S* score).

### 2.2.4 Pharmacophore design

The Pharmacophore Query Editor in MOE2009.10 was used to build pharmacophore models with a set of constraints on the location and type of ligand annotations using the default Unified scheme. Using appropriate protein-ligand complexes from docking studies as templates, points of interactions between the ligand and S100P dimeric binding site were selected to create two three-point 3D pharmacophore queries. Both queries are the same except in the descriptors used to define them. In the first "stringent" query, the two hydrogen bond acceptors (HBAs) were defined as anionic, whilst in the second "relaxed" query, the same acceptors were defined as hydrogen bond acceptors. All query distances and angles remained the same.

### 2.2.5 Virtual screening and clustering

Two databases were screened using the pharmacophore queries described in section 2.2.4. The MOE Conformation Database within the MOE software package which contains 3D conformations for approximately 650,000 lead-like compounds, and lead-like compounds from the ZINC database [232] which had more than three million compounds at the time of access (November, 2011).

Since the MOE database came ready-to-use, no further processing was applied to the compounds prior to virtually screening them. The database was screened using both pharmacophore queries. The same Dock settings were applied during the virtual screening. Additionally, the pharmacophores generated in section 2.2.4 were used as constraints to guide molecule selection.

Lead-like compounds from the ZINC database were downloaded in their SMILES (Simplified Molecular-Input Line-Entry System) format [302] before converting them into 3D molecules using the conformational import tool in MOE2011.10. During the conformational import, molecules were "washed and filtered" to deprotonate/protonate acids/bases; filter out reactive groups; omit compounds with a molecular weight >600, sum of acceptor and donor atoms exceeding 12, rotatable bonds of more than 7, chiral centres greater than 4, more than 8 rings, and d-hybrids. The molecules satisfying the inclusion criteria were minimised to their lowest energy conformations using the default molecular mechanics settings within the conformational import tool. Minimised compounds were saved in the MOE molecular database (MBD).

A similarity search using a Tanimoto Coefficient threshold of 0.7 was carried out between the MOE and ZINC databases to filter out the MOE compounds that had already been screened from the ZINC database. This was carried out using ChemAxon's Instant JChem application (ChemAxon JChem Software Suite, Version 5.8.2). The final ZINC compound database, with a total of 765,278 molecules, was screened using the same Dock settings outlined in section 2.2.3.

Compounds that match the pharmacophore restraints from the virtual screenings of both databases – "hits" – were clustered in ChemAxon's Library Maximum Common Substructure (MCS) application (Version 5.8.2) [303]. Library MCS clustered the "hits" by matching all atom types, bond types, charge, and whole rings, using the default minimum MCS size of 9.

### 2.2.6 Rescoring "hits" generated from virtually screening studies in "native" S100P

The three mutated amino acid residues in the NMR ensemble of S100P: A6, S85, and T92 were mutated back to their original amino acids T6, C85, and A92 respectively using the Protein Builder tool in MOE. "Hits" identified from virtual screening were rescored in the "native" protein to compare any differences or otherwise between the *S* scores obtained from the virtual screenings.

The Rescore protocol in the Dock panel of MOE2014.08 was used to rescore "hits" in the native protein. The binding sites were restricted to the sites used in the original virtual screening. "Hits" were rescored using the GBVI/WSA scoring function while the placement method was set to None.

## 2.3 RESULTS AND DISCUSSION

### 2.3.1 Identification, manipulation and analysis of experimental (3D) S100P structures

Selecting the most appropriate experimental 3D structure on which to carry out structure-based drug design (SBDD) studies is a crucial step in any drug discovery project. In the case of S100P, there were two entries deposited in the PDB database using two different methods, a high resolution X-ray crystal structure, (PDB Accession Code 1J55) and an NMR ensemble containing sixteen conformers (PDB Accession Code 1OZO). Both methods can generate high quality structures although it has been claimed that X-Ray crystallography, being the older and more established method of the two, produces structures of better quality [304]. For proteins that crystallise readily, this method provides an excellent choice for structure determination.

Alternatively, NMR solution spectroscopy is ideal for studying protein folding and dynamics [305, 306]. An advantage of this technique over X-ray crystallography is its ability to study proteins in solution, which gives an indication of protein dynamics. However, unlike X-ray crystallography, structure determination using NMR spectroscopy is limited to small proteins as large proteins can present overlapping peaks in the NMR spectra that are difficult to assign [307].

#### 2.3.1.1 **S100P X-ray structure**

The X-ray crystal structure of S100P (PBD ID 1J55) was resolved at 2 Å with an *R*-value and free *R*-value (*R*free) of 0.214 and 0.267 respectively [159]. The *R*-values give an indication of the quality of the model. The smaller this value, the better the fit of the model to experimental data [306]. However, because calculation of the *R*-value could be biased following refinement of the model to better fit experimental data, the free *R*-value was introduced [308]. The free *R*-value is similar to the *R*-value in definition except it is calculated on a small subset of experimental observations that was not used in the refinement of the model [309]. In an ideal model, the difference between the two values is <0.05 [308]. For 1J55, the difference between *R*- and free *R*- values is 0.053. This falls within the accepted range for a model to be considered an "accurate model" of the protein [310], thus making 1J55 an acceptable model to be used as a SBDD studies.

The X-Ray crystal structure of S100P 1J55 has two calcium ions bound to the EF-hands with a classical pentagonal bipyramid geometry [159]. There are four helices present: helix 1 (E3-Y18), helix 2 (L30-E40), helix 3 (A53-61), and helix 4 (F71-A92) (Figure 2.2). Since no electron density was observed for loop residues 46-51, and C-terminus residue 95, these are absent from the final model of the protein deposited in the PDB. The absence of these residues could impact on SBDD studies using this model, as any efforts to remodel the missing residues could introduce error into the system that could compound errors in subsequent drug design studies.

S100P 1J55 was deposited in the PDB in its monomeric form. However, the authors provided information in the accompanying PDB Header for *in silico* generation of the biologically relevant dimeric form of the protein. The S100P dimer is formed by hydrophobic interactions related by a crystallographic 2-fold axis and hydrogen bonding between hydrophilic residues [159]. Three hydrophobic contacts are present at the dimeric interface all involving the 1/1′ and 4/4′ helices. The first contact is provided by residues L4, A7, M10, I11, and V14 in helices 1 and 1′, with hydrogen bonding from E3 and D13 stabilising the contact; the second contact comes from residues F71, I75, A79 and H86 in the 4/4′ helices, with hydrogen bonding between S72 and S83 reinforcing contact. This second hydrophobic contact also extends into the hydrophobic core of the protein which is formed by residues F74, F77, V78 and I81. The third and final contact involves interactions between L4, M10, I12, and F15 of helices 1/4′, with V78, T82 and H86 of helices 4/1′ (Figure 2.3). These residues involved in these hydrophobic contacts on the S100P dimeric interface will be crucial in putative binding pockets identified on the protein for subsequent SBDD studies.

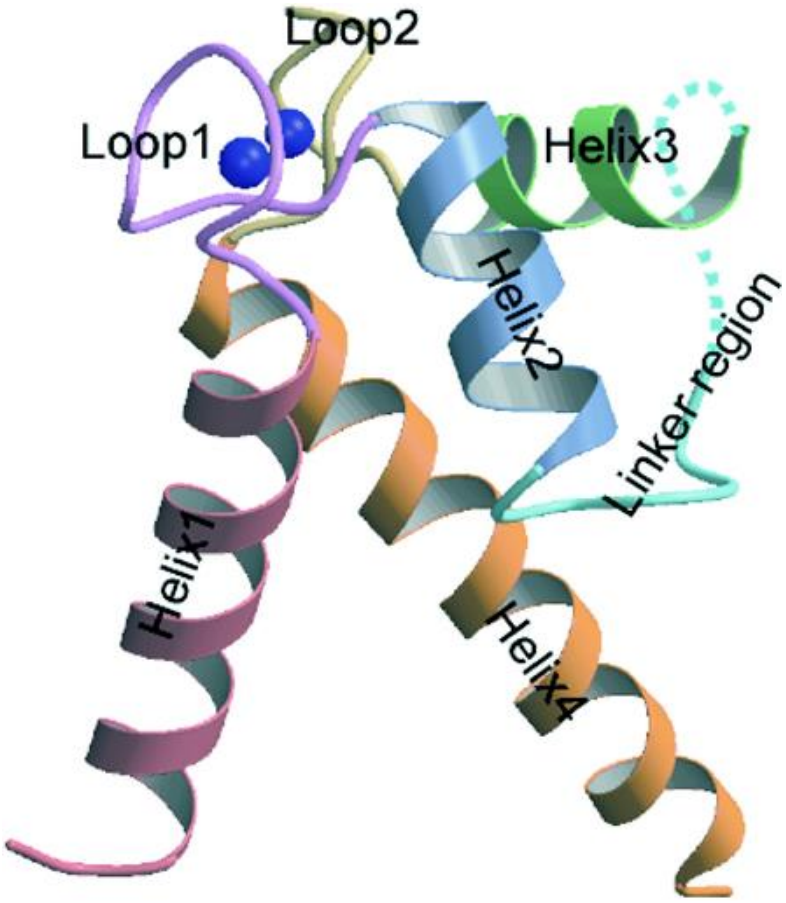



**Figure 2.2**: Monomeric X-ray crystal structure of S100P (1J55) showing helices 1-4 and the calcium ions (blue spheres). (Reprinted from Zhang *et al*. [159] with permission from Elsevier Limited, 2015).

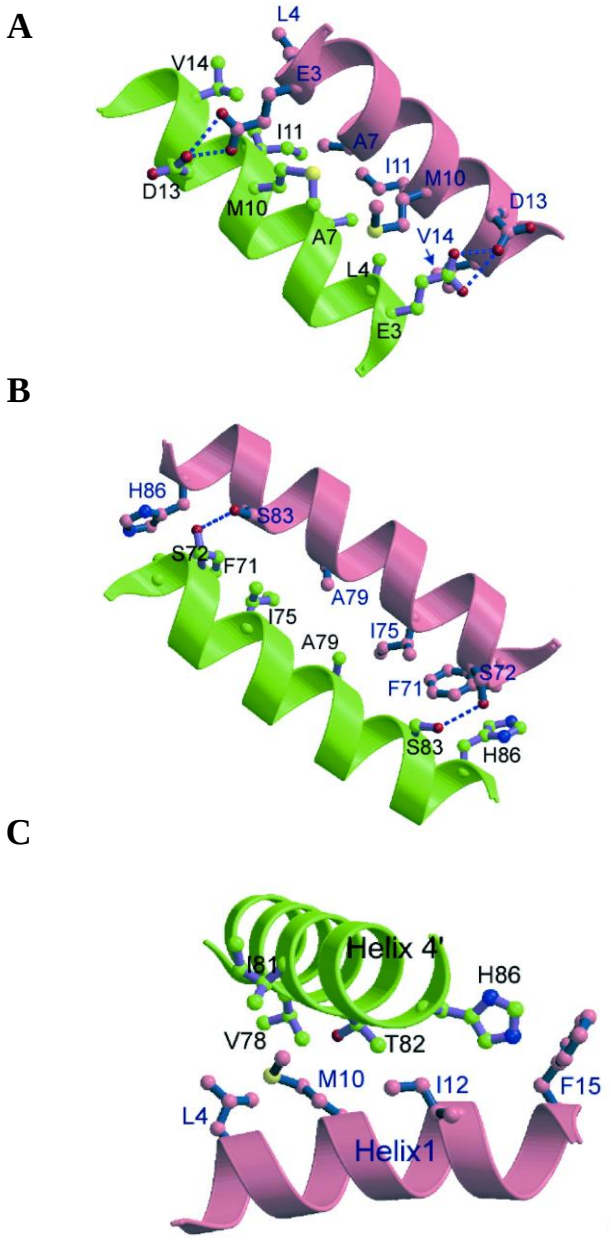



**Figure 2.3**: Hydrophobic contacts at the dimeric interface of two subunits of 1J55. A) The first contact between helices 1/1′. B) The second contact between helices 4/4′. C) Third contact between helices 1 and 4′. The residues forming the hydrophobic contacts are shown in ball and stick representation. Segments from each subunit of the protein are depicted in pink and green respectively. (Reprinted from Zhang *et al*. [159] with permission from Elsevier Limited, 2015).

#### 2.3.1.2 **S100P NMR ensemble**

Unlike the crystal structure, structures of the NMR ensemble of S100P (PDB ID 1OZO) were determined in the absence of $Ca^{2+}$ by triple resonance NMR spectroscopy [298]. Triple resonance NMR spectroscopy uses three types of atomic nuclei – $^{1}H$, $^{15}N$ and $^{13}C$ – to assign resonances to atoms forming the backbone of proteins especially those proteins enriched with $^{13}C$ and $^{15}N$ [311]. Instead of defining a single set of atomic coordinates, NMR spectroscopy uses restraints that cover a range of allowed distances and angles, to thoroughly explore all regions of conformational space that a protein could adopt [306, 312]. Hence an ensemble of viable structures, as opposed to a single definitive model, are generated via this method. A huge advantage of using NMR spectroscopy to determine protein structure is its ability to provide information on the average structure of the protein as well as its dynamics in solution [313]. In addition, the NMR model revealed the intrinsically disordered regions in S100P i.e. residues 46-51 [314] that were not resolved in the X-ray model of the protein.

In comparison to X-ray crystallography, quality assessment of NMR structures is more complicated as there is no associated *R* or free-*R* value by which to gauge the accuracy of the generated models. Instead, from the initially generated conformers, a sub-group with the least residual error values, the ensemble, is selected to represent the NMR structure of the protein [305]. Calculation of the pairwise root mean square deviation (RMSD) of the C-α atoms or heavy backbone atoms of the ensemble members gives an indication of the precision, and quality, of the NMR data [305, 307]. Lower RMSD values signify a better overall fit to the experimental data [306]. A mean RMSD

value of 1.0 Å for C-α atoms is comparable to a precision of approximately 2.0 Å for an X-ray crystal structure [307]. Precision is defined as the reproducibility of the experiment.

Fifty conformers were initially calculated for 1OZO and from these, 16 members with the lowest energy and least restraint violations were selected to form the final ensemble [298]. The structure deemed most representative of the ensemble was conformer 15. Mean RMSD values for this member were 1.35 Å and 0.89 Å for backbone atoms of residues of 3-90 and 3-86 respectively, and 2.01 Å and 1.39 Å for heavy atoms of 3-90 and 3-86 respectively [298]. These values indicate that models in 1OZO are suitable templates to use for SBDD studies.

A second way of assessing the quality of NMR structures is to look at the geometric quality of the models. Since the quality of NMR structures is dependent on the force fields and refinement procedures used during optimisation of the models, it is critical that the resulting models contain minimal errors with respect to their geometry [306]. Ramachandran plots provide valuable information on the stereochemical quality of a protein by mapping the distribution of backbone dihedral angles ψ and φ onto the four regions of the plot: favoured, additionally allowed, generously allowed and disallowed [306, 315]. In MOE, three different colours are used to represent the distribution of torsion angles of different residues in the four quadrant of the Ramachandran plot (Figure 2.4). Small green squares represent residues with sterically

hindered sidechain atoms that are clustered in the core region. Yellow squares represent residues in the allowed region and outliers are marked as red crosses.

Ramachandran assessment of each of the conformers in the1OZO ensemble show a number of residues that lie outside of the allowed regions of the Phi-Psi plot (Appendix). For conformer 15, the most representative model in the ensemble, most of the amino acids fall within the favoured regions. There are however 13 outliers (Figure 2.4, Table 2.1). Eight of the 13 residues that demonstrate unusual stereochemistry are located in the $Ca^{2+}$-binding region [141]. However, it is not known if the absence of $Ca^{2+}$ in the NMR-derived structure of S100P is responsible for the unusual conformation of these residues in the Ramachandran plot. These residues were found in the favoured regions in the Ramachandran plot of 1J55 (Figure 2.4B).

**Table 2.1**: Ramachandran data for conformer 15 of 1OZO showing the outliers and their φ/ψ angles. Residues involve in $Ca^{2+}$-binding are highlighted in blue. Residues absent in the X-ray crystal structure 1J55 are shown in red. (Data obtained from MOE).

| Model | Chain | Residue | Psi | Phi |
|---|---|---|---|---|
| 15 | A | S21 | -68.5 | 122.3 |
| | A | T25 | 127.2 | 96.5 |
| | A | Q26 | 86.2 | -56.5 |
| | A | K49 | -44.1 | -159.4 |
| | A | D50 | -72.3 | 19.2 |
| | A | A53 | -71.3 | 41.4 |
| | A | H86 | 9.9 | -54.0 |
| | B | T25 | 156.7 | 81.6 |
| | B | K49 | 101.1 | 146.3 |
| | B | K51 | 5.6 | 46.6 |
| | B | D62 | -34.7 | -31.1 |
| | B | D66 | -41.4 | -161.7 |
| | B | A67 | -39.0 | 169.3 |

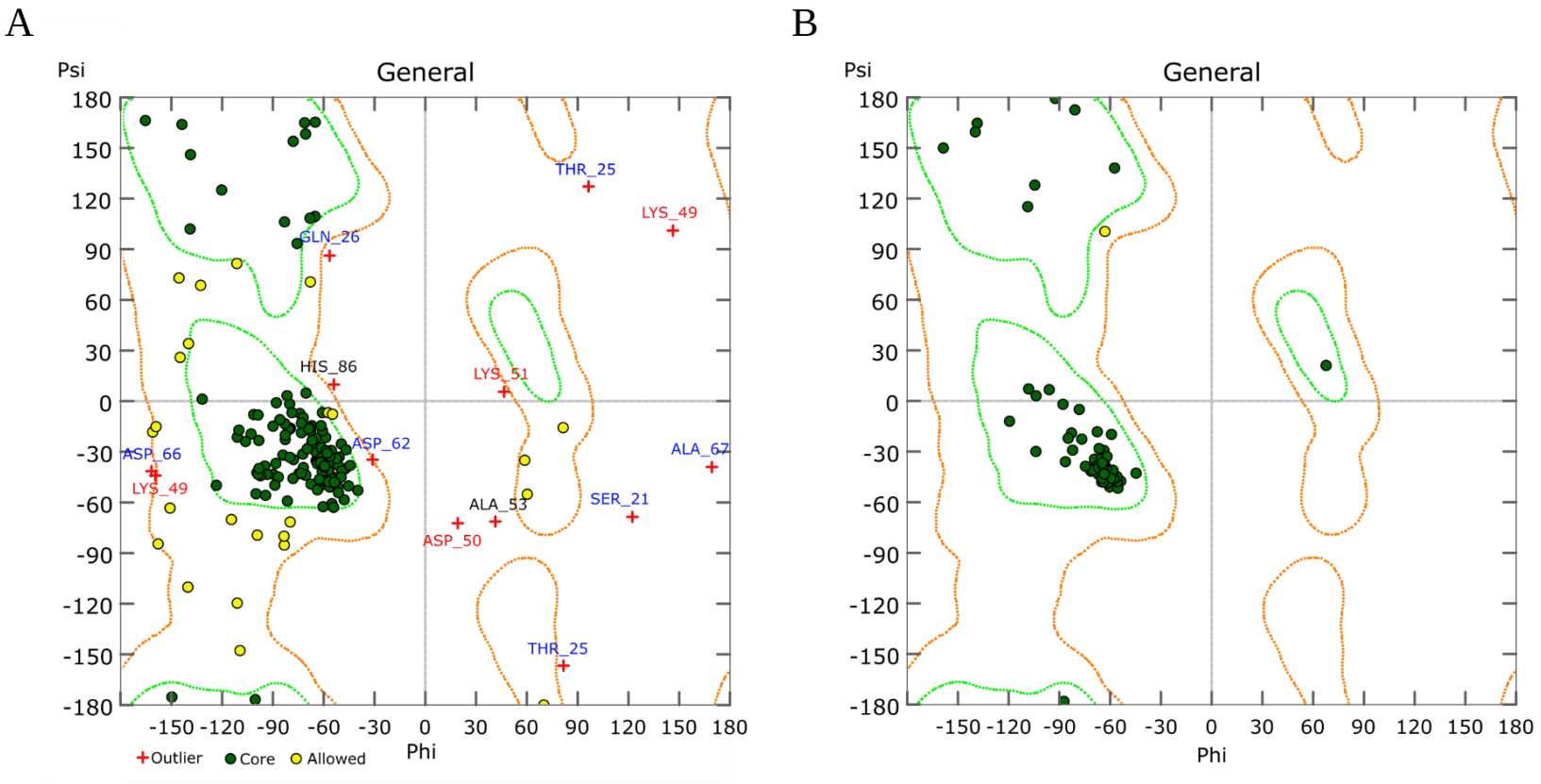



**Figure 2.4**: Ramachandran plot for S100P showing residues in allowed (orange contours) and core (green contours) regions. A) Conformer 15 of 1OZO. Outlier residues involved in $Ca^{2+}$-binding are highlighted in blue. Outlier residues absent in the X-ray crystal structure 1J55 are shown in red. B) Ramachandran plot for 1J55. (Plots generated in MOE).

The S100P structures in the NMR ensemble were resolved as dimers in contrast to the crystal structure. There were three mutations in the ensemble members: residue 6 (T→A), residue 85 (C→S) and residue 92 (A→T). It is not known why these mutations are present but the initial templates the authors used in their calculation of the NMR ensemble have similar mutations in the same positions [141, 161].

Despite these mutations, the geometric data and RSMD values for the NMR ensemble point to it being a good template for SBDD work. It is however important to note that the end-product of both X-ray diffraction studies and NMR spectroscopy is a model that is based on observed experimental data. This model could be a good or bad model depending on the accuracy of the experimental data on which it was determined from, as well as other factors during the modelling and refining of the data [313]. If there are uncertainties and/or errors present in the data or in the model, which is then used as a template for subsequent SBDD studies, this could lead to amplification of original errors resulting in inaccurate results. Given that independent X-Ray crystallographic- and NMR spectroscopy-derived models of S100P yielded similar models of the protein, gives confidence that both are accurate representations of the true nature of the protein, and that the models are good ones. However, since the protein was not fully represented in the crystal structure due to missing flexible regions, and since the model was monomeric, 1J55 was deemed undesirable as a template for SBDD studies. The NMR ensemble of S100P was therefore advanced for further *in silico* drug design studies. Nevertheless, information from the X-ray crystal structure was used as a guide in SBDD studies carried out using the NMR structures.

### 2.3.2 Identification of putative dimeric bindings sites on S100P

#### 2.3.2.1 Potential binding sites on 1OZO

With the exception of the calcium ion-binding regions observed when calcium ions were co-crystallised in the X-ray crystal structure of S100P (1J55), there is no direct experimental evidence for small-molecule binding sites on the S100P protein. Identifying putative binding pockets where small molecules can bind and hence disrupt the interaction between S100P and RAGE was therefore a critical step in the progression of the project. Understanding the nature of these potential binding cavities will play a central role in the rational design of a therapeutic agent targeting S100P.

The use of a number of independent geometry- and energy-based cavity detection algorithms to identify potential binding pockets on S100P was done in order to increase confidence in the results generated and to ensure that the final binding site(s) selected were determined by consensus, thus eliminating the potential for bias present in any single algorithm. This is additionally important given that the pocket-detecting programs used were not developed using NMR structures. This is largely due to the lower numbers of NMR structures, when compared to X-Ray crystal structures, available in the PDB. As of May 18$^{th}$ 2015, of the 101,060 protein structures in the PDB archive, 90,662 (~90%) were determined by X-ray crystallography, 9,597 (9.5%) by NMR and the rest by other methods such as electron microscopy, and small angle X-ray scattering. This does not detract from the use of NMR structures in drug discovery as high quality NMR structures, where available, have been shown to be good templates for drug design studies [304, 316].

Given that an NMR structure was used in this research, and to counteract criticism that the tools used to identify putative binding cavities were not suitable for such structures it was imperative that all conformers in the NMR ensemble were subjected to both energy- and geometry-based pocket-searching programs in order to demonstrate that similar pockets were identified by both methods, and that an objective consensus could be reached with respect to the most appropriate pocket at the S100P dimeric interface for further study.

Four cavity-detection algorithms were investigated as part of this research: Pocket-Finder, Q-SiteFinder, Fpocket and Site Finder. These four tools use either energy-based (Q-SiteFinder) or geometry-based (Pocket-Finder, Fpocket and Site Finder) methods to determine potential binding sites on macromolecules. In a large-scale validation study Fpocket and Site Finder were found to successfully predict the expected binding site for a group of diverse high quality X-ray crystal structure protein-ligand complexes in their top five predicted pockets with a success rate of more than 80% [262]. When Q-SiteFinder was used to predict binding sites in protein structures without a co-complexed ligand, the expected pocket was predicted in the top three sites for 86% of the proteins studied [261]. In contrast, Pocket-Finder, a geometry-based pocket-detecting algorithm developed to compare the performance of the energy-based algorithm, was found to be less successful than Q-SiteFinder in predicting putative binding sites, but performed better than LIGSITE [317], the pocket-detecting algorithm it was based on [261]. All four algorithms demonstrate a high success rate in predicting

the correct site in their top 5 predictions, and as such were credible tools with which to identify putative binding sites on S100P.

Both Q-SiteFinder and Pocket-Finder programmes returned the ten largest pockets detected for each S100P conformer in the NMR ensemble (Table 2.2 and 2.3). Although the same number of pockets were returned by both programmes, they were ranked differently and thus, pocket 1 for conformer 1 in Q-SiteFinder does not correspond to pocket 1 for the same conformer in Pocket-Finder. That said, some of the pockets identified by Pocket-Finder were subsets of those identified by Q-SiteFinder (discussed below).

The volumes observed for putative binding pockets is not constant. Individual pocket volumes differed between the programmes used because each algorithm uses a different method to calculate pocket volumes. In addition, the volumes between similar pockets on different conformers of the ensemble will also differ. This is due to the variation in protein conformation that is representative of the dynamic nature of the protein in solution. Developers of Q-SiteFinder and Pocket-Finder used different parameters to calculate site volume [261]. In general, volumes estimated by Q-SiteFinder on all conformers of 1OZO were 3-12 times bigger than those estimated by Pocket-Finder. Q-SiteFinder however was reported to have a higher accuracy and precision rate than Pocket-Finder in predicting binding sites, with predicted sites mapping accurately onto ligand coordinates [261].

**Table 2.2**: Potential binding pockets and their estimated volumes identified by Q-SiteFinder for each of the S100P conformers in the 1OZO ensemble. Pockets on the dimeric interface are highlighted in yellow.

| S100P NMR ensemble Conformer | Protein volume (Å³) | Predicted site volumes (Å³) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| 1 | 18546 | 346 | 331 | 297 | 236 | 228 | 129 | 167 | 163 | 129 | 124 |
| 2 | 18536 | 366 | 238 | 252 | 252 | 190 | 214 | 170 | 153 | 172 | 129 |
| 3 | 18591 | 469 | 358 | 326 | 270 | 179 | 179 | 174 | 163 | 136 | 103 |
| 4 | 18610 | 341 | 259 | 271 | 207 | 211 | 216 | 165 | 119 | 130 | 116 |
| 5 | 18456 | 327 | 252 | 220 | 279 | 206 | 180 | 153 | 147 | 135 | 110 |
| 6 | 18495 | 456 | 324 | 235 | 300 | 258 | 206 | 217 | 141 | 128 | 114 |
| 7 | 18608 | 313 | 204 | 203 | 181 | 194 | 170 | 171 | 154 | 146 | 127 |
| 8 | 18619 | 321 | 296 | 251 | 255 | 178 | 198 | 144 | 149 | 157 | 163 |
| 9 | 18559 | 487 | 459 | 340 | 269 | 152 | 156 | 107 | 135 | 115 | 110 |
| 10 | 18618 | 783 | 275 | 254 | 195 | 180 | 184 | 141 | 114 | 130 | 116 |
| 11 | 18503 | 388 | 217 | 220 | 149 | 148 | 134 | 131 | 125 | 116 | 129 |
| 12 | 18500 | 371 | 356 | 297 | 240 | 237 | 200 | 185 | 136 | 148 | 102 |
| 13 | 18556 | 381 | 253 | 353 | 324 | 207 | 155 | 199 | 164 | 108 | 109 |
| 14 | 18641 | 255 | 308 | 177 | 189 | 166 | 185 | 161 | 124 | 115 | 145 |
| 15 | 18566 | 349 | 321 | 285 | 190 | 176 | 202 | 182 | 198 | 167 | 110 |
| 16 | 18474 | 509 | 386 | 289 | 254 | 224 | 133 | 122 | 106 | 106 | 126 |

**Table 2.3**: Potential binding pockets and their estimated volumes identified by Pocket-Finder for each of the S100P conformers in the 1OZO ensemble. Pockets on the dimeric interface are highlighted in yellow.

| S100P NMR ensemble conformer | Protein volume (Å³) | Predicted site volumes (Å³) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| 1 | 18546 | 67 | 51 | 44 | 28 | 22 | 21 | 24 | 20 | 20 | 16 |
| 2 | 18536 | 101 | 67 | 56 | 52 | 54 | 44 | 38 | 29 | 30 | 25 |
| 3 | 18591 | 58 | 54 | 51 | 36 | 27 | 24 | 23 | 20 | 15 | 14 |
| 4 | 18610 | 77 | 66 | 47 | 51 | 44 | 40 | 35 | 37 | 28 | 20 |
| 5 | 18456 | 88 | 59 | 54 | 49 | 47 | 37 | 31 | 25 | 23 | 24 |
| 6 | 18495 | 148 | 79 | 70 | 55 | 44 | 38 | 35 | 20 | 17 | 15 |
| 7 | 18608 | 86 | 63 | 70 | 41 | 38 | 35 | 19 | 19 | 17 | 17 |
| 8 | 18619 | 75 | 44 | 42 | 38 | 24 | 20 | 22 | 18 | 17 | 16 |
| 9 | 18559 | 71 | 62 | 58 | 39 | 35 | 31 | 28 | 25 | 23 | 17 |
| 10 | 18618 | 65 | 62 | 34 | 32 | 21 | 20 | 20 | 14 | 17 | 13 |
| 11 | 18503 | 90 | 72 | 61 | 51 | 46 | 41 | 39 | 40 | 35 | 26 |
| 12 | 18500 | 65 | 58 | 58 | 55 | 49 | 45 | 36 | 34 | 32 | 29 |
| 13 | 18556 | 105 | 82 | 56 | 52 | 54 | 48 | 33 | 33 | 26 | 24 |
| 14 | 18641 | 86 | 76 | 67 | 40 | 26 | 19 | 18 | 16 | 14 | 15 |
| 15 | 18566 | 70 | 68 | 47 | 47 | 48 | 50 | 39 | 30 | 28 | 25 |
| 16 | 18474 | 65 | 65 | 59 | 55 | 50 | 34 | 34 | 23 | 17 | 14 |

Similar to Pocket-Finder, both Fpocket and Site Finder employ geometric approaches to search for putative binding pockets. Geometric pocket-detecting programs use the geometry of the protein surface to compute potential binding sites [259, 265, 318, 319].

In contrast to Q-SiteFinder and Pocket-Finder, where 10 pockets were identified for each conformer of 1OZO, Fpocket identified between seven and 12 pockets in total (Table 2.4). Despite this disparity in the number of pockets identified in each conformer, some pockets identified by Fpocket were similar to those previously identified by both Q-SiteFinder and Pocket-Finder (as discussed below).

The output for putative sites identified by Site-Finder was different to the three pocket-detecting algorithms discussed above. In addition to the residues involved in the identified sites, the program also returned the number of alpha spheres in the site which is indicated as Size in Table 2.5. The number of hydrophobic and sidechain contact atoms in the receptor are indicated under Hyd and Side respectively. In upgrades of MOE, sites were ranked according to their propensity for ligand binding (PLB) i.e. druggability (Table 2.5B, [320]).

**Table 2.4**: Potential binding pockets identified by Fpocket for each of the S100P conformers in the 1OZO NMR ensemble. Pockets on the dimeric interface are highlighted in yellow.

| | | Real volume approximation (Å³) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **S100P Protein** | **No. of pockets detected** | **Pocket 0** | **Pocket 1** | **Pocket 2** | **Pocket 3** | **Pocket 4** | **Pocket 5** | **Pocket 6** | **Pocket 7** | **Pocket 8** | **Pocket 9** | **Pocket 10** | **Pocket 11** |
| 1 | 9 | 603 | 592 | 573 | 462 | 226 | 131 | 163 | 249 | 550 | | | |
| 2 | 8 | 272 | 256 | 179 | 405 | 541 | 427 | 135 | 101 | | | | |
| 3 | 8 | 877 | 740 | 659 | 663 | 384 | 413 | 149 | 548 | | | | |
| 4 | 10 | 727 | 665 | 751 | 672 | 209 | 483 | 140 | 712 | 613 | 433 | | |
| 5 | 9 | 919 | 271 | 213 | 115 | 228 | 467 | 159 | 427 | 90 | | | |
| 6 | 7 | 875 | 876 | 160 | 161 | 547 | 151 | 633 | | | | | |
| 7 | 11 | 1196 | 504 | 317 | 556 | 497 | 286 | 179 | 323 | 131 | 310 | 571 | |
| 8 | 8 | 994 | 508 | 477 | 483 | 587 | 156 | 607 | 84 | | | | |
| 9 | 8 | 489 | 575 | 288 | 401 | 168 | 410 | 92 | 431 | | | | |
| 10 | 7 | 941 | 958 | 561 | 416 | 300 | 166 | 149 | | | | | |
| 11 | 8 | 884 | 247 | 613 | 305 | 218 | 267 | 146 | 78 | | | | |
| 12 | 10 | 808 | 240 | 217 | 608 | 319 | 619 | 134 | 639 | 128 | 313 | | |
| 13 | 10 | 309 | 199 | 490 | 378 | 550 | 258 | 473 | 165 | 96 | 112 | | |
| 14 | 6 | 714 | 954 | 160 | 314 | 147 | 139 | | | | | | |
| 15 | 7 | 958 | 447 | 302 | 198 | 653 | 341 | 880 | | | | | |
| 16 | 12 | 232 | 310 | 473 | 373 | 461 | 458 | 330 | 630 | 273 | 203 | 120 | 312 |

**Table 2.5**: Pockets identified by two different versions of MOE's Site Finder on conformer 15 of 1OZO. A) MOE2010.09. B) MOE2011.10.Pockets on the dimeric interface are highlighted in yellow.

**A**

| Site | Size* | Hyd[δ] | Side[†] | Residues |
|---|---|---|---|---|
| 1 | 101 | 22 | 82 | 1: (H86 K91 T92 K95) 2: (I12 F15 S16 S19 G20 S21 E22 Q26 D70 F71) |
| 2 | 77 | 19 | 64 | 1: (I12 F15 S16 S19 S21 Q26 F71) 2: (H86 K87 Y88 F89) |
| 3 | 106 | 11 | 81 | 1: (M1 E5 A6 M8 G9 M10 I12 D13 S16) 2: (I81 T82 S85 H86 F89 T92) |
| 4 | 47 | 8 | 32 | 2: (K30 K49 D50 K51 D52 A53 V54 D55) |
| 5 | 55 | 6 | 40 | 1: (L41 P42 G43 F44 A84 S85 E90 K91) 2: (M1 T2 E5) |

*Number of contributing spheres
[δ]Number of hydrophobic points
[†]Number of sidechain contact atoms

**B**

| Site | Size | PLB[θ] | Hyd | Side | Residues |
|---|---|---|---|---|---|
| 1 | 85 | 1.20 | 10 | 66 | 1: (M1 E5 A6 M8 G9 I12) 2: (I81 T82 S85 H86 F89 T92) |
| 2 | 90 | 0.81 | 21 | 74 | 1: (H86 K91 T92 K95) 2: (I12 F15 S16 S19 G20 S21 E22 Q26 F71) |
| 3 | 74 | 0.23 | 19 | 62 | 1: (I12 F15 S16 S19 S21 Q26 F71) 2: (H86 F89) |
| 4 | 34 | -0.72 | 3 | 27 | 1: (M1 A6 G9 M10 I12 D13 S16) 2: (H86 F89) |
| 5 | 55 | -1.52 | 6 | 40 | 1: (L41 P42 G43 F44 A84 S85 E90 K91) 2: (M1 T2 E5) |

[θ]Propensity for ligand binding score

Pockets at the dimeric interface identified by each of the algorithms on all conformers of 1OZO were comparable, e.g. pocket 3 from Pocket-Finder on conformer 15 is a subset of pocket 1 on the same conformer from Q-SiteFinder (Figure 2.5). The same is true for two pockets identified by Site Finder on the same conformer. In MOE 2010.09, Site Finder predicted two sites that were subsets of pockets 1 and 8 identified by Q-SiteFinder (Figure 2.5). These sites were ranked first and third respectively according to their ability to bind small drug-like molecules or PLB score [320].

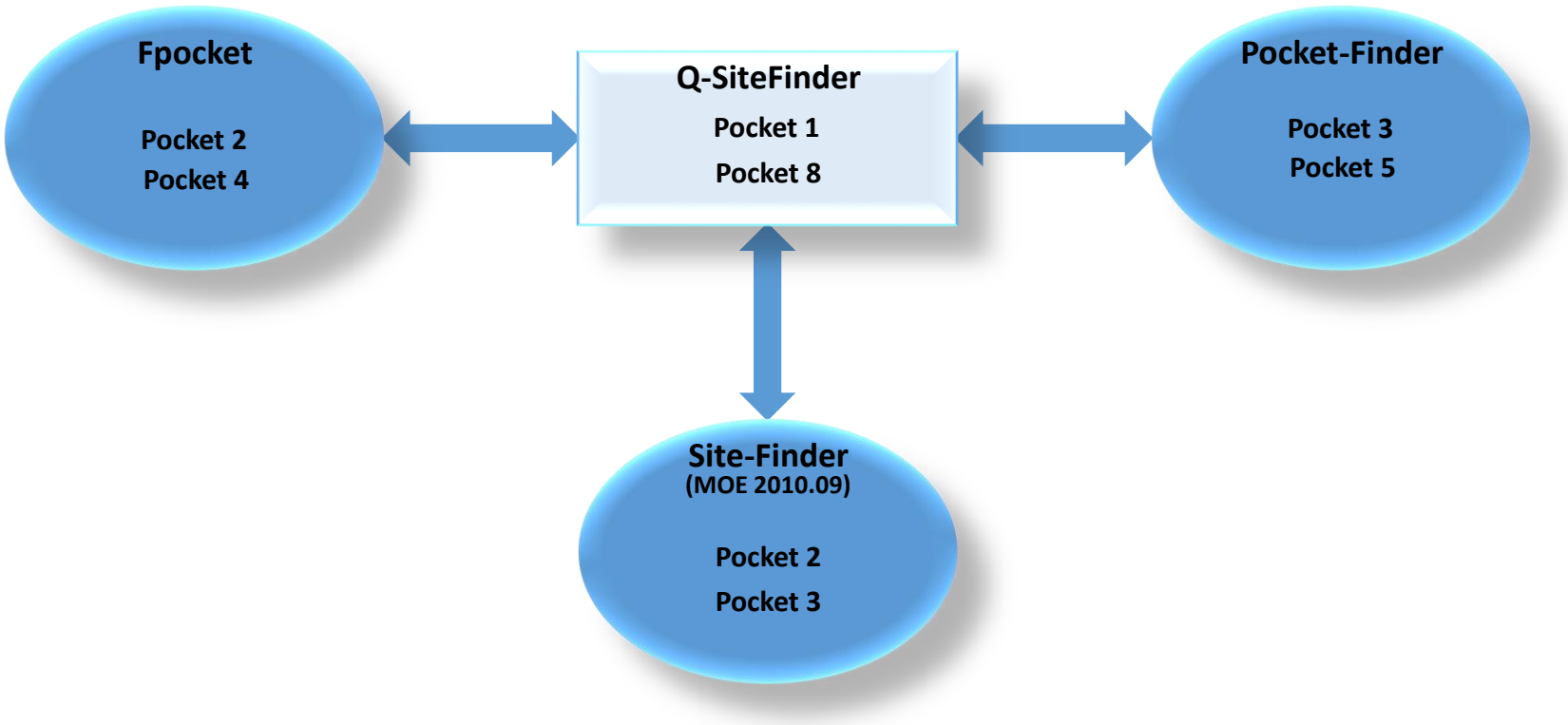



**Figure 2.5**: Overlap between pockets identified by Q-SiteFinder, Site Finder, Pocket-Finder and Fpocket on the dimeric interface of conformer 15 of the NMR ensemble of S100P. All pockets shown here identified by Fpocket, Site Finder and Pocket-Finder are subsets of pockets 1 and 8 identified by Q-SiteFinder on conformer 15 of 1OZO.

#### 2.3.2.2 **Binding sites on the dimeric interface of 1OZO**

It has been shown that the biological function of S100P is dependent upon dimerization [284, 292, 321, 322]. Hence, it was hypothesised that cromolyn exerted its effect on S100P-RAGE interaction by binding at or near to the homodimeric interface, thus preventing the protein interacting effectively with its receptor. The significance of the dimeric interface was highlighted in a recent report that showed residues that form part of the dimer interface of S100P – residues 83-94, and residues 2-14 – as important in binding to RAGE [294]. Consequently, only putative sites identified at the dimeric interface of the S100P 1OZO NMR ensemble were examined in detail in subsequent docking and virtual screening studies.

The pocket-detection algorithms identified two cavities at the dimeric interface for each conformer in the NMR ensemble, one on each side of the protein. However, whilst there were similarities between these pockets identified for a given conformer, none were identical e.g., pockets identified by Site Finder at the dimeric interface of conformer 15 were not identical with respect to either size or amino acid composition (Table 2.5). This lack of symmetry at the dimeric interface of members of the NMR ensemble of S100P is in contrast to a recent publication by Penumutchu *et al*. [323] who identified two symmetrical pockets on an NMR ensemble of S100P that they resolved using the coordinates of 1J55, the crystal structure of S100P. Manipulation of the monomeric form of the protein will result in symmetrical dimeric interfaces given that this manipulation is a simple superposition of the monomer. S100P is a dynamic protein, and modelling a dynamic structure on a static structure has the potential to introduce

flaws into the system. The NMR ensemble 1OZO is direct experimental evidence against the symmetric nature of the S100P dimeric binding site.

However, a second report published by the same group [294] identified residues E5, D13, F44, Y88 and F89 as playing a vital role in S100P-RAGE binding. These residues were also identified in the dimeric interface of some of the pockets identified in this study. In addition, residues that form the hydrophobic core that is exposed upon $Ca^{2+}$ binding to S100P [141] were also present in sites identified at the dimer interface in this study. This hydrophobic core, formed mainly by residues F15, F71, and F74 [159], has been reported to be important in target interaction in S100P [138, 284]. All this evidence points to the S100P NMR ensemble as a valid template in drug design studies.

#### 2.3.2.3 **Potential binding sites on the X-ray crystal structure of S100P (1J55)**

The functional dimer form of 1J55 was generated by the UCSF Chimera package [299] using the atomic coordinates included in the PDB file of the protein. A symmetrical dimer resulted with four bound calcium ions (Figure 2.6). The generated dimer formed was an exact and opposite mirror image of the original monomer. Given that this symmetry was not observed in any of the structures of the NMR ensemble, there is a possibility that the dimeric interface created using the coordinates of 1J55 could have some limitations and have introduced some errors into the system. Moreover, when the resolved dimer was investigated for potential binding sites using Q-SiteFinder, identified pockets had smaller volumes compared to those of the NMR conformers

(Figure 2.7). These pockets were too small to accommodate cromolyn, a known inhibitor of the S100P-RAGE interaction, and the ligand to be used for molecular docking [195]. The volume of cromolyn is 375 $Å^3$ (calculated from http://www.molinspiration.com/cgi-bin/properties).

Sequence-only superposition of the C-α atoms of the 1J55 monomer with Chain A of conformer 15 of 1OZO, the most representative of the NMR ensemble, using the Protein Superpose tool in MOE gave a RMSD of 3.75 Å (Figure 2.8). Structural superposition of the two models using the Flexible structure Alignment by Changing AFPs (Aligned Fragment Pairs) with Twists (FATCAT) [324] however resulted in a RMSD of 2.73 Å. Although these values may seem high for models of the same protein (and suggest significantly different structures), this apparent discrepancy could be due to several factors. Regions of variability (traffic-light colour coded in Figure 2.8) are those in the flexible loop regions. These regions have high *B*-values in the X-ray model, demonstrating the uncertainty in the location of these atoms in the crystal lattice. In addition, the superposition was only carried out over 88 residues as 1J55 has part of its linker region missing as well as the last terminal residue in its model. Also of note is the position of helix 2 and length of helix 4 in the apo and $Ca^{2+}$-bound states of the protein. Molecular dynamics (MD) simulations have been used to show a change in the relative position of this helix in apo-S100P and $Ca^{2+}$-bound S100P [325]. The length of helix 4 is believed to be extended by six residues upon $Ca^{2+}$ binding compared to the apo state [298]. All these factors could contribute to the high RMSD value.

**Figure 2.6**: The dimerised structure of 1J55 (Chain A cyan, Chain B red) generated by the USCF Chimera software [299] showing four bound calcium ions (green spheres). Red and cyan spheres represent water molecules from the crystal structure.

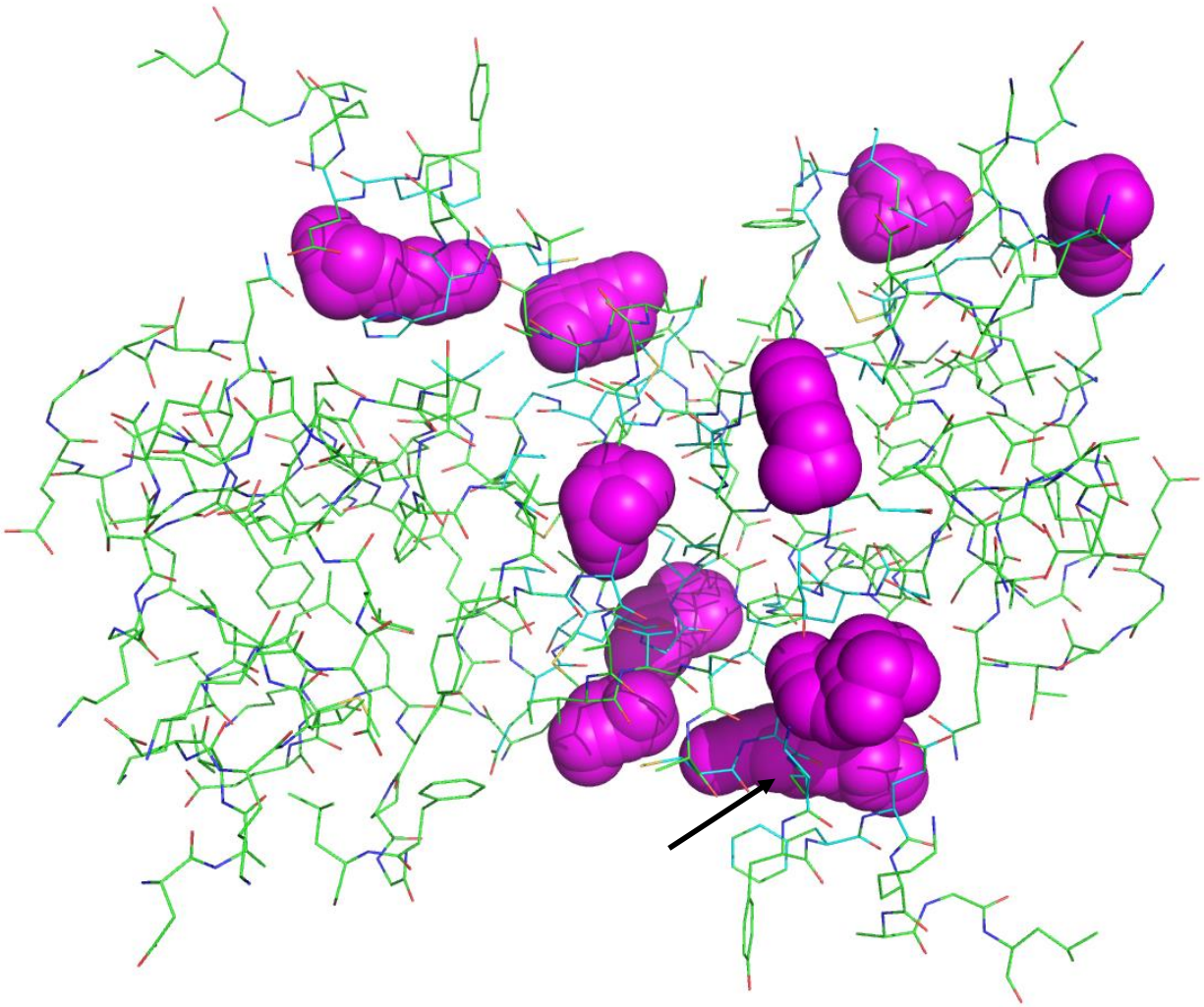


**Figure 2.7**: Putative binding pockets (magenta spheres) on the dimerised structure of 1J55 predicted by Q-SiteFinder. The largest predicted pocket (arrow) has a volume of 123 $Å^3$ which is far too small to accommodate cromolyn (375 $Å^3$).

**A**

**B**

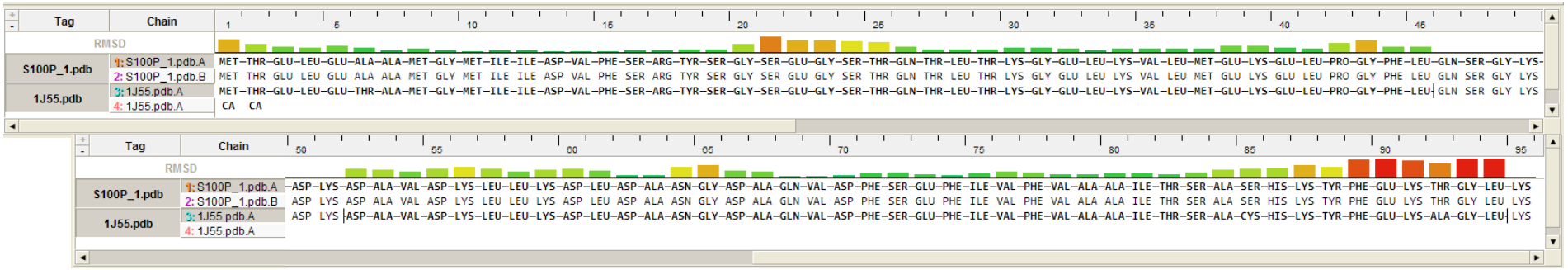



**Figure 2.8**: Superposition of C-α atoms of 1J55 with Chain A of conformer 15 of the 1OZO ensemble in MOE. A) Cartoon model of monomeric 1J55 (cyan) with bound $Ca^{2+}$ (green spheres) superposed with Chain A of conformer 15 (red). B) RMSD bar graphs of the superposed atoms. Green colours indicate high conformation similarity while red indicate low similarity in backbone conformation.

### 2.3.3 Molecular docking of cromolyn to the S100P dimeric interface

Initial dockings studies were focused on sites located at the dimer interface that had a volume close to or larger than the calculated ligand volume e.g. pocket 1 of conformers 9 and 15 (Table 2.6). However, in spite of their volume, docking cromolyn to these pockets showed that only part of the ligand was found to fit into the cavity, with the remainder protruding into solvent (Figure 2.9). Closer examination of the S100P conformers showed that for some of them there was more than one potential binding site at the dimer interface, and sites that had been identified as separate sites by the pocket-detection algorithms were actually close enough to be thought of as two lobes of a larger binding pocket (Figure 2.10). The pocket-detection algorithms work by identifying regions of tight atomic packing (Fpocket, Site Finder), or by using a probe of a certain radius (Pocket-Finder), or clusters with favourable binding energy (Q-SiteFinder). It is thus feasible that two pockets that are close to each other but are separated by a few atoms that do not fall within the algorithms' criteria for inclusion within a putative site will be identified as two separate sites instead of one. Where these pockets exist, they were combined to form the larger cavity and the dockings experiments repeated.

Analysis of all conformers in the 1OZO NMR ensemble showed that only one member, conformer 15, was able to yield a combined cavity at the dimeric interface that was of sufficiently large volume to accommodate cromolyn, and that this occurred for only one of the two cavities at the dimeric interface in the structure. These two sites, Pockets 1 and 8 in Table 2.6, were also identified by the other pocket-detection

algorithms on the same conformer (Table 2.7). As such, only this conformer was carried forwards for docking studies.

**Table 2.6**: Volumes of potential binding pockets identified at the dimer interface of conformers 1OZO (highlighted in yellow) by Q-SiteFinder.

| 1OZO conformer | Predicted site volumes (Å$^3$) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| 1 | 346 | 331 | 297 | 236 | 228 | 129 | 167 | 163 | 129 | 124 |
| 2 | 366 | 238 | 252 | 252 | 190 | 214 | 170 | 153 | 172 | 129 |
| 3 | 469 | 358 | 326 | 270 | 179 | 179 | 174 | 163 | 136 | 103 |
| 4 | 341 | 259 | 271 | 207 | 211 | 216 | 165 | 119 | 130 | 116 |
| 5 | 327 | 252 | 220 | 279 | 206 | 180 | 153 | 147 | 135 | 110 |
| 6 | 456 | 324 | 235 | 300 | 258 | 206 | 217 | 141 | 128 | 114 |
| 7 | 313 | 204 | 203 | 181 | 194 | 170 | 171 | 154 | 146 | 127 |
| 8 | 321 | 296 | 251 | 255 | 178 | 198 | 144 | 149 | 157 | 163 |
| 9 | 487 | 459 | 340 | 269 | 152 | 156 | 107 | 135 | 115 | 110 |
| 10 | 783 | 275 | 254 | 195 | 180 | 184 | 141 | 114 | 130 | 116 |
| 11 | 388 | 217 | 220 | 149 | 148 | 134 | 131 | 125 | 116 | 129 |
| 12 | 371 | 356 | 297 | 240 | 237 | 200 | 185 | 136 | 148 | 102 |
| 13 | 381 | 253 | 353 | 324 | 207 | 155 | 199 | 164 | 108 | 109 |
| 14 | 255 | 308 | 177 | 189 | 166 | 185 | 161 | 124 | 115 | 145 |
| 15 | 349 | 321 | 285 | 190 | 176 | 202 | 182 | 198 | 167 | 110 |
| 16 | 509 | 386 | 289 | 254 | 224 | 133 | 122 | 106 | 106 | 126 |

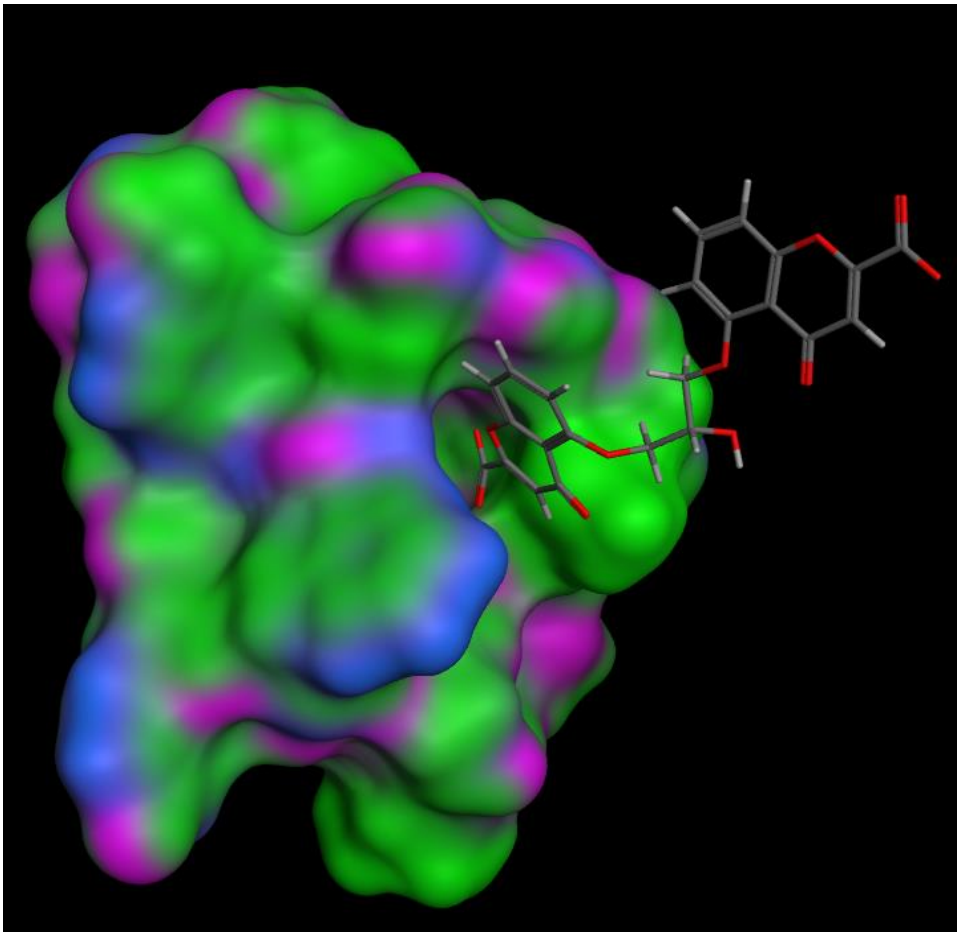


**Figure 2.9**: Docked cromolyn in the first predicted pocket of conformer 9 of the S100P 1OZO NMR ensemble. Only part of the ligand is docked to the protein with the rest protruding into the solvent.

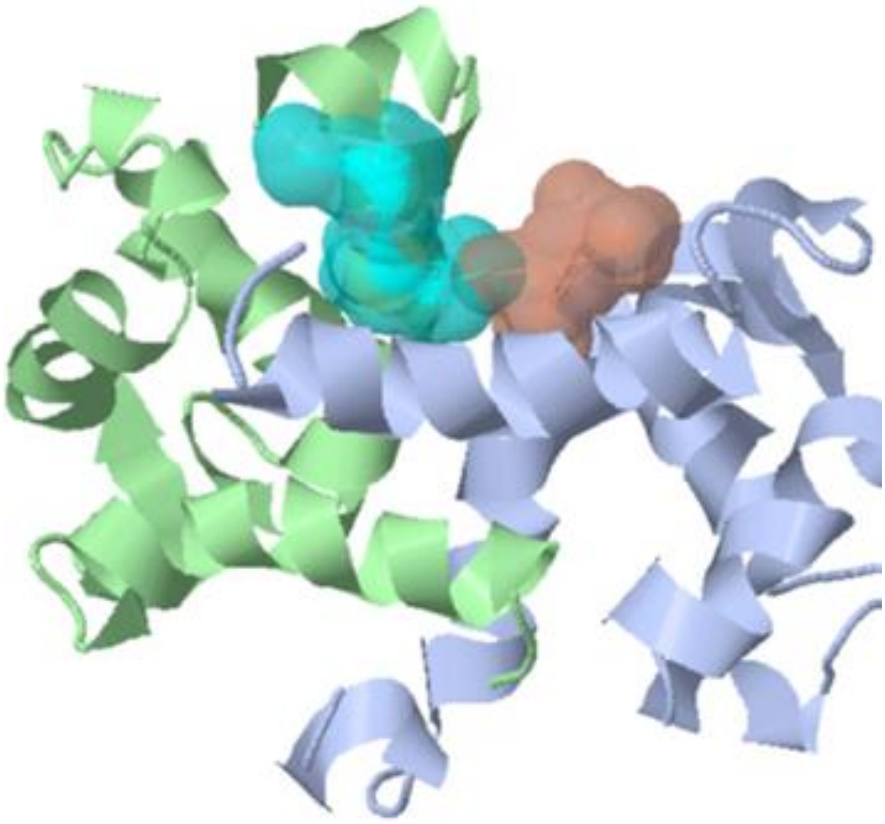


**Figure 2.10**: Two pockets (cyan and brown) at the dimeric interface of conformer 15 of 1OZO. These were initially identified by Q-SiteFinder as distinct individual pockets which, upon closer inspection were observed to overlap with one another. As a result, they were combined as one large pocket that was used to dock cromolyn.

**Table 2.7**: Pockets identified by Site Finder, Pocket-Finder and Fpocket on the dimeric interface of conformer 15 of the NMR ensemble of S100P were also identified by Q-SiteFinder.

| | | | **Amino acid residues at the dimeric interface of conformer 15** | |
|---|---|---|---|---|
| **Algorithm** | **Pocket** | **Pocket size ($Å^{3}$)** | **Chain A** | **Chain B** |
| **Q-SiteFinder** | 1 | 349 | M1 T2 E5 A6 M8 G9 I12 F71 S72 I75 | F44 V78 A79 A80 I81 T82 S83 A84 S85 H86 K87 Y88 F89 K91 T92 G93 L94 K95 |
| | 8 | 198 | G9 I11 I12 D13 F15 S16 S19 S21 Q26 F71 S72 F74 I75 V78 | T82 H86 |
| **Site Finder (MOE 2010.09)** | 2 | 77* | M1 E5 A6 M8 G9 M10 I12 D13 S16 | I81 T82 S85 H86 F89 T92 |
| | 3 | 106* | I12 F15 S16 S19 S21 Q26 F71 | H86 L87 Y88 F89 |
| **Pocket-Finder** | 2 | 76 | T82 S83 H86 T92 | I12 F15 F71 I75 |
| | 3 | 67 | M1 E5 G9 I12 | T82 S85 H86 T92 |
| | 5 | 26 | T2 L4 E5 M8 | I11 V14 E40 L41 F74 V78 I81 |
| **Fpocket** | 2 | 302 | M1 E5 G9 I12 | S85 H86 F89 T92 |
| | 4 | 653 | I12 F15 S16 S19 S21 Q26 F71 | H86 F89 |

*Number of contributing spheres

Docking studies of cromolyn into conformer 15 of 1OZO revealed possible binding interactions between the protein and ligand at the dimeric interface. Visual analyses of the ligand poses showed that in most cases, the first ranked ligand pose did not have the optimal binding interactions with the protein. Indeed, the ligand pose which was subsequently used to design pharmacophore queries was ranked 18$^{th}$ out of the 36 poses returned. The observed interactions between the ligand and protein involved three residues of Chain B: F89, T82 and T92. There was hydrophobic contact between F89 with the chromone moiety of the ligand while both threonine residues were involved in hydrogen bonding interaction with both carboxylate oxygens at one end of the ligand (Figure 2.11).

The predicted involvement of F89 in binding interactions with the ligand is significant as this residue has been reported to provide important hydrophobic contact that is crucial in target recognition in S100A1 and S100B [326, 327]. S100P shares a sequence identity of approximately 50% and 44% with S100A1 [136] and S100B [159] respectively. Thus, evidence of the involvement of this residue in a $\pi$-$\pi$ interaction with cromolyn is a credible starting point when designing potential inhibitors against S100P as blocking this hydrophobic contact with small molecules could have the potential to inhibit S100P binding to its target, RAGE.

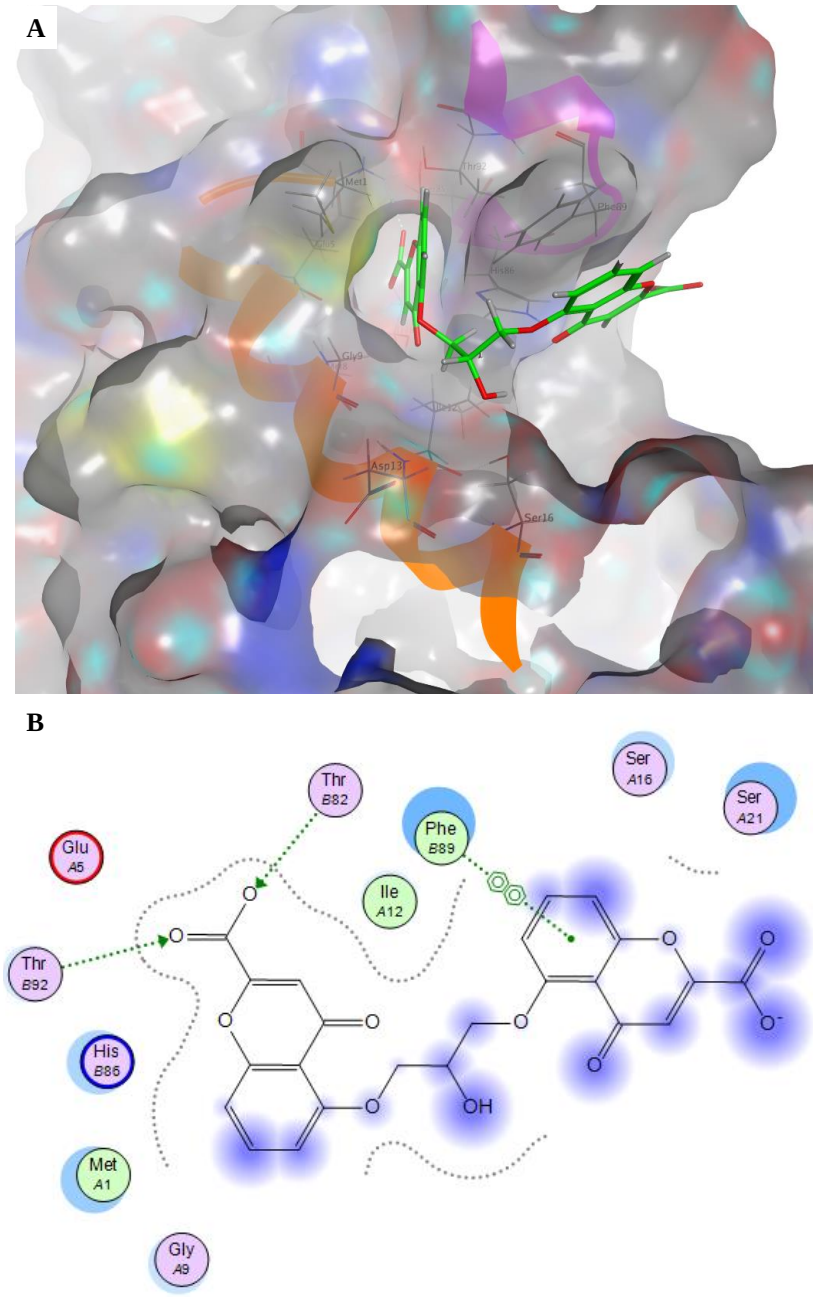



**Figure 2.11**: Predicted binding interactions between S100P and the ligand cromolyn. A) Docked cromolyn (ball and stick in green) in the overlapped sites at the dimeric interface of conformer 15 of S100P 1OZO NMR ensemble (shown here with the molecular surface and residues forming the binding sites). B) Predicted binding interactions between cromolyn and S100P.

### 2.3.4 Pharmacophore generation, virtual screening and hit clustering

From the observed binding interaction between cromolyn and S100P (Figure 2.11), a three-point pharmacophore model was derived comprising of an aromatic centre (Aro), and two anionic acceptors (Ani&Acc) (Figure 2.12). These features represent the hydrophobic contact between F89 and the chromone moiety of cromolyn, and the hydrogen bonding between one carboxylate group of the ligand and T82 and T92. This was deemed a "stringent" pharmacophore query due to the anionic requirement applied to both hydrogen bond acceptor features. This means that "hits" fulfilling the pharmacophoric constraints cannot only be hydrogen bond acceptors, but must also be anionic. This stringency was evidenced from the number of "hits" obtained when the MOE database of lead-like compounds was screened with this query.

Out of the 653,214 compounds in the database, only 52 came up as '"hits"', 0.008% of total compounds (Table 2.8). As expected, most of these "hits" were acids, a fundamental requirement for fulfilling the anionic constraint of the pharmacophore query (Figure 2.13). The average molecular weight of hit compounds was 389 compared to 468 for cromolyn. All "hits" possess rings within them, ranging from two to six, demonstrating the diversity in their structures.

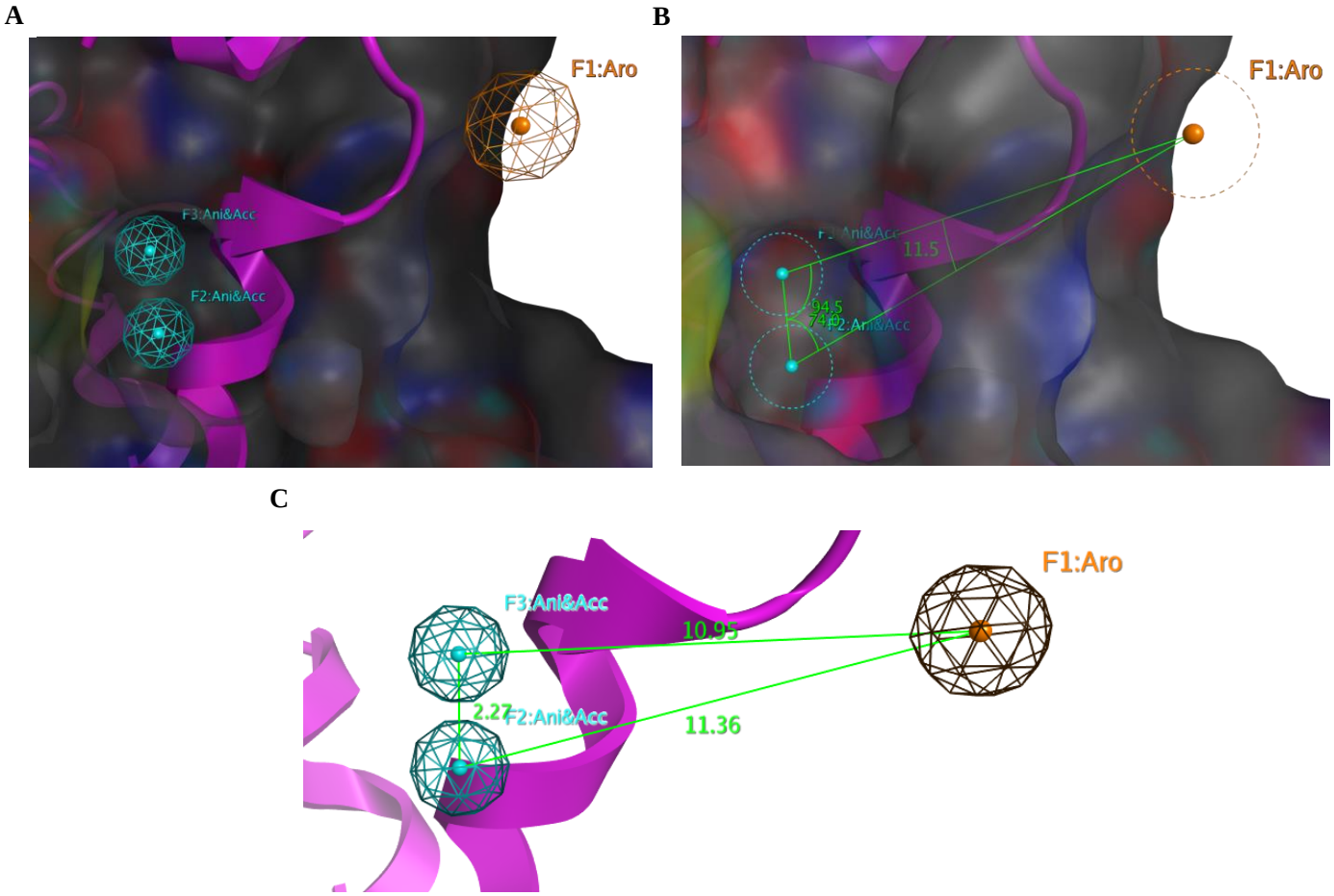



**Figure 2.12**: Pharmacophore model derived from the predicted binding interactions between S100P and cromolyn. A) The three pharmacophore features represented by spheres in S100P (with molecular surface). B) Angles between the pharmacophore points. C) Distances (in Å) between pharmacophore features.

**Table 2.8**: "Hits" obtained when the MOE database was screened with a "stringent" pharmacophore query. "Hits" sorted according to their commercial vendor ID. Molecular descriptors – weight, number of rings, LogP, total polar surface area (TPSA), H-Donors and H-Acceptors – were calculated using the Descriptor feature in MOE.

| Molecule ID | Rings | Molecular weight | LogP | TSPA | H-Donors | H-Acceptors |
|---|---|---|---|---|---|---|
| Akos LT-1098 X 2900 (**6**) | 2 | 340.21 | 4.48 | 65.97 | 0 | 1 |
| Akos LT-1098 X 4644 (**7**) | 2 | 271.32 | 3.23 | 65.97 | 0 | 1 |
| Art-Chem UZI/6188182 (**8**) | 4 | 332.31 | 2.56 | 119.19 | 2 | 3 |
| Asinex ASN 03791583 (**9**) | 4 | 405.43 | 5.08 | 83.73 | 0 | 3 |
| Asinex BAS 02011727 (**10**) | 4 | 441.85 | 3.86 | 139.76 | 1 | 3 |
| Asinex BAS 02380039 (**11**) | 3 | 408.46 | 4.36 | 103.71 | 1 | 4 |
| Biofocus 144_5936_7270_6684 (**12**) | 4 | 376.44 | 3.03 | 78.26 | 0 | 2 |
| Biofocus 144_5940_7270_6684 (**13**) | 4 | 406.46 | 3.02 | 87.49 | 0 | 3 |
| Chem T&I BOBM0003527 (**14**) | 4 | 377.33 | 1.49 | 127.17 | 1 | 5 |
| Chem T&I UZI/4036985 (**15**) | 2 | 342.44 | 3.56 | 69.23 | 1 | 1 |
| Chembridge 5241336 (**16**) | 3 | 433.45 | 2.57 | 166.33 | 1 | 4 |
| Chembridge 5311676 (**17**) | 3 | 403.42 | 2.62 | 157.10 | 2 | 4 |
| Chembridge 5862836 (**18**) | 3 | 440.27 | 5.38 | 90.82 | 1 | 3 |
| Chembridge 6640941 (**19**) | 3 | 368.39 | 4.48 | 103.18 | 1 | 3 |
| Chembridge 6688677 (**20**) | 3 | 430.85 | 5.02 | 120.25 | 1 | 4 |
| Chembridge 6897450 (**21**) | 4 | 414.37 | 4.15 | 126.07 | 0 | 2 |
| Chembridge 7230553 (**22**) | 3 | 434.21 | 3.75 | 115.84 | 1 | 4 |
| Chembridge 7266811 (**23**) | 3 | 413.79 | 3.41 | 115.84 | 1 | 4 |
| Chembridge 7356270 (**24**) | 6 | 439.40 | 3.73 | 123.33 | 0 | 2 |
| Chembridge 7778804 (**25**) | 3 | 389.36 | 4.20 | 118.59 | 1 | 2 |
| Chembridge 7926943 (**26**) | 3 | 359.37 | 2.13 | 110.84 | 1 | 3 |
| Chemdiv 1805-1431 (**27**) | 4 | 398.40 | 4.21 | 112.33 | 1 | 4 |
| Chemdiv 6049-3009 (**28**) | 5 | 439.49 | 6.00 | 90.17 | 2 | 2 |
| Chemstar CHS 2401453 (**29**) | 3 | 404.36 | 7.21 | 136.64 | 1 | 2 |
| Comgenex CGX-3084890 (**30**) | 2 | 358.42 | 2.80 | 96.28 | 1 | 3 |
| Enamine T0508-3103 (**31**) | 2 | 335.34 | 4.51 | 117.41 | 2 | 3 |
| Enamine T0509-5138 (**32**) | 3 | 412.24 | 2.20 | 113.94 | 2 | 3 |
| Enamine T0514-7139 (**33**) | 3 | 431.84 | 1.65 | 141.18 | 2 | 5 |
| FCHC 29368 (**34**) | 4 | 383.39 | 6.79 | 97.19 | 1 | 3 |

**Table 2.8** cont.: "Hits" obtained when the MOE database was screened with a stringent pharmacophore query. "Hits" sorted according to their commercial vendor ID. Cromolyn is included at the bottom for comparison. Molecular descriptors – weight, number of rings, LogP, total polar surface area (TPSA), H-Donors and H-Acceptors – were calculated using the Descriptor feature in MOE. Cromolyn is included at the bottom for comparison.

| Molecule ID | Rings | Molecular weight | LogP | TSPA | H-Donors | H-Acceptors |
|---|---|---|---|---|---|---|
| InterBioScreen STOCK5S-92825 (**35**) | 3 | 432.45 | 3.42 | 151.45 | 2 | 6 |
| InterBioScreen STOCK3S-82963 (**36**) | 4 | 386.39 | 2.10 | 112.30 | 1 | 4 |
| InterBioScreen STOCK5S-29098 (**37**) | 4 | 419.46 | 2.37 | 114.26 | 1 | 4 |
| InterBioScreen STOCK5S-52328 (**38**) | 3 | 372.40 | 2.58 | 104.76 | 1 | 3 |
| InterBioScreen STOCK5S-64425 (**39**) | 3 | 365.37 | 2.67 | 122.72 | 3 | 3 |
| InterBioScreen STOCK5S-65135 (**40**) | 3 | 424.52 | 1.88 | 110.18 | 0 | 4 |
| InterBioScreen STOCK5S-78564 (**41**) | 4 | 396.38 | 1.92 | 109.11 | 0 | 3 |
| InterBioScreen STOCK5S-87979 (**42**) | 3 | 386.41 | 2.19 | 145.55 | 0 | 3 |
| InterBioScreen STOCK5S-88540 (**43**) | 3 | 373.43 | 2.65 | 114.65 | 0 | 4 |
| Labotest LT00103793 (**44**) | 3 | 347.35 | 3.87 | 92.70 | 2 | 2 |
| Labotest LT00786495 (**45**) | 2 | 372.38 | 3.98 | 127.41 | 1 | 2 |
| Life Chemicals F3074-0048 (**46**) | 3 | 387.44 | 2.97 | 102.26 | 1 | 4 |
| Lithuania 1008485 (**47**) | 3 | 439.51 | 5.25 | 91.26 | 0 | 4 |
| Maybridge HTS 05485 (**48**) | 3 | 327.32 | 1.97 | 99.72 | 2 | 3 |
| Maybridge S 11067 (**49**) | 3 | 346.41 | 3.84 | 93.95 | 1 | 3 |
| Otava 7018770072 (**50**) | 3 | 409.42 | 2.54 | 120.28 | 2 | 4 |
| Peakdale 3002075 (**51**) | 4 | 314.32 | 3.49 | 70.84 | 0 | 2 |
| Pharmeks PHAR103163 (**52**) | 5 | 395.40 | 3.79 | 102.01 | 2 | 4 |
| Princeton Biomolecular OSSK_552866 (**53**) | 2 | 383.41 | 0.96 | 150.82 | 2 | 3 |
| Princeton Biomolecular OSSK_562305 (**54**) | 4 | 409.42 | 4.87 | 98.33 | 2 | 2 |
| Scientific Exchange R-093770 (**55**) | 2 | 326.36 | 2.17 | 105.91 | 1 | 3 |
| Specs AE-641/01104012 (**56**) | 3 | 428.30 | 3.71 | 119.17 | 0 | 3 |
| Specs AH-487/15148047 (**57**) | 5 | 401.42 | 4.83 | 88.16 | 0 | 2 |
| **AVERAGE** | **3** | **389** | **3** | **111** | **1** | **3** |
| **Cromolyn (4)** | **4** | **466** | **-1.08** | **171.55** | **1** | **9** |

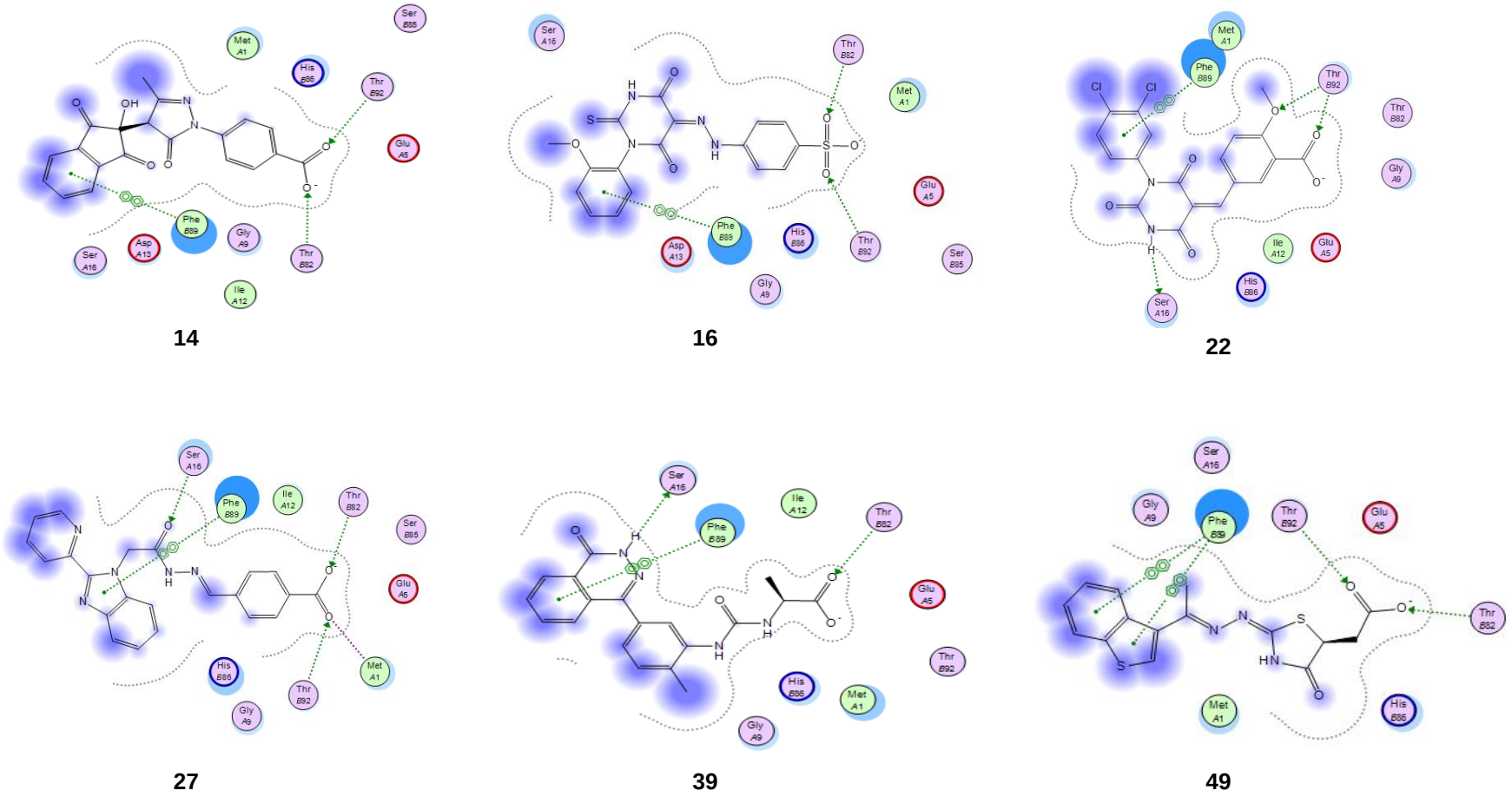



**Figure 2.13**: Some of the "hits" identified from the MOE database. Most of these "hits" were acidic because the HBA groups on the pharmacophore query that was used to identify them was also labelled as anionic.

The second "relaxed" pharmacophore query derived from the binding interactions between conformer 15 and cromolyn has the same geometric constraints as the "stringent" pharmacophore query but different pharmacophoric features. These features were defined as HBA and an aromatic centre. Screening the MOE database with this "relaxed" pharmacophore query resulted in a total of 4,619 "hits" (0.7%), again demonstrating the rigidity the anionic feature conferred on the first pharmacophore. There were 4,789 "hits" from the ZINC lead-like database giving a total of 9,408 "hits" from both databases. Clustering of these "hits" based on their chemical similarities defined as a 9-atom maximum common substructure [303] produced 299 clusters altogether with 77 singletons (Table 2.9).

#### 2.3.4.1 **Hit selection for biological screening**

Due to budgetary constraints and compound availability, only a small number of "hits" were purchased for biological testing. The first set of 13 compounds purchased (Figure 2.14) came from the 52 "hits" obtained from the virtual screening of the MOE database using the "stringent" pharmacophore query and were pragmatically selected with respect to their price and availability. Key_Organcis_12T_0223 (**58**) identified in a previous screen by Kirton (unpublished) alongside two of its analogues (**58a** and **58b**), and compound **7**, one of the "hits" identified from the MOE database (Table 2.8) were synthesised in-house (discussed in Chapter 3). Hence, a total of 17 hit compounds were subsequently tested against S100P-expressing and non-S100P expressing pancreatic cancer cells (Chapter 3).

The second set of compounds purchased for testing came from members of individual clusters from the less stringent screening of the MOE database. To ensure diversity, 52 compounds with the best *S* score from each cluster were procured (Table 2.10, Figure 2.15).

**Table 2.9**: Total number of "hits" and clusters obtained from screening the MOE and ZINC databases of lead-like compounds with a 3-point less "stringent" pharmacophore query.

| Database | Total # of compounds | Total # of "hits" | Clusters | Singletons |
|---|---|---|---|---|
| MOE | 653,214 | 4,619 | 129 | 24 |
| ZINC | 765,278 | 4,789 | 170 | 53 |
| **Total** | **1,418,492** | **9,408** | **299** | **77** |

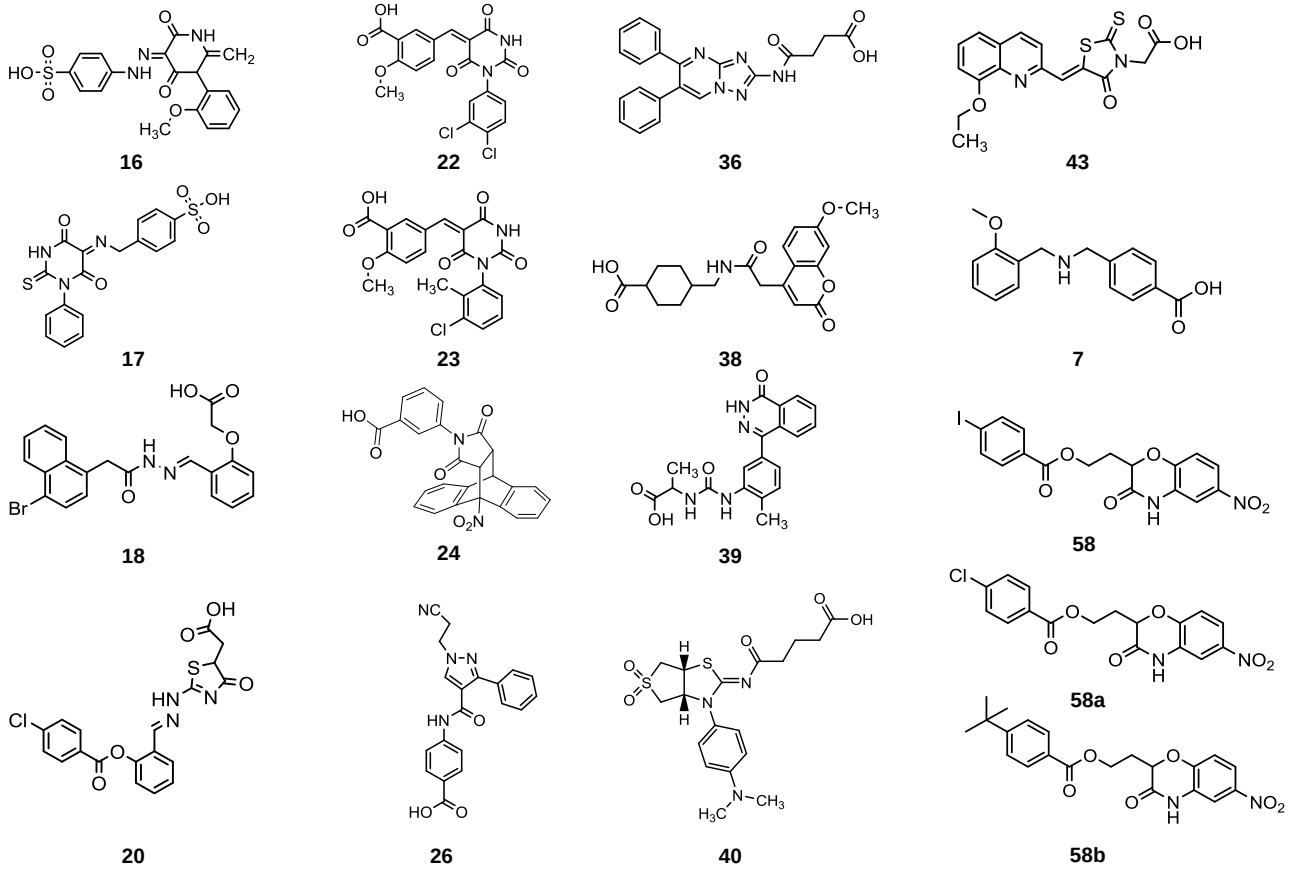



**Figure 2.14**: The first set of compounds from the MOE "stringent" virtual screening that were subjected to biological screening against pancreatic cancer cells. Compound **58** and two of its analogues, **58a** and **58b**, and compound **7**, were synthesised in-house making a total of 17 compounds that were subsequently subjected to biological screening.

**Table 2.10**: "Hits" purchased from the virtual screening of the MOE database using the less "stringent" pharmacophore query and objective selection criteria. Molecular descriptors – weight, number of rings, LogP, total polar surface area (TPSA), H-Donors and H-Acceptors – were calculated using the Descriptor feature in MOE.

| Molecule ID | Rings | Molecular weight | LogP | TPSA | H-donors | H-acceptors |
|---|---|---|---|---|---|---|
| LT-1074 X 2785 90 (**59**) | 3 | 251.3 | 3.02 | 66.49 | 2 | 1 |
| LT-1167 X 603 (**60**) | 3 | 309.2 | 2.58 | 55.88 | 5 | 0 |
| LT-1167 X 815 (**61**) | 3 | 356.3 | 1.36 | 108.48 | 4 | 1 |
| Chembridge 6376403 (**62**) | 4 | 375.4 | 4.19 | 66.46 | 3 | 1 |
| Chembridge 6656937 (**63**) | 3 | 416.9 | 5.20 | 103.18 | 3 | 1 |
| Chembridge 7948674 (**64**) | 3 | 284.3 | 2.14 | 59.81 | 3 | 2 |
| Chembridge 7967237 (**65**) | 3 | 292.3 | 2.57 | 59.81 | 4 | 1 |
| Chembridge 7971920 (**66**) | 5 | 407.4 | 4.10 | 110.67 | 6 | 3 |
| Chembridge 7988575 (**67**) | 2 | 381.4 | 1.00 | 131.93 | 3 | 1 |
| Chembridge 7993610 (**68**) | 4 | 362.4 | 3.02 | 96.45 | 3 | 1 |
| ASN 06346992 (**69**) | 2 | 264.3 | 1.29 | 91.94 | 3 | 0 |
| ASN 06747799 (**70**) | 4 | 387.4 | 1.68 | 105.37 | 6 | 1 |
| BAS 00243347 (**71**) | 2 | 294.3 | 3.51 | 107.09 | 3 | 1 |
| BAS 02943034 (**72**) | 4 | 432.6 | 5.77 | 58.33 | 0 | 0 |
| BAS 03609804 (**73**) | 3 | 398.4 | 0.95 | 98.73 | 5 | 1 |
| Bionet 10G-910 (**74**) | 2 | 323.7 | 4.08 | 53.02 | 4 | 3 |
| Bionet 8M-521S (**75**) | 4 | 313.4 | 2.37 | 107.56 | 2 | 1 |
| PHAR058776 (**76**) | 2 | 313.1 | 3.49 | 90.37 | 4 | 1 |
| PHAR087402 (**77**) | 4 | 346.4 | 3.64 | 91.08 | 4 | 1 |
| STOCK1S-56176 (**78**) | 3 | 398.9 | 5.45 | 61.53 | 5 | 1 |
| STOCK1S-87136 (**79**) | 2 | 278.3 | 1.54 | 67.98 | 4 | 0 |
| STOCK2S-82643 (**80**) | 2 | 251.3 | -0.35 | 111.08 | 5 | 0 |
| STOCK3S-05798 (**81**) | 2 | 238.3 | -2.47 | 96.43 | 1 | 0 |
| STOCK3S-50234 (**82**) | 2 | 365.4 | 3.07 | 109.28 | 5 | 2 |
| STOCK3S-56652 (**83**) | 3 | 326.4 | 2.59 | 100.29 | 7 | 0 |
| STOCK3S-76708 (**84**) | 2 | 307.3 | 1.21 | 80.10 | 2 | 0 |
| STOCK3S-82963 (**85**) | 4 | 386.4 | 2.10 | 112.30 | 2 | 1 |
| STOCK5S-26863 (**86**) | 2 | 304.3 | -0.01 | 118.64 | 2 | 0 |

**Table 2.10 cont**.: "Hits" purchased from the virtual screening of the MOE database using the less stringent pharmacophore query and objective selection criteria. Molecular descriptors – weight, number of rings, LogP, total polar surface area (TPSA), H-Donors and H-Acceptors – were calculated using the Descriptor feature in MOE.

| Molecule ID | Rings | Molecular weight | LogP | TPSA | H-donors | H-acceptors |
|---|---|---|---|---|---|---|
| STOCK5S-43994 (**87**) | 2 | 275.2 | -1.10 | 114.26 | 4 | 1 |
| STOCK5S-58409 (**88**) | 2 | 261.3 | 1.69 | 88.28 | 3 | 2 |
| STOCK5S-74616 (**89**) | 3 | 425.5 | 1.95 | 116.17 | 3 | 0 |
| STOCK5S-79215 (**90**) | 3 | 413.5 | 1.75 | 106.94 | 4 | 1 |
| STOCK5S-88609 (**91**) | 3 | 359.4 | 2.35 | 114.65 | 3 | 0 |
| STOCK5S-92145 (**92**) | 3 | 343.4 | 2.48 | 105.42 | 6 | 2 |
| STOCK5S-59895 (**93**) | 4 | 389.5 | 0.18 | 101.63 | 5 | 0 |
| STOCK5S-33478 (**94**) | 5 | 395.5 | 3.05 | 74.04 | 4 | 0 |
| F2009-0120 (**95**) | 3 | 371.4 | 1.25 | 104.27 | 4 | 0 |
| F0532-0705 (**96**) | 4 | 388.5 | 4.49 | 61.20 | 3 | 0 |
| F2510-0310 (**97**) | 3 | 346.4 | 1.40 | 78.38 | 5 | 1 |
| AE-848/11422545 (**98**) | 4 | 372.4 | 3.13 | 106.37 | 5 | 1 |
| AO-623/14653070 (**99**) | 2 | 299.3 | 1.25 | 118.08 | 5 | 0 |
| AK-968/15610255 (**100**) | 2 | 345.4 | 3.80 | 82.12 | 7 | 1 |
| AO-080/43378361 (**101**) | 2 | 370.4 | 3.96 | 87.69 | 2 | 1 |
| AQ-405/42300151 (**102**) | 6 | 379.4 | 6.42 | 62.05 | 6 | 1 |
| Peakdale 1001920 (**103**) | 4 | 418.5 | 3.86 | 69.16 | 1 | 0 |
| Peakdale 1002562 (**104**) | 2 | 290.3 | -0.10 | 78.27 | 2 | 0 |
| Peakdale 1003002770 (**105**) | 2 | 235.2 | -0.89 | 114.06 | 5 | 2 |
| HTS 01757 (**106**) | 3 | 383.45 | 3.72 | 82.44 | 6 | 0 |
| HTS 08111 (**107**) | 2 | 292.32 | 1.10 | 93.21 | 5 | 4 |
| KM 07099 (**108**) | 4 | 359.39 | 2.84 | 95.18 | 2 | 1 |
| SEW 00462 (**109**) | 3 | 324.82 | 2.91 | 64.70 | 4 | 0 |
| SPB 03404 (**110**) | 3 | 362.80 | 3.84 | 90.47 | 5 | 0 |
| **AVERAGE** | **3** | **341.74** | **2.39** | **90.95** | **1** | **4** |

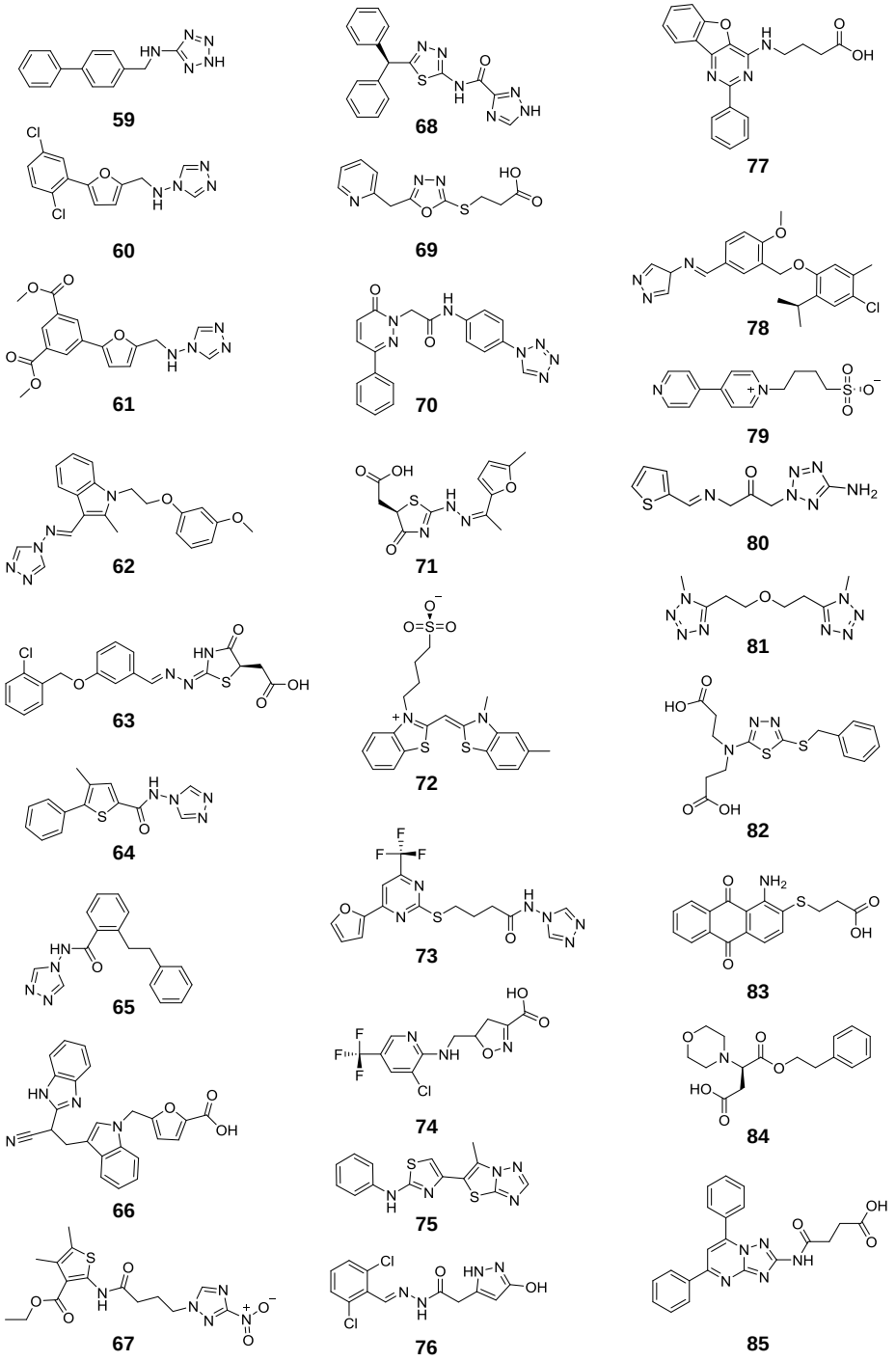



**Figure 2.15**: Structure of "hits" purchased from the virtual screening of the MOE database using the less "stringent" pharmacophore query and objective selection criteria.

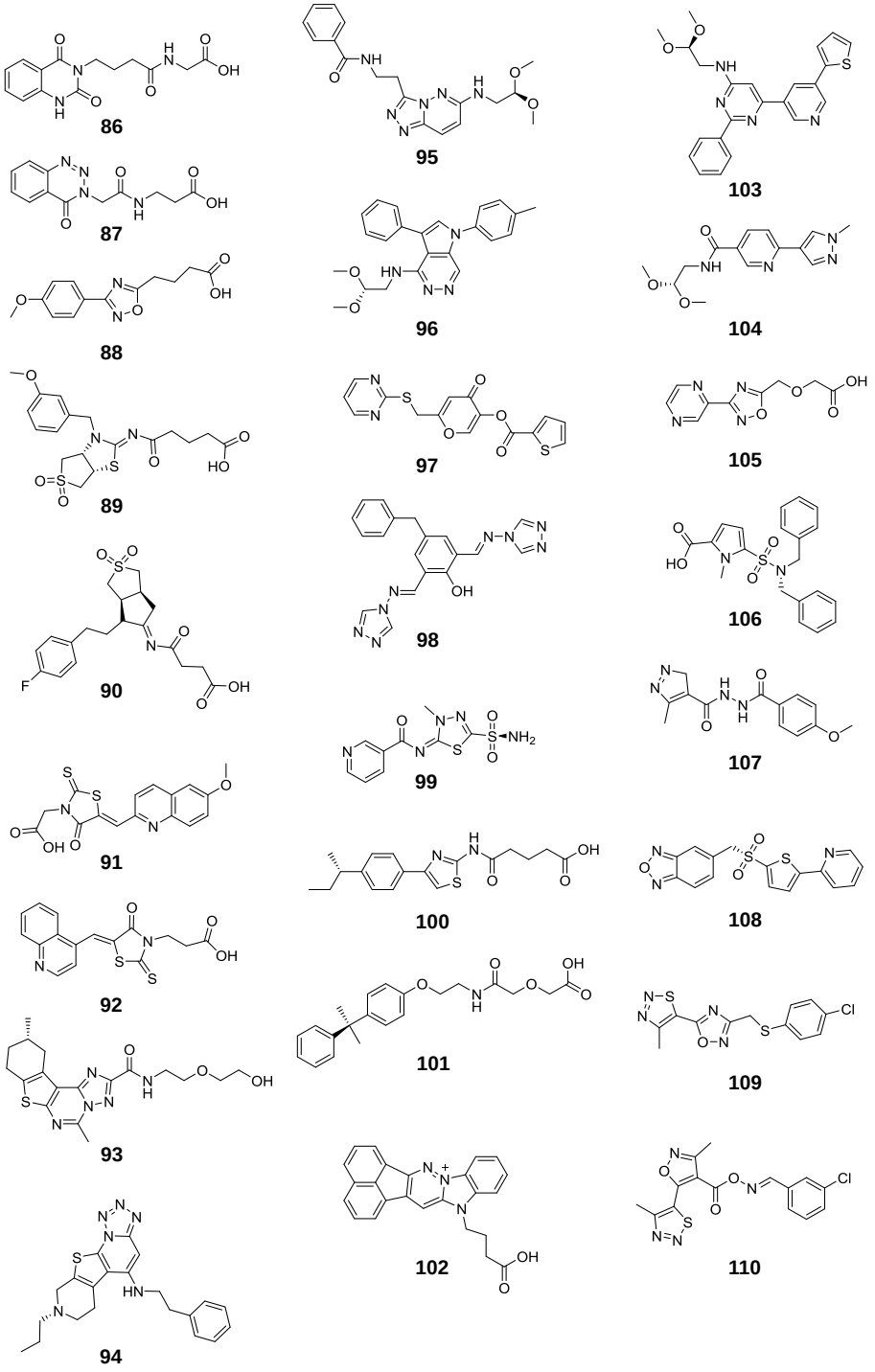



**Figure 2.15** cont.: Structure of "hits" purchased from the virtual screening of the MOE database using the less stringent pharmacophore query and objective selection criteria.

#### 2.3.4.2 **Rescoring "hits" generated from virtual screening studies in "native" S100P**

The NMR ensemble of S100P which was used in this project as a template for structure-based drug design studies has three mutations: T6A, C85S and A92T. To assess the impact of these mutations on the "hits" obtained from the virtual screening studies, the 52 purchased "hits" were rescored in the "native" conformer 15 of 1OZO and their *S* score compared to their original *S* score. Except for three "hits", compounds **96**, **99** and **104**, all the other "hits" returned negative *S* scores similar to those obtained in the original 1OZO conformer 15 (Figure 2.16). The positive *S* score for these three "hits" could be indicative of unfavourable interaction with the native protein. It is possible that there may be some penalties – i.e. high binding energy – for potential clashes with the protein surface, although there is no evidence of such clashes from observation of the compounds' interaction with the protein (Figure 2.17). In comparison, "hits" like compounds **74** and **80** seemed to show favourable interaction with the native protein than the mutated conformer 15 of the NMR ensemble. While the pharmacophore was a useful constraint in the virtual screening, the interaction of the protein and these "hits" goes beyond the three pharmacophoric points. A correlation between the predicted binding affinity and the actual activity from biological screening can only be made once all the compounds are screened and these two factors are compared.

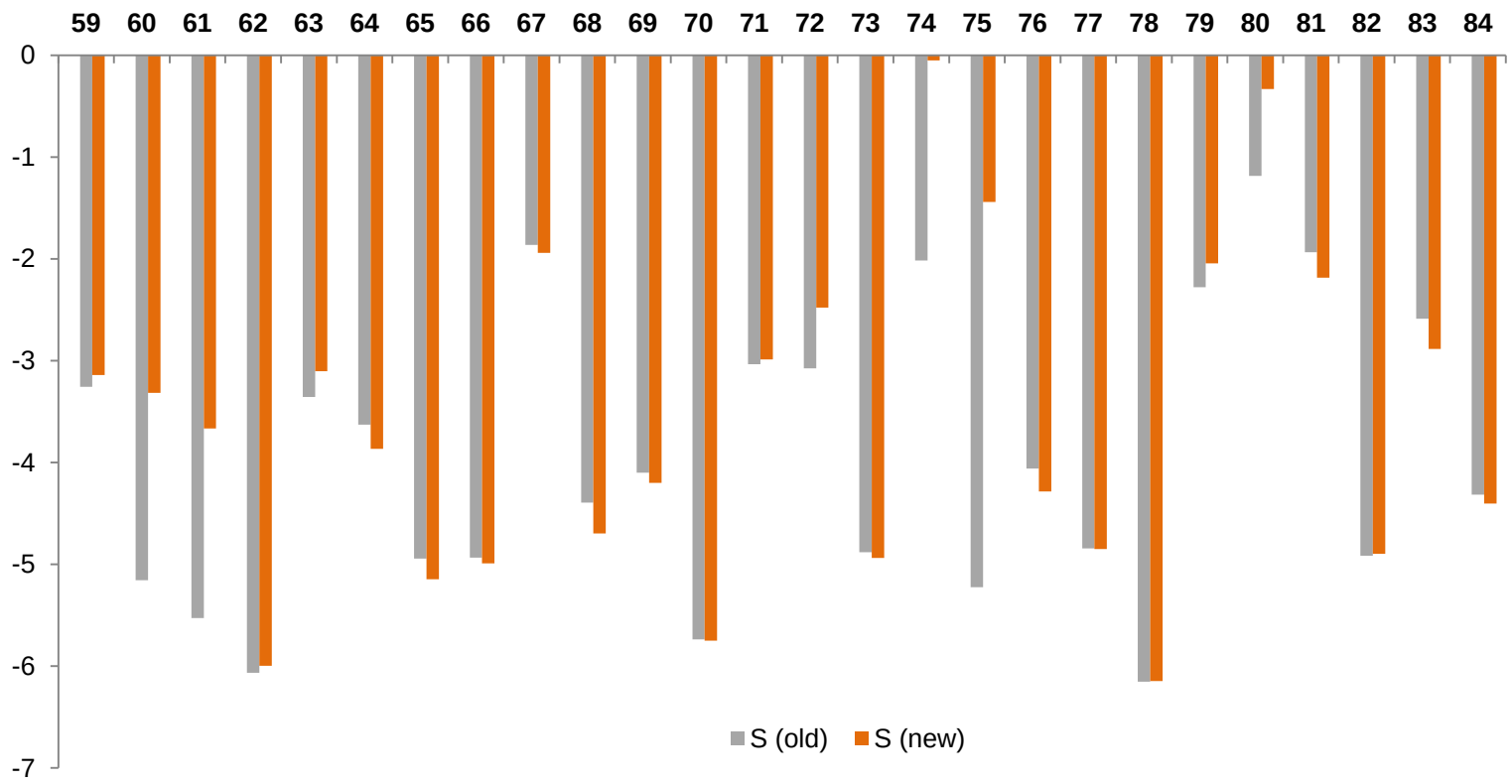



**Figure 2.16**: Comparison of *S* score of the 52 purchased "hits" obtained from using the original S100P NMR conformer 15 (old) and the *S* score obtained from using the "native" S100P (new).

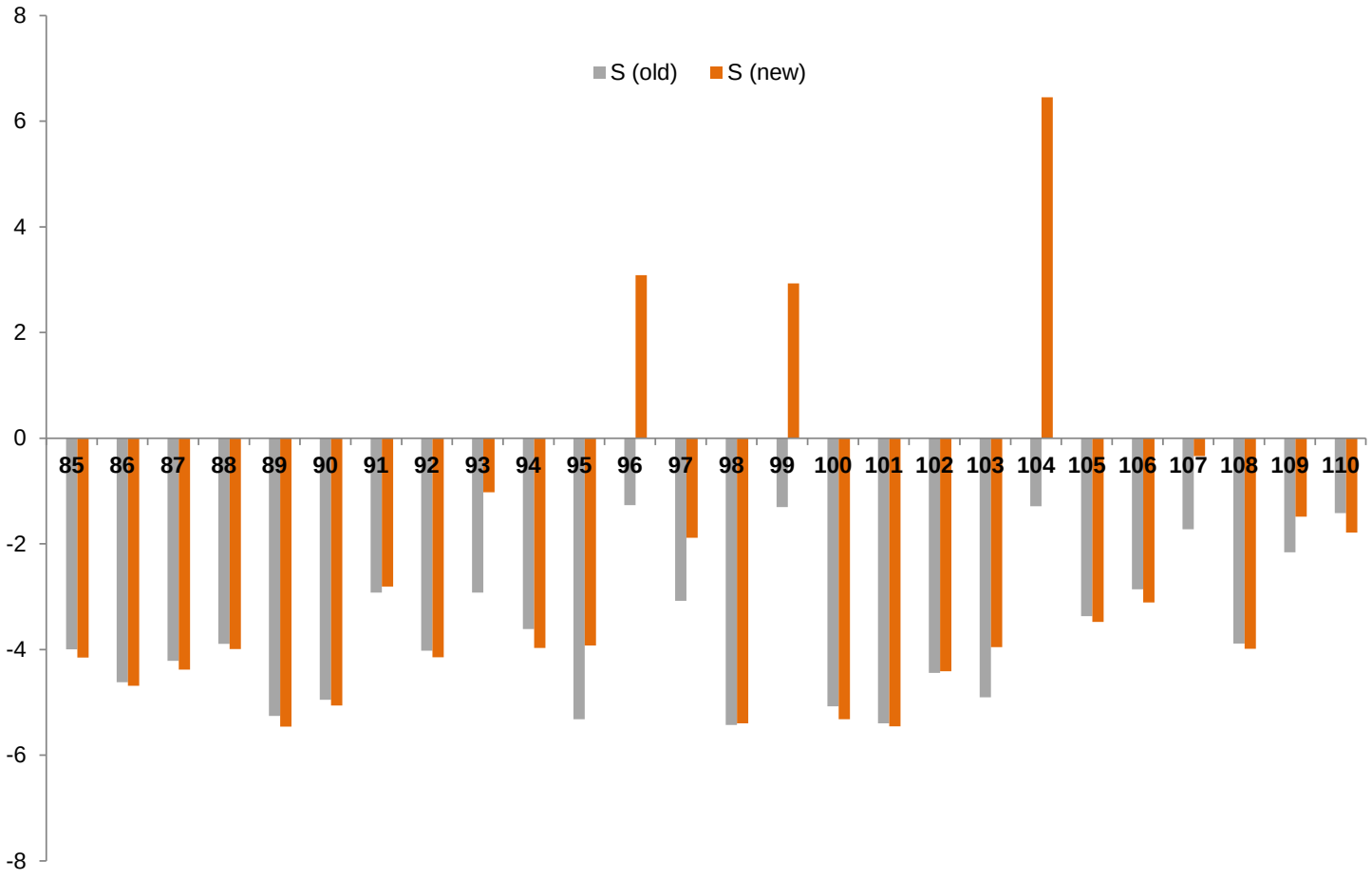



**Figure 2.16** cont.: Comparison of *S* score of the 52 purchased "hits" obtained from using the original S100P NMR conformer 15 (old) and the *S* score obtained from using the "native" S100P (new).

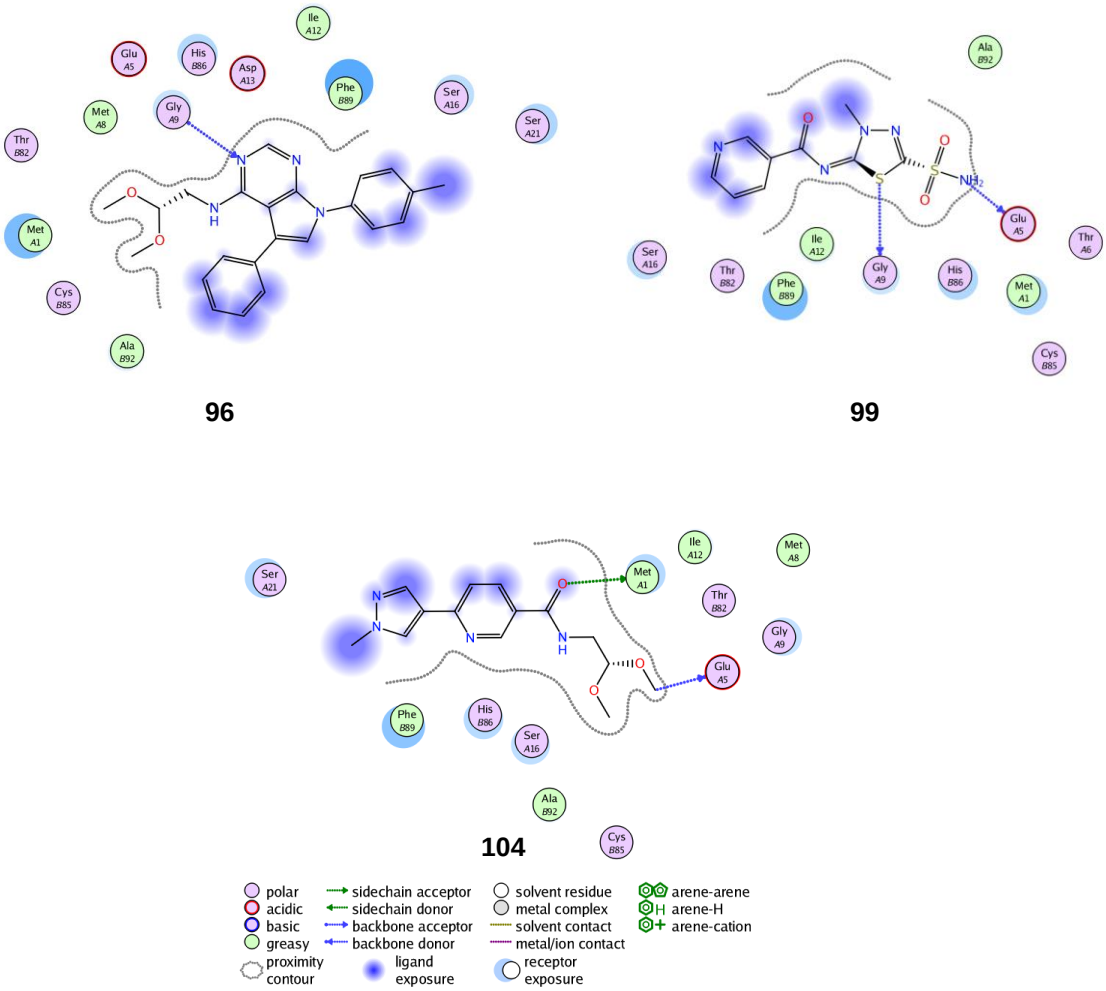



**Figure 2.17**: Predicted binding interaction between compounds **96**, **99** and **104** with "native" S100P. All three "hits" had positive *S* score when rescored in the "native" protein.

## 2.4 CONCLUSION

Publically available 3D experimental structures of S100P, 1J55 and 1OZO were analysed for suitability for use in structure-based drug design (SBDD) studies. The NMR ensemble 1OZO was selected over the X-ray crystal structure because it represented the biologically active dimeric form of S100P compared to the X-ray crystal structure. In addition, all the residues were resolved albeit with three mutations.

Using computational methods, it was found that only one conformer (conformer 15) of the NMR ensemble of S100P was suitable for use as template for SBDD studies. Both geometric- and energy-based pocket-detection algorithms identified potential cavities at the dimeric interface of this conformer adequate to accommodate cromolyn during docking studies. Using observed interactions between ligand and protein as a template, a pharmacophore query was designed and used as a constraint for subsequent virtual screening studies. Seventeen compounds from the generated "hits" form a diverse collection of chemical compounds that will be screened against pancreatic cancer cells (Chapter 3) in the hope of identifying a potential lead therapeutic candidate against this lethal cancer.

Despite the mutations present in the NMR structures, there was little difference between the predicted binding affinities of "hits" with the "mutated" S100P NMR conformer and the "native" protein, again confirming the usefulness of the NMR ensemble as a template for *in silico* drug design studies.

# CHAPTER 3: Synthesis and Biological Evaluation of "Hits" from Virtual Screening Studies

# 3. PREFACE

Having identified virtual screening "hits" that have the potential to interact with the S100P dimeric interface (Chapter 2, Section 2.3.4), two compounds – **7** and **58** (Figure 3.1) – were synthesised, characterised and subjected to biological screening against the pancreatic cancer cell lines BxPC-3 and Panc-1. This chapter details the syntheses and biological evaluation of these two hit compounds. In addition to the two synthesised compounds, the biological evaluation of 13 commercially available compounds identified during the virtual screen (Figure 2.14) is also reported in this chapter.

Biological screening of the "hits" was carried out in collaboration with Dr Tatjana Crnogorac-Jurcevic at Barts Cancer Centre, Queen Mary, University of London. The biological screening, and the initial evaluation of the data, was carried out by Nasir Mahmoud [328]; interpretation of the results was carried out by the author.

The results from the syntheses and biological screening are discussed in Sections 3.2 and 3.3 respectively. The experimental details of the syntheses are given at the end of this chapter in Section 3.5.

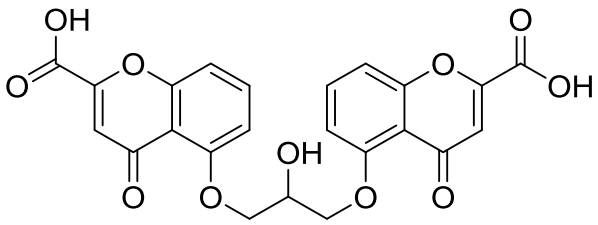
## 3.1 INTRODUCTION

The synthesis of two compounds identified by the virtual screening studies in Chapter 2, that were prohibitively expensive (**7**) and unavailable for purchase (**58**) are described below. Compound **7** is a dibenzyl derivative while **58** is a substituted benzoic acid (benzoxazin-2-yl)ethyl ester. In addition, two analogues of **58** (**58a** and **58b**, Figure 3.1) were also synthesised.

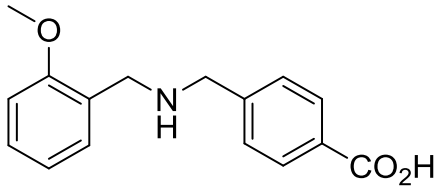
In total, the syntheses of the above four compounds are discussed in this chapter. Viable synthetic routes and identification of readily available starting materials (RASMs) were identified using retrosynthetic disconnection approaches [329, 330]. Once readily available starting materials were determined, published synthetic routes were identified using Reaxys®, (version 1.7.8; www.elsevier.com/reaxys), which were followed and adapted where applicable. Where there were no published records for synthetic routes, logical synthetic routes were designed and implemented.

The four synthesised compounds were combined with the 13 "hits" purchased from the virtual screening studies (Chapter 2), to give 17 compounds in total. These compounds are structurally and/or chemically dissimilar to cromolyn, the compound with limited bioavailability previously reported to inhibit S100P binding to RAGE [195]. Hence, if any of the compounds demonstrate inhibition of cell proliferation and/or cell invasion in the biological assays they will provide novel starting points for drug discovery investigations.

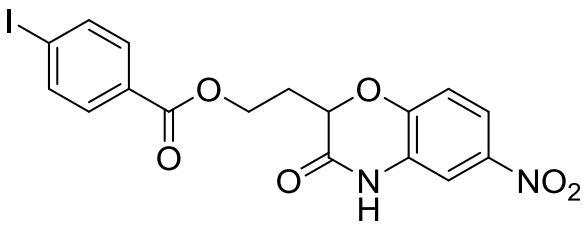
The 17 compounds were assessed for biological activity in separate screening studies that examined cell proliferation and cell invasion in two pancreatic cancer cells – S100P-expressing BxPC-3 and Panc-1 which does not produce S100P.



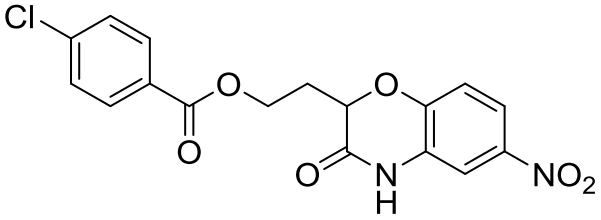
**Cromolyn (4)**



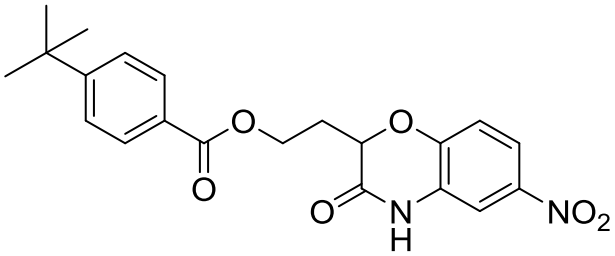
**7**



**58**



**58a**



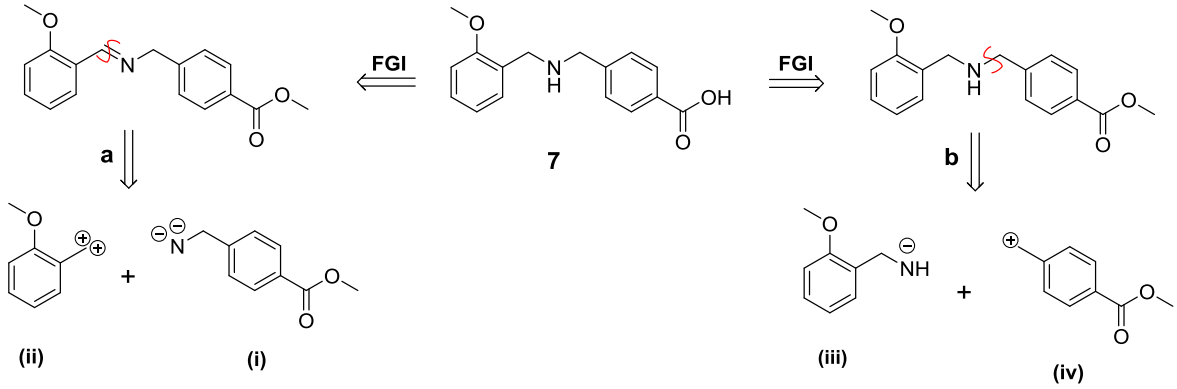
**58b**

**Figure 3.1**: Structure of cromolyn (**4**) and the two "hits" from the virtual screening studies (**7** and **58**) together with the analogues (**58a** and **58b**) of compound **58**. The known inhibitor of the S100P-RAGE interaction, cromolyn, is shown for comparison purposes.

## 3.2 RESULTS AND DISCUSSION

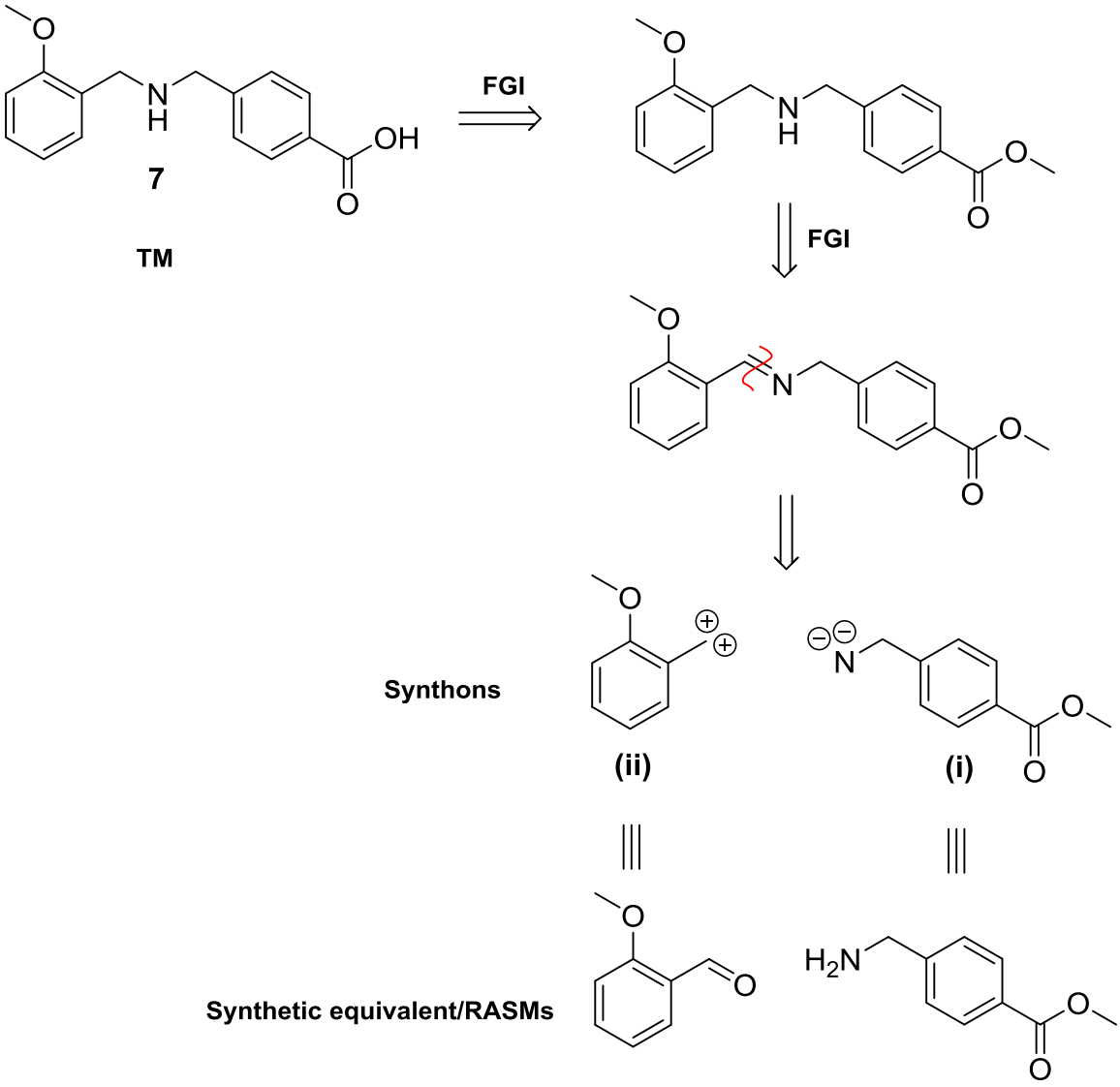
### 3.2.1 Synthesis of 4-({[(2-methoxyphenyl)methyl]amino}methyl)benzoic acid: compound 7

Retrosynthetic analysis was applied to identify potential precursors to compound **7**. Following functional group conversions (FGI) of the amino linker to the imine, and the carboxylic acid group to the ester, the C=N or the N-C bond could be disconnected in the direction of either **a** or **b** (Scheme 3.1).



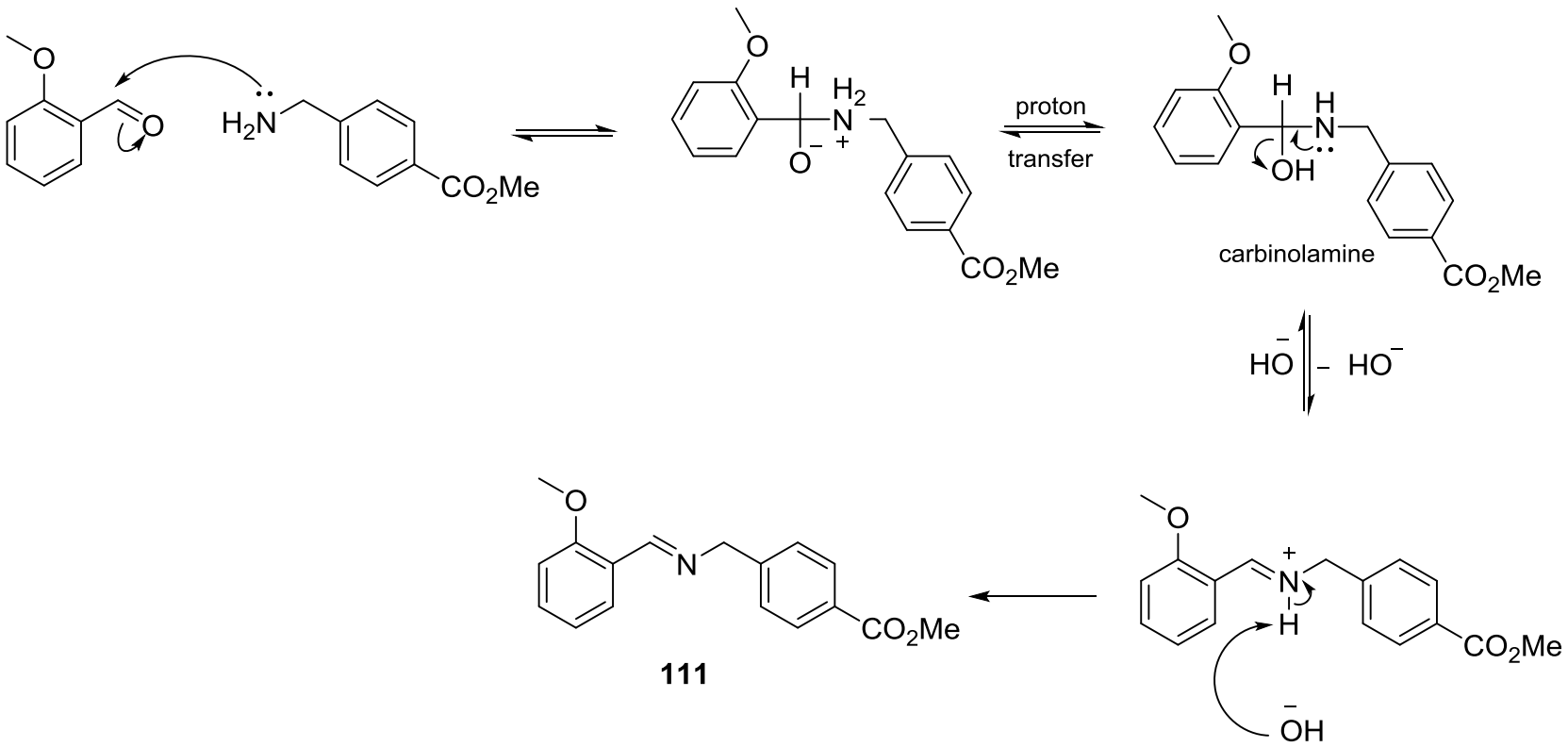

**Scheme 3.1**. Possible disconnections on compound **7** to identify starting materials.

The respective synthetic equivalents for synthons (**i**) and (**iii**), and (**ii**) and (**iv**) are an amine and an aldehyde respectively. The first attempt at synthesising the target molecule **7** was carried out via route **a**. The corresponding synthetic equivalents for this disconnection were 4-(aminomethyl)benzoate for (**i**), after functional group interconversion and 2-methoxybenzaldehyde for (**ii**) (Scheme 3.2).




**Scheme 3.2**: Retrosynthetic analysis on compound **7** to identify simple precursor molecules via route a. (TM: target molecule, FGI: functional group interconversion, RASMs: readily available starting materials).

2-Methoxybenzaldehyde was condensed with 4-(aminomethyl)benzoate, the corresponding methyl ester of 4-(aminomethyl)benzoic acid via a nucleophilic addition reaction at the carbonyl centre, with the subsequent elimination of a water molecule. The nucleophilic nitrogen of the amine group of 4-(aminomethyl)benzoate attacks the electrophilic carbon of the aldehyde to form a carbinolamine intermediate (Scheme 3.3) [331]. Dehydration of this intermediate generates the imine **111** as a pale yellow oil in 95% yield. As this was a slow and reversible reaction, anhydrous $MgSO_4$ was added to "mop up" the water by-product formed thus driving the synthesis in the direction of the required product. The structure of the imine was confirmed by the presence of the imine proton as a singlet downfield at $\delta$ 8.8 ppm in the $^1$H-NMR (Section 3.5.2).

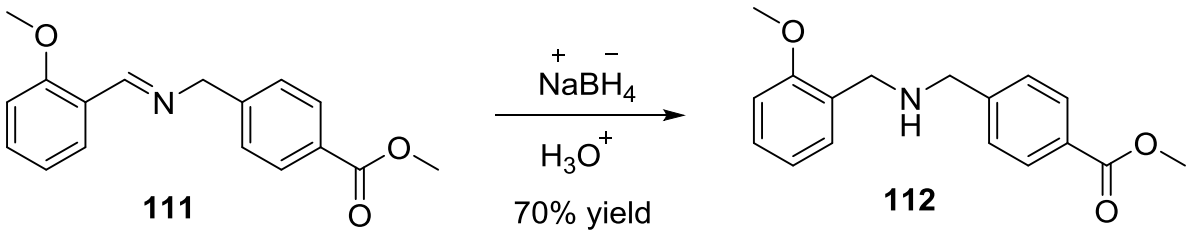



**Scheme 3.3**: Mechanism for the formation of the imine **111** via a carbinolamine intermediate.

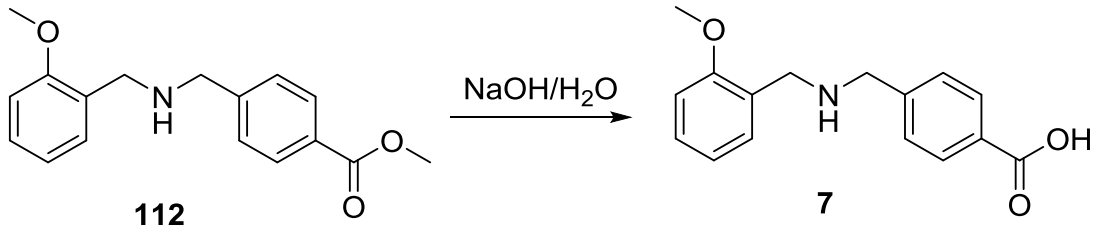
The isolated imine **111** was reduced using sodium borohydride to give the corresponding amine **112** as pale yellowish oil in 70% crude yield (Scheme 3.4). The amine was confirmed on proton NMR by the absence of the imine proton at δ 8.8 ppm and the presence of both pairs of methylene protons on either side of the amine group as singlets at δ 3.74 and 3.61 ppm (Section 3.5.2.1). Compound **112** could be synthesised from **6** in a one-pot reductive amination reaction where the aldehyde is condensed with the amine to form the imine, which is subsequently reduced using a reducing agent. However, the use of sodium borohydride, the reducing agent used here, is unsuitable in this case as it would have also reduced the aldehyde to its corresponding alcohol.

The final step in the synthesis of the target compound **7** involved functional group interconversion; the ester group of **112** was hydrolysed under basic conditions to give the desired acid **7** as a pale yellow solid with a yield of 71% from the final reaction step (Scheme 3.5 and 3.6). The absence of the carbon peak and the three methyl proton peaks of the ester from both $^{13}C$ and $^{1}H$-NMR spectra respectively, and the presence of a broad OH peak at 3368 $cm^{-1}$ confirmed the formation of the acid product (Section 3.5.2.2).

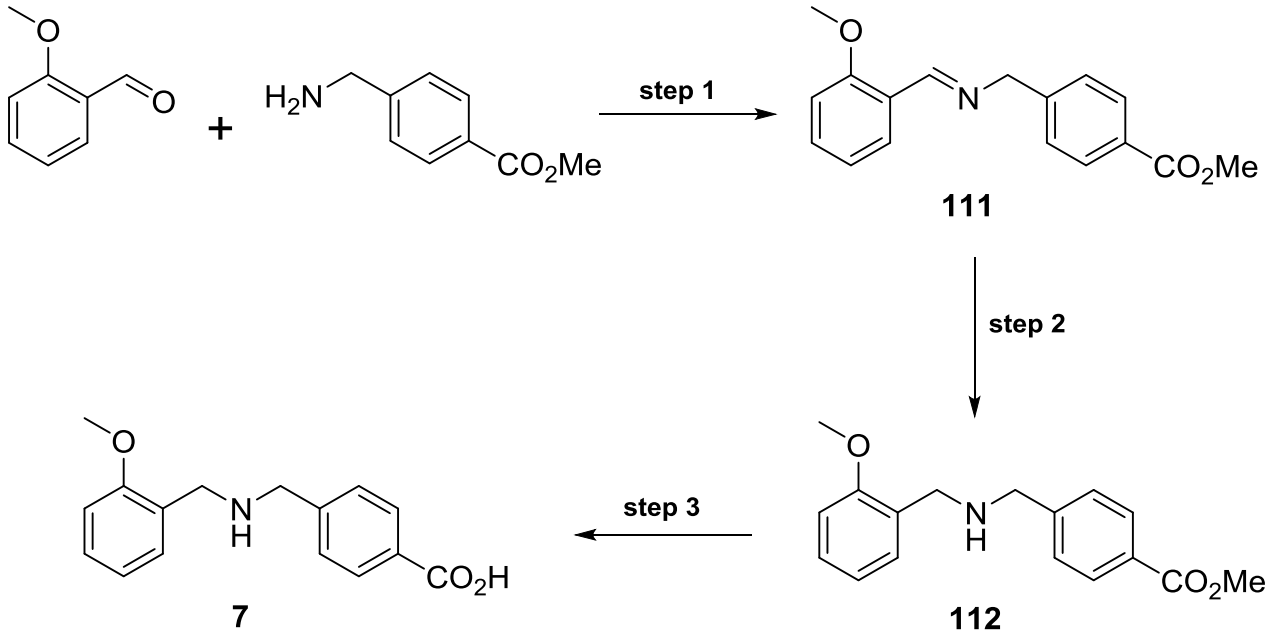
The second route **b** (Scheme 3.1) was not employed as route **a** succeeded in generating the target compound **7**.




**Scheme 3.4**: Reduction of the imine **111** to the corresponding amine **112**.




**Scheme 3.5**: Base hydrolysis of compound **112** to give the required acid **7**.

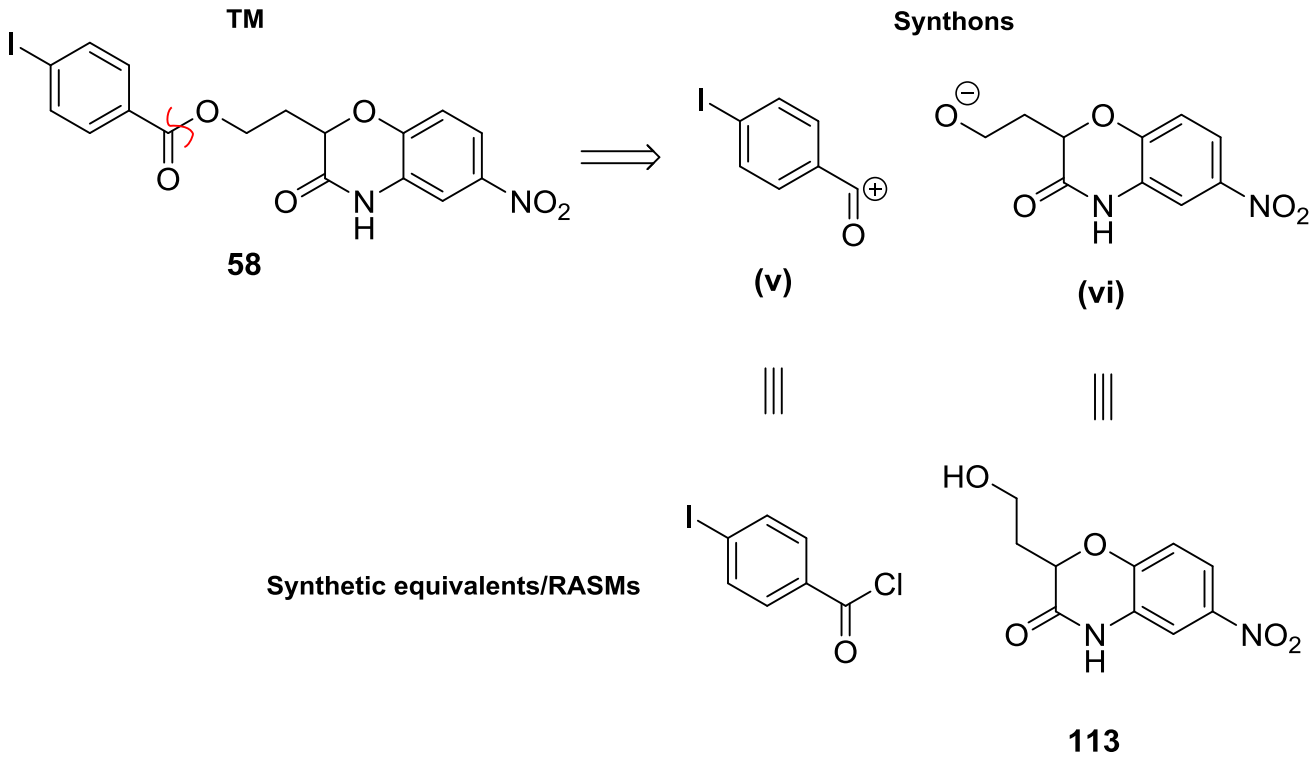



**Scheme 3.6**: Summary of steps involved in the synthesis of compound **7**.

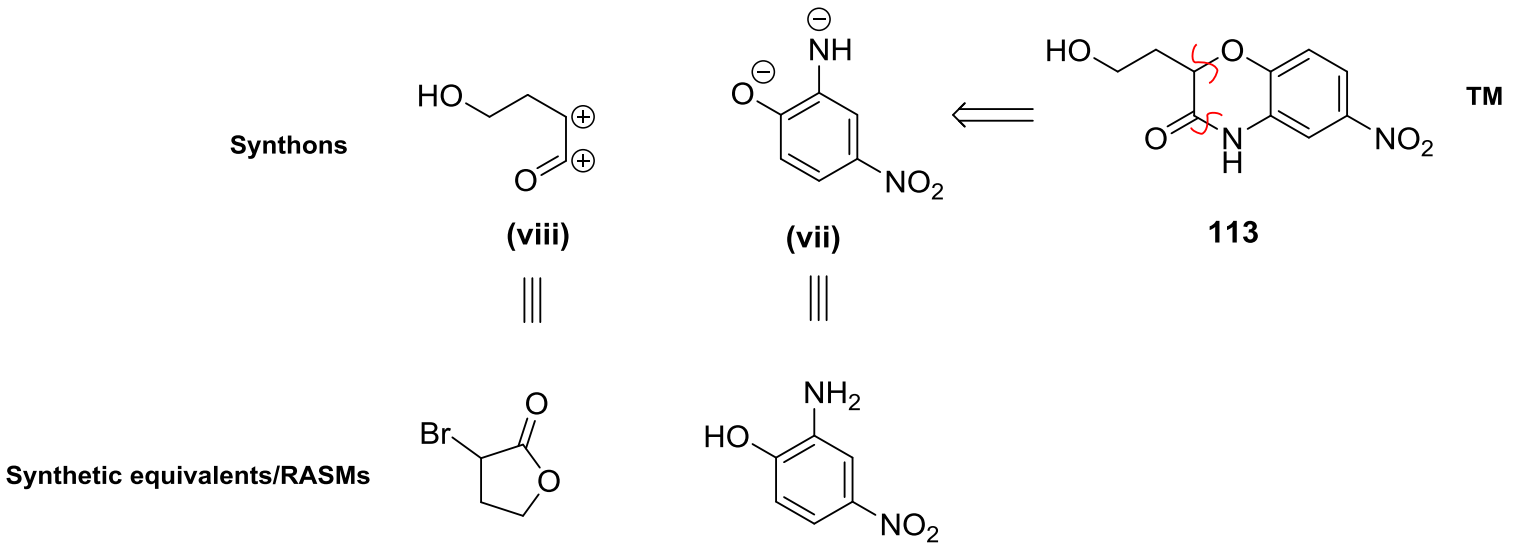
### 3.2.2 Synthesis of 2-(6-nitro-3-oxo-3,4-dihydro-2*H*-1,4-benzoxazin-2-yl)ethyl 4-iodobenzoate: compound 58 and analogues

Retrosynthetic analysis of 2-(6-nitro-3-oxo-3,4-dihydro-2*H*-1,4-benzoxazin-2-yl)ethyl 4-iodobenzoate (**58**) identified two main synthons (**v**) and (**vi**) from the disconnection of the ester bond (Scheme 3.7). The synthetic equivalents for (**v**) and (**vi**) were an acyl chloride and an alcohol respectively, which react in an addition elimination resulting in a substitution reaction to form the ester compound **58**. Further retrosynthetic analysis on the alcohol **113** identified two additional bonds that can be logically disconnected; the amide bond between the carbonyl group and the amine, and the bond between C8 and the ether oxygen. A suitable synthetic equivalent for synthon (**vii**) is found in 2-amino-4-nitrophenol. The synthetic equivalent for the remaining synthon (**viii**) with the two electropositive carbon centres is α-bromo-γ-butyrolactone (Scheme 3.8). The presence of a carbonyl group and a bromine atom in this compound ensures that the carbons directly attached to them are electron deficient, thus providing an appropriate synthetic equivalent.

Having identified these starting materials, it was simple to put together a synthetic route to get to the target molecule i.e. synthesis of the alcohol (**113**) before reacting this with the acyl chloride to obtain the target molecule **58**. The use of the acyl chloride instead of the corresponding carboxylic acid, in the second step proceeds via an irreversible reaction. This is advantageous in driving the synthesis to completion [332].

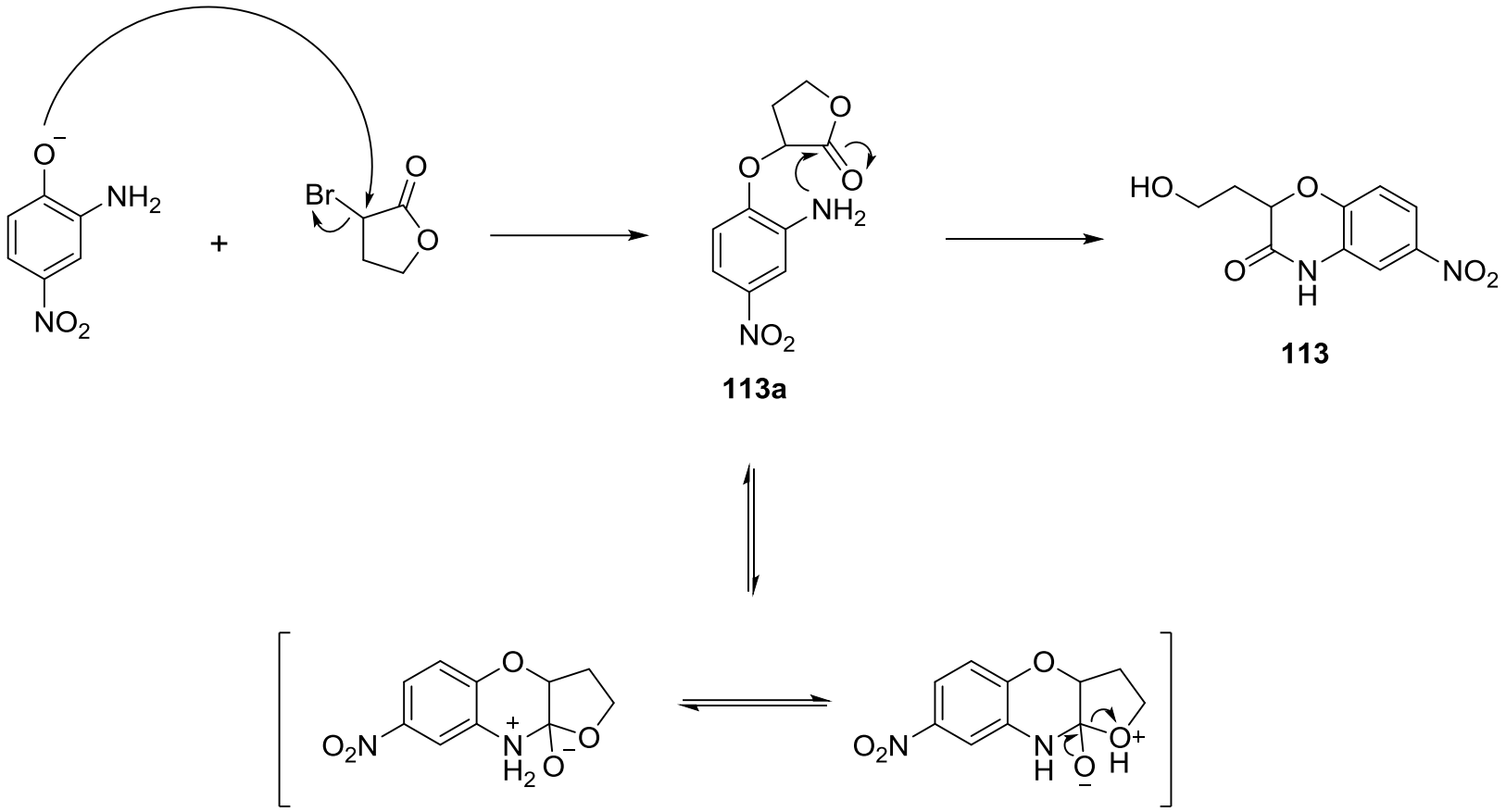



**Scheme 3.7**: Retrosynthetic analysis on compound **58** to identify readily available starting materials. Compound **113** was synthesised first prior to synthesising the target molecule.




**Scheme 3.8**: Retrosynthetic analysis on compound **113** to identify readily available starting materials (RASMs).

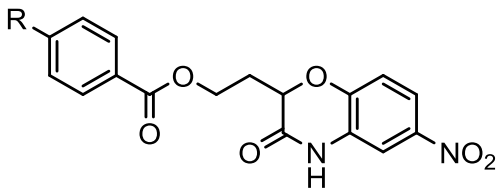
The alcohol **113** (Scheme 3.8) was synthesised by adapting the method of Frechette and Beach [333]. α-Bromo-γ-butyrolactone was reacted with 2-amino-4-nitrophenol in anhydrous tetrahydrofuran to afford **113** in 72% crude yield as a brick-red solid after workup. The reaction was carried out under reflux conditions, and is believed to proceed via attack of the nucleophilic oxygen of the 2-amino-4-nitrophenol on the electron deficient α-carbon centre of the carbonyl compound, formed by the departing bromine [333]. Close proximity of the carbonyl and aromatic amine groups leads to formation of the intermediate **113a** (Scheme 3.9) with subsequent ring-opening of the tetrahydrofuran ring giving the benzoxazine crude product as a brick-red solid in 72% yield. Recrystallisation from ethyl acetate gave **113** as yellow fluffy crystals with a melting point of 169-170 °C and a final yield of 56%. The structure of the product was confirmed by $^{1}$H-NMR by the presence of the chiral C-H proton of the morpholinone ring at δ 4.85 which is absent in either of the starting materials (Section 3.5.3).




**Scheme 3.9**: Synthesis of compound **113**. (Adapted from Frechette and Beach [333]).

To obtain the required compound **58**, the intermediate **113** was reacted with 4-iodobenzoyl chloride in an acylation reaction to form the ester. This reaction was initially attempted in dichloromethane (DCM) but solubility issues with the alcohol resulted in incomplete reactions with no product isolated. When tetrahydrofuran (THF) was substituted for DCM, the reaction proceeded to completion but the crude product contained a mixture of the acid chloride, the required ester product and impurities. The ester product was recovered via column purification, in a final yield of 8%. The chloro and *tert*-butyl analogues (**58a** and **58b** respectively) were obtained in a similar manner, with yields of 5 and 36% respectively after column purification (Table 3.1). The crude products of both **58** and **58a** were dark sticky brown and cream solids respectively, whereas **58b** was a brown oil. Column chromatography using diethyl ether as eluent afforded **58**, **58a** and **58b** as a pale yellow, a greenish-yellow and a cream solid respectively with narrow melting points ranges (Table 3.1). All products were characterised by $^{1}$H- and $^{13}$C-NMR, IR and LC-MS (Sections 3.5.4, 3.5.4.1 and 3.5.4.2). All three compounds exhibited the characteristic C=O stretch of the ester functional group at around 1700 $cm^{-1}$ in addition to the C=O stretch of the amide around 1680 $cm^{-1}$ in infra-red spectra (Sections 3.5.4, 3.5.4.1, and 3.5.4.2).

**Table 3.1**: Yields and melting points of compound **58** and analogues (**58a** and **58b**).




| Compound | R | Yield (%) | Melting point (°C) |
|---|---|---|---|
| **58** | I | 8 | 188-190 |
| **58a** | Cl | 5 | 218-220 |
| **58b** | *t*-Bu | 36 | 88-90 |

## 3.3 BIOLOGICAL EVALUATION OF "HITS"

Purchased compounds and those synthesised in-house were screened for activity against BxPC-3 (ATCC® CRL-1687™) and Panc-1 (ATCC® CRL-1469™) human pancreatic cancer cells (American Type Culture Collection, Teddington, Middlesex, UK). BxPC-3 cells have been reported to express endogenous S100P [121, 177, 334], the therapeutic target protein for which this research aims to design inhibitors against its interaction with RAGE. Alternatively, Panc-1 cells are non-S100P expressing cells [177, 334], making them an ideal S100P-negative control. The use of an S100P-negative control should help explain if any observed activity against these cells is S100P-related or not. The presence or absence of endogenous S100P in both cell types has been previously reported [121, 177, 192, 290].

Proliferation studies were carried out using the MTS (3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium) colorimetric assay. MTS is based on the presence of mitochondrial dehydrogenase enzymes in metabolically active viable cells that reduce the salt to a soluble coloured formazan dye product which is then read spectrophotometrically [335]. Although specifically designed to measure metabolic activity, results from an MTS assay are sometimes correctly or incorrectly used to indirectly infer the proliferative activity of cells, as less active cells are more likely to be less proliferative. The assay is also used loosely to describe cell viability, although it has been argued that a multiplex of different assays that also measure other properties such as cytotoxicity from the same experimental well give an overall picture of cell viability than a single assay [336, 337].

### 3.3.1 Effect of cromolyn (4), siS100P on proliferation and invasion of pancreatic cancer cells

Prior to testing the anti-proliferative effect of the screening compounds on pancreatic cancer cells, it was found that silencing the S100P gene in BxPC-3 cells has little to no effect on their proliferation (Figure 3.2, [328]).

This result is however in contrast to findings published by Arumugam *et al*. [177] who reported that S100P stimulates pancreatic cancer cell proliferation and survival. Their work showed a correlation between high levels of S100P and pancreatic cancer cell proliferation, survival, migration, and invasion in both *in vitro* and *in vivo* models. The effect of knock-down S100P on the proliferation of BxPC-3 is marginal (Figure 3.2) after 72 h compared to control, and non-existent at 24 and 48 h [328], implying that proliferation of these cells may be independent of S100P.

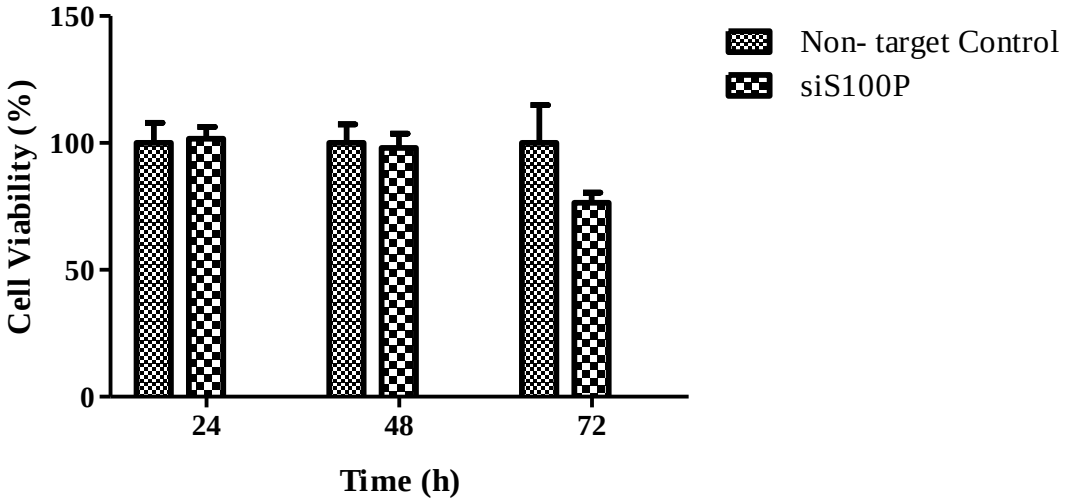



**Figure 3.2**: Effect of knock-down S100P in BxPC-3 cell proliferation. BxPC-3 cells were stably transfected with non-target control siRNA or siS100P. Cells were plated at an equal density and cultured from 24-74 h before cell numbers were determined using the MTS assay. Results are expressed as mean + SEM. Cell viability was expressed as a percentage for each treatment group relative to the negative control (1% v/v DMSO in culture media) according to the following equation: Growth (%) = $OD_{490}$(sample)/$OD_{490}$(control) x 100. P-value was determined relative to control using un-paired t-test. None of the differences were significant.

Interestingly, cromolyn (**4**), the small anti-allergy molecule that has been reported to bind to S100P and inhibit its interaction with RAGE [195, 334], was also found to have little effect on the proliferation of BxPC-3 cells at concentrations up to 1 mM [328]. An effect is only observed at higher concentrations of 2 mM, although this was not statistically significant (Figure 3.3). On its own, cromolyn (**4**) does not exert a significant effect on the proliferation of these cells. An increase in anti-proliferative activity is only observed in combination with gemcitabine [195], the current therapeutic agent used in the treatment of pancreatic cancer.

Knocked-down S100P was found to significantly reduce invasion of BxPC-3 cells compared to control (Figure 3.4). This effect is in agreement with published data that indicate that pancreatic, breast, colon and prostate cancer metastasis is mediated via an S100P-related mechanism [177, 181, 182, 226, 289, 338].

No anti-proliferative or anti-invasive effects were observed in Panc-1 cells following treatment with cromolyn (data not shown).

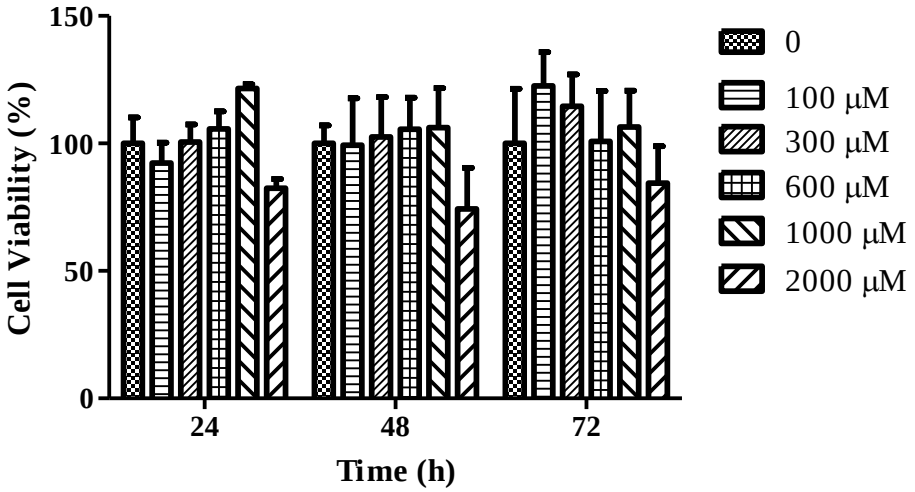
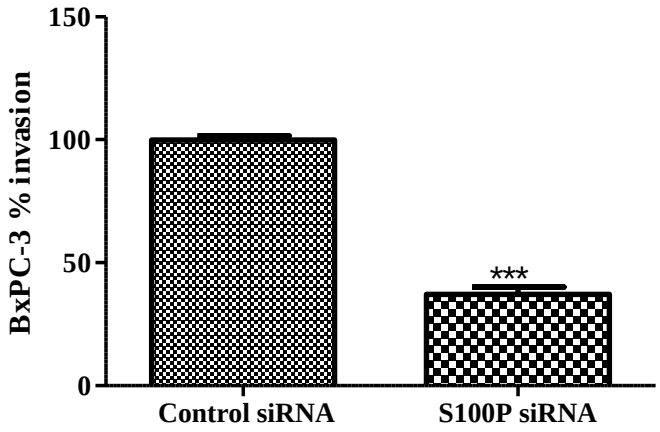
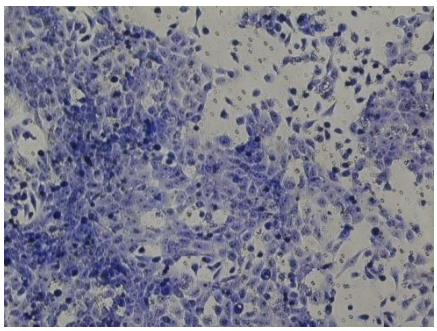



**Figure 3.3**: Anti-proliferative effect of cromolyn (**4**) on BxPC-3 cells. Cells (1.0 x $10^3$ cells/well) were cultured in the presence or absence of cromolyn and cell proliferation was analysed using the MTS assay after 24, 48 and 72 h. Results are expressed as mean + SEM. Cell viability was expressed as a percentage for each treatment group relative to the negative control (1% v/v DMSO in culture media) according to the following equation: Growth (%) = $OD_{490}$(sample)/$OD_{490}$(control) x 100. None of the differences were significant [328].

**A**




**B**



**C**

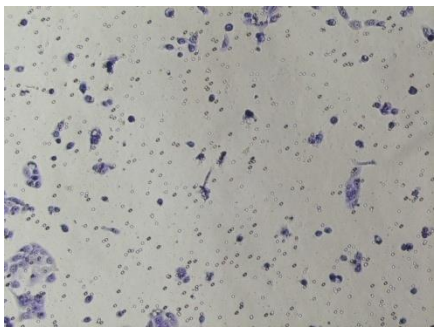


**Figure 3.4**: Effect of S100P on BxPC-3 invasion. A) BxPC-3 cells stably transfected with control siRNA or S100P siRNA were placed in serum-free culture media and added into the upper compartment of an invasion chamber. After 48 h, cells in the upper chamber were removed and cells that had invaded onto the lower surface of the membrane were stained using Giemsa stain. Cells in three different areas were counted for invasion studies. Results are expressed as mean + SEM (***$p$<0.001). B) Non-target control siRNA. C) siS100P. Photographs of representative membranes for cells after Giemsa staining viewed at x10 objective lens [328].

### 3.3.2 Effect of screening compounds on proliferation and invasion of pancreatic cancer cells

Of the 17 compounds examined, seven compounds (Table 3.2) were investigated for effect on both anti-proliferative (BxPC-3 and Panc-1) and anti-invasive (BxPC-3) properties. As no invasion was observed for Panc-1 cells, the compounds were not screened for anti-invasive effects on these cells. A summary of the data for the compounds investigated on BxPC-3 cell line for anti-invasive effects, and on both Panc-1 and BxPC-3 cell lines for anti-proliferative effects is presented in Table 3.2.

All compounds were tested for anti-invasion activity against BxPC-3. However, four compounds (**7**, **58**, **58a** and **58b)**, were not tested for anti-proliferative activity on the same cell lines, while seven (compounds **7**, **16**, **22**, **26**, **38, 39**, **40**) were not tested on Panc-1 cells for the same activity (Table 3.2). Of the seven compounds screened for both anti-invasive and anti-proliferative properties, five compounds, **17**, **18**, **20**, **24**, **43**, showed activity against invasion of BxPC-3 cells (Figure 3.5). At 100 μM, both **15** and **24** had a significant effect relative to their respective controls ($p<0.05$, Figure 3.5). Interestingly, compound **24** had an effect at 100 μM that is comparable to cromolyn (**4**) at the same concentration (Figure 3.5D). At 500 μM all five compounds showed a considerable anti-invasive effect ($p<0.001$) compared to controls. This effect however may be attributed to the high concentration of these compounds which could be detrimental to the survival of the cells.

**Table 3.2**: Screening compounds tested against the proliferation and invasion of pancreatic cancer.

| | **Invasion** | **Proliferation** | |
|---|---|---|---|
| **Compound** | **BxPC-3** | **BxPC-3** | **Panc-1** |
| **7** | Tested | ND[1] | ND[1] |
| **16** | Tested | Tested | ND[1] |
| **17** | Tested | Tested | Tested |
| **18** | Tested | Tested | Tested |
| **20** | Tested | Tested | Tested |
| **22** | Tested | Tested | ND[1] |
| **23** | Tested | Tested | Tested |
| **24** | Tested | Tested | Tested |
| **26** | Tested | Tested | ND[1] |
| **36** | Tested | Tested | Tested |
| **38** | Tested | Tested | ND[1] |
| **39** | Tested | Tested | ND[1] |
| **40** | Tested | Tested | ND[1] |
| **43** | Tested | Tested | Tested |
| **58** | Tested | ND[1] | Tested |
| **58a** | Tested | ND[1] | Tested |
| **58b** | Tested | ND[1] | Tested |

[1]ND: Not determined. Due to time constraints, some compounds were not tested on either or both cells for anti-proliferative activity.

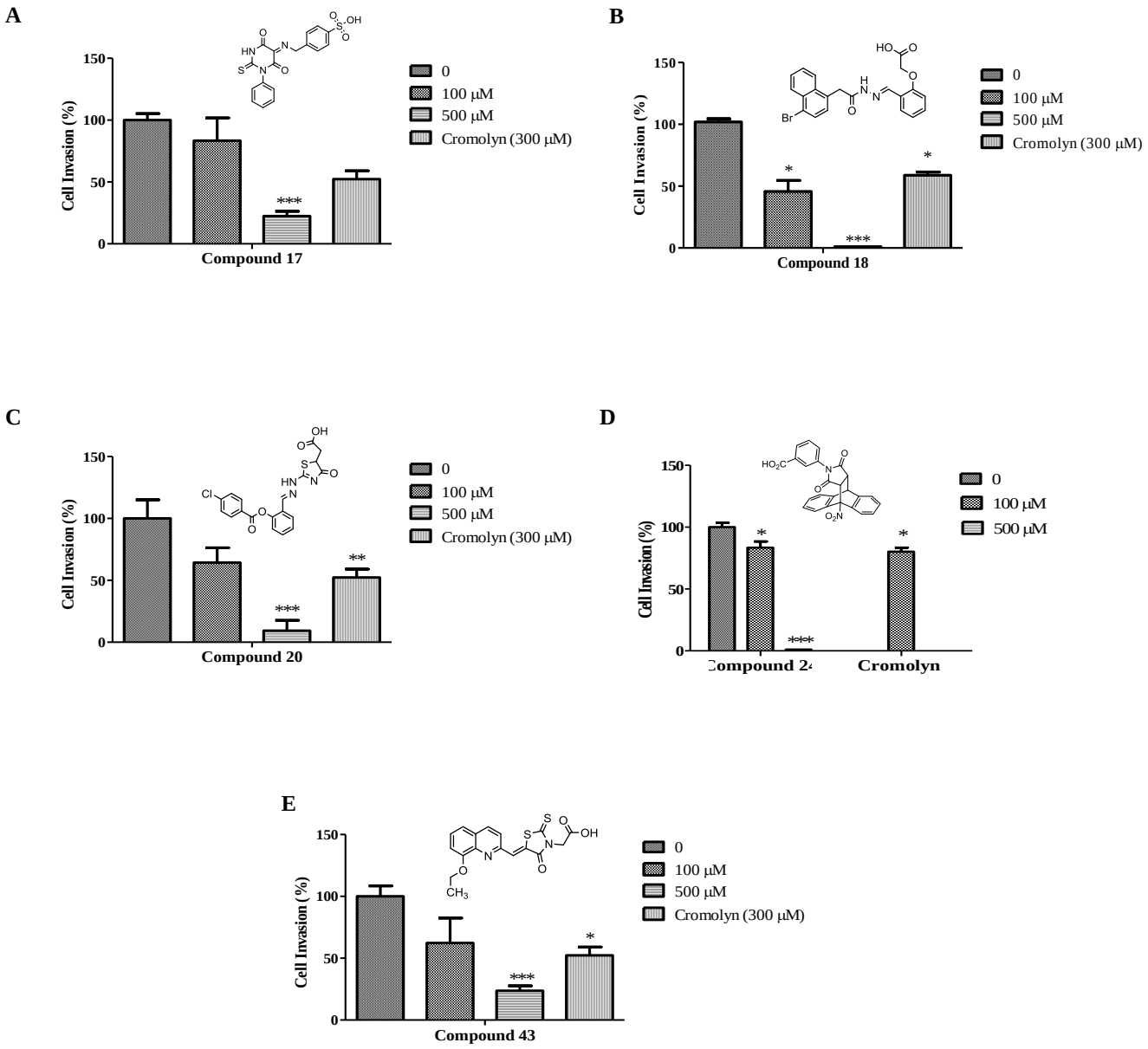



**Figure 3.5**: Effect of compounds screened on invasion of BxPC-3 pancreatic cancer cells. A) Compound **17**. B) Compound **18**. C) Compound **20**. D) Compound **24**. E) Compound **43**. Compounds were added to serum-free medium in the upper compartment of an invasion chamber along with the BxPC-3 cells (2.5 x $10^4$ cells/well), media with FBS (10% v/v) was added to lower chamber. After 48 hours, cells in the upper chamber were removed and cells that had invaded onto the lower surface of the membrane were stained Giemsa stain and cells in 5 different fields were counted for invasion studies. Results are expressed as the mean + SEM. P value of <0.05 (*), <0.01(**) and <0.001(***) was determined relative to control using 2-Way ANOVA.

Similar to cromolyn, many of the compounds examined did not show an effect on the proliferation of BxPC-3 cells. Compounds **18**, **20** and **26** had an effect at 500 μM, although the cells' response appear to be similar at all three time points for both compounds **20** and **26** (Figure 3.6). Again, similar to the response of the cells for the invasion assay at this concentration (Figure 3.5), it is not known if the effect observed here is due to the cells dying from being too stressed after being exposed to an environment of such high compound concentration.

Of the compounds examined for activity on the proliferation of Panc-1 cells, five compounds showed an effect at 500 μM (Figure 3.7). Compound **18** exerted a similar effect on the proliferation of both cells at 500 μM. Compound **58b**, the *tert*-butyl analogue of compound **58**, showed a significant concentration dependent effect on proliferation of Panc-1 cells at 100 and 500 μM ($p<0.05$ and 0.001 respectively) compared to control, although the cells seem to be recovering over time (Figure 3.7E).

The lack of anti-proliferative effect from these compounds on Panc-1 cells is not surprising as they are designed to interact with S100P. However, the lack of difference between their anti-proliferative effects on both cell lines is further evidence that S100P may not play a role in pancreatic cancer cell proliferation, contrary to published reports [151, 177].

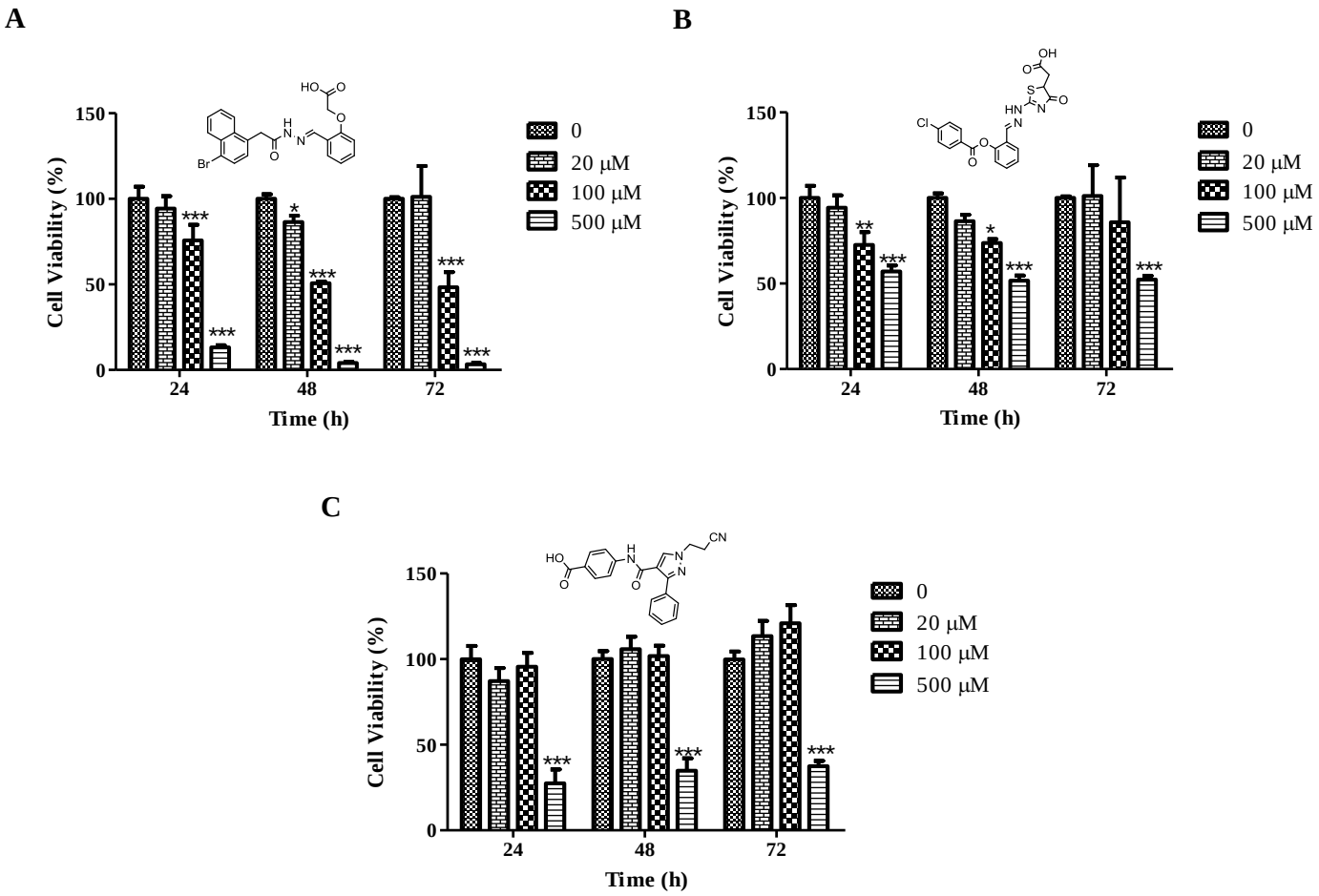



**Figure 3.6**: Effect of compounds screened on proliferation of BxPC-3 cells. A) Compound **18**. B) Compound **20**. C) Compound **26**. Cells ($1.0 \times 10^3$ cells/well) were cultured in the presence or absence of screening compounds and cell proliferation analysed using the MTS assay after 24, 48, and 72 h. Results are expressed as the mean + SEM. P value of <0.05 (*), <0.001(**) and <0.001(***) relative to control using 2-Way ANOVA. Cell viability was expressed as a percentage for each treatment group relative to the negative control (1% v/v DMSO in culture media) according to the following equation: Growth (%) = $OD_{490}$(sample)/ $OD_{490}$(control) x 100.

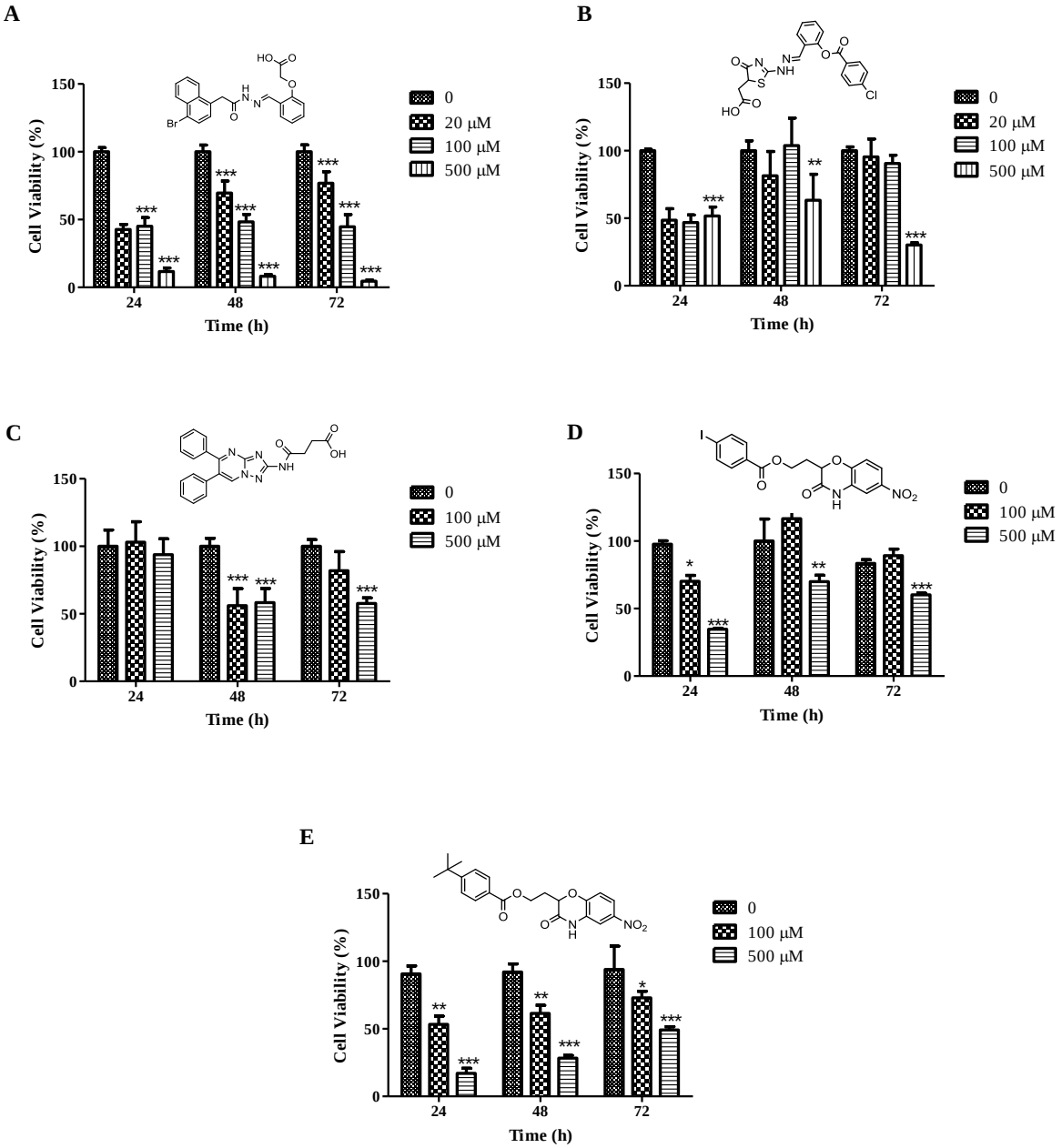



**Figure 3.7**: Effect of compounds screened on proliferation of Panc-1 cells. A) Compound **18**. B) Compound **20**. C) Compound **36**. D) Compound **58**. E) Compound **58b**. Cells ($1.0 \times 10^3$ cells/well) were cultured in the presence or absence of screening compounds and cell proliferation analysed using the MTS assay after 24, 48, and 72 h. Results are expressed as the mean + SEM. P value of <0.05 (*), <0.001(**) and <0.001(***) relative to control using 2-Way ANOVA. Cell viability was expressed as a percentage for each treatment group relative to the negative control (1% DMSO v/v in culture media) according to the following equation: Growth (%) = $OD_{490}$(sample)/$OD_{490}$(control) x 100.

The effect of the screened compounds on proliferation of both cell lines indicates that the mechanism by which these compounds are exerting their effect on the cells is not mediated by S100P. If the mechanism were S100P-dependent, it is expected that compounds that demonstrated an inhibition to the invasive ability of BxC-3 cells would also demonstrate a similar inhibitive effect on the proliferative capability of these cells. In contrast, no effect on both migration and proliferation would be observed with the S100P-negative cell line, Panc-1. This effect was not observed for any of the compounds examined. Since most of the compounds showed a comparable effect on proliferation of BxPC-3 cells to that of cromolyn (**4**), it could be that proliferation of these cells is independent of S100P, but more work needs to be performed to conclusively confirm this observation as it is contrary to current published data.

However, five compounds were identified that show activity against the invasion of BxPC-3 cells consistent with the experimental findings that S100P is involved in metastatic disease progression [177, 181, 226]. These five compounds, **17**, **18**, **20**, **24**, and **43**, will serve as good templates on which to design anti-metastatic compounds against pancreatic cancer.

## 3.4 CONCLUSION

Four compounds, **7**, **58** (its analogues **58a** and **58b)**, were synthesised, isolated and characterised. They formed part of the 17 compounds from the virtual screening studies that were screened for activity against the invasion and proliferation of pancreatic cancer cells.

Biological evaluation of the 17 hit compounds has identified five compounds that show promise with respect to inhibiting the invasion potential of BxPC-3 cells. Since these cells secrete endogenous S100P, this inhibitory effect is promising as there is scope to further exploit these compounds to increase potency against this novel therapeutic target. One such compound of interest is compound **24**. At 100 μM, it demonstrated an inhibitory effect comparable to that of cromolyn (**4**) at the same concentration. As this was one of the compounds originally purchased, the aim will be to synthesise, purify, characterised and retest it against these cells for reproducibility (Chapter 4).

The results from the proliferation assay studies are inconclusive as to the involvement of S100P in pancreatic cancer cell proliferation. Since compounds that show inhibitory activity against invasion of BxPC-3 cells did not reproduce similar effects on proliferation of the same cells, or the opposite results on Panc-1 cells, it is difficult to conclude that proliferation in pancreatic cancer is S100P-mediated based on these results, despite literature evidence to the contrary [151, 177].

## 3.5 EXPERIMENTAL

### 3.5.1 General methodology: Synthesis and Instrumentation

Reagents for the chemical synthesis unless specified were purchased from Sigma-Aldrich (Gillingham, Dorset, UK) and were used without further purification. All solvents were purchased from Fisher Scientific (Loughborough, UK). Where necessary, solvents were dried using activated 3 Å (methanol, ethanol, dichloromethane) and 4 Å (diethyl ether) molecular sieves.

$^{1}$H-NMR and $^{13}$C-NMR spectra were recorded on a JEOL ECA/54/SSS (400 or 600 MHz) spectrometer using (TMS) as internal standard. Deuterated solvents used for compound analysis are indicated with individual compound data. Raw NMR data were processed with ACD/NMR Processor Academic Edition Version 12.01 (Advanced Chemistry Development, Inc., Toronto, Canada, www.acdlabs.com, 2014). Chemical shifts are given in ppm relative to tetramethylsilane and *J* values (where given) are in Hz. Spectral splitting patterns are given as singlet (s), doublet (d), doublet of doublets (dd), triplet (t), quadruplet (q), quartet of quartets (qq), multiplet overlapped (m), broad (br).

Infrared spectra were recorded in Scimitar 800 FT-IR spectrometer (Varian Inc.), Nicolet 6700 FT-IR Smart iTR (Thermo Scientific) spectrometer using a Golden Gate™ Diamond ATR adapter, or using a Perkin Elmer FT-IR/FIR Spectrometer Frontier (Version 10.03.07), with samples prepared as thin films on the universal ATR sampling accessory.

Liquid chromatography–mass spectrometry (LC-MS) was determined using a Varian 1200L Quadrupole LC/MS/MS system equipped with Electrospray Ionisation (ESI) (Agilent Technologies, USA) and using a Varian Pursuit 50 mm x 4.6 mm 5 micron pore size C18 reverse phase column. Samples were run at a flow rate of 0.25 mL/min for 20 minutes using water/formic acid (0.1% v/v): acetonitrile/formic acid (0.1% v/v) mobile phases. The gradient at which the samples were ran is as follows:

0:00 minutes: 80% formic acid and water/formic acid and acetonitrile
10:00 minutes: 10% formic acid and water/formic acid acetonitrile
16:00 minutes: 10% formic acid and water/formic acid acetonitrile
16:30 minutes: 80% formic acid and water/formic acid acetonitrile
20:00 minutes: 80% formic acid and water/formic acid acetonitrile

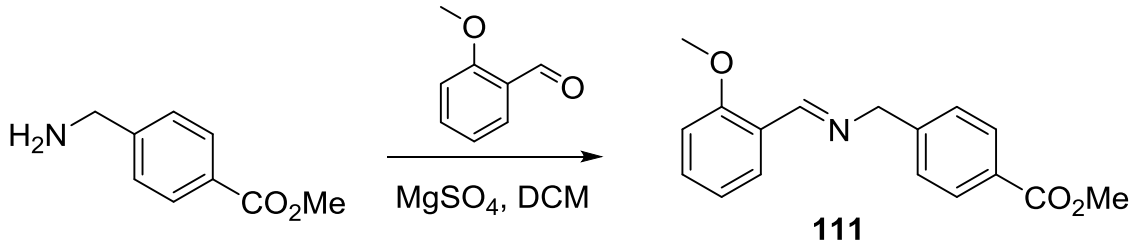
The Mass Spectrometer acquires both positive and negative ions from masses 50 Da to 1000 Da.

Thin-layer chromatography (TLC) was carried out using Macherey-Nagel 60 Å (250 μm thick) flexible polyester sheet silica gel plates pre-coated with fluorescent indicator $UV_{254}$ (TLC-sheets POLYGRAM® SIL G/$UV_{254}$, Fisher, Loughborough, UK). Dichloromethane (DCM) was used as the mobile phase unless otherwise stated. Column chromatography for compounds **55**, **55a** and **55b** was carried out using silica gel (high-purity grade, pore size 60 Å, 230-400 mesh particle size, 40-63 μm particle size, for flash chromatography, Sigma-Aldrich, Gillingham, Dorset, UK) in a 30 cm long column using diethyl ether ($Et_2O$) as eluent. Samples were dry-loaded by dissolving in either dichloromethane (DCM; **58** and **58b**)) or chloroform (**58a**) before adsorbing unto silica gel.

Melting points (M.p.) were measured in open capillaries using a Griffin melting point apparatus and are uncorrected. Where known, melting points from published literature are shown next to that determined for each compound in this study.

All structures, unless otherwise specified, were generated and named using ChemDraw Ultra 14. (PerkinElmer, Massachusetts, US).

### 3.5.2 Synthesis of Methyl 4-{[(*E*)-[(2-methoxyphenyl)methylidene]amino]methyl}benzoate (111). (Adapted from Mohamed *et al*. [339])




**111**

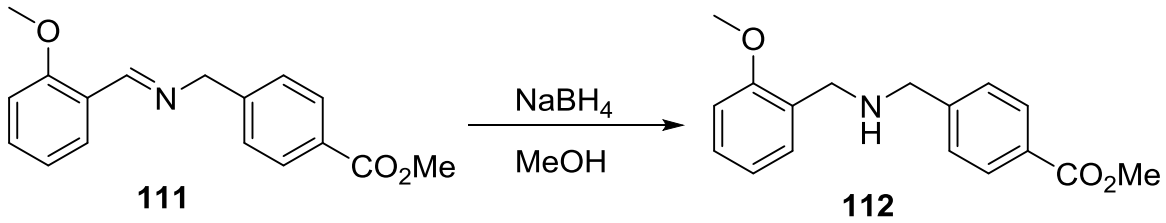
$C_{17}H_{17}NO_3$
Mol. Wt: 283.33

Methyl (4-aminomethyl)benzoate was obtained from its hydrochloride salt by stirring the salt with sodium carbonate in water and extracting the ester in diethyl ether. The extracted benzoate (0.21 g, 1.3 mmol, 1 eq) was reacted with 2-methoxybenzaldehyde (0.17 g, 1.3 mmol, 1 eq) in dry DCM (3 mL) in the presence of a drying agent (anhydrous $MgSO_4$). The reaction mixture was stirred overnight for a total reaction time of 22 h. The flask contents were filtered through $K_2CO_3$ and the solvent removed under vacuum to afford **111** as a yellow oil which was used in the next step without further purification. Yield 0.34 g, 95%; **$^1$H-NMR (600 MHz, DMSO-*d*$_6$) δ (ppm)** 8.80 (1 H, s, 1 HC=N), 7.89 - 7.92 (2 H, m, *para*-Ar-H), 7.85 (1 H, dd, *J* = 7.7, 1.7 Hz, Ar-H), 7.41 - 7.45 (3 H, m, 2 *para*-Ar-H, Ar-H), 7.08 (1 H, d, *J* = 8.5 Hz, Ar-H), 6.94 - 6.98 (1 H, m, Ar-H), 4.82 (2 H, s, $CH_2$), 3.83 (3 H, s, $CH_3$ ester), 3.81 (3 H, s, $CH_3$ ether). **$^{13}$C-NMR (150 MHz, DMSO-*d*$_6$) δ (ppm)** 166.69 (C=O ester), 159.08 (C=N), 158.19 (CO ether), 146.09 (Ar-C), 132.96 (Ar-CH), 129.83 (*para*-Ar-CH), 128.60 (Ar-C), 128.54 (*para*-Ar-CH), 127.19 (Ar-CH),

124.29 (Ar-C), 121.10 (Ar-CH), 112.41 (Ar-CH), 64.32 ($CH_2$), 56.20 ($CH_3$ ether), 52.58 ($CH_3$ ester).

### 3.5.2.1 Reduction of 111 to Methyl 4-({[(2-methoxyphenyl)methyl]amino}methyl) benzoate (112).

(Adapted from Blackburn and Taylor [340])



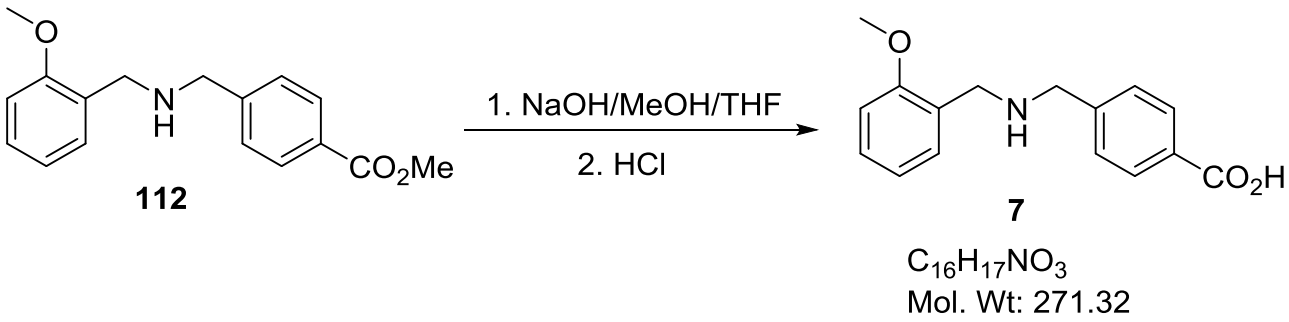

**111** **112**

$C_{17}H_{19}NO_3$
Mol. Wt: 285.34

The imine **111** (0.3 g, 1.1 mmol, 1 eq) was reduced to the amine under a nitrogen atmosphere using sodium borohydride (0.05 g, 1.3 mmol, 1.1 eq). Briefly, the imine was added to methanol (5 mL) and stirred at room temperature before cooling to 0 °C in an ice bath. Sodium borohydride ($NaBH_4$) was added in aliquots and the mixture stirred for another 0.5 h. After this time methanol was removed under vacuum and the residue diluted with $Et_2O$ (30 mL) and saturated $NaHCO_3$ (30 mL). The organic layer was extracted with $Et_2O$ (2 x 20 mL). The combined organic layers were dried over $MgSO_4$, the mixture was filtered and the solvent removed *in vacuo* to give the amine **112** as a pale yellowish oil, yield 0.23 g, 70%. **$^1$H-NMR (600 MHz, DMSO-*d*$_6$) δ (ppm)** 7.86 - 7.89 (2 H, d, *J* = 8.0 Hz, *para*-Ar-H), 7.46 (2 H, d, *J* = 8.0 Hz, *para*-Ar-H), 7.30 (1 H, d, *J* = 7.3 Hz, Ar-H), 7.18 (1 H, m, Ar-H),

6.91 (1 H, d, *J* = 8.3 Hz, Ar-H), 6.88 (1 H, m, Ar-H), 3.81 (3 H, s, $CH_3$ ester), 3.74 (2 H, s, $CH_2$), 3.72 (3 H, s, $CH_3$ ether), 3.61 (2 H, $CH_2$). **$^{13}$C-NMR (150 MHz, DMSO-$d_6$) δ (ppm)** 166.76 (C=O ester), 157.51 (CO ether), 147.45 (Ar-C), 129.58 (*para*-Ar-CH), 129.17 (Ar-CH), 128.80 (Ar-C), 128.53 (*para*-Ar-CH), 128.41 (Ar-CH), 128.27 (Ar-C), 120.61 (Ar-CH), 110.94 (Ar-CH), 61.07 ($CH_2$), 55.71 ($CH_3$), 52.50 ($CH_2$), 47.29 ($CH_3$ ester).

### 3.5.2.2 Hydrolysis of 112 to 4-({[(2-Methoxyphenyl)methyl]amino}methyl)benzoic acid (7)



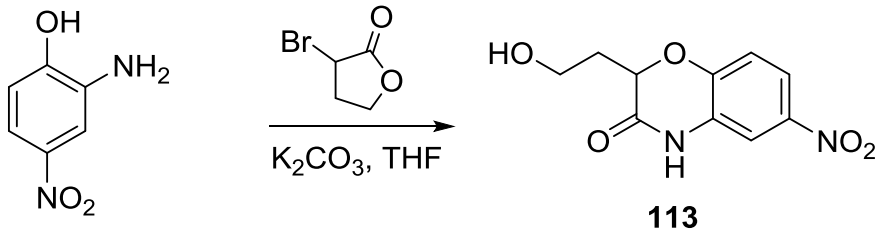

**112** **7**

$C_{16}H_{17}NO_3$
Mol. Wt: 271.32

The amine **112** (0.22 g, 0.78 mmol) was stirred under reflux in MeOH/THF mixture (1:1 v/v) and NaOH (1 M, 1.5 mL) for 1.5 h, then left overnight at room temperature whilst stirring. After a total reaction time of 20.5 h, the reaction was quenched with de-ionised water (2 mL) and the organic solvents removed under vacuum. HCl (1 M) was added to the concentrated residue to adjust the pH. A precipitate was formed at pH 7 which was filtered off, using a Hirsch funnel, to afford **7** as a pale yellow solid, yield 0.15 g, 71%. M.p. 190-192 °C. **$^1$H-NMR (600 MHz, MeOH-$d_3$) δ (ppm)** 7.95 - 7.99 (2 H, m, para-Ar-H), 7.43 (2 H, d, *J* = 8.3 Hz, para-Ar-H), 7.40 (1 H, dd, *J* = 7.6, 1.7 Hz, Ar-H), 7.31 (1 H, dd, *J* = 7.6, 1.7 Hz, Ar-

H), 7.05 (1 H, d, *J* = 8.3 Hz, Ar-H), 6.98 (1 H, td, *J* = 7.4, 1.0 Hz, Ar-H), 4.18 (2 H, s, $CH_2$), 4.13 (2 H, s, $CH_2$), 3.86 (3 H, s, $CH_3$). **$^{13}$C-NMR (150 MHz, MeOH-$d_3$) δ (ppm)** 172.57 (C=O acid), 158.05 (CO ether), 138.37 (Ar-C), 133.70 (Ar-C), 131.23 (Ar-CH), 131.15 (Ar-CH), 129.67 (para-Ar-CH), 129.00 (para-Ar-CH), 120.63 (Ar-CH), 119.67 (Ar-C), 110.69 (Ar-CH), 54.72 ($CH_3$), 50.42 ($CH_2$), 46.20 ($CH_2$). **IR $\nu_{max}$/cm$^{-1}$**: 3368.08 (OH stretch), 2920.39 (CH stretch), 2850.51 (CH stretch), 2188.30, 1716.84 (C=O stretch), 1604.93 (NH bend). LC-MS (ESI) found ***m/z*** [M + H]$^+$: 272, $C_{16}H_{17}NO_3$ requires 271.

### 3.5.3 Synthesis of 2-(2-Hydroxyethyl)-6-nitro-3,4-dihydro-2*H*-1,4-benzoxazin-3-one (113). (Adapted from Frechette and Beach [333])




**113**

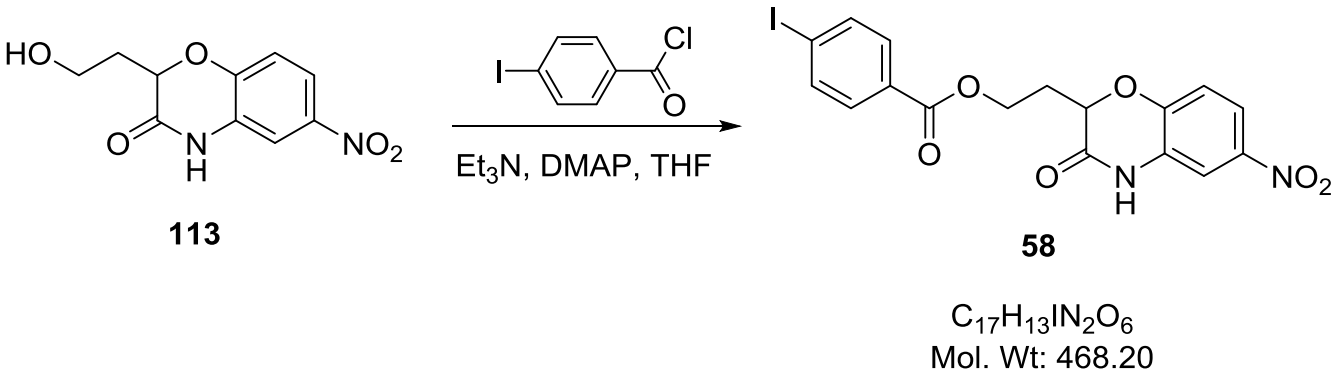
$C_{10}H_{10}N_2O_5$
Mol. Wt: 238.20

2-Amino-4-nitrophenol (3.00 g, 19.48 mmol, 1 eq) and potassium carbonate (4.03 g, 29.22 mmol, 1.5 eq) were stirred in anhydrous tetrahydrofuran (THF, 40 mL) at room temperature (r.t.) for 50 minutes under a nitrogen gas atmosphere. α-Bromo-γ-butyrolactone (3.86 g, 23.38 mmol, 1.2 eq) was added dropwise and the resulting mixture heated under reflux for 5.5 h. The reaction was quenched with 80 mL ice-cold deionised water and the solvent removed under vacuum. The aqueous portion was cooled on ice before filtering off the solid to afford **113** as a brick-red solid. Recrystallisation from ethyl acetate gave fluffy yellow crystals, yield 2.61 g (56%). M.p. 169-170 °C (lit. 173 °C [333]); $R_f$ 0.81 (ethyl acetate: cyclohexane: ethanoic acid, 5:4:1 v/v/v). **$^1$H-NMR (600 MHz, DMSO-$d_6$) δ (ppm)** 11.01 (1 H, br. s., NH), 7.81 (1 H, dd, *J* = 8.3, 2.3 Hz, Ar-H), 7.70 (1 H, d, *J* = 2.3 Hz, Ar-H), 7.14 (1 H, d, *J* = 8.3 Hz, Ar-H), 4.85 (1 H, dd, *J* = 9.2, 4.6 Hz, CH), 4.65 (1 H, t, *J* = 5.5 Hz, OH), 3.45 - 3.62 (2 H, m, $CH_2$), 1.79 - 2.03 (2 H, m, $CH_2$). **$^{13}$C-NMR (150 MHz, MeOH-$d_3$) δ (ppm)** 166.97 (C=O amide), 148.29 (Ar-CO), 142.81 (Ar-$CNO_2$), 127.55 (Ar-CNH), 119.36 (Ar-CH), 116.83 (Ar-CH), 110.78 (Ar-CH), 74.33

(aliphatic-CHO), 56.75 (aliphatic-$CH_2$), 33.46 (aliphatic-$CH_2$). **IR $\nu_{max}$/cm$^{-1}$**: 3596 (OH stretch), 3187 (CH stretch), 3118 (CH stretch), 3047 (CH stretch), 2955 (CH stretch), 2882 (CH stretch), 1696 (C=O amide), 1605 (C=C aromatic stretch), 1516 ($NO_2$ asymmetrical stretch), 1491 ($CH_2$ bend), 1394 ($NO_2$ symmetrical stretch), 1331 (C-O stretch). LC-MS (ESI) found ***m/z*** $[M - H]^+$: 237, $C_{10}H_{10}N_2O_5$ requires 238.

### 3.5.4 Synthesis of 2-(6-Nitro-3-oxo-3,4-dihydro-2*H*-1,4-benzoxazin-2-yl)ethyl 4-iodobenzoate (58).

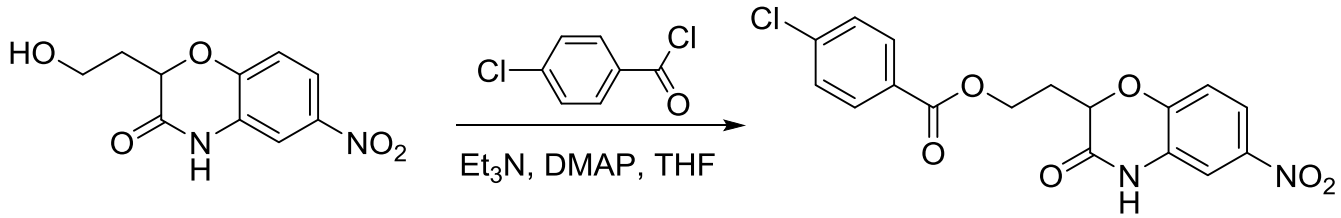
(Adapted from Berry *et al*. [341])




Compound **113** (1.0 g, 4.2 mmol, 1 eq), triethylamine (1.4 mL, 0.4 g/mL, 5.0 mmol, 1.2 eq) and 4-dimethylaminopyridine (DMAP, 0.30 g, 2.1 mmol, 0.5 eq) were stirred under a nitrogen atmosphere in anhydrous tetrahydrofuran (THF, 120 mL) at room temperature for at least 30 minutes. The mixture was then cooled in an ice bath to 0 °C and 4-iodobenzoyl chloride (1.12 g, 4.20 mmol, 1 eq) was added, in portions, with stirring, to the mixture. The mixture was then heated under reflux for a total reaction time of 5 h. After completion of the reaction as judged by TLC, the reaction mixture was poured into a flask containing 100 mL deionised ice-cold $H_2O$. The

aqueous phase was extracted with DCM three times (2x 30 mL, 1x 20 mL). The combined organic phases were washed with 10% HCl (2.74 M, 40 mL) to remove the triethylamine base, followed by saturated sodium hydrogen carbonate ($NaHCO_3$, 40 mL). The extracted organic layer was dried over anhydrous magnesium sulphate ($MgSO_4$), filtered, and the solvent removed under vacuum to afford **58** as a sticky brown solid. Column chromatography (diethyl ether) afforded the pure ester as a pale yellow solid, yield 0.15 g, 8%. M.p. 188-190 °C; $R_f$ 0.43 (diethyl ether). **$^1$H-NMR (600 MHz, DMSO-*d*$_6$) δ (ppm)** 11.10 (1 H, s, NH), 7.86 - 7.89 (2 H, m, *para*-Ar-H), 7.79 (1 H, dd, *J* = 8.7, 2.8 Hz, Ar-H), 7.67 (1 H, d, *J* = 2.3 Hz, Ar-H), 7.63 - 7.68 (2 H, m, *para*-Ar-H), 7.10 (1 H, d, *J* = 8.9 Hz, Ar-H), 5.03 (1 H, dd, *J* = 7.6, 4.6 Hz, CH), 4.42 (2 H, t, *J* = 6.3 Hz, $CH_2$ aliphatic), 2.24 - 2.39 (2 H, m, $CH_2$). **$^{13}$C-NMR (150 MHz, DMSO-*d*$_6$) δ (ppm)** 174.53 (C=O amide), 165.72 (C=O ester), 154.57 (Ar-CO), 145.67 ($CNO_2$), 138.26 (*para*-Ar-CH), 131.41 (*para*-Ar-CH), 129.01 (Ar-C), 127.81 (Ar-C), 119.67 (Ar-CH), 117.34 (Ar-CH), 111.17 (Ar-CH), 101.12 (Ar-I), 74.56 (CH), 61.19 ($CH_2$ aliphatic), 30.07 ($CH_2$ aliphatic). **IR $\nu_{max}$/cm$^{-1}$**: 3607 (NH stretch), 3257 (CH stretch), 3088 (CH stretch), 1704 (C=O ester), 1673 (C=O amide), 1633, 1606 (C=C aromatic stretch), 1588, 1517 ($NO_2$ asymmetrical stretch), 1495 ($CH_2$ bend), 1473, 1393 ($NO_2$ symmetrical stretch). LC-MS (ESI) found ***m/z*** $[M - H]^-$: 467 $C_{17}H_{13}IN_2O_6$ requires 468.

### 3.5.4.1 Synthesis of 2-(6-Nitro-3-oxo-3,4-dihydro-2*H*-1,4-benzoxazin-2-yl)ethyl 4-chlorobenzoate (58a)



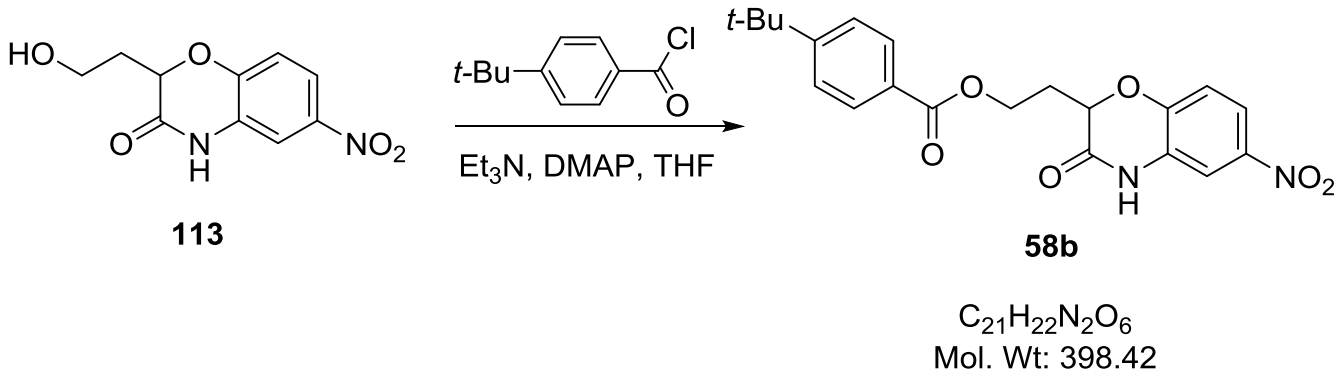

**113** **58a**

$C_{17}H_{13}ClN_2O_6$
Mol. Wt: 376.75

Compound **113** (0.52 g, 2.2 mmol, 1 eq), $Et_3N$ (0.4 mL, 0.7 g/mL, 2.6 mmol, 1.2 eq), DMAP (0.1 g, 1.1 mmol, 0.5 eq), 4-chlorobenzoyl chloride (0.38 g, 0.28 mL, 2.2 mmol, 1 eq). Procedure identical to the synthesis of compound **58**. Column chromatography (diethyl ether) afforded **58a** as a greenish-yellow solid, yield 0.042 g, 5%. M.p. 218-220 °C; $R_f$ 0.82 (diethyl ether). **$^1$H-NMR (600 MHz, DMSO-*d*$_6$) δ (ppm)** 11.09 (1 H, s, NH), 7.88 - 7.93 (2 H, m, *para*-Ar-H), 7.80 (1 H, dd, *J* = 8.9, 2.75 Hz, Ar-H), 7.67 (1 H, d, *J* = 2.8 Hz, Ar-H), 7.53 - 7.58 (2 H, m, *para*-Ar-H), 7.11 (1 H, d, *J* = 8.9 Hz, Ar-H), 5.04 (1 H, dd, *J* = 7.6, 4.47 Hz, CH), 4.43 (2 H, t, *J* = 6.2 Hz, $CH_2$ aliphatic), 2.25 - 2.40 (2 H, m, $CH_2$ aliphatic). **$^{13}$C-NMR (150 MHz, DMSO-*d*$_6$) δ (ppm)** 165.69 (C=O amide), 165.31 (C=O ester), 148.66 (Ar-CO), 142.52 ($CNO_2$), 138.85 (*para*-Ar-CH), 131.59 (*para*-Ar-CH), 129.44 (Ar-C), 128.92 (Ar-C), 128.24 (Ar-C), 119.74 (Ar-CH), 117.37 (Ar-CH), 111.15 (Ar-CH), 74.57 (CH), 61.23 ($CH_2$ aliphatic), 30.07 ($CH_2$ aliphatic). **IR $\nu_{max}$/cm$^{-1}$**: 3275 (NH stretch), 1716 (C=O ester), 1687 (C=O amide), 1524 ($NO_2$ asymmetrical stretch), 1488 ($CH_2$

bend), 1378 ($NO_2$ symmetrical stretch). LC-MS (ESI) found ***m/z*** $[M - H]^-$: 375, $C_{17}H_{13}ClN_2O_6$ requires 376.

### 3.5.4.2 Synthesis of 2-(6-Nitro-3-oxo-3,4-dihydro-2*H*-1,4-benzoxazin-2-yl)ethyl 4-tert-butylbenzoate (58b)




$C_{21}H_{22}N_2O_6$
Mol. Wt: 398.42

Compound **113** (0.5 g, 2.1 mmol, 1 eq), $Et_3N$ (0.35 mL, 0.73 g/mL, 2.5 mmol, 1.2 eq), DMAP (0.13 g, 1.1 mmol, 0.5 eq), 4-*tert*-butylbenzoyl chloride (0.41 g, 0.41 mL, 2.1 mmol, 1 eq). Procedure identical to the synthesis of compound **58**. Column chromatography (diethyl ether) afforded **58b** as a pale brown solid, yield 0.31 g, 37%. M.p. 88-90 °C; $R_f$ 0.82 (diethyl ether) 0.84. **$^1$H-NMR (600 MHz, DMSO-*d*$_6$) δ (ppm)** 11.11 (1 H, s, NH), 7.81 - 7.84 (2 H, m, *para*-Ar-H), 7.78 (1 H, dd, *J* = 8.9, 2.8 Hz, Ar-H), 7.66 (1 H, d, *J* = 2.8 Hz, Ar-H), 7.47 - 7.50 (2 H, m. *para*-Ar-H), 7.11 (1 H, d, *J* = 8.9 Hz, Ar-H), 5.03 (1 H, dd, *J* = 7.6, 4.47 Hz, CH), 4.41 (2 H, t, *J* = 6.2 Hz, $CH_2$ aliphatic), 2.25 - 2.38 (2 H, m, $CH_2$ aliphatic), 1.29 (9 H, s, $CH_3$). **$^{13}$C-NMR (150 MHz, DMSO-*d*$_6$) δ (ppm)** 166.06 (C=O ester), 165.68 (C=O amide), 156.91 (Ar-C), 148.65 (Ar-CO), 142.49 ($CNO_2$), 129.63 (*para*-Ar-CH), 128.21 (Ar-C), 127.35 (Ar-C), 126.03 (*para*-Ar-CH), 119.70 (Ar-CH), 117.34

(Ar-CH), 111.13 (Ar-CH), 74.59 (CH), 60.70 ($CH_2$ aliphatic), 35.37 (C-*t*Bu), 31.33 ($CH_3$), 30.20 ($CH_2$ aliphatic). **IR $\nu_{max}$/cm$^{-1}$**: 3251 (NH stretch), 1700 (C=O ester), 1685 (C=O amide), 1502 ($NO_2$ asymmetrical stretch), 1365 ($NO_2$ symmetrical stretch). LC-MS (ESI) found ***m/z*** $[M - H]^-$: 396, $C_{21}H_{22}N_2O_6$ requires 398.

# CHAPTER 4: Synthesis and Biological Evaluation of ChemBridge 7356270 (24) and Analogues

# 4. PREFACE

ChemBridge 7356270 (**24**) was one compound from the initial screening studies that showed promising inhibitory activity against the invasion of the S100P-expressing pancreatic cancer cells BxPC-3. At 100 μM, the compound exerted an effect on BxPC-3 cells that is comparable to the proposed S100P-RAGE interaction inhibitor cromolyn at the same concentration (Figure 3.5D, pg. 135).

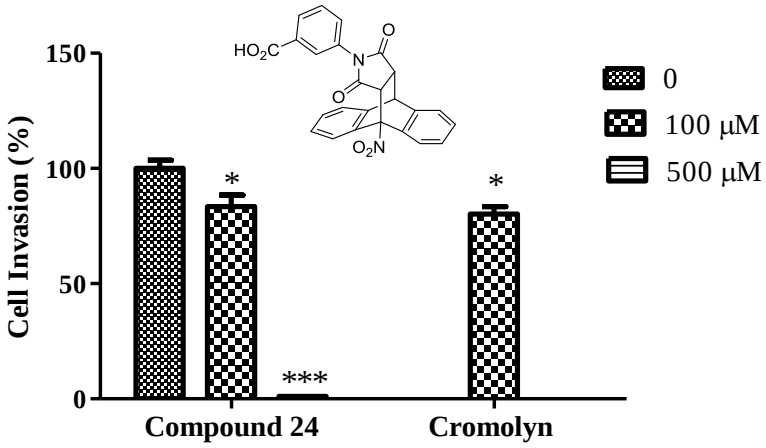



**Figure 3.5D**: Effect of compound **24** on the invasion of BxPC-3 cells. At 100 μM, compound **24** showed a significant effect on the invasion of these cells relative to the non-treated control. This effect is comparable to that of cromolyn at the same concentration. **P*<0.05; ****P*<0.001. (Mahmoud [328]).

Compound **24** showed a weak but significant effect ($p<0.001$) on proliferation of BxPC-3 cells at the highest concentration of 500 μM after exposure to the compound for 24 h (Chapter 3). This anti-proliferation effect was not observed on Panc-1 cells on exposure for 24 h. The effect of cromolyn on proliferation of BxPC-3 were not reproducible by Mahmoud [328] and in earlier work carried out in Dr Crnogorac-Jurcevic's laboratory (unpublished results) at Queen Mary University of London (QMUL). Work by the same laboratory also found that siS100P had little effect on the proliferation of BxPC-3 (Chapter 3) casting doubt on the role of S100P in the proliferation of pancreatic cancer cells [328]. It could therefore be possible that inhibition of S100P could affect pancreatic cancer cell migration/invasion and survival without significantly affecting proliferation.

In this chapter, the design, synthesis and characterisation of ChemBridge 7356270 (**24**) and its analogues will be discussed. Two synthetic routes to the target compound are presented, followed by biological screening studies of five analogues, including compound **24**, against pancreatic cancer cells. Preliminary findings from chick chorioallantioc membrane (CAM) assay studies, used to assess the effect of the compounds on angiogenesis, are also presented.

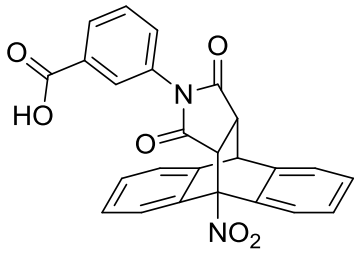
The experimental work detailing the syntheses and biological screening, and analytical data pertaining to the analogues synthesised form the final section of the chapter.

## 4.1 INTRODUCTION

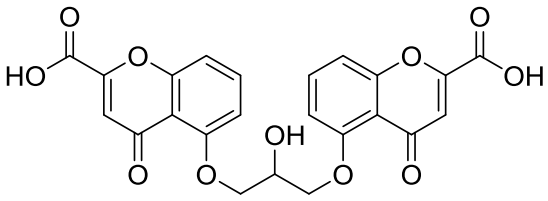
The discovery of compound **24** as a potential lead candidate for a therapeutic agent against pancreatic cancer is promising. Given the lethality of this cancer, any compound that shows even marginal potency against the disease opens up many possibilities for further optimisation on the chemical structure in order to improve its therapeutic effect.

Structurally, compound **24** is quite distinct from cromolyn (**4**) and its analogue C5-OH (**5**) (Figure 4.1), sharing a maximum common substructure Tanimoto similarity coefficient of 0.17 and 0.16 with both compounds respectively [282, 342]. Compounds **4** and **5** have been reported to inhibit the growth and invasion of pancreatic cancer cells via disruption of the S100P-RAGE interaction [195, 197]. This structural diversity between **24** and **4** and **5** bodes well in the search for novel inhibitors of S100P as it ensures exploitation of the chemical space that could be targeted in pancreatic cancer therapy.

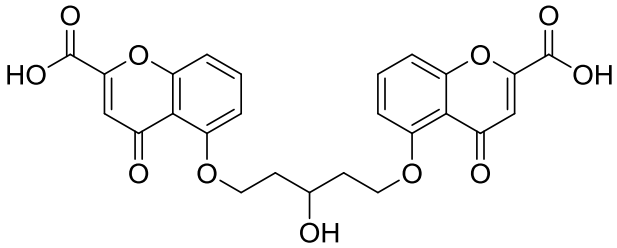
Although compound **24** is commercially available, this will be first time, to the best knowledge of this author, that its synthesis, purification and characterisation are discussed. The approach used to design synthetic routes for the compound will be shown. The compound's *in vitro* activity, as well as four of its analogues, against target-specific and non-specific pancreatic cancer cells will also be discussed.




**24**




**Cromolyn (4)**



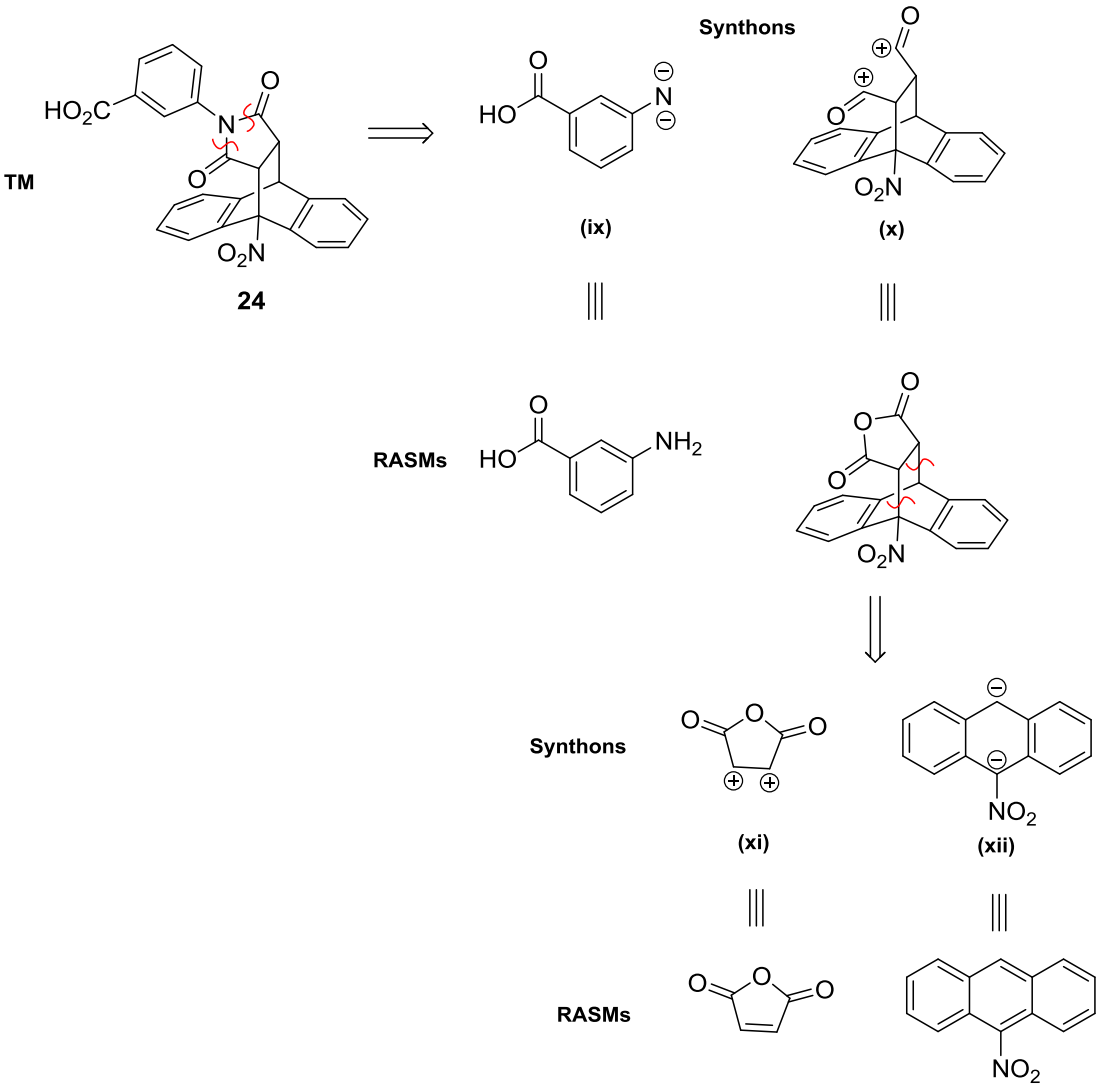

**C5OH (5)**

**Figure 4.1**: Compound **24** is structurally different from cromolyn (**4**) and its analogue C5OH (**5**), the two compounds shown to bind to S100P [195, 197].

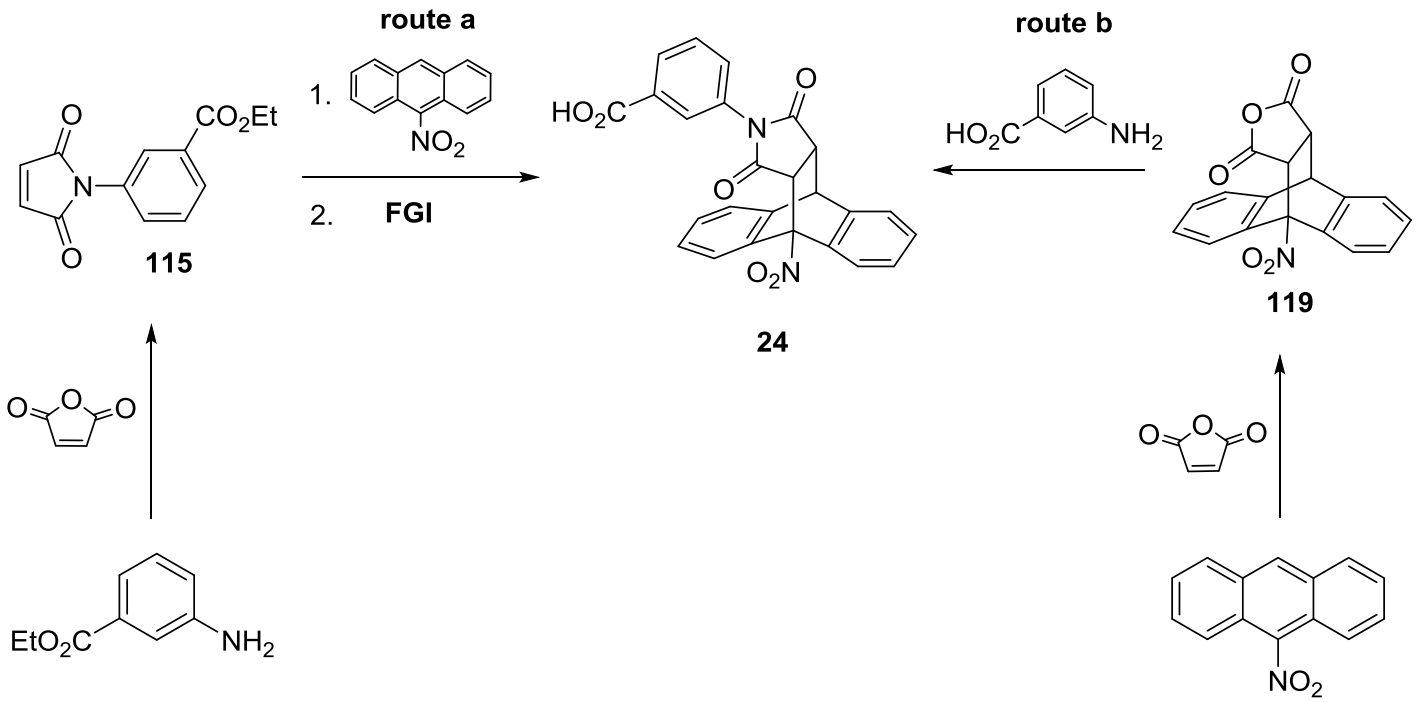
## 4.2 RESULTS AND DISCUSSION

### 4.2.1 Strategy for the synthesis of compound 24

Retrosynthetic analysis on 3-(9-nitro-12,14-dioxo-9,10-dihydro-9,10-[3,4]epipyrroloanthracen-13-yl)benzoic acid (**24**) identified four synthons (**ix**), (**x**), (**xi**) and (**xii**) (Scheme 4.1). Disconnection of the two O=C-N bonds on the pyrrolidine-2,5-dione moiety led to 3-aminobenzoic acid being identified as the appropriate synthetic equivalent for synthon (**ix**), and 9-nitroanthracene-maleic anhydride cycloadduct as the synthetic equivalent for (**x**). Further disconnection of the cycloadduct resulted in maleic anhydride and 9-nitroanthracene as the corresponding synthetic equivalents for the generated synthons (**xi**) and (**xii**) respectively (Scheme 4.1). Having determined the simplest readily available starting materials (RASMs) to the target molecule **24**, two synthetic routes **a** and **b** were identified using Reaxys® (version 1.7.8; Elsevier; 2012, www.elsevier.com/reaxys, Scheme 4.2). Both routes involve the reaction of maleic anhydride either with ethyl aminobenzoate initially, followed by 9-nitroanthracene (route **a**) or alternatively with 9-nitrotoanthracene initially, followed by the aminobenzoic acid (route **b**). The first step of route **a** involves an addition/elimination reaction to yield the intermediate **115**, which will subsequently react with 9-nitroanthracene in a Diels–Alder cycloaddition to give **24**. Alternatively, route **b** starts with the Diels–Alder cycloaddition reaction between maleic anhydride and 9-nitroanthracene to form the cycloadduct **119** (Scheme 4.2). The final step in this route involves a simultaneous addition/elimination reaction with the amine to form the target compound **24** (Section 4.2.3).




**Scheme 4.1**: Retrosynthetic analysis on target molecule (TM) **24** to identify readily available starting materials (RASMs).



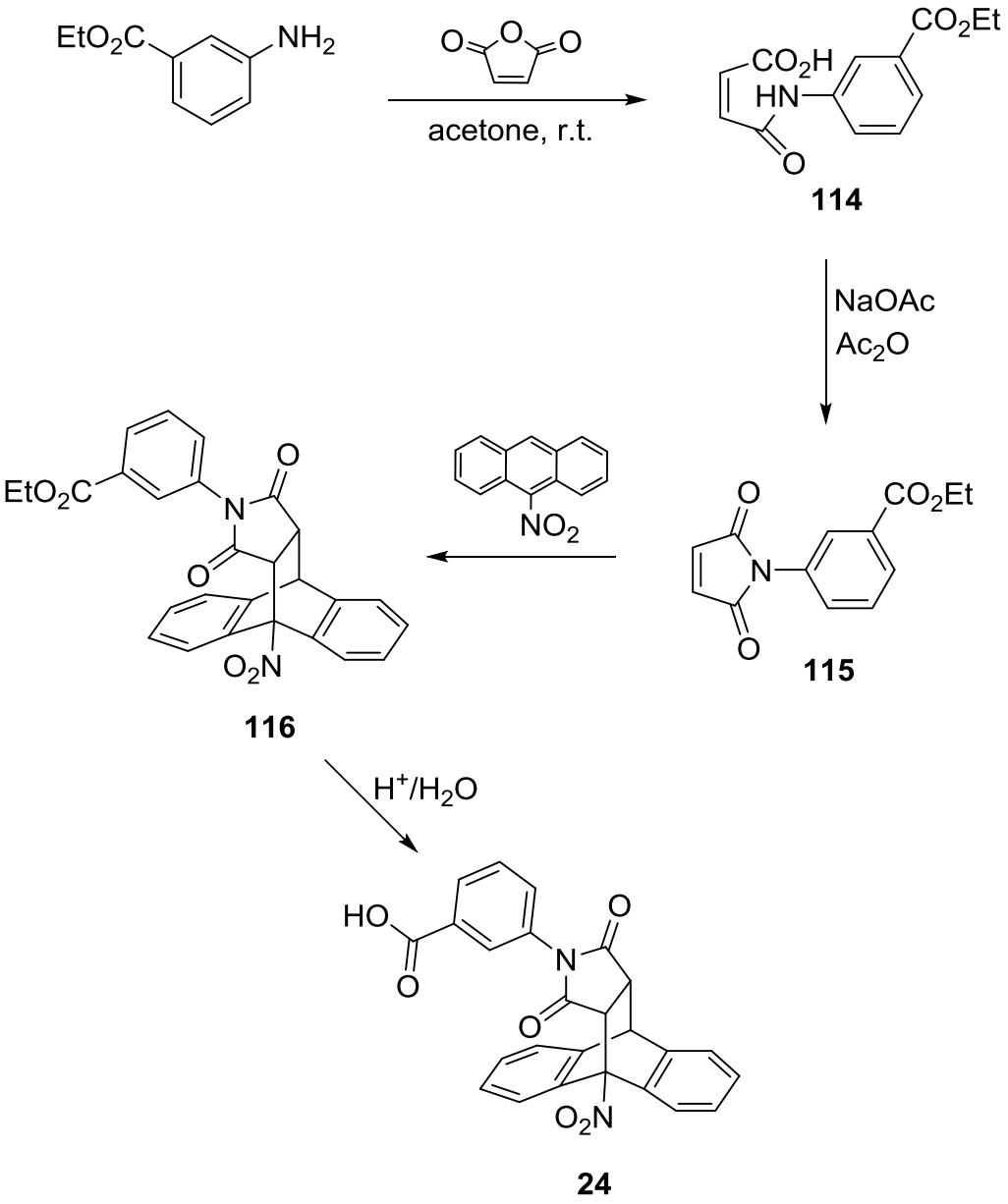

**Scheme 4.2**: Overview of the synthetic routes identified for **24**. (FGI: functional group interconversion).

### 4.2.2 Route a

In the first approach (Scheme 4.3), the ethyl ester of 3-aminobenzoic acid was reacted with maleic anhydride to obtain **114**. The acid **114** was then dehydrated to give the maleimide **115**. Using the ester instead of the acid in the first step circumvented solubility problems encountered in subsequent steps when 3-aminobenzoic acid was used as the initial reagent. The first synthetic step was carried out at room temperature to yield the intermediate product **114** and the second under simple reflux conditions to give intermediate **115**. Yields for intermediates **114** and **115** were reproducible, and the intermediates were not purified prior to the next synthetic step being carried out (Scheme 4.3). The next stage of the synthesis involved heating the maleimide **115** under reflux with 9-nitroanthracene in a Diels–Alder cycloaddition, to give the ester cycloadduct **116** of the target compound.

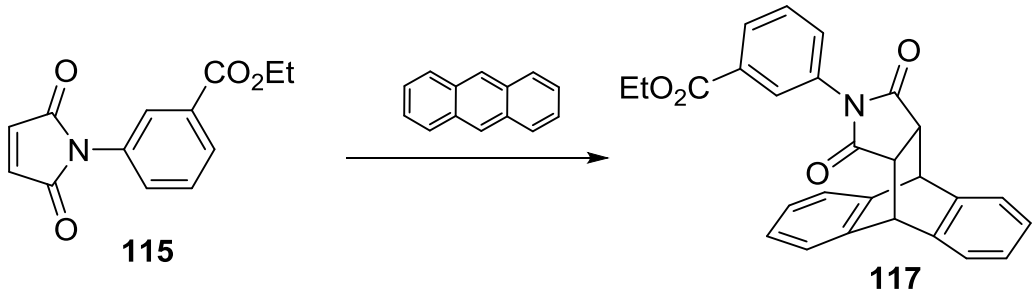
Diels–Alder synthesis is an important [4+2] cycloaddition reaction that is commonly employed to generate new cycloadducts by reacting a diene, a compound with two conjugated carbon double bonds, in its *cis* conformation, with a dienophile [343-346]. The reaction was initially described by Otto Diels and Kurt Alder in 1928 [347], for which they were awarded a Nobel Prize in 1950. The Diels–Alder reaction has been widely applied in synthesis to give highly regioselective and stereogenic compounds [348-350]. Its simplicity has made it an ideal and important route for forming new unsaturated six-membered rings and the use of catalysts has seen the production of highly enantioselective products [351-353].

The first attempt at synthesising the target compound **24** via the Diels–Alder reaction was carried out by adapting the method of Bova *et al*. [354] but was not successful. In this method, the maleimide **115** was reacted with 9-nitroanthracene under reflux in xylene but no product was isolated. Microwave-assisted synthesis with and without solvent proved ineffective. Both toluene and xylene, the two non-polar solvents used, are poor microwave energy absorbers [355] and could not therefore attain a high enough temperature required to facilitate the reaction. It was hoped that the use of microwave-assisted synthesis using such solvents would allow the direct transfer of microwave energy to the reactants thus allowing the reaction to occur. The use of silica gel in Diels–Alder reactions using microwave energy has been previously reported with moderate yields [356, 357]. However, when the Diels–Alder reaction between **115** and 9-nitroanthracene was carried out with silica gel using the microwave, no product was isolated, the starting diene was recovered suggesting no reaction had occurred.




**Scheme 4.3**: First synthetic route (route a) for compound **24**.

While 9-nitroanthracene has been cited in literature for many Diels–Alder reactions [358-360], the poor reactivity between this diene and the maleimide **115** could be attributed to the former's poor electron density due to the electron withdrawing effect of the nitro group on the π system. Electron-withdrawing groups on position C9 of anthracene have been reported to slow the rate of the Diels-Alder reaction [361]. It is also possible that the bulky *N*-aryl group of **115** was sterically interfering with the π system of the diene. Similar interactions between *ortho*-substituents of the *N*-aryl group and the π system have been previously reported to have a stabilising or destabilising effect on the π system [362]. Absence of the electron withdrawing effect of the nitro group was evident when anthracene was used as the diene. Anthracene reacted well with **115** under reflux conditions to give the anthracene-maleimide cycloadduct ester **117** (Scheme 4.4). Hydrolysis of the ester to the acid did not go to completion, but gave a mixture of both the acid and ester. No attempts were made to separate the two as a simpler route to forming the target compound (route **b**) had been identified.



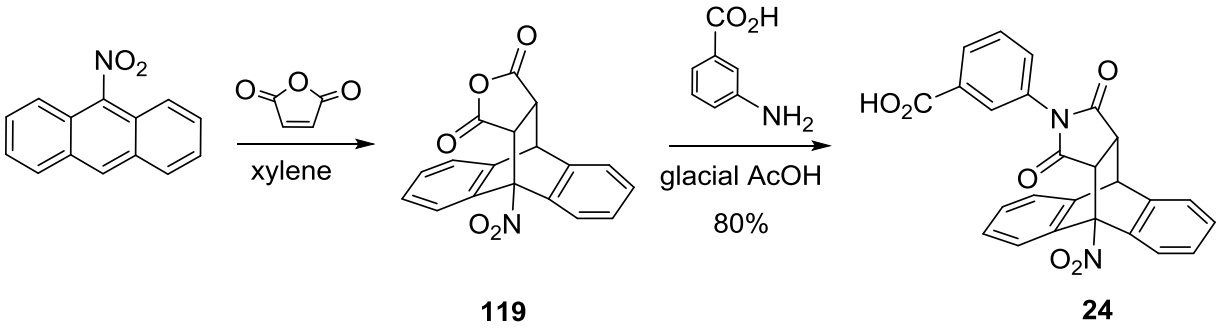

**Scheme 4.4**: Reaction of anthracene with **115** to give the anthracene-maleimide cycloadduct ester **117**.

### 4.2.3 Route b

A second approach to generate the target compound was employed after the first attempt (Scheme 4.3) failed to generate the required product **24**. In the second route, maleic anhydride was first reacted with 9-nitroanthracene under reflux as described by Wade [363] to give the Diels–Alder cycloadduct **119** (Scheme 4.5). The successful reaction between maleic anhydride and 9-nitroanthracene demonstrates the possibility that failure of the latter to react with the dienophile **115** (route **a**, Scheme 4.3) could be due to the steric inference of the *N*-aryl group of the dienophile with the π system of the diene in addition to the possible electron withdrawing effect of the nitro group of the diene on the π system.

The cycloadduct product **119** was confirmed as the desired intermediate by the appearance of three distinct signals for the three chemically different protons in the $^{1}$H-NMR as a result of the formation of a new 6-membered ring (Section 4.5.2.1). Since these signals were different from the singlet observed for the olefinic protons of maleic anhydride, their appearance in the $^{1}$H-NMR spectrum served as a benchmark in confirming the success or otherwise of the Diels–Alder cycloaddition reaction.

In the final step of this route, the cycloadduct **119** was reacted directly with 3-aminobenzoic acid in glacial acetic acid [364] to give the desired product **24** as a white solid in 80% yield (Scheme 4.5). $^{1}$H- and $^{13}$C-NMR, IR and LC-MS analyses all confirmed the product as compound **24** (Section 4.5.2.1.1).




**Scheme 4.5**: Synthesis of compound **24** via route b.

### 4.2.4 Analogues of compound 24

Once the target compound was confirmed as compound **24**, route **b** was employed to synthesise analogues by varying three main groups R, R′ and X on the target compound (Figure 4.2).

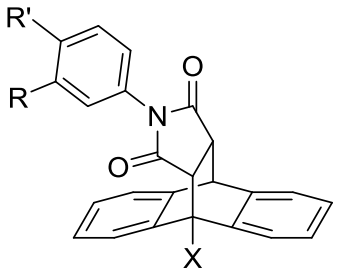



**Figure 4.2**: General structure of analogues of compound **24** that were synthesised by varying the X, R and R′ groups.

In the first instance, analogues were synthesised using 9-nitroanthracene (X = $NO_2$) as the diene in the Diels–Alder reaction. The substituents on the amine (Scheme 4.2) were varied between positions 3 and 4 of the phenyl ring. *Ortho*-substituted amines were not considered for synthesis because of the potential interaction between the *ortho*-substituent on the *N*-aryl phenyl ring and the π system of one of the benzene rings of the dihydroanthracene group if the latter adopts a conformation perpendicular to the succinimide group [362]. In addition, docking studies showed that an *ortho* substituent on the *N*-phenyl ring could prevent that part of the compound from interacting with the "smaller putative site" of S100P (Figure 4.3). A total of 19 analogues were synthesised with X = $NO_2$ (**24a-24s**), and *meta*- and *para*-substituted amines in addition to compound **24** (Table 4.1, Section 4.5.2.1.1).

The next set of analogues were synthesised using anthracene (X = H) as the diene replacing 9-nitroanthracene. Each analogue synthesised with the nitro group had a corresponding analogue without the group (Table 4.2, Section 4.5.2.2.1). Twenty compounds were synthesised from the anthracene-maleic anhydride cycloadduct (**121-121s**) bringing to 40 the total number of analogues synthesised including compound **24**. It was not surprising to find some of the synthesised anthracene-maleic anhydride derived analogues, **121**, **121d**, **121e**, **121j**, **121n**, **121n**, **121p**, **121q**, and **121r** in the literature [365-368] as this cycloadduct is commonly encountered in many Diels–Alder reactions [369-371]. These analogues were previously synthesised to study their photoactivatability ([365], **121e**, **121n**, **121p**, **121q**, **121r**) or their inclusion behaviour ([366], **121**, **121d**, **121e**, **121n**) but not for their biological activity.

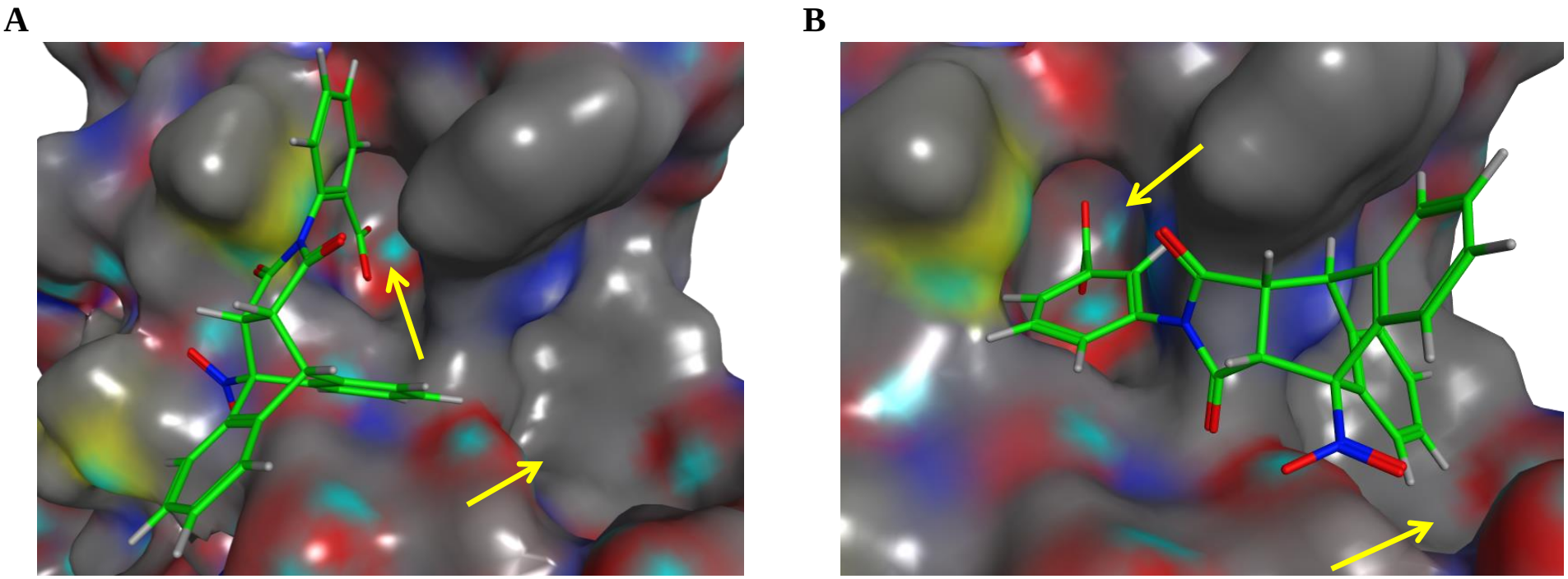
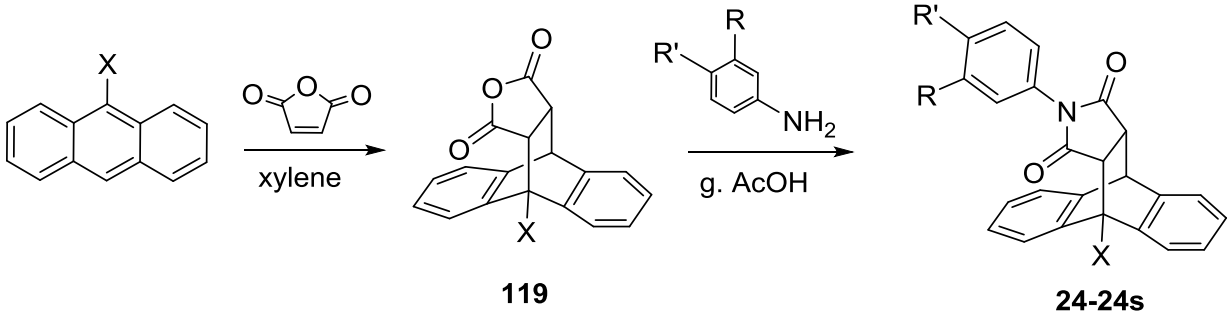



**Figure 4.3**: *Ortho*-substituted analogue of compound **24**. A) Modelling of the interaction of the *ortho*-substituted analogue of compound **24** (in green) shows the effect of *ortho*- substitution preventing interaction with the putative binding pockets of S100P (arrows). B) Predicted interaction of compound **24**, the carboxylic acid group in the *meta* position, with putative binding pockets of S100P (arrows).

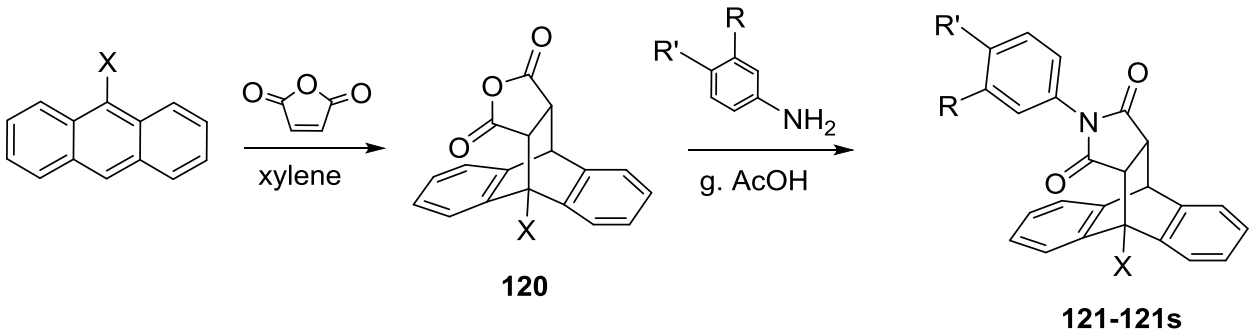
**Table 4.1**: Analogues of compound **24** where X = $NO_2$. Yields (from the Diels–Alder step), melting point (M.p.) and compound appearance are shown.




| Compound | R | R' | Yield* (%) | M.p. (°C) | Appearance |
|---|---|---|---|---|---|
| **24** | $CO_2H$ | H | 80 | 271-272 | white crystals |
| **24a** | $CO_2Et$ | H | 84 | 202-203 | white flakes |
| **24b** | F | H | 81 | 241-243 | cream flakes |
| **24c** | Cl | H | 82 | 230-232 | cream solid |
| **24d** | $NO_2$ | H | 93 | 240-242 | pale yellow powder |
| **24e** | H | H | 72 | 110-112 | cream flakes, static |
| **24f**† | $CH_3$ | H | 91 | 209-211 | cream powder |
| **24g**† | MeO | H | 72 | 262-265 | brown solid |
| **24h**† | $Me_2N$ | H | 97 | 213-215 | cream powder |
| **24i**† | CN | H | 97 | 310-311 | cream powder |
| **24j** | H | $CO_2H$ | 80 | 344-345 | cream flakes |
| **24k** | H | $CO_2Et$ | 78 | 255-256 | white flakes |
| **24l** | H | F | 88 | 223-225 | cream solid |
| **24m** | H | Cl | 91 | 270-273 | cream flakes, static |
| **24n** | H | $NO_2$ | 78 | 316-318 | fluffy cream powder |
| **24o**† | H | $CH_3$ | 75 | 258-262 | white solid |
| **24p**† | H | MeO | 85 | 249-250 | grey-purple solid |
| **24q**† | H | $Me_2N$ | 74 | 265-266 | light grey powder |
| **24r** | H | I | 84 | 305-308 | white crystalline powder |
| **24s** | H | *t*-Bu | 82 | 238-240 | cream solid |

†Compounds synthesised by Kulikowska [372] and Grewal [373] as part of their final year project for the degree of Master of Pharmacy, University of Hertfordshire. *Yields shown are from the reaction of maleic-anthracene adduct and the amine.

**Table 4.2**: Analogues of compound **24** where X = H. Yields (from the Diels–Alder step), melting point and compound appearance are shown.




| Compound | R | R' | Yield* (%) | Mp (°C) | Appearance |
|---|---|---|---|---|---|
| **121** | $CO_2H$ | H | 95 | 285-288 | white solid |
| **121a** | $CO_2Et$ | H | 96 | 205-207 | pale pink powder |
| **121b** | F | H | 87 | 228-230 | white flakes |
| **121c** | Cl | H | 92 | 238-240 | cream flakes |
| **121d** | $NO_2$ | H | 95 | 282-285 | white powder |
| **121e** | H | H | 83 | 211-215 | white solid |
| **121f†** | $CH_3$ | H | 61 | 199-200 | cream solid |
| **121g†** | MeO | H | 91 | 200-201 | sand-coloured solid |
| **121h†** | $Me_2N$ | H | 93 | 228-230 | cream solid |
| **121i†** | CN | H | 99 | 262-264 | white solid |
| **121j** | H | $CO_2H$ | 86 | 354-355 | white powder |
| **121k** | H | $CO_2Et$ | 92 | 220-222 | white powder |
| **121l** | H | F | 96 | 250-253 | pinkish solid |
| **121m** | H | Cl | 94 | 275-278 | fluffy white powder |
| **121n** | H | $NO_2$ | 83 | 294-295 | cream crystals |
| **121o†** | H | $CH_3$ | 94 | 220-222 | white powder |
| **121p†** | H | MeO | 87 | 240-243 | light purple solid |
| **121q†** | H | $Me_2N$ | 96 | 243-245 | light purple powder |
| **121r** | H | I | 89 | 303-307 | white solid |
| **121s** | H | *t*-Bu | 93 | 268-271 | white solid |

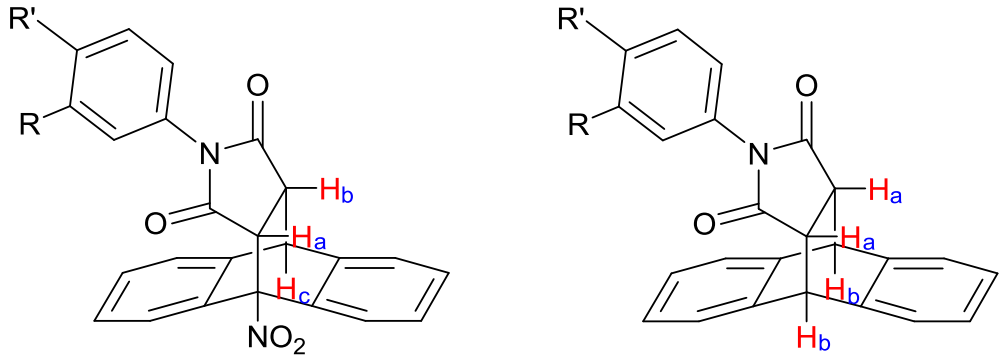
†Compounds synthesised by Kulikowska [372] and Grewal [373] as part of their final year project for the degree of Master of Pharmacy, University of Hertfordshire. *Yields shown are from the reaction of maleic-anthracene adduct and the amine.

Similar to compound **24**, analogues with the X = $NO_2$ substituent (Table 4.1) exhibited three distinct signals upfield in their $^1$H-NMR spectra due to asymmetry from the nitro group, which coupled to one another. Where defined, the first signal appeared as a doublet of doublets (dd) in the region δ = 3.53-3.76 ppm with coupling constants of ~ 9.0 and 3.0 Hz. This signal comes from proton $H_b$ of the succinimide moiety (Figure 4.4) and is split into a doublet of doublets by protons $H_a$ and $H_c$ respectively (Section 4.5.2.1.1). The large *J* coupling constant (*J* ~ 9.0 Hz) comes from the adjacent vicinal proton $H_a$ of the succinimide moiety whose signal, a doublet, lies downfield between δ = 4.4-4.6 ppm. Proton $H_c$ of 9,10-dihydroanthrancene appears as a doublet when the peak is defined, at around δ 5.0-5.1 ppm with a coupling constant of *J*~3.0 Hz from proton $H_b$. These coupling constant values are comparable to those reported for similar 9-substituted Diels–Alder anthracene-cycloadducts [360].

Anthracene-derived analogues (**121-121s**), where X = H, in comparison to the corresponding nitro derivatives, only had two proton peaks from protons $H_a$ and $H_b$ (Figure 4.4) on their $^1$H-NMR spectra from the newly formed six-membered ring of the Diels–Alder cycloadduct (Section 4.5.2.2) due to symmetry. These signals are in agreement with similar anthracene-maleic anhydride-derivatives published in the literature [365, 366, 369, 370, 374]. Both peaks appear as multiplets, with the first peak occurring between δ = 3.34-3.51 ppm. This signal originates from protons $H_a$ on the succinimide moiety. The second signal lies upfield between δ = 4.85-4.89 and comes from protons $H_b$ of the newly formed 5,10-dihydroanthracene cycloadduct. No coupling

constants were available for these signals as they appeared as multiplets (Section 4.5.2.2.1).

In the $^{13}$C-NMR spectra of the 9-nitroanthracene analogues (**24-24s**), the two succinimide carbonyl carbon atoms appear as two signals between δ = 175 and 172 ppm, whereas in the anthracene-derived analogues (**121-121s**), they are appear as a single signal in the same region due to symmetry (Section 4.5.2.2.1). In compounds **24-24s**, there are 18 aromatic carbon signals (6 carbon signals per aromatic ring) on the $^{13}$C-NMR spectra. In comparison, compounds **121-121s** exhibited 12 carbon signals (6 from the *N*-substituted phenyl ring and 6 from the aromatic anthracene rings) in their $^{13}$C-NMR spectra due to symmetry.




**24-24s** **121-121s**

**Figure 4.4**: Signals from $H_a$, $H_b$ and $H_c$. Three distinct signals were observed for protons $H_a$, $H_b$ and $H_c$ in the $^1$H-NMR spectra of all analogues with X = $NO_2$ (compounds **24-24s**), while only two distinct signals were observed for the corresponding analogues where X = H (compounds **121-121s**).

Infrared (IR) analysis of all the nitro analogues (X = $NO_2$, **24-24s**) show distinct sharp $NO_2$ peaks at around 1550 $cm^{-1}$ (asymmetric stretch) and 1340 (±50) $cm^{-1}$ (symmetric stretch). Melting points for the analogues varied, with **24j** (X = $NO_2$, R = H, R′ = $CO_2H$) and its corresponding analogue **121j** (X = H, R = H, R′ = $CO_2H$) recording the highest melting points between 354-355 °C (Table 4.1 and 4.2 respectively). Both compounds are *para*- substituted and exhibited a considerable difference, greater than 50 °C, in their melting points compared to the *meta*- substituted isomers. This difference is not surprising as the presence of symmetry in organic compounds has been reported to correlate with high melting points [375, 376]. The high melting point in symmetrical molecules is believed to be due to the more compact arrangement of the molecules in the crystal lattice which in turn requires higher energy to disrupt [376, 377]. Indeed, of the 32 analogues synthesised that have both *para*- and *meta*-isomers, all but three of the *para*-isomers, **24b**, **24g**, and **121d**, have higher melting points than their corresponding *meta*- substituted counterparts. It is not known what is responsible for this disparity in the melting points of the three isomers, but cases where *meta*- or *ortho*- substituted isomers have a higher melting point than their *para*-substituted have been reported [378].

Although most of the analogues are not easily ionisable, except for those with acidic groups such as compounds **24**, **24j**, **121** and **121j**, liquid chromatography-mass spectrometry (LC-MS) analysis was still able to pick up the molecular ion, either in negative or positive mode. In the LC-MS spectra of intermediates **119** and **120**, an *m/z*

of + 31 was observed for both (Sections 4.5.2.1 and 4.5.2.2). These compounds were dissolved in methanol and the *m/z* indicates formation of a methanol-adduct.

### 4.2.5 Analogue synthesis for high-throughput biological screening

The rationale behind the synthesis of the above analogues was to create a library of compounds similar to compound **24** that could be used in high-throughput biological screening. By varying the R and R′ groups on the phenyl ring, the effect of substituents in position 3 or 4 on biological activity could be compared and contrasted for further studies such as during structure-activity relationship (SAR) studies. By replacing the carboxylic acid group in **24** with groups such as H, $CO_2Et$, Cl, F, MeO, CN, $NO_2$, I, *t*-Bu, and $Me_2N$, a variety of groups with different physicochemical properties to the acid are sampled. For example, replacing the $CO_2H$ group with a $CH_3$ group removes the potential for hydrogen bonding and potential ionic interactions between the compound and the target protein. However, the presence of the hydrophobic $CH_3$ group could enhance hydrophobic interactions with a lipophilic region of the protein that would otherwise not have been possible with the acid group. Substitution at the *para*- position provides a series of compounds that could be compared with their *meta*- substituent counterparts, to assess whether substituent position on the phenyl ring impacts on biological activity. For SAR purposes, the *para*-substituted substituents synthesised lie in different quadrants of the Craig plot (Figure 4.5) compared to the carboxylic acid group. The Craig plot [379] is a two-dimensional map that correlates two physicochemical properties, the Hammett substituent constant ($\sigma$) and the substituent hydrophobicity constant ($\pi$), for *para*-substituted groups on a phenyl ring. The Hammett

substituent constant provides a measure of a substituent's electronic properties while the substituent's π-value provides its hydrophobicity relative to hydrogen [201]. This information is useful as a guide to synthesising more compounds based on biological activity and calculation of which properties may need improving. Sixteen of the analogues synthesised (**24l**-**24s**, **121l**-**121s**) have substituents which are included in the Craig plot (highlighted in Figure 4.5) which should aid in future SAR calculations.

By synthesising analogues with different substituents and physicochemical properties, it is hoped that high-throughput biological screening will facilitate the process of identifying which groups of compounds could be progressed for further optimisation studies in the search for a therapeutic agent against pancreatic cancer.

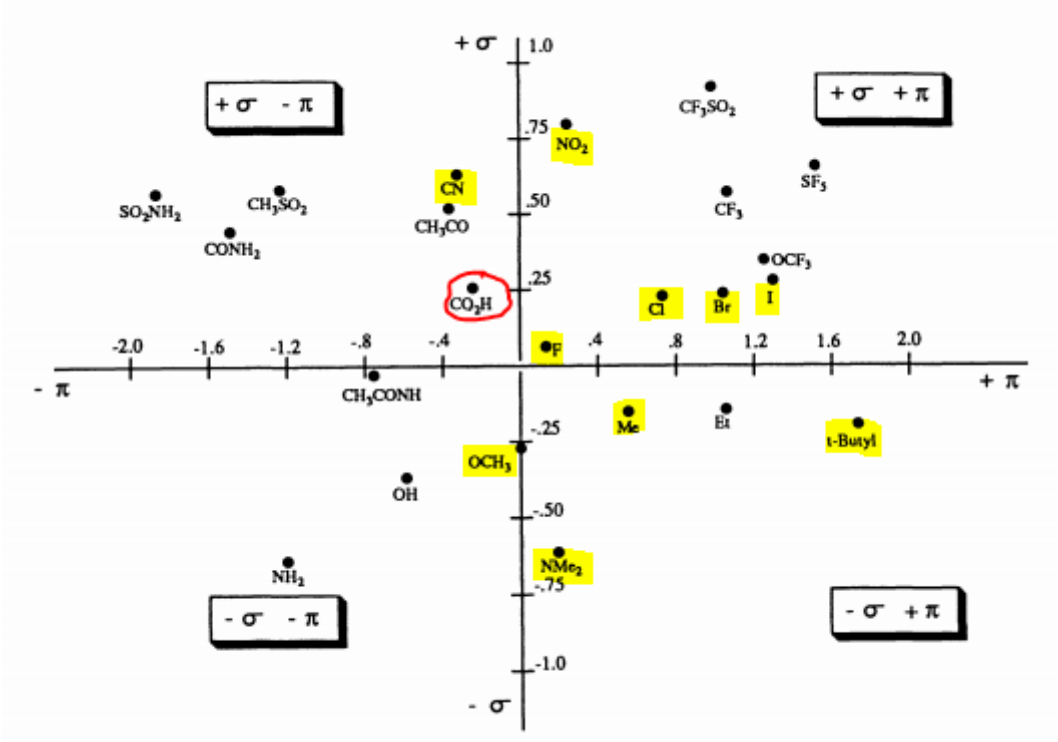



**Figure 4.5**: Craig plot showing σ and π values for *para*-substituted phenyl substituents. Sixteen of the synthesised analogues (**24l-24s**, **121l-121s**) have substituents which are included in the Craig plot. These substituents (highlighted in yellow) have σ and π properties that vary from the carboxylic acid group (circled in red). (Graph adapted from Patrick [201]).

### 4.2.6 Selecting analogues for biological screening

Five compounds including the "hit" compound **24** were selected, from the 40 synthesised, for biological screening against pancreatic cancer cells (Figure 4.6, Table 4.3). Compound **24a** is the ester analogue of **24**, and has a similar molecular weight to cromolyn; compound **24e** on the other hand lacks a substituent on the phenyl ring, and was selected to examine the absence of the carboxylic acid group on biological activity. Compound **24j**, the *para*-substituted analogue of **24**, is selected to examine the effect of substituent position on biological activity, and compound **121**, which lacks a $NO_2$ group, was selected to assess whether this group is necessary for activity.

**Table 4.3**: Physicochemical properties of cromolyn and compounds selected for biological screening. MW: Molecular Weight, tPSA: total Polar Surface Area, Log S: aqueous solubility, clog P: octanol-water partition coefficient, pKa: acid dissociation constant. Properties calculated in MOE

| Compound | MW | tPSA | Log S | clog P | pKa |
|---|---|---|---|---|---|
| **Cromolyn** | 468.37 | 165.89 | -4.24 | 1.48 | 2.58 |
| **24** | 440.41 | 126.49 | -5.21 | 2.80 | 3.84 |
| **24a** | 468.13 | 115.49 | -5.93 | 3.74 | - |
| **24e** | 396.40 | 89.19 | -5.26 | 2.74 | - |
| **24j** | 440.41 | 126.49 | -5.34 | 2.80 | 3.71 |
| **121** | 395.41 | 74.68 | -4.66 | 3.10 | 3.87 |

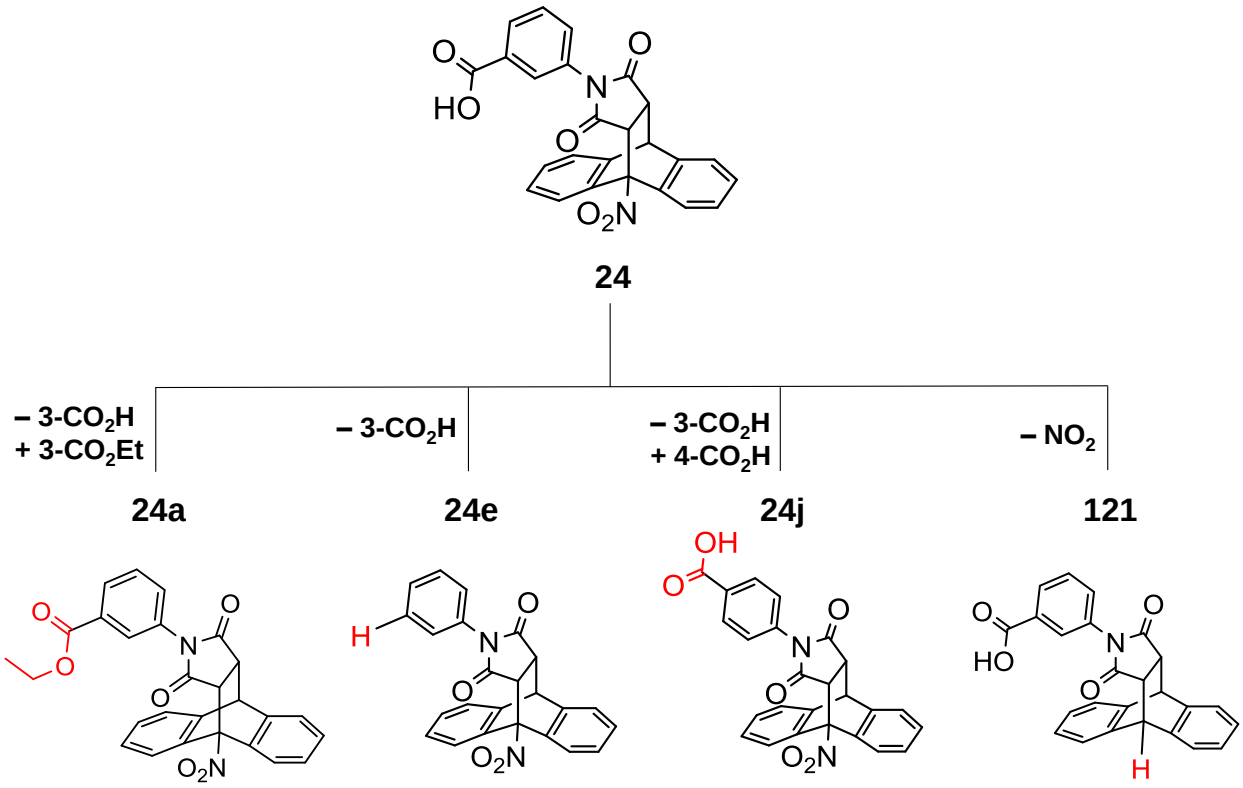



**Figure 4.6**: Compounds selected for biological screening. Highlighted in red are the differences between analogues and "hit" compound **24**.

## 4.3 BIOLOGICAL SCREENING

### 4.3.1 Growth of pancreatic cancer cells

The response of pancreatic cancer cells BxPC-3 and Panc-1 (Figure 4.7) in growth culture medium was assessed prior to screening the effect of compound **24** and analogues against them. A study of the growth of both cell lines for optimal cell proliferation studies was carried out using a 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2*H*-tetrazolium (MTS)-based assay (Promega, Southampton, UK). This assay exploits the presence of a mitochondrial dehydrogenase enzyme in metabolically active viable cells that reduces the salt to a coloured soluble formazan dye product which is then determined spectrophotometrically [335]. In short, a high absorbance reading corresponds to a higher metabolic activity, which can also be interpreted as an increase in cell numbers. This will be discussed further in the sections below.

It was found that both cell lines grow optimally when seeded at a density of 1 x $10^4$ cells/well in 96-well plates (Figure 4.8). At this density, the cells grew exponentially for a period of 120 hours (5 days) before reaching a plateau and entering the death phase around day 7 (Figure 4.8). Furthermore, at a density of 1 x $10^4$ cells/well, most of the cells were attached overnight (~23 hours, Figure 4.9). These cells are adherent cell lines and as such need to attach to grow. For screening purposes, seeding at 1 x $10^4$ cells/well seemed to provide the best seeding density for attachment during the exponential growth phase. The growth characteristics of the cells when seeded at 1 x $10^3$ cells/well are

similar to that at 1 x $10^4$ cells/well. However, due to the lower cell number per surface area per well, the cells at the former density took longer to attach and divide compared to the latter (Figure 4.9), which is most likely due to the limited cell-to-cell interaction and signalling at the lower concentration. At a density of 1 x $10^5$ cells/well, both BxP-3 and Panc-1 cells showed a rapid decline in cell number (i.e. metabolic activity) possibly due to cell death caused by the cells competing for limited nutrients in 100 μL of growth medium, lack of growing area that they require for adherence as well as exposure to a cocktail of by-products/metabolites/intracellular enzymes excreted by the other growing or dead cells (Figure 4.8 and 5).

In conjunction with the MTS assay, the CytoTox-ONE™ Homogeneous Membrane Integrity Assay (Promega, Southampton, UK) was used to determine viability of both cell lines during their growth. However, the results obtained were found to be, inconclusive. This assay relies on the release of the enzyme lactate dehydrogenase (LDH) from compromised cell membranes of dead cells into the culture medium which is then measured fluorometrically [380]. The presence of foetal bovine serum (FBS, 10% v/v) in the growth medium has been reported to interfere with this particular assay due to endogenous LDH in FBS [381, 382] yielding a high signal-to-noise ratio. In this study, the signal-to-noise ratio was still high even after accounting for background noise from media-only controls. As the assay was carried out in multiplex with the MTS assay, which performed satisfactorily in the presence of FBS, lowering the serum concentration to optimise it for LDH release studies would require assessing the

response of the cells to such conditions which would be counter to the serum concentration for optimal cell growth as recommended by the vendor.

In addition to possible serum interference, phenol red present in culture media has also been reported to interfere with the LDH assay [383]. Indeed, many papers have reported carrying out LDH studies using serum-free/reduced serum and phenol red-free culture medium [384-386]. This is an avenue that could be pursued in future if using this assay to determine the cytotoxicity of the compounds screened.

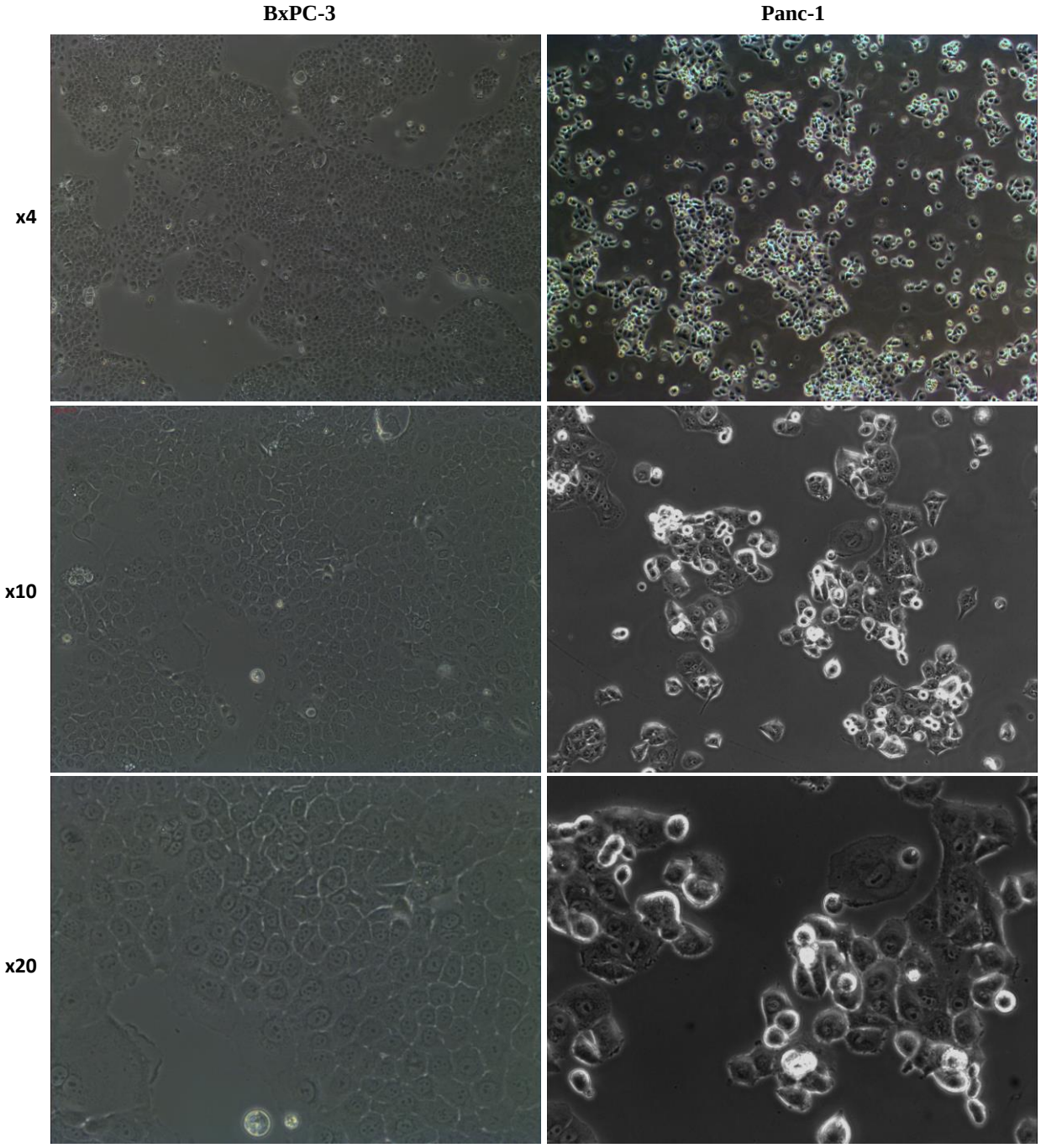
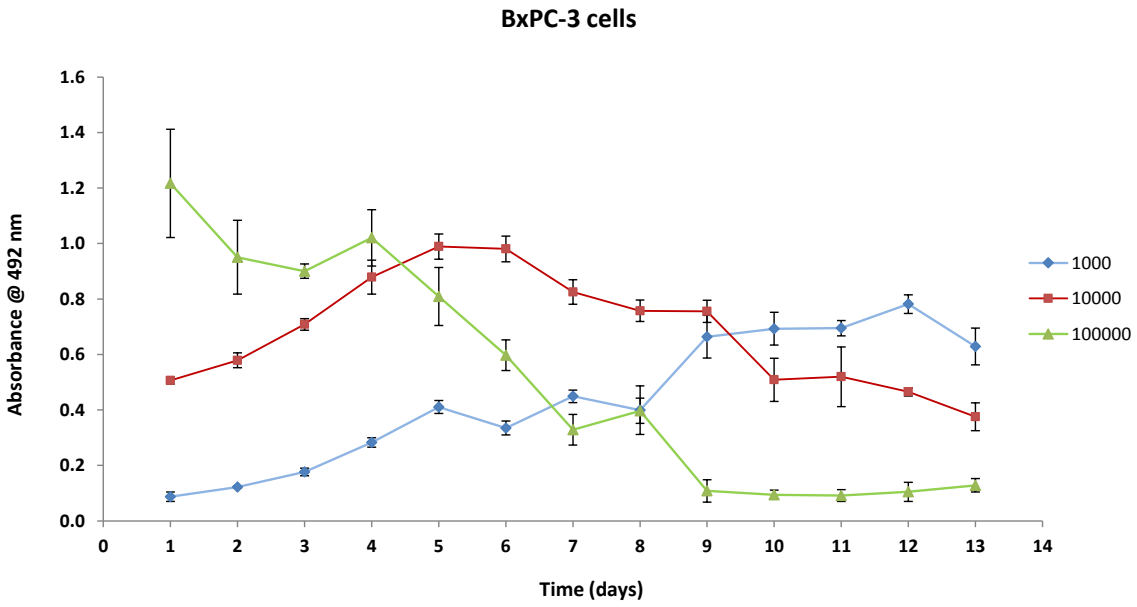



**Figure 4.7**: Pancreatic cancer cells BxPC-3 and Panc-1 adhered at the bottom of T-75 $cm^2$ flasks. Cells were cultured in complete growth medium supplemented with FBS (10% v/v), Penicillin/Streptomycin (200 U) solution, and L-glutamine (2 mM). Images taken with a GXCAM-9 digital microscope C-mount camera (GT Vision, Suffolk, UK) mounted on an Olympus CKX41 microscope at different magnifications.

**A**




**B**

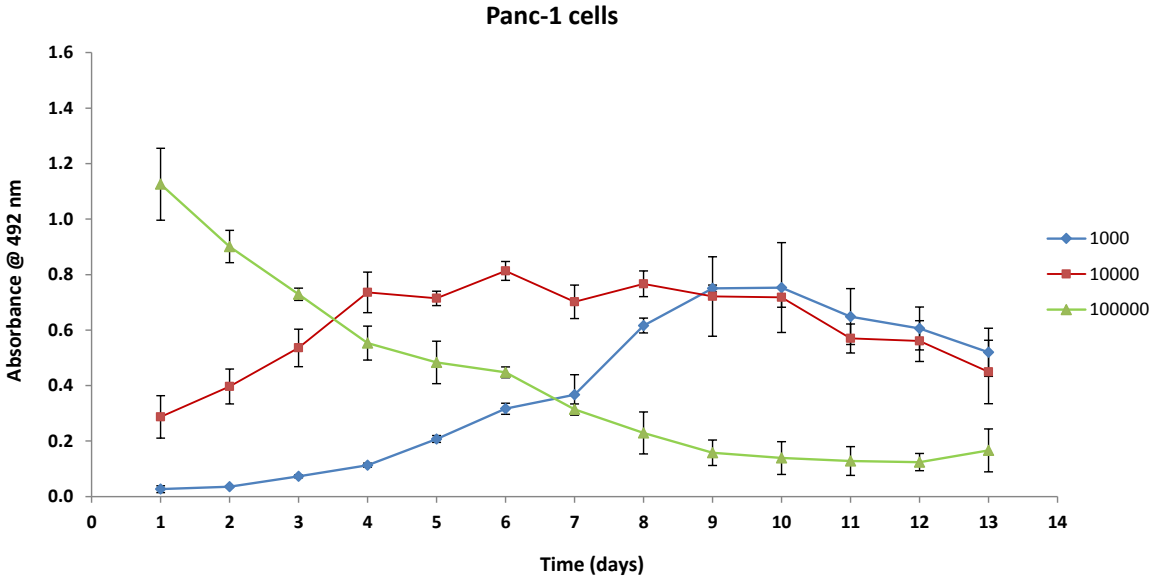



**Figure 4.8**: Batch growth curve for pancreatic cancer cells over a period of two weeks. A) BxPC-3 cells. B) Panc-1 cells. Cells were plated in 96-well plates at a density of 1000, 10,000 and 100,000 cells/well (100 µM) in complete cell culture growth medium. Samples were assessed using the MTS reagent (Promega) according to the manufacturer's instructions before reading absorbance at 492 nm using a Multiskan Ascent 96/384 plate reader (Thermo Scientific).

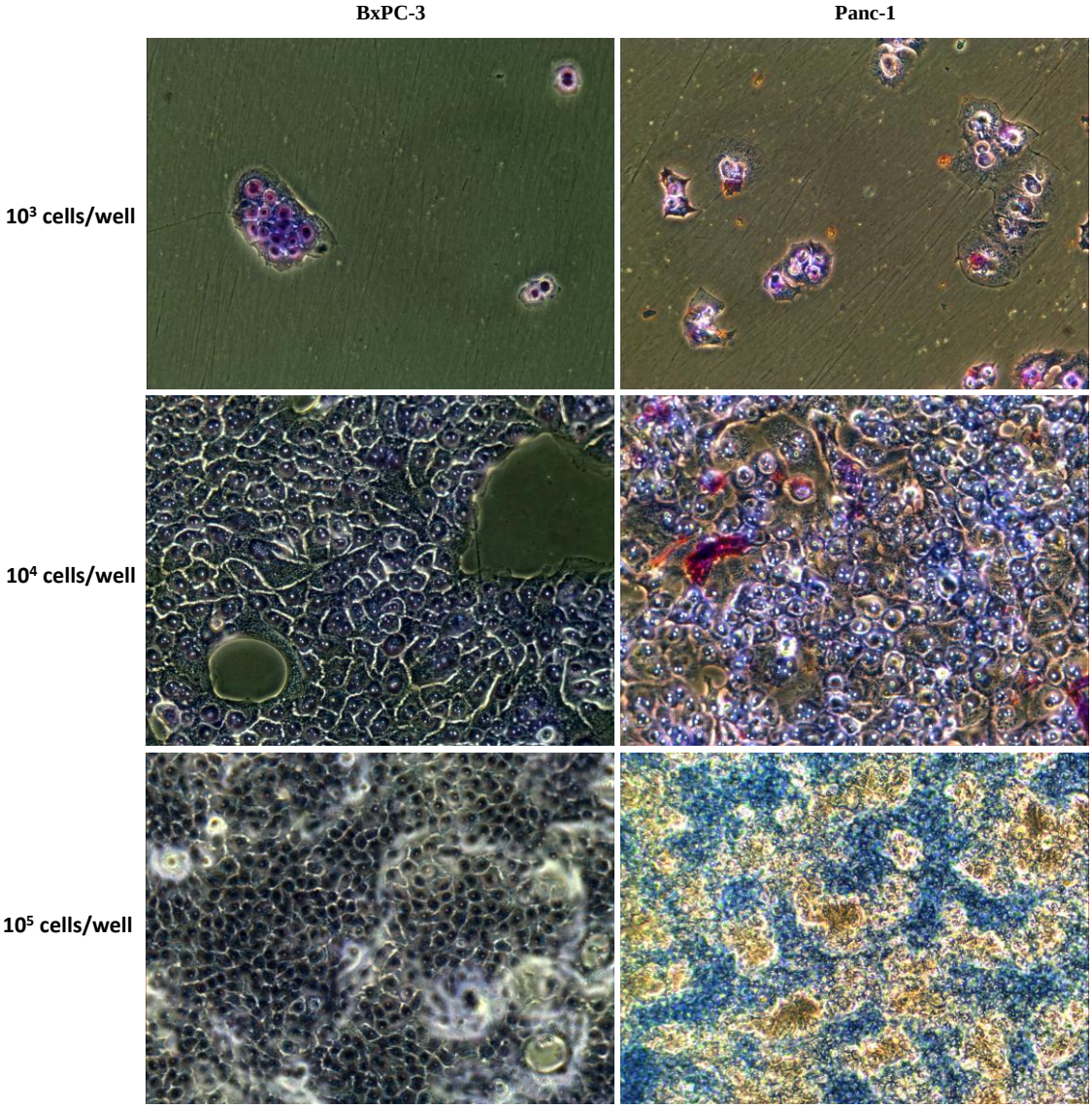



**Figure 4.9**: Representative images of attachment of pancreatic cancer cells 48 h after seeding in 96-well plates for batch growth curve studies. After reading MTS absorbance for proliferation studies, cells were washed with Phosphate Buffered Saline (PBS), fixed with paraformaldehyde (3.7% w/v) at room temperature for 15 minutes followed by another wash with PBS. After fixation they were permeabilised with Triton X-100 (0.01% v/v) for 15 minutes, washed with PBS before staining the cytoplasm and nuclei with May-Grünwald (0.25% v/v) and Giemsa (0.4% v/v) stains respectively. Images taken with a GXCAM-9 digital microscope C-mount camera (GT Vision, Suffolk, UK) mounted on an Olympus CKX41 microscope at x20 magnification (objective lens).

### 4.3.2 Effect of cromolyn (4), compounds 24, 24a, 24e, 24j and 121 on pancreatic cancer cells

#### 4.3.2.1 Proliferation studies

Cromolyn (**4**) has been reported to inhibit proliferation of pancreatic cancer cells by binding to S100P cells in both *in vitro* and *in vivo* models [195, 334, 387]. To determine this effect, pancreatic cancer cells BxPC-3 and Panc-1 were incubated with **4** at different concentrations and the effect on proliferation assessed using the MTS assay. Compound **4** was found to have no meaningful statistical effect on the proliferation of both these cell lines at low concentrations (1, 10 and 100 μM) compared to the no-treatment control ($p>0.05$, Figure 4.10). At the high concentration of 1000 μM, there was a significant effect on proliferation of both cell lines ($p<0.05$) although, on closer observation, this effect seems to be constant over the 72 h period. This effect could be a result of general toxicity of the compound to the cells from such a high concentration rather than a specific anti-proliferative effect. These findings are similar to results by Mahmoud [328] but contrary to those of Arumugam *et al*. [195]. The latter group observed no effect on the proliferation of Panc-1 cells from **4**, but Panc-1 cells transfected with S100P were found to be inhibited by the compound at 100 μM after 48 h. In the same work, they also reported that **4** inhibited proliferation of BxPC-3 cells at 10 and 100 μM ($p = 0.002$ and $p<0.001$ respectively). Since the current studies were not carried out in exactly the same way as Arumugam *et al*.'s work, it is not known if the discrepancies observed herein are due to disparate methodologies.

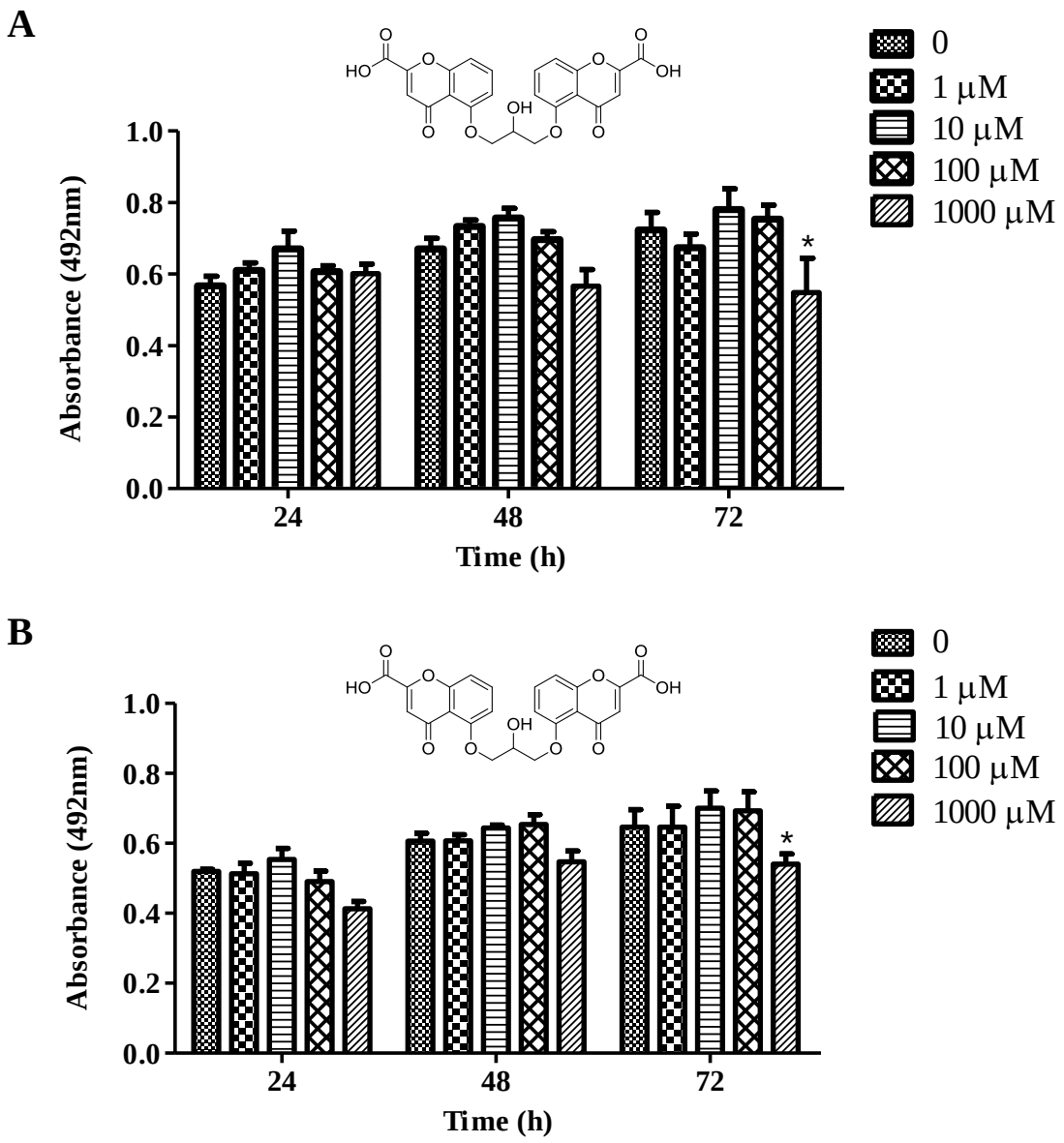



**Figure 4.10**: Effect of cromolyn (**4**) on proliferation of pancreatic cancer cells BxPC-3 (A) and Panc-1 (B). Cells were seeded overnight (1 x $10^4$ cells/well in 100 μL) in 96-well plates in complete medium supplemented with FBS (10% v/v), L-glutamine (2 mM) and Penicillin/Streptomycin (200 U) solution. Cromolyn was added to the wells and cell proliferation assessed using the MTS assay after 24, 48, and 72 h. Results are expressed as mean + SEM of three independent experiments with n = 12. P value of <0.05(*) was determined relative to control using 2-Way ANOVA. Post-test comparisons were made using Bonferroni test at 95% confidence interval.

In Arumugam *et al.*'s proliferation studies, they plated 96-well plates at a density of 1 x $10^3$ cells/well, compared to 1 x $10^4$ cells/well used in the present study, and they measured the effect of **4** on the cells' proliferation after 48 h [195]. Other than the differences in seeding density, the discrepancies in the effect of **4** on these cells between this work and Arumugam *et al.*'s is something that warrants further investigation.

When compound **24** was screened for anti-proliferation properties on BxPC-3 cells, there was a significant inhibition of growth relative to control at all three concentrations (1, 10 and 100 μM) after 72 h ($p$<0.0001, Figure 4.11). However, a similar effect was also observed on the Panc-1 cells, casting doubt upon a solely S100P-mediated effect. Compound **24a**, which has an ester group in place of the carboxyl group, had a more pronounced anti-proliferative effect on BxPC-3 cells after 24 h compared to compound **24** (Figure 4.12A). This initial decrease in cell metabolic activity may be due to exposure of the cells to a foreign stimulus resulting in a lower proliferative ability. After 48 h, the cells appear to be recovering albeit only slightly compared to the control. A dose-dependent decrease in proliferation was observed for all three concentrations (1, 10 and 100 μM) after 72 h with the highest effect seen at 100 μM. This concentration-dependent effect of **24a** was absent on Panc-1 cells, although the compound still interestingly exerted an effect on proliferation at all three concentrations (1, 10 and 100 μM) relative to control (Figure 4.13A).

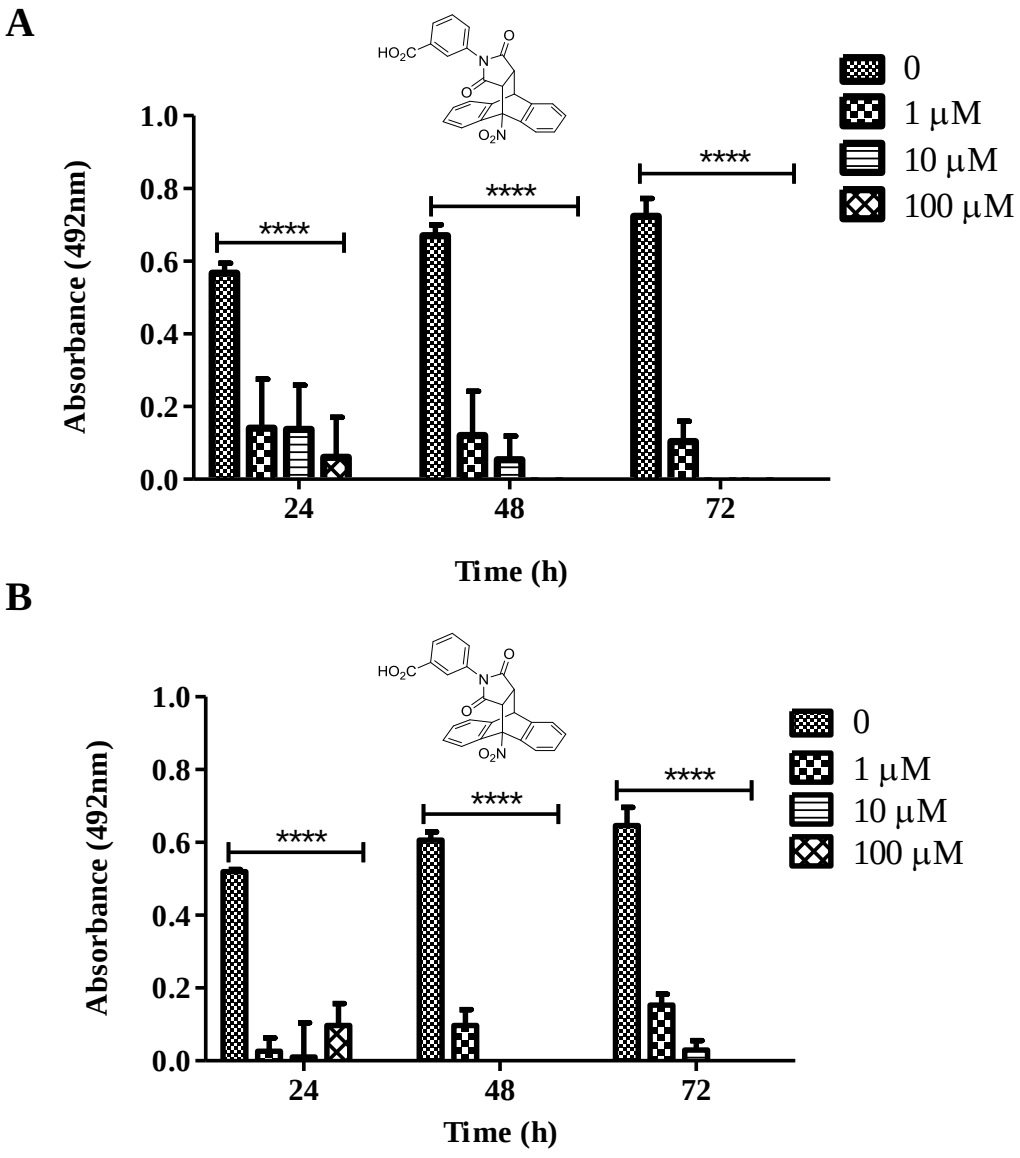



**Figure 4.11**: Effect of compound **24** on proliferation of pancreatic cancer cells BxPC-3 (A) and Panc-1 (B). Cells were seeded overnight (1 x $10^4$ cells/well in 100 μL) in 96-well plates in complete medium supplemented with FBS (10% v/v), L-glutamine (2 mM) and Penicillin/Streptomycin (200 U) solution. Compound **24** was added to the wells and cell proliferation assessed using the MTS assay after 24, 48, and 72 h. Results are expressed as mean + SEM of three independent experiments with n = 12. P value of <0.0001 (****) was determined relative to control using 2-Way ANOVA. Post-test comparisons were made using Bonferroni test at 95% confidence interval.

A similar dose-dependent effect on proliferation of BxPC-3 was seen with **24e** after 72 h (Figure 4.12B). This analogue lacks a substituent on the phenyl ring when compared to **24** and **24a**. Interestingly, the effect on proliferation of Panc-1 cells closely resembled that exerted by **24a**, the ester analogue. At 24 h, the anti-proliferative effect on Panc-1 cells was striking but the cells seem to be recovering after 72 h albeit marginally relative to control (Figure 4.13B).

The regioisomer of compound **24**, compound **24j**, with the carboxyl group in the *para* position produced a more pronounced anti-proliferative effect on BxPC-3 cells at 1 and 10 μM compared to **24** where the acid group is in the *meta* position (Figure 4.12C). Although still viable, as assessed by trypan blue counting, no metabolic activity was observed for the BxPC-3 cells at 1 μM after 72 h, making **24j** the most potent inhibitor at this concentration in comparison to **24**, **24a** and **24e**. Similar effects were observed on the proliferation of Panc-1 cells calling into question whether the mechanism by which this compound is acting is S100P-mediated (Figure 4.13C).

Compound **121**, which lacks the $NO_2$ group, inhibited proliferation of BxPC-3 cells at low concentrations, 1 and 10 μM, relative to control (Figure 4.12D). Again, similar to compounds **24**, **24a**, **24e** and **24j**, inhibition was also observed against Panc-1 cells (Figure 4.13D).

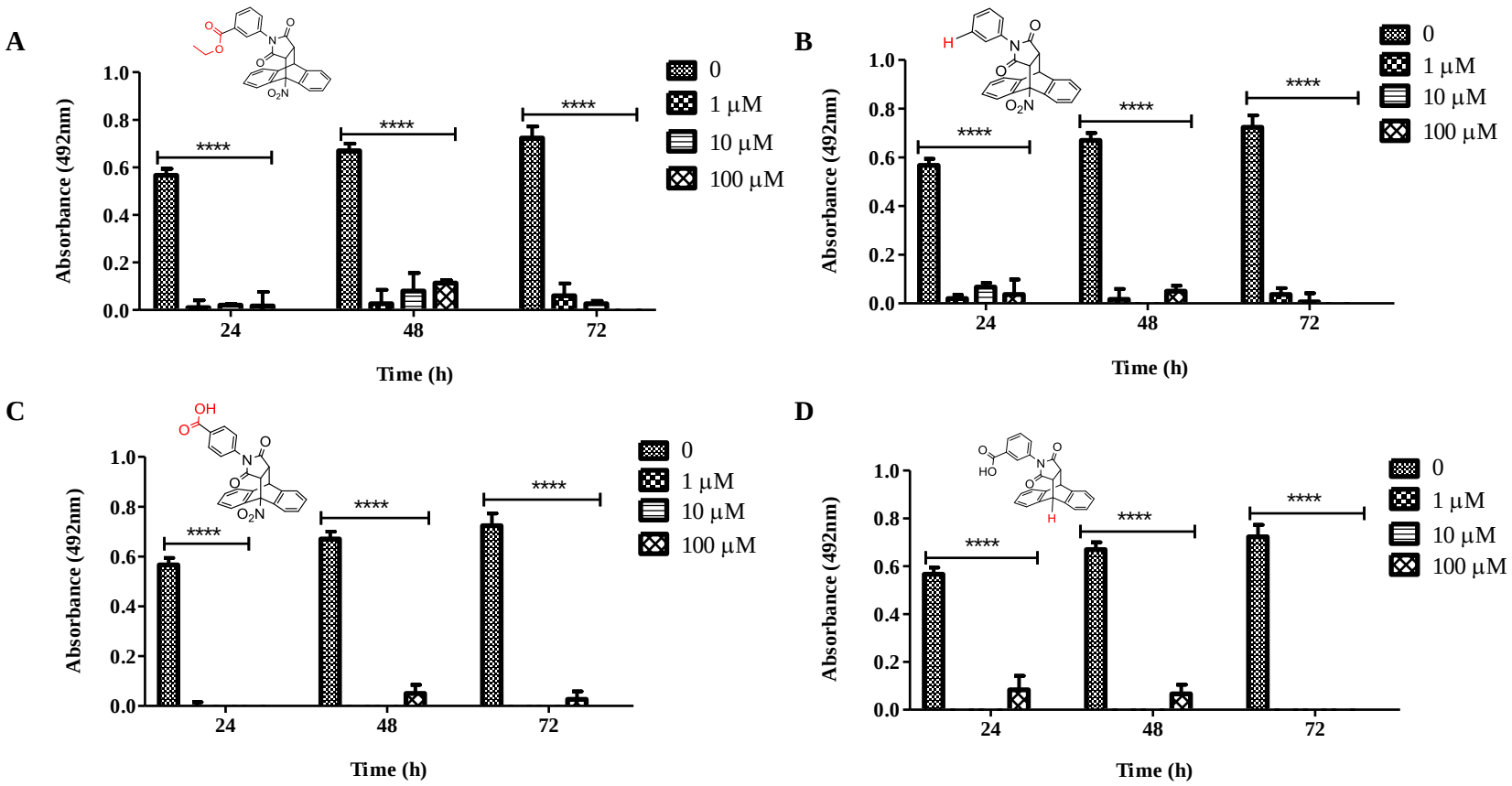



**Figure 4.12**: Effect of hit compounds on proliferation of BxPC-3 cells. A) Compound **24a**. B) Compound **24e**. C) Compound **24j**. D) Compound **121**. Cells were seeded overnight (1 x $10^4$ cells/well in 100 µL) in 96-well plates in complete medium supplemented with FBS (10% v/v), L-glutamine (2 mM) and Penicillin/Streptomycin (200 U) solution. Compounds were added to the wells and cell proliferation assessed using the MTS assay after 24, 48, and 72 h. Results are expressed as mean + SEM of three independent experiments with n = 12. P value of <0.0001 (****) was determined relative to control using 2-Way ANOVA**.** Post-test comparisons were made using Bonferroni test at 95% confidence interval

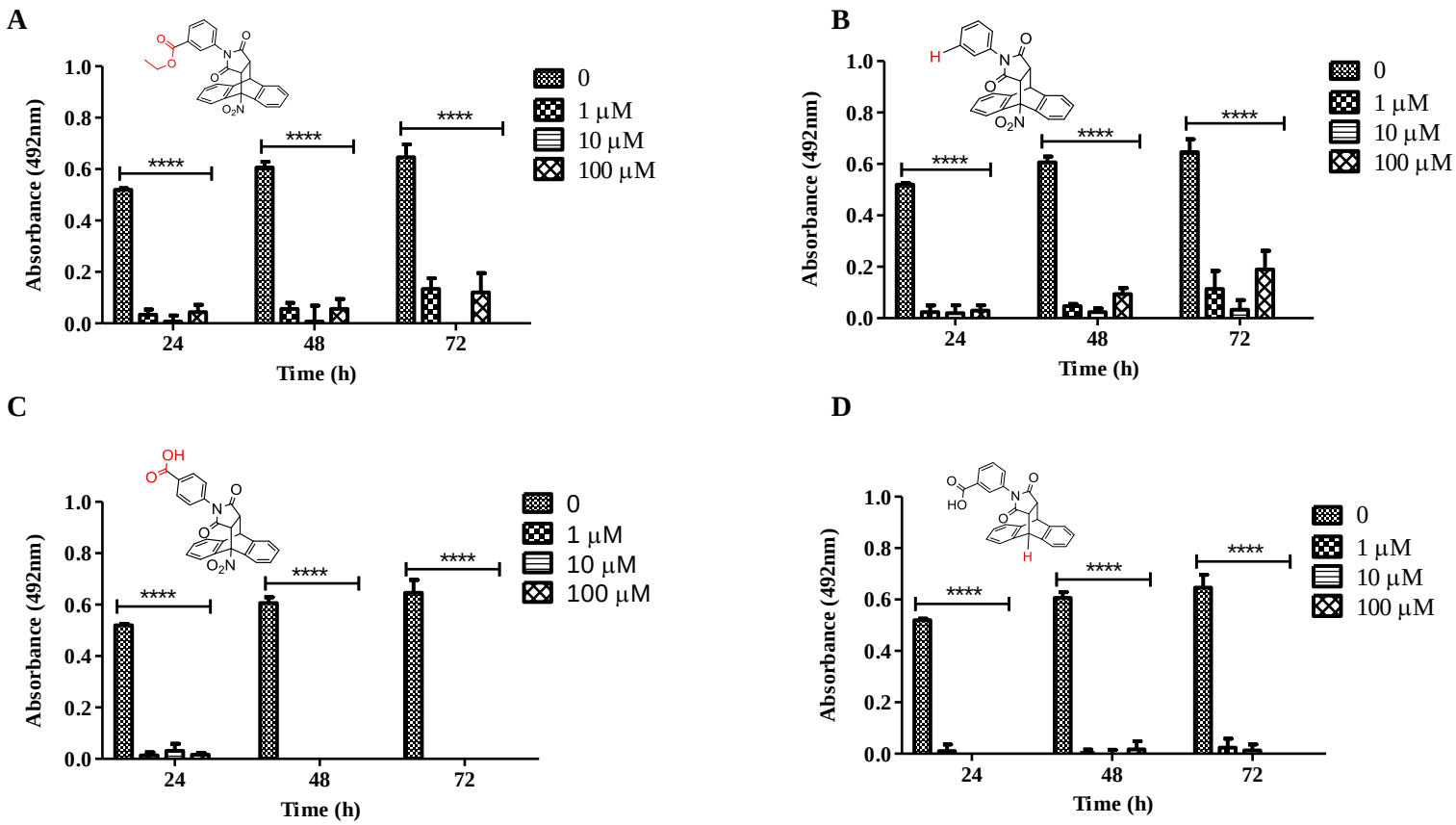



**Figure 4.13**: Effect of hit compounds on proliferation of Panc-1 cells. A) Compound **24a**. B) Compound **24e**. C) Compound **24j**. D) Compound **121**. Cells were seeded overnight (1 x $10^4$ cells/well in 100 µL) in 96-well plates in complete medium supplemented with FBS (10% v/v), L-glutamine (2 mM) and Penicillin/Streptomycin (200 U) solution. Compounds were added to the wells and cell proliferation assessed using the MTS assay after 24, 48, and 72 h. Results are expressed as mean + SEM of three independent experiments with n = 12. P value of <0.0001 (****) was determined relative to control using 2-Way ANOVA. Post-test comparisons were made using Bonferroni test at 95% confidence interval.

Despite the effect of these compounds on the proliferation of Panc-1 cells, which casts doubt on an S100P-mediated effect on BxPC-3 cells, it is encouraging that the compounds are exerting some anti-proliferative effects at low concentrations of 1 and 10 μM relative to control. What will be interesting to see is how the cells respond to the compounds at nanomolar concentrations, a task that was not carried out in this study due to time limitations, as this would provide an opportunity to further optimise them to improve their potency. Ideally, high potency is desired for a drug candidate at low concentrations in order to avoid unwanted and potentially adverse side effects. Low-concentration dosing could also be beneficial to solubility and stability of the compounds, important properties that are useful further down the drug development and formulation stages [388].

There were some solubility issues in the aqueous-based culture media for compounds **24a**, **24e** and **24j** at 100 and 1000 μM (Figure 4.14). Although soluble in dimethyl sulfoxide (DMSO), a precipitate was observed for all three compounds at 100 and 1000 μM once they were added to wells containing the cells in culture media. These precipitates sometimes resulted in erroneously high MTS absorbance readings, for e.g. compound **24a** at 100 μM for Panc-1 cells (Figure 4.13B). To identify the source of the abnormally high MTS readings at these concentrations, the compounds were added to culture media (both RPMI and DMEM) in the absence of cells in 96-well plates before adding the MTS reagent and reading the absorbance. As suspected, there were high absorbance readings after addition of the MTS reagent to these three compounds in culture medium (Figure 4.15). Visual analysis of images of **24j** taken before and after

addition of the MTS reagent showed what appeared to be salt deposits on the crystals of the compound, suggesting formation of a complex between the reagent and the compound (Figure 4.15B and C). Since these solubility issues were only encountered at high concentrations (100 and 1000 μM), these concentration values will be omitted in future studies.

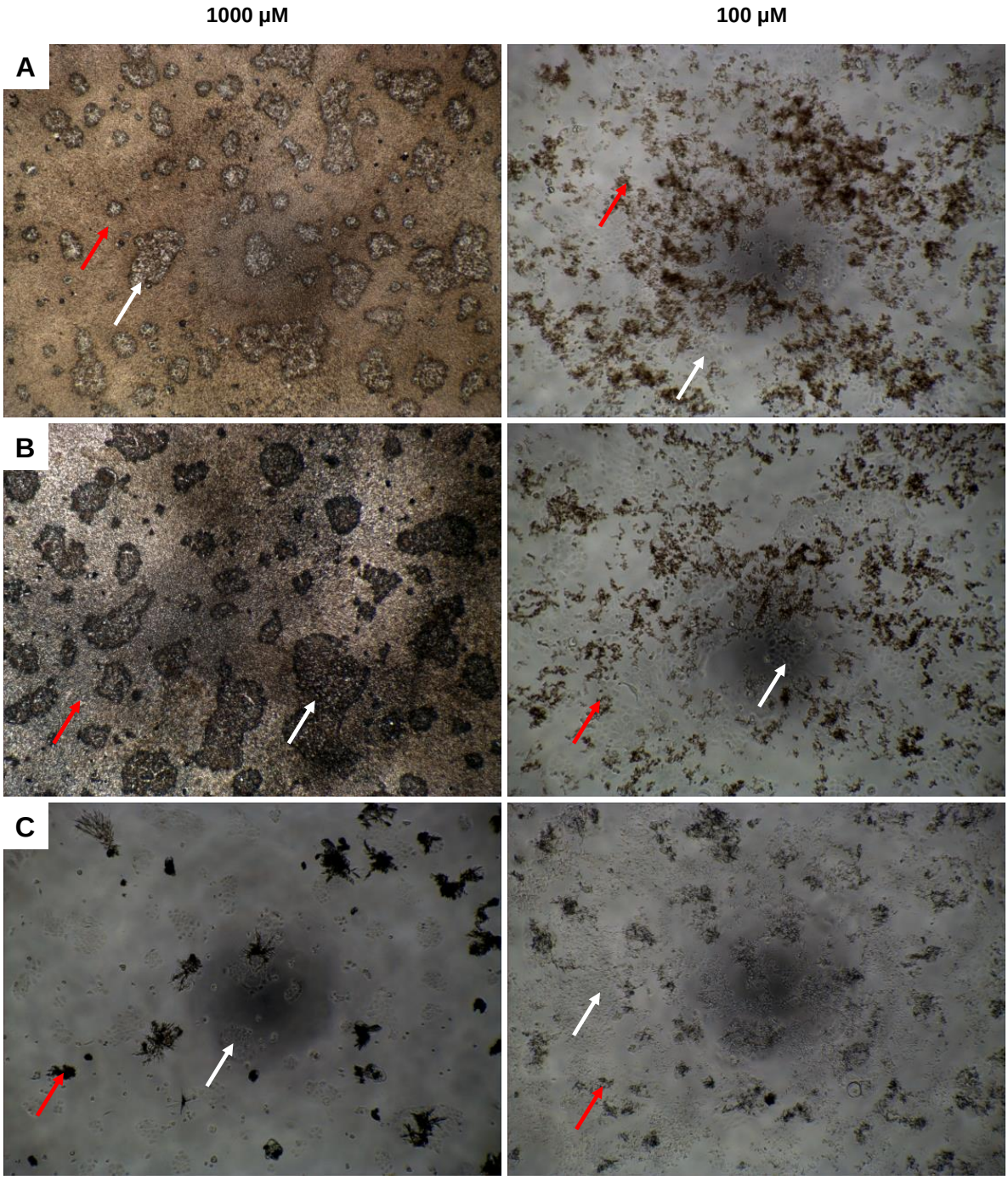



**Figure 4.14**: Precipitation of compounds (red arrows) when added to BxPC-3 cells (white arrows) in RPMI-1640 culture medium in 96-well plates. A) **24a**. B) **24e**. C) **24j**. Images taken with a GXCAM-9 digital microscope C-mount camera (GT Vision, Suffolk, UK) mounted on an Olympus CKX41 microscope at x4 magnification (objective lens).

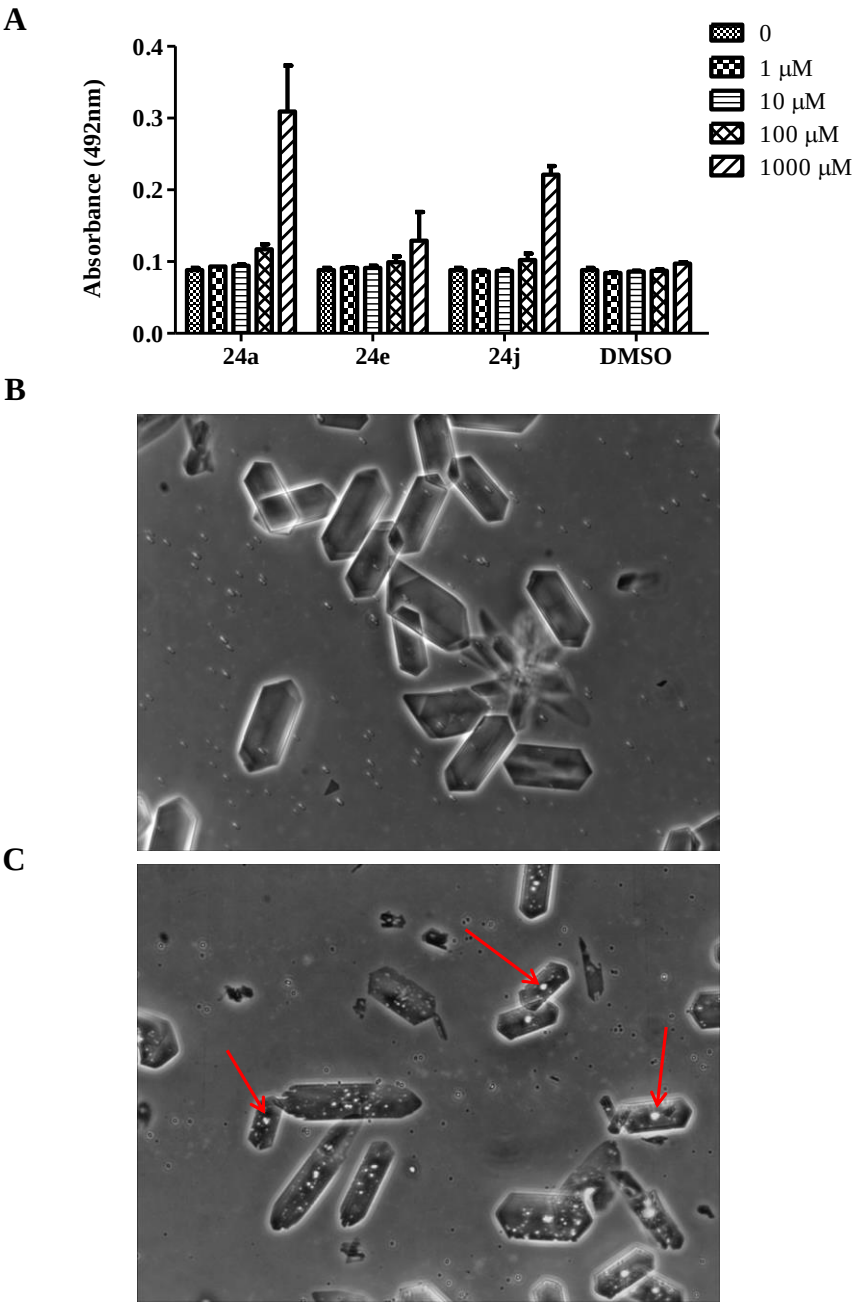



**Figure 4.15**: Solubility issues encountered with compounds **24a**, **24e** and **24j**. A) Absorbance reading for the compounds in RPMI-1640 culture medium (in the absence of cells) following the addition of the MTS reagent. B) Compound **24j** in RPMI-1640 medium before addition of the MTS reagent (x40). C) Compound **24j** in RPMI-1640 after addition of MTS reagent showing **24j** crystals in complex with reagent (arrows) (x40).

#### 4.3.2.2 **Membrane integrity studies**

Cell membrane integrity studies carried out using CytoTox-ONE™ Homogeneous Membrane Integrity Assay (Promega, Southampton, UK), were inconclusive for all the compounds (**24**, **24a**, **24e**, **24j** and **121**) examined. There was a high signal-to-noise ratio at all concentrations even after subtracting values obtained with media-only controls for background absorbance. As discussed previously (Section 4.3.1), the high interference observed here could be due to many factors including the presence of foetal bovine serum and the phenol red dye in the culture medium, both of which have been reported to interfere with the lactate dehydrogenase (LDH) enzyme in the assay.

Results from the membrane integrity studies, together with those from the MTS assay studies, would have provided an insight into how the compounds were exerting their effect on the cells. If the compounds were cytotoxic, cell membranes of both cell lines would be compromised and a high fluorescence reading would be expected from the LDH assay studies, the cells were viable at the start of the experiment as confirmed by the Trypan blue exclusion assay and therefore had intact cell membranes at the start of the experiment. The inconclusive results from the LDH assay investigations indicate that no conclusions can be drawn on the cytotoxicity or otherwise of these compounds. Further work to optimise this assay, or use of a different assay is necessary in the future in order to clarify the situation. Despite this setback with the cytotoxicity studies, the MTS results suggest that the compounds may be cytostatic, i.e. inhibition of cell growth by inhibiting the cells' metabolic ability. This cytostatic property is promising at this

stage of the drug discovery process as it indicates possible alternative strategies for compound efficacy optimisation against pancreatic cancer cells. Indeed, in the absence of a cure, cytostatic anti-cancer drugs used to stabilise the disease may be the way forward with lethal, incurable cancers such as pancreatic cancer [389].

#### 4.3.2.3 **Chick Chorioallantoic Membrane (CAM) Assay**

Angiogenesis plays a crucial role in cancer growth and progression by providing a rich network of blood vessels that supply oxygen and nutrients to the proliferating cells [390-392]. Inhibition of tumour angiogenesis therefore offers another therapeutic avenue of targeting cancer cell proliferation, migration and metastasis. Indeed, many anti-cancer drugs that are developed to inhibit tumour growth, migration and invasion do so by indirectly inhibiting proangiogenic factors [393, 394].

The chick chorioallantioc membrane (CAM) assay was employed to assess the effect of screening compounds on blood vessel formation. Traditionally, this assay has served as a useful model to study angiogenesis [395, 396], tumour growth and migration [397-400]. Compounds **24**, **24a**, **24e**, **24j** and **121** were tested for anti-angiogenic properties adapting the method described in [401]. The compounds were added to the chorioallantioc membranes of 4-day old developing embryos via a cut window through the shell and incubated until day 10 when representative images are taken to assess angiogenesis. Of the five compounds tested, preliminary observations show only compound **24** appeared to affect angiogenesis (Figure 4.16). However, this effect was not reproducible as it was difficult to retain viable embryos by day 10 of the

investigation despite efforts to minimise injury to the embryo during the study. As a result of the high attrition rate associated with this assay, insufficient data was generated for statistical analysis, even for compound **24**. Approximately 95% of embryos treated with the compounds on day 4 after first incubation did not make it to day 10, when they were assessed for anti-angiogenic effects. Due to time limitations, the experiment was only repeated twice, and on both occasions, most of the embryos had died by day 10. Despite this setback, the initial results obtained with compound **24** demonstrate the feasibility of this assay to study angiogenesis, a crucial process in tumour development. Once optimised in future studies, this assay, together with the proliferation and cytotoxicity assays, should provide a comprehensive understanding of how the screened compounds are affecting pancreatic cancer cell lines.

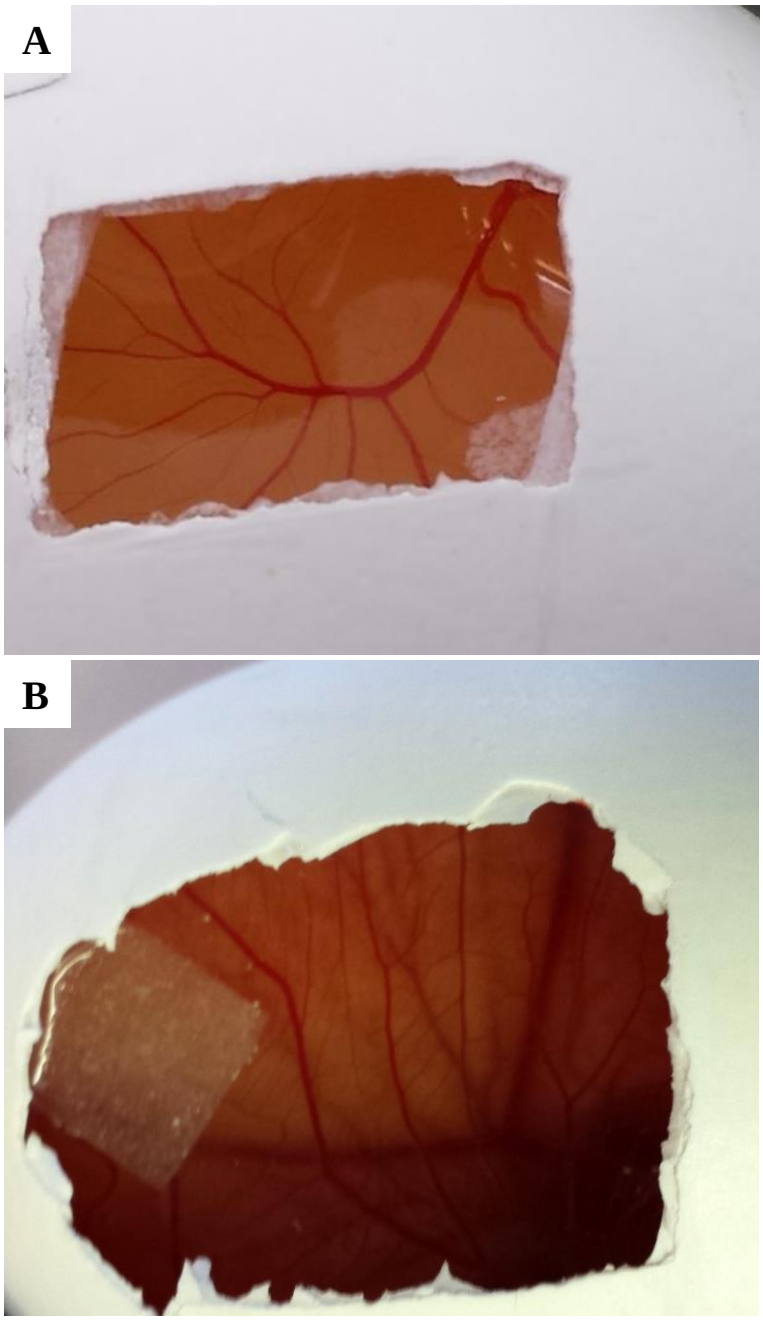



**Figure 4.16**: Chick chorioallantioc membrane (CAM) assay for compound **24**. A) Developing embryo on day 4 following incubation showing the developing blood vessels. B) Compound **24** (10 μM 20 μL pre-absorbed on filter paper) on the CAM membrane of a 10-day old embryo.

## 4.4 CONCLUSIONS

Compound **24** and 39 analogues were successfully synthesised, isolated and characterised following a two-step synthetic route involving a Diels–Alder cycloaddition and an addition/elimination reaction. Compounds structures and purity were confirmed by characterisation via $^{1}$H- and $^{13}$C-NMR, melting point analysis, and LC-MS spectrometry. By varying the substituents in positions 3 and 4 of the phenyl ring of compound **24**, as well as removing the nitro group, it is hoped that the diverse collection of analogues synthesised will provide a library of compounds that could be used in high throughput biological screening studies against pancreatic cancer cells. The results from these screens will be valuable in the development of a structure-activity relationship (SAR) model for these compounds in their optimisation for increased potency.

Five compounds, including compound **24**, were advanced for biological screening against the proliferation of pancreatic cancer cells. Compound **24** was found to inhibit proliferation of both BxPC-3 and Panc-1 cells at all concentrations of 1, 10 and 100 μM ($p$<0.0001). Its ester analogue, **24a**, also inhibited proliferation of both cell lines at all three concentrations. The same pattern of inhibition was observed for **24e** (no substitution on the phenyl ring system), **24j** (4-$CO_2H$) and **121** (no $NO_2$ group). Interestingly, all four analogues seemed to have a greater effect on proliferation of both cell lines compared to **24**. Since all five compounds showed inhibitory activity at 1 μM relative to control ($p$<0.0001), it will be interesting to see their effect at sub-micromolar concentrations. Testing the compounds at sub-micromolar concentrations will also allow

the assessment of their $IC_{50}$ values, an important property that will be useful when optimising them for potency.

Early indications from the biological screening, although not definitively conclusive, suggest that these compounds have the potential to contribute to pancreatic cancer therapy. Preliminary results demonstrate that the five compounds examined affect the proliferative capability of pancreatic cancer cells to an extent. Given that there are 35 more analogues yet to be examined, the prospect of finding more hits with even better activity than the five compounds investigated remains high. The five compounds examined all have the potential for further optimisation to improve potency in the next stage of the drug discovery process, in the hope that one or more candidates will emerge as a lead therapeutic candidate for the treatment of pancreatic cancer.

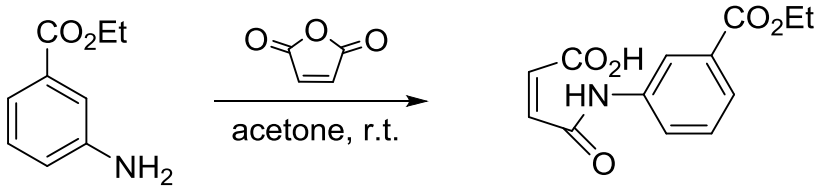
## 4.5 EXPERIMENTAL

### 4.5.1 General Methodology: Synthesis and Instrumentation

Similar procedure as described in Section 3.5.1.

#### 4.5.1.1 Synthesis of compound 24: Route a

4.5.1.1.1 Synthesis of (*Z*)-4-((3-(ethoxycarbonyl)phenyl)amino)-4-oxobut-2-enoic acid (**114**). (Adapted from Sortino *et al*. [402])



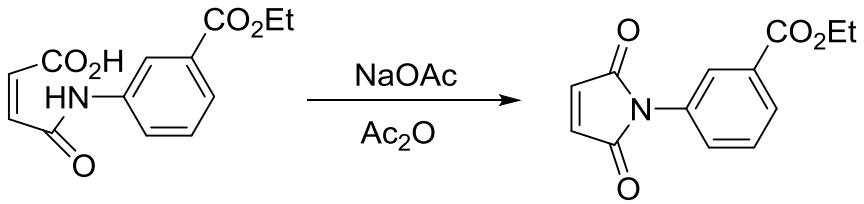

**114**

$C_{13}H_{13}NO_5$
Mol. Wt: 263.25

A pale brown solution of maleic anhydride (0.61 g, 6.20 mmol) and ethyl-3-aminobenzoate (1.01 g, 6.11 mmol) in acetone (30 mL) was stirred at room temperature for 4.5 h until a precipitate formed. The precipitate was filtered off to give **114** as a fluffy cream powder, yield 1.49 g, 93%, which was used in the next step without further purification. **M.p.** 154-155 °C. **$^1$H-NMR (400 MHz, acetone-*d*$_6$) δ (ppm)** 10.64 (br. s, 1H, OH), 8.27 - 8.41 (1 H, m, Ar-H), 7.99 (1 H, m, Ar-H ), 7.77 - 7.92 (1 H, m, Ar-H), 7.53 (1 H, t, *J* = 8.0 Hz, Ar-H), 6.69 (1 H, d, *J* = 12.6 Hz, alkenyl-H), 6.39 (1 H, d, *J* =

12.6 Hz, alkenyl-H), 4.36 (1 H, q, *J* = 7.1 Hz, $CH_2$), 1.36 (3 H, t, *J* = 7.1 Hz, $CH_3$). **$^{13}$C-NMR (100 MHz, acetone-*d*$_6$) δ (ppm)** 178.34 (C=O acid), 174.11 (C=O ester), 173.49 (C=O amide), 134.60 (alkenyl-CH), 132.49 (alkenyl-CH), 131.52 (Ar-C), 129.35(Ar-C), 126.56, 125.99 (Ar-C), 124.82 (Ar-C), 121.28 (Ar-C), 60.91 ($CH_2$), 13.75 ($CH_3$). **IR $\nu_{max}$/cm$^{-1}$**: 3303 (N-H stretch), 3223 (OH stretch), 3158, 3124, 3093 (C-H stretch), 1720 (C=O stretch, acid), 1678 (C=O stretch, ester), 1627 (C=O amide), 1106 (C-O stretch), 1018 (C-N stretch). LC-MS (ESI) found ***m/z*** $[M - H]^-$: 262, $C_{13}H_{13}NO_5$ requires 263.

4.5.1.1.2 Synthesis of ethyl 3-(2,5-dioxo-2,5-dihydro-1*H*-pyrrol-1-yl)benzoate (**115**). (Adapted from Sortino *et al*. [402])



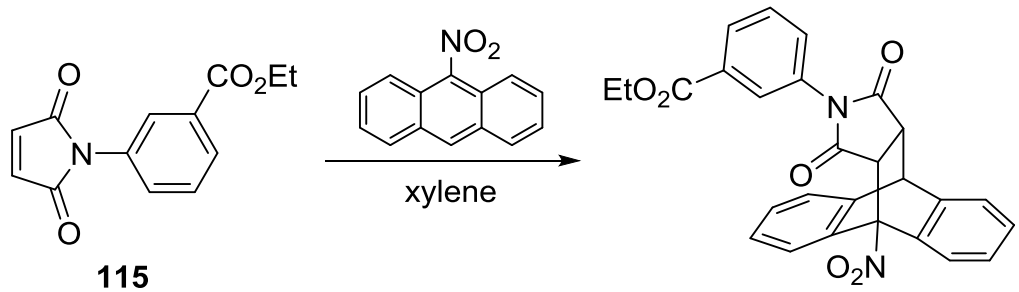

**114** **115**

$C_{13}H_{11}NO_4$
Mol. Wt: 245.23

Sodium acetate (0.03 g, 0.31 mmol) was added to a stirred solution of 4-((3-(ethoxycarbonyl)phenyl)amino)-4-oxobut-2-enoic acid (**114**) (0.35 g, 1.33 mmol) in acetic anhydride (20 mL) and the contents heated under reflux at 140 °C for 0.5 h. The pale yellow solution was cooled to room temperature while stirring for another 2 h. The mixture was poured into ice-cold sodium hydrogen carbonate (10% w/v, 40 mL) in an

ice-bath while stirring vigorously until a precipitate formed. The precipitate was extracted with $Et_2O$ (1x 75 mL, 1x 50 mL, 1x 25 mL). The combined ethereal extracts were dried over anhydrous $MgSO_4$, filtered, $Et_2O$ removed under vacuum and the product was precipitated from the remaining acetic anhydride solution by the addition of ice-cold distilled water. The precipitate was filtered under suction and dried to give **115** as pale orange flakes. The product was used in the next step without further purification. Yield 0.23 g, 66%. M.p. 109-110 °C. **$^1$H-NMR (400 MHz, DMSO-*d*$_6$) δ (ppm)** 7.91 - 8.02 (2 H, m, Ar-H), 7.59 - 7.72 (2 H, m, Ar-H), 7.21 (2 H, s, alkenyl-H), 4.34 (2 H, q, *J* = 7.14 Hz, $CH_2$), 1.33 (3 H, t, *J* = 7.14 Hz, $CH_3$). **$^{13}$C-NMR (100 MHz, DMSO-*d*$_6$) δ (ppm)** 170.32 (2 C=O amide), 165. 61 (C=O ester), 135.34 (2 C, olefinic), 132.60 (Ar-C), 131.94 (Ar-C), 131.20 (Ar-C), 129.99 (Ar-C), 128.77 (Ar-C), 127.74 (Ar-C), 61.65 ($CH_2$), 14.71 ($CH_3$). **IR $\nu_{max}$/cm$^{-1}$**: 3091, 2993 (C-H stretch), 1716, 1707 (C=O stretch), 1020 (C-O stretch). LC-MS (ESI) found ***m/z*** $[M - H]^-$: 244, $C_{13}H_{11}NO_4$ requires 245.

4.5.1.1.3 Attempted synthesis of ethyl 3-(9-nitro-12,14-dioxo-9,10-dihydro-9,10-[3,4]epipyrroloanthracen-13-yl)benzoate (**116**). (Adapted from Bova *et al*. [354])



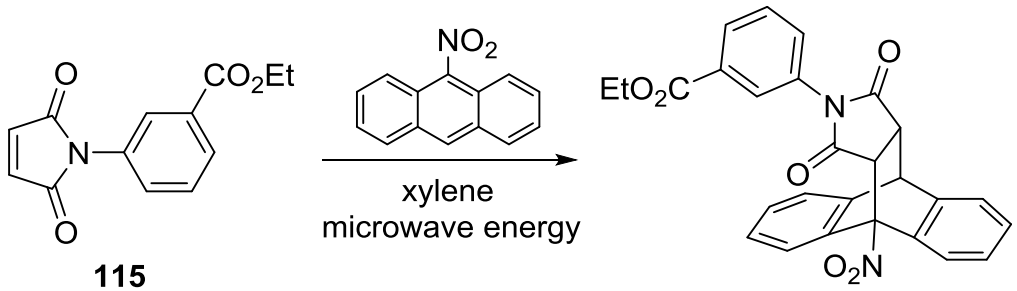

**116**

$C_{27}H_{20}N_2O_6$

Mol. Wt: 468.47

9-Nitroanthracene (0.22 g, 1.00 mmol) was added to a stirred solution of **115** (0.25 g, 1.00 mmol) in xylene (15 mL). The contents were heated under reflux for 3 days after which the reaction was cooled to room temperature with stirring. The mixture was then placed in an ice-bath while stirring for 45 minutes. After this time the solvent was removed under vacuum to yield **116** as a sticky orange solid, which proton NMR analysis indicated the material to be mainly unreacted diene. No further analysis was carried out on the product.

#### 4.5.1.2 **Attempted microwave-assisted synthesis of ethyl 3-(9-nitro-12,14-dioxo-9,10-dihydro-9,10-[3,4]epipyrroloanthracen-13-yl)benzoate (116)**




**115**

**116**

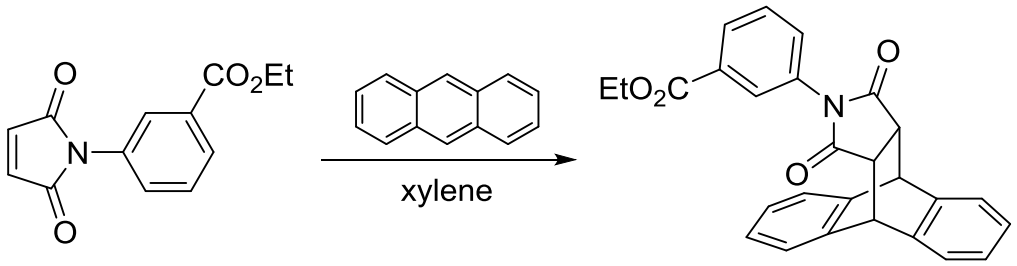
$C_{27}H_{20}N_2O_6$
Mol. Wt: 468.47

Ethyl 3-(2,5-dioxo-2,5-dihydro-1*H*-pyrrol-1-yl)benzoate (0.10 g, 0.50 mmol), 9-nitroanthrocene (0.18 g, 0.80 mmol) and xylene (4 mL) were placed in a 5 mL round bottom vial fitted with a magnetic stirrer. The contents were subjected to microwave (MW) irradiation in a Biotage Microwave Synthesizer (Biotage Limited, Hengoed,

Wales, UK) for 3 h at low absorption, at a temperature of 220 °C and 400 W. When the reaction was complete (monitored by TLC using hexane:*t*-butyl methyl ether, 4:1, v/v), the mixture was cooled to room temperature then placed in an ice-bath with stirring, which resulted in the formation of a precipitate. The precipitate was filtered analysis of which by proton NMR indicated the presence of unreacted 9-nitroanthracene.

#### 4.5.1.3 **Replacing substituted diene (9-$NO_2$) with unsubstituted diene (9-H)**

#### 4.5.1.4 **Synthesis of ethyl 3-(12,14-dioxo-9,10-dihydro-9,10-[3,4]epipyrroloanthracen-13-yl)benzoate (117)**



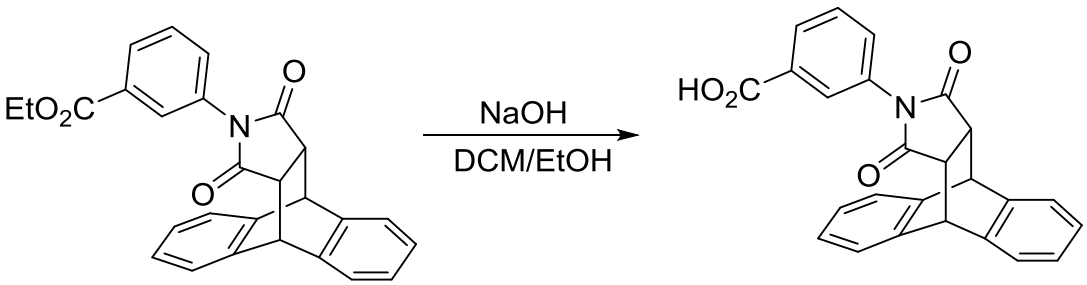

**115** **117**

$C_{27}H_{21}NO_4$
Mol. Wt: 423.47

The diene anthracene (0.25 g, 1.39 mmol) was added to a stirred solution of **115** (0.17 g, 0.70 mmol) in xylene (15 mL). The contents were heated under reflux at 140 °C for ~4 h. After this time, the reaction mixture was cooled to room temperature then placed in an ice-bath to precipitate out **117** as cream flakes. Yield 0.13 g, 44%. M.p. 203-205 °C. **$^1$H-NMR (400 MHz, acetone-$d_6$) δ (ppm)** 8.06 - 8.09 (1 H, m, Ar-H), 7.92

- 7.94 (1 H, m, Ar-H), 7.43 - 7.53 (4 H, m, Ar-H), 7.20 - 7.24 (4 H, m, Ar-H), 7.17 - 7.18 (1 H, m, Ar-H), 6.77 - 6.80 (1 H, m, Ar-H), 4.89 - 4.90 (2 H, m, CH dihydroanthracene), 4.34 (2 H, q, $J$ = 7.3 Hz, $CH_2$), 3.47 (2 H, m, CH succinimide), 1.36 (3 H, t, $J$ = 7.0 Hz, $CH_3$). **$^{13}$C-NMR (100 MHz, acetone-$d_6$) δ (ppm)** 175.62 (2 C=O amide), 164.91 (C=O ester), 141.97 (Ar-C), 139.68 (Ar-C), 131.40 (Ar-C), 131.30 (Ar-C), 129.12 (Ar-C), 129.02 (Ar-C), 128.17 (Ar-C), 127.85 (Ar-C), 126.85 (Ar-C), 126.61 (Ar-C), 125.51 (Ar-C), 124.38 (Ar-C), 60.95 ($CH_2$ aliphatic), 47.25 (2 CH dihydroanthracene), 45.83 (2 CH succinimide), 13.75 ($CH_3$). **IR $\nu_{max}$/cm$^{-1}$**: 3052, 2988, 2961 (C-H stretch), 1775, 1718, 1706 (C=O stretch, amide and ester). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 424, $C_{27}H_{21}NO_4$ requires 423.

#### 4.5.1.5 Attempted synthesis of 3-(12,14-dioxo-9,10-dihydro-9,10-3,4]epipyrroloanthracen-13-yl)benzoic acid (118)




**117**

**118**

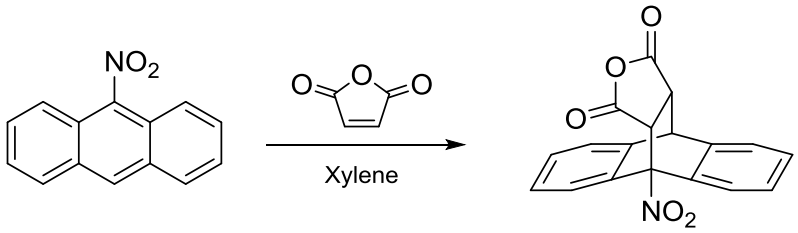
$C_{25}H_{17}NO_4$

Mol. Wt: 395.41

Compound **117** (0.05 g, 0.13 mmol) was dissolved in dichloromethane/ethanol (9:1 v/v) and stirred for 5 minutes at room temperature to give a pale yellow solution. Sodium hydroxide (2 mL, 1 M) was added while stirring resulting in a cloudy yellow solution. After 75 minutes of stirring at room temperature, the mixture was heated under reflux at 45 °C for 4 h after which a white precipitate was formed. The reaction was cooled to room temperature, the precipitate was filtered off and the residue taken up in deionised water (10 mL). The aqueous solution was extracted with diethyl ether (2x 20 mL) which was discarded. The aqueous phase was retained, acidified with HCl (1 M) to pH 2-3, and extracted with ethyl acetate (2x 20 mL, 1x 10 mL). The combined organic phases were dried over anhydrous magnesium sulphate, filtered, and the solvent removed to afford **118** as a white powder, yield 0.04 g, 77%. The product contained both the starting ester as well as some of the acid. No purification was carried out as the second route (route **b**) gave the desired product.

### 4.5.2 Route b

4.5.2.1 **Synthesis of 9-nitro-9,10-dihydro-9,10-[3,4]furanoanthracene-12,14-dione (119).** (Adapted from Wade [363]).



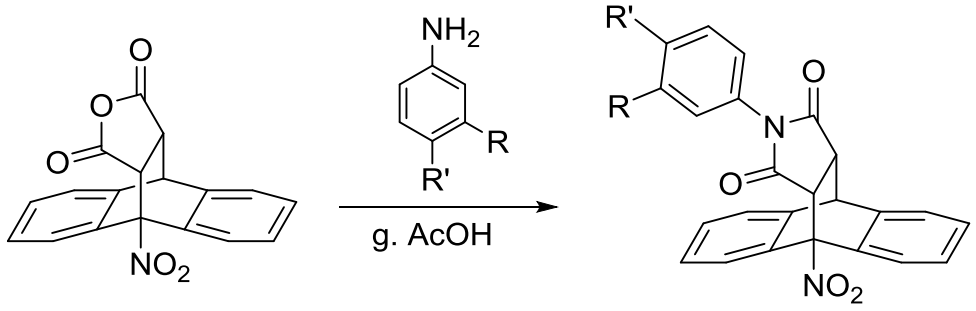

**119**

$C_{18}H_{11}NO_5$
MW: 321.29

Maleic anhydride (0.46 g, 4.71 mmol) was added to a stirred solution of 9-nitroanthracene (1.03 g, 4.61 mmol) in xylene (25 mL) and the mixture heated under reflux at 140-143 °C until all the diene disappeared as indicated by TLC (DCM). The reaction mixture was cooled to room temperature before placing on ice with stirring. A precipitate separated out upon cooling, which was filtered, washed with ice-cold xylene to give the desired Diels-Alder cycloadduct **119** as a pale yellow powder which was used in the next step without further purification. $R_f$ 0.78 ($Et_2O$). Yield 0.97 g, 65%. M.p. 239-240 °C (lit. 244-245 °C, [370]). **$^1$H-NMR (400 MHz, DMSO-*d*$_6$) δ (ppm)** 7.65 - 7.69 (1 H, m, Ar-H), 7.50 - 7.58 (2 H, m, Ar-H), 7.37 - 7.44 (3 H, m, Ar-H), 7.32 - 7.37 (1 H, m, Ar-H), 7.05 (1 H, dd, *J* = 7.5, 0.5 Hz, Ar-H), 5.11 (1 H, d, *J* = 3.0 Hz, CH dihydroanthracene), 4.66 (1 H, d, *J* = 9.3 Hz, CH succinimide), 3.94 (1 H, dd, *J* = 9.3, 3.2 Hz, CH succinimide). **$^{13}$C-NMR (100 MHz, DMSO-*d*$_6$) δ (ppm)** 170.32 (C=O

carboxyl), 168.99 (C=O carboxyl), 139.47 (Ar-C), 137.17 (Ar-C), 136.93 (Ar-C), 134.15 (Ar-C), 129.49 (Ar-C), 129.44 (Ar-C), 128.40 (Ar-C), 128.17 (Ar-C), 126.11 (Ar-C), 125.60 (Ar-C), 123.37 (Ar-C), 120.81 (Ar-C), 93.24 ($CNO_2$), 51.14 (CH dihydroanthracene), 49.25 (CH succinimide), 44.79 (CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 2988, 2969 (C-H stretch), 1781 (C=O stretch), 1556 ($NO_2$ asymmetric), 1361 ($NO_2$ symmetric). LC-MS (ESI) found ***m/z*** [M – H + MeOH]$^-$: 352, $C_{18}H_{11}NO_5$ requires 321.

4.5.2.1.1 Synthesis of 3-(9-nitro-12,14-dioxo-9,10-dihydro-9,10-[3,4]epipyrroloanthracen-13-yl)benzoic acid, [R = $CO_2H$, R′=H] (**24**) and analogues. (Adapted from Lima *et al*. [364] and Perry and Parveen [403]).




**119** **24-24s**

| Compound | R | R' | Compound | R | R' |
|---|---|---|---|---|---|
| **24** | $CO_2H$ | H | **24j** | H | $CO_2H$ |
| **24a** | $CO_2Et$ | H | **24k** | H | $CO_2Et$ |
| **24b** | F | H | **24l** | H | F |
| **24c** | Cl | H | **24m** | H | Cl |
| **24d** | $NO_2$ | H | **24n** | H | $NO_2$ |
| **24e** | H | H | **24o**† | H | $CH_3$ |
| **24f**† | $CH_3$ | H | **24p**† | H | MeO |
| **24g**† | MeO | H | **24q**† | H | $Me_2N$ |
| **24h**† | $Me_2N$ | H | **24r** | H | I |
| **24i**† | CN | H | **24s** | H | *t*-Bu |

†Compounds synthesised by Kulikowska [372] and Grewal [373] as part of their final year project for the degree of Master of Pharmacy, University of Hertfordshire.

**General procedure for the synthesis of 24-24s**

The corresponding 3- or 4-substituted aniline (1.2-1.7 eq) was added to a stirred solution of the maleic anhydride-9-nitroanthracene cycloadduct (**119**, 1eq) in glacial acetic acid (10 mL). The reaction mixture was heated under reflux at 120 °C for 2-3 h then cooled to room temperature. In some reactions, a precipitate formed upon cooling to room temperature (**24**, **24b**, **24d**, **24f**, **24j**, **24m**, **24n**, **24r**) while none was present for others (**24a**, **24c**, **24e**, **24g**, **24h**, **24i**, **24k**, **24l**, **24o**, **24p**, **24q**, **24s**). De-ionised water was added to quench all the reactions resulting in the formation of a precipitate in all cases. The precipitates were washed with ice-cold de-ionised water and dried under suction to yield the desired products. Yields shown are from the final Diels-Alder step, not from overall synthesis. Where known, reference melting points are shown next to that determined for the respective compound.

**Synthesis of 3-(9-nitro-12,14-dioxo-9,10-dihydro-9,10-[3,4]epipyrroloanthracen-13-yl)benzoic acid (24) [R = $CO_2H$, R′ = H]**.

Compound **119** (0.07 g, 0.22 mmol, 1 eq) was combined with the 3-aminobenzoic acid (0.20 g, 1.5 mmol, 6.7 eq) in 10 mL glacial acetic acid according to the procedure above. Recrystallisation in glacial acetic acid afforded **24** as white crystals, yield 1.07g, 96%. **M.p.** 272-273 °C. **$^1$H NMR (600 MHz, acetone-$d_6$) δ (ppm)** 7.92 - 7.99 (1 H, m, Ar-H), 7.67 - 7.76 (2 H, m, Ar-H), 7.32 - 7.51 (6 H, m, Ar-H), 7.25 (1 H, m, Ar-H), 7.08 (1 H, m, Ar-H), 6.68 - 6.85 (1 H, m, Ar-H), 5.05 (1 H, d, $J$ = 2.7

Hz, CH dihydroanthracene), 4.60 (1 H, d, *J* = 8.9 Hz, CH succinimide), 3.74 (1 H, dd, *J* = 8.9, 3.4 Hz, CH succinimide). **$^{13}$C-NMR (150 MHz, acetone-$d_6$) δ (ppm)** 174.05 (C=O), 172.93 (C=O), 166.63 (C=O acid), 139.79 (Ar-C), 137.60 (Ar-C), 137.45 (Ar-C), 134.37 (Ar-C), 132.20 (Ar-C), 131.61 (Ar-C), 130.99 (Ar-C), 129.68 (Ar-C), 129.13 (Ar-C), 128.78 (Ar-C), 128.48 (Ar-C), 127.97 (Ar-C), 127.52 (Ar-C), 127.47 (Ar-C), 125.47 (Ar-C), 124.71 (Ar-C), 123.79 (Ar-C), 120.55 (Ar-C), 93.69 ($CNO_2$), 49.41 (CH dihydroanthracene), 47.91 (CH succinimide), 45.78 (CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 3535 (OH stretch), 1783, 1706 (C=O stretch), 1552 ($NO_2$ asymmetric), 1398 ($NO_2$ symmetric), 1202 (CO stretch). LC-MS (ESI) found ***m/z*** $[M - H]^-$: 439, $C_{25}H_{16}NO_6$ requires 440.

**Synthesis of ethyl 3-(9-nitro-12,14-dioxo-9,10-dihydro-9,10-[3,4]epipyrroloanthracen-13-yl)benzoate (24a) [R = $CO_2$Et, R′ = H]**.

Compound **119** (0.05 g, 0.22 mmol, 1 eq), ethyl 3-aminobenzoate (0.03 g, 0.18 mmol, 1.2 eq) in 10 mL glacial acetic acid according to the procedure described above. Recrystallisation in glacial acetic acid afforded **24a** as white flakes, yield 0.06 g, 85%. **M.p.** 202-203 °C.**$^{1}$H-NMR (400 MHz, acetone-$d_6$) δ (ppm)** 7.94 - 7.96 (1 H, m, Ar-H), 7.68 - 7.75 (2 H, m , Ar-H), 7.35 - 7.52 (6 H, m, Ar-H), 7.16 - 7.17 (1 H, m, Ar-H), 7.08 - 7.11 (1 H, m, Ar-H), 6.79 - 6.81 (1 H, m, Ar-H), 5.07 (1 H, d, *J* = 3.4 Hz, CH dihydroanthracene), 4.62 (1 H, d, *J* = 8.8 Hz, CH succinimide), 4.33 (2 H, q, *J* = 7.1 Hz, $CH_2$ aliphatic), 3.75 - 3.77 (1 H, dd, *J* = 8.8, 3.1 Hz, CH succinimide), 1.40 (3 H, t, *J* = 7.1 Hz, $CH_3$). **$^{13}$C-NMR (150 MHz, acetone-$d_6$) δ (ppm)** 174.06 (C=O), 172.93 (C=O),

164.76 (C=O ester), 139.76 (Ar-C), 137.58 (Ar-C), 137.48 (Ar-C), 134.39 (Ar-C), 132.23 (Ar-C), 131.52 (Ar-C), 131.16 (Ar-C), 129.46 (Ar-C), 129.22 (Ar-C), 128.79 (Ar-C), 128.49 (Ar-C), 127.70 (Ar-C), 127.53 (Ar-C), 127.47 (Ar-C), 125.48 (Ar-C), 124.72 (Ar-C), 123.81 (Ar-C), 120.55 (Ar-C), 93.69 ($CNO_2$), 61.02 ($CH_2$ aliphatic), 49.44 (CH dihydroanthracene), 47.94 (CH succinimide), 45.78 (CH succinimide), 13.72 ($CH_3$). **IR $\nu_{max}$/cm$^{-1}$**: 2984, 2903 (CH stretch), 1782 (C=O stretch), 1710 (C=O stretch), 1551 ($NO_2$ asymmetric), 1363 ($NO_2$ symmetric), 1209 (CO stretch). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 469, $C_{27}H_{20}NO_6$ requires 468.

**Synthesis of 13-(3-fluorophenyl)-9-nitro-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (24b) [R = F, R′ = H]**.

Compound **119** (0.07 g, 0.23 mmol, 1 eq), 3-fluoroaniline (0.03 g, 0.27 mmol, 1.2 eq) in 10 mL glacial acetic acid according to the procedure described above. Recrystallisation in glacial acetic acid afforded **24b** as a cream solid, yield 0.08 g, 81%. **M.p.** 241-243 $^oC$. **$^1$H-NMR (400 MHz, DCM-$d_2$) δ (ppm)** 7.74 (1 H, d, *J* = 6.8 Hz, Ar-H), 7.54 (1 H, d, *J* = 7.0 Hz, Ar-H), 7.01 - 7.46 (8 H, m, Ar-H), 6.35 (1 H, d, *J* = 7.7 Hz, Ar-H), 6.12 - 6.31 (1 H, m, Ar-H), 4.93 (1 H, d, *J* = 3.1 Hz, CH dihydroanthracene), 4.45 (1 H, d, *J* = 9.0 Hz, CH succinimide), 3.56 (1 H, dd, *J* = 9.0, 3.1 Hz, CH succinimide). **$^{13}$C-NMR (100 MHz, DCM-$d_2$) δ (ppm)** 173.73 (C=O), 172.37 (C=O), 163.65 (Ar-CF), 138.96 (Ar-C), 137.13 (Ar-C), 136.49 (Ar-C), 133.93 (Ar-C), 130.34 (Ar-C), 128.88 (Ar-C), 128.66 (Ar-C), 127.80 (Ar-C), 127.68 (Ar-C), 125.36 (Ar-C), 124.36 (Ar-C), 123.90 (Ar-C), 122.24 (Ar-C), 120.91 (Ar-C), 116.18 (Ar-C), 115.97

(Ar-C), 113.79 (Ar-C), 93.42 ($CNO_2$), 49.03 (CH dihydroanthracene), 47.77 (CH succinimide), 45.99 (CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 3075, 3040, 3015, 2973, 2890 (CH stretch), 1706 (C=O stretch), 1551 ($NO_2$ asymmetric), 1383 ($NO_2$ symmetric). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 415, $C_{24}H_{15}FNO_4$ requires 414.

**Synthesis of 13-(3-chlorophenyl)-9-nitro-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (24c) [R = Cl, R′ = H]**.

Compound **119** (0.08 g, 0.23 mmol, 1 eq) was combined with 3-chloroaniline (0.04 g, 0.28 mmol, 1.2 eq) in 10 mL glacial acetic acid according to the procedure described above. Recrystallisation in glacial acetic acid afforded **24c** as a cream solid, yield 0.08 g, 82%. **M.p.** 230-232 $^oC$. **$^1$H-NMR (400 MHz, DMSO-$d_6$) δ (ppm)** 7.71 (1 H, d, *J* = 7.2 Hz, Ar-H), 7.61 - 7.66 (1 H, m, Ar-H), 7.31 - 7.51 (7 H, m, Ar-H), 7.06 (1 H, d, *J* = 7.1 Hz, Ar-H), 6.40 - 6.52 (2 H, m, Ar-H), 5.07 (1 H, d, *J* = 3.0 Hz, CH dihydroanthracene), 4.49 (1 H, d, *J* = 8.8 Hz, CH succinimide), 3.69 (1 H, dd, *J* = 8.7, 3.2 Hz, CH succinimide). **$^{13}$C-NMR (100 MHz, DMSO-$d_6$) δ (ppm)** 174.63 (C=O), 173.44 (C=O), 134.30 (Ar-C), 133.04 (Ar-C), 131.38 (Ar-C), 1245 (Ar-C), 125.51 (Ar-C), 129.29 (Ar-C), 128.99 (Ar-C), 128.03 (Ar-C), 127.95 (Ar-C), 126.84 (Ar-C), 125.96 (Ar-C), 125.83 (Ar-C), 125.67 (Ar-C), 125.47 (Ar-C), 123.80 (Ar-C), 122.72 (Ar-C), 120.75 (Ar-C), 120.65 (Ar-C), 93.74 ($CNO_2$), 49.63 (CH dihydroanthracene), 48.01 (CH succinimide), 45.29 (CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 3115, 2981, 2966 (CH stretch), 1717 (C=O stretch), 1547 ($NO_2$ asymmetric), 1383 ($NO_2$ symmetric). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 431, $C_{24}H_{15}ClNO_4$ requires 430.

**Synthesis of 9-Nitro-13-(3-nitrophenyl)-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (24d) [R = $NO_2$, R′ = H]**.

Compound **119** (0.06 g, 0.20 mmol, 1 eq) was combined with 3-nitroaniline (0.04 g, 0.30 mmol, 1.5 eq) in 10 mL glacial acetic acid according to the procedure described above. Recrystallisation in glacial acetic acid afforded **24d** as a pale yellow powder, yield 0.08 g, 93%. **M.p.** 240-242 $^{o}$C.**$^{1}$H-NMR (400 MHz, DCM-*$d_2$*) δ (ppm)** 8.15 (1 H, m, Ar-H), 7.68 - 7.78 (1 H, m, Ar-H), 7.47 - 7.59 (2 H, m, Ar-H), 7.26 - 7.46 (6 H, m, Ar-H), 7.06 - 7.16 (1 H, m, Ar-H), 6.94 - 6.98 (1 H, m, Ar-H), 4.95 (1 H, d, *J* = 3.0 Hz, CH dihydroanthracene), 4.50 (1 H, dd, *J* = 8.9, 1.3 Hz, CH succinimide), 3.61 - 3.63 (1 H, m, CH succinimide). **$^{13}$C-NMR (150 MHz, DCM-*$d_2$*) δ (ppm)** 173.55 (C=O), 172.23 (C=O), 148.41 (Ar-$CNO_2$), 138.77 (Ar-C), 136.96 (Ar-C), 136.39 (Ar-C), 133.86 (Ar-C), 132.37 (Ar-C), 132.01 (Ar-C), 130.10 (Ar-C), 128.96 (Ar-C), 128.80 (Ar-C), 127.93 (Ar-C), 127.76 (Ar-C), 125.41 (Ar-C), 124.40 (Ar-C), 123.93 (Ar-C), 123.72 (Ar-C), 121.67 (Ar-C), 120.95 (Ar-C), 93.42 ($CNO_2$), 49.13 (CH dihydroanthracene), 47.88 (CH succinimide), 46.04 (CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 3094, 3080, 3049, 2976, 2971, 2959, 2887, (CH stretch), 1775, 1709 (C=O stretch), 1548, 1529 ($NO_2$ asymmetric), 1386, 1341 ($NO_2$ symmetric). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 442, $C_{24}H_{15}N_3O_6$ requires 441.

**Synthesis of 9-nitro-13-phenyl-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (24e) [R = H, R′ = H]**.

Compound **119** (0.11 g, 0.33 mmol, 1 eq) was combined with aniline (0.04 g, 0.39 mmol, 1.2 eq) in 10 mL glacial acetic acid according to the procedure described above. Recrystallisation in glacial acetic acid afforded **24e** as cream-coloured flakes, yield 0.09 g, 72%. **M.p.** 111-112 °C. **$^{1}$H-NMR (400 MHz, DMSO-*d*$_6$) δ (ppm)** 7.71 (1 H, d, *J* = 7.3 Hz, Ar-H), 7.62 - 7.67 (1 H, m, Ar-H), 7.30 - 7.50 (8 H, m, Ar-H), 7.05 (1 H, d, *J* = 7.6 Hz, Ar-H), 6.42 - 6.47 (2 H, m, Ar-H), 5.07 (1 H, d, *J* = 2.6 Hz, CH dihydroanthracene), 4.50 (1 H, dd, *J* = 8.8, 1.5 Hz, CH succinimide), 3.66 - 3.68 (1 H, m, CH succinimide). **$^{13}$C-NMR (100 MHz, DMSO-*d*$_6$) δ (ppm)** 174.96 (C=O), 173.75 (C=O), 139.98 (Ar-C), 137.67 (Ar-C), 137.67 (Ar-C), 134.27 (Ar-C), 131.84 (Ar-C), 129.58 (2 *para*-Ar-C), 129.38 (Ar-C), 129.27 (Ar-C), 128.98 (Ar-C), 127.99 (Ar-C), 127.92 (Ar-C), 126.99 (2 *para*-Ar-C), 125.94 (Ar-C), 125.42 (Ar-C), 123.79 (Ar-C), 120.73 (Ar-C), 93.76 ($CNO_2$), 49.54 (CH dihydroanthracene), 47.95 (CH succinimide), 45.28 (CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 3070, 2972, 2903 (CH stretch), 1783, 1713 (C=O stretch), 1597, 1550 ($NO_2$ asymmetric), 1387 ($NO_2$ symmetric). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 397, $C_{24}H_{16}N_2O_4$ requires 396.

**Synthesis of 9-nitro-13-(*m*-tolyl)-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (24f) [R = CH$_3$, R′ = H]**.

Compound **119** (0.50 g, 1.56 mmol, 1 eq), *m*-toluidine (0.17 g, 1.59 mmol, 1 eq) in 20 mL glacial acetic acid according to the procedure described above.

Recrystallisation in glacial acetic acid afforded **24f** as a cream solid, yield 0.58 g 91%. **M.p.** 209-211 °C. **$^{1}$H-NMR (400 MHz, DMSO-*d*$_6$) δ (ppm)** 7.68 - 7.73 (1 H, m, Ar-H), 7.62 - 7.67 (1 H, m, Ar-H), 7.46 - 7.50 (1 H, m, Ar-H), 7.39 - 7.45 (3 H, m, Ar-H), 7.33 - 7.39 (1 H, m, Ar-H), 7.22 (1 H, m, Ar-H), 7.15 (1 H, m, Ar-H), 7.05 (1 H, m, Ar-H), 6.23 (1 H, m, Ar-H), 6.19 (1 H, s, Ar-H), 5.06 (1 H, d, *J* = 3.0 Hz, CH dihydroanthracene), 4.48 (1 H, d, *J* = 8.7 Hz, CH succinimide), 3.65 (1 H, dd, *J* = 8.8, 3.1 Hz, CH succinimide), 2.21 (3 H, s, $CH_3$). **$^{13}$C-NMR (100 MHz, DMSO-*d*$_6$) δ (ppm)** 175.00 (C=O), 173.80 (C=O), 139.97 (Ar-C), 139.09 (Ar-C), 137.70 (Ar-C), 137.54 (Ar-C), 134.29 (Ar-C), 131.82 (Ar-C), 129.99 (Ar-C), 129.37 (Ar-C), 129.26 (Ar-C), 128.95 (Ar-C), 127.99 (Ar-C), 127.91 (Ar-C), 127.48 (Ar-C), 125.97 (Ar-C), 125.41 (Ar-C), 124.09 (Ar-C), 123.82 (Ar-C), 120.73 (Ar-C), 93.76 ($CNO_2$), 49.53 (CH dihydroanthracene), 47.96 (CH succinimide), 45.28 (CH succinimide), 21.23 ($CH_3$). **IR $\nu_{max}$/cm$^{-1}$**: 3079, 3039, 2982, 2921 (CH stretch), 1787 (C=O), 1716 (C=O), 1549 ($NO_2$ asymmetric), 1382 ($NO_2$ symmetric). LC-MS (ESI) found ***m/z*** $[M - H]^-$: 409, $C_{25}H_{18}N_2O_4$ requires 410.

**Synthesis of 13-(3-methoxyphenyl)-9-nitro-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (24g) [R = MeO, R′ = H]**.

Compound **119** (0.52 g, 1.62 mmol, 1 eq) was combined with 3-methoxyaniline (0.24 g, 1.95 mmol, 1.2 eq) in 20 mL glacial acetic acid according to the procedure described above. Recrystallisation in glacial acetic acid afforded **24g** as a brown granules, yield 0.50 g, 72%. **M.p.** 262-265 °C. **$^{1}$H-NMR (600 MHz, DMSO-*d*$_6$) δ**

**(ppm)** 7.65 - 7.68 (1 H, m, Ar-H), 7.59 - 7.62 (1 H, m, Ar-H), 7.43 - 7.46 (1 H, m, Ar-H), 7.36 - 7.41 (3 H, m, Ar-H), 7.32 (1 H, m, Ar-H), 7.22 (1 H, t, *J* = 8.2 Hz, Ar-H), 7.00 - 7.03 (1 H, m, Ar-H), 6.88 (1 H, dd, *J* = 8.4, 2.6 Hz, Ar-H), 6.04 (1 H, m, Ar-H), 5.80 (1 H, t, *J* = 2.2 Hz, Ar-H), 5.02 (1 H, d, *J* = 3.1 Hz, CH dihydroanthracene), 4.44 (1 H, d, *J* = 8.8 Hz, CH succinimide), 3.61 - 3.63 (3 H, m, $CH_3$), 3.60 - 3.61 (1 H, m, CH succinimide). **$^{13}$C-NMR (150 MHz, DMSO-*d*$_6$) δ (ppm)** 174.78 (C=O), 173.59 (C=O), 159.94 (Ar-CO), 139.91 (Ar-C), 137.68 (Ar-C), 137.49 (Ar-C), 134.28 (Ar-C), 132.87 (Ar-C), 130.38 (Ar-C), 129.22 (Ar-C), 128.91 (Ar-C), 127.95 (Ar-C), 127.86 (Ar-C), 125.92 (Ar-C), 125.38 (Ar-C), 123.80 (Ar-C), 120.25 (Ar-C), 119.22 (Ar-C), 115.06 (Ar-C), 112.66 (Ar-C), 93.74 ($CNO_2$), 55.87 ($CH_3$), 49.51 (CH dihydroanthracene), 47.93 (CH succinimide), 45.28 (CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 3070, 2952, 2833 (CH stretch), 1783, 1716 (C=O), 1550 ($NO_2$ asymmetric), 1386 ($NO_2$ symmetric).

**Synthesis of 13-(3-(dimethylamino)phenyl)-9-nitro-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (21h) [R = Me$_2$N. R′ = H]**.

Compound **119** (0.50 g, 1.56 mmol, 1 eq) combined with *N*, *N*-dimethyl-*m*-phenylenediamine (0.39 g, 1.86 mmol, 1.2 eq) in 20 mL glacial acetic acid according to the procedure described above. Recrystallisation in glacial acetic acid afforded **24h** as a cream solid, yield 0.66 g, 97%. **M.p.** 213-215 °C. **$^{1}$H-NMR (400 MHz, DMSO-*d*$_6$) δ (ppm)** 7.70 (1 H, m, Ar-H), 7.63 - 7.67 (1 H, m, Ar-H), 7.46 - 7.50 (1 H, m, Ar-H), 7.38 - 7.44 (3 H, m, Ar-H), 7.32 - 7.38 (1 H, m, Ar-H), 7.11 (1 H, m, Ar-H), 7.03 - 7.07 (1 H, m, Ar-H), 6.65 (1 H, m, Ar-H), 5.79 (1 H, m, Ar-H), 5.43 (1 H, t, *J* = 2.2 Hz, Ar-H),

5.05 (1 H, d, *J* = 3.1 Hz, CH dihydroanthracene), 4.47 (1 H, d, *J* = 8.8 Hz, CH succinimide), 3.62 (1 H, dd, *J* = 8.8, 3.1 Hz, CH succinimide), 2.78 (6 H, s, 2 $CH_3$). **$^{13}$C-NMR (100 MHz, DMSO-*d*$_6$) δ (ppm)** 175.02 (C=O), 173.81 (C=O), 151.25 (Ar-CN), 140.00 (Ar-C), 137.75 (Ar-C), 137.59 (Ar-C), 134.33 (Ar-C), 132.78 (Ar-C), 129.83 (Ar-C), 129.24 (Ar-C), 128.88 (Ar-C), 127.98 (Ar-C), 127.85 (Ar-C), 125.96 (Ar-C), 125.39 (Ar-C), 123.85 (Ar-C), 120.71 (Ar-C), 114.45 (Ar-C), 113.07 (Ar-C), 110.63 (Ar-C), 93.79 ($CNO_2$), 49.49 (CH dihydroanthracene), 47.93 (CH succinimide), 45.31 (CH succinimide), C in $CH_3$ overlapped with solvent peaks. **IR $\nu_{max}$/cm$^{-1}$**: 3075, 2911, 2822 (CH stretch), 2358 (CN), 1788, 1717 (C=O), 1548 ($NO_2$ asymmetric), 1389, 1354 ($NO_2$ symmetric). LC-MS (ESI) found ***m/z*** $[M - H]^-$: 438, $C_{26}H_{21}N_3O_4$ requires 439.

**Synthesis of 3-(9-nitro-12,14-dioxo-9,10-dihydro-9,10-[3,4]epipyrroloanthracen-13-yl)benzonitrile (24i) [R = CN, R′ = H]**.

Compound **119** (0.50 g, 1.56 mmol, 1 eq) was combined with 3-aminobenzonitrile (0.22 g, 1.86 mmol, 1.2 eq) in 20 mL glacial acetic acid according to the procedure described above. Recrystallisation in glacial acetic acid afforded **24i** as a cream solid, yield 0.64 g, 98%. **M.p.** 310-3.11 °C. **$^{1}$H-NMR (600 MHz, DMSO-*d*$_6$) δ (ppm)** 7.82 (1 H, m, Ar-H), 7.68 (1 H, dd, *J* = 7.5, 1.0 Hz, Ar-H), 7.55 - 7.62 (2 H, m, Ar-H), 7.44 - 7.48 (1 H, m, Ar-H), 7.36 - 7.41 (3 H, m, Ar-H), 7.33 (1 H, m, Ar-H), 7.02 (1 H, M, Hz, Ar-H), 6.86 - 6.88 (1 H, m, Ar-H), 6.75 (1 H, m, Ar-H), 5.05 (1 H, d, *J* = 3.1 Hz, CH dihydroanthracene), 4.48 (1 H, d, *J* = 8.768 Hz, CH succinimide), 3.69 (1 H, dd, *J* = 8.8, 3.1 Hz, CH succinimide). **$^{13}$C-NMR (150 MHz, DMSO-*d*$_6$) δ (ppm)**

174.50 (C=O), 173.31 (C=O), 139.76 (Ar-C), 137.59 (Ar-C), 137.33 (Ar-C), 134.22 (Ar-C), 133.28 (Ar-C), 132.44 (Ar-C), 131.82 (Ar-C), 131.26 (Ar-C), 130.15 (Ar-C), 129.28 (Ar-C), 129.01 (Ar-C), 128.02 (Ar-C), 127.94 (Ar-C), 125.95 (Ar-C), 125.44 (Ar-C), 123.77 (Ar-C), 120.74 (Ar-C), 118.02 (CN), 112.56 (Ar-C), 93.71 ($CNO_2$), 49.63 (CH dihydroanthracene), 47.99 (CH succinimide), 45.27 (CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 3101, 3064, 2995, 2924 (CH stretch), 2231 (CN stretch), 1788, 1716 (C=O), 1547 ($NO_2$ asymmetric), 1387 ($NO_2$ symmetric). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 420, $C_{25}H_{15}N_3O_4$ requires 421.

**Synthesis of 4-(9-nitro-12,14-dioxo-9,10-dihydro-9,10-[3,4]epipyrroloanthracen-13-yl)benzoic acid (24j) [R = H, R′ = $CO_2H$]**.

Compound **119** (0.05 g, 0.16 mmol, 1 eq) was combined with 4-aminobenzoic acid (0.03 g, 0.16 mmol, 1 eq) in 10 mL glacial acetic acid according to the procedure described above. Recrystallisation in glacial acetic acid afforded **24j** as cream-coloured flakes, yield 0.06 g 80%. **M.p.** 343-345 °C.**$^1$H-NMR (400 MHz, acetone-$d_6$) δ (ppm)** 7.97 (2 H, d, *J* = 8.9 Hz, *para*-Ar-H), 7.69 - 7.78 (2 H, m, Ar-H), 7.47 - 7.51 (1 H, m, Ar-H), 7.35 - 7.45 (4 H, m, Ar-H), 7.08 - 7.12 (1 H, m, Ar-H), 6.73 (2 H, d, *J* = 8.8 Hz, *para*-Ar-H), 5.05 - 5.08 (1 H, m, CH dihydroanthracene), 4.62 (1 H, d, *J* = 8.9 Hz, CH succinimide), 3.76 (1 H, dd, *J* = 8.9, 3.1 Hz, CH succinimide). **$^{13}$C-NMR (150 MHz, acetone-$d_6$) δ (ppm)** 173.91 (C=O), 172.76 (C=O), 165.94 (C=O ester), 139.77 (Ar-C), 137.58 (Ar-C), 137.41 (Ar-C), 135.70 (Ar-C), 134.33 (Ar-C), 130.86 (2 *para*-Ar-C), 130.08 (Ar-C), 128.80 (Ar-C), 128.52 (Ar-C), 127.53 (Ar-C), 127.49 (Ar-C), 126.57 (2

*para*-Ar-C), 125.47 (Ar-C), 124.71 (Ar-C), 123.77 (Ar-C), 120.54 (Ar-C), 93.69 ($CNO_2$), 49.39 (CH dihydroanthracene), 47.89 (CH succinimide), 45.77 (CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 3284 (OH stretch), 3076, 2970, 2903 (CH stretch), 1781, 1702 (C=O stretch), 1550 ($NO_2$ asymmetric), 1389 ($NO_2$ symmetric). LC-MS (ESI) found ***m/z*** $[M - H]^-$: 439, $C_{25}H_{16}N_2O_6$ requires 440.

**Synthesis of ethyl 4-(9-nitro-12,14-dioxo-9,10-dihydro-9,10-[3,4]epipyrroloanthracen-13-yl)benzoate (24k) [R = H, R′ = $CO_2$Et]**.

Compound **119** (0.05 g, 0.16 mmol, 1 eq) ethyl-4-aminobenzoate (0.03 g, 0.21 mmol, 1.3 eq) in 10 mL glacial acetic acid according to the procedure described above. Recrystallisation in glacial acetic acid afforded **24k** as white flakes, yield 0.06 g, 78%. **M.p.** 255-256 $^oC$. **$^1$H-NMR (400 MHz, acetone-$d_6$) δ (ppm)** 7.92 - 7.97 (2 H, m, *para*-Ar-H), 7.68 - 7.77 (2 H, m, Ar-H), 7.47 - 7.51 (1 H, m, Ar-H), 7.36 - 7.46 (4 H, m, Ar-H), 7.10 (1 H, d, *J* = 6.1 Hz, Ar-H), 6.70 - 6.74 (2 H, m, *para*-Ar-H), 5.07 (1 H, d, *J* = 3.1 Hz, CH dihydroanthracene), 4.62 (1 H, d, *J* = 8.9 Hz, CH succinimide), 4.32 (2 H, q, *J* = 7.1 Hz, $CH_2$ aliphatic), 3.76 (1 H, dd, *J* = 8.8, 3.0 Hz, CH succinimide), 1.33 (3 H, t, *J* = 7.1 Hz, $CH_3$). **$^{13}$C-NMR (150 MHz, acetone-$d_6$) δ (ppm)** 173.85 (C=O), 172.73 (C=O), 164.98 (C=O ester), 139.76 (Ar-C), 137.57 (Ar-C), 137.41 (Ar-C), 134.34 (Ar-C), 130.63 (Ar-C), 129.78 (2 *para*-Ar-C), 128.79 (Ar-C), 128.51 (Ar-C), 127.53 (Ar-C), 127.48 (Ar-C), 126.61 (2 *para*-Ar-C), 125.46 (Ar-C), 124.71 (Ar-C), 123.77 (Ar-C), 120.54 (Ar-C), 113.38 (Ar-C), 93.68 ($CNO_2$), 60.91 ($CH_2$ aliphatic), 49.39 (CH dihydroanthracene), 47.89 (CH succinimide), 45.78 (CH succinimide), 13.65 ($CH_3$). **IR**

**$\nu_{max}$/cm$^{-1}$**: 2989 (CH stretch), 1785, 1710 (C=O stretch), 1550 ($NO_2$ asymmetric), 1394 ($NO_2$ symmetric). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 469, $C_{27}H_{20}N_2O_6$ requires 468.

**Synthesis of 13-(4-fluorophenyl)-9-nitro-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (24l) [R = H, R′ = F]**.

Compound **119** (0.10 g, 0.31 mmol, 1 eq) was combined with 4-fluoroaniline (0.04 g, 0.37 mmol, 1.2 eq) in 10 mL glacial acetic acid according to the procedure described above. Recrystallisation in glacial acetic acid afforded **24l** as a cream solid, yield 0.11 g, 88%. **M.p.** 238-240 ºC. **$^1$H-NMR (600 MHz, acetone-$d_6$) δ (ppm)** 7.66 - 7.75 (2 H, m, *para*-Ar-H), 7.44 - 7.48 (1 H, m, Ar-H), 7.34 - 7.42 (4 H, m, Ar-H), 7.05 - 7.10 (3 H, m, Ar-H), 6.54 - 6.59 (2 H, m, *para*-Ar-H), 5.02 (1 H, d, *J* = 2.7 Hz, CH dihydroanthracene), 4.56 (1 H, d, *J* = 8.9 Hz, CH succinimide), 3.70 (1 H, dd, *J* = 8.9, 2.7 Hz, CH succinimide). **$^{13}$C-NMR (150 MHz, acetone-$d_6$) δ (ppm)** 174.10 (C=O), 172.97 (C=O), 163.03 (Ar-CF), 139.80 (Ar-C), 137.61 (Ar-C), 137.48 (Ar-C), 134.40 (Ar-C), 128.79 (Ar-C), 128.76 (Ar-C), 128.74 (Ar-C), 128.41 (2 *para*-Ar-C), 128.04 (Ar-C), 127.51 (2 *para*-Ar-C), 127.41 (Ar-C), 125.45 (Ar-C), 124.69 (Ar-C), 123.79 (Ar-C), 120.53 (Ar-C), 115.77 (Ar-C), 115.62 (Ar-C), 93.66 ($CNO_2$), 49.29 (CH dihydroanthracene), 47.77 (CH succinimide), 45.75 (CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 2981, 2970, 2893 (CH stretch), 1782, 1710 (C=O stretch), 1551 ($NO_2$ asymmetric), 1395 ($NO_2$ symmetric). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 415, $C_{24}H_{15}FN_2O_4$ requires 441.

**Synthesis of 13-(4-chlorophenyl)-9-nitro-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (24m) [R = H, R′ = Cl]**.

Compound **119** (0.07 g, 0.23 mmol, 1 eq) was combined with 4-chloroaniline (0.05 g, 0.37 mmol, 1.6 eq) in 10 mL glacial acetic acid according to the procedure described above. Recrystallisation in glacial acetic acid afforded **24m** as a cream solid, yield 0.09 g, 90%. **M.p.** 270-273 $^{o}$C. **$^{1}$H-NMR (400 MHz, DCM-$d_2$) δ (ppm)** 7.70 - 7.75 (1 H, m, Ar-H), 7.52 - 7.56 (1 H, m, Ar-H), 7.26 - 7.42 (7 H, m, Ar-H), 7.08 - 7.13 (1 H, m, Ar-H), 6.43 - 6.49 (2 H, m, *para*-Ar-H), 4.92 (1 H, d, *J* = 3.1 Hz, CH dihydroanthracene), 4.44 (1 H, d, *J* = 8.9 Hz, CH succinimide), 3.55 (1 H, dd, *J* = 8.9, 3.1 Hz, CH succinimide). **$^{13}$C-NMR (150 MHz, DCM-$d_2$) δ (ppm)** 173.84 (C=O), 172.47 (C=O), 138.96 (Ar-C), 137.14 (Ar-C), 136.49 (Ar-C), 134.80 (Ar-C), 133.92 (Ar-C), 129.62 (Ar-C), 129.29 (2-*para*-Ar-C), 128.86 (Ar-C), 128.60 (Ar-C), 127.76 (Ar-C), 127.68 (3 Ar-C), 125.34 (Ar-C), 124.34 (Ar-C), 123.89 (Ar-C), 120.90 (Ar-C), 93.43 ($CNO_2$), 49.02 (CH dihydroanthracene), 47.76 (CH succinimide), 45.99 (CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 2981, 2971, 2887 (CH stretch), 1781, 1709 (C=O stretch), 1551 ($NO_2$ asymmetric), 1389 ($NO_2$ symmetric). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 431, $C_{24}H_{15}ClN_2O_4$ requires 430.

**Synthesis of 9-nitro-13-(4-nitrophenyl)-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (24n) [R = H, R′ = $NO_2$]**.

Compound **119** (0.07 g, 0.21 mmol, 1 eq) was combined with 4-nitroaniline (0.04 g, 0.29 mmol, 1.4 eq) in 10 mL glacial acetic acid according to the procedure described above. Recrystallisation in glacial acetic acid afforded **24n** as fluffy cream powder, yield 0.07 g, 78%. **M.p.** 316-318 $^{o}$C. **$^{1}$H-NMR (400 MHz, DCM-*d*$_2$) δ (ppm)** 8.12 - 8.16 (2 H, m, *para*-Ar-H), 7.72 - 7.76 (1 H, m, Ar-H), 7.53 - 7.56 (1 H, m, Ar-H), 7.30 - 7.44 (5 H, m, Ar-H), 7.10 - 7.14 (1 H, m, Ar-H), 6.78 - 6.83 (2 H, m, *para*-Ar-H), 4.94 (1 H, d, *J* = 3.1 Hz, CH dihydroanthracene), 4.49 (1 H, d, *J* = 9.0 Hz, CH succinimide), 3.60 (1 H, dd, *J* = 9.0, 3.1 Hz, CH succinimide). **$^{13}$C-NMR (100 MHz, DCM-*d*$_2$) δ (ppm)** 170.62 (C=O), 169.30 (C=O), 144.68 (Ar-C), 136.03 (Ar-C), 134.20 (Ar-C), 133.66 (Ar-C), 133.60 (Ar-C), 131.07 (Ar-C), 126.19 (Ar-C), 125.97 (Ar-C), 125.11 (Ar-C), 124.99 (Ar-C), 124.30 (2 *para*-Ar-C), 122.59 (Ar-C), 121.62 (Ar-C), 121.54 (2 *para*-Ar-C), 121.11 (Ar-C), 118.17 (Ar-C), 90.62 ($CNO_2$), 46.33 (CH dihydroanthracene), 45.06 (CH succinimide), 43.25 (CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 3121, 3087, 2978, 2855 (CH stretch), 1781, 1712 (C=O stretch), 1551, 1521 ($NO_2$ asymmetric), 1344 ($NO_2$ symmetric). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 442, $C_{24}H_{15}N_3O_6$ requires 441.

**Synthesis of 9-nitro-13-(p-tolyl)-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (24o), [R = H, R′ = CH$_3$]**.

Compound **119** (0.50 g, 1.56 mmol, 1 eq) was combined with *p*-toluidine (0.19 g, 1.77 mmol, 1.1 eq) in 20 mL glacial acetic acid according to the procedure described above. Recrystallisation in glacial acetic acid afforded **24o** as white powder, yield 0.48

g. 75%. **M.p.** 258-262 °C. **$^1$H-NMR (400 MHz, DMSO-$d_6$) δ (ppm)** 7.70 (1 H, dd, *J* = 7.3, 1.1 Hz, Ar-H), 7.60 - 7.66 (1 H, m, Ar-H), 7.45 - 7.49 (1 H, m, Ar-H), 7.38 - 7.44 (3 H, m, Ar-H), 7.33 - 7.38 (1 H, m, Ar-H), 7.14 (2 H, m, *J* = 8.1 Hz, *para*-Ar-H), 7.1 (1 H, d, *J* = 7.4 Hz, Ar-H), 6.32 (2 H, m, *J* = 8.2 Hz, *para*-Ar-H), 5.06 (1 H, d, *J* = 3.0 Hz, CH dihydroanthracene), 4.47 (1 H, d, *J* = 8.7 Hz, CH succinimide), 3.64 (1 H, dd, *J* = 8.8, 3.1 Hz, CH succinimide), 2.25 (3 H, s, $CH_3$). **$^{13}$C-NMR (100 MHz, DMSO-$d_6$) δ (ppm)** 175.02 (C=O), 173.81 (C=O), 140.00 (Ar-C), 138.99 (Ar-C), 137.67 (Ar-C), 137.56 (Ar-C), 134.26 (Ar-C), 129.99 (2 *para*-Ar-C), 129.25 (Ar-C), 129.22 (Ar-C), 128.94 (Ar-C), 127.98 (Ar-C), 127.88 (Ar-C), 126.69 (2 *para*-Ar-C), 125.92 (Ar-C), 125.41 (Ar-C), 123.78 (Ar-C), 120.72 (Ar-C), 93.76 ($CNO_2$), 49.47 (CH dihydroanthracene), 47.89 (CH succinimide), 45.27 (CH succinimide), 21.21 ($CH_3$). **IR $\nu_{max}$/cm$^{-1}$**: 3486, 3049, 2966, 2918 (CH stretch), 1784, 1719 (C=O stretch), 1553, 1514 ($NO_2$ asymmetric), 1391 ($NO_2$ symmetric). LC-MS (ESI) found ***m/z*** [M – H]$^-$: 409, $C_{25}H_{18}N_2O_4$ requires 410.

**Synthesis of 13-(4-methoxyphenyl)-9-nitro-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (24p) [R = H, R′ = MeO]**.

Compound **119** (0.50 g, 1.56 mmol, 1 eq) was combined with 4-methoxyaniline (0.23 g, 1.87 mmol, 1.2 eq) in 20 mL glacial acetic acid according to the procedure described above. Recrystallisation in glacial acetic acid afforded **24p** as a grey-purple solid, yield 0.56 g, 84%**. M.p.** 249-250 °C. **$^1$H-NMR (600 MHz, DMSO-$d_6$) δ (ppm)** 7.65 - 7.67 (1 H, m, Ar-H), 7.58 - 7.62 (1 H, m, Ar-H), 7.41 - 7.45 (1 H, m, Ar-H), 7.34

- 7.39 (3 H, m, Ar-H), 7.32 (1 H, td, *J* = 7.7, 1.20 Hz, Ar-H), 7.01 (1 H, d, *J* = 7.7 Hz, Ar-H), 6.81 - 6.89 (2 H, m, *para*-Ar-H), 6.26 - 6.34 (2 H, m, *para*-Ar-H), 5.01 (1 H, d, *J* = 3.1 Hz, CH dihydroanthracene), 4.43 (1 H, d, *J* = 8.8 Hz, CH succinimide c), 3.68 (3 H, s, $CH_3$), 3.59 (1 H, dd, *J* = 8.8, 3.1 Hz, CH succinimide). **$^{13}$C-NMR (150 MHz, DMSO-*d*$_6$) δ (ppm)** 175.09 (C=O), 173.87 (C=O), 159.73 (Ar-CO), 139.97 (Ar-C), 137.54 (Ar-C), 134.27 (Ar-C), 129.20 (Ar-C), 128.87 (Ar-C), 128.12 (2 *para*-Ar-CH), 127.93 (Ar-C), 127.82 (Ar-C), 127.30 (Ar-C), 125.88 (Ar-C), 125.34 (Ar-C), 124.31 (Ar-C), 123.75 (Ar-C), 120.68 (Ar-C), 114.76 (2 *para*-Ar-CH), 93.74 ($CNO_2$), 55.88 ($CH_3$), 49.40 (CH dihydroanthracene), 47.81 (CH succinimide), 45.28 (CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 3367, 2960, 2909 (CH stretch), 1782, 1702 (C=O stretch), 1549 ($NO_2$ asymmetric), 1395 ($NO_2$ symmetric). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 427, $C_{25}H_{18}N_2O_5$ requires 426.

**Synthesis of 13-(4-(dimethylamino)phenyl)-9-nitro-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (21q) [R = H, R′ = Me$_2$N]**.

Compound **119** (0.50 g, 0.1.56 mmol, 1 eq) was combined with *N*,*N*-dimethyl-1,4-diamine (0.25 g, 1.84 mmol, 1.2 eq) in 20 mL glacial acetic acid according to the procedure described above. Recrystallisation in glacial acetic acid afforded **24q** as light-grey solid, yield 0.50 g, 73%. **M.p.** 265-266 ºC. **$^1$H-NMR (600 MHz, DMSO-*d*$_6$) δ (ppm)** 7.66 (1 H, dd, *J* = 7.4, 1.0 Hz, Ar-H), 7.57 - 7.61 (1 H, m, Ar-H), 7.40 - 7.43 (1 H, m, Ar-H), 7.3 - 7.39 (3 H, m, Ar-H), 7.31 (1 H, td, *J* = 7.6, 1.3 Hz, Ar-H), 7.00 (1 H, d, *J* = 7.6 Hz, Ar-H), 6.52 - 6.57 (2 H, m, para-Ar-H), 6.15 - 6.19 (2 H, m, para-Ar-H),

4.99 (1 H, d, *J* = 3.1 Hz, CH dihydroanthracene), 4.40 (1 H, d, *J* = 8.8 Hz, CH succinimide), 3.55 (1 H, dd, *J* = 8.8, 3.1 Hz, CH succinimide), 2.81 (6 H, s, 2 $CH_3$). **$^{13}$C-NMR (150 MHz, DMSO-*d*$_6$) δ (ppm)** 175.32 (C=O), 174.10 (C=O), 150.76 (Ar-CO), 140.04 (Ar-C), 137.68 (Ar-C), 137.60 (Ar-C), 134.28 (Ar-C), 129.17 (Ar-C), 128.82 (Ar-C), 127.91 (Ar-C), 127.76 (Ar-C), 127.39 (2 *para*-Ar-CH), 125.86 (Ar-C), 125.30 (Ar-C), 123.75 (Ar-C), 120.65 (Ar-C), 119.99 (Ar-C), 112.38 (2 *para*-Ar-CH), 93.76 ($CNO_2$), 49.28 (CH dihydroanthracene), 47.71 (CH succinimide), 45.28 (CH succinimide), 2 $CH_3$ overlapped with solvent peaks. **IR $\nu_{max}$/cm$^{-1}$**: 3074, 2980, 2890, 2816 (CH stretch), 1703 (C=O stretch), 1545 ($NO_2$ asymmetric), 1362 ($NO_2$ symmetric). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 440, $C_{26}H_{21}N_3O_4$ requires 439.

**Synthesis of 13-(4-iodophenyl)-9-nitro-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (24r) [R = H, R′ = I]**.

Compound **119** (0.11 g, 0.35 mmol, 1 eq) was combined with 4-iodoaniline (0.12 g, 0.54 mmol, 1.6 eq) in 10 mL glacial acetic acid according to the procedure described above. Recrystallisation in glacial acetic acid afforded **24r** as white crystalline solid, yield 0.15 g, 84%. **M.p.** 305-308 $^{o}$C. **$^{1}$H-NMR (400 MHz, DMSO-*d*$_6$) δ (ppm)** 7.66 - 7.77 (3 H, m, Ar-CH), 7.59 - 7.65 (1 H, m, Ar-CH), 7.30 - 7.49 (5 H, m, Ar-CH), 7.05 (1 H, d, *J* = 7.5 Hz, Ar-CH), 6.23 - 6.32 (2 H, m, *para*-Ar-CH), 5.06 (1 H, d, *J* = 2.6 Hz, CH dihydroanthracene), 4.48 (1 H, d, *J* = 8.8 Hz, CH succinimide), 3.66 (1 H, dd, *J* = 8.7, 2.6 Hz, CH succinimide). **$^{13}$C-NMR (100 MHz, DMSO-*d*$_6$) δ (ppm)** 174.65 (C=O), 173.44 (C=O), 139.89 (Ar-C), 138.50 (2 *para*-Ar-C), 137.60 (Ar-C),

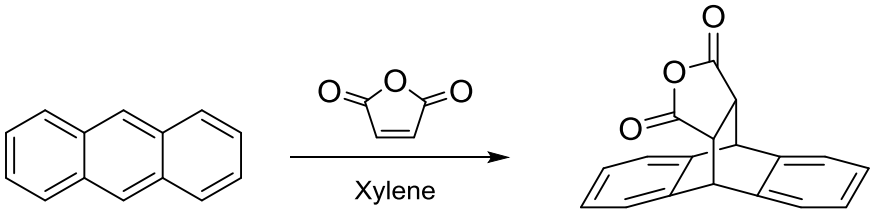
137.47 (Ar-C), 134.20 (Ar-C), 131.46 (Ar-C), 128.99 (Ar-C), 129.28 (Ar-C), 128.91 (2 *para*-Ar-C), 128.01 (Ar-C), 127.94 (Ar-C), 125.90 (Ar-C), 125.41 (Ar-C), 123.73 (Ar-C), 120.73 (Ar-C), 95.62 (Ar-CI), 93.74 ($CNO_2$), 49.59 (CH dihydroanthracene), 47.97 (CH succinimide), 45.28 (CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 2968, 2901 (CH stretch), 1778, 1706 (C=O stretch), 1550 ($NO_2$ asymmetric), 1387 ($NO_2$ symmetric). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 523, $C_{24}H_{15}IN_2O_4$ requires 522.

**Synthesis of 13-(4-(*tert*-butyl)phenyl)-9-nitro-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (24s) [R = H, R′ = *t*-Bu]**.

Compound **119** (0.10 g, 0.32 mmol, 1 eq) was combined with *t*-butylaniline (0.06 g, 0.38 mmol, 1.2 eq) in 15 mL glacial acetic acid according to the procedure described above. Recrystallisation in glacial acetic acid afforded **24s** as a cream solid, yield 0.13 g, 89%. **M.p.** 238-240 $^oC$. **$^1$H-NMR (400 MHz, DCM-$d_2$) δ (ppm)** 7.72 - 7.75 (1 H, m, Ar-CH), 7.51 - 7.55 (1 H, m, Ar-CH), 7.28 - 7.43 (7 H, m, Ar-CH), 7.08 - 7.12 (1 H, m, Ar-CH), 6.36 - 6.41 (2 H, m, *para*-Ar-CH), 4.92 (1 H, d, *J* = 3.1 Hz, CH dihydroanthracene), 4.43 (1 H, d, *J* = 9.0 Hz, CH succinimide), 3.53 (1 H, dd, *J* = 8.9, 3.1 Hz, CH succinimide), 1.26 (9 H, s, 3 $CH_3$). **$^{13}$C-NMR (150 MHz, DCM-$d_2$) δ (ppm)** 174.33 (C=O), 172.92 (C=O), 152.38 (Ar-C-C), 139.09 (Ar-C), 137.23 (Ar-C), 136.57 (Ar-C), 133.95 (Ar-C), 128.82 (Ar-C), 128.58 (Ar-C), 128.42 (Ar-C), 127.72 (Ar-C), 127.61 (Ar-C), 126.18 (2 *para*-Ar-C), 125.90 (2 *para*-Ar-C), 125.37 (Ar-C), 124.34 (Ar-C), 123.89 (Ar-C), 120.86 (Ar-C), 93.47 ($CNO_2$), 48.98 (CH dihydroanthracene), 47.74 (CH succinimide), 45.97 (CH succinimide), 34.65 (*t*-Bu-C),

30.92 (3 $CH_3$). **IR $\nu_{max}$/cm$^{-1}$**: 2981, 2902, 2875 (CH stretch), 1783, 1714 (C=O stretch), 1547 ($NO_2$ asymmetric), 1392 ($NO_2$ symmetric). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 453, $C_{28}H_{24}N_2O_4$ requires 452.

#### 4.5.2.2 **Synthesis of 9,10-dihydro-9,10-[3,4]furanoanthracene-12,14-dione (120)**



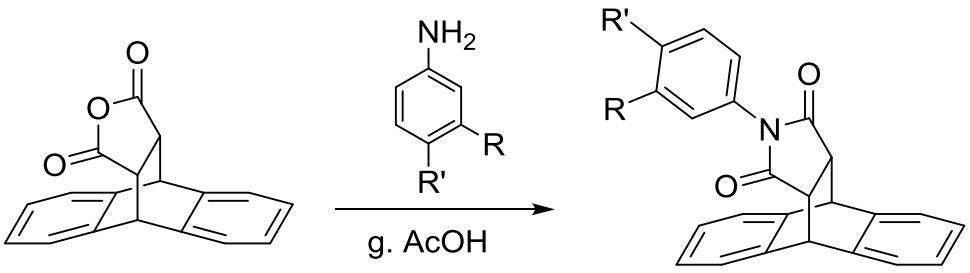

**120**

$C_{18}H_{12}O_3$
MW: 276.29

Maleic anhydride (1.00 g, 10 mmol) was added to a stirring solution to anthracene (1.82 g, 10.2 mmol) in xylene (25 mL) and the mixture was heated under reflux at 140-143 °C until all the diene disappeared as indicated by TLC. The reaction mixture was cooled to room temperature before placing on ice while stirring. The precipitate was filtered, washed with ice-cold xylene to give the desired Diels-Alder cycloadduct **120** as sand-coloured granules, yield 2.69, 95%. The cycloadduct was used in the next step without further purification. $R_f$ 0.87 ($Et_2O$); M.p. 259-260 °C (lit. 261-263 °C [404]; 265-266 °C [374], 253-256 °C [405]). **$^1$H-NMR (400 MHz, acetone-$d_6$) δ (ppm)** 7.47 - 7.51 (2 H, m, Ar-CH), 7.33 - 7.38 (2 H, m, Ar-CH), 7.17 - 7.22 (4 H, m, Ar-CH), 4.89 - 4.92 (2 H, m, CH dihydroanthracene), 3.73 - 3.75 (2 H, m, CH succinimide). **$^{13}$C-NMR (100 MHz, acetone-$d_6$) δ (ppm)** 171.22 (2 C=O), 141.51 (2 Ar-C), 139.41 (2 Ar-C), 127.23 (2 Ar-C), 126.79 (2 Ar-C), 125.09 (2 Ar-C), 124.48 (2

Ar-C), 48.24 (2 CH dihydroanthracene), 45.29 (2 CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 3077, 3026, 2970 (C-H stretch), 1782 (C=O stretch). LC-MS (ESI) found ***m/z*** [M – H + MeOH]$^-$: 307, $C_{18}H_{12}O_3$ requires 276.

4.5.2.2.1 Synthesis of 3-(12,14-dioxo-9,10-dihydro-9,10-[3,4]epipyrroloanthracen-13-yl)benzoic acid (**121**) and analogues




**120** **121-121s**

| Compound | R | R' | Compound | R | R' |
|---|---|---|---|---|---|
| **121** | $CO_2H$ | H | **121j** | H | $CO_2H$ |
| **121a** | $CO_2Et$ | H | **121k** | H | $CO_2Et$ |
| **121b** | F | H | **121l** | H | F |
| **121c** | Cl | H | **121m** | H | Cl |
| **121d** | $NO_2$ | H | **121n** | H | $NO_2$ |
| **121e** | H | H | **121o**† | H | $CH_3$ |
| **121f**† | $CH_3$ | H | **121p**† | H | MeO |
| **121g**† | MeO | H | **121q**† | H | $Me_2N$ |
| **121h**† | $Me_2N$ | H | **121r** | H | I |
| **121i**† | CN | H | **121s** | H | *t*-Bu |

†Compounds synthesised by Kulikowska [372] and Grewal [373] as part of their final year project for the degree of Master of Pharmacy, University of Hertfordshire

**General procedure for the synthesis of 121-121s**

Maleic anhydride-anthracene cycloadduct (**120**, 1 eq), and the appropriate 3- or 4-substituted aniline (1.2-1.5 eq) were stirred under reflux in glacial acetic acid (10 mL) between 1-5 h then cooled to room temperature. In some reactions, a precipitate formed upon cooling to room temperature (**121**, **121b**, **121c**, **121d**, **121e**, **121j**, **121k**, **121l**, **121m**, **121n**, **121r**) while none was present for others (**121a**, **121f**, **121g**, **121h**, **121i**, **121o**, **121p**, **121q**, **121s**). De-ionised water (25-40 mL) was added to the reaction mixture, which resulted in a precipitate forming. This was filtered, washed with ice-cold de-ionised water and dried. Recrystallisation was carried out in glacial acetic acid. Yields shown are from the final Diels-Alder step, not from overall synthesis. Although analogues **121**, **121d**, **121e**, **121j**, **121n**, **121r**, **121p**, and **121q** have been previously synthesised, their melting points were however not reported by the authors [365-368].

**Synthesis of 3-(12,14-dioxo-9,10-dihydro-9,10-[3,4]epipyrroloanthracen-13-yl)benzoic acid (121) [R = $CO_2H$, R′ = H]**.

Compound **120** (0.08 g, 0.29 mmol, 1 eq) was combined with 3-aminobenzoic acid (0.06 g, 0.41 mmol, 1.4 eq) in 10 mL glacial acetic acid according to the procedure described above. Compound **121** was obtained as a white solid, yield 0.11 g, 95%. M.p. 285-288 $^{o}$C. **$^{1}$H-NMR (400 MHz, acetone-$d_6$) δ (ppm)** 7.93 - 7.97 (1 H, m, Ar-H), 7.52 (1 H, dd, *J* = 5.4, 3.3 Hz, Ar-H), 7.41 - 7.48 (1 H, m, Ar-H), 7.28 - 7.35 (4 H, m, Ar-H), 7.21 (4 H, m, Ar-H), 6.73 - 6.79 (1 H, m, Ar-H), 4.88 - 4.91 (2 H, m, CH dihydroanthracene), 3.47 (2 H, m, CH succinimide ), 1.95 (1 H, s, OH). **$^{13}$C-NMR (100**

**MHz, acetone-$d_6$) δ (ppm)** 175.60 (2 C=O), 165.77 (C=O acid), 141.99 (2 Ar-C), 139.65 (2 Ar-C), 132.79 (Ar-C), 131.44 (Ar-C), 131.12 (Ar-C), 129.34 (Ar-C), 128.93 (Ar-C), 128.10 (Ar-C), 126.84 (2 Ar-C), 126.60 (2 Ar-C), 125.00 (2 Ar-C), 124.37 (2 Ar-C), 47.23 (2 CH dihydroanthracene), 45.82 (2 CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 3321 (OH stretch), 1774, 1704 (C=O stretch). LC-MS (ESI) found ***m/z*** $[M - H]^-$: 394, $C_{25}H_{17}NO_4$ requires 395.

**Synthesis of ethyl 3-(12,14-dioxo-9,10-dihydro-9,10-[3,4]epipyrroloanthracen-13-yl)benzoate (121a) [R = $CO_2$Et, R′ = H]**.

Compound **120** (0.11 g, 0.41 mmol, 1 eq) was combined with ethyl-3-aminobenzoate (0.08 g, 0.51 mmol, 1.3 eq) in 15 mL glacial acetic acid according to the procedure described above. Compound **121a** was obtained as a pale pink powder, yield 0.17 g, 96%. **M.p.** 205-207 $^oC$. **$^1$H-NMR (400 MHz, acetone-$d_6$) δ ppm** 7.89 - 7.96 (1 H, m, Ar-H), 7.49 - 7.55 (2 H, m, Ar-H), 7.45 (1 H, t, *J* = 7.9 Hz, Ar-H), 7.30 - 7.37 (2 H, m, Ar-H), 7.15 - 7.26 (5 H, m, Ar-H), 6.76 - 6.82 (1 H, m, Ar-H), 4.89 (2 H, m, CH dihydroanthracene), 4.33 (2 H, q, *J* = 7.1 Hz, $CH_2$ aliphatic), 3.47 (2 H, dd, *J* = 2.1, 1.3 Hz, CH succinimide), 1.36 (3 H, t, *J* = 7.1 Hz, $CH_3$). **$^{13}$C-NMR (150 MHz, acetone-$d_6$) δ (ppm)** 175.63 (2 C=O), 164.90 (C=O ester), 132.81 (Ar-C), 131.39 (Ar-C), 132.81 (Ar-C), 131.39 (Ar-C), 131.29 (Ar-C), 129.12 (Ar-C), 129.02 (2 Ar-C), 127.83 (2 Ar-C), 126.84 (2 Ar-C), 126.61 (2 Ar-C), 125.01 (2 Ar-C), 124.37 (2 Ar-C), 60.94 ($CH_2$ aliphatic), 47.24 (2 CH dihydroanthracene), 45.81 (2 CH succinimide), 13.73 ($CH_3$). **IR**

**$\nu_{max}$/cm$^{-1}$**: 3070, 3042, 2992 (CH stretch), 1775, 1718, 1703 (C=O stretch). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 424, $C_{27}H_{21}NO_4$ requires 423.

**Synthesis of 13-(3-fluorophenyl)-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (121b) [R = F, R′ = H]**.

Compound **120** (0.10 g, 0.37 mmol, 1 eq) was combined with 3-fluoroaniline (0.05 g, 0.44 mmol, 1.2 eq) in 10 mL glacial acetic acid according to the procedure described above. Compound **121b** was obtained as white flakes, yield 0.12 g, 87%. **M.p.** 228-230 $^oC$. **$^1$H-NMR (400 MHz, DCM-$d_2$) δ (ppm)** 7.41 - 7.45 (2 H, m, Ar-H), 7.17 - 7.34 (7 H, m, Ar-H), 6.97 - 7.05 (1 H, m, Ar-H), 6.29 - 6.36 (1 H, m, Ar-H), 6.24 (1 H, m, Ar-H), 4.83 - 4.89 (2 H, m, CH dihydroanthracene), 3.35 - 3.41 (2 H, m, CH succinimide). **$^{13}$C-NMR (100 MHz, DCM-$d_2$) δ (ppm)** 175.60 (2 C=O), 163.67 (Ar-CF), 141.40 (2 Ar-C), 139.01 (2 Ar-C), 133.8 (Ar-C), 130.19 (Ar-C), 127.18 (2 Ar-C), 126.83 (2 Ar-C), 125.07 (2 Ar-C), 124.38 (2 Ar-C), 122.42 (Ar-C), 115.60 (Ar-C), 114.15 (Ar-C), 47.13 (2 CH dihydroanthracene), 45.89 (2 CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 3083, 3041, 3020, 2974 (CH stretch), 1774, 1702 (C=O stretch). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 370, $C_{24}H_{16}FNO_2$ requires 369.

**Synthesis of 13-(3-chlorophenyl)-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (121c) [R = Cl, R′ = H]**.

Compound **120** (0.09 g, 0.31 mmol, 1 eq) was combined with 3-chloroaniline (0.05 g, 0.37 mmol, 1.2 eq) in 10 mL glacial acetic acid according to the procedure described above. Compound **121c** was obtained as a cream solid, yield 0.11 g, 92%. M.p. 238-240 $^{o}$C. **$^{1}$H-NMR (400 MHz, DMSO-*d*$_6$) δ (ppm)** 7.49 - 7.54 (2 H, m, Ar-H), 7.36 - 7.45 (2 H, m, Ar-H), 7.30 - 7.34 (2 H, m, Ar-H), 7.17 - 7.25 (4 H, m, Ar-H), 6.42 - 6.49 (2 H, m, Ar-H), 4.87 (2 H, s, CH dihydroanthracene), 3.42 (2 H, m, CH succinimide). **$^{13}$C-NMR (100 MHz, DMSO-*d*$_6$) δ (ppm)** 176.24 (2 C=O), 142.05 (2 Ar-C), 139.86 (2 Ar-C), 133.63 (Ar-C), 133.46 (Ar-C), 131.21 (Ar-C), 129.15 (Ar-C), 127.22 (2 Ar-C), 126.98 (3 Ar-C), 125.95 (Ar-C), 125.38 (2 Ar-C), 124.98 (2 Ar-C), 47.29 (2 CH dihydroanthracene), 45.44 (2 CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 3077 (CH stretch), 2960 (CH stretch), 1774, 1708 (C=O stretch). LC-MS (ESI) found ***m/z*** [M + H]$^{+}$: 386, $C_{24}H_{16}ClNO_2$ requires 385.

**Synthesis of 13-(3-nitrophenyl)-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (121d) [R = NO$_2$, R′ = H]**.

Compound **120** (0.10 g, 0.38 mmol, 1 eq) was combined with 3-nitroaniline (0.08 g, 0.57 mmol, 1.5 eq) in 10 mL glacial acetic acid according to the procedure described above. Compound **121d** was obtained as a white solid, yield 0.14 g, 95%. **Mp** 282-285 $^{o}$C. **$^{1}$H-NMR (400 MHz, DCM-*d*$_2$) δ (ppm)** 8.11 - 8.15 (1 H, m, Ar-CH), 7.42 - 7.52 (3 H, m, Ar-H), 7.31 - 7.38 (3 H, m, Ar-H), 7.19 - 7.27 (4 H, m, Ar-CH), 6.92 - 7.13 (1 H, m, Ar-H), 4.88 (2 H, s, CH dihydroanthracene), 3.40 - 3.44 (2 H, m, CH succinimide). **$^{13}$C-NMR (150 MHz, DCM-*d*$_2$) δ (ppm)** 175.41 (2 C=O), 148.40 (Ar-

$CNO_2$), 141.19 (2 Ar-C), 138.89 (2 Ar-C), 132.65 (Ar-C), 132.54 (Ar-C), 129.91 (Ar-C), 127.30 (2 Ar-C), 126.90 (2 Ar-C), 125.09 (2 Ar-C), 124.41 (2 Ar-C), 123.36 (Ar-C), 121.77 (Ar-C), 47.21 (2 CH dihydroanthracene), 45.93 (2 CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 3081, 3048, 3011, 2957 (CH stretch), 1775, 1710 (C=O stretch), 1530 ($NO_2$ asymmetric), 1387 ($NO_2$ symmetric). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 397, $C_{24}H_{16}N_2O_4$ requires 396.

**Synthesis of 13-phenyl-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (121e) [R = H, R′ = H]**.

Compound **120** (0.05 g, 0.18 mmol, 1 eq) was combined with aniline (0.02 g, 0.21 mmol, 1.2 eq) in 10 mL glacial acetic acid according to the procedure described above. Compound **121e** was obtained as a white solid, yield 0.05 g, 83%. **Mp** 212-215 $^oC$. **$^1$H-NMR (400 MHz, DMSO-$d_6$) δ (ppm)** 7.51 (2 H, dd, *J* = 5.3, 3.3 Hz, Ar-H), 7.28 - 7.36 (5 H, m, Ar-H), 7.21 (4 H, m, Ar-H), 6.37 - 6.47 (2 H, m, Ar-H), 4.86 (2 H, s, CH dihydroanthracene), 3.41 (2 H, s, CH succinimide). **$^{13}$C-NMR (150 MHz, DCM-$d_2$) δ (ppm)** 175.97 (2 C=O), 141.51 (2 Ar-C), 139.07 (2 Ar-C), 131.79 (Ar-C), 128.97 (2 *para*-Ar-C), 128.70 (Ar-C), 127.09 (2 Ar-C), 126.76 (2 *para*-Ar-C), 126.61 (2 Ar-C), 125.05 (2 Ar-C), 124.33 (2 Ar-C), 47.13 (2 CH dihydroanthracene), 45.89 (2 CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 3069, 3038, 2969 (CH stretch), 1776, 1710 (C=O stretch). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 352, $C_{24}H_{17}NO_2$ requires 351.

**Synthesis of 13-(m-tolyl)-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (121f) [R = $CH_3$, R′ = H]**.

Compound **120** (0.50 g, 0.81 mmol, 1 eq) was combined with *m*-toluidine (0.19 g, 1.77 mmol, 1 eq) in 20 mL glacial acetic acid according to the procedure described above. Compound **121f** was obtained as a cream solid, yield 0.40 g, 60%. **Mp** 199-200 °C. **$^1$H-NMR (400 MHz, acetone-$d_6$) δ ppm** 7.47 - 7.55 (2 H, m, Ar-H), 7.29 - 7.34 (2 H, m, Ar-H), 7.19 - 7.25 (4 H, m, Ar-H), 7.13 - 7.19 (1 H, m, Ar-H), 7.07 - 7.11 (1 H, m, Ar-H), 6.29 - 6.33 (1 H, m, Ar-H), 6.26 - 6.29 (1 H, m, Ar-H), 4.87 (2 H, m, CH dihydroanthracene), 3.42 (2 H, m, CH succinimide), 2.22 (3 H, s, $CH_3$). **$^{13}$C-NMR (150 MHz, acetone-$d_6$) δ (ppm)** 175.71 (2 C=O), 142.07 (2 Ar-C), 139.74 (2 Ar-C), 138.49 (Ar-C), 132.55 (Ar-C), 128.90 (Ar-C), 128.42 (Ar-C), 127.42 (Ar-C), 126.72 (2 Ar-C), 126.55 (2 Ar-C), 125.02 (2 Ar-C), 124.32 (2 Ar-C), 123.91 (Ar-C), 47.09 (2 CH dihydroanthracene), 45.82 (2 CH succinimide), 20.23 ($CH_3$). **IR $\nu_{max}$/cm$^{-1}$**: 3049, 3014, 2961, 2916 (CH stretch), 1773, 1703 (C=O stretch). LC-MS (ESI) found ***m/z*** $[M - H]^-$: 364, $C_{25}H_{19}NO_2$ requires 365.

**Synthesis of 13-(3-methoxyphenyl)-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (121g) [R = MeO, R′ = H]**.

Compound **120** (0.50 g, 1.81 mmol, 1 eq) was combined with 3-methoxyaniline (0.27 g, 2.19 mmol, 1.2 eq) in 20 mL glacial acetic acid according to the procedure described above. Compound **121g** was obtained as a sand-coloured solid, yield 0.63 g,

91%. **Mp** 200-201 °C. **$^{1}$H-NMR (400 MHz, acetone-$d_6$) δ (ppm)** 7.46 - 7.55 (2 H, m, Ar-H), 7.29 - 7.36 (2 H, m, Ar-H), 7.16 - 7.26 (5 H, m, Ar-H), 6.85 - 6.93 (1 H, m, Ar-H), 6.17 - 6.23 (1 H, m, Ar-H), 5.88 - 5.94 (1 H, m, Ar-H), 4.87 (2 H, m, CH dihydroanthracene), 3.68 (3 H, s, $CH_3$), 3.42 (2 H, m, CH succinimide). **$^{13}$C-NMR (150 MHz, DMSO-$d_6$) δ (ppm)** 176.37 (2 C=O), 159.89 (Ar-CO), 142.12 (2 Ar-C), 139.85 (2 Ar-C), 133.45 (Ar-C), 130.20 (Ar-C), 127.15 (2 Ar-C), 126.90 (2 Ar-C), 125.35 (2 Ar-C), 124.90 (2 Ar-C), 119.41 (Ar-C), 114.73 (Ar-C), 112.83 (Ar-C), 55.84 (C-MeO), 47.16 (2 CH dihydroanthracene), 45.43 (2 CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 3011, 2948, 2905, 2832 (CH stretch), 1778, 1713 (C=O stretch), 1025 (CO stretch). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 382, $C_{25}H_{19}NO_3$ requires 381.

**Synthesis of 13-(3-(dimethylamino)phenyl)-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (121h) [R = Me$_2$N, R′ = H]**.

Compound **120** (0.50 g, 1.81 mmol, 1 eq) was combined with *N*,*N*-dimethy-1,3-diamine (0.45 g, 2.15 mmol, 1.2 eq) in 20 mL glacial acetic acid according to the procedure described above. Compound **121h** was obtained as cream crystals, yield 0.67 g, 94%. **M.p.** 255-256 °C. **$^{1}$H-NMR (400 MHz, DCM-$d_2$) δ (ppm)** 7.41 - 7.46 (2 H, m, Ar-H), 7.31 - 7.36 (2 H, m, Ar-H), 7.17 - 7.24 (4 H, m, Ar-H), 7.11 (1 H, t, *J* = 8.1 Hz, Ar-H), 6.61 (1 H, dd, *J* = 8.4, 2.6 Hz, Ar-H), 5.85 (1 H, m, Ar-H), 5.40 (1 H, t, *J* = 2.1 Hz, Ar-H), 4.83 - 4.88 (2 H, m, CH dihydroanthracene), 3.32 - 3.37 (2 H, m, CH succinimide), 2.82 (6 H, s, $CH_3$). **$^{13}$C-NMR (150 MHz, DCM-$d_2$) δ (ppm)** 176.14 (2 C=O), 151.19 (Ar-CN), 141.57 (2 Ar-C), 139.22 (2 Ar-C), 132.71 (Ar-C), 129.34 (Ar-

C), 126.98 (2 Ar-C), 126.73 (2 Ar-C), 125.11 (2 Ar-C), 124.33 (2 Ar-C), 114.08 (Ar-C), 112.57 (Ar-C), 110.35 (Ar-C), 47.12 (2 CH dihydroanthracene), 45.93 (2 CH succinimide), 40.17 (2 $CH_3$). **IR $\nu_{max}$/cm$^{-1}$**: 3474 (NH stretch), 3031 2960 2880 2804 (CH stretch), 1777, 1713 (C=O stretch). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 395, $C_{26}H_{22}N_2O_2$ requires 394.

**Synthesis of 3-(12,14-dioxo-9,10-dihydro-9,10-[3,4]epipyrroloanthracen-13-yl)benzonitrile (121i) [R = CN, R′ = H]**.

Compound **120** (0.50 g, 1.81 mmol, 1 eq) was combined with 3-aminobenzonitrile (0.26 g, 2.20 mmol, 1.2 eq) in 20 mL glacial acetic acid according to the procedure described above. Compound **121i** was obtained as a white solid, yield 0.67 g, 98%. **M.p.** 262-264 °C. **$^1$H-NMR (600 MHz, DMSO-$d_6$) δ (ppm)** 7.75 - 7.83 (1 H, m, Ar-H), 7.52 - 7.59 (1 H, m, Ar-H), 7.45 - 7.51 (2 H, m, Ar-H), 7.26 - 7.32 (2 H, m, Ar-H), 7.14 - 7.21 (4 H, m, Ar-H), 6.81 - 6.86 (1 H, m, Ar-H), 6.72 - 6.74 (1 H, m, Ar-H), 4.85 (2 H, m, CH dihydroanthracene), 3.42 (2 H, m, CH succinimide). **$^{13}$C-NMR (150 MHz, DMSO-$d_6$) δ (ppm)** 176.10 (2 C=O), 141.95 (2 Ar-C), 139.80 (2 Ar-C), 133.03 (Ar-C), 132.92 (Ar-C), 131.97 (Ar-C), 131.10 (Ar-C), 130.29 (Ar-C), 127.23 (2 Ar-C), 126.96 (2 Ar-C), 125.36 (2 Ar-C), 124.97 (2 Ar-C), 118.13 (Ar-C), 112.42 (Ar-C), 47.28 (2 CH dihydroanthracene), 45.41 (2 CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 3076, 29750 (CH stretch), 1785, 1713 (C=O stretch). LC-MS (ESI) found ***m/z*** $[M - H]^-$: 375, $C_{25}H_{16}N_2O_2$ requires 376.

**Synthesis of 4-(12,14-dioxo-9,10-dihydro-9,10-[3,4]epipyrroloanthracen-13-yl)benzoic acid (121j) [R = H, R′ = $CO_2H$]**.

Compound **120** (0.50 g, 1.81 mmol, 1 eq) was combined with 4-aminobenzoic acid (0.30 g, 2.17 mmol, 1.2 eq) in 15 mL glacial acetic acid according to the procedure described above. Precipitate formed after 25 minutes of reaction. Compound **121j** was obtained as a white solid. Yield 0.61 g, 86%. **M.p.** 354-355 $^oC$. **$^1$H-NMR (400 MHz, acetone-$d_6$) δ (ppm)** 7.95 (2 H, d, *J* = 8.8 Hz, *para*-Ar-H), 7.50 - 7.54 (2 H, m, Ar-H), 7.31 - 7.35 (2 H, m, Ar-H), 7.19 - 7.23 (4 H, m, Ar-H), 6.72 (2 H, d, *J* = 8.6 Hz, *para*-Ar-H), 4.89 (2 H, d, *J* = 1.6 Hz, CH dihydroanthracene), 3.47 (2 H, m, CH succinimide). **$^{13}$C-NMR (100 MHz, acetone-$d_6$) δ (ppm)** 175.66 (2 C=O), 171.39 (C=O acid), 140.26 (2 Ar-C), 139.65 (2 Ar-C), 133.07 (Ar-C), 129.97 (2 *para*-Ar-C), 128.64 (Ar-C), 127.27 (Ar-C), 126.62 (2 Ar-C), 125.01 (2 Ar-C), 124.75 (Ar-C), 124.37 (2 *para*-Ar-C), 110.79 (2 Ar-C), 47.21 (2 CH dihydroanthracene), 45.83 (2 CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 3283 (OH stretch), 3020, 2972 (CH stretch), 1771, 1728, 1698 (C=O stretch), 1105 (C-O stretch). LC-MS (ESI) found ***m/z*** $[M - H]^-$: 394, $C_{25}H_{17}NO_4$ requires 395.

**Synthesis of ethyl 4-(12,14-dioxo-9,10-dihydro-9,10-[3,4]epipyrroloanthracen-13-yl)benzoate (121k) [R = H, R′ = $CO_2$Et]**.

Compound **120** (0.15 g, 0.56 mmol, 1 eq) was combined with ethyl-4-aminobenzoate (0.11 g, 0.69 mmol, 1.2 eq) in 10 mL glacial acetic acid according to the procedure described above. Compound **121k** was obtained as a white solid, yield 0.22 g,

92%. **M.p.** 220-222 ºC. **$^{1}$H-NMR (400 MHz, acetone-$d_6$) δ (ppm)** 7.89 - 7.98 (2 H, m, *para*-Ar-H), 7.45 - 7.56 (2 H, m, Ar-H), 7.28 - 7.38 (2 H, m, Ar-H), 7.16 - 7.24 (4 H, m, Ar-H), 6.67 - 6.76 (2 H, m, *para*-Ar-H), 4.78 - 4.99 (2 H, m, CH dihydroanthracene), 4.32 (2 H, q, *J* = 7.1 Hz, $CH_2$ aliphatic), 3.47 (2 H, m, CH succinimide), 1.33 (3 H, t, *J* = 7.1 Hz, $CH_3$). **$^{13}$C-NMR (100 MHz, acetone-$d_6$) δ (ppm)** 175.42 (2 C=O), 165.09 (C=O ester), 141.95 (2 Ar-C), 139.62 (2 Ar-C), 136.42 (Ar-C), 130.27 (Ar-C), 129.65 (2 *para*-Ar-C), 126.85 (2 Ar-C), 126.67 (2 *para*-Ar-C), 126.61 (2 Ar-C), 124.99 (2 Ar-C), 124.37 (2 Ar-C), 60.86 ($CH_2$ aliphatic), 47.20 (2 CH dihydroanthracene), 45.82 (2 CH succinimide), 13.67 ($CH_3$). **IR $\nu_{max}$/cm$^{-1}$**: 3132, 3067, 2958 (CH stretch), 1777, 1707, 1608 (C=O stretch). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 424, $C_{27}H_{21}NO_4$ requires 423.

**Synthesis of 13-(4-fluorophenyl)-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (121l) [R = H, R′ = F]**.

Compound **120** (0.13 g, 0.48 mmol, 1 eq) was combined with 4-fluoroaniline (0.06 g, 0.57 mmol, 1.2 eq) in 10 mL glacial acetic acid according to the procedure described above. Compound **121l** was obtained as a pale pink solid, yield 0.17 g, 96%. **M.p.** 250-253 ºC. **$^{1}$H-NMR (400 MHz, DCM-$d_2$) δ (ppm)** 7.40 - 7.46 (2 H, m, Ar-H), 7.29 - 7.34 (2 H, m, Ar-H), 7.17 - 7.24 (4 H, m, Ar-H), 6.94 - 7.02 (2 H, m, Ar-H), 6.42 - 6.49 (2 H, m, Ar-H), 4.82 - 4.88 (2 H, m, CH dihydroanthracene), 3.35 - 3.38 (2 H, m, CH succinimide). **$^{13}$C-NMR (100 MHz, DCM-$d_2$) δ (ppm)** 175.93 (2 C=O), 141.41 (2 Ar-C), 139.05 (2 Ar-C), 128.52 (Ar-C), 128.43 (2 *para*-Ar-C), 127.12 (2 Ar-C), 126.80 (2 Ar-C), 125.05 (2 Ar-C), 124.37 (2 Ar-C), 116.07 (Ar-C), 115.84 (2 *para*-Ar-C),

47.10 (2 CH dihydroanthracene), 45.88 (2 CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 2975, 2891 (CH stretch), 1777, 1708 (C=O stretch). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 370, $C_{24}H_{16}FNO_2$ requires 369.

**Synthesis of 13-(4-chlorophenyl)-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (121m) [R = H, R′ = Cl]**.

Compound **120** (0.15 g, 0.53 mmol, 1 eq) was combined with 4-chloroaniline (0.09 g, 0.67 mmol, 1.3 eq) in 10 mL glacial acetic acid according to the procedure described above. Compound **121m** was obtained as fluffy white powder, yield 0.19 g, 94%. **M.p.** 275-278 $^oC$. **$^1$H-NMR (400 MHz, DCM-$d_2$) δ (ppm)** 7.37 - 7.48 (2 H, m, *para*-Ar-H), 7.10 - 7.37 (8 H, m, Ar-H), 6.40 - 6.49 (2 H, m, *para*-Ar-H), 4.85 (2 H, s, CH dihydroanthracene), 3.31 - 3.41 (2 H, m, CH succinimide). **$^{13}$C-NMR (150 MHz, DCM-$d_2$) δ (ppm)** 175.71 (2 C=O), 141.38 (2 Ar-C), 139.00 (2 Ar-C), 134.41 (Ar-C), 130.24 (Ar-C), 129.16 (2 *para*-Ar-C), 127.86 (2 *para*-Ar-C), 127.13 (2 Ar-C), 126.80 (2 Ar-C), 125.04 (2 Ar-C), 124.36 (2 Ar-C), 47.10 (2 CH dihydroanthracene), 45.88 (2 CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 2981, 2973, 2885 (CH stretch), 1774, 1702 (C=O stretch). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 386, $C_{24}H_{16}ClNO_2$ requires 385.

**Synthesis of 13-(4-nitrophenyl)-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (121n) [R = H, R′ = $NO_2$]**.

Compound **120** (0.10 g, 0.37 mmol, 1 eq) was combined with 4-nitroanline (0.06 g, 0.45 mmol, 1.2 eq) in 10 mL glacial acetic acid according to the procedure described

above. Compound **121n** was obtained as cream crystals, yield 0.12 g, 83%. **M.p.** 294-295 $^{o}$C. **$^{1}$H-NMR (400 MHz, acetone-$d_6$) δ (ppm)** 8.17 - 8.23 (2 H, m, *para*-Ar-H), 7.49 - 7.55 (2 H, m, Ar-H), 7.31 - 7.37 (2 H, m, Ar-H), 7.17 - 7.25 (4 H, m, Ar-H), 6.89 - 6.96 (2 H, m, *para*-Ar-H), 4.85 - 4.96 (2 H, m, CH dihydroanthracene), 3.51 (2 H, m, CH succinimide). **$^{13}$C-NMR (151 MHz, acetone-$d_6$) δ ppm** 175.28 (2 C=O), 147.17 (Ar-$CNO_2$), 141.83 (2 Ar-C), 139.56 (2 Ar-C), 137.85 (Ar-C), 127.45 (2 *para*-Ar-C), 126.92 (2 Ar-C), 126.66 (2 Ar-C), 124.98 (2 *para*-Ar-C), 124.41 (2 Ar-C), 123.97 (2 Ar-C), 47.24 (2 CH dihydroanthracene), 45.77 (2 CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 3480, 3121, 3082, 3017, 2973, 2860 (CH stretch), 1777, 1708 (C=O stretch), 1522 ($NO_2$ asymmetric), 1344 ($NO_2$ symmetric). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 397, $C_{24}H_{16}N_2O_4$ requires 396.

**Synthesis of 13-(*p*-tolyl)-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (121o) [R = H, R′ = $CH_3$]**.

Compound **120** (0.50 g, 1.81 mmol, 1 eq) was combined with *p*-toluidine (0.19 g, 1.77 mmol, 1 eq) in 20 mL glacial acetic acid according to the procedure described above. Compound **121o** was obtained as a white solid, yield 0.62 g, 94%. **M.p.** 220-222 °C. **$^{1}$H-NMR (400 MHz, DMSO-$d_6$) δ (ppm)** 7.49 - 7.53 (2 H, m, Ar-H), 7.27 - 7.32 (2 H, m, Ar-H), 7.20 (4 H, m, Ar-CH), 7.12 (2 H, m, *para*-Ar-H), 6.21 - 6.39 (2 H, m, *para*-Ar-H), 4.85 (2 H, s, CH dihydroanthracene), 3.38 - 3.40 (2 H, m, CH succinimide), 2.25 (3 H, $CH_3$). **$^{13}$C-NMR (100 MHz, DMSO-$d_6$) δ (ppm)** 176.59 (2 C=O), 142.21 (2 Ar-C), 139.86 (2 Ar-C), 138.58 (Ar-C), 129.83 (2 *para*-Ar-C), 129.77 (Ar-C), 127.17 (2

*para*-Ar-C), 126.92 (2 Ar-C), 126.84 (2 Ar-C), 125.35 (2 Ar-C), 124.93 (2 Ar-C), 47.13 (2 CH dihydroanthracene), 45.41 (2 CH succinimide), 21.20 ($CH_3$). **IR $\nu_{max}$/cm$^{-1}$:** 3075, 3048, 2982, 2928, 2862 (CH stretch), 1768 1704 (C=O stretch). LC-MS (ESI) found ***m/z*** $[M - H]^-$: 364, $C_{25}H_{19}NO_2$ requires 365.

**Synthesis of 13-(4-methoxyphenyl)-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (121p) [R = H, R′ = MeO]**.

Compound **120** (0.50 g, 1.81 mmol, 1 eq) was combined with 4-methoxyanline (0.27 g, 2.19 mmol, 1.2 eq) in 20 mL glacial acetic acid according to the procedure described above. Compound **121p** was obtained as a light purple solid, yield 0.60 g, 87%. **M.p.** 240-243 °C. **$^1$H-NMR (400 MHz, DCM-*d*$_2$) δ (ppm)** 7.38 - 7.48 (2 H, m, Ar-H), 7.28 - 7.35 (2 H, m, Ar-H), 7.15 - 7.26 (4 H, m, Ar-H), 6.74 - 6.84 (2 H, m, *para*-Ar-H), 6.31 - 6.40 (2 H, m, *para*-Ar-H), 4.81 - 4.89 (2 H, m, CH dihydroanthracene), 3.74 (3 H, s, MeO), 3.35 (2 H, m, CH succinimide). **$^{13}$C-NMR (150 MHz, DMSO-*d*$_6$) δ (ppm)** 176.36 (2 C=O), 159.22 (Ar-CO), 141.88 (2 Ar-C), 139.56 (2 Ar-C), 127.93 (2 *para*-Ar-CH), 126.81 (2 Ar-C), 126.57 (2 Ar-C), 125.00 (2 Ar-C), 124.57 (3 Ar-C), 114.33 (2 *para*-Ar-CH), 55.58 (MeO), 46.76 (2 CH dihydroanthracene), 45.11 (2 CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 3393, 3012, 2949, 2905, 2832 (CH stretch), 1779, 1710 (C=O stretch). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 380, $C_{25}H_{19}NO_3$ requires 381.

**Synthesis of 13-(4-(dimethylamino)phenyl)-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (121q) [R = H, R′ = $Me_2N$]**.

Compound **120** (0.50 g, 1.81 mmol, 1 eq) was combined with *N*,*N*-dimethyl-1,4-diamine (0.30 g, 2.20 mmol, 1.2 eq) in 20 mL glacial acetic acid according to the procedure described above. Compound **121q** was obtained as a light purple powder, yield 0.68 g, 95%. **Mp** 243-245 °C. **$^1$H-NMR (400 MHz, DMSO-$d_6$) δ (ppm)** 7.48 - 7.53 (2 H, m, Ar-H), 7.26 - 7.31 (2 H, m, Ar-H), 7.16 - 7.23 (4 H, m, Ar-H), 6.53 - 6.63 (2 H, m, *para*-Ar-H), 6.15 - 6.24 (2 H, m, *para*-Ar-H), 4.84 (2 H, m, CH dihydroanthracene), 3.35 (2 H, m, CH succinimide), 2.85 (6 H, s, 2 $CH_3$). **$^{13}$C-NMR (150 MHz, DMSO-$d_6$) δ (ppm)** 176.89 (2 C=O), 150.63 (Ar-CN), 142.25 (2 Ar-C), 139.86 (2 Ar-C), 127.53 (2 *para*-Ar-C), 127.06 (2 Ar-C), 126.84 (2 Ar-C), 125.28 (2 Ar-C), 124.83 (2 Ar-C), 120.68 (Ar-C), 112.38 (2 *para*-Ar-C), 46.94 (2 CH dihydroanthracene), 45.43 (2 CH succinimide), 21.80 (2 $CH_3$). **IR $\nu_{max}$/cm$^{-1}$**: 3477 (NH stretch), 3075, 3040, 3017, 2974, 2864 (CH stretch), 1773, 1698 (C=O stretch). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 395, $C_{26}H_{22}N_2O_2$ requires 394.

**Synthesis of 13-(4-iodophenyl)-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (121r) [R = H, R′ = I]**.

Compound **120** (0.11 g, 0.38 mmol, 1 eq) was combined with 4-iodoaniline (0.10 g, 0.46 mmol, 1.2 eq) in 10 mL glacial acetic acid according to the procedure described above. Compound **121r** was obtained as a white solid, yield 0.16 g, 90%.

**M.p.** 303-307 °C. **$^{1}$H-NMR (400 MHz, DMSO-*d*$_6$) δ (ppm)** 7.68 - 7.75 (2 H, m, *para*-Ar-H), 7.47 - 7.53 (2 H, m, Ar-H), 7.26 - 7.32 (2 H, m, Ar-H), 7.19 (4 H, m, Ar-H), 6.23 - 6.30 (2 H, m, *para*-Ar-H), 4.86 (2 H, s, CH dihydroanthracene), 3.40 (2 H, s, CH succinimide). **$^{13}$C-NMR (100 MHz, DMSO-*d*$_6$) δ (ppm)** 176.23 (2 C=O), 142.12 (Ar-C), 139.81 (Ar-C), 138.35 (2 *para*-Ar-C), 132.01 (Ar-C), 129.03 (2 *para*-Ar-C), 127.22 (2 Ar-C), 126.95 (2 Ar-C), 125.31 (2 Ar-C), 124.94 (2 Ar-C), 95.17 (Ar-CI), 47.24 (2 CH dihydroanthracene), 45.41 (2 CH succinimide). **IR $\nu_{max}$/cm$^{-1}$**: 2973 (CH stretch), 1700 (C=O stretch). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 478, $C_{24}H_{16}INO_2$ requires 477.

**Synthesis of 13-(4-(*tert*-butyl)phenyl)-9,10-dihydro-9,10-[3,4]epipyrroloanthracene-12,14-dione (121s) [R = H, R′ = *t*-Bu]**.

Compound **120** (0.14 g, 0.51 mmol, 1 eq) was combined with *t*-butyl aniline (0.09 g, 0.61 mmol, 1.2 eq) in 10 mL glacial acetic acid according to the procedure described above. Compound **121s** was obtained as a white solid, yield 0.19 g, 93%. **M.p.** 268-271 °C. **$^{1}$H-NMR (400 MHz, DCM-*d*$_2$) δ (ppm)** 7.43 (2 H, m, Ar-H), 7.29 - 7.34 (4 H, m, Ar-H), 7.17 - 7.25 (4 H, m, Ar-H), 6.35 - 6.40 (2 H, m, *para*-Ar-H), 4.84 - 4.86 (2 H, m, CH dihydroanthracene), 3.36 (2 H, m, CH succinimide), 1.26 (9 H, s, $CH_3$). **$^{13}$C-NMR (150 MHz, DCM-*d*$_2$) δ (ppm)** 176.14 (2 C=O), 151.98 (Ar-C), 141.54 (2 Ar-C), 139.10 (2 Ar-C), 129.06 (Ar-C), 127.06 (2 Ar-C), 126.74 (2 Ar-C), 126.08 (2 *para*-Ar-C), 126.03 (2 *para*-Ar-C), 125.05 (2 Ar-C), 124.32 (2 Ar-C), 47.10 (2 CH dihydroanthracene), 45.89 (2 CH succinimide), 34.60 (quaternary C), 30.94 (3 $CH_3$). **IR**

**$\nu_{max}$/cm$^{-1}$**: 2963, 2903, 2868 (CH stretch), 1773, 1707 (C=O stretch). LC-MS (ESI) found ***m/z*** $[M + H]^+$: 408, $C_{28}H_{25}NO_2$ requires 407.

### 4.5.3 General methodology: biological screening

#### 4.5.3.1 Cell culture

All reagents for *in vitro* studies were purchased from Sigma Aldrich (Poole, Dorset, UK) unless otherwise specified. Human pancreatic cancer cells, Panc-1 (ATCC® CRL-1469™) and BxPC-3 (ATCC® CRL-1687™), were purchased from LGC Standards (LGC Limited., Middlesex, UK). Both were grown as monolayers in their respective culture media – Dulbecco Modified Eagle Medium (DMEM) for Panc-1 and RPMI-1640 for BxPC-3 – supplemented with Foetal Bovine Serum (FBS, 10% v/v), Penicillin/Streptomycin solution (200 U), and L-glutamine (2 mM). Both cells were maintained in T75 $cm^2$ flasks at 37 °C in a 5% humidified $CO_2$ atmosphere and passaged at least three times a week when they had reached about 70% confluency using a standardised trypsinisation protocol.

#### 4.5.3.2 Growth of pancreatic cancer cells

To determine the optimal assay conditions for both BxPC-3 and Panc-1 cells, the cells were plated in 96-well plates and their growth monitored over a period of two weeks to obtain "batch growth curve" characteristics. The cells were seeded at a density of 1000, 10,000 and 100,000 cells/well in triplicates overnight. CellTiter 96® AQueous Non-Radioactive Cell Proliferation Assay (MTS) reagent (Promega, UK) was added to the wells at 24 h and the plates incubated at 37 °C in a 5% humidified $CO_2$ atmosphere. Plates were read at an absorbance at 492 nm 90 minutes after addition of the MTS

reagent using a Multiskan Ascent 96/384 plate reader (Thermo Scientific, Loughborough, UK). This procedure was repeated daily over a period of two weeks.

#### 4.5.3.3 **Preparation of compounds**

Cromolyn was purchased as the disodium salt (Sigma Aldrich, Poole, Dorset, UK) with ≥95% purity and the drug solutions were prepared in distilled water (10 Mm). Compounds to be screened (**24**, **24a**, **24e**, **24j** and **121**) were prepared by dissolving in sterile-filtered DMSO Hybri-Max™ (≥99.7%, Sigma Aldrich, Poole, Dorset, UK). The stock solutions of cromolyn and the compounds to be screened were filtered through 0.22 μm sterile filters and stored at –20 °C prior to use. For the proliferation studies, stock solutions (1000 μM) were prepared in the appropriate cell culture medium and serially diluted as required.

#### 4.5.3.4 **Proliferation studies**

Pancreatic cancer cells (1 x $10^4$ cells/mL) in 100 μL complete cell culture medium were plated in clear 96-well plates (Sigma-Aldrich, Poole, UK) and incubated overnight at 37 °C in a 5% $CO_2$ atmosphere. Cell viability was confirmed prior to seeding for the MTS assay study using the Trypan blue exclusion assay. Cromolyn sodium salt (Sigma Aldrich, Poole, UK), compound **24** and its analogues **24a**, **24e**, **24j** and **121** were prepared from 10 mM stock solutions in serum-free medium and serially diluted to give concentrations of 1, 10, 100 and 1000 μM. Control samples for these investigations included cells incubated in the absence of drugs, cells incubated with the addition of the solvent DMSO at the corresponding concentration(s) used for the drug

solutions and media-only samples. Cells were incubated with the test compounds for 24, 48 and 72 h. At each of the three time points, CellTiter 96® AQueous Non-Radioactive Cell Proliferation Assay (MTS) (Promega, Southampton, UK) was added to each well (1:5 v/v). The plates were incubated for 90 minutes before reading the absorbance at 492 nm using a Multiskan Ascent 96/384 plate reader (Thermo Scientific, Loughborough, UK). Each concentration had four replicates. The experiments were repeated twice.

For data analysis, the background absorbance from media-only wells was subtracted from the absorbance from the test wells. The absorbance from control wells, containing DMSO at concentrations of 1, 10, 100 and 1000 μM was subtracted from the absorbance values obtained with wells treated with the sample compounds before an average of all four replicates was taken. The results are expressed as mean ±SEM of the three independent experiments with n = 12. Two-way ANOVA with P values of <0.05(*), <0.01 (**), <0.001 (***) and <0.0001 (****) relative to control were determined using GraphPad Prism 6.00 trial version for Windows (GraphPad Software, San Diego, CA, USA, www.graphpad.com) and post-test comparisons were made using Bonferroni test at 95% confidence interval.

#### 4.5.3.5 CytoTox-ONE™ Homogeneous Membrane Integrity Assay: Lactate Dehydrogenase (LDH) studies

Cell membrane integrity studies were carried out in multiplex with the MTS assay studies. Prior to adding the MTS reagent to the wells, 50 μL of supernatant was removed from the wells of the clear plate and pipetted into a black 96-well plate.

CytoTox-ONE™ Homogeneous Membrane Integrity Assay (Promega, Southampton, UK) was added to each well (50 μL, 1:1 v/v) and the plate incubated at room temperature (~ 22 °C) in the dark for 10 minutes. To stop enzyme activity, 25 μL of Stop Solution (Promega, Southampton, UK) was added to each well and the fluorescence read at an excitation wavelength of 560 nm and an emission wavelength of 590 nm using GloMax®-Multi Detection System plate reader (Promega, Southampton, UK).

For data analysis, background noise (media only wells) was subtracted from all wells. The absorbance of wells containing DMSO at concentrations of 1, 10, 100 and 1000 μM was subtracted from wells containing the sample compounds before an average of all four replicates was taken. The results are expressed as mean ±SEM with of three independent experiments with n =12. P values of <0.05(*), <0.01 (**), <0.001 (***) and <0.0001 (****) were determined relative to control using 2-Way ANOVA with post-test comparisons made using Bonferroni test at the 95% confidence interval.

#### 4.5.3.6 **Chick Chorioallantoic Membrane (CAM) assay for angiogenesis studies**

Fertilised eggs (Dekalb White) were purchased from Henry Stewart & Co. Limited (Fakenham, Norfolk, UK) and left to settle in the laboratory overnight at room temperature. The eggs were incubated in a Brinsea Mini Eco Egg incubator (Brinsea Products Limited, Weston Super Mare, UK) at 37.5 °C, with a “no turning” setup. Following four days of incubation, the eggs were candled to confirm the presence of a developing embryo and a small volume of egg (~3 mL) removed from the blunt end of

the egg using a sterile needle and syringe under aseptic conditions and discarded. A small rectangular window (~2.5 x 2.0 cm) was cut into the shell of the egg using a Dremel drill to expose the developing embryo before the compounds (10 μM 20 μL pre-absorbed on sterile 1 x 1 $cm^2$ Whatman filter paper), were loaded onto the sample. Compounds, **24**, **24a**, **24e**, **24j** and **121**, were prepared as described in Section 4.5.3 and diluted in sterile PBS. The window is sealed with low-adhesive Scotch tape and eggs incubated at 37.5 °C. On day ten after first incubation, the eggs are removed from the incubator and representative images of the embryo and associated blood vessel network captured using GXCAM-9 digital microscope C-mount camera (GT Vision, Suffolk, UK) mounted on an Olympus CKX41 microscope at x3 magnification (objective lens).

# CHAPTER 5: General Discussion and Conclusions

# 5. General Discussion

Pancreatic cancer remains one of the most difficult cancers to diagnose and treat with very poor five-year survival rates [4, 50, 91]. Despite advances made in the treatment of many other cancers that have led to improvement in survival rates, pancreatic cancer still remains "the silent killer", with little means of detecting the disease in its early stages.

The main aim of this project was to identify compounds with the potential to disrupt the interaction between S100P and its receptor, RAGE, in the hope of finding a potential therapeutic candidate against pancreatic cancer. S100P is a $Ca^{2+}$-binding protein that has been validated [125, 150, 175, 177, 181, 185] and shown to be a druggable therapeutic target in pancreatic cancer [195, 225, 228].

The NMR-derived 3D experimental structure of S100P (PDB ID 1OZO) was used for structure-based drug design (SBDD) studies (Chapter 2) in the investigation to identify potential inhibitors of the S100P-RAGE interaction. The selection of the NMR ensemble over the X-ray crystal structure (PDB ID 1J55) was due to the former representing a complete experimental structure of the active (i.e. dimeric) form of the protein. To the best of this author's knowledge, there is very limited data available on the application of the 3D-NMR ensemble of S100P for SBDD hence the rationale behind the present work. In addition, with the exception of the recently published papers of Penumutchu *et al*. [294, 323], there are very limited computational studies on the

experimental S100P structures to elucidate the mechanism by which small molecules interact with it.

Using *in silico* methods, conformer 15 of 1OZO was identified as the most suitable member of the NMR ensemble for use as a template for SBDD studies (Chapter 2). This member was selected after a number of pocket-detection algorithms independently identified similar potential pockets at the dimeric interface of the conformer that were sufficiently large enough to accommodate cromolyn (**4**), the ligand reported to bind to and inhibit the protein's interaction with its receptor RAGE [195].When this work was carried out there was no experimental information on the binding pockets of S100P, although it had been shown that interaction with target proteins is dependent on its dimerization [137, 284, 292, 322]. In 2014, the two reports by Penumutchu *et al*. showed the interaction of **4** with S100P as well as the protein's interaction with RAGE [294, 323]. The authors identified residues on the dimeric interface of S100P involving helix 4 and helix 1′ (E5, D13, F44, Y88 and F89) that play a crucial role in S100P binding to RAGE, and residues F44, F89, A84, I81, I12, G9, M8, E5 and C85 as those involved in the binding of cromolyn to S100P [294, 323].

These residues, with the exception of C85 which was mutated to S85 in the NMR ensemble, are present in the putative dimeric binding site of conformer 15 identified by the studies presented in this thesis (Chapter 2). In retrospect, this validates conformer 15 of the NMR ensemble as a credible starting point for SBDD studies given

that the putative binding site has now been shown to correlate with direct experimental evidence [294, 323].

Docking studies (Chapter 2) showed three main points of interaction between compound **4** and S100P: two hydrogen bonding interactions involving residues T82 and T92 with the two oxygen atoms of one of the carboxylate groups of compound **4**, and a hydrophobic contact between F89 and one of the chromone moieties. Two pharmacophore models, "stringent" and "relaxed" models, were generated from the predicted binding interactions. Both models have the same geometric constraints (angles and distances) but different pharmacophoric features (Chapter 2). Both have two features defined as hydrogen bond acceptors and an aromatic centre. In the "stringent" pharmacophore model, the two acceptor features were further classified as "anionic". In the "relaxed" model, they were defined as hydrogen bond acceptors (Chapter 2).

Both the "stringent" and "relaxed" pharmacophore models were used to virtually screen 653,214 lead-like compounds in the MOE database resulting in 52 and 4,619 "hits" respectively. The huge difference in the number of "hits" obtained from the two virtual screens demonstrates the rigidity that the requirement for an anionic feature in the "stringent" pharmacophore model conveys on the results returned. Retention of the anionic feature could have served to reduce the variety in the hit compounds returned from the virtual screening experiment, which in turn would hamper drug discovery efforts, and as such it was important to widen the scope of the virtual screen by defining the "relaxed" pharmacophore. Virtual screening of the ZINC database containing

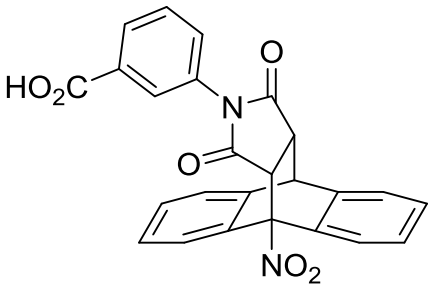
765,278 lead-like compounds with the "relaxed" pharmacophore generated 4,789 "hits". A total of 9,408 "hits" were obtained from the virtual screening of both databases using the "relaxed" pharmacophore (Chapter 2).

To obtain a diverse library of compounds for biological screening, the 9,408 "hits" from the virtual screening studies were clustered based on their maximum common substructure similarity [281, 282]. A total of 299 clusters were obtained with 77 singletons. Sixty-five "hits", 13 from the "stringent" pharmacophore search and 52 from the "relaxed" pharmacophore query, were selected from the clusters and purchased. Purchased compounds were selected from different clusters to ensure maximum diversity. Although commercially available from specialist vendors, these compounds are considered "novel" as their synthesis, purification, and characterisation is not usually available in the public domain. Four "hit" compounds, **7**, **58**, **58a** and **58b**, were synthesised and characterised in-house as described in Chapter 3.

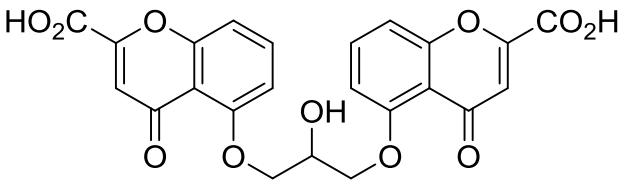
The first round of biological screening involving 17 compounds and carried out in collaboration with Dr Crnogorac-Jurcevic's laboratory at Queen Mary, University of London, identified five compounds, **17**, **18**, **20**, **24**, and **43**, that demonstrated promising activity against the invasion of the S100P-expressing pancreatic cancer cell line BxPC-3 (Chapter 3). Compound **24**, in particular, showed a significant inhibition of the invasion of these cells at a concentration of 100 μM that was comparable to compound **4** at the same concentration ($p<0.05$). As this was one of the compounds initially purchased, and since repurchasing it in sufficient quantity for further explorative studies was

prohibitively expensive, it was synthesised in-house via two synthetic routes (Chapter 4).

Structurally, compound **24** shares little similarity to compound **4** (Figure 5.1). The maximum common substructure Tanimoto similarity between the two compounds is 0.17 meaning they share a 17% similarity. The former is a 9-$NO_2$ anthracene-maleic anhydride-derived *N*-aryl substituted succinimide compound, whilst the latter is a symmetrical acid consisting of two chromone moieties joined together by an alcohol chain. Retrosynthetic analysis of compound **24** led to the identification of two potential synthetic routes **a** and **b** (Chapter 4, Section 4.2.1).




**24**




**4**

**Figure 5.1**: Compound **24**, which showed an inhibitory effect on the invasion of BxPC-3 cells comparable to compound **4**, is structurally different from the latter.

In route **a**, ethyl-3-aminobenzoate was combined with maleic anhydride in a two-step reaction process to give the maleimide intermediate **115**. The Diels-Alder reaction of the dienophile **115** with 9-nitroanthracene did not however yield the desired target compound **24**, possibly due to interaction of the *N*-aryl substituent of the dienophile with the π system of the 9-$NO_2$ anthracene diene [362], or, the effect of the electron withdrawing nitro group on the π system [361]. In route **b**, maleic anhydride was combined with 9-nitroanthracene in a Diels-Alder reaction to give the cycloadduct intermediate **119**. Reaction of **119** with 3-aminobenzoic acid led to the successful isolation of the target compound **24** in 80% yield from the Diels–Alders step (Chapter 4, Section 4.5.2.1.1). Thirty-nine analogues were also synthesised by varying substituents in the *meta* (R) or *para* (R′) positions on the phenyl ring, and with or without the nitro group (X) on the dihydroanthracene ring (Figure 5.2), in excellent reproducible yields ranging from 61-97%. All compounds were characterised via $^{1}$H- and $^{13}$C-NMR, LC-MS spectrometry, and IR spectroscopy to confirm that they are the desired products.

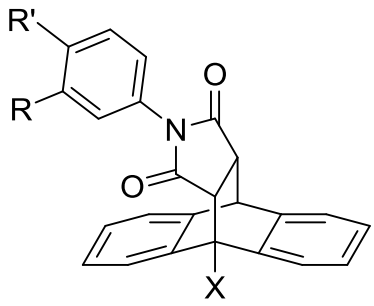



**Figure 5.2**: Analogues of compound **24** were synthesised by varying different substituents on R, R′ and X (Chapter 4).

Due to time limitations, five of the 40 synthesised compounds, **24**, **24a**, **24e**, **24j** and **121**, were biologically screened against pancreatic cancer cell lines BxPC-3 and Panc-1 (Chapter 4). These compounds were selected to assess the effect of substituting the carboxylic acid group with an ester group (**24a**), a hydrogen (**24e**), and in the *para* position (**24j**) on biological activity relative to the main "hit" compound **24**. Compound **121**, which lacks the nitro group, was included to assess the effect of this group on biological activity relative to **24**. The compounds were assessed for cytotoxicity and for effect on the proliferative capability of BxPC-3 and Panc-1 cells using the LDH and MTS assays respectively.

All five compounds demonstrated significant inhibitory activity against the proliferation of S100P-expressing BxPC-3 cells relative to control ($p<0.0001$) at a concentration of 1, 10 and 100 μM. Interestingly, a similar effect was also observed on Panc-1 cells, casting doubt on an S100P-specific effect from these compounds based on these results. The effect of compound **4** on the proliferation of both cell lines, despite published reports [195, 334], was not reproduced in this study, nor in numerous similar studies in Dr Crnogorac-Jurcevic's laboratory (unpublished results). The inability to reproduce the reported effect of compound **4** on the proliferation of these cells points to the conclusion that S100P-RAGE may not be involved in cell proliferation, merely metastasis.

Preliminary studies on the effect of the five compounds investigated on angiogenesis using chick chorioallantioc membrane (CAM) assay, showed initial

promising results for compound **24**. However, additional studies are required to validate these initial findings.

# 6. Conclusions

In this project, computational methods were successfully employed to identify putative sites on the NMR ensemble of S100P with potential to bind small molecules. This was the first project of its kind to do so on the NMR-derived structure of the protein, and has shown the feasibility of using an NMR ensemble for structure-based drug design studies provided due care is taken to select the most appropriate of the ensemble members for use as a template. Conformer 15 of the NMR ensemble was identified as having the most appropriate "druggable" binding site for SBDD studies, a hypothesis subsequently confirmed by recent experimental studies by others [294, 323], and was used as template to predict S100P's interaction with compound **4** via *in silico* docking studies. The predicted binding interactions between the ligand and protein were used to design 3D pharmacophore models that were used to perform virtual screens on 1,418,492 lead-like compounds, and returned molecules fulfilling the pharmacophoric constraints with a hit rate of 0.7%. All "hits" identified are novel in that their synthesis, characterisation and/or potential biological activity against pancreatic cancer cells have not been previously reported.

A number of "hits" were successfully synthesised, isolated and characterised with excellent yields. The novelty of these compounds means that this is the first time their syntheses have been shown successfully, albeit by following established synthetic routes. Compound **24**, one of the "hits" that demonstrated promising biological activity, was successfully synthesised and characterised having been initially purchased during

the first round of biological screening. Thirty-nine analogues of this compound were also successfully synthesised to form a library for future SAR and QSAR studies.

Biological screening to validate 17 of the "hit" compounds identified from the virtual screening studies show five compounds with the capability of reducing pancreatic cancer cell metastasis *in vitro* to the same extent as the marker substrate. This represents a 29% success rate, which can be higher considering that the 52 purchased "hits" and the 35 synthesised analogues of compound **24** are yet to be tested, and shows the significant level of enrichment in identifying potential molecules of interest that is afforded by the rational approach to drug design employed in this thesis when compared with random screening of libraries of compounds [271, 406, 407]. This is both exciting and encouraging as it demonstrates the relative success of the project at this stage of the drug discovery process, where attrition can be very high [408], and the potential and opportunity to develop a number of chemically distinct hit-compounds into lead molecules. With lead optimisation, these compounds, could provide a catalyst for identifying that elusive therapeutic agent for the treatment of pancreatic cancer which is desperately needed.

In summary, this thesis has successfully demonstrated the close interaction between the fields of computational science, synthetic chemistry and biological technology to address a growing demand within drug discovery. In particular, this work has documented potential avenues of further research and confirmed possible novel therapeutic candidates to tackle pancreatic cancer, one of the deadliest cancers in the Western world.

# 7. Future work

The journey of a drug from early discovery to when it enters the market could take up to 10 years or more [409]. This project is in the early discovery stage and has identified "hits" that show promising *in vitro* activity against pancreatic cancer cell lines. These initial results provide tangible starting points for a number of follow-on researchers to develop a clinical candidate. Some of the steps that will be required in order to achieve this goal are outlined below.

First, anti-proliferative activity of the compounds on pancreatic cancer cells should be investigated at nanomolar concentrations to establish the minimal effective concentration of these compounds. The membrane integrity assay used here to assess the cytotoxicity of the compounds could be optimised by investigating the effect that reduced foetal bovine serum will have on the assay and cells. Failing this, alternative assays, such as Promega's MultiTox-Glo Multiplex Cytotoxicity Assay (Promega, Southampton, UK), which simultaneously measures both viability and cytotoxicity, could be explored.

Once the potency of the compounds against pancreatic cancer cells is established, they should be tested for toxicity on normal healthy cells. This is important as it will allow the assessment of the structures of the compounds, identify potential toxophores and alteration of the molecules so that they retain their activity but reduce toxicity if the latter is shown to be a problem. It will also be interesting to assess the

effect of the compounds against other cancer cell lines such as lung, breast, colon and prostate, in which S100P has been shown to be highly expressed.

In addition to toxicity studies, mutagenesis studies will provide insight into the mechanism by which these compounds are exerting their effect. If, when the compounds are tested at nanomolar concentrations and they show an S100P-effect, mutating certain genes on the protein could offer an explanation into this mechanism. NMR spectroscopy or X-ray crystallography studies of the protein in complex with the compounds could be another avenue to provide an understanding of this mechanism in lieu of mutagenesis studies. The compounds can also be tested at the genetic level using microarray analysis to assess their effect on signalling pathways that are activated via S100 binding to RAGE.

*De novo* synthesis could be employed to design new leads using existing compounds that show potency as templates or, generate novel fragments into the identified putative sites on S100P that have the potential to interact with key residues within the pocket. *De novo* design and synthesis are complementary to virtual screening studies. Both methods are excellent means of generating lead candidates although with *de novo* design, synthetic feasibility should be considered due to the novelty of the designed compound(s).

Upon completion of the biological screening of the 52 purchased "hits" (compounds **59**-**110**) identified from the virtual screening (Chapter 2), analogues of compounds showing promising activity should be synthesised to build a library for

SAR/QSAR studies. In addition, analogues of four of the "hits" that show anti-metastatic properties (Chapter 3) could be synthesised, similar to the synthesis of analogues of compound **24**, to build libraries for SAR studies.

If the chick chorioallantioc membrane (CAM) assay is going to be employed to further investigate the possible mechanism by which these compounds are acting, it would be useful to look at adapting the method of Dohle *et al*. [410]. This group did the assay *ex ovo*, which gave them easy access to the developing embryo and its rich blood network. This approach will also limit the rate of dead embryos at day 10 (as experienced in this work (Chapter 4).

# References


1. WHO. *Cancer Fact sheet No. 297*. 2014 February 2015 [cited 2014 18th June ]; Available from: http://goo.gl/3lQNWA.

2. Cancer, I.A.f.R.o., *World cancer report 2014*. 2014: Geneva, Switzerland

3. Ferlay, J., Steliarova-Foucher, E., Lortet-Tieulent, J., *et al.*, *Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012.* European Journal of Cancer, 2013. **49**(6): p. 1374-1403.

4. Siegel, R., Ma, J., Zou, Z., *et al.*, *Cancer statistics, 2014.* CA: A Cancer Journal for Clinicians, 2014. **64**(1): p. 9-29.

5. Yabroff, K.R., Lund, J., Kepka, D., *et al.*, *Economic burden of cancer in the United States: estimates, projections, and future research.* Cancer Epidemiology Biomarkers & Prevention, 2011. **20**(10): p. 2006-2014.

6. Luengo-Fernandez, R., Leal, J., Gray, A., *et al.*, *Economic burden of cancer across the European Union: a population-based cost analysis.* The Lancet Oncology, 2013. **14**(12): p. 1165-1174.

7. Sankaranarayanan, R., Ramadas, K., and Qiao, Y.-l., *Managing the changing burden of cancer in Asia.* BMC Medicine, 2014. **12**(1): p. 3-20.

8. Kang, S., Koh, E.S., Vinod, S.K., *et al.*, *Cost analysis of lung cancer management in South Western Sydney.* Journal of Medical Imaging and Radiation Oncology, 2012. **56**(2): p. 235-241.

9. Mariotto, A.B., Yabroff, K.R., Shao, Y., *et al.*, *Projections of the cost of cancer care in the United States: 2010–2020.* Journal of the National Cancer Institute, 2011.

10. England, UK. *2010 to 2015 Government Policy: cancer research and treatment*. 2013. [Cited 8th October, 2015]. Available from: https://goo.gl/kUMPHz.

11. Hanly, P., Soerjomataram, I., and Sharp, L., *Measuring the societal burden of cancer: The cost of lost productivity due to premature cancer-related mortality in Europe.* International Journal of Cancer, 2015. **136**(4): p. E136-E145.

12. Tingstedt, B., Andersson, E., Flink, A., *et al.*, *Pancreatic cancer, healthcare cost, and loss of productivity: a register-based approach.* World Journal of Surgery, 2011. **35**(10): p. 2298-2305.

13. Hanly, P.A. and Sharp, L., *The cost of lost productivity due to premature cancer-related mortality: an economic measure of the cancer burden.* BMC Cancer, 2014. **14**(1): p. 224-233.

14. Feng, Z., Hu, W., Rajagopal, G., *et al., The tumor suppressor p53: cancer and aging.* Cell Cycle, 2008. **7**(7): p. 842-847.

15. Greenblatt, M., Bennett, W., Hollstein, M., *et al., Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis.* Cancer Research, 1994. **54**(18): p. 4855-4878.

16. Hwang, S.-J., Cheng, L.S.-C., Lozano, G., *et al., Lung cancer risk in germline p53 mutation carriers: association between an inherited cancer predisposition, cigarette smoking, and cancer risk.* Human Genetics, 2003. **113**(3): p. 238-243.

17. Fan, R., Wu, M.-T., Miller, D., *et al., The p53 codon 72 polymorphism and lung cancer risk.* Cancer Epidemiology Biomarkers & Prevention, 2000. **9**(10): p. 1037-1042.

18. Ford, D., Easton, D.F., Bishop, D.T., *et al., Risks of cancer in BRCA1-mutation carriers.* The Lancet, 1994. **343**(8899): p. 692-695.

19. Ford, D., Easton, D., Stratton, M., *et al., Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families.* The American Journal of Human Genetics, 1998. **62**(3): p. 676-689.

20. Thompson, D., Easton, D.F., and Consortium, B.C.L., *Cancer incidence in BRCA1 mutation carriers.* Journal of the National Cancer Institute, 2002. **94**(18): p. 1358-1365.

21. Slamon, D.J., Godolphin, W., Jones, L.A., *et al., Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer.* Science, 1989. **244**(4905): p. 707-712.

22. Amosenko, F., Brzhezovskiy, V.Z., Lyubchenko, L., *et al., Analysis of mutations of the RET proto-oncogene in patients with medullary thyroid carcinoma.* Russian Journal of Genetics, 2003. **39**(6): p. 706-711.

23. Ueki, T., Fujimoto, J., Suzuki, T., *et al., Expression of hepatocyte growth factor and its receptor c-met proto-oncogene in hepatocellular carcinoma.* Hepatology, 1997. **25**(4): p. 862-866.

24. Barras, D., *BRAF Mutation in Colorectal Cancer: An Update.* Biomarkers in Cancer, 2015. **7**(Suppl 1): p. 9-12.

25. Lichtenstein, P., Holm, N.V., Verkasalo, P.K., *et al., Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland.* New England Journal of Medicine, 2000. **343**(2): p. 78-85.

26. Richiardi, L., Pettersson, A., and Akre, O., *Genetic and environmental risk factors for testicular cancer.* International Journal of Andrology, 2007. **30**(4): p. 230-241.

27. Hecht, S.S., *Tobacco smoke carcinogens and lung cancer.* Journal of the National Cancer Institute, 1999. **91**(14): p. 1194-1210.

28. Pleasance, E.D., Stephens, P.J., O'Meara, S., *et al., A small-cell lung cancer genome with complex signatures of tobacco exposure.* Nature, 2010. **463**(7278): p. 184-190.

29. CRUK *The National Cancer Research Institute (NCRI) Cancer research in the UK 2002-2011: An overview of the research funded by NCRI Partners*. 2013 [Cited on 12th April, 2015]; Available from: http://goo.gl/baehBS.

30. NIH. *Estimates of funding for various research, condition, and disease categories (RCDC).* 2015 [Cited 9th October 2015]; Available from: http://report.nih.gov/categorical_spending.aspx.

31. Goodman, L.S., Wintrobe, M.M., Dameshek, W., *et al., Nitrogen mustard therapy: Use of methyl-bis (beta-chloroethyl) amine hydrochloride and tris (beta-chloroethyl) amine hydrochloride for Hodgkin's disease, lymphosarcoma, leukemia and certain allied and miscellaneous disorders.* Journal of the American Medical Association, 1946. **132**(3): p. 126-132.

32. van Zutphen, S. and Reedijk, J., *Targeting platinum anti-tumour drugs: Overview of strategies employed to reduce systemic toxicity.* Coordination Chemistry Reviews, 2005. **249**(24): p. 2845-2853.

33. Winocur, G., Vardy, J., Binns, M.A., *et al., The effects of the anti-cancer drugs, methotrexate and 5-fluorouracil, on cognitive function in mice.* Pharmacology Biochemistry and Behavior, 2006. **85**(1): p. 66-75.

34. Sitzia, J. and Huggins, L., *Side effects of cyclophosphamide, methotrexate, and 5-fluorouracil (CMF) chemotherapy for breast cancer.* Cancer Practice, 1998. **6**(1): p. 13-21.

35. Botchkarev, V.A. *Molecular mechanisms of chemotherapy-induced hair loss*. in *Journal of Investigative Dermatology Symposium Proceedings*. Nature Publishing Group. 2003. **8**(1): p. 72-75.

36. Peto, J., Gilham, C., Fletcher, O., *et al., The cervical cancer epidemic that screening has prevented in the UK.* The Lancet, 2004. **364**(9430): p. 249-256.

37. Youlden, D.R., Cramb, S.M., Dunn, N.A., *et al., The descriptive epidemiology of female breast cancer: an international comparison of screening, incidence, survival and mortality.* Cancer Epidemiology, 2012. **36**(3): p. 237-248.

38. Team, N.L.S.T.R., *Reduced lung-cancer mortality with low-dose computed tomographic screening.* The New England Journal of Medicine, 2011. **365**(5): p. 395-409.

39. Gøtzsche, P.C. and Nielsen, M., *Screening for breast cancer with mammography.* Cochrane Database Systematic Reviews, 2011. **1**(6): p. 1-84.

40. Afshar-Oromieh, A., Malcher, A., Eder, M., *et al., PET imaging with a [68Ga] gallium-labelled PSMA ligand for the diagnosis of prostate cancer: biodistribution in humans and first evaluation of tumour lesions.* European Journal of Nuclear Medicine and Molecular Imaging, 2013. **40**(4): p. 486-495.

41. Verma, S., Turkbey, B., Muradyan, N., *et al., Overview of dynamic contrast-enhanced MRI in prostate cancer diagnosis and management.* American Journal of Roentgenology, 2012. **198**(6): p. 1277-1288.

42. Lortet-Tieulent, J., Renteria, E., Sharp, L., *et al., Convergence of decreasing male and increasing female incidence rates in major tobacco-related cancers in Europe in 1988–2010.* European Journal of Cancer, 2015. **51**(9): p. 1144-1163.

43. Janssen-Heijnen, M.L. and Coebergh, J.-W.W., *The changing epidemiology of lung cancer in Europe.* Lung Cancer, 2003. **41**(3): p. 245-258.

44. Steliarova-Foucher, E., O'Callaghan, M., Ferlay, J., *et al., The European Cancer Observatory: a New Data Resource.* European Journal of Cancer, 2015. **51**(9): p. 1131-1143.

45. Malvezzi, M., Bertuccio, P., Levi, F., *et al., European cancer mortality predictions for the year 2014.* Annals of Oncology, 2014. p. 1-7.

46. Torre, L.A., Bray, F., Siegel, R.L., *et al., Global cancer statistics, 2012.* CA: A Cancer Journal for Clinicians, 2015. **65**(2): p. 87-108.

47. Peto, R., Boreham, J., Clarke, M., *et al., UK and USA breast cancer deaths down 25% in year 2000 at ages 20-69 years.* Lancet, 2000. **355**(9217): p. 1822.

48. O'Connor, S.J., *Review of the incidence, prevalence, mortality and causative factors for lung cancer in Europe.* European Journal of Cancer, 2011. **47**: p. S346-S347.

49. Petersen, G.M. and Boffetta, P., *Carcinogenesis of pancreatic cancer: challenges, collaborations, progress.* Molecular Carcinogenesis, 2012. **51**(1): p. 1-2.

50. Ma, J., Siegel, R., and Jemal, A., *Pancreatic cancer death rates by race among US men and women, 1970–2009.* Journal of the National Cancer Institute, 2013. **105**(22): p. 1694-1700.

51. Cardin, D.B. and Berlin, J.D., *Pancreas cancer on the rise: are we up to the challenge?* Journal of the National Cancer Institute, 2013. **105**(22): p. 1675-1676.

52. Rogers, K., *The digestive system*. 2010, New York, USA: The Rosen Publishing Group.

53. Lerch, M.M., Albrecht, E., Ruthenbürger, M., *et al.*, *Pathophysiology of alcohol-induced pancreatitis.* Pancreas, 2003. **27**(4): p. 291-296.

54. Mudan, S., *Diseases of the pancreas.* Trends in Urology & Men's Health, 2013. **4**(4): p. 35-39.

55. Slack, J., *Developmental biology of the pancreas.* Development, 1995. **121**(6): p. 1569-1580.

56. Helmberger, T.K. and Manfredi, R., *Diseases of the Pancreas*, in *Diseases of the Abdomen and Pelvis 2014–2017*. 2014, Springer. p. 127-133.

57. Sarner, M. and Cotton, P., *Classification of pancreatitis.* Gut, 1984. **25**(7): p. 756-759.

58. Paxton, J.R. and Payne, J.H., *Symptoms of acute pancreatitis.* California Medicine, 1947. **67**(4): p. 241-243.

59. Frossard, J.-L., Steer, M.L., and Pastor, C.M., *Acute pancreatitis.* The Lancet, 2008. **371**(9607): p. 143-152.

60. Toh, S., Phillips, S., and Johnson, C., *A prospective audit against national standards of the presentation and management of acute pancreatitis in the South of England.* Gut, 2000. **46**(2): p. 239-243.

61. Etemad, B. and Whitcomb, D.C., *Chronic pancreatitis: diagnosis, classification, and new genetic developments.* Gastroenterology, 2001. **120**(3): p. 682-707.

62. Mitchell, C., *Chronic Pancreatitis.* Medicine, 2003. **31**(3): p. 122-125.

63. Braganza, J., Hunt, L., and Warwick, F., *Relationship between pancreatic exocrine function and ductal morphology in chronic pancreatitis.* Gastroenterology, 1982. **82**(6): p. 1341-1347.

64. Ammann, R.W., Heitz, P.U., and Kloppel, G., *Course of alcoholic chronic pancreatitis: a prospective clinicomorphological long-term study.* Gastroenterology, 1996. **111**(1): p. 224-231.

65. Ammann, R.W., Muellhaupt, B., and Group, Z.P.S., *The natural history of pain in alcoholic chronic pancreatitis.* Gastroenterology, 1999. **116**(5): p. 1132-1140.

66. Key, C., *Cancer of the Pancreas.* SEER survival monograph: cancer survival among adults: US SEER program, 1988. **2001**.

67. Li, D., Xie, K., Wolff, R., *et al.*, *Pancreatic cancer.* The Lancet, 2004. **363**(9414): p. 1049-1057.

68. Kloppel, G., Hruban, R.H., Longnecker, D.S., *et al.*, in *Pathology and genetics of tumours of the digestive system*, Hamilton, S.R., *et al.*, Editors. 2000, IARC press International Agency for Research on Cancer (IARC) 69372 Lyon, France.

69. Hruban, R.H., Goggins, M., Parsons, J., *et al.*, *Progression model for pancreatic cancer.* Clinical Cancer Research, 2000. **6**(8): p. 2969-2972.

70. Hruban, R.H., Maitra, A., Kern, S.E., *et al.*, *Precursors to pancreatic cancer.* Gastroenterology Clinics of North America, 2007. **36**(4): p. 831-849.

71. Hruban, R.H., Takaori, K., Canto, M., *et al.*, *Clinical importance of precursor lesions in the pancreas.* Journal of Hepato-Biliary-Pancreatic Surgery, 2007. **14**(3): p. 255-263.

72. Tanaka, M., Chari, S., Adsay, V., *et al.*, *International consensus guidelines for management of intraductal papillary mucinous neoplasms and mucinous cystic neoplasms of the pancreas.* Pancreatology, 2006. **6**(1-2): p. 17-32.

73. Hruban, R.H., Takaori, K., Klimstra, D.S., *et al.*, *An illustrated consensus on the classification of pancreatic intraepithelial neoplasia and intraductal papillary mucinous neoplasms.* The American Journal of Surgical Pathology, 2004. **28**(8): p. 977-987.

74. Vincent, A., Herman, J., Schulick, R., *et al.*, *Pancreatic cancer.* The Lancet, 2011. **378**(9791): p. 607-620.

75. Hruban, R.H., Takaori, K., Klimstra, D.S., *et al.*, *An illustrated consensus on the classification of pancreatic intraepithelial neoplasia and intraductal papillary*

*mucinous neoplasms.* The American Journal of Surgical Pathology, 2004. **28**(8): p. 977-987.

76. Chang, D.K., Merrett, N.D., and Biankin, A.V., *Improving outcomes for operable pancreatic cancer: is access to safer surgery the problem?* Journal of Gastroenterology and Hepatology, 2008. **23**(7pt1): p. 1036-1045.

77. McDermott, M.T., *Endocrine secrets*. 2009, Philadelphia: Elsevier Mosby.

78. Kaltsas, G.A., Besser, G.M., and Grossman, A.B., *The diagnosis and medical management of advanced neuroendocrine tumors.* Endocrine Reviews, 2004. **25**(3): p. 458-511.

79. Nieto, J., Grossbard, M.L., and Kozuch, P., *Metastatic pancreatic cancer 2008: is the glass less empty?* Oncologist, 2008. **13**(5): p. 562-576.

80. Ferlay, J., Autier, P., Boniol, M., *et al.*, *Estimates of the cancer incidence and mortality in Europe in 2006.* Annals of Oncology, 2007. **18**(3): p. 581-592.

81. Howlader, N., Noone, A., Krapcho, M., *et al.*, *SEER Cancer Statistics Review, 1975-2009 (Vintage 2009 Populations)*, National Cancer Institute. Bethesda, MD. 2012, MD.

82. Office of National Statistics (ONS). *Pancreatic cancer incidence, mortality and survival, England, 1971-2011*. Cancer Statistics Registrations, England (Series MB1) , No. 42, 2011 2013 Available from: http://goo.gl/baehBS.

83. Yachida, S., Jones, S., Bozic, I., *et al.*, *Distant metastasis occurs late during the genetic evolution of pancreatic cancer.* Nature, 2010. **467**(7319): p. 1114-1117.

84. Raimondi, S., Lowenfels, A.B., Morselli-Labate, A.M., *et al.*, *Pancreatic cancer in chronic pancreatitis; aetiology, incidence, and early detection.* Best Practice & Research Clinical Gastroenterology, 2010. **24**(3): p. 349-358.

85. Lowenfels, A.B., Maisonneuve, P., Cavallini, G., *et al.*, *Pancreatitis and the risk of pancreatic cancer.* New England Journal of Medicine, 1993. **328**(20): p. 1433-1437.

86. National Health Service (NHS). *Diagnosing pancreatic cancer* 2012 10 July, 2014 [cited 28th April 2014]; Available from: http://www.nhs.uk/Conditions/Cancer-of-the-pancreas/Pages/Diagnosis.aspx.

87. Siddiqui, A.A., Brown, L.J., Hong, S.-K.S., *et al.*, *Relationship of pancreatic mass size and diagnostic yield of endoscopic ultrasound-guided fine needle aspiration.* Digestive Diseases and Sciences, 2011. **56**(11): p. 3370-3375.

88. Itoi, T., Tsuchiya, T., Itokawa, F., *et al., Histological diagnosis by eus-guided fine-needle aspiration biopsy in pancreatic solid masses without on-site cytopathologist: a single-center experience.* Digestive Endoscopy, 2011. **23**(s1): p. 34-38.

89. Chang, K.J., Nguyen, P., Erickson, R.A., *et al., The clinical utility of endoscopic ultrasound–guided fine-needle aspiration in the diagnosisand staging of pancreatic carcinoma.* Gastrointestinal Endoscopy, 1997. **45**(5): p. 387-393.

90. Evans, D.B., Abbruzzese, J.L., and Rich, T.R., *Cancer of the pancreas,* in *Cancer principles and practice of oncology*, DeVita, V.T., Hellman, S., and Rosenberg, S.A., Editors. 1997, Lippincott Raven: Philadelphia. p. 1054-1087.

91. Kalser, M.H., Barkin, J., and Macintyre, J.M., *Pancreatic cancer. Assessment of prognosis by clinical presentation.* Cancer, 1985. **56**(2): p. 397-402.

92. Logsdon, C.D., Simeone, D.M., Binkley, C., *et al., Molecular profiling of pancreatic adenocarcinoma and chronic pancreatitis identifies multiple genes differentially regulated in pancreatic cancer.* Cancer Research, 2003. **63**(10): p. 2649.

93. Bloomston, M., Frankel, W.L., Petrocca, F., *et al., MicroRNA expression patterns to differentiate pancreatic adenocarcinoma from normal pancreas and chronic pancreatitis.* JAMA: the Journal of the American Medical Association, 2007. **297**(17): p. 1901-1908.

94. Rosty, C., Geradts, J., Sato, N., *et al., p16 Inactivation in pancreatic intraepithelial neoplasias (PanINs) arising in patients with chronic pancreatitis.* The American Journal of Surgical Pathology, 2003. **27**(12): p. 1495.

95. Evans, D.B., Catalano, P., Charnsangavej, C., *et al., AJCC cancer staging manual*. 6th Ed, Greene, F.L., *et al.* Editors. 2002. **1**: Springer.

96. National Institute for Cliniacal Execellence (NICE). *Guidance on the use of gemcitabine for the treatment of pancreatic cancer*. 2001 [Cited 29th April 2015; NICE technology appraisal guidance 25]. Available from: http://goo.gl/UsuGgR.

97. Iqbal, N., Lovegrove, R., Tilney, H., *et al., A comparison of pancreaticoduodenectomy with pylorus preserving pancreaticoduodenectomy: a meta-analysis of 2822 patients.* European Journal of Surgical Oncology (EJSO), 2008. **34**(11): p. 1237-1245.

98. Iqbal, N., Lovegrove, R., Tilney, H., *et al., A comparison of pancreaticoduodenectomy with extended pancreaticoduodenectomy: a meta-analysis of 1909 patients.* European Journal of Surgical Oncology (EJSO), 2009. **35**(1): p. 79-86.

99. Crist, D.W., Sitzmann, J.V., and Cameron, J.L., *Improved hospital morbidity, mortality, and survival after the Whipple procedure.* Annals of Surgery, 1987. **206**(3): p. 358-365.

100. Pedrazzoli, S., Beger, H.G., nther, u., *et al.*, *A surgical and pathological based classification of resective treatment of pancreatic cancer.* Digestive Surgery, 1999. **16**(4): p. 337-345.

101. Zerbi, A., Fossati, V., Parolini, D., *et al.*, *Intraoperative radiation therapy adjuvant to resection in the treatment of pancreatic cancer.* Cancer, 1994. **73**(12): p. 2930-2935.

102. National Institute for Clinical Excellence (NICE). NHS *Technology Appraisal No. 25*. 2001 [Cited 26th April 2015]. Available from: http://guidance.nice.org.uk/TA25

103. Oettle, H., Post, S., Neuhaus, P., *et al.*, *Adjuvant chemotherapy with gemcitabine vs observation in patients undergoing curative-intent resection of pancreatic cancer: a randomized controlled trial.* JAMA, 2007. **297**(3): p. 267-277.

104. Burris, H.A., Moore, M.J., Andersen, J., *et al.*, *Improvements in survival and clinical benefit with gemcitabine as first-line therapy for patients with advanced pancreas cancer: a randomized trial.* Journal of Clinical Oncology, 1997. **15**(6): p. 2403-2413.

105. Von Hoff, D.D., Ramanathan, R.K., Borad, M.J., *et al.*, *Gemcitabine plus nab-paclitaxel is an active regimen in patients with advanced pancreatic cancer: a phase I/II trial.* Journal of Clinical Oncology, 2011. **29**(34): p. 4548-4554.

106. Department of Health. England, N. *National Cancer Drugs Fund List ver5.1*. 2015 [Cited 15th October 2015]. Available from: http://www.england.nhs.uk/wp-content/uploads/2015/09/ncdf-list-sept15.pdf

107. Bachet, J.-B., Chibaudel, B., Bonnetain, F., *et al.*, *A randomized phase II study of weekly nab-paclitaxel plus gemcitabine or simplified LV5FU2 as first-line therapy in patients with metastatic pancreatic cancer: the AFUGEM GERCOR trial.* BMC Cancer, 2015. **15**(1): p. 653-663.

108. Heinemann, V., Boeck, S., Hinke, A., *et al.*, *Meta-analysis of randomized trials: evaluation of benefit from gemcitabine-based combination chemotherapy applied in advanced pancreatic cancer.* BMC Cancer, 2008. **8**(1): p. 82-92.

109. Louvet, C., André, T., Lledo, G., *et al.*, *Gemcitabine combined with oxaliplatin in advanced pancreatic adenocarcinoma: final results of a GERCOR multicenter phase II study.* Journal of Clinical Oncology, 2002. **20**(6): p. 1512-1518.

110. Regine, W., Winter, K., Abrams, R., *et al., RTOG 9704 a phase III study of adjuvant pre and post chemoradiation (CRT) 5-FU vs. gemcitabine (G) for resected pancreatic adenocarcinoma.* Journal of Clinical Oncology, 2006. **24**(18 suppl): p. 4007.

111. Von Hoff, D.D., Ervin, T., Arena, F.P., *et al., Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine.* New England Journal of Medicine, 2013. **369**(18): p. 1691-1703.

112. Conlon, K.C., Klimstra, D.S., and Brennan, M.F., *Long-term survival after curative resection for pancreatic ductal adenocarcinoma. Clinicopathologic analysis of 5-year survivors.* Annals of Surgery, 1996. **223**(3): p. 273-279.

113. Hezel, A.F., Kimmelman, A.C., Stanger, B.Z., *et al., Genetics and biology of pancreatic ductal adenocarcinoma.* Genes & Development, 2006. **20**(10): p. 1218-1249.

114. Crnogorac-Jurcevic, T., Missiaglia, E., Blaveri, E., *et al., Molecular alterations in pancreatic carcinoma: expression profiling shows that dysregulated expression of S100 genes is highly prevalent.* The Journal of Pathology, 2003. **201**(1): p. 63-74.

115. Crnogorac–Jurcevic, T., Gangeswaran, R., Bhakta, V., *et al., Proteomic analysis of chronic pancreatitis and pancreatic adenocarcinoma.* Gastroenterology, 2005. **129**(5): p. 1454-1463.

116. Jiang, G., Cao, F., Ren, G., *et al., PRSS3 promotes tumour growth and metastasis of human pancreatic cancer.* Gut, 2010. **59**(11): p. 1535-1544.

117. Campagna, D., Cope, L., Lakkur, S.S., *et al., Gene expression profiles associated with advanced pancreatic cancer.* International Journal of Clinical and Experimental Pathology, 2008. **1**(1): p. 32-43.

118. Dowen, S.E., Crnogorac-Jurcevic, T., Gangeswaran, R., *et al., Expression of S100P and its novel binding partner S100PBPR in early pancreatic cancer.* The American Journal of Pathology, 2005. **166**(1): p. 81-92.

119. Costello, E., *Molecular Biological Understanding of Development of Pancreatic Cancer.* The Pancreas, 2008. p. 583-590.

120. Perkins, G.L., Slater, E.D., Sanders, G.K., *et al., Serum tumor markers.* American Family Physician, 2003. **68**(6): p. 1075-1088.

121. Ohuchida, K., Mizumoto, K., Egami, T., *et al., S100P is an early developmental marker of pancreatic carcinogenesis.* Clinical Cancer Research, 2006. **12**(18): p. 5411-5416.

122. Goonetilleke, K.S. and Siriwardena, A.K., *Systematic review of carbohydrate antigen (CA 19-9) as a biochemical marker in the diagnosis of pancreatic cancer.* European Journal of Surgical Oncology, 2007. **33**(3): p. 266-270.

123. Locker, G.Y., Hamilton, S., Harris, J., *et al., ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer.* Journal of Clinical Oncology, 2006. **24**(33): p. 5313-5327.

124. Radon, T.P., Massat, N.J., Jones, R., *et al., Identification of a Three-Biomarker Panel in Urine for Early Detection of Pancreatic Adenocarcinoma.* Clinical Cancer Research, 2015. **21**(15): p. 3512-3521.

125. Deng, H., Shi, J., Wilkerson, M., *et al., Usefulness of S100P in diagnosis of adenocarcinoma of pancreas on fine-needle aspiration biopsy specimens.* American Journal of Clinical Pathology, 2008. **129**(1): p. 81-88.

126. Nakata, K., Nagai, E., Ohuchida, K., *et al., S100P is a novel marker to identify intraductal papillary mucinous neoplasms.* Human Pathology, 2010. **41**(6): p. 824-831.

127. Moore, B.W., *A soluble protein characteristic of the nervous system.* Biochemical and Biophysical Research Communications, 1965. **19**(6): p. 739.

128. Schafer, B.W. and Heizmann, C.W., *The S100 family of EF-hand calcium-binding proteins: functions and pathology.* Trends in Biochemical Sciences, 1996. **21**(4): p. 134-140.

129. Zimmer, D.B., Cornwall, I.H., Landar, A., *et al., The S100 protein family: History, function, and expression.* Brain Research Bulletin, 1995. **37**(4): p. 417-429.

130. Hermann, A., Donato, R., Weiger, T.M., *et al., S100 calcium binding proteins and ion channels.* Frontiers in Pharmacology, 2012. **3**(67).

131. Donato, R., *Functional roles of S100 proteins, calcium-binding proteins of the EF-hand type.* Biochimica *et* Biophysica Acta (BBA)-Molecular Cell Research, 1999. **1450**(3): p. 191-231.

132. Schaub, M.C. and Heizmann, C.W., *Calcium, troponin, calmodulin, S100 proteins: From myocardial basics to new therapeutic strategies.* Biochemical and Biophysical Research Communications, 2008. **369**(1): p. 247-264.

133. Kretsinger, R.H. and Nockolds, C.E., *Carp muscle calcium-binding protein.* Journal of Biological Chemistry, 1973. **248**(9): p. 3313.

134. Heizmann, C., Ackermann, G., and Galichet, A., *Pathologies involving the S100 proteins and RAGE.* Calcium Signalling and Disease, 2008. p. 93-138.

135. Gerke, V. and Weber, K., *The regulatory chain in the p36-kd substrate complex of viral tyrosine-specific protein kinases is related in sequence to the S-100 protein of glial cells.* The EMBO Journal, 1985. **4**(11): p. 2917-2920.

136. Wang, G., Zhang, S., Fernig, D.G., *et al.*, *Heterodimeric interaction and interfaces of S100A1 and S100P.* Biochemical Journal, 2004. **382**(Pt 1): p. 375-383.

137. Koltzscher, M. and Gerke, V., *Identification of hydrophobic amino acid residues involved in the formation of S100P homodimers in vivo.* Biochemistry, 2000. **39**(31): p. 9533-9539.

138. Gribenko, A., Guzman-Casado, M., Lopez, M.M., *et al.*, *Conformational and thermodynamic properties of peptide binding to the human S100P protein.* Protein Science, 2002. **11**(6): p. 1367-1375.

139. Santamaria-Kisiel, L., Rintala-Dempsey, A.C., and Shaw, G.S., *Calcium-dependent and -independent interactions of the S100 protein family.* Biochemical Journal, 2006. **396**(2): p. 201-214.

140. Zimmer, D.B., Wright Sadosky, P., and Weber, D.J., *Molecular mechanisms of S100-target protein interactions.* Microscopy Research and Technique, 2003. **60**(6): p. 552-559.

141. Gribenko, A. and Makhatadze, G.I., *Oligomerization and divalent ion binding properties of the S100P protein: a $Ca^{2+}$/$Mg^{2+}$-switch model1.* Journal of Molecular Biology, 1998. **283**(3): p. 679-694.

142. Heil, A., Nazmi, A.R., Koltzscher, M., *et al.*, *S100P is a novel interaction partner and regulator of IQGAP1.* Journal of Biological Chemistry, 2011. **286**(9): p. 7227-7238.

143. Mbele, G.O., Deloulme, J.C., Gentil, B.J., *et al.*, *The zinc- and calcium-binding S100B interacts and co-localizes with IQGAP1 during dynamic rearrangement of cell membranes.* Journal of Biological Chemistry, 2002. **277**(51): p. 49998-50007.

144. Heizmann, C.W. and Cox, J.A., *New perspectives on S100 proteins_ a multi-functional $Ca^{2+}$-, $Zn^{2+}$- and $Cu^{2+}$ binding protein family.* BioMetals, 1998. **11**: p. 383-397.

145. Schmidt, A.M., Yan, S.D., Yan, S.F., *et al.*, *The multiligand receptor RAGE as a progression factor amplifying immune and inflammatory responses.* Journal of Clinical Investigation, 2001. **108**(7): p. 949-956.

146. Leclerc, E., Fritz, G., Vetter, S.W., *et al.*, *Binding of S100 proteins to RAGE: an update.* Biochimica *et* Biophysica Acta, 2009. **1793**(6): p. 993-1007.

147. Bierhaus, A., Humpert, P.M., Morcos, M., *et al.*, *Understanding RAGE, the receptor for advanced glycation end products.* Journal of Molecular Medicine (Berlin, Germany), 2005. **83**(11): p. 876-886.

148. Sasaki, N., Takeuchi, M., Chowei, H., *et al.*, *Advanced glycation end products (AGE) and their receptor (RAGE) in the brain of patients with Creutzfeldt-Jakob disease with prion plaques.* Neuroscience Letters, 2002. **326**(2): p. 117-120.

149. Katakami, N., Matsuhisa, M., Kaneto, H., *et al.*, *Decreased endogenous secretory advanced glycation end product receptor in type 1 diabetic patients.* Diabetes Care, 2005. **28**(11): p. 2716-2721.

150. Fuentes, M.K., Nigavekar, S.S., Arumugam, T., *et al.*, *RAGE activation by S100P in colon cancer stimulates growth, migration, and cell signaling pathways.* Diseases of the Colon and Rectum, 2007. **50**(8): p. 1230-1240.

151. Arumugam, T., Simeone, D.M., Schmidt, A.M., *et al.*, *S100P stimulates cell proliferation and survival via receptor for activated glycation end products (RAGE).* Journal of Biological Chemistry, 2004. **279**(7): p. 5059-5065.

152. Bartling, B., Hofmann, H.S., Weigle, B., *et al.*, *Down-regulation of the receptor for advanced glycation end-products (RAGE) supports non-small cell lung carcinoma.* Carcinogenesis, 2005. **26**(2): p. 293-301.

153. Hsieh, H.L., Schäfer, B.W., Sasaki, N., *et al.*, *Expression analysis of S100 proteins and RAGE in human tumors using tissue microarrays.* Biochemical and Biophysical Research Communications, 2003. **307**(2): p. 375-381.

154. Brodersen, D.E., Etzerodt, M., Madsen, P., *et al.*, *EF-hands at atomic resolution: the structure of human psoriasin (S100A7) solved by MAD phasing.* Structure, 1998. **6**(4): p. 477-489.

155. Brodersen, D.E., Nyborg, J., and Kjeldgaard, M., *Zinc-binding site of an S100 protein revealed. Two crystal structures of $Ca^{2+}$-bound human psoriasin (S100A7) in the $Zn^{2+}$-loaded and $Zn^{2+}$-free states.* Biochemistry, 1999. **38**(6): p. 1695-1704.

156. Drohat, A.C., Baldisseri, D.M., Rustandi, R.R., *et al., Solution structure of calcium-bound rat S100B (ββ) as determined by nuclear magnetic resonance spectroscopy.* Biochemistry, 1998. **37**(9): p. 2729-2740.

157. Ishikawa, K., Nakagawa, A., Tanaka, I., *et al., The structure of human MRP8, a member of the S100 calcium-binding protein family, by MAD phasing at 1.9 A resolution.* Acta Crystallographica Section D: Biological Crystallography, 2000. **56**(5): p. 559-566.

158. Otterbein, L.R., Kordowska, J., Witte-Hoffmann, C., *et al., Crystal Structures of S100A6 in the $Ca^{2+}$-Free and $Ca^{2+}$-Bound States: The Calcium Sensor Mechanism of S100 Proteins Revealed at Atomic Resolution.* Structure, 2002. **10**(4): p. 557-567.

159. Zhang, H., Wang, G., Ding, Y., *et al., The Crystal Structure at 2Å Resolution of the $Ca^{2+}$-binding Protein S100P.* Journal of Molecular Biology, 2003. **325**(4): p. 785-794.

160. Dempsey, A.C., Walsh, M.P., and Shaw, G.S., *Unmasking the annexin I interaction from the structure of Apo-S100A11.* Structure, 2003. **11**(7): p. 887-897.

161. Drohat, A.C., Tjandra, N., Baldisseri, D.M., *et al., The use of dipolar couplings for determining the solution structure of rat apo-S100B (ββ).* Protein Science, 1999. **8**(4): p. 800-809.

162. Kilby, P.M., Van Eldik, L.J., and Roberts, G.C.K., *The solution structure of the bovine S100B protein dimer in the calcium-free state.* Structure, 1996. **4**: p. 1041-1052.

163. Réty, S., Sopkova, J., Renouard, M., *et al., The crystal structure of a complex of p11 with the annexin II N-terminal peptide.* Nature Structural & Molecular Biology, 1999. **6**(1): p. 89-95.

164. Rustandi, R.R., Baldisseri, D.M., Inman, K.G., *et al., Three-dimensional solution structure of the calcium-signaling protein apo-S100A1 as determined by NMR.* Biochemistry, 2002. **41**(3): p. 788-796.

165. Bhattacharya, S., Large, E., Heizmann, C.W., *et al., Structure of the $Ca^{2+}$/S100B/NDR kinase peptide complex: insights into S100 target specificity and activation of the kinase.* Biochemistry, 2003. **42**(49): p. 14416-14426.

166. Inman, K.G., Yang, R., Rustandi, R.R., *et al., Solution NMR structure of S100B bound to the high-affinity target peptide TRTK-12.* Journal of Molecular Biology, 2002. **324**(5): p. 1003-1014.

167. Itou, H., Yao, M., Fujita, I., *et al., The crystal structure of human MRP14 (S100A9), a $Ca^{2+}$-dependent regulator protein in inflammatory process1.* Journal of Molecular Biology, 2002. **316**(2): p. 265-276.

168. Réty, S., Osterloh, D., Arié, J.P., *et al., Structural basis of the $Ca^{2+}$-dependent association between S100C (S100A11) and its target, the N-terminal part of annexin I.* Structure, 2000. **8**(2): p. 175-184.

169. Donato, R., *Intracellular and extracellular roles of S100 proteins.* Microscopy Research and Technique, 2003. **60**(6): p. 540-551.

170. Donato, R., *S100: a multigenic family of calcium-modulated proteins of the EF-hand type with intracellular and extracellular functional roles.* The International Journal of Biochemistry & Cell Biology, 2001. **33**(7): p. 637-668.

171. Parkkila, S., Pan, P.W., Ward, A., *et al., The calcium-binding protein S100P in normal and malignant human tissues.* BMC Clinical Pathology, 2008. **8**: p. 2-10.

172. Bartling, B., Rehbein, G., Schmitt, W.D., *et al., S100A2-S100P expression profile and diagnosis of non-small cell lung carcinoma: impairment by advanced tumour stages and neoadjuvant chemotherapy.* European Journal of Cancer, 2007. **43**(13): p. 1935-1943.

173. Diederichs, S., Bulk, E., Steffen, B., *et al., S100 family members and trypsinogens are predictors of distant metastasis and survival in early-stage non-small cell lung cancer.* Cancer Research, 2004. **64**(16): p. 5564-5569.

174. Guerreiro, D.S.I.D., Hu, Y., Russo, I., *et al., S100P calcium-binding protein overexpression is associated with immortalization of human breast epithelial cells in vitro and early stages of breast cancer development in vivo.* International Journal of Oncology, 2000. **16**(2): p. 231-271.

175. Hamada, S., Satoh, K., Hirota, M., *et al., Calcium-binding protein S100P is a novel diagnostic marker of cholangiocarcinoma.* Cancer Science, 2011. **102**(1): p. 150-156.

176. Hunter, M.J. and Chazin, W.J., *High level expression and dimer characterization of the S100 EF-hand proteins, migration inhibitory factor-related proteins 8 and 14.* Journal of Biological Chemistry, 1998. **273**(20): p. 12427-12435.

177. Arumugam, T., Simeone, D.M., Van Golen, K., *et al., S100P promotes pancreatic cancer growth, survival, and invasion.* Clinical Cancer Research, 2005. **11**(15): p. 5356-5364.

178. Zhou, Y., Frey, T.K., and Yang, J.J., *Viral calciomics: interplays between $Ca^{2+}$ and virus.* Cell Calcium, 2009. **46**(1): p. 1-17.

179. Becker, T., Gerke, V., Kube, E., *et al., S100P, a novel $Ca^{2+}$-binding protein from human placenta.* European Journal of Biochemistry, 1992. **207**(2): p. 541-547.

180. Emoto, Y., Kobayashi, R., Akatsuka, H., *et al., Purification and characterization of a new member of the S-100 protein family from human placenta.* Biochemical and Biophysical Research Communications, 1992. **182**(3): p. 1246-1253.

181. Basu, G.D., Azorsa, D.O., Kiefer, J.A., *et al., Functional evidence implicating S100P in prostate cancer progression.* International Journal of Cancer, 2008. **123**(2): p. 330-339.

182. Hammacher, A., Thompson, E.W., and Williams, E.D., *Interleukin-6 is a potent inducer of S100P, which is up-regulated in androgen-refractory and metastatic prostate cancer.* International Journal of Biochemistry and Cell Biology, 2005. **37**(2): p. 442-450.

183. Sato, N. and Hitomi, J., *S100P expression in human esophageal epithelial cells: human esophageal epithelial cells sequentially produce different S100 proteins in the process of differentiation.* The Anatomical Record, 2002. **267**(1): p. 60-69.

184. Beer, D.G., Kardia, S.L.R., Huang, C.C., *et al., Gene-expression profiles predict survival of patients with lung adenocarcinoma.* Nature Medicine, 2002. **8**(8): p. 816-824.

185. Dong, L., Wang, F., Yin, X., *et al., Overexpression of S100P promotes colorectal cancer metastasis and decreases chemosensitivity to 5-FU in vitro.* Molecular and Cellular Biochemistry, 2014. **389**(1-2): p. 257-264.

186. Namba, T., Homan, T., Nishimura, T., *et al., Up-regulation of S100P expression by non-steroidal anti-inflammatory drugs and its role in anti-tumorigenic effects.* Journal of Biological Chemistry, 2009. **284**(7): p. 4158-4167.

187. Onyeagucha, B.C., Mercado-Pimentel, M.E., Hutchison, J., *et al., S100P/RAGE signaling regulates microRNA-155 expression via AP-1 activation in colon cancer.* Experimental Cell Research, 2013. **319**(13): p. 2081-2090.

188. Sparvero, L.J., Asafu-Adjei, D., Kang, R., *et al., RAGE (Receptor for Advanced Glycation Endproducts), RAGE ligands, and their role in cancer and inflammation.* Journal of Translational Medicine, 2009. **7**(1): p. 17-37.

189. MOE, *Chemical Computing Group, Inc.,*. 2014.08: Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2015.

190. Sato, N., Fukushima, N., Matsubayashi, H., *et al., Identification of maspin and S100P as novel hypomethylation targets in pancreatic cancer using global gene expression profiling.* Oncogene, 2004. **23**(8): p. 1531-1538.

191. Arumugam, T. and Logsdon, C.D., *S100P: a novel therapeutic target for cancer.* Amino Acids, 2011. **41**(4): p. 893-899.

192. Crnogorac-Jurcevic, T., Missiaglia, E., Blaveri, E., *et al.*, *Molecular alterations in pancreatic carcinoma: expression profiling shows that dysregulated expression of S100 genes is highly prevalent.* Journal of Pathology, 2003. **201**(1): p. 63-74.

193. Goonesekere, N.C., Wang, X., Ludwig, L., *et al.*, *A Meta Analysis of Pancreatic Microarray Datasets Yields New Targets as Cancer Genes and Biomarkers.* PloS One, 2014. **9**(4).

194. Zhang, L., Farrell, J.J., Zhou, H., *et al.*, *Salivary transcriptomic biomarkers for detection of resectable pancreatic cancer.* Gastroenterology, 2010. **138**(3): p. 949-957 e941-947.

195. Arumugam, T., Ramachandran, V., and Logsdon, C.D., *Effect of cromolyn on S100P interactions with RAGE and pancreatic cancer growth and invasion in mouse models.* Journal of the National Cancer Institute, 2006. **98**(24): p. 1806-1818.

196. Choby, G.W. and Lee, S. *Pharmacotherapy for the treatment of asthma: current treatment options and future directions*. in *International Forum of Allergy & Rhinology*. 2015. **5**(S1): p.S35-S40.

197. Arumugam, T., Ramachandran, V., Maxwell, D., *et al.*, *Designing and Developing S100P Inhibitor 5-methyl Cromolyn (C5OH) for Pancreatic Cancer Therapy.* Molecular Cancer Therapeutics, 2013. **12**(5): p. 654-662.

198. Moss, G., Jones, K., Ritchie, J.T., *et al.*, *Plasma levels and urinary excretion of disodium cromoglycate after inhalation by human volunteers.* Toxicology and Applied Pharmacology, 1971. **20**(2): p. 147-156.

199. Ashton, M., Clark, B., Jones, K., *et al.*, *The absorption, metabolism and excretion of disodium cromoglycate in nine animal species.* Toxicology and Applied Pharmacology, 1973. **26**(3): p. 319-328.

200. Bevan, S. *Overview of drug discovery & development*. nd [cited 16th July 2014]; Available from: http://goo.gl/mc9nvj.

201. Patrick, G.L., *An introduction to medicinal chemistry*. 2013: 5th Ed. Oxford University Press. Oxford, UK. ISBN 9780199697397.

202. Imming, P., Sinning, C., and Meyer, A., *Drugs, their targets and the nature and number of drug targets.* Nature Reviews Drug Discovery, 2007. **6**(2): p. 1474-1776.

203. O'BRIEN, R.J. and Nunn, P.P., *The need for new drugs against tuberculosis: obstacles, opportunities, and next steps.* American Journal of Respiratory and Critical Care Medicine, 2001. **163**(5): p. 1055-1058.

204. Jamieson, E.R. and Lippard, S.J., *Structure, recognition, and processing of cisplatin-DNA adducts.* Chemical Reviews, 1999. **99**(9): p. 2467-2498.

205. Creemers, G., Lund, B., and Verweij, J., *Topoisomerase I inhibitors: topotecan and irenotecan.* Cancer Treatment Reviews, 1994. **20**(1): p. 73-96.

206. Smith, C., *Drug target validation: Hitting the target.* Nature, 2003. **422**(6929): p. 341-347.

207. Wiederschain, D., Susan, W., Chen, L., *et al.*, *Single-vector inducible lentiviral RNAi system for oncology target validation.* Cell Cycle, 2009. **8**(3): p. 498-504.

208. Xin, H., Bernal, A., Amato, F.A., *et al.*, *High-throughput siRNA-based functional target validation.* Journal of Biomolecular Screening, 2004. **9**(4): p. 286-293.

209. Asahi, M., Asahi, K., Jung, J.-C., *et al.*, *Role for Matrix Metalloproteinase 9 After Focal Cerebral Ischemia: Effects of Gene Knockout and Enzyme Inhibition With BB-94.* Journal of Cerebral Blood Flow and Metabolism, 2000. **20**(12): p. 1681-1689.

210. Kurreck, J., *Antisense and RNA interference approaches to target validation in pain research.* Current Opinion in Drug Discovery and Development, 2004. **7**(2): p. 179-187.

211. Ilag, L.L., Ng, J.H., Beste, G., *et al.*, *Emerging high-throughput drug target validation technologies.* Drug Discovery Today, 2002. **7**(18): p. S136-S142.

212. Zhang, W.-W., *The use of gene-specific IgY antibodies for drug target discovery.* Drug Discovery Today, 2003. **8**(8): p. 364-371.

213. Blank, M. and Blind, M., *Aptamers as tools for target validation.* Current Opinion in Chemical Biology, 2005. **9**(4): p. 336-342.

214. Craik, D.J., Swedberg, J.E., Mylne, J.S., *et al.*, *Cyclotides as a basis for drug design.* Expert Opinion on Drug Discovery, 2012. **7**(3): p. 179-194.

215. Kopec, K.K., Bozyczko-Coyne, D., and Williams, M., *Target identification and validation in drug discovery: the role of proteomics.* Biochemical Pharmacology, 2005. **69**(8): p. 1133-1139.

216. Yu, L., Coelho, J.E., Zhang, X., *et al.*, *Uncovering multiple molecular targets for caffeine using a drug target validation strategy combining A2A receptor knockout mice with microarray profiling.* Physiological Genomics, 2009. **37**(3): p. 199-210.

217. Chakraborty, C., Hsu, C.H., Wen, Z.H., *et al.*, *Zebrafish: a complete animal model for in vivo drug discovery and development.* Current Drug Metabolism, 2009. **10**(2): p. 116-124.

218. Xiao, A., Wang, Z., Hu, Y., *et al.*, *Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish.* Nucleic Acids Research, 2013. **41**(14): p. 464-674.

219. Terstappen, G.C. and Reggiani, A., *In silico research in drug discovery.* Trends in Pharmacological Sciences, 2001. **22**(1): p. 23-26.

220. Jenkins, J.L., Bender, A., and Davies, J.W., *In silico target fishing: Predicting biological targets from chemical structure.* Drug Discovery Today: Technologies, 2007. **3**(4): p. 413-421.

221. Bolívar, J.C., Machens, F., Brill, Y., *et al.*, *'In silico expression analysis', a novel PathoPlant web tool to identify abiotic and biotic stress conditions associated with specific cis-regulatory sequences.* Database: The Journal of Biological Databases and Curation, 2014. **2014**: p. 30-40.

222. Khanfar, M.A., AbuKhader, M.M., Alqtaishat, S., *et al.*, *Pharmacophore modeling, homology modeling, and in silico screening reveal mammalian target of rapamycin inhibitory activities for sotalol, glyburide, metipranolol, sulfamethizole, glipizide, and pioglitazone.* Journal of Molecular Graphics and Modelling, 2013. **42**: p. 39-49.

223. Floudas, C.A., *Computational methods in protein structure prediction.* Biotechnology and Bioengineering, 2007. **97**(2): p. 207-213.

224. Hopkins, A.L. and Groom, C.R., *The druggable genome.* Nature Reviews Drug Discovery, 2002. **1**(9): p. 727-730.

225. Arumugam, T., Ramachandran, V., Gomez, S.B., *et al.*, *S100P-derived RAGE antagonistic peptide reduces tumor growth and metastasis.* Clinical Cancer Research, 2012. **18**(16): p. 4356-4364.

226. Barry, S., Chelala, C., Lines, K., *et al.*, *S100P is a metastasis-associated gene that facilitates transendothelial migration of pancreatic cancer cells.* Clinical & Experimental Metastasis, 2013. **30**(3): p. 251-264.

227. Chandramouli, A., Mercado-Pimentel, M.E., Hutchinson, A., *et al.*, *The induction of S100P expression by the Prostaglandin E(2) (PGE(2))/EP4 receptor signaling pathway in colon cancer cells.* Cancer Biology & Therapy, 2010. **10**(10): p. 1056-1066.

228. Dakhel, S., Padilla, L., Adan, J., *et al.*, *S100P antibody-mediated therapy as a new promising strategy for the treatment of pancreatic cancer.* Oncogenesis, 2014. **3**(3): p. 92-101.

229. Kapetanovic, I., *Computer-aided drug discovery and development (CADDD): in silico-chemico-biological approach.* Chemico-Biological Interactions, 2008. **171**(2): p. 165-176.

230. Chen, Y.-C., *Beware of docking!* Trends in Pharmacological Sciences, 2015. **36**(2): p. 78-95.

231. Berman, H.M., Westbrook, J., Feng, Z., *et al.*, *The protein data bank.* Nucleic Acids Research, 2000. **28**(1): p. 235-242.

232. Irwin, J.J. and Shoichet, B.K., *ZINC-a free database of commercially available compounds for virtual screening.* Journal of Chemical Information and Modeling, 2005. **45**(1): p. 177-182.

233. Allen, F.H., *The Cambridge Structural Database: a quarter of a million crystal structures and rising.* Acta Crystallographica Section B: Structural Science, 2002. **58**(3): p. 380-388.

234. Veselovsky, A. and Ivanov, A., *Strategy of computer-aided drug design.* Current Drug Targets-Infectious Disorders, 2003. **3**(1): p. 33-40.

235. Barbaro, R., Betti, L., Botta, M., *et al.*, *Synthesis, biological evaluation, and pharmacophore generation of new pyridazinone derivatives with affinity toward α1-and α2-adrenoceptors 1.* Journal of Medicinal Chemistry, 2001. **44**(13): p. 2118-2132.

236. Brown, R.D. and Martin, Y.C., *Use of structure-activity data to compare structure-based clustering methods and descriptors for use in compound selection.* Journal of Chemical Information and Computer Sciences, 1996. **36**(3): p. 572-584.

237. Patterson, D.E., Cramer, R.D., Ferguson, A.M., *et al.*, *Neighborhood behavior: a useful concept for validation of "molecular diversity" descriptors.* Journal of Medicinal Chemistry, 1996. **39**(16): p. 3049-3059.

238. Speck-Planche, A., Kleandrova, V.V., Luan, F., *et al.*, *Multi-target drug discovery in anti-cancer therapy: fragment-based approach toward the design of*

*potent and versatile anti-prostate cancer agents.* Bioorganic & Medicinal Chemistry, 2011. **19**(21): p. 6239-6244.

239. Langer, T. and Wolber, G., *Pharmacophore definition and 3D searches.* Drug Discovery Today: Technologies, 2004. **1**(3): p. 203-207.

240. Lyne, P.D., *Structure-based virtual screening: an overview.* Drug Discovery Today, 2002. **7**(20): p. 1047-1055.

241. Cao, X., Jiang, J., Zhang, S., *et al.*, *Discovery of natural estrogen receptor modulators with structure-based virtual screening.* Bioorganic & Medicinal Chemistry Letters, 2013. **23**(11): p. 3329-3333.

242. Wang, L., Yang, C., Lu, W., *et al.*, *Discovery of new potent inhibitors for carbonic anhydrase IX by structure-based virtual screening.* Bioorganic & Medicinal Chemistry Letters, 2013. **23**(12): p. 3496-3499.

243. Leach, A.R., Shoichet, B.K., and Peishoff, C.E., *Prediction of protein-ligand interactions. Docking and scoring: successes and gaps.* Journal of Medicinal Chemistry, 2006. **49**(20): p. 5851-5855.

244. Kontoyianni, M., McClellan, L.M., and Sokol, G.S., *Evaluation of docking performance_comparative data on docking algorithms.* Journal of Medicinal Chemistry, 2004. **47**(3): p. 558-565.

245. Warren, G.L., Andrews, C.W., Capelli, A.-M., *et al.*, *A critical assessment of docking programs and scoring functions.* Journal of Medicinal Chemistry, 2006. **49**(20): p. 5912-5931.

246. Tame, J.R., *Scoring functions: a view from the bench.* Journal of Computer-Aided Molecular Design, 1999. **13**(2): p. 99-108.

247. David, L., Nielsen, P.A., Hedstrom, M., *et al.*, *Scope and limitation of ligand docking: Methods, scoring functions and protein targets.* Current Computer-Aided Drug Design, 2005. **1**(3): p. 275-306.

248. Langer, T. and Krovat, E., *Chemical feature-based pharmacophores and virtual library screening for discovery of new leads.* Current Opinion in Drug Discovery & Development, 2003. **6**(3): p. 370-376.

249. Kontoyianni, M., McClellan, L.M., and Sokol, G.S., *Evaluation of docking performance: comparative data on docking algorithms.* Journal of Medicinal Chemistry, 2004. **47**(3): p. 558-565.

250. Böhm, H.J. and Stahl, M., *The use of scoring functions in drug discovery applications.* Reviews in Computational Chemistry, 2003. **18**: p. 41-87.

251. Kellenberger, E., Rodrigo, J., Muller, P., *et al.*, *Comparative evaluation of eight docking tools for docking and virtual screening accuracy.* Proteins: Structure, Function, and Bioinformatics, 2004. **57**(2): p. 225-242.

252. Morris, G.M., Goodsell, D.S., Halliday, R.S., *et al.*, *Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function.* Journal of Computational Chemistry, 1998. **19**(14): p. 1639-1662.

253. Verdonk, M.L., Cole, J.C., Hartshorn, M.J., *et al.*, *Improved protein–ligand docking using GOLD.* Proteins: Structure, Function, and Bioinformatics, 2003. **52**(4): p. 609-623.

254. Friesner, R.A., Banks, J.L., Murphy, R.B., *et al.*, *Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy.* Journal of Medicinal Chemistry, 2004. **47**(7): p. 1739-1749.

255. Kramer, B., Rarey, M., and Lengauer, T., *Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking.* Proteins: Structure, Function, and Bioinformatics, 1999. **37**(2): p. 228-241.

256. Spitzer, R. and Jain, A.N., *Surflex-Dock: Docking benchmarks and real-world application.* Journal of Computer-Aided Molecular Design, 2012. **26**(6): p. 687-699.

257. Oleg, T. and Olson, A. J., *AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, eficient optimization, and multithreading.* Journal of Computational Chemistry, 2010. **31**(2): p. 455-461.

258. Vilar, S., Cozza, G., and Moro, S., *Medicinal chemistry and the molecular operating environment (MOE): application of QSAR and molecular docking to drug discovery.* Current Topics in Medicinal Chemistry, 2008. **8**(18): p. 1555-1572.

259. Le Guilloux, V., Schmidtke, P., and Tuffery, P., *Fpocket: an open source platform for ligand pocket detection.* BMC Bioinformatics, 2009. **10**: p. 168-178.

260. Labute, P. and Santavy, M., *Locating binding sites in protein structures.* Journal of Chemical Computing Group, 2007 [Cited 5th May 2015]. Available from https://goo.gl/bFc5Zy.

261. Laurie, A.T. and Jackson, R.M., *Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites.* Bioinformatics, 2005. **21**(9): p. 1908-1916.

262. Schmidtke, P., Souaille, C., Estienne, F., *et al.*, *Large-scale comparison of four binding site detection algorithms.* Journal of Chemical Information and Modeling, 2010. **50**(12): p. 2191-2200.

263. An, J., Totrov, M., and Abagyan, R., *Pocketome via comprehensive identification and classification of ligand binding envelopes.* Molecular & Cellular Proteomics, 2005. **4**(6): p. 752-761.

264. Poupon, A., *Voronoi and Voronoi-related tessellations in studies of protein structure and interaction.* Current Opinion in Structural Biology, 2004. **14**(2): p. 233-241.

265. Labute, P. and Santavy, M., *Locating binding sites in protein structures.* Chemical Computing Group, Inc.: Montreal, Quebec, Canada, 2001.

266. Laurie, R., Alasdair, T., and Jackson, R.M., *Methods for the prediction of protein-ligand binding sites for structure-based drug design and virtual ligand screening.* Current Protein and Peptide Science, 2006. **7**(5): p. 395-406.

267. Levitt, D.G. and Banaszak, L.J., *POCKET: A computer graphies method for identifying and displaying protein cavities and their surrounding amino acids.* Journal of Molecular Graphics, 1992. **10**(4): p. 229-234.

268. Mannhold, R., Kubinyi, H., Folkers, G., *et al.*, *Virtual Screening: Volume 48-Principles, Challenges, and Practical Guidelines*. 2011: John Wiley & Sons. Weinheim, Germany. ISBN: 9783527326365.

269. Gong, L.-L., Fang, L.-H., Peng, J.-H., *et al.*, *Integration of virtual screening with high-throughput screening for the identification of novel Rho-kinase I inhibitors.* Journal of Biotechnology, 2010. **145**(3): p. 295-303.

270. Jenkins, J.L., Kao, R.Y., and Shapiro, R., *Virtual screening to enrich hit lists from high-throughput screening: A case study on small-molecule inhibitors of angiogenin.* Proteins: Structure, Function, and Bioinformatics, 2003. **50**(1): p. 81-93.

271. Paiva, A.M., Vanderwall, D.E., Blanchard, J.S., *et al.*, *Inhibitors of dihydrodipicolinate reductase, a key enzyme of the diaminopimelate pathway of Mycobacterium tuberculosis.* Biochimica *et* Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology, 2001. **1545**(1): p. 67-77.

272. Boisclair, M.D., Egan, D.A., Huberman, K., *et al.*, *High-Throughput Screening in Industry*, in *Anticancer Drug Development Guide*. Springer, 2004. p. 23-39.

273. Persidis, A., *High-throughput screening.* Nature Biotechnology, 1998. **16**(5): p. 488-489.

274. Macarron, R., Banks, M.N., Bojanic, D., *et al.*, *Impact of high-throughput screening in biomedical research.* Nature reviews Drug discovery, 2011. **10**(3): p. 188-195.

275. Cao, X., Jiang, J., Zhang, S., *et al.*, *Discovery of natural estrogen receptor modulators with structure-based virtual screening.* Bioorganic and Medicinal Chemistry Letters, 2013. **23**(11): p. 3329-3333.

276. Poh, M.K., Yip, A., Zhang, S., *et al.*, *A small molecule fusion inhibitor of dengue virus.* Antiviral Research, 2009. **84**(3): p. 260-266.

277. Oprea, T.I., Davis, A.M., Teague, S.J., *et al.*, *Is there a difference between leads and drugs? A historical perspective.* Journal of Chemical Information and Computer Sciences, 2001. **41**(5): p. 1308-1315.

278. Lipinski, C.A., *Drug-like properties and the causes of poor solubility and poor permeability.* Journal of Pharmacological and Toxicological Methods, 2000. **44**(1): p. 235-249.

279. Lipinski, C.A., *Lead-and drug-like compounds: the rule-of-five revolution.* Drug Discovery Today: Technologies, 2004. **1**(4): p. 337-341.

280. Downs, G.M. and Barnard, J.M., *Clustering methods and their uses in computational chemistry.* Reviews in Computational Chemistry, 2002. **18**: p. 1-40.

281. Willett, P., *Chemoinformatics – similarity and diversity in chemical libraries.* Current Opinion in Biotechnology, 2000. **11**(1): p. 85-88.

282. Cao, Y., Jiang, T., and Girke, T., *A maximum common substructure-based algorithm for searching and predicting drug-like compounds.* Bioinformatics, 2008. **24**(13): p. i366-i374.

283. Koltzscher, M., Neumann, C., Konig, S., *et al.*, *$Ca^{2+}$-dependent binding and activation of dormant ezrin by dimeric S100P.* Molecular Biology of the Cell, 2003. **14**(6): p. 2372-2384.

284. Tutar, Y., *Dimerization and ion binding properties of S100P protein.* Protein and Peptide Letters, 2006. **13**(3): p. 301-306.

285. Crnogorac-Jurcevic, T., Efthimiou, E., Nielsen, T., *et al.*, *Expression profiling of microdissected pancreatic adenocarcinomas.* Oncogene, 2002. **21**(29): p. 4587-4594.

286. Friess, H., Ding, J., Kleeff, J., *et al., Microarray-based identification of differentially expressed growth-and metastasis-associated genes in pancreatic cancer.* Cellular and Molecular Life Sciences CMLS, 2003. **60**(6): p. 1180-1199.

287. Sato, N., Fukushima, N., Matsubayashi, H., *et al., Identification of maspin and S100P as novel hypomethylation targets in pancreatic cancer using global gene expression profiling.* Oncogene, 2004. **23**(8): p. 1531-1538.

288. Torres Schor, A.P., Carvalho, F.M., Kemp, C., *et al., S100P calcium-binding protein expression is associated with high-risk proliferative lesions of the breast.* Oncology Reports, 2006. **15**(1): p. 3-6.

289. Wang, G., Platt-Higgins, A., Carroll, J., *et al., Induction of metastasis by S100P in a rat mammary model and its association with poor survival of breast cancer patients.* Cancer Research, 2006. **66**(2): p. 1199-1207.

290. Bulk, E., Hascher, A., Liersch, R., *et al., Adjuvant therapy with small hairpin RNA interference prevents non–small cell lung cancer metastasis development in mice.* Cancer Research, 2008. **68**(6): p. 1896-1904.

291. Terris, B., Blaveri, E., Crnogorac-Jurcevic, T., *et al., Characterization of gene expression profiles in intraductal papillary-mucinous tumors of the pancreas.* The American Journal of Pathology, 2002. **160**(5): p. 1745-1754.

292. Austermann, J., Nazmi, A.R., Muller-Tidow, C., *et al., Characterization of the $Ca^{2+}$ -regulated ezrin-S100P interaction and its role in tumor cell migration.* Journal of Biological Chemistry, 2008. **283**(43): p. 29331-29340.

293. Lines, K.E., Chelala, C., Dmitrovic, B., *et al., S100P-binding protein, S100PBP, mediates adhesion through regulation of cathepsin Z in pancreatic cancer cells.* The American Journal of Pathology, 2012. **180**(4): p. 1485-1494.

294. Penumutchu, S.R., Chou, R.-H., and Yu, C., *Structural Insights into Calcium-Bound S100P and the V Domain of the RAGE Complex.* PloS One, 2014. **9**(8): p. e103947-e103962.

295. Hwang, Y.S. and Chmielewski, J., *Development of low molecular weight HIV-1 protease dimerization inhibitors.* Journal of Medicinal Chemistry, 2005. **48**(6): p. 2239-2242.

296. Shultz, M.D., Ham, Y.-W., Lee, S.-G., *et al., Small-molecule dimerization inhibitors of wild-type and mutant HIV protease: a focused library approach.* Journal of the American Chemical Society, 2004. **126**(32): p. 9886-9887.

297. Bowman, M.J., Byrne, S., and Chmielewski, J., *Switching between allosteric and dimerization inhibition of HIV-1 protease.* Chemistry & Biology, 2005. **12**(4): p. 439-444.

298. Lee, Y.C., Volk, D.E., Thiviyanathan, V., *et al.*, *Letter to the Editor: NMR structure of the Apo-S100P protein.* Journal of Biomolecular NMR, 2004. **29**(3): p. 399-402.

299. Pettersen, E.F., Goddard, T.D., Huang, C.C., *et al.*, *UCSF Chimera - a visualization system for exploratory research and analysis.* Journal of Computational Chemistry, 2004. **25**(13): p. 1605-1612.

300. Labute, P., *Protonate3D: assignment of ionization states and hydrogen coordinates to macromolecular structures.* Proteins: Structure, Function, and Bioinformatics, 2009. **75**(1): p. 187-205.

301. Corbeil, C.R., Williams, C.I., and Labute, P., *Variability in docking success rates due to dataset preparation.* Journal of Computer-Aided Molecular Design, 2012. **26**(6): p. 775-786.

302. Weininger, D., *SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules.* Journal of Chemical Information and Computer Sciences, 1988. **28**(1): p. 31-36.

303. Raymond, J.W. and Willett, P., *Maximum common subgraph isomorphism algorithms for the matching of chemical structures.* Journal of Computer-Aided Molecular Design, 2002. **16**(7): p. 521-533.

304. Schneider, M., Fu, X., and Keating, A.E., *X-ray vs. NMR structures as templates for computational protein design.* Proteins: Structure, Function, and Bioinformatics, 2009. **77**(1): p. 97-110.

305. Wüthrich, K., *NMR studies of structure and function of biological macromolecules (Nobel Lecture).* Angewandte Chemie International Edition, 2003. **42**(29): p. 3340-3363.

306. Spronk, C.A., Nabuurs, S.B., Krieger, E., *et al.*, *Validation of protein structures derived by NMR spectroscopy.* Progress in Nuclear Magnetic Resonance Spectroscopy, 2004. **45**(3): p. 315-337.

307. Wagner, G., Hyberts, S.G., and Havel, T.F., *NMR structure determination in solution: a critique and comparison with X-ray crystallography.* Annual Review of Biophysics and Biomolecular Structure, 1992. **21**(1): p. 167-198.

308. Brünger, A.T., *Free R value: a novel statistical quantity for assessing the accuracy of crystal structures.* Nature, 1992. **355**: p. 472-475.

309. Davis, A.M., Teague, S.J., and Kleywegt, G.J., *Application and limitations of X-ray crystallographic data in structure-based ligand and drug design.* Angewandte Chemie, International Edition in English, 2003. **42**(24): p. 2718-2736.

310. Hoch, J., *An amateur looks at error analysis in the determination of protein structure by NMR,* in *Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy*, Hoch, J., Poulsen, F., and Redfield, C., Editors. 1991, Springer US. p. 253-267.

311. Kay, L.E., Ikura, M., Tschudin, R., *et al.*, *Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins.* Journal of Magnetic Resonance, 1990. **89**(3): p. 496-514.

312. Zhao, D. and Jardetzky, O., *An assessment of the precision and accuracy of protein structures determined by NMR: dependence on distance errors.* Journal of Molecular Biology, 1994. **239**(5): p. 601-607.

313. Laskowski, R.A., *Structural quality assurance.* Methods of Biochemical Analysis, 2003. **44**: p. 273-303.

314. Permyakov, S.E., Ismailov, R.G., Xue, B., *et al.*, *Intrinsic disorder in S100 proteins.* Molecular Biosystems, 2011. **7**(7): p. 2164-2180.

315. Kleywegt, G.J. and Jones, T.A., *Phi/Psi-chology: Ramachandran revisited.* Structure, 1996. **4**(12): p. 1395-1400.

316. Huang, S.Y. and Zou, X., *Efficient molecular docking of NMR structures: Application to HIV-1 protease.* Protein Science, 2007. **16**(1): p. 43-51.

317. Hendlich, M., Rippmann, F., and Barnickel, G., *LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins.* Journal of Molecular Graphics and Modelling, 1997. **15**(6): p. 359-363.

318. Liang, J., Woodward, C., and Edelsbrunner, H., *Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design.* Protein Science, 1998. **7**(9): p. 1884-1897.

319. Zhou, W. and Yan, H., *Alpha shape and Delaunay triangulation in studies of protein-related interactions.* Briefings in Bioinformatics, 2014. 15(1): p. 54-64.

320. Soga, S., Shirai, H., Kobori, M., *et al.*, *Use of amino acid composition to predict ligand-binding sites.* Journal of Chemical Information and Modeling, 2007. **47**(2): p. 400-406.

321. Koltzscher, M. and Gerke, V., *Identification of hydrophobic amino acid residues involved in the formation of S100P homodimers in vivo.* Biochemistry, 2000. **39**: p. 9533-9539.

322. Heizmann, C.W., Fritz, G., and Schafer, B., *S100 proteins: structure, functions and pathology.* Frontiers in Bioscience, 2002. **7**(1): p. 13561-11368.

323. Penumutchu, S.R., Chou, R.-H., and Yu, C., *Interaction between S100P and the anti-allergy drug cromolyn.* Biochemical and Biophysical Research Communications, 2014. **454**(3): p. 404-409.

324. Ye, Y. and Godzik, A., *Flexible structure alignment by chaining aligned fragment pairs allowing twists.* Bioinformatics, 2003. **19**(suppl 2): p. 246-255.

325. Austermann, J., Nazmi, A.R., Heil, A., *et al.*, *Generation and characterization of a novel, permanently active S100P mutant.* Biochimica *et* Biophysica Acta (BBA)-Molecular Cell Research, 2009. **1793**(6): p. 1078-1085.

326. Osterloh, D., Ivanenkov, V.V., and Gerke, V., *Hydrophobic residues in the C-terminal region of S100A1 are essential for target protein binding butnot for dimerization.* Cell Calcium, 1998. **24**(2): p. 137-151.

327. Garbuglia, M., Verzini, M., Rustandi, R.R., *et al.*, *Role of the C-terminal extension in the interaction of S100A1 with GFAP, tubulin, the S100A1-and S100B-inhibitory peptide, TRTK-12, and a peptide derived from p53, and the S100A1 inhibitory effect on GFAP polymerization.* Biochemical and Biophysical Research Communications, 1999. **254**(1): p. 36-41.

328. Mahmoud, N., *Testing of novel small molecules against S100P in pancreatic cancer,* 2012. Unpublished Masters Dissertation, Queen Mary University of London, London.

329. Corey, E.J., *Robert Robinson Lecture. Retrosynthetic thinking—essentials and examples.* Chemical Society Reviews, 1988. **17**: p. 111-133.

330. Corey, E.J., *The logic of chemical synthesis: Multistep synthesis of complex carbogenic molecules.* Angewandte Chemie, International Edition in English, 1991. **30**(5): p. 455-465.

331. Taylor, P., *Mechanism and Synthesis*. Molecular World. 2002, Cambridge, UK: Royal Society of Chemistry.

332. Sonntag, N.O., *The Reactions of Aliphatic Acid Chlorides.* Chemical Reviews, 1953. **52**(2): p. 237-416.

333. Frechette, R.F. and Beach, M.J., *The Preparation of 2-Hydroxyethyl-2, 3-dihydro-2H-1, 4-benzoxazin-3 (4H)-one Derivatives.* Synthetic Communications, 1998. **28**(18): p. 3471-3478.

334. Kim, C.-E., Lim, S.-K., and Kim, J.-S., *In vivo antitumor effect of cromolyn in PEGylated liposomes for pancreatic cancer.* Journal of Controlled Release, 2012. **157**(2): p. 190-195.

335. O'Toole, S.A., Sheppard, B.L., McGuinness, E.P.J., *et al.*, *The MTS assay as an indicator of chemosensitivity/resistance in malignant gynaecological tumours.* Cancer Detection and Prevention, 2003. **27**(1): p. 47-54.

336. Niles, A., Scurria, M., Bernad, L., *et al.*, *Using protease biomarkers to measure viability and cytotoxicity.* Cell Notes, 2007. **19**: p. 16-20.

337. Riss, T.L. and Moravec, R.A., *Cell proliferation assays: improved homogeneous methods used to measure the number of cells in culture.* CELIS, JE Cell Biology: a Labotatory Handbook. San Diego, CA.: Elsevier, 2005. p. 25-32.

338. Whiteman, H.J., Weeks, M.E., Dowen, S.E., *et al.*, *The role of S100P in the invasion of pancreatic cancer cells is mediated through cytoskeletal changes and regulation of cathepsin D.* Cancer Research, 2007. **67**(18): p. 8633-8642.

339. Mohamed, M.A., Yamada, K., and Tomioka, K., *Accessing the amide functionality by the mild and low-cost oxidation of imine.* Tetrahedron Letters, 2009. **50**(26): p. 3436-3438.

340. Blackburn, L. and Taylor, R.J.K., *In Situ Oxidation-Imine Formation-Reduction Routes from Alcohols to Amines.* Organic Letters, 2001. **3**(11): p. 1637-1639.

341. Berry, D.J., DiGiovanna, C.V., and Murugan, R., *Catalysis by 4-dialkylaminopyridines.* ARKIVOC: Free Online Journal of Organic Chemistry, 2001. **2**: p. 944-964.

342. Chen, X. and Reynolds, C.H., *Performance of similarity measures in 2D fragment-based similarity searching: comparison of structural descriptors and similarity coefficients.* Journal of Chemical Information and Computer Sciences, 2002. **42**(6): p. 1407-1414.

343. Danishefsky, S. and Kitahara, T., *Useful diene for the Diels-Alder reaction.* Journal of the American Chemical Society, 1974. **96**(25): p. 7807-7808.

344. Kagan, H.B. and Riant, O., *Catalytic asymmetric Diels-Alder reactions.* Chemical Reviews, 1992. **92**(5): p. 1007-1019.

345. Nicolaou, K.C., Snyder, S.A., Montagnon, T., *et al.*, *The Diels–Alder reaction in total synthesis.* Angewandte Chemie International Edition, 2002. **41**(10): p. 1668-1698.

346. Weinreb, S.M. and Staib, R.R., *Synthetic aspects of Diels-Alder cycloadditions with heterodienophiles.* Tetrahedron, 1982. **38**(21): p. 3087-3128.

347. Diels, O. and Alder, K., *Synthesen in der hydroaromatischen Reihe.* Justus Liebigs Annalen der Chemie, 1928. **460**(1): p. 98-122.

348. Corey, E.J., Imwinkelried, R., Pikul, S., *et al.*, *Practical enantioselective Diels-Alder and aldol reactions using a new chiral controller system.* Journal of the American Chemical Society, 1989. **111**(14): p. 5493-5495.

349. Singleton, D.A. and Martinez, J.P., *High reactivity, regioselectivity, and endo-stereoselectivity of vinyl boranes in Diels-Alder reactions.* Journal of the American Chemical Society, 1990. **112**(20): p. 7423-7424.

350. Yoshizawa, M., Tamura, M., and Fujita, M., *Diels-Alder in aqueous molecular hosts: unusual regioselectivity and efficient catalysis.* Science, 2006. **312**(5771): p. 251-254.

351. Bah, J. and Franzén, J., *Carbocations as Lewis Acid Catalysts in Diels–Alder and Michael Addition Reactions.* Chemistry-A European Journal, 2014. **20**(4): p. 1066-1072.

352. Evans, D.A. and Wu, J., *Enantioselective rare-earth catalyzed quinone Diels-Alder reactions.* Journal of the American Chemical Society, 2003. **125**(34): p. 10162-10163.

353. Kobayashi, S. and Ishitani, H., *Lanthanide (III)-Catalyzed Enantioselective Diels-Alder Reactions. Stereoselective Synthesis of Both Enantiomers by Using a Single Chiral Source and a Choice of Achiral Ligands.* Journal of the American Chemical Society, 1994. **116**(9): p. 4083-4084.

354. Bova, S., Saponara, S., Rampa, A., *et al.*, *Anthracene based compounds as new L-type $Ca^{2+}$ channel blockers: design, synthesis, and full biological profile.* Journal of Medicinal Chemistry, 2009. **52**(5): p. 1259-1262.

355. Perreux, L. and Loupy, A., *A tentative rationalization of microwave effects in organic synthesis according to the reaction medium, and mechanistic considerations.* Tetrahedron, 2001. **57**(45): p. 9199-9223.

356. Díaz-Ortiz, A., Carrillo, J., Díez-Barra, E., *et al.*, *Diels-Alder cycloaddition of vinylpyrazoles. Synergy between microwave irradiation and solvent-free conditions.* Tetrahedron, 1996. **52**(27): p. 9237-9248.

357. Loupy, A., Maurel, F., and Sabatié-Gogová, A., *Improvements in Diels–Alder cycloadditions with some acetylenic compounds under solvent-free microwave-assisted conditions: experimental results and theoretical approaches.* Tetrahedron, 2004. **60**(7): p. 1683-1691.

358. Alston, P.V., Ottenbrite, R.M., and Newby, J., *Regioselectivity in the Diels-Alder reaction of 9-substituted anthracenes.* The Journal of Organic Chemistry, 1979. **44**(26): p. 4939-4943.

359. Meek, J.S., Wilgus, D.R., and Dann, J.R., *Diels-Alder Reactions of 9-Substituted Anthracens. 1 IV. 9-Nitroanthracene and 9-Anthramide.* Journal of the American Chemical Society, 1960. **82**(10): p. 2566-2569.

360. Adams, H., Bawa, R.A., McMillan, K.G., *et al.*, *Asymmetric control in Diels–Alder cycloadditions of chiral 9-aminoanthracenes by relay of stereochemical information.* Tetrahedron: Asymmetry, 2007. **18**(8): p. 1003-1012.

361. Bawa, R.A. and Jones, S., *Synthesis and Diels–Alder reactions of 9-(4-benzyloxazolin-2-yl) anthracene.* Tetrahedron, 2004. **60**(12): p. 2765-2770.

362. Raimondi, L., Benaglia, M., and Cozzi, F., *Aliphatic C–H/π and heteroatom/π interactions in N-Aryl-3, 4-(9′, 10′-dihydroanthracene-9′, 10′-diyl) succinimides.* European Journal of Organic Chemistry, 2014. **2014**(23): p. 4993-4998.

363. Wade, L.G.J., *The Diels-Alder reaction of anthracene with maleic anhydride*, in *Modular Laboratory Program in Chemistry*, Neidig, H.A. and Jeffers, J., Editors. 1998, Chemical Education Resources, Inc.: Pennsylvania, USA.

364. Lima, L.d.M., Castro, P., Machado, A.L., *et al.*, *Synthesis and anti-inflammatory activity of phthalimide derivatives, designed as new thalidomide analogues.* Bioorganic & Medicinal Chemistry, 2002. **10**(9): p. 3067-3073.

365. Thapaliya, E.R., Captain, B., and Raymo, F.i.M., *Photoactivatable anthracenes.* The Journal of Organic Chemistry, 2014. **79**(9): p. 3973-3981.

366. Weber, E., Finge, S., and Csoeregh, I., *Modular design of hosts involving a rigid succinimide framework and N-bonded lateral groups. Crystalline inclusion properties and crystal structures of inclusion compounds with dioxane, methanol, and DMF.* The Journal of Organic Chemistry, 1991. **56**(26): p. 7281-7288.

367. Bratko, I., Mallet-Ladeira, S., Saffon, N., *et al.*, *9, 10-Dihydroanthracenyl structures: original ligands for the synthesis of polymetallic complexes through selective π-coordination.* Dalton Transactions, 2013. **42**(4): p. 1136-1143.

368. Cava, M. and Schlessinger, R., *2, 3-Naphthoquinonoid heterocycles. II. A highly stabilized thiadiazole analog of anthracene.* Tetrahedron Letters, 1964. **5**(50): p. 3815-3818.

369. Sanhes, D., Favier, I., Saffon, N., *et al.*, *Stereo-specific synthesis of hydroanthracene-dicarboximides.* Tetrahedron Letters, 2008. **49**(47): p. 6720-6723.

370. Bartlett, P.D. and Cohen, S.G., *Bicyclic structures prohibiting the walden inversion. some bicyclo [2, 2, 2] octane derivatives with substituents at the bridge-head1.* Journal of the American Chemical Society, 1940. **62**(5): p. 1183-1189.

371. Roberts, R. and Yavari, F., *Kinetics and mechanism of Diels-Alder additions of tetracyanoethylene to anthracene derivatives-III: Effect of substituents on rates and intermediate complex formation.* Tetrahedron, 1981. **37**(15): p. 2657-2662.

372. Kulikowska, A., *Synthesis of compounds designed to inhibit the protein S100P, leading to a possible treatment of pancreatic cancer,* 2015. Unpublished Master of Pharmacy Dissertation, University of Hertfordshire, Hatfield.

373. Grewal, A., *Synthesis of computer-designed inhibitors of the calcium-binding protein S100P, with potential therapeutic effects in pancreatic cancer,* 2015. Unpublished Master of Pharmacy Dissertation, University of Hertfordshire, Hatfield.

374. Wu, X., Kang, W., Zhu, D., *et al.*, *Synthesis, crystal structure and biological activities of two novel organotin (IV) complexes constructed from 12-(4-methylbenzoyl)-9, 10-dihydro-9, l0-ethanoanthracene-11-carboxylic acid.* Journal of Organometallic Chemistry, 2009. **694**(18): p. 2981-2986.

375. Yalkowsky, S.H., Krzyzaniak, J.F., and Myrdal, P.B., *Relationships between melting point and boiling point of organic compounds.* Industrial & Engineering Chemistry Research, 1994. **33**(7): p. 1872-1877.

376. Brown, R. and Brown, R., *Melting point and molecular symmetry.* Journal of Chemical Education, 2000. **77**(6): p. 724.

377. Pinal, R., *Effect of molecular symmetry on melting temperature and solubility.* Organic & Biomolecular Chemistry, 2004. **2**(18): p. 2692-2699.

378. Gavezzotti, A., *Molecular symmetry, melting temperatures and melting enthalpies of substituted benzenes and naphthalenes.* Journal of the Chemical Society, Perkin Transactions 2, 1995. (7): p. 1399-1404.

379. Craig, P.N., *Interdependence between physical parameters and selection of substituent groups for correlation studies.* Journal of Medicinal Chemistry, 1971. **14**(8): p. 680-684.

380. Chan, F.K.-M., Moriwaki, K., and De Rosa, M.J., *Detection of necrosis by release of lactate dehydrogenase activity*, in *Immune Homeostasis*. 2013, Springer. p. 65-70.

381. Capps, R.D., Batsakis, J.G., Briere, R.O., *et al.*, *An automated colorimetric (Tetrazolium salt) assay for serum lactate dehydrogenase.* Clinical Chemistry, 1966. **12**(7): p. 406-413.

382. Rothwell, D., Jendrzejczak, B., Becker, M., *et al.*, *Lactate dehydrogenase activities in serum and plasma.* Clinical Chemistry, 1976. **22**(7): p. 1024-1026.

383. Kendig, D.M. and Tarloff, J.B., *Inactivation of lactate dehydrogenase by several chemicals: implications for in vitro toxicology studies.* Toxicology in Vitro, 2007. **21**(1): p. 125-132.

384. Farrugia, W., Nicholls, L., and Rice, G., *Effect of bacterial endotoxin on the in vitro release of type II phospholipase-A2 and prostaglandin E2 from human placenta.* Journal of Endocrinology, 1999. **160**(2): p. 291-296.

385. Margaron, P., Sorrenti, R., and Levy, J.G., *Photodynamic therapy inhibits cell adhesion without altering integrin expression.* Biochimica *et* Biophysica Acta (BBA)-Molecular Cell Research, 1997. **1359**(3): p. 200-210.

386. D'Mello, S.R., Galli, C., Ciotti, T., *et al.*, *Induction of apoptosis in cerebellar granule neurons by low potassium: inhibition of death by insulin-like growth factor I and cAMP.* Proceedings of the National Academy of Sciences, 1993. **90**(23): p. 10989-10993.

387. Ma, Y., Hwang, R.F., Logsdon, C.D., *et al.*, *Dynamic mast cell–stromal cell interactions promote growth of pancreatic cancer.* Cancer Research, 2013. **73**(13): p. 3927-3937.

388. Hughes, J., Rees, S., Kalindjian, S., *et al.*, *Principles of early drug discovery.* British Journal of Pharmacology, 2011. **162**(6): p. 1239-1249.

389. Rixe, O. and Fojo, T., *Is cell death a critical end point for anticancer therapies or is cytostasis sufficient?* Clinical Cancer Research, 2007. **13**(24): p. 7280-7287.

390. Ichihara, E., Kiura, K., and Tanimoto, M., *Targeting angiogenesis in cancer therapy.* Acta Medicinae Okayama, 2011. **65**(6): p. 353-362.

391. Eichhorn, M., Kleespies, A., Angele, M., *et al., Angiogenesis in cancer: molecular mechanisms, clinical impact.* Langenbeck's Archives of Surgery, 2007. **392**(3): p. 371-379.

392. Dimova, I., Popivanov, G., and Djonov, V., *Angiogenesis in cancer-general pathways and their therapeutic implications.* Journal of Balkan Union of Oncology, 2014. **19**(1): p. 15-21.

393. Kerbel, R., Viloria-Petit, A., Klement, G., *et al., 'Accidental'anti-angiogenic drugs: anti-oncogene directed signal transduction inhibitors and conventional chemotherapeutic agents as examples.* European Journal of Cancer, 2000. **36**(10): p. 1248-1257.

394. Schirner, M., *Anti-angiogenic chemotherapeutic agents.* Cancer and Metastasis Reviews, 2000. **19**(1-2): p. 67-73.

395. Ribatti, D., *Chicken chorioallantoic membrane angiogenesis model.* Cardiovascular Development, 2012. **843**: p. 47-57.

396. Ribatti, D., Vacca, A., Roncali, L., *et al., The chick embryo chorioallantoic membrane as a model for in vivo research on anti-angiogenesis.* Current Pharmaceutical Biotechnology, 2000. **1**(1): p. 73-82.

397. Deryugina, E.I. and Quigley, J.P., *Chick embryo chorioallantoic membrane model systems to study and visualize human tumor cell metastasis.* Histochemistry and Cell Biology, 2008. **130**(6): p. 1119-1130.

398. Ribatti, D., *The chick embryo chorioallantoic membrane as a model for tumor biology.* Experimental Cell Research, 2014. **328**(2): p. 314-324.

399. Kunzi-Rapp, K., Genze, F., Küfer, R., *et al., Chorioallantoic membrane assay: vascularized 3-dimensional cell culture system for human prostate cancer cells as an animal substitute model.* The Journal of Urology, 2001. **166**(4): p. 1502-1507.

400. Lokman, N.A., Elder, A.S., Ricciardelli, C., *et al., Chick chorioallantoic membrane (CAM) assay as an in vivo model to study the effect of newly identified molecules on ovarian cancer invasion and metastasis.* International Journal of Molecular Sciences, 2012. **13**(8): p. 9959-9970.

401. Saif, J., Schwarz, T.M., Chau, D.Y., *et al., Combination of injectable multiple growth factor–releasing scaffolds and cell therapy as an advanced modality to enhance tissue neovascularization.* Arteriosclerosis, Thrombosis, and Vascular Biology, 2010. **30**(10): p. 1897-1904.

402. Sortino, M., Garibotto, F., Cechinel Filho, V., *et al.*, *Antifungal, cytotoxic and SAR studies of a series of N-alkyl, N-aryl and N-alkylphenyl-1, 4-pyrrolediones and related compounds.* Bioorganic and Medicinal Chemistry, 2011. **19**(9): p. 2823-2834.

403. Perry, C.J. and Parveen, Z., *The cyclisation of substituted phthalanilic acids in acetic acid solution. A kinetic study of substituted N-phenylphthalimide formation.* Journal of the Chemical Society, Perkin Transactions 2, 2001. (4): p. 512-521.

404. Bachmann, W.E. and Scott, L., *The reaction of anthracene with maleic and fumaric acid and their derivatives and with citraconic anhydride and mesaconic acid.* Journal of the American Chemical Society, 1948. **70**(4): p. 1458-1461.

405. Marsh, B.J., Adams, H., Barker, M.D., *et al.*, *Enantioselective catalytic desymmetrization of maleimides by temporary removal of an internal mirror plane and stereoablative over-reduction: synthesis of (R)-pyrrolam A.* Organic Letters, 2014. **16**(14): p. 3780-3783.

406. Doman, T.N., McGovern, S.L., Witherbee, B.J., *et al.*, *Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B.* Journal of Medicinal Chemistry, 2002. **45**(11): p. 2213-2221.

407. Shoichet, B.K., McGovern, S.L., Wei, B., *et al.*, *Lead discovery using molecular docking.* Current Opinion in Chemical Biology, 2002. **6**(4): p. 439-446.

408. Kola, I. and Landis, J., *Can the pharmaceutical industry reduce attrition rates?* Nature Reviews Drug discovery, 2004. **3**(8): p. 711-716.

409. Hughes, J.P., Rees, S., Kalindjian, S.B., *et al.*, *Principles of early drug discovery.* British Journal of Pharmacology, 2011. **162**(6): p. 1239-1249.

410. Dohle, D.S., Pasa, S.D., Gustmann, S., *et al.*, *Chick ex ovo culture and ex ovo CAM assay: how it really works.* Journal of Visualized Experiments: JoVE, 2009. (33): p. 1620-1625.

# Appendix

**Ramachandran data for conformers of the NMR ensemble 1OZO** (data generated in MOE)

| Model | Chain | Residue | Psi | Phi |
|---|---|---|---|---|
| **1** | A | S 21 | -66 | 131.3 |
| | A | S 24 | -20.9 | 179.6 |
| | A | T 25 | 173 | 95.2 |
| | A | K 49 | -57.7 | 179.8 |
| | A | D 50 | 44.7 | -7.1 |
| | A | D 62 | -33.7 | -29.3 |
| | A | D 66 | -38.8 | -159.6 |
| | A | A 67 | -8.5 | 153.9 |
| | A | T 92 | 71.5 | 128.7 |
| | A | G 93 | -63.3 | -22.8 |
| | B | S 19 | 20 | -63.6 |
| | B | S 21 | -57.4 | -167.5 |
| | B | S 24 | -15.3 | -163.8 |
| | B | T 25 | -163.4 | 83.2 |
| | B | K 49 | -75 | -170.5 |
| | B | K 51 | -0.2 | 42.6 |
| | B | D 62 | -14.7 | -43.8 |
| | B | D 66 | -5.7 | 52.7 |
| | B | A 67 | -4.3 | -177.4 |
| | B | H 86 | -80.4 | -19.5 |
| | B | F 89 | -119.6 | -153.5 |
| **2** | A | T 25 | -161.2 | 81.7 |
| | A | K 49 | -118.7 | 172.1 |
| | A | K 51 | -6.1 | 49.5 |
| | A | D 62 | -35.5 | -28.7 |
| | A | A 67 | -50.8 | 144.5 |
| | B | E 22 | -116.7 | 30.3 |
| | B | T 25 | -102.8 | 118.9 |
| | B | K 49 | 112.7 | 110.2 |
| | B | D 66 | -50.4 | -178.5 |
| | B | A 67 | -1.2 | 163.5 |
| | B | H 86 | -52.7 | -19.3 |
| | B | F 89 | -17.4 | 53 |
| **3** | A | S 21 | -69.6 | 115.6 |

| | | | | |
|---|---|---|---|---|
| | A | T 25 | -102.9 | 112.6 |
| | A | K 49 | -90.5 | -153 |
| | A | A 53 | -77 | 46 |
| | A | K 87 | 24.6 | 37.1 |
| | A | T 92 | -38.5 | -164.2 |
| | B | S 24 | -127.6 | -176.7 |
| | B | P 42 | 10 | -38.5 |
| | B | K 49 | -7.6 | -172.2 |
| | B | D 66 | -42.2 | -165.2 |
| | B | A 67 | -3.5 | 157.2 |
| | B | H 86 | -88.8 | -49.7 |
| | B | K 87 | -85.7 | 48 |
| **4** | A | S 21 | -71.2 | 97.4 |
| | A | K 49 | -100.3 | -170 |
| | A | A 67 | -51 | 145 |
| | B | S 16 | -91.2 | -72.5 |
| | B | R 17 | -82.2 | 17.3 |
| | B | Y 18 | -88.2 | -40.4 |
| | B | S 24 | -16.1 | 126.5 |
| | B | T 25 | 172.2 | 117.2 |
| | B | K 49 | -68.1 | -171.6 |
| | B | K 51 | -14.3 | 139.2 |
| | B | G 65 | -67.5 | -150.4 |
| | B | A 67 | -52 | 137.8 |
| | B | K 91 | -94.8 | -77.8 |
| **5** | A | S 21 | -66.5 | 113.1 |
| | A | D 66 | -59.6 | -162.1 |
| | A | A 67 | -51.4 | 146.2 |
| | A | T 92 | 65.5 | 112.3 |
| | B | S 24 | -69.3 | 145.2 |
| | B | T 25 | 52.4 | 179.7 |
| | B | K 49 | -108.9 | 173.9 |
| | B | G 65 | -72.8 | -157.5 |
| | B | D 66 | -50.6 | -160.2 |
| | B | A 67 | 3.9 | 158.3 |
| | B | H 86 | 11.2 | -54.1 |
| | B | T 92 | 95.1 | -45.9 |
| **6** | A | S 21 | -83.1 | 105.9 |
| | A | S 24 | 11 | -24.5 |
| | A | K 49 | -18.7 | -163.4 |
| | A | D 62 | -36.3 | -23.1 |

| | | | | |
|---|---|---|---|---|
| | A | A 67 | -50.8 | 143.4 |
| | A | K 87 | -74.8 | 42.1 |
| | A | F 89 | 18 | 41.1 |
| | B | E 22 | 17 | 37.8 |
| | B | S 24 | 30.8 | -63.8 |
| | B | K 49 | 105.2 | 118.9 |
| | B | K 51 | -29.4 | 134 |
| | B | A 67 | -43 | 173.2 |
| | B | E 90 | -24.1 | 32.3 |
| | B | T 92 | 60.4 | 144 |
| **7** | A | S 21 | -74.8 | 126.1 |
| | A | S 24 | 21.5 | 35.8 |
| | A | T 27 | 153.8 | -33.7 |
| | A | P 42 | -14.3 | -34.2 |
| | A | D 50 | 65.3 | 15.1 |
| | A | K 51 | -10.9 | 46.2 |
| | A | D 62 | -24.9 | -31.4 |
| | A | A 67 | -36.5 | 168.8 |
| | A | K 87 | -7.9 | 47.3 |
| | A | F 89 | -131 | -155.7 |
| | B | S 24 | 38.2 | -61.2 |
| | B | T 27 | 140.8 | 17.9 |
| | B | K 49 | 92.5 | 111.1 |
| | B | D 62 | -22.6 | -37.3 |
| | B | A 67 | 6.2 | 167.2 |
| | B | H 86 | -13.6 | -42.4 |
| | B | F 89 | -9.7 | 48.7 |
| | B | K 91 | 142.7 | 166.2 |
| | B | G 93 | -22.6 | 47.2 |
| **8** | A | S 21 | -22.1 | 130.8 |
| | A | S 24 | -118.3 | -164.3 |
| | A | T 27 | 126.8 | 17.1 |
| | A | K 49 | -85 | -177.1 |
| | A | D 62 | -35.3 | -27.6 |
| | A | D 62 | -35.3 | -27.6 |
| | A | A 67 | -30.9 | 162.5 |
| | B | S 19 | 17.6 | -61.5 |
| | B | E 22 | -117.4 | 101.8 |
| | B | T 25 | -12.1 | 124.9 |
| | B | T 27 | 132.4 | 21.9 |
| | B | Q 46 | -30.6 | -32.8 |

| | | | | |
|---|---|---|---|---|
| | B | K 49 | 138.3 | 74.6 |
| | B | D 52 | -106.9 | -73 |
| | B | D 62 | -35.3 | -27.3 |
| | B | A 67 | -44.7 | 142.7 |
| | B | K 91 | -85.6 | 17.6 |
| **9** | A | S 21 | -22.7 | 132.9 |
| | A | T 27 | 129 | 29.9 |
| | A | A 53 | -69.2 | 47.6 |
| | A | A 67 | -51.9 | 146.2 |
| | A | K 87 | 70.8 | -58 |
| | A | K 91 | -114.4 | -51.6 |
| | A | S 19 | 30.5 | -63.4 |
| | B | S 21 | -19.1 | -165.1 |
| | B | S 24 | -139.9 | -169.6 |
| | B | T 27 | 126 | 33.2 |
| | B | L 28 | 82.4 | -61.2 |
| | B | D 50 | -36.8 | 134.3 |
| | B | D 62 | -34.6 | -30.4 |
| | B | A 67 | -32.4 | 166.3 |
| **10** | A | S 21 | -21.7 | 121 |
| | A | T 27 | 136.4 | 30.3 |
| | A | D 50 | 13.4 | -51.7 |
| | A | D 66 | -91.4 | -157.6 |
| | A | Y 88 | 0.4 | 141.6 |
| | A | F 89 | 143.7 | 67.2 |
| | A | E 90 | -74.8 | 53.7 |
| | A | L 94 | -54.5 | 119.2 |
| | B | E 22 | -148.3 | 100.2 |
| | B | T 25 | -41.6 | 125.9 |
| | B | T 27 | 132.2 | 28.9 |
| | B | D 50 | 45.1 | -62.3 |
| | B | D 62 | -33.8 | -25.4 |
| | B | A 67 | -54.2 | 151 |
| | B | K 87 | -19.8 | 55 |
| | B | Y 88 | -2.4 | -177.3 |
| **11** | A | S 21 | -16.8 | 129.9 |
| | A | T 25 | -20.7 | 140.5 |
| | A | T 27 | 138.9 | 88.9 |
| | A | D 50 | 20.5 | -56.7 |
| | A | G 65 | -70.3 | -155.9 |
| | A | A 67 | -47.7 | 160.5 |

| | | | | |
|---|---|---|---|---|
| | A | Y 88 | -7.9 | -177.8 |
| | A | F 89 | 3.6 | 44.3 |
| | B | T 25 | -23.3 | 138.4 |
| | B | T 27 | 126.7 | 66.1 |
| | B | F 89 | 94.1 | 114.7 |
| | B | K 91 | -58.9 | -16.4 |
| | B | L 94 | -28.1 | 140.2 |
| **12** | A | S 21 | -16.8 | 129.9 |
| | A | T 25 | -20.7 | 140.5 |
| | A | T 27 | 138.9 | 88.9 |
| | A | D 50 | 20.5 | -56.7 |
| | A | G 65 | -70.3 | -155.9 |
| | A | A 67 | -47.7 | 160.5 |
| | A | Y 88 | -7.9 | -177.8 |
| | A | F 89 | 3.6 | 44.3 |
| | B | T 25 | -23.5 | 138.4 |
| | B | T 27 | 126.7 | 66.1 |
| | B | F 89 | 94.1 | 114.7 |
| | B | K 91 | -58.9 | -16.4 |
| | B | L 94 | -28.1 | 140.2 |
| **13** | A | S 21 | -68.7 | 113.7 |
| | A | S 24 | 13.8 | -3.1 |
| | A | K 49 | 93.8 | 89.8 |
| | A | K 51 | -25.8 | 139.1 |
| | A | D 62 | -38.2 | -25.9 |
| | A | G 65 | -69.6 | -152.2 |
| | A | A 67 | -51.3 | 151.3 |
| | A | F 89 | -2.8 | 38.3 |
| | A | L 94 | 77.8 | 136.7 |
| | B | E 22 | 37.1 | 29.3 |
| | B | S 24 | 42.9 | -32.4 |
| | B | T 27 | 142.1 | 15.8 |
| | B | K 49 | 120.8 | 123.2 |
| | B | D 62 | -12.6 | -44 |
| | B | D 66 | 0.7 | 50.6 |
| | B | A 67 | -11.4 | 176 |
| | B | L 94 | 8.8 | 48 |
| **14** | A | S 24 | 5.2 | -0.7 |
| | A | T 27 | 139 | -6.8 |
| | A | K 49 | 98 | 146.5 |
| | A | K51 | -12.9 | 59.9 |

| | | | | |
|---|---|---|---|---|
| | A | D 62 | -43.6 | -19.1 |
| | A | A 67 | -54.3 | 142 |
| | A | K 87 | -5.6 | -48.3 |
| | B | E 22 | -170.6 | 42.2 |
| | B | S 24 | 1.1 | 44.3 |
| | B | T 27 | 142.5 | 47.7 |
| | B | K 49 | -60.5 | -169.7 |
| | B | D 50 | 5.2 | 41.4 |
| | B | D 62 | -37.2 | -29 |
| | B | D 66 | -60.8 | -162.3 |
| | B | A 67 | -46.6 | 149.5 |
| | B | K 87 | -15.2 | 54.1 |
| **16** | A | S 21 | -55.2 | 125.2 |
| | A | D 50 | -76.8 | 53.9 |
| | A | K 51 | -26.7 | 172.6 |
| | A | D62 | -38 | -24.3 |
| | A | G 65 | -70.7 | -153.4 |
| | A | A 67 | -51.8 | 148.2 |
| | A | K 87 | 4.9 | 45.4 |
| | A | Y 88 | 114.3 | 55.5 |
| | B | S 24 | -152.9 | 134.2 |
| | B | Q 26 | 31.4 | 36.3 |
| | B | D 62 | -14.1 | -42.6 |
| | B | A 67 | 4.3 | -167.7 |
| | B | F 89 | 17.7 | -64.7 |
| | B | T 92 | -112.1 | -160.4 |

## Publications

1. **Camara**, Rossiter, Chau, Crnograc-Jurcevic and Kirton. Identifying Potential Inhibitors of the S100P-RAGE Interaction to Combat Pancreatic Cancer Progression (2015). *Journal of Computer Aided Molecular Design* (manuscript submitted).
2. **Ramatoulie Camara**, David Y. S. Chau, Deborah Ogbeni, Pryank Patel, Louise S. Mackenzie, Tatjana Crnogorac-Jurcevic, Stewart B. Kirton and Sharon Rossiter. *In silico*, synthetic and *in vitro* studies of potential inhibitors of S100P/RAGE interaction in pancreatic cancer (2015). *MedChemComm* (manuscript under revision)

## Posters

1. **R. Camara**, D. Ogbeni, L. Mackenzie, T. Crnogorac-Jurcevic, S. Rossiter, D. Chau, S. Kirton. Design, synthesis and evaluation of S100P-rage antagonists: implications for pancreatic cancer (2015). *Frontiers in Medicinal Chemistry*, Antwerp, Belgium
2. **R. Camara**, D. Ogbeni, L. Mackenzie, T. Crnogorac-Jurcevic, S. Rossiter, D. Chau, S. Kirton. Design, synthesis and evaluation of S100P-rage antagonists: implications for pancreatic cancer (2015). *2nd EFMC-Young Medicinal Chemist Symposium*, Antwerp, Belgium.
3. **R. Camara**, S. Rossiter, N. Mahmoud, T. Crnogorac-Jurcevic, S.B. Kirton. Design, synthesis and evaluation of inhibitors of the S100P/RAGE interaction as potential anti-pancreatic cancer agents (2014). *EFMC-ISM 23rd Symposium in Medicinal Chemistry*, Lisbon, Portugal.
4. **R. Camara**, S. Rossiter, N. Mahmoud, T. Crnogorac-Jurcevic, S.B. Kirton. Design, synthesis and evaluation of inhibitors of the S100P/RAGE interaction as potential anti-pancreatic cancer agents (2014). *1st EFMC-Young Medicinal Chemist Symposium*, Lisbon, Portugal.
5. **R. Camara**, S. Rossiter, N. Mahmoud, T. Crnogorac-Jurcevic, S.B. Kirton. S100P/RAGE inhibitors as potential therapeutics for pancreatic cancer. *UK PharmSci*, University of Hertfordshire, Hatfield.

6. **Ramatoulie Camara**, Sharon Rossiter, Nasir Mahmoud, Tatjana Crnogorac-Jurcevic, Stewart B. Kirton. Identification of novel inhibitors of S100P – a protein implicated in pancreatic cancer – using computational and synthetic methods (2014). *University of Hertfordshire School of Life and Medical Science Research Conference*, University of Hertfordshire, Hatfield.
7. **R. Camara** *et al*. In search of S100P/RAGE inhibitors as therapeutic agents in pancreatic cancer using structure-based drug design (SBDD) (2013). *7th RSC BMC Postgraduate Symposium*, University of Cambridge, Cambridge.
8. **Ramatoulie Camara**, Sharon Rossiter, Nasir Mahmoud, Tatjana Crnogorac-Jurcevicb, Stewart B. Kirton. Structure-Based Drug Discovery (SBDD) using an NMR ensemble: the case of S100P (2013). *Young Modeller's Forum*, SOAS, University of London, London.
9. **Ramatoulie Camara**, Sharon Rossiter, Vishal Kholi and Stewart B. Kirton. Rational design of S100P/RAGE inhibitors: in search of a therapeutic agent in the fight against pancreatic cancer (2013). *17th RSC/SCI Medicinal Chemistry Symposium*, University of Cambridge, Cambridge.
10. **Ramatoulie Camara**, Sharon Rossiter, Vishal Kholi and Stewart B. Kirton. *In Silico* Structure-Based Drug Design (SBDD) on S100P: Identification of a lead compound towards a therapeutic intervention in pancreatic cancer (2013). *6th Joint Sheffield Conference on Chemoinformatics*, University of Sheffield, Sheffield.
11. **Ramatoulie Camara**, Sharon Rossiter, Vishal Kholi and Stewart B. Kirton. Protein informatics on S100P for structure-based drug design (2012). *12th Symposium of the European Calcium Society*, Toulouse, France.

## Patent

Inventors: Camara, Kirton and Rossiter. Compounds and their uses (2015). Patent number: GB1510123.1