LEABHARLANN CHOLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH | TRINITY COLLEGE LIBRARY DUBLIN
Ollscoil Átha Cliath | The University of Dublin

Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Synthesis and Biological Evaluation of Diaromatic Guanidinium-like Derivatives as Protein Kinase Inhibitors for the Treatment of Cancer

A thesis presented to the University of Dublin

for the degree of Doctor of Philosophy

by

Elena Diez Cecilia, BSc

Under the supervision of Prof. Isabel Rozas

School of Chemistry

Trinity College Dublin

January 2013

TRINITY COLLEGE
2 4 MAY 2013
LIBRARY DUBLIN

Thesis 10046

Declaration

This work comprises a doctoral thesis submitted for the consideration of Trinity College Dublin.

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work, with due acknowledgement and reference given to the work of others, where appropriate.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

.....

Elena Diez Cecilia

January 2013

Trinity College Dub

Acknowledgements

This thesis would not have been possible without the continuous support and friendship of many people, who made my time in Trinity College an unforgettable experience.

First and foremost, I offer my sincere thanks to my supervisor Prof. Isabel Rozas, who always supported me, encouraged me and guided me throughout my Ph.D. Thanks for believing in me and giving me the opportunity to work with you. Also, I would like to thank you for always being there to answer all my questions, for your friendship, for being a very approachable "boss" and for your positive attitude.

I would like to extend my deep gratitude to my co-supervisor Prof. Daniela Zisterer for her invaluable assistance in all biochemistry issues. Without her knowledge, guidance and advice I would not have been able to learn so much about one of my passions, biochemistry.

Thanks to the past and present members of the Rozas group, which one by one have been becoming doctors, Dr. Alessandra Cordeiro Machado (my kind Brazilian-Spanish friend), Dr. Fernando Blanco (my helpful and best Spanish friend), Dr. Patrick Nagle (my welcoming friend), Dr. Daniel O'Donovan (my multi-faceted and good friend), Dr. Amila Kahvedžić (mi chica and almost my sister), Dr. Aoife Flood (my little dancer and great friend), Dr. Caitriona McKeever (my big hearted and generous crazy friend), Dr. Brendan Kelly (my unconditional supporter), Julian (my synthetic Irish-American friend), Paddy (my charming party animal friend) and Michela (my fun and hard-working friend). It has often been said that pleasure and work do not go together, but working with you guys has proven the opposite. Thanks a million for all your help, support and friendship, and for the good times and laughs inside and outside the lab. You were one of my biggest supports during my time in Ireland, and made it less difficult being far from home. You became my second family, the Rozas family!

I cannot forget my colleges in biochemistry, the past and present members of the DZ group Dr. Seema Nathwani, Dr. Lisa Greene, Dr. Sandra Bright and especially Jenny for all her patience, encouragement, advice and support during my time in cell culture. I learnt so much from you! Thanks as well for your friendship and the good times with "the girls" Jade, Emma, Claire and Kasia.

I would like to show my appreciation to Dr. John O'Brien and Dr. Manuel Ruether for their help and expertise in NMR work and dealing with my "nasty and amorphous samples". Thanks also go to Dr. Martin Feeney for his help with mass spectrometry and to Dr. Tom McCabe for the X-ray crystallographic analysis. Thanks to all the staff in the chemistry office, in particular Prof. Grayson, the staff at the Cocker and SNIAM labs, the biochemistry facilities such as Flow Cytometry and confocal. Thanks for making TCD a great place to work. Thanks also to The Irish Research Council for funding this project.

I would especially like to acknowledge the fantastic work of my collaborators in Queens University (Belfast, U.K) – in particular Robert Carson, the University of Cardiff (U.K) and the Molecular Design Group in TCD – in particular Fernando Blanco. Your contribution to this work was impeccable and much appreciated. You really helped me to develop my project further.

I would like to thank the staff at Universidad de Alcala de Henares (Spain) for my initial scientific formation and for awakening my passion in medicinal chemistry.

I cannot thank my family enough, my parents and my brother have been a constant source of support, encouragement, kindness, happiness and love throughout these years far from home. Thanks for always teaching me that hard work is the key to success and that I should focus on the good things in life and enjoy them as much as I can. Even though I was not home that much during the years of my Ph.D, you were always present in my heart. I would like to dedicate this work to you as a small token of my appreciation and gratitude for all the things you have done for me. Os quiero.

Thanks as well to my friends at home for always being there for me despite the distance. Thanks for charging my batteries with energy and fun after every visit and for being like we met every day, despite having months apart. Os quiero.

Last but not least, I would like to thank my colleague, my friend, my walking English dictionary, my proof reader, my cheerleader, my soul mate, my boyfriend and my love; all these in one person, you Brendan. I would need a thousand pages to express my gratitude for all you have helped me with and to thank you enough for all you have done for me. You and your family have become my main pillar during these years that helped me to overcome any difficulties and adversities I encountered, and have also shared all my achievements and happiness with me. Muchas gracias.

Summary

The aim of a chemotherapeutic drug is to specifically target a biomolecule involved in the development of cancer whilst not affecting healthy tissues. Unfortunately, this is a difficult task due to the complexity of the biochemical processes and regulatory mechanisms entailed. Targeting systems which are only up-regulated or mutated in cancer cells can help to achieve selectivity. A quintessential example of this approach is inhibiting the protein kinase family, which are enzymes involved in the pathophysiology of cancer. Their activity is regulated in normal cells; however, they are often over-expressed or mutated in cancer cells, leading to malfunctioning of signaling networks.

There are several protein kinase inhibitors (PKI) on the market for the treatment of cancer, for example sorafenib. Considering the unknown target of previous cytotoxic compounds in Rozas' group and their similarities with sorafenib, five new families of guanidinium-based derivatives related to sorafenib were synthesised. During these syntheses a novel simultaneous reduction of nitro and carbonyl groups was discovered.

Several different biochemical assays were performed to investigate the effects of these new compounds on a number of cell lines and biological systems, as potential anticancer drugs. Thus, viability assays, apoptosis evaluation and different kinase inhibition tests were performed to assess their cytotoxicity and pro-apoptotic ability as well as to evaluate their role as potential protein kinase inhibitors.

Firstly, the influence on cell viability of these five families of compounds was tested in a wide range of cancer cell lines such as promyelocytic leukaemia HL-60, liver carcinoma HUH-7, colorectal cancer RKO and breast cancer MCF-7 and tamoxifen-resistant MCF-7 cell lines. Secondly, microscopy and flow cytometry studies were carried out to investigate the apoptosis-inducting activity of these derivatives. Finally, to elucidate the molecular target of our compounds, and considering the targets of sorafenib and the biological related pathways, several kinase tests were performed in the RAF/MEK/ERK pathway, p38 MAPK, VEGFR, CK-1δ, GSK-3, AKT and STAT-3.

Computational docking studies were also performed to analyse the binding interactions of our compounds with different possible targets (RAF and MEK) to rationalise our biological results and to allow for improvements to be made in the anticancer activity of future compounds.

Abbreviations

AGC:	protein kinase group containing PKA, G and C
AKT:	AKR mouse thymoma kinase
ALL:	Acute Lymphocytic Leukaemia
AML:	Acute Myeloid Leukaemia
ATP:	Adenosine Triphosphate
BCL-2:	B-cell Lymphoma gene 2
BCR-ABL:	Break point Cluster Region Abelson
Boc:	*tert*-Butoxycarbonyl
Boc_2O:	Di-*tert*-Butyldicarbonate
B-RAF:	B-Raf proto-oncogene serine/threonine kinase, also known as v-raf murine sarcoma viral oncogene homolog B1
Bs:	Broad singlet
$CDCl_3$:	Deuterated Cloroform
CDK:	Cyclin-Dependent Kinase
CK-1δ:	Casein Kinase 1 δ
CMGC:	Protein kinase group containing CDK, MAPK, GSK3, CLK families
CML:	Chronic Myeloid Leukaemia
C-RAF:	RAF-1, v-raf1 murine leukaemia viral oncogene homolog 1
DCM:	Dichloromethane
ddH_2O:	Distilled water
DGF:	Conserved Asp-Phe-Gly motif in the ATP-binding cleft
DME:	1,2-Dimethoxyethane

DMF: Dimethylformamide

DMSO: Dimethyl sulfoxide

EGFR: Epidermal Growth Factor Receptor

ERK: Extracellular signal-regulated kinase

ES: Electrospray

EI: Electroimpact

Et_3N: Triethylamine

EtOAc: Ethyl Acetate

FACs: Flow Cytrometry

FLT-3: Fms-related tyrosine kinase 3

GIST: Gastrointestinal Stromal Tumours

GSK-3: Glycogen Synthase Kinase 3

GTP: Guanosine Tri-Phosphate

HB: Hydrogen Bond

HCC: Hepatocellular Carcinoma

HRMS: High-Resolution Mass Spectrometry

IC_{50}: Half Maximal Inhibitory Concentration

IPA: Isopropanol or 2-Propanol

IR: Infrared

K_d: Quantitative dissociation constant of drug-protein complex

K_M^{ATP}: Michaelis constant that measure the ATP binding affinity of a kinase

K_{off}: Dissociation constant of drug-protein complex

K_{on}: Formation constant of drug-protein complex

MAPK:	Mitogen-Activated Protein Kinase
MEK:	Mitogen-Activated Protein Kinase/Extracellular Signal-Regulated Kinase
Mp:	Melting Point
mTOR:	Mammalian Targeted of Rapamycin
NaH:	Sodium hydride
NMR:	Nuclear Magnetic Resonance
NRTK:	Non-Receptor Tyrosine Kinases
NSCLC:	Non Small Cell Lung Cancer
P38 MAPK:	p38 Mitogen-Activated Protein Kinase
P53:	Protein 53 or Tumour Protein 53
PARP:	Poly (ADP-Ribose) Polymerase
PC:	Pancreatic Cancer
Pd/C:	Palladium on Carbon
PI3K:	Phosphatidylinositol 3-kinases
PK:	Protein Kinase
PKI:	Protein Kinase Inhibitor
RAF-1:	or C-RAF; Rapidly Accelerated Fibrosarcoma-1
RAS:	Renin-Angiotensin System
RCC:	Renal Cell Carcinoma
RTK:	Receptor Tyrosine Kinase
Ser/Thr:	Serine/Threonine
SAR:	Structure Activity Relationship
STAT-3:	Signal Transducers and Activator of Transcription 3

TFA:	Trifluoroacetic Acid
TFAA:	Tifluoroacetic Anhydride
TK:	Tyrosine Kinase
TKI:	Tyrosine Kinase Inhibitor
TLC:	Thin Layer Chromatography
tR:	Retention Time
UV:	Ultraviolet
VEGF:	Vascular Endothelial Growth Factor
VEGFR:	Vascular Endothelial Growth Factor Receptor
WT:	Wild Type

Table of Contents

Chapter 1 – Introduction **1**

1.1. Origin of Cancer 1

1.2. Apoptosis 3

1.3. Chemotherapeutic Agents 5

1.3.1. DNA targeting 5

1.3.2. Antitubulin Agents 8

1.3.3. Hormonal Therapies 9

1.3.4. Tumour-Targeting Strategies 10

1.3.5. Biological Agents 11

1.3.6. Molecularly Targeted Agents 12

1.4. Protein Kinases and their Relevance in Cancer 17

1.4.1. The Kinome 18

1.4.2. Tyrosine Kinase Group 20

1.4.2.1. Receptor Tyrosine Kinases 20

1.4.2.2. Non Receptor Tyrosine Kinases 24

1.4.3. Serine Threonine Kinases 25

1.4.3.1. Mitogen Activated Protein Kinase (MAPK) Pathway 26

1.4.3.2. Casein Kinase-1δ and Glycogen Synthase-3 29

1.5. Protein Kinase Inhibitors (PKIs) 31

1.5.1. Types of PKIs 31

1.5.2. Therapeutic Strategies of PKIs 36

1.5.3. Influence of PKIs in Medicinal Chemistry 38

1.5.4. Targeting the Active *versus* the Inactive Form 43

1.5.5. ATP-competitive *versus* Non-Competitive Inhibitors 43

1.5.6. Specific *versus* Multi-Targeted Inhibitors 44

1.5.7. MAPK inhibitors 48

1.5.8. Tyrosine Kinase Inhibitors 48

1.6. Sorafenib (Nexevar®, BAY-43-9006) 49

1.6.1. Discovery and Development of Sorafenib 49

1.6.2. Mode of Action 51

1.6.3. Chemical Interactions between Sorafenib and B-RAF 53

1.6.4. Original Synthesis of Sorafenib 54

1.6.5. Limitations of Sorafenib 55
1.7. Future and Challenges of Kinase Drug Discovery 56
1.8. Previous Work Leading Up to the Present Research 60
1.9. References 63

Chapter 2 – Objectives **69**
2.1. Chemical Objectives 70
2.2. Biological Objectives 73
2.3. Computational Objectives 75
2.4. References 76

Chapter 3 – Chemical Synthesis **77**
3.1. Introduction 77
3.2. Synthesis of 3,4'-Diphenyl Amines 80
3.2.1. Synthesis of 3-(4-aminobenzoyl)aniline 80
3.2.2. Synthesis of 3-(4-aminobenzyl)aniline 82
3.2.3. Synthesis of 3-(4-aminoanilino)aniline 89
3.3. Synthesis of 4,4'- and 3,4'-*Bis*-Guanidine-like Derivatives – Families 1 and 2 . 101
3.3.1. General Aspects of Families 1 and 2 101
3.3.2. Mercury(II) Chloride Guanidylation and 2-Iminoimidazolidylation Method 103
3.3.3. Alternative Guanidylation-like Methods 105
3.3.4. Synthesis of 4,4'- and 3,4'-*Bis*-Boc-Protected Guanidine-like Derivatives 111
3.3.5. Synthesis of the Hydrochloride Salts of the 4,4'- and 3,4'-*bis*-Guanidine-like Derivatives 119
3.3.6. Conformational Aspects of *Bis*-Boc-protected and Hydrochloride Salts of Aryl Guanidine Derivatives 122
3.4. Synthesis of Phenyl Guanidine Derivatives – Families 3A, 3B, 4A and 4B 125
3.4.1. General Aspects of Families 3A, 3B, 4A and 4B 125
3.4.2. Literature Methods for the Preparation of *N,N'*-Disubstituted Guanidines 127
3.4.3. Synthesis of *N*-(*tert*-butoxycarbonyl)-*N'*-phenylthiourea and *N*-(*tert*-butoxycarbonyl)-*N'*-(4-chloro-3-trifluoromethylphenyl) thiourea 132

3.4.4. Synthesis of Asymmetrically Substituted *para*-Phenyl Guanidine Derivatives (Family 3A and 4A) 133
3.4.5. Synthesis of Asymmetrically Substituted *meta*-Phenyl Guanidine Derivatives (Families 3B and 4B) 137
3.4.6. Synthesis of the Hydrochloride Salts of the Asymmetrically Substituted Phenyl Guanidine Derivatives from Families 3A, 3B, 4A and 4B 139
3.4.7. Isomerism in *N*-(*tert*-Butoxycarbonyl)-*N'*-phenyl Guanidine Compounds and their Hydrochloride Salts 141
3.5. Synthesis of Acetamide Derivatives – Family 5 152
3.5.1. General Aspects of Family 5 152
3.5.2. Synthesis of the Boc-Protected Acetamide Derivatives 153
3.5.3. Synthesis of the Hydrochloride Salts of the Acetamide Derivatives 155
3.6. Conclusions 157
3.7. References 159

Chapter 4 – Biological Evaluation of Potential Protein Kinase Inhibitors 162
4.1. Introduction 162
4.2. Cytotoxicity Evaluation of the Novel Compounds 164
4.2.1. Evaluation of the Cytotoxicity of Compounds from Families 1-5 in the HL-60 Cell Line 168
4.2.2. Evaluation of the Cytotoxicity of Selected Compounds in the HUH-7 Cell Line 173
4.2.3. Evaluation of the Cytotoxicity of Selected Compounds in the RKO Cell Line 176
4.2.4. Evaluation of the Cytotoxicity of Selected Compounds in the MCF-7 Cell Line 181
4.3. Apoptosis Evaluation of the Novel Compounds 187
4.3.1. Cell Morphology Studies 187
4.3.1.1. Morphological Studies of the HL-60 Cell Line 189
4.3.1.2. Morphological Studies of the HUH-7 Cell Line 193
4.3.1.3. Morphological Studies of the MCF-7 and RKO Cell Lines 196
4.3.2. Flow Cytometry 197
4.3.2.1. Flow Cytometry Evaluation of 73a and Sorafenib in the HL-60 Cell Line 200

4.3.3. Role of Poly(ADP-ribose) Polymerase (PARP) Cleavage in Apoptosis.... 207
4.4. Protein Kinase Evaluation 210
4.4.1. Casein Kinase 1δ and Glycogen Synthase Kinase-3 Studies 210
4.4.2. RAF/MEK/ERK pathway, p38 MAPK and VEGFR Study 211
4.4.3. LanthaScreen® Europium (Eu) Kinase Binding Assay for B-RAF 213
4.4.4. Effect of Selected Compounds in the Expression Levels of Various Signalling Proteins in RKO Cells..... 218
4.5. Conclusions..... 221
4.6. References..... 224

Chapter 5 – Computational Modelling 227
5.1. Introduction..... 227
5.2. Comparison of B-RAF, C-RAF and MEK-1 228
5.3. Docking of Synthesised Compounds to Systems of Interest 233
5.4. Docking and Minimisation of Synthesised Compounds in MEK-1 (3PP1)..... 235
5.4.1. Protein Model Preparation 235
5.4.2. Docking of Crystal Ligand and Novel Compounds into MEK-1 Minimised Structure..... 235
5.4.3. MMPBSA Analysis of Ligand-Protein Complexes..... 237
5.5. Conclusions..... 247
5.6. References..... 249

Chapter 6 – Conclusions and Future Work 250
6.1. Conclusions..... 250
6.1.1. Synthesis 250
6.1.2. Biochemistry..... 252
6.1.3. Computational Chemistry 254
6.2. Future Work..... 255
6.2.1. Future Biological Application of our Compounds 255
6.2.2. Future Protein Kinase Inhibitors..... 256
6.3. References..... 259

Chapter 7 – Experimental **260**
7.1. Synthetic Chemistry 260
7.1.1. Material and Methods 260
7.1.2. Purity Assessment of Hydrochloride Salts 260
7.1.3. General Procedures 261
7.1.4. Synthesis and Characterisation 264
7.2. Biochemical Procedures 316
7.2.1. General Procedures 316
7.2.2. Growth and Maintenance of the Cell Lines 317
7.2.3. Cell Viability Assays 319
7.2.4. Cell Morphology Studies 321
7.2.5. Flow Cytometry (Determination of DNA content) 322
7.2.6. Kinase Studies 323
7.2.7. LanthaScreen® Eu Kinase Binding Assay for B-RAF 326
7.2.8. Western Blot Preparation 327
7.3. Computational Chemistry Procedures 329
7.3.1. Proton Affinity and Molecular Electrostatic Potential 329
7.3.2. Isomerism Study Protocol 329
7.3.3. Molecular Docking Protocol 330
7.3.4. Minimisation of MEK-1 Structure in AMBER 12.0 330
7.3.5. Minimisation of Docking Poses and MMPBSA Analysis 331
7.4. References 332

Appendix ..... **i**

Chapter 1 – Introduction

1.1. Origin of Cancer

Cancer is a term used to describe diseases in which control of cell division during the cell proliferation process is lost, leading to abnormal and uncontrolled cell division.[1] This set of diseases is one of the most pressing health challenges of the 21st century. The International Agency for Research on Cancer (IARC) states that, by the year 2030, almost 21.4 million new cases will be diagnosed annually and that cancer will annually kill more than 13.2 million people.[2]

Cancer may spread from the original site to other adjacent tissues, or even to other parts of the body, through the blood (invasion) and lymph (metastasis) systems. Cancer cells divide much faster than normal cells, thereby increasing the risk of spreading.[3] These cells have subtle physical differences compared to normal cells; for example their cellular walls are less rigid and their nuclei are larger. Furthermore, cancer cells are not dependent on normal growth factors or inhibitory signalling and they evade apoptosis (programmed cell death). Additionally, cancer cells induce angiogenesis, the growth of new blood vessels, to form tumours and expand their generation, whereas, normal cells depend on blood vessels to supply oxygen and nutrients, maintaining a constant rate of division.[1]

Resistance to apoptosis can be acquired by cancer cells through a variety of strategies. The most commonly occurring loss of a pro-apoptotic regulator through mutation involves the p53 tumour suppressor gene. The resulting functional inactivation of the p53 protein is seen in greater than 50% of human cancers and results in the removal of a key component of the DNA damage sensor that can induce the effector cascade. Signals evoke by other abnormalities including hypoxia and oncogene hyper-expression, are also funneled in part via p53 to the apoptotic machinery. Disabling the p53 protein, enable the cells to continue multiplying for additional generations.[4]

Although cancers of different origins have distinct features, they all share in common the fact that they stem from deoxyribonucleic acid (DNA) damage, caused endogenously or exogenously. This damage can be caused by several factors including ultraviolet radiation, cellular metabolism, viral infection, replication errors and/or chemical exposure. In normal

cells, DNA suffers thousands of lesions per day, which, if left unattended, can lead to mutations or cell death. If these damaged cells continue proliferating and do not die by apoptosis, the accumulation of mutations in cells over time can lead to cancer, in which uncontrolled growth of cancer cells generates aggregates called tumours. This process is illustrated in the Figure 1.1.

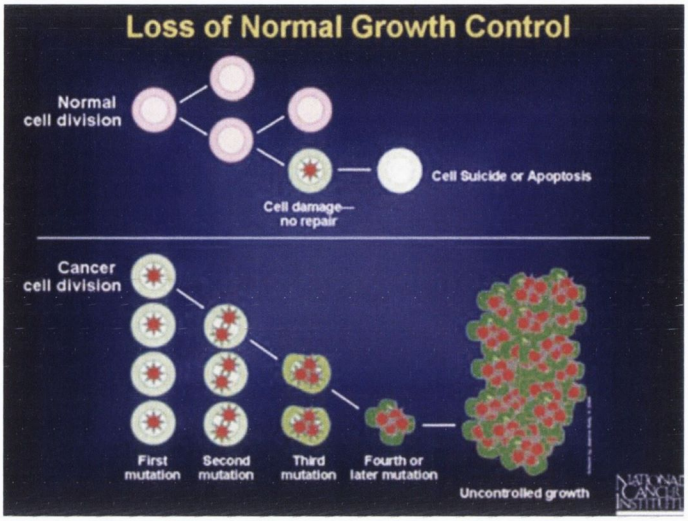


Figure 1.1. Cell division process.[5]

1.2. Apoptosis

Apoptosis plays an important role in development as well as in normal cell turnover. In addition, cellular apoptosis plays a key role in the treatment of cancer as well as other diseases, as it stops the uncontrolled proliferation of cancer cells. The term apoptosis is often used synonymously with programmed cell death (from a Greek word meaning "falling off," as leaves from a tree), implying that death results from the regulated activation of a pre-existing death programme that is encoded in the genome.[6] Further studies suggest that apoptosis consists of three successive stages: (i) commitment to death triggered by extracellular or intracellular signals, (ii) execution of cell killing by activation of intracellular proteases, (iii) removal of dead cells through engulfment of cell corpses by other cells, followed by degradation of these dead cells within lysosomes of phagocytic cells.[7]

Molecular biology studies of apoptosis have rapidly unveiled the mechanisms of apoptotic machinery over the years. The process of apoptosis is controlled by a diverse range of cell signals, which may originate either extracellularly (extrinsic inducers) or intracellularly (intrinsic inducers) (Figure 1.2). In addition, anticancer drugs can activate lipid-dependent signalling pathways that result in decreased apoptosis threshold, or modulate other cytoprotective pathways such as nuclear factor-κB (NF-κB), heat shock proteins, and cell cycle regulatory pathways. On the one hand, regulation of apoptosis can be controlled by cessation of extracellular signals which negatively regulate apoptosis, because they are required for the survival of cells such as cytokines. It can also be controlled through upregulated expression of pro-apoptotic receptors on cancer cells, whose subsequent interaction with their ligands activates apoptotic signalling pathways. These receptors include Fas and the tumor-necrosis-factor (TNF)-related apoptosis-inducing ligand (TRAIL) receptors.[8] Other extracellular signals include toxins, hormones, nitric oxide and growth factors (eg. NF-κB), that must either cross the plasma membrane or be transduced in order to affect a response.[9,10] Study of the pathways which regulate cell death from outside the cell is complicated by the fact that the same pathways are often used to transmit signals required for growth and differentiation.[11]

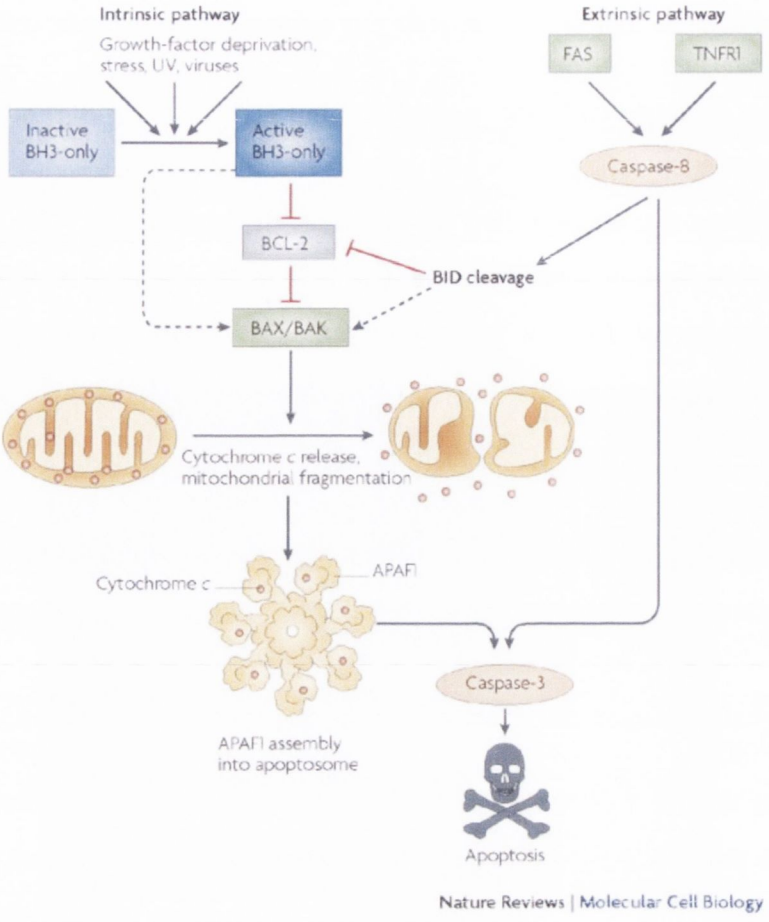


Figure 1.2. Intrinsic and extrinsic apoptosis pathways.[12]

On the other hand, a cell initiates intracellular apoptotic signalling in response to diverse stressful stimuli and metabolic disturbances by triggering apoptotic programs. Examples include cell death caused by drugs, heat, radiation, hypoxia and viral infections.[8] Generally, agents that damage DNA, such as ionizing radiation and certain xenobiotics, lead to activation of p53-mediated mechanisms which commit cells to apoptosis. These mechanisms include the transcriptional upregulation of pro-apoptotic proteins (*e.g.* the BCL-2 family members Bid and Bax)[13] which release mitochondrial pro-apoptotic proteins such as cytochrome C, which go on to activate a series of proteases called caspases. Other stresses induce increased activity of stress-activated protein kinases (SAPKs), which result ultimately in apoptotic commitment.[14]

Although there are many mechanisms of apoptotic signal transduction, it seems certain that they converge on common death programs involving the sequential activation of a series of proteases known as caspases, particularly initiator caspases-8 and -9, which cleave and

activate downstream caspases. These caspases cleave and inactivate many structural and repair components of the cell resulting in its demise.

1.3. Chemotherapeutic Agents

The aim of a chemotherapeutic drug is to specifically target a biomolecule involved in cancer cell replication whilst avoiding affecting any healthy tissues. Unfortunately, this goal is complicated by the complexity of the biochemical processes and regulatory mechanisms entailed. Notwithstanding recent advances in anti-cancer research, the toxicity of modern chemotherapeutic agents and the side effects they induce mean that there is still very much a need for new anti-cancer therapies. Nowadays, there is a continuous search for novel anti-cancer targets and non-classical chemotherapeutic drugs, not indiscriminately cytotoxic, are continuously being discovered. These may, for example, target specific enzymes or inhibit pathways, such as those that trigger apoptosis, which subsequently lead to the death of the cancer cell. The list of these agents is extensive and they can be classified according to their specific targets and mode of action.

In general there are six major groups of anticancer agents: DNA-interactive, antitubulin, molecularly targeted, hormonal, tumour-targeting and biological.[15] This project is based on the study of molecules that target signalling pathways, or more specifically, protein kinases. A brief explanation of the different types of chemotherapeutic agents will follow, with a more detailed analysis of the targets in question.

1.3.1. DNA-Interactive Agents

DNA-interactive agents exert their effect by interacting with DNA through specific modes of action. The structure of the helices of DNA is thereby disrupted and processes such as replication and repair are inhibited through a variety of mechanisms. Some of such drugs work by interacting with the DNA minor groove intercalating between the pairs of DNA bases or intercalating between the base pairs, whereas others alkylate these bases or cross-link the DNA strands. Further modes of action of these agents include the inhibition of topoisomerases (regulators of the winding of DNA) or DNA cleavage.[15]

Typically, intercalators are aromatic, flat structures and have a rigid conformation. These drugs bind according to the 'nearest neighbour exclusion principle' which states that such a molecule can bind to a maximum loading of one ligand per pair of bases.[16] The most common families of intercalators are anthracyclines, anthracenes and phenoxazines.

On the other hand, minor groove binders (Figure 1.3) target one of the two indentations in the DNA double helix (the major and the minor grooves) which result from the asymmetric attachment of the phosphodiester backbone to the base pairs.

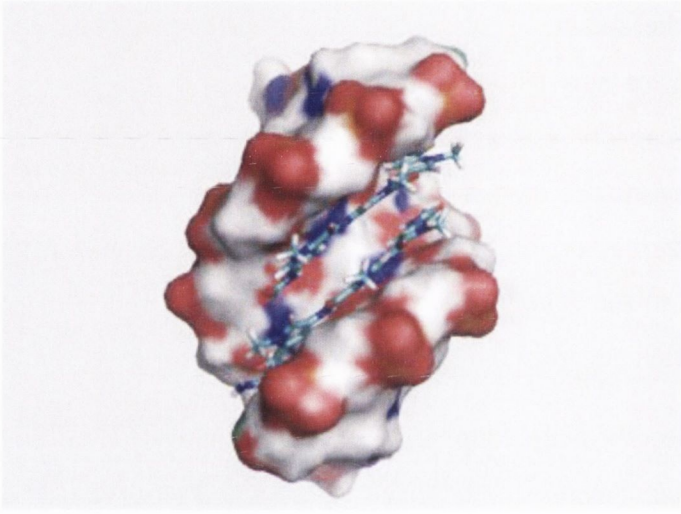


Figure 1.3. A minor groove binder in the double helix of DNA.[17]

Classically, they are concave in shape, flexible, cationic in nature and must possess interactive features, such as hydrogen bond (HB) donors and acceptors, to bind to DNA.[18] Examples of minor grooves binders include furamidine (Figure 1.4, a), distamycin and some *bis*-2-aminoimidazolinium and *bis*-guanidinium derivatives synthesised by the Rozas group in TCD (Figure 1.4, b).[19]

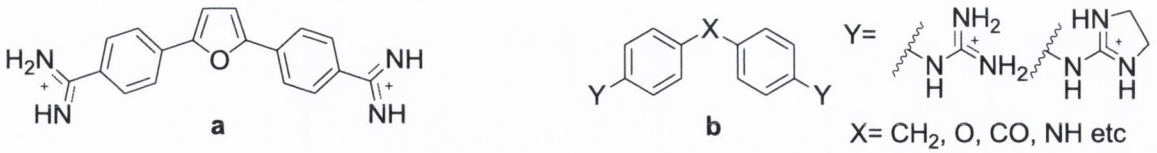


Figure 1.4. Structure of (a) furamidine and (b) *bis*-2-aminoimidazolinium or *bis*-guanidinium DNA minor groove binders with potent *in vitro* antitrypanosomal and antiplasmodial activity, synthesised in the Rozas group.

Moreover, alkylating agents interact with the major groove of DNA, methylating the guanine bases, predominantly at the N^7-position and, to a lesser degree, the O^6-position.[20,21] Leading alkylating drugs are dicarbazine, temozolomide and procarbazine. Similarly, cross-linking agents can function by alkylating two nucleophilic functional groups on the DNA bases of the two complementary strands of the duplex (interstrand) or within a strand (intrastrand).[22] Leading agents that are used clinically are nitrogen mustards, aziridines, epoxides, platinium complexes, carbinolamines and mitomycin-c (Figure 1.5).

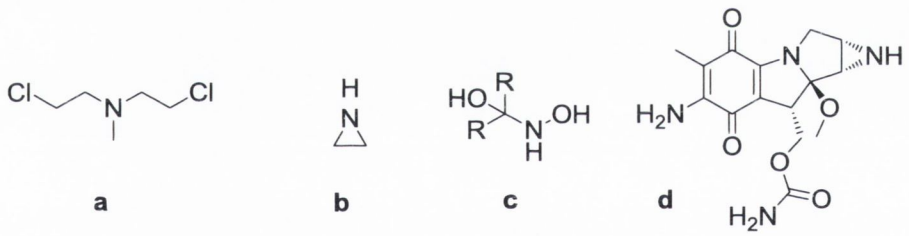


Figure 1.5. Cross-linking chemoterapeutic agents clinically used: (a) nitrogen mustard, (b) aziridine, (c) carbinolamines, and (d) mitomycin-c.

Another group of agents that interact with DNA is the topoisomerase I and II inhibitors. These enzymes are responsible of regulating the overwinding or underwinding of DNA. Hence, these inhibitors work by trapping the DNA-enzyme complex so that processes such as transcription and translation cannot occur and subsequent cell death results.[22] Examples of topoisomerase I inhibitors are topotecan and irinotecan and topoisomerase II inhibitors include etoposide, ellipticene and amsacrine.[15]

The final group of DNA binding agents is that of the DNA cleaving agents. They bind by a selective radical mechanism to the sugar moieties of both DNA strands, resulting in a double-strand break.[23] The mechanism of these radiomimetic drugs is based on oxidation of the deoxyribose ring in DNA, initiated by hydrogen abstraction at the C-1', C-4' and C-5' positions. The best known example of a drug that acts in this manner is bleomycin.[23]

1.3.2. Antitubulin Agents

Another class of chemotherapeutic drugs that have shown anti-tumour characteristics is that of the antitubulin agents which bind to microtubules – polymers of tubulin essential in the formation of the mitotic spindle and microtubule assembly – during the mitotic step of the cell division process.[24] Agents which interrupt this procedure block mitosis and lead to cell death. Vinca alkaloids, identified over 50 years ago, and taxanes, first isolated almost 40 years ago, are the most important families of chemotherapeutic drugs currently administered for a large range of solid tumours and haematological malignancies (Figure 1.6).[25-26]

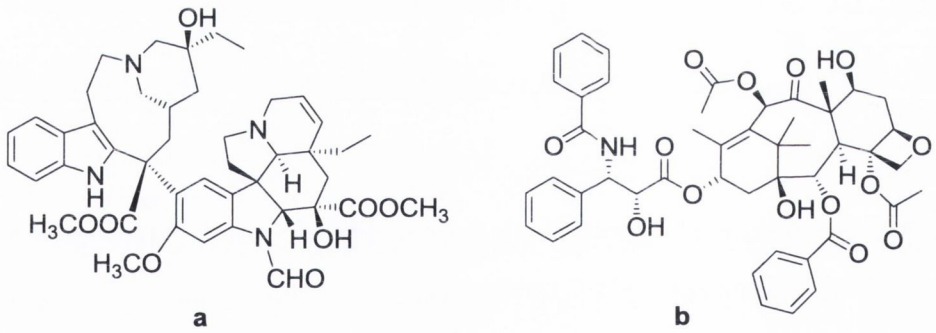


Figure 1.6. (a) Vinblastine, a vinca alkaloid and (b) taxol, a taxane.

The two types of agents previously mentioned are categorised as "conventional anticancer drugs" because they are non-specific and they target DNA-processing or cell division indiscriminately. Consequently, they do not show any selectivity for tumour cells over normal ones, leading to serious side effects. Because of this, a more selective generation of

anticancer drugs was developed, called molecularly targeted agents, which will be discussed later.

1.3.3. Hormonal Therapies

Hormonal therapy is used for several types of cancers occurring in hormonally responsive tissues, including breast, prostate, endometrium and adrenal cortex. Therefore, manipulation of the endocrine system by administration of hormones or drugs that act as hormone antagonists will stop the cancer growing or even cause cell death.[27] Surgical removal of the gland that produces the hormone is required in the treatment of such cancers.

An example of hormonal therapy is tamoxifen (Figure 1.7 a), a non-esteroidal estrogen antagonist that is used for the treatment of breast cancer. Tamoxifen competes with estrogen for binding to the estrogen receptors (ERs) thus preventing estrogen activation and subsequent tumour growth. Much research has gone into the discovery of novel selective ER modulators, as a replacement for tamoxifen, with more potent activity such as tamoxifen-like triphenylethylene derivatives (Figure 1.7 b), which are based on benzothiophenes, and steroidal anti-estrogens.[15]

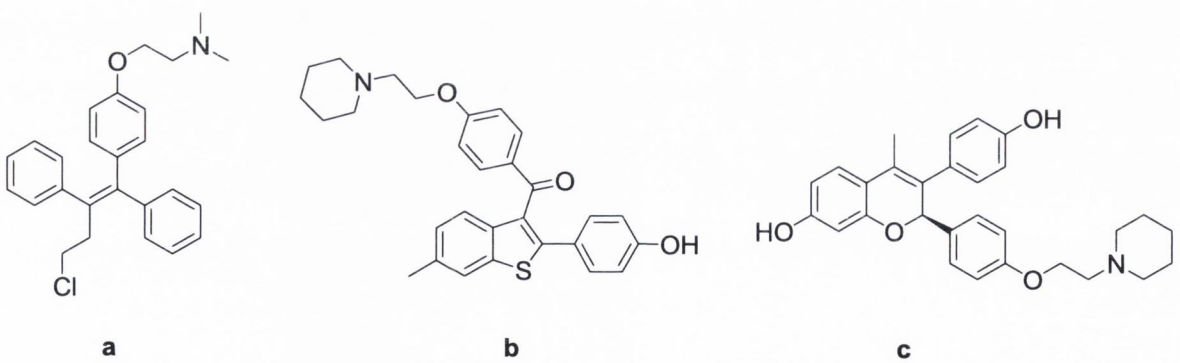


Figure 1.7. (a) Tamoxifen structure, (b) raloxifene, a benzothiophene based compound, (c) EM-652 (SCH 57068) steroidal anti-estrogen compound.

1.3.4. Tumour-Targeting Strategies

Current modern approaches to cancer treatment also include tumour-targeting strategies. Advances in achieving selectivity towards tumour cells have been made, for example, in the use of antibodies to lead a pro-drug/drug-releasing agent to tumour cells. Another approach involves targeting tumour vasculature by using anti-angiogenesis drugs.[28] These molecules interfere with tumour growth and survival by blocking the generation of new blood vessels, thereby preventing the development of the tumour. An example of such an agent is sorafenib.

Other approaches, such as modulating oxygen concentration and drug delivery techniques, are also used to target the tumour selectively over healthy tissue. Some pro-drugs are activated under the hypoxic conditions which are typical of tumour cells; they experience a bio-reduction under such conditions which results in release of the active drug, and such an activation is not possible in healthy cells.[29] Furthermore, novel drug delivery techniques such as nanotechnology, gene therapy, photodynamic therapy (PDT), intracranial wafers and ultrasound have been developed to release drugs specifically at the tumour site without harming healthy tissues.[30]

Photodynamic therapy is a treatment which uses photosensitising agents along with light, to kill cancer cells. The drugs only work after they have been activated or "turned on" by light of specific wavelength. Photodynamic therapy may also be called photo-radiation therapy, phototherapy, or photo-chemotherapy. Depending on the part of the body being treated, the photosensitising agent is either put into the bloodstream or the skin. When the drug is absorbed specifically by the cancer cells and light is applied to the area to be treated, the light causes the drug to react with oxygen, forming singlet oxygen which kills the cancer cells.[31] However, PDT has some limitations such as the fact that the activating light cannot pass through more than ~1 cm of tissue, making some tumours inaccessible to this treatment strategy. For this reason, PDT is a local treatment and generally cannot be used to treat cancer that has spread (metastasised), or large tumours.[32] Some of the PDT drugs approved for cancer treatment are: porfimer sodium (Photofrin®, for which the central porfirin moiety is illustrated in Figure 1.8 a) for oesophageal cancer and aminolevulinic acid (ALA or Levulan®, Figure 1.8 b) and its methyl ester (Metuixia®, Figure 1.8 c) for skin problems which can lead to cancer.

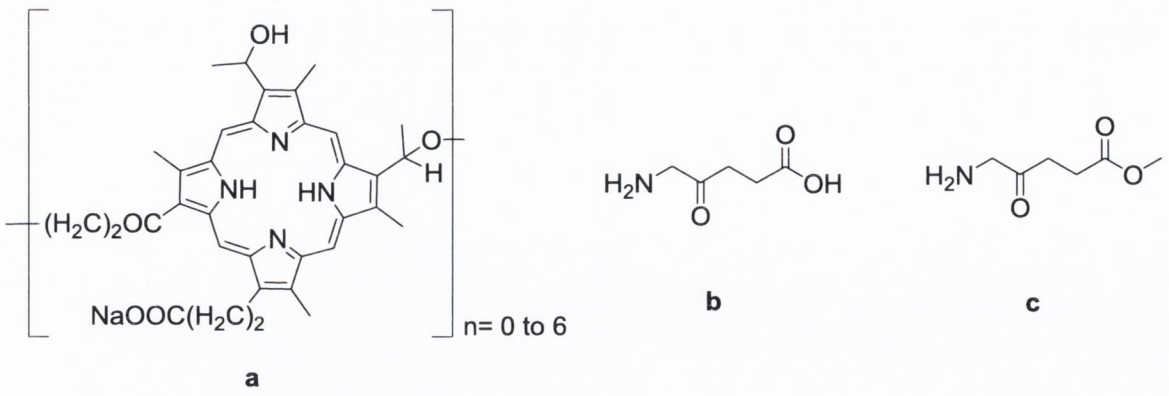


Figure 1.8. PDT drugs approved for cancer treatment: (a) Photofrin®, (b) Levulan®, (c) Metuixia®.

1.3.5. Biological Agents

Another variety of antineoplasic agents requires the utilisation of biological systems such as proteins or glycoproteins to modify the immune system and control the growth of the tumour. This may be achieved by enhancing the potency of the immune system cells (macrophages or natural killer cells) or by the employment of vaccines. These agents are produced either by extraction from human cells or by genetic engineering process. Biological response modifiers include such agents as antibodies, interferons and interleukins and may be used to control processes that facilitate the growth of tumours. Vaccines represent the fastest growth area for biological therapeutics.[15] However, from more than 100 cancer vaccines that are researched, less than half will ever reach commercialization. A success story is the human papiloma virus (HPV) vaccine to prevent cervical cancer. HPV vaccination has the potential to reduce cervical cancer deaths around the world by as much as two-thirds, and to prevent anal cancer in males and females.[33] While a promising approach to the treatment of cancer, it is worth noting that an over-exposure to these agents could lead to an allergic reaction and the desensitisation of the immune system.[34]

1.3.6. Molecularly Targeted Agents

Conventional chemotherapy treatment, although directed toward certain macromolecules or enzymes, typically does not discriminate effectively between rapidly dividing normal cells and tumour cells, thus leading to several toxic side effects. In contrast, targeted therapies interfere with molecular targets which have a role in tumour growth or progression. These targets are usually located in tumour cells, although some like the antiangiogenic agents may target other cells (*i.e.* endothelial cells). Thus, targeted therapies have a high specificity toward malignant cells, providing a broader therapeutic window with less toxicity. They are also often useful in combination therapies with cytotoxic chemotherapy or radiation to produce additive or synergistic anticancer activity, because their toxicity profiles generally do not overlap with traditional cytotoxic chemotherapeutics. Thus, targeted therapies represent a promising approach to cancer therapy, and they are already leading to beneficial clinical effects.

There are multiple types of targeted therapies available, including monoclonal antibodies, protein kinases inhibitors and inhibitors of growth factor receptors. Some examples are: inhibitors of the RAS pathway, inhibitors of the cell cycle, proteasomes and mTOR (mammalian target of rapamycin). Protein kinase inhibitors will be discussed in more detail in the next section as they are the main focus of this project.

It is thought that 30% of all human tumours are caused by a mutated RAS oncogene. RAS causes uncontrolled cell division when it loses its GTPase activity (hydrolysis of GTP to GDP) as a result of genetic mutations.[35] For this reason, inhibitors of this G-protein signal such as farnesyl transferase (FTase) inhibitors will decrease the activity of RAS protein and offer a target for anticancer drugs.[36] RAS require farnesylation (lipid modification) for its membrane attachment and function; hence, inhibition of farnesylation of RAS proteins was identified as a potential mechanism of inhibiting their function, thereby controlling cancer-cell growth. Some examples of drugs which operate through this mechanism of action are tipifarnib and lonafarnib (Figure 1.9).

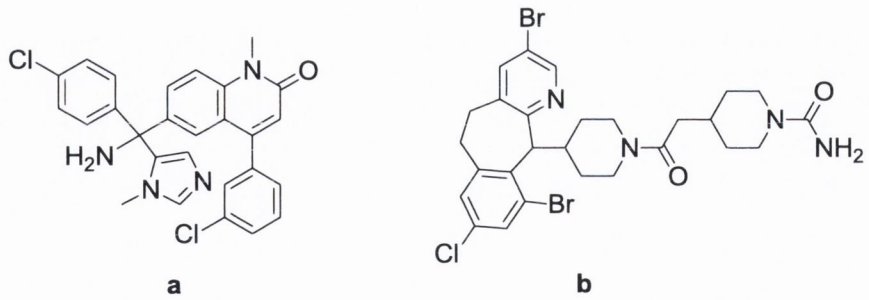


Figure 1.9. Inhibitors of RAS G-protein signalling; (a) tipifarnib, (b) lonafarnib.

Further downstream in the signalling cascade of the RAS, other proteins can be found such as the rapidly accelerated fibrosarcoma (RAF), mitogen-activated protein kinase/extracellular signal-regulated kinases (MEK) and extracellular signal-regulated kinases (ERK). Therefore, inhibiting any of these downstream proteins will also inhibit this RAS/RAF/MEK/ERK pathway and that is an attractive prospect for targeted chemotherapeutic agents.

In addition to these, another type of molecularly targeted agents is that of the cell cycle inhibitors. Progression through each phase of the cell cycle (Figure 1.10) is regulated by heterodimers formed by cyclin-dependent kinases (CDKs) and their regulatory partner proteins, the cyclins.[37] Together, they coordinate the cellular events which occur throughout the cell cycle, especially during the G_1 phase (post-mitotic phase) when the major period of cell growth takes place. Dysregulation of cell cycle control due to aberrant CDK activity is a common feature of most cancer types.

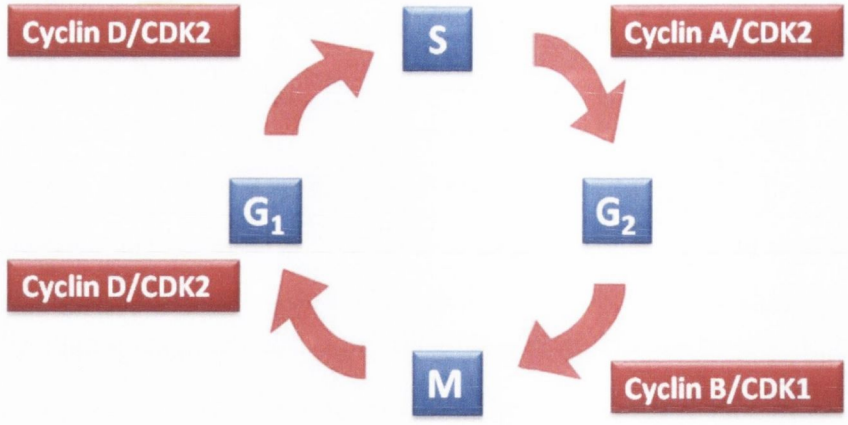


Figure 1.10. Role of cyclins and CDKs in the cell cycle.

Intensive research on small molecules targeting cell cycle regulatory proteins has identified many candidate inhibitors that are able to arrest proliferation and induce apoptosis in neoplastic cells, offering a promising new strategy to treat cancer. Flavopiridol (Figure 1.11, a) is a prototype inhibitor of CDKs and was the first one to enter clinical trials.[38]

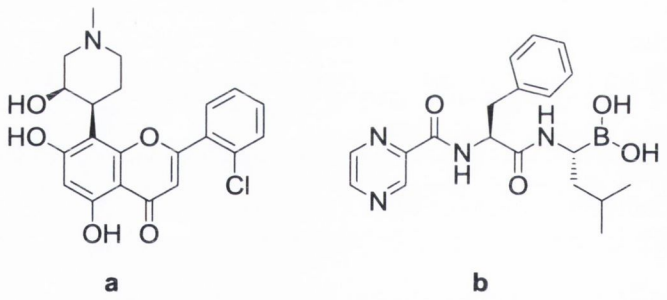


Figure 1.11. (a) Flavopiridol, a cell cycle inhibitor and (b) bortezomib, a proteosome inhibitor.

Similarly, proteasome inhibitors selectively attack proteasomes, which are complex proteins that degrade unnecessary or damaged proteins. This disposal pathway is crucial for tumourogenesis, as elimination of these key control proteins is known to halt cell cycle progression. For example, a proteasome inhibitor is involved in the deactivation of the pro-survival protein NF-κB, which is elevated in various types of cancer, by blocking

degradation of its inhibitory protein IκB. Then, NF-κB remains inactive and cannot elicit pro-survival effects, promoting apoptosis in tumour cells.[39] Bortezomib (Velcade®) is the first therapeutic proteasome inhibitor which was tested in humans for the treatment of multiple myeloma (Figure 1.11, b). Velcade® can also be administered in combination with other therapeutic agents for the treatment of other types of cancer.[40]

The final type of molecularly targeted agents is that of the mTOR pathway inhibitors. The mTOR protein plays an important role in the transmission of signals mediated by phosphatidylinositol-3-kinase (PI3K).[41] This PI3K/mTOR pathway is commonly over-expressed in cancer cells. Repression of mTOR produces a false signal indicating that the cell has a lack of nutrients and growth factors. Consequently, dramatic reprogramming, prevention of cell growth and arrest of cell division is initiated leading to tumour regression. Drugs that operate by this mode of action include the macrolides rapamycin and everolimus (Figure 1.12).[42]

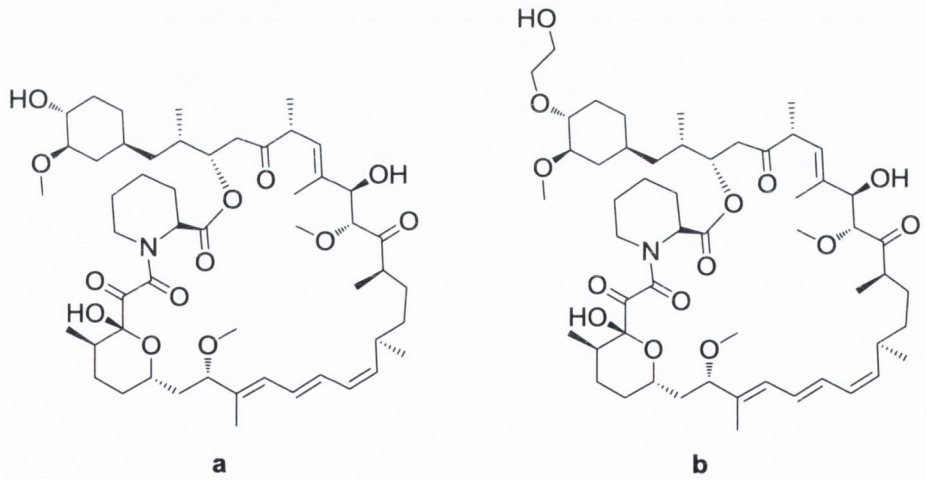


Figure 1.12. mTOR inhibitors (a) rapamycin and (b) everolimus.

As the complexity of signalling in cancers becomes better understood, therapeutic strategies to inhibit mTOR together with other proliferation pathways hold significant promise. Such strategies involve the combination of derivatives of rapamycin and inhibitors which prevent feedback pathway activation because of the correlation between

the pathways. They are inhibitors of epidermal growth factor receptor (EGFR1/2), insulin growth factor-1 receptor (IGF-1R), PI3K, B-RAF, and MEK.[43]

In conclusion, several antitumour agents for the treatment of cancer and several approaches to targeting cancer cells more selectively have been presented. Therefore, it is obvious that the current therapeutic strategies are very diverse and intricate as a result of the complexities of the disease itself. Nowadays, the so-called "War on Cancer" is still very much in progress, as is the need for the continuous development of new research tools and therapies in the area of chemoprevention.

1.4. Protein Kinases (PKs) and their Relevance in Cancer.

PKs are key regulators of cell function. Their mechanism of action consists of the chemical addition of a phosphate group to substrate proteins. Phosphorylation by a particular kinase involves the removal of a phosphate group from ATP and its covalent attachment to the hydroxyl group of an amino acid residue of the substrate protein. Protein phosphorylation is a universal switching mechanism that regulates nearly all cellular processes. PKs direct the activity, localisation and overall function of many proteins, and serve to orchestrate the activity of almost all cellular processes.

Protein kinases are essential enzymes in cellular signalling that regulate cell growth, proliferation, metabolism, differentiation, migration, signal transduction and transcription.[44] All these processes are crucial molecular targets for the design of current cancer drugs. In response to specific extracellular metabolic signals, PKs activate target proteins by phosphorylation and hence expand the signal into cellular growth cascades.[45]

There are several cases in which the abnormal regulation of kinases can trigger the development of cancer, owing to the importance that kinases hold in signalling pathways, they are summarised in Figure 1.13.

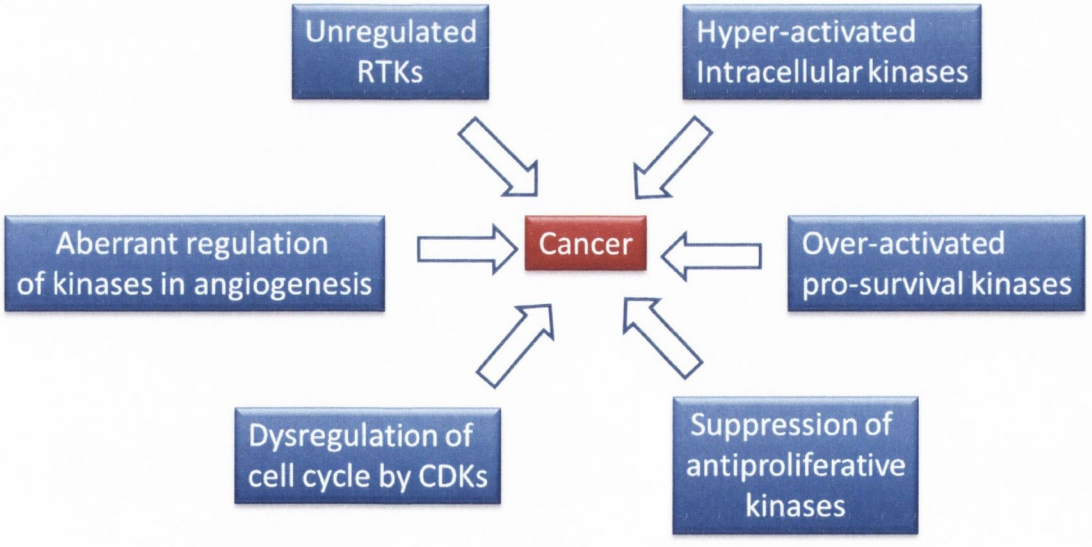


Figure 1.13. Influence of kinases in the generation of cancer.

In conclusion, dysregulation of kinases, together with autocrine growth factor production, are important in promoting the autonomous growth and proliferation, evasion of apoptosis, and invasive phenotype that are characteristic of tumour cells.[4]

1.4.1. The Kinome

The kinome describes the kinase component of the human genome, and was compiled by Manning *et al* in 2002.[46] Of the 518 human PKs, 478 are eukaryotic protein kinases (ePKs) belonging to a single super-family whose catalytic domains are related in sequence,[47] and 40 are "atypical" PKs. The ePKs can be clustered into eight main groups, families and sub-families of increasing sequence similarity and biochemical function, which are illustrated in the kinome tree shown in Figure 1.14.

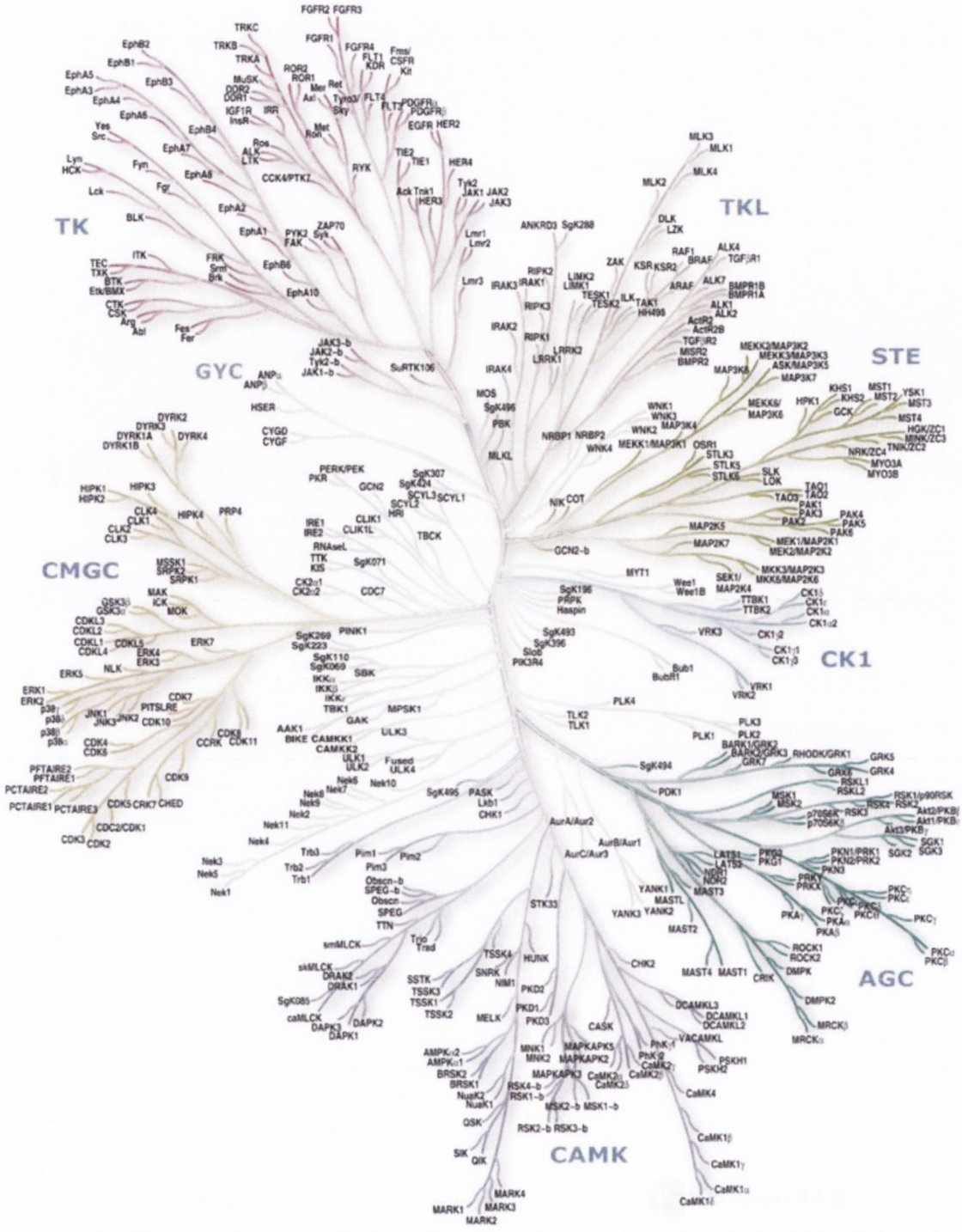


Figure 1.14. Kinome tree.[48]

The main PK families are: AGC (containing protein kinase A, G and C (PK-A, G and C)), CAMK (calcium/calmodulin-dependent protein kinases), CMGC (containing cyclin-dependent kinases -CDK-, mitogen-activated protein kinases -MAPK-, glycogen synthase kinases 3 -GSK3- and cdc2-like kinases -CLK-), TK (tyrosine kinases), STE (including

homologues of the yeast sterile kinases), CK1 (casein kinase 1), TKL (tyrosine kinase-like) and RGC (receptor guanylate cyclase).[49] Over half of the kinase genes identified can be associated with diseases and 164 kinases are frequently found in tumours. This fact underscores once again the importance of the PK family in cellular signalling and the relevance of their dysregulation to tumour biology.

A description of all of the families of PK is outside the scope of this work. However, a description of those families of the kinome most relevant to this work will be presented in more detail in the following sections.

1.4.2. Tyrosine Kinase Group (TKs)

TKs are one of the most widely studied and significant kinase families with respect to cancer biology. They have been divided into two major groups: (i) transmembrane receptor tyrosine kinase (RTK), characterised by membrane localisation and by the presence of an extracellular domain, and (ii) non-receptor tyrosine kinase (NRTKs), mainly located in the cell cytoplasm, as integral components of the signalling cascades triggered by receptor TKs and other cell-surface receptors.[50] In humans, there are around 90 TKs, which can be divided into 58 RTKs and 32 NRTKs.

1.4.2.1. Receptor Tyrosine Kinases (RTKs)

The RTKs are key sensors of the extracellular environment, transducing signals across the membrane through transmembrane receptors to a diverse range of intracellular signalling networks. Depending on the cells/tissues and ligands/receptors involved, these internal signals may modulate processes such as mitogenesis, metabolism, migration and angiogenesis.

Each RTK contains an extracellular N-terminal segment, where the growth factors bind, a transmembrane segment and an intracellular C-terminal segment, which contains the kinase domain and tyrosine phosphorylation sites. Binding of the external growth factor induces monomeric receptors to dimerise and this dimerisation induces activation of the TK activity through autophosphorylation. This, then leads to a cascade of phosphorylation events involving different PKs which act as intracellular transducers or transcription factor

regulators, modifying gene expression (Figure 1.15).[51] Signal transduction can be downregulated by protein tyrosine phosphatases, by dephosphorylation or by other mechanisms. The RTKs are involved in several biological processes in all types of cells, encompassing cellular growth, proliferation, survival, differentiation and motility. Given these crucial biological roles of RTKs, their dysregulation is a common feature in a broad range of cancers.[52] For instance, RTKs act as drivers of tumour growth because they are the switch which initiates complex and interconnected intracellular signalling cascades that promote cell growth, proliferation and survival.[4] Aberrant activation of these receptors results in a dysregulation of the kinases in downstream pathways, allowing the tumour to proliferate. Another characteristic of this abnormal behaviour of RTKs is that they are no longer modified by external signals that control their expression, such as growth factors or hormones. Receptor over-expression may enable the cancer cell to become hyper-responsive to external levels of growth factors that normally would not trigger proliferation. This means that they are kept active continuously, sending the wrong message to downstream kinases, resulting in a ligand-independent signalling that causes distorted cell responses and, ultimately, proliferation of a tumour. Some examples of unregulated RTKs that generate different types of cancer are the deviant behaviour of the KIT receptor in gastrointestinal stromal tumour (GISTs),[53] Ableson leukemia oncogene cellular homolog (ABL) in chronic myelogenous leukemia (CML),[54] epithermal growth factor receptor (EGFR) in gastric, brain, breast, endometrial and colorectal tumours.[55] Amplification or over-expression of the v-erb-b2 erythroblastic leukaemia viral oncogene homolog (ERBB2) gene occurs in approximately 30% of breast cancers, while HER2/neu receptor is over expressed in stomach and mammary carcinoma.[56]

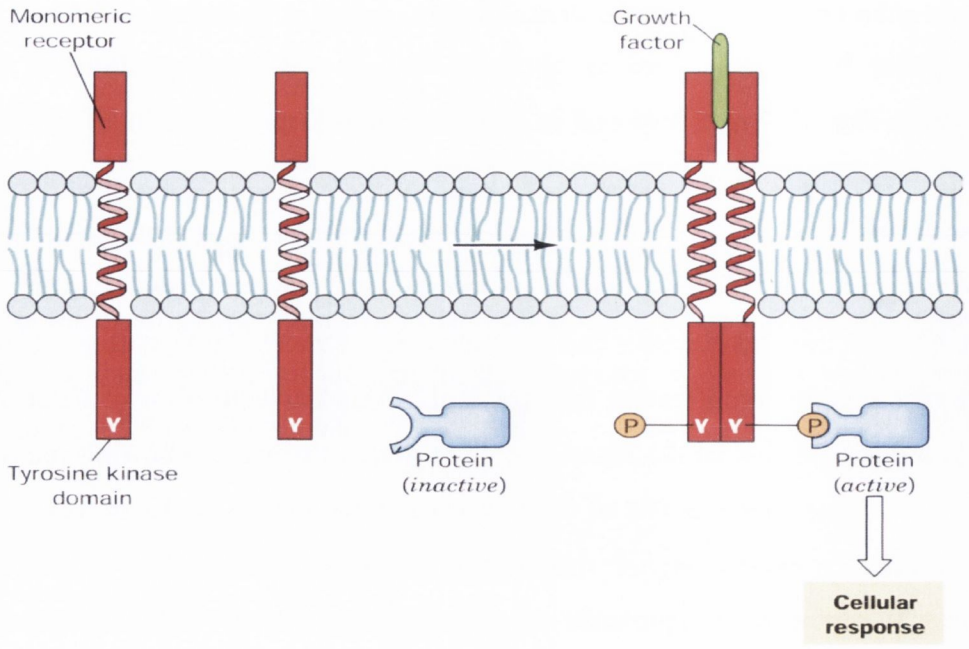


Figure 1.15. Activation process of RTKs.[57]

Many RTKs and intracellular kinases are significant regulators of pro-survival signalling, such as B-cell lymphoma gene 2 (BCL-2) proteins and NFκB transcription factors, so when they are over-activated in cancer cells, they provide resistance to cell death signals.

From the 20 families of RTKs those more implicated in cancer are: EGFR family, IR family (insulin receptor), PDGFR family (platelet- derived growth factor receptor), FGF receptor family (fibroblast growth factor), VEGFR family (vascular endothelial growth factor), HGF receptor family (hepatocyte growth factor), TRK receptor (tropomyosin-related kinases) and EPH receptors (erythropoietin-producing hepatocellular carcinoma).[58]

The RTKs were among the first kinases to be targeted for therapeutic purposes (EGFR was the first RTK targeted in oncology, during the 1980s).[59] Since then, many RTKs have been found to be over-expressed, amplified, mutated and/or translocated in human tumours. Therefore, several RTK inhibitors are currently approved for the treatment of different cancers and others are being evaluated in clinic. The RTKs are also vital in promoting tumour angiogenesis, the inhibition of which was proposed as an anticancer strategy over 30 years ago.[60] From the families listed above, the only ones relevant to this study are the PDGF and VEGFR families, thus, they will be discussed next.

The growth of new blood vessels or support of the existing capillary networks allows tumours to grow in size and to expand to other tissues. Under conditions of low pH, hypoxia, poor nutrient availability, and inflammatory stimuli, both cancer cells and tumour stromal cells produce VEGF and other pro-angiogenic factors, promoting the recruitment of new blood capillaries that can further support tumour growth. A major contributor to this process is VEGFR, which promotes the proliferation, survival and formation of vascular tissue, as well as the migration of endothelial cells. There are five VEGFR endogenous ligands and three receptor subtypes. The ligands are VEGF A, B, C, D and placental growth factor (PlGF) and the receptors are VEGFR1 (FLT1), VEGFR2 (KDR) and VEGFR3 (FLT4).[61] Each of the ligands has a characteristic receptor binding pattern, producing different cellular responses. Furthermore, each receptor has a slightly different function. Nevertheless, the common activity of these members of the VEGF family is the initiation of angiogenesis, allowing the tumour to grow and proliferate.

There are two strategies which can be employed to inhibit the VEGF pathway: (i) targeting the VEGF ligands and (ii) targeting the VEGF receptors. Generally, inhibitors of this VEGF pathway stop the generation of blood vessels that assist the expansion of the tumour and subsequently, tumours have no nutrients and oxygen to proliferate. The inhibition strategies of PK will be described in more detail in section 1.5.2.

On the other hand, PDGF family members contribute to multiple tumour-associated processes and are involved in the control of cell growth, proliferation and cellular differentiation.[62] Over-expression or constitutive activation of PDGF signalling has been linked to various diseases involving excessive cell proliferation such as various fibroses and multiple tumour types such as gliobastoma (brain tumour) and GIST.[63] Additionally, PDGF signalling is decisive for maintenance of the perivascular structure that surrounds and supports tumour blood vessels, leading to interest in PDGFR as targets for antiangiogenic cancer therapies. The PDGFR consist of a disulfide linked dimer which has five isoform-homodimers A, B, C, D-polypeptide chains and heterodimer PDGFAB.[63] When these isoforms bind to the receptors PDGFRα, PDGFRβ or PDGFRα-PDGFRβ they induce dimerisation, which allows autophosphorylation. Phosphorylation of the tyrosine residues is involved in the regulation of catalytic activity and downstream signalling pathways. Examples of other members of the PDGF family are the stem cell factor receptor KIT and FLT3 (FMS-like tyrosine kinase 3). Over-expression of KIT is observed in melanoma, AML (acute myelogenous leukemia), and over 90% of GIST tumours.[64] In

addition to this, FLT3 is often mutationally activated in AML and lymphocytic leukaemias.[65]

The concept of tumour angiogenesis was first proposed in the 1970s and, since then, many anti-vascular agents have been evaluated as anticancer therapies in the clinic. These efforts have met with some success, and both antibody and small molecules targeting angiogenesis have been approved in several oncology indications. Moreover, the new generation of anti-angiogenesis therapeutics, involving the disruption of both VEGFR and PDGFR signalling, have been shown to be more effective than blocking each receptor alone. This novel class of anti-angiogenesis inhibitors is called "angiokinase inhibitors" and an example of this is sorafenib.[66]

In conclusion, studies on the roles of RTK in cancer signalling are presently in a dynamic phase where collaborations between chemists, oncologists, pathologists, and tumour biologists are predicted to be highly productive.

1.4.2.2. Non Receptor Tyrosine Kinases (NRTKs)

The NRTKs are cytoplasmic proteins that do not directly interact with extracellular ligands. However, they are tightly coupled to signalling cascades downstream of a diverse range of RTKs, G-protein coupled receptors and cytokine receptors. Their dysregulation underlies a number of pathologies including multiple types of neoplasia. There are six NRTK families: rous sarcoma oncogene cellular homolog (SRC), TEC, ABL, C-terminal SRC kinase (CSK), Janus kinase (JAK) and Syk-related tyrosine kinase (SYK).[58] Only those NRTK families relevant for this study will be discussed herein. The first family of NRTKs to be identified was SRC.[67] Members of the SRC family have a number of cellular functions (such as gene transcription, cell movement and differentiation)[68] and are also the switch of activation for other kinases downstream, such as the RAS pathway, PI3K and signal transducer and activator of transcription (STAT), amongst others. Its mutation is associated with many tumour types.

The first NRTK inhibitor to reach the clinic was an inhibitor of ABL, imatinib mesylate, commercially known as Gleevec®. This small molecule kinase inhibitor revolutionised the treatment of chronic myelogenous leukaemia (CML) and also served to enhance the interest in kinase inhibitors as anticancer drugs. The ABL kinase regulates cytoskeletal

function, progression through the cell cycle, differentiation and cell death. The clinical importance of ABL emerged through the discovery of the "Philadelphia chromosome", which is involved in a translocation between chromosomes 9 and 22, leading to the head-to-tail fusion of the BCR and ABL genes, generating break point cluster region abelson (BCR-ABL) fusion protein.[69] This 9:22 translocation occurs in more than 90% of CML cases and Gleevec® inhibits it.

Another NRTK family distinguished for its use in oncology is the JAK family. The activation of JAK results in the activation and phosphorylation of some of the main kinases pathways such as MAPK signalling proteins, PI3K/AKT pathway and STAT (signal transducer and activator of transcription) proteins, which enter the nucleus and regulate gene transcription.[70] Hyper-activated signalling due to dysregulation of downstream kinases, like in the PI3K/AKT/mTOR pathway is observed in many cancers such as breast cancer,[71] NSCLC[72] and RCC,[73] and it correlates with tumour growth and survival.

1.4.3. Serine/Threonine Kinases (Ser/Thr kinases)

The differences between Ser/Thr kinases and TK reside in the amino acid that they phosphorylate. This means that Ser/Thr kinases phosphorylate the serine and threonine residues of their protein substrate. Among the most significant members of this group are the RAF family of kinases that are part of the RAS/MAPK pathway, which plays a key role in cancer biology. Other relevant Ser/Thr kinase members are CK-1, GSK-3, PK-A, PK-B, PK-C and also receptor protein Ser/Thr kinases like TGFβ (transforming growth factor), which has an enigmatic role in cancer biology and has been suggested as a potential target for therapeutic intervention.[74] Only those Ser/Thr kinases more relevant for this study, MAPK, GSK-3β and CK-1δ, will be discussed in more detail.

1.4.3.1. MAPK Pathway (Mitogen-Activated Protein Kinase)

The MAPK pathway is highly relevant to this project as the principal aim of our research is to synthesise novel molecules which inhibit it. For this reason, a comprehensive review has been carried out to properly understand the mechanisms of this pathway.

The MAPKs belong to the CMGC kinome group, as previously mentioned. They are down-stream effectors of a signalling mechanism which is highly conserved from yeast to humans. These kinases control signalling pathways that are key regulators of several aspects of normal cell growth and malignant transformation. Besides the important roles in cellular processes, MAPKs have been found to play key physiological roles, such as regulating embryonic development and maintaining tissue homeostasis.[75] Dysregulation of this pathway in tumours contributes to many hallmarks of cancer, including immortalisation, uncontrolled proliferation, metastasis, angiogenesis and evasion of apoptosis.[76] It is aberrant in 30% of all cancers[77] due to activating mutations in the RAS genes or to alterations in upstream or downstream signalling components.[76]

They comprise three major subgroups: ERK, p38 MAPK and JNK (c-Jun N-terminal kinases).[75] Despite the diversity in function and upstream signalling events, MAPKs are always activated by a highly conserved mechanism that involves a sequence cascade of phosphorylation starting with MAPKKK, which catalyses the phosphorylation of MAPKK, which subsequently phosphorylates the last MAPK kinase, which translocate inside the nucleus and modify transcription factors, resulting in a biological response (Figure 1.16). Each phosphorylation step responds to different extracellular signals (growth factors or UV light or stress) which stimulate an intracellular pathway activator.

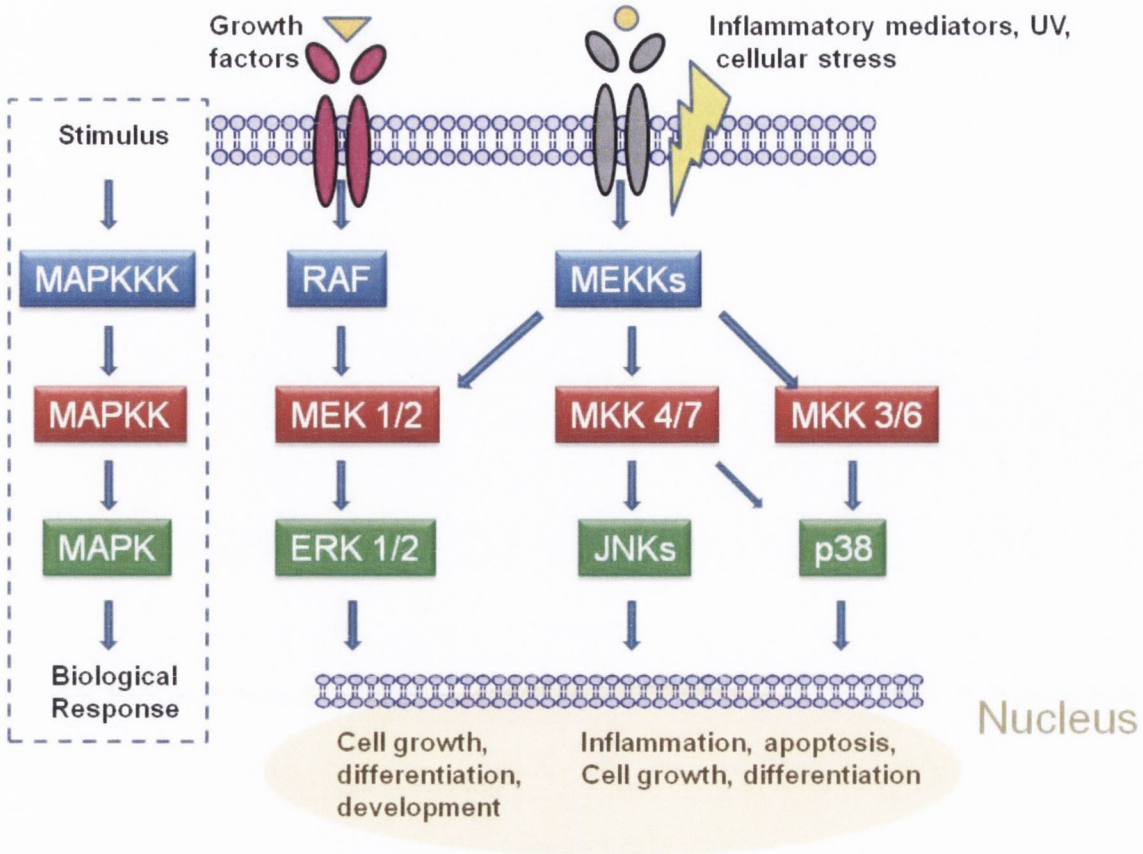


Figure 1.16. MAPK signalling pathways ERK, JNK and p38.

Of all the systems of the MAPK pathway, the most relevant for this study is the ERK/MAPK pathway, which is activated extracellularly by growth factors or hormones, resulting in the activation of RTKs located in the plasma membrane. These RTKs can then induce the phosphorylation of RAS GTPase, which recruits and phosphorylates RAF. Next, RAF phosphorylates downstream MEK1/2, which further activates ERK1/2. The latter translocates to the nucleus where it regulates gene expression through the activation of several key transcription factors.[78] There are several isoforms of RAF: A-RAF, B-RAF and C-RAF (also known as RAF-1). All of these isoforms share a common structure, with the kinase domain in the C-terminal half of the protein and two additional conserved sequences (CR1 and CR2) in the N-terminal half;[79] however, B-RAF has an extended N-terminal region. The three conserved regions of RAF kinases are preserved across isoforms and species (Figure 1.17). The CR1 sequence has a RAS binding domain (RBD) and a cysteine rich domain (CRD), while CR2 is Ser/Thr rich (S/T rich) and CR3 is the catalytic

kinase domain.[79] B-RAF has higher kinase activity than C-RAF and A-RAF, and it is the most important proto-oncogene among the RAF kinases.

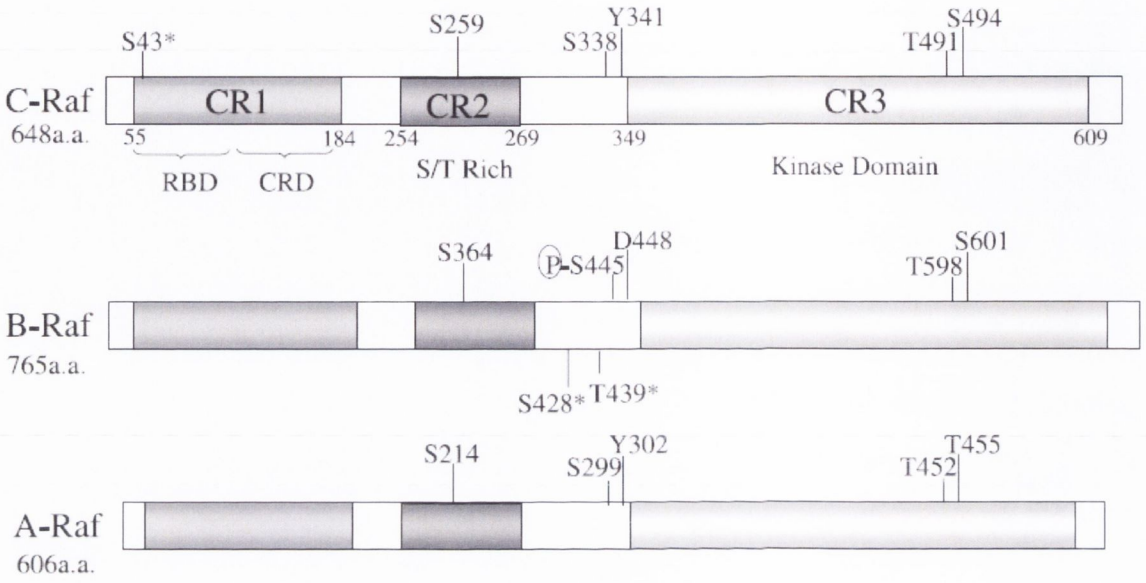


Figure 1.17. Conserved structure of the RAF kinase family (C-RAF, B-RAF and A-RAF).[79] Conserved sites of activating regulatory phosphorylation are shown in with a number; other phosphorylation sites that are not conserved are marked with an asterisk (*) and these are sites of inhibitory phosphorylation.

While A-RAF and C-RAF are rarely mutated in human tumour cells, B-RAF is mutated in 7% of cancers.[80] Mutations in B-RAF (especially ^{V600E} B-RAF, which results in an amino acid substitution at position 600 in B-RAF, from a valine V to a glutamic acid E) have been identified in a number of malignancies, with the highest incidence occurring in malignant melanoma (60-80%), papillary thyroid cancer (35-70%), colorectal cancer (~10%) and a wide variety of other malignancies at lower frequencies.[58] Thus, RAF isoforms are considered an important target for antitumour drug development.

Additionally, other relevant MAPK pathways are JNK and p38 MAPK (Figure 1.16). Their kinase networks are highly interrelated, *i.e.* there are very few stimuli that can elicit JNK activation without simultaneously activating p38, or vice versa.[81] They play a central role in the regulation of the biosynthesis and actions of cytokines (key proinflammatory

mediators), which are important regulators of the immune response to infection as well as cellular stress.[82]

There are other intracellular signal transduction cascades besides MAPK, such as the PI3K/PTEN (protein tyrosine phosphatase) pathways, mTOR signalling, the PIM (provirus integration site for moloney murine leukaemia virus) kinases and PK-C pathways. Some of these pathways interrelate with the MAPK pathway; however, an explanation of these transduction pathways is beyond the scope of this study.

In conclusion, MAPKs have emerged as promising targets for therapeutics, and great efforts are being made to explore new functions and mechanisms regulated by MAPKs in disease conditions.

1.4.3.2. Casein Kinase 1δ (CK-1δ) and Glycogen Synthase Kinase 3-β (GSK-3β)

Both GSK-3 and CK-1δ are multifunctional Ser/Thr kinases. The former especially is implicated in many cellular processes involved in proliferation, differentiation and survival, and has been pursued as a therapeutic target in diabetes, neurological disorders and inflammation. On the other hand, CK-1δ also has several cellular functions, such as regulation of membrane transport, cell division, DNA repair, control of circadian rhythms, and nuclear localisation.[83]

As well as inhibiting glycogen synthesis, GSK-3 is a key regulator of the WNT/β-catein pathway, and as CK-1δ, has been associated with tumour formation. Both GSK-3 and CK-1δ are involved in the modification of the WNT pathway which leads to dephosphorylation and nuclear translocation of β-catenin, where it promotes morphogenetic signalling.[84] In the nucleus, β-catenin activates transcription factors, such as AP-1,[85] the same transcription factor that the RAF/MEK/ERK pathway targets (Figure 1.18).

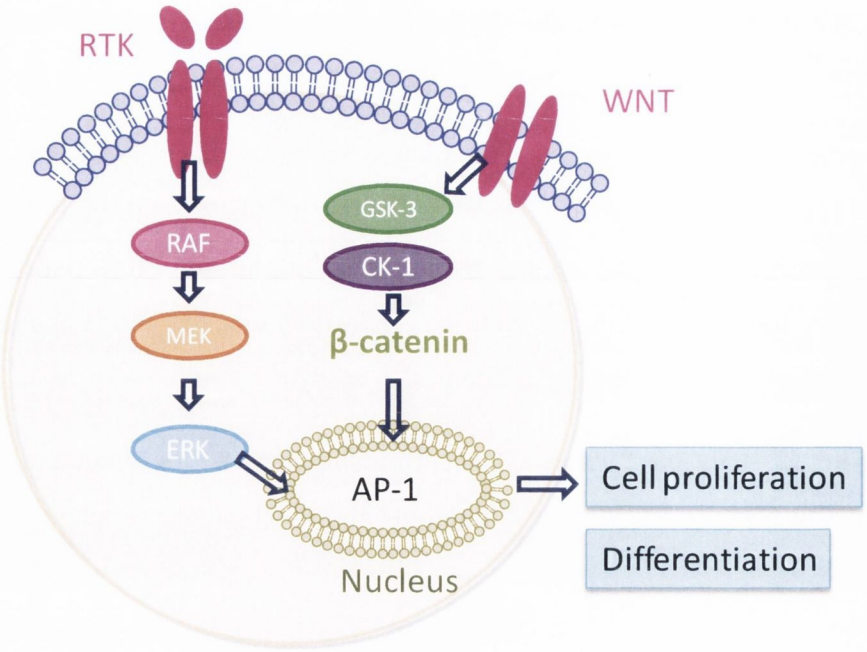


Figure 1.18. RAF/MEK/ERK and WNT signalling pathways and their cellular effects.

In light of this connection between RAF/MEK/ERK and GSK-3 and CK-1 we decided to investigate these two pathways during the study of our new PK inhibitors.

1.5. Protein Kinase Inhibitors (PKIs)

Protein kinases play a critical role in the modulation of growth factor signalling. Activated forms of these enzymes can cause increases in tumour cell proliferation and growth, induce anti-apoptotic effects, and promote angiogenesis and metastasis. In addition to activation by growth factors, PK activation by somatic mutation is a common mechanism of tumourogenesis.[86] For these reasons PKIs opened new routes for the treatment of cancer and the development of anti-cancer drugs. There are a lot of aspects to take into consideration for the design of new PKIs, the major characteristics are discussed in this section.

1.5.1. Types of PKIs

There are three main types of PKI classified as type I, II and I ½, according to how they interact with the domains of the kinase. The core of the kinase is composed of: (i) the hinge region, where the ATP binding site is allocated, (ii) the catalytic site, where the DFG (Asp-Phe-Gly) motif is present which can assume the active DFG-in or the inactive DFG-out conformation, (iii) a so-called "gatekeeper" residue that flanks the hydrophobic pocket, (iv) helix N-terminal and (v) helix C-terminal.

The initial strategy was to develop compounds that could mimic ATP binding to the kinase. This led to the first generation (type I) of PKI. They are ATP-competitive and bind to the ATP site through the formation of one to three HBs (one acceptor and two donors) with the backbone residues of the hinge region and a hydrophobic moiety (see Figure 1.20 b). A representative example of this type of compound is sunitinib (see structure in Figure 1.23 f). The problem with these inhibitors is that the hinge region is highly conserved across all members of a kinase group, so cross-reactivity is a frequent occurrence. In a recent study, a general scaffold or pharmacophore, that is maintained in several libraries of PKI, was proposed to target the hinge region (Figure 1.18).[87]

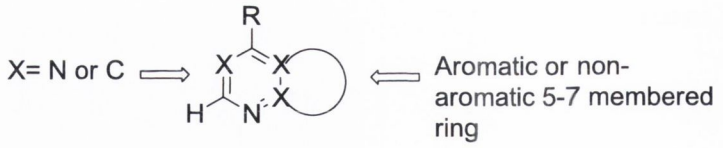


Figure 1.19. General scaffold to target the hinge region in several libraries of PKI.

This indicates that to increase selectivity when targeting kinases, it is necessary to bind to other regions in addition to the hinge region. This problem is addressed in the type II inhibitors. Type II inhibitors – apart from binding to the hinge region – bind to the DFG "out" domain. This DFG domain is the conserved triad of amino acids at the beginning of the activation loop of the kinase, which consists of Gly-Phe-Asp aminoacids.[88] If these amino acids are in this order, the DFG is in the inactive, or "out", form. However, the conformation of the DFG "in" is characterised by an almost 180 degree rotation of the conserved DFG motif (Asp-Phe-Gly) to the inactive form.[89] The different position of the DFG residues in the "out" form results in the opening of an additional cavity, the allosteric site, which is hydrophobic in nature and is specifically targeted by type II inhibitors. Therefore, these compounds not only bind to the ATP region through a set of HBs to the hinge region – like type I compounds – but also establish hydrogen bonding interactions with the back cavity (conserved glutamic amino acid residue) and hydrophobic interactions within the allosteric site (see Figure 1.20 c). This allosteric site is commonly referred to as the "selectivity pocket", as it is different in each kinase, so selectivity between kinases can often be found by interacting with this region.

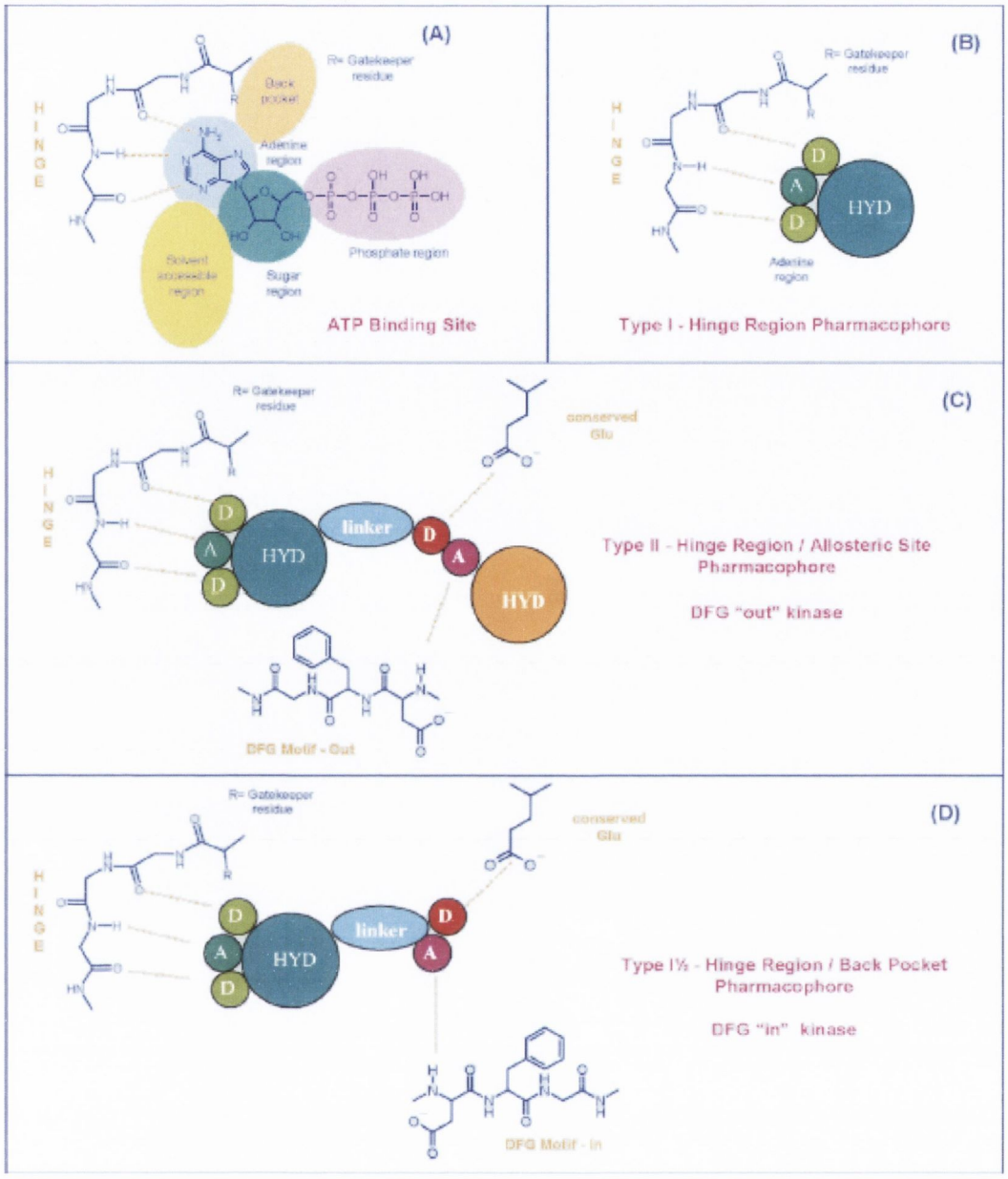


Figure 21.20. Types of PKIs, type I, II and I ½.[88] Note: D: HB donor, A: HB acceptor, HYD: hydrophobic moieties which occupy the adenine (green) and allosteric site (orange).

Type II inhibitors can be viewed as an extension of the type I inhibitors, where the hydrogen bonding part is connected with the allosteric site through a linker. For this reason, type II inhibitors are more selective than type I inhibitors. A quintessential example of this type of inhibitors is sorafenib (see structure in Figure 1.23 d). However, the main problem with targeting the DFG "out" form is that it is more vulnerable to mutations in the

kinase domain, leading to loss of activity of the inhibitor. To avoid this problem the next family of inhibitors, type I ½, bind to the DFG "in" form.

The type I ½ inhibitors are in between type I and II inhibitors. Hence, they form HBs with the hinge and the back cavity, but these inhibitors interact with the DFG "in" form of the activation loop of the kinase (see Figure 1.20 d). However, they do not form hydrophobic interactions with the allosteric site as type II inhibitors do. Type I ½ inhibitors are usually sought for kinases that are involved frequently in mutations. An example of type I ½ inhibitors is lapatinib (see structure in Figure 1.23 h).

Recently, another type of kinase inhibitor has emerged, the type III inhibitors. They are non-ATP competitive so they block kinase activity without displacing ATP. They do not target the ATP binding site of a kinase, but bind to allosteric sites of the protein.[90] There are cases where the inhibitor binds only to allosteric sites of the protein or simultaneously binds to both the allosteric site of the protein and to the ATP or ADP. An example of the latter is the MEK inhibitor PD318088, which crystal structure of the binding to MEK-1 and ATP is shown in Figure 1.21.[91]

PD318088

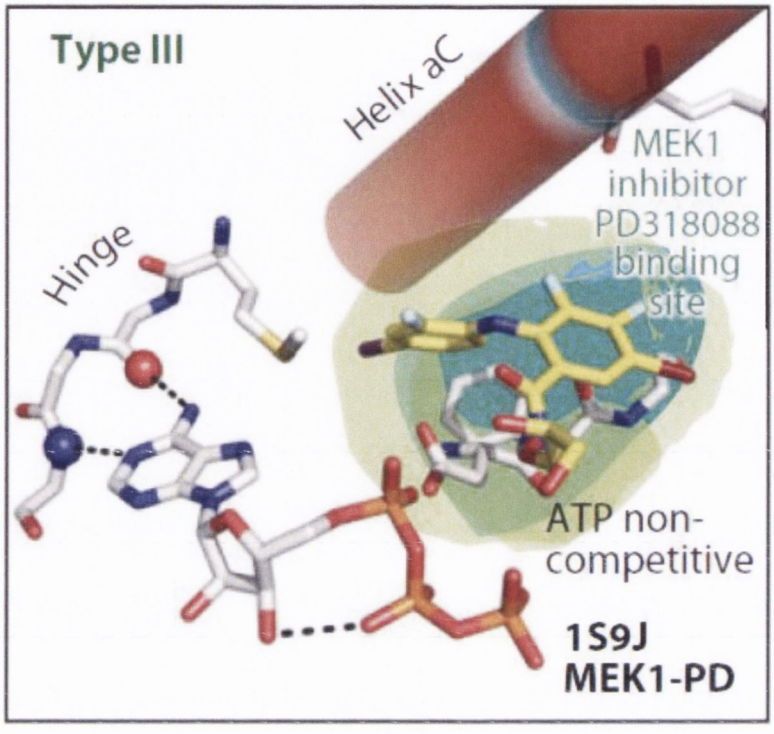


Figure 1.21. Type III inhibitor PD318088 (yellow) which blocks MEK-1 kinase activity without displacing ATP. PDB ID 1S9J.[91]

These inhibitors do not require any typical hinge binding motifs and can target the kinase regardless of its activation state. As allosteric sites are highly specific for a given kinase, type III compounds can be expected to exert a high degree of selectivity and potency. For this reason, there are continuous efforts to bring more candidates like these from laboratory benches to the clinic.

1.5.2. Therapeutic Strategies of PKIs

There are different therapeutic strategies for which PKIs exhibit their inhibitory kinase activity. The two main classes are small organic molecules and antibodies. Small molecules normally have a molecular weight of <1,000 D, and are produced by organic synthesis or from natural products. Examples of synthesised small molecules are sorafenib and imatinib, and an example of a natural product source is rapamycin. Small molecules can target one kinase selectively (single target) or several kinases promiscuously (multitarget). However, antibodies are monoclonal and recognise a single epitope on a single target only. The drawback of this therapy is that antibodies are not cell permeable, so their activity is restricted to kinases on the cell surface. This does not occur with small molecules, thus, they exhibit a broader scope of action. Some of the more prominent examples of antibodies are the anti-HER2 antibody trastuzumab (Herceptin®), the anti-EGFR antibody cetuximab (Erbitux®), and the anti-VEGF antibody bevacizumab (Avastin®).[92]

The molecular mechanism of small molecules and antibodies is different. While small molecules inhibit the catalytic activity of kinases intracellularly (Figure 1.22, number 3), antibodies inhibit ligand binding (Figure 1.22, number 1) and/or receptor activation extracellularly (Figure 1.22, number 2). For example, the antiangiogenic anti-VEGF antibody, bevacizumab, is designed to directly bind to the VEGF ligand and prevent its interactions with the VEGF receptor, thereby inhibiting its biological activity. Other comparisons between small molecules and antibodies are the cost of manufacture and pharmacokinetic properties. Antibodies are more costly than small molecules but the latter have longer half-life at the target.

Apart from the previously mentioned mechanisms of action of the PKIs – *i.e.* targeting ATP or allosteric sites of the kinase (Figure 1.22, numbers 5, 6, 7) and inhibiting ligand binding or receptor activation (Figure 1.22, numbers 1, 2 and 3) – PKIs exhibit other therapeutic strategies. An additional mode of action of PKIs is to target the RNA of a gene and inhibit its expression; this is called antisense therapy (Figure 1.22, number 4).[93] For example, C-RAF has been targeted this way in clinical studies, however no drugs have resulted as yet.[93] Furthermore, small molecules can also influence the stability of PKs by binding to chaperone proteins (Figure 1.22, number 8). In many cancers chaperones are overactive, stabilising faulty proteins that drive the growth of cancer cells. Some PKIs

inhibit these chaperones, stopping cancer cells from growing. An example of a PKI with this mode of action is geldanamycin.[94]

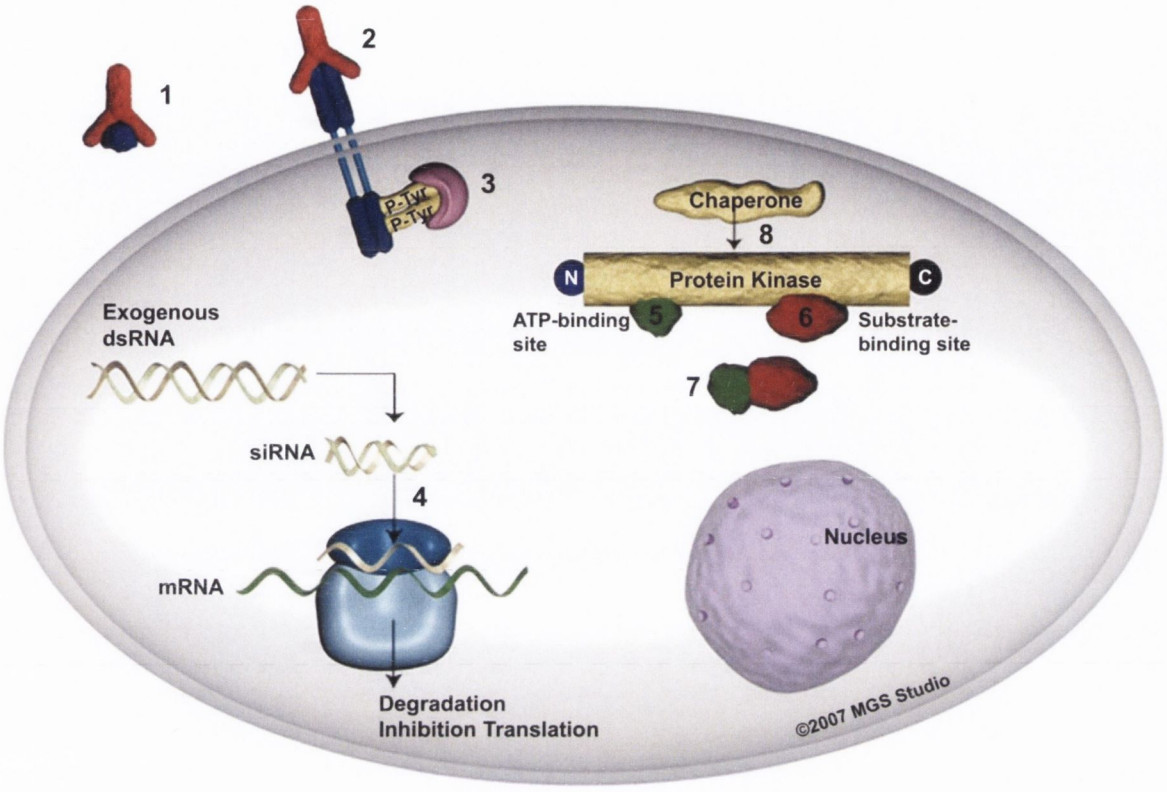


Figure 1.22. Mechanism of action of PKIs: Antibodies can prevent a small molecule from binding to its RTK either (1) by interacting with the ligand or (2) by blocking the binding site on the receptor; small molecules can also (3) block the intracellular part of the receptor, or (4) target the RNA of a gene, stopping its translation; competitive inhibitors of the (5) ATP-binding site, (6) the allosteric or substrate-binding site or (7) both, and (8) inhibition of stabilising chaperone proteins are also employed.[95]

1.5.3. Influence of PKIs in Medicinal Chemistry

Over the last few years PKs have been among the most actively studied pharmaceutical targets. In 2005, it was estimated that one out of every three discovery efforts was targeting PKs.[96] However, not all kinases have received equal attention and a large majority of kinases have not been studied at all. Certainly TK, CMGC and AGC kinases have been intensively researched and there is an extensive amount of kinase inhibitor publications and patents that deal with them. However, it is worth noting that there are entire branches of the kinome which remain "untouched" and for which no small molecule inhibitor is known. Unfortunately, many of these kinases have uncharacterised function and, as a result, researchers and pharmaceutical companies have not been motivated to develop new inhibitors of these kinases. With an ever-growing range of kinase inhibitors, broad assay panels and in combination with crystallography efforts, kinase scientists are well positioned to identify molecules that can elucidate the function of un-targeted kinases and discover new targets, which could ultimately help patients.

To illustrate the fact that there is a lot of "un-targeted kinome", 13 out of 14 PKI currently approved by the FDA target only TKs or Ser/Thr kinases, and not any other branch of the kinome (Table 1.1). The exception is Fasudil, an inhibitor of Rho-associated protein kinase, ROCK (a member of the commonly targeted AGC kinase family), approved for acute cardiovascular disease in Japan. All of these inhibitors are small molecules, whose structures are represented in Figure 1.23. The only kinase inhibitor that is not a small molecule is ziv-aflibercept, which is a protein comprised of segments of the extracellular domains of human VEGFR-1, -2 fused to the constant region (Fc) of human IgG1, which also has potential anti-angiogenesis activity.[97]

Table 1.1. Small molecule PKIs currently approved by FDA. [49,98-99]

Name	Manufacturer and FDA approval date	Approved for	Target
Imatinib (Gleevec®)	Novartis 2001	CML, ALL, GIST and some others	BCR-ABL,c- KIT, PDGFR
Gefitinib (Iressa®)	AstraZeneca 2003	NSCLC	EGFR
Erlotinib (Tarceva®)	Genentech/ Roche 2004	NSCLC, PC	EGFR
Sorafenib (Nexevar®)	Bayer 2005	HCC, RCC	B-RAF, C-RAF, VEGFR-2,3, PDGFRβ, c-KIT, FLT3
Dasatinib (Sprycel®)	BMS 2006	CML, ALL	BCR-ABL, SRC- family, PDGFR β, c-KIT
Sunitinib (Sutent®)	Pfizer 2006	GIST, RCC	VEGFR-1,2,3, PDGFR α,β, c-KIT, FLT-3, RET
Nilotinib (Tasigna®)	Novartis 2007	CML	BCR-ABL, c-KIT, PDGFR α, β
Lapatinib (Tykerb®)	GSK 2007	Breast cancer	EGFR, HER-1,2
Pazopanib (Votrient®)	GSK 2012	RCC	VEGFR-1,2,3
Everolimus (Afinitor®)	Novartis 2012	Breast cancer	mTOR,
Bosutinib (Bosulif®)	Pfizer 2012	CML	BCR-ABL, SRC-family
Axatinib (Inlyta®)	Pfizer 2012	Adv RCC	VEGFR-1,2,3, PDGFR, c-KIT
Ziv-aflibercept (Zaltrap®)	Sanofi Aventis 2012	Colorectal cancer	VEGFR-1,2, PLGF

Note: CML – Chronic Myelogenous Leukemia, ALL – Acute Lymphoblastic Leukemia, MDS – Myelodysplastic Syndromes, GIST – Gastrointestinal Stromal Tumours, BCR-ABL – Break point Cluster Region Abelson, NSCLC – Non-small cell lung cancer, EGFR – Epidermal Growth Factor Receptor, RCC – Renal Cell Carcinoma, HCC – Hepatocellular Carcinoma, PC – Pancreatic Cancer, PDGFR – Platelet-derived Growth Factor Receptor, VEGFR – Vascular Endothelial Growth Factor Receptor, RET – Rearranged during Transfection, FLT3 – Fms-like tyrosine kinase, mTOR – Mammalian Target Of Rapamycin, PLGF – Placental Growth Factor,

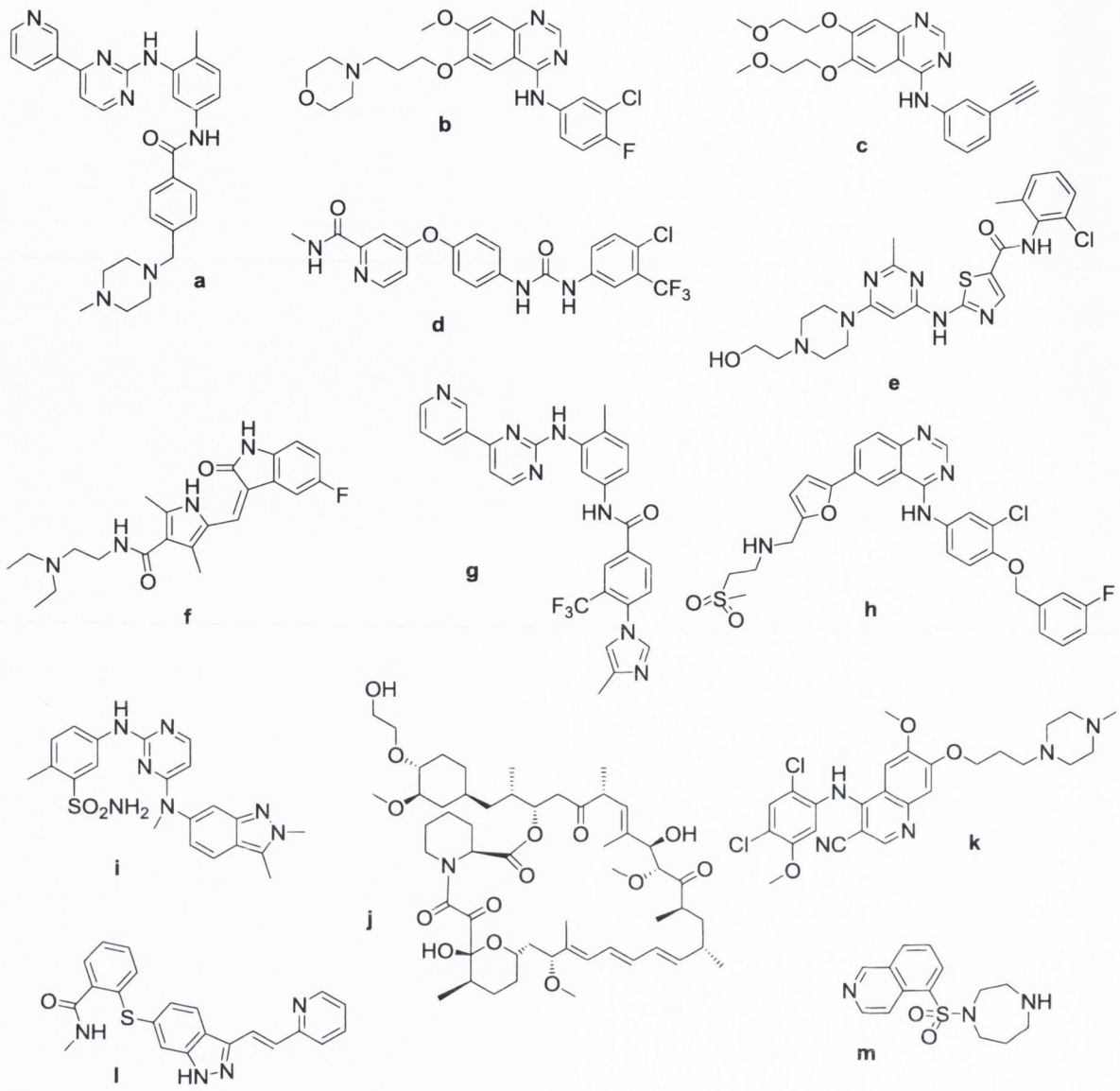


Figure 1.23. Structures of kinase inhibitors: (a) imatinib, (b) gefitinib, (c) erlotinib, (d) sorafenib, (e) dasatinib, (f) sunitinib, (g) nilotinib, (h) lapatinib, (i) pazopanib, (j) everolimus, (k) bosutinib, (l) axatinib, (m) fasudil.

Looking at all of these FDA approved kinase inhibitors together, it is obvious to wonder why there are so many commercialised drugs with the same target and what the improvements of the newest drugs show over those first released to the market. For example, if we compare the drugs targeting BCR-ABL in CML, such as imatinib, dasatinib, nilotinib and bosutinib (Figure 1.23 a, e, g, k), the former was the first potent inhibitor that dramatically altered the natural course of the disease. However, some patients

developed resistance to imatinib therapy and for this reason, dasatinib, nilotinib and bosutinib were developed. In contrast to imatinib, which binds only to the inactive form of the ABL kinase domain, dasatinib can bind to both the active and inactive forms, being active against many imatinib resistant kinase domain mutations of BCR-ABL.[100] Furthermore, nilotinib was developed through rational modifications of imatinib. The result is a better topological fit to the receptor that inhibits ABL 30-fold more potently than imatinib. The key design strategy for nilotinib was to replace the basic, polar 4-(methylpiperazinyl)-toluamide pharmacophore element in imatinib, with less polar, substituted phenylanilide, with the reasoning being that the imidazole would maintain the H-bond interactions, while the substituents would allow the exploration of new binding interactions.[101] The recently approved ABL inhibitor, bosutinib, is also an alternative to solving the imatinib-resistance problem, due to the insertion of a 4-anilinoquinoline group.[102] Additionally, the nitrile group introduced in bosutinib can be used to study electrostatic differences in the ATP-binding sites of kinases, allowing for the rational design of more selective inhibitors in the future.

From all EGFR inhibitors, gefitinib (Figure 1.23 b) was the first oral one to come on the market for the treatment of NSCLC. The Trial709 study by ISEL (iressa survival evaluation in lung cancer) showed that gefitinib did not confer an overall survival advantage.[103] Erlotinib (Figure 1.23 c) and gefitinib are drugs with differing efficacy; although structurally related (same quinazoline core) and with the same ability to interfere with the EGFR-mediated cascade, erlotinib and gefitinib have some pharmacokinetic differences which could account for the observed discrepancies in activity. Erlotinib is, for example, less susceptible than gefitinib to the metabolism mediated by cytochrome P450 and this could account for its lower clearance. In addition, erlotinib inhibits the activity of wild type EGFR at lower concentrations than those needed by gefitinib.[104]

Then, in 2007, lapatinib (Figure 1.23 h) was approved as it binds more effectively to EGFR. This is because lapatinib has a slower off-rate (K_{off}) than erlotinib and genitinib.[105] In other words, the half-life of the complex of lapatinib with EGFR is higher. This could be due to lapatinib being a type II inhibitor that binds to the inactive form of the kinase, and erlotinib being a type I inhibitor that binds to the active form only. The difference is that type II inhibitors – besides binding to the hinge region – also bind to the deep pocket of the kinase, increasing the potency of the interaction. This is correlated with a prolonged

downregulation of EGFR phosphorylation in tumour cells, halting the constitutive proliferation of cancer cells.

In the case of VEGFR inhibitors, most patients receive sunitinib (Figure 1.23 f) as first-line treatment, but it is possible that the disease progresses during initial VEGFR-TKI therapy. Therefore, there is a need to continue the treatment with a different VEGFR inhibitor, or alternatively the mechanism of action of treatment may be switched to an mTOR inhibitor. Sunitinib, sorafenib, pazopanib (Figure 1.23 i) and the monoclonal antibody bevacizumab, are not considered suitable options for use as second-line therapies following failure of an initial VEGFR-TKI, according to recent studies.[106] The recently approved axitinib (Figure 1.23 l) and the mTOR inhibitor, everolimus (Figure 1.23 j) are the most commonly used options of second line treatment for progressing patients. Additionally, just on the 3rd of August, 2012 the FDA approved a combination of ziv-aflibercept (a recombinant protein) and 5-fluorouracil for the treatment of patients who have had problems of resistance in previous treatments.[107]

As can be seen in Table 1.1, the therapeutic application of all of these TK and Ser/Thr kinase inhibitors is centred on cancer. However, efforts have been increasingly directed to other areas, such as inflammatory conditions, autoimmune and cardiovascular diseases and viral infections. To evolve in the development of PKIs, the discovery of novel, potent and selective inhibitors which target the unexplored kinases need to be investigated. Ultimately, the goal of basic research and clinical trials should be to identify new targets and agents which can lead to complete remission of diseases rather than to patent extension. To achieve this, it is necessary that scientists and pharmaceutical companies work together and share knowledge to explore the diversity of the kinome. This year, several conferences have taken place to reveal some of the information that industry possesses about untargeted kinases so the entire scientific community can access it. With an ever-growing range of kinase inhibitor chemotypes, broad assay panels and combining medicinal chemistry, screening, computational chemistry and X-ray crystallography, scientists are well positioned to identify molecules that target the unexplored kinases and potentially discover a new generation of kinase inhibitors for current and future diseases.

1.5.4. Targeting the Active *versus* the Inactive Form

Kinases are typically regulated and undergo an activation process in which the inactive state is converted to an active state. This process of activation involves a structural change, meaning that both forms are sufficiently different as to be targeted selectively. Most protein kinases share a DFG motif in the ATP site that can assume two distinct conformations, as we previously described, the active DFG-in and the inactive DFG-out states. Generally, active kinases have high structural homology, whereas the inactive forms have much more structural diversity.[58] Therefore, if specific inhibitors are sought, it would be preferable to target the inactive form; however, this form is more susceptible to mutations. Some inhibitors of the active kinase form (*e.g.* the EGFR inhibitors erlotinib and genitinib) show quite high specificity for their target and, conversely, some inhibitors targeting the inactive conformation (*e.g.* the ABL inhibitor imatinib) show significant cross-reactivity with other kinases.

Taking everything into consideration, there is not a standard preference for targeting the active or inactive form of the kinase, so all these described factors have to be taken into consideration in the design of new PKIs.

1.5.5. ATP-Competitive *versus* Non-Competitive Inhibitors

Although most kinase inhibitors described to date are ATP-competitive – *i.e.* they bind to the ATP binding site – there is increasing interest in the development of kinase allosteric inhibitors which are non-ATP competitive and bind outside of that binding site, commonly referred to as allosteric inhibitors. The drawback of ATP-competitive inhibitors is that the ATP binding site is highly conserved among the kinase family members, increasing the possibility of cross-reactivity with other kinases. However, the non-competitive inhibitors may provide higher specificity and improved therapeutic efficacy. Their mechanisms of action varied, with some being competitive with the target's protein substrates, some binding to the substrate itself and inhibiting its interaction with an upstream activating kinase, and others binding to a site outside of the ATP pocket, causing conformational changes at a second, distal site, which induce an inactive conformation.[108]

A way to verify the best way of targeting a certain protein is by checking its Michaelis constant (K_M^{ATP}). This is a parameter that measures the ATP binding affinity of a

kinase,[109] varying from 1 μM to 1000 μM. When a kinase has high K_M^{ATP} value, ATP binding is weak and it will be easier to target its binding site through ATP-competitive compounds. However, if the K_M^{ATP} value is low it might be worthwhile considering a screen for non ATP-competitive compounds. The high intracellular concentration of ATP represents a challenge for ATP-competitive small molecules, which must compete effectively with ATP for binding to the target. A table with different K_M^{ATP} values of different kinases can be found in the literature.[109] For example, the K_M^{ATP} values of MEK-1 and RAF-1 are 5.6 and 11.6 μM, respectively. They are quite low values in comparison to the K_M^{ATP} of mTOR of 1000 μM; hence they are more suited to non-competitive inhibition. Promising data exists which suggests that the use of an allosteric kinase inhibitor in combination with an ATP competitive inhibitor of the same target can improve efficacy and may help to overcome resistance to either compound alone.[110]

1.5.6. Specific *versus* Multi-Targeted Inhibitors

Since the beginning of the development of kinase inhibitors, the issue of whether to target one or many kinases has received much attention. The selectivity profile has important applications for both efficacy and safety. From a purely academic point of view, highly specific inhibitors allow more precise manipulation of intra-cellular signalling pathways. However, even if an inhibitor is highly specific for a single kinase, it should be noted that nearly all small molecules can have activity against other pharmacological targets.[58]

However, most cancers have multiple aberrations in kinase signalling and this necessitates inhibition of several signalling pathways in order to achieve therapeutic benefits. There may also be unintentional, but clinically useful, off-target activity for multi-target kinase inhibitors such as the inhibition of KIT by imatinib. However, it must be kept in mind that there are drawbacks associated with multi-targeted inhibitors with respect to specific inhibitors, such as several side effects and less control of the pharmacodynamics.

Figure 1.24 illustrates the results obtained from a study[111] where the target selectivity of some of the most clinically used kinase inhibitors was evaluated. In this study, each compound was screened against a panel of 287 human kinases at a single concentration of 10 μM to identify candidate kinase targets, and for each interaction observed in this primary screen a quantitative dissociation constant (K_d) was determined.[111] The K_d is

calculated as the K_{off}/K_{on} ratio, where K_{off} is the dissociation constant and K_{on} is the formation constant of the interaction between the kinase and the compound under investigation. Only interactions with $K_d < 3$ μM are shown in the diagram, with a red dot. The selectivity of the kinase inhibitors ranges from very high (*e.g* sunitinib or erlotinib) to relatively promiscuous (*e.g.* sorafenib, dasatinib or imatinib). Staurosporine is the most promiscuous kinase inhibitor of all, interacting with nearly every kind of kinase (see structures of all the PKI mentioned in Figure 1.23 and staurosporine structure in 1.24 top)

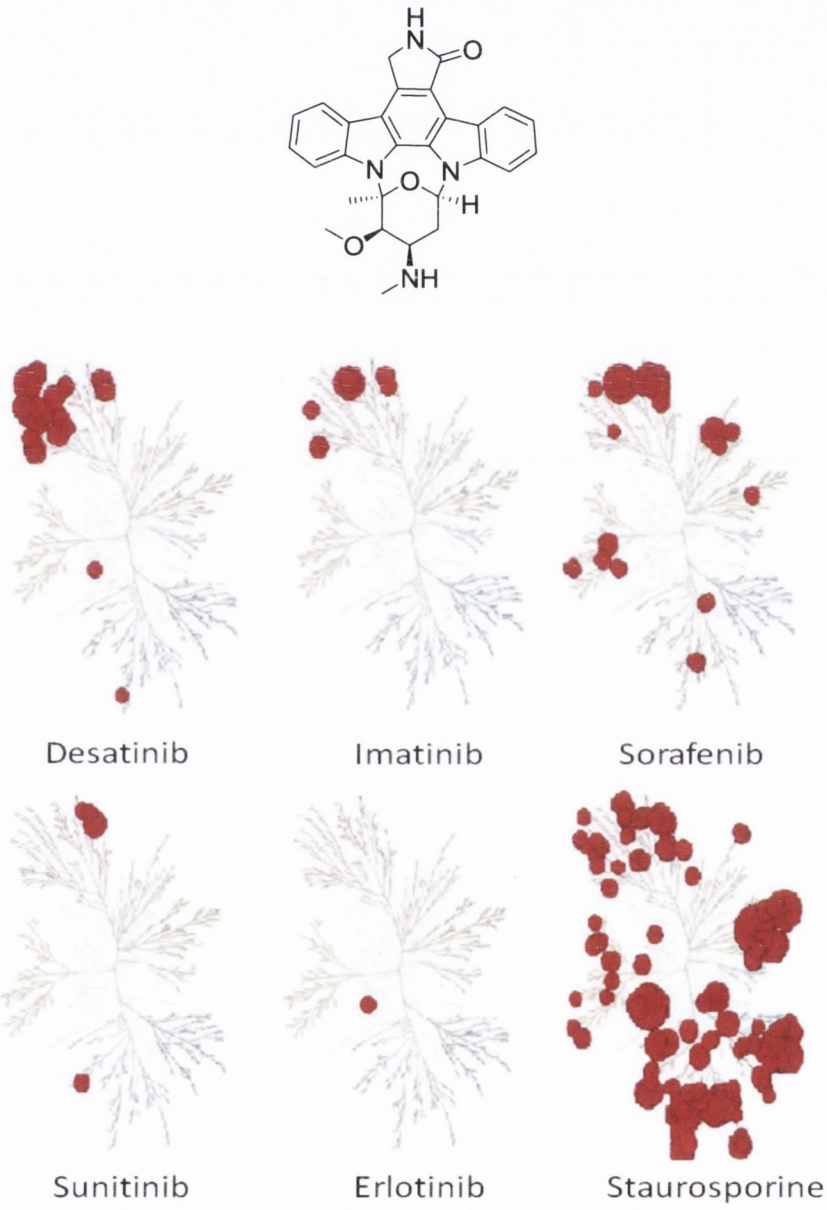


Figure 1.24. Staurosporine structure (*top*), small molecule-kinase interaction maps for dasatinib, imatinib, sorafenib, sunitinib, erlotinib and staurosporine kinase inhibitors (*bottom*). Structures shown in Figure 1.23.[111]

As an example of K_d values, the K_d of sorafenib for B-RAF is 540 nM and 59 nM for VEGFR-2. However, a more selective inhibitor erlotinib has a K_d of 0.67 nM for EGFR.[111]

Potency and selectivity could be addressed through optimising binding kinetics. In order to do so, the binding mode of the drug with the kinase should be taken into account, to allow for optimisation of target residence-times (the lifetime of the drug–target complex). Conformational adaptation can greatly influence the residence time of a drug on its target; long residence times can lead to sustained pharmacology and may also mitigate off-target toxicity, helping selectivity and efficacy.[112] In other words, more stable drug-target complexes lead to slower off-rates (K_{off}) and, hence, longer half-lives. As a result, the K_d decreases, generating a more selective inhibition of the kinase. This can be achieved by forming covalent interactions between the drug and the kinase target, as is the case with irreversible inhibitors.[50] Use of such inhibitors has given promising results with respect to overcoming tumour resistance and competition within class selectivity, both of which are encountered with reversible TK inhibitors.

Another way to measure the selectivity of a kinase inhibitor against a family of kinases is by using the Gini coefficient, in which the sum of all the percentage inhibition values is calculated as a function of the cumulative fraction of kinases tested (Figure 1.25).[113] A completely unselective inhibitor will occupy the diagonal line of the diagram and its Gini coefficient will be close to zero (*e.g* staurosporine), whereas a highly selective inhibitor will show a Gini coefficient close to 1 and its profile will show a dramatic increase in cumulative inhibition for a few kinases at the far right of the graph (*e.g.* PD184352).

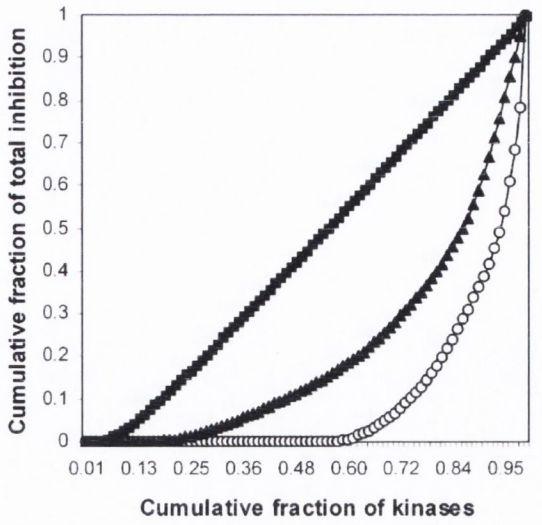


Figure 1.25. Gini coeficient of a non-selective multi-target inhibitor (staurosporine, diagonal), a highly selective inhibitor (PD184352, white circles) and moderate selective inhibitor (AG1024, black triangles).[113]

Since the "founder of chemotherapy", Paul Ehrlich, described drugs as molecules that would eliminate disease *precisely* and *efficiently* by aiming to a specific target as a "magic bullet", oncologists have been trying to find the new generation of target-specific PKIs.[114] Unfortunately, there are several reasons indicating that the existence of a magic bullet is not possible. One such reason is that a selective inhibitor cannot target just one mutated kinase pathway because there are lots of different kinases and they are often interrelated. Moreover, a general drawback of target-specific monotherapy derivatives is that a single genetic alteration is sufficient to confer target resistance to an individual tumour cell, and this can eventually lead to relapse of the disease. The ability of kinases to mutate in response to the selective pressure created by drug treatment provides a strong rationale for hitting more than one essential target at the same time in the tumour cells.

Additionally, cancer is a complex disease formed of several aberrations, so there is not a magic bullet that can target all the abnormalities brought about by the cancer at once, without side effects or developing some kind of drug-resistance. For these reasons, nowadays the idea of multi-targeted therapy is achieved with either combinations of specific drugs or a single promiscuous drug that acts on a set of disease-relevant proteins only. The latter strategy can be quite unselective, generating off-target effects and toxicity. The combined therapy approach will be described in more detail in section 1.8.

1.5.7. MAPK Inhibitors

Due to the diverse and important roles of the MAPK pathways in inflammatory responses and carcinogenesis (see section 1.3.3.1), it has been subjected to intensive research in an effort to identify novel inhibitors for cancer treatment. Among them, sorafenib (RAF inhibitor), PD184352, PD0325901 and ARRY-142886 (MEK1/2 inhibitors) have reached the market or clinical trial stages.[115] MEK inhibitors may be even more efficacious than inhibitors of other kinases upstream, since they are solely dependent on proliferative signals to drive tumourogenesis. Intriguingly, no inhibitors of ERK1 or ERK2 have been described so far.

Given the roles of inflammation in promoting tumour initiation and progression, many pharmaceutical companies have carried out screens for compounds that block cytokine production. As a result, many compounds have been found to inhibit p38 with high selectivity; however, very few amongst them have progressed beyond Phase I clinical trials, largely due to toxic side effects.[115]

The diverse functions of JNK in proliferation and cell death are also targets for therapies. A small compound (CC-930) that binds to the JNK ATP-pocket is under clinical trials for treatment of lung fibrosis disease by Celgene.[116] On the other hand, as the MAPK signalling pathways crosstalk with each other, it should be taken into consideration whether inhibiting a single MAPK gene or a single MAPK pathway will result in unexpected outcomes. A solution to this challenge may lie in a combination of two drugs, which target specifically one gene or pathway.[75] Such a combined use would have a more effective outcome for cancer therapy.

1.5.8. Tyrosine Kinases Inhibitors (TKIs)

As has already been mentioned, TKs are associated with the cytoplasmic domains of growth factor receptors (RTK) as well as oncoproteins (TK). They have the potential to cause cell proliferation dysregulation if mutated or hyper-expressed. For this reason, they represent an excellent target for the development of anti-cancer drugs. It has been shown that TKIs are more likely to show cross-inhibition of intimately related kinases than Ser/Thr kinase inhibitors.[117] This is because the binding site of TKs is closely related, generating unexpected off-target interactions. However, a large number of inhibitors have

been identified and many show promising cytostatic activity. There are different types of TKIs, for example those that inhibit BCR-ABL, SRC, JAK2, EGFR, IGF1R, FLT-3 and KIT and receptor tyrosine kinase inhibitors (RTKIs) such as those of VEGFR and PDGFR. Table 1.1 contains a list of the currently FDA approved and marketed TKIs with their targets.

1.6. Sorafenib (Nexavar®, BAY-43-9006)

1.6.1. Discovery and Development of Sorafenib

Sorafenib was approved by the FDA in December 2005 for the treatment of advanced renal cell carcinoma (RCC) and hepatocellular carcinoma (HCC).[118] As a disease, RCC represents approximately 2% of all adult malignant tumours and is the sixth leading cause of cancer.[119-120] Also, HCC is the sixth most common cancer and the third leading cause of cancer death in the world.[121]

The number of individuals diagnosed with RCC and HCC has dramatically increased over the past five decades. Several factors may explain this rise including significant improvements in diagnosis and increased occurrence of other issues like smoking, hypertension, obesity and diet, which have contributed to the rising incidence of these cancers.[122]

Nowadays, there are different drugs on the market for the treatment of RCC and HCC; for example, bevacizumab, sunitinib, pazopanib and sorafenib.[123] The discovery and development of sorafenib took place over 11 years, from initial lead compound to approval (see timeline in Figure 1.26). The results from the Phase II and III trials established orally administered sorafenib as a safe and effective treatment for metastatic RCC. Sorafenib treatment significantly prolonged the median progression-free survival; four-fold compared to placebo treatment.[124] The most common side effects were dermatological, gastrointestinal and fatigue related. Currently ongoing are Phase I, II and III trials of sorafenib as monotherapy or in combination with various chemotherapeutic regimens in many different cancers including melanoma, hepatocellular, pancreatic, biliary, renal, breast, glioblastoma, prostate, lymphoma, multiple myeloma, AML, myelodysplastic syndromes, NSCLC, colorectal and neuroendocrine.[58]

Sorafenib is a multikinase inhibitor which stops tumour growth and disrupts tumour microvasculature through antiproliferative, antiangiogenic and pro-apoptotic effects.[125] Anticancer agents with multiple targets might be more effective against cancers with multiple molecular drivers, such as RCC and HCC, and it could be less likely that resistance to these drugs will develop in tumours than to agents targeting a single pathway.

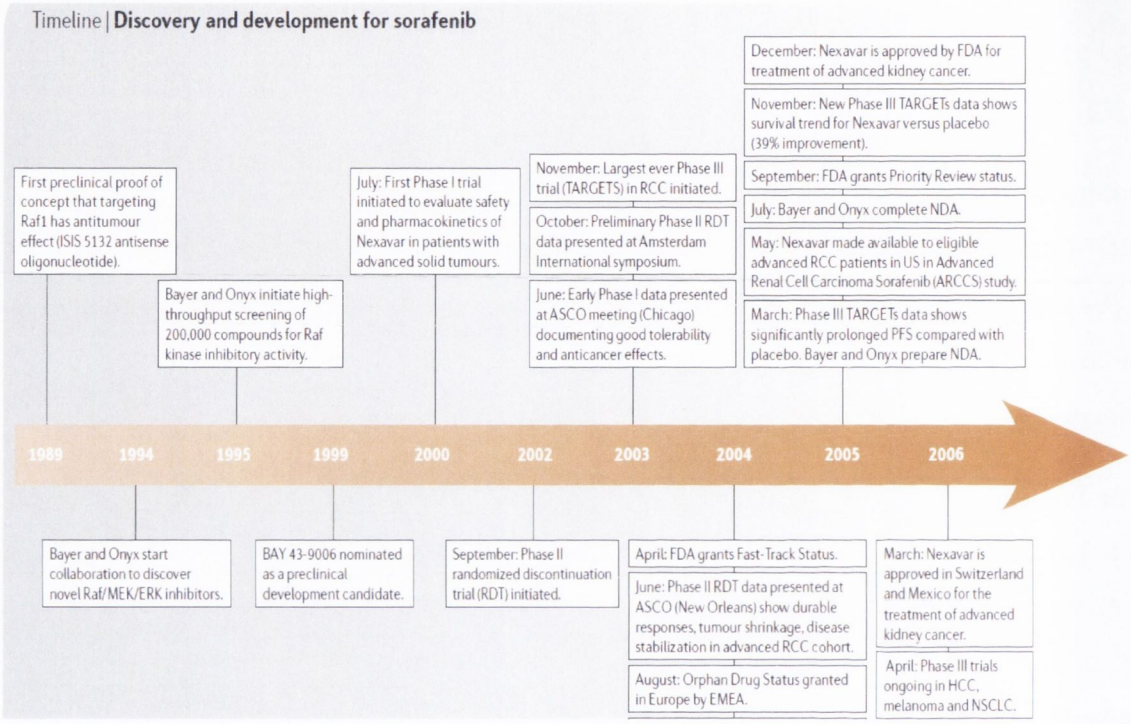


Figure 1.26. Timeline of the discovery and development of sorafenib.[118]

It is conceivable that sorafenib's clinical activity is due to its capacity to inhibit multiple kinases involved in ubiquitous signalling pathways, which are dysregulated in cancer. With its novel mechanism of action, favourable safety profile and combinability demonstrated so far (with several standard chemotherapeutic agents such as carboplatin, 5-fluorouracil -5-FU- or paclitaxel), sorafenib has potential clinical utility in a broad range of tumour types.

1.6.2. Mode of Action

Sorafenib is the first oral multi-kinase inhibitor that targets Ser/Thr kinases and RTKs in both the tumour cell and tumour vasculature. In pre-clinical models sorafenib inhibited tumour cell proliferation through targeting of the RAF/MEK/ERK pathway at the level of the Ser/Thr kinase, RAF. In tumour cells sorafenib also targets the C-KIT and FLT-3 receptors.[118]

Additionally, in preclinical models, sorafenib also exerted an anti-angiogenic effect on the blood vessels that feed cancer cells with nutrients and oxygen; hence, stopping the growth of the tumour. Anti-angiogenic drugs are cytostatic as opposed to cytotoxic agents and, hence, they inhibit tumour growth by depleting the essentials required for its survival.[126] The angiogenic process starts with the activation of blood vessels near the tumour by cell survival growth factors such as VEGF and PDGF. This process leads to self-growth of blood vessels, migration and tumour formation. In preclinical models, sorafenib demonstrated inhibition of angiogenesis through the blockage of the RTKs VEGFR 1,2 and PDGFR-β, and their associated signalling cascades including the RAF/MEK/ERK pathway, which helps to slow or prevent the growth of blood vessels.[118]

Sorafenib targets multiple PK *in vitro* and *in vivo.* It has been identified as a potent inhibitor of the RAF Ser/Thr kinase isoforms, with an order of potency based on their IC_{50} values (half maximal inhibitory concentration) of RAF-1 or C-RAF > B-RAF > oncogenic ^{V600E} B-RAF (mutated).[127,128] In biochemical assays, sorafenib has been reported to inhibit C-RAF, B-RAF, ^{V600E} B-RAF, VEGFR1,2,3, PDGFR β, FLT-3, c-Kit and p38 MAPK with IC_{50} values of 6, 22, 38, 26, 90, 20, 57, 33, 68 and 38 nM, respectively.[118,127]Also of note, the compound has low to no inhibitory activity on ERK1, MEK1, EGFR, HER2, PKB, PKA, CDK1/cyclin B and PKCα,β.[127]

By targeting RAF, the RAF/MEK/ERK pathway is downregulated because the phosphorylation cascade is interrupted and ERK cannot phosphorylate cytosolic and nuclear substrates, thereby interrupting the proliferation of cancer cells.[125] The inhibition of this pathway has been observed in kidney, liver, colon, pancreatic and breast tumour cell lines expressing wild type or mutant B-RAF.[127] However, the activity of sorafenib in melanoma, which is governed by mutations in the oncogenic B-RAF, is currently unclear. Sorafenib can inhibit ^{V600E} B-RAF *in vitro*, while it cannot do so *in vivo*, establishing that sorafenib does not target ^{V600E} B-RAF directly but actually inhibits the tumour driven by

this protein.[129] It has been shown that sorafenib inhibits ^{V600E} B-RAF by driving it into a complex with C-RAF but this does not occur *in vivo*.[130]

The mechanism of action of sorafenib is illustrated in Figure 1.27.

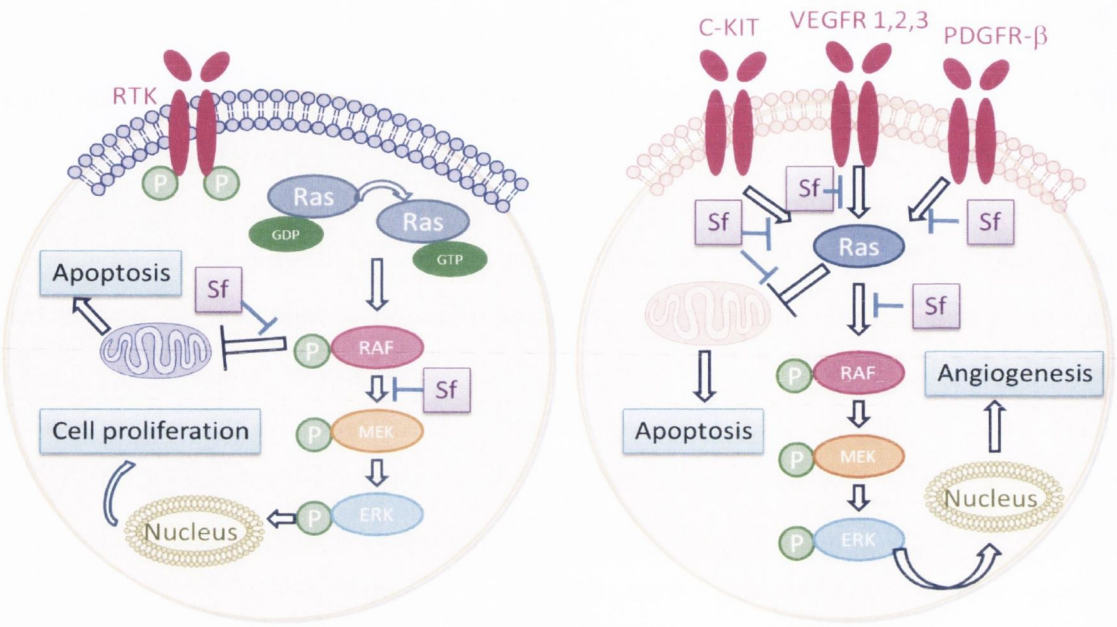


Figure 1.27. Mechanism of action of sorafenib (Sf) in the tumour cell (*left*) and tumour vasculature (*right*).

Other recently discovered targets of sorafenib found in the literature show that it inhibits the anti-apoptotic protein MCL-1 (myeloid cell leukaemia sequence 1) and AKT phosphorylation in prostate cancer.[131] Also, sorafenib is shown to enhance apoptosis in the HCC cell line HepG2, induced by TRAIL through downregulating MCL-1 expression.[132]

STAT3 (Signal transducers and activators of transcription 3) is constitutively active in human pancreatic cancer cells, 50% of AML and other cancers, and can promote cell growth and apoptosis resistance that contribute to tumourogenesis. It has been proven that sorafenib inhibits STAT-3 activation, inducing apoptosis in pancreatic cancer cells.[133] Additionally, sorafenib blocks SRC kinase-mediated STAT-3 phosphorylation and decreases the expression of the apoptosis regulatory proteins MCL-1 and BCL-2, which are associated with increased apoptosis in HL60 cells.[134]

1.6.3. Chemical Interactions between Sorafenib and B-RAF

Sorafenib targets B-RAF in a type II TK inhibitor manner because it attaches to the ATP binding site of the kinase by forming four HBs and hydrophobic interactions to the allosteric site of the domain in the "out" form (Asp, Phe, Gly). The pyridyl ring occupies the ATP binding pocket, interacting with three aromatic residues of B-RAF: Trp530 (in the hinge region), Phe582 (at the catalytic loop) and Phe594 (in the DFG motif). At the opposite end of the molecule, the 3-trifluoromethyl-4-chlorophenyl ring interacts with the hydrophobic pocket between the DFG motif and the catalytic loop. Moreover, the central ring contacts with the aliphatic chains of Lys482, Leu513 and Thr528. All these van der Waals interactions are less significant than the four HBs formed: (i) between the N-H of the urea group and the carboxylate chain of the catalytic Glu500 residue, (ii) between the carbonyl moiety and the Asp593 N-H of the DFG motif, (iii) between the N atom of the pyridyl ring and an N-H of Cys531 of the hinge region, and (iv) the *N*-methylcarbamoyl group contacting with the carbonyl of Cys531 in a bidentate fashion.[128] These interactions can be seen in Figure 1.28.

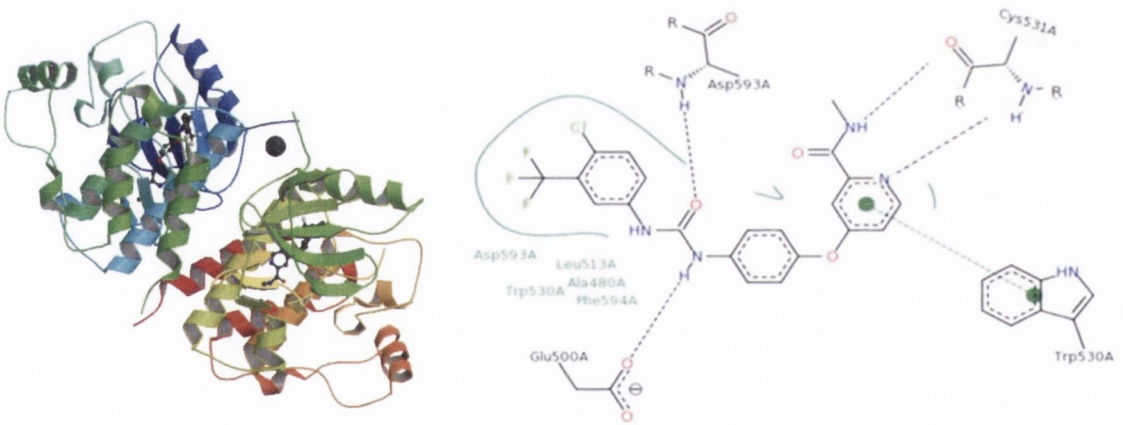
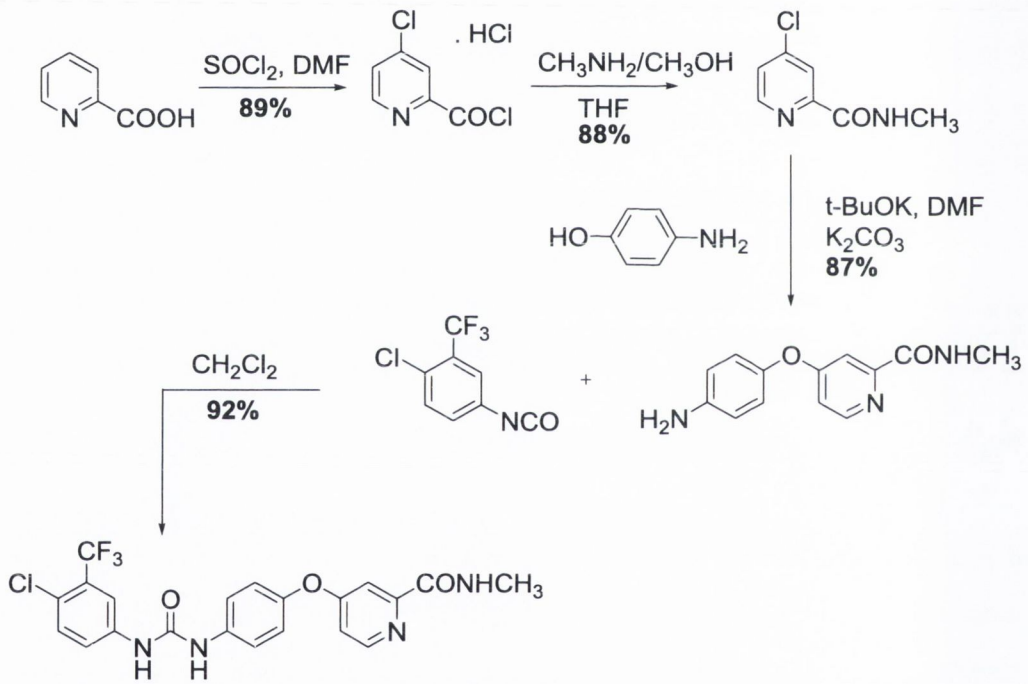


Figure 1.28. Sorafenib molecule crystallised with wild type B-RAF (*left*) and the ligand's interactions with the kinase shown in PDB 1UWH (*right*).[128]

1.6.4. Original Synthesis of Sorafenib

Originally, sorafenib was synthesised in four steps with an overall yield of 63%.[135-136] The first step consists of the reaction of picolinic acid with thionyl chloride in DMF to give 4-chloropyridine-2-carbonyl chloride. This was then directly converted into the amide by reaction with a solution of methylamine. Nucleophilic aromatic substitution of this with 4-aminophenol in the presence of potassium *tert*-butoxide in DMF yields the pyridyloxyaniline, which is finally condensed with 4-chloro-3-trifluoromethylphenyl isocyanate in CH_2Cl_2 to obtain sorafenib.

Scheme 1.1.



Subsequently, a plethora of different protocols for the synthesis of sorafenib was published. Some examples are the following: (i) unsymmetrical *N*,*N*'-diarylureas and the utilisation of a Pd-catalyzed reaction for the arylation of ureas were used to afford sorafenib in 86% yield over two steps,[137] (ii) unsymmetrical *bis*-aryl ureas from *N*-carbamoylimidazoles were used to obtain sorafenib over two steps in 50% yield,[138] (iii) attack of diarylcarbamates by substituted amines to generate the urea-based compound of sorafenib,[139] (iv) halogenation, amidation and etherification of picolinic acid, as starting

material, after which sorafenib was obtained by urea formation in 56% yield,[140] (v) sorafenib tosylate was synthesised from 2-picolinic acid by chlorination, amidation, nucleophilic substitution to obtain 4-(4-aminophenoxy)-2-(methylcarbamoyl)-pyridine, which was subjected to condensation with 4-chloro-3-(trifluoromethyl)-phenylisocyanate and salt formation in 64% yield.[141]

1.6.5. Limitations of Sorafenib

Sorafenib has some significant drawbacks, including modest, transient benefits, and toxicity challenges. Many clinical studies have indicated several limitations to the application of sorafenib as a single agent in various cancers. Treatment with sorafenib is not associated with complete responses, a second line treatment being necessary, or administration in combination with other anti-cancer agents.[142] Owing to these reasons and the potential of sorafenib to synergise with other anticancer therapies, its combination with other targeted agents and chemotherapy has been widely explored, with promising results.[143] In addition, it has also shown synergistic results when combined with radiation.[143]

Although sorafenib represents a much needed treatment option for patients with HCC, it also produces toxicities that may significantly affect patients' quality of life. High rates of dermatologic side effects are commonly reported with sorafenib, the most clinically significant being hand-foot skin reaction. Like other angiogenesis inhibitors, sorafenib also has known side effects, including skin-related toxicities, hypertension, proteinuria, diarrhoea, and an increased risk of thromboembolism and bleeding events.[144]

For these reasons, there is a need for the development of new analogues of sorafenib that improve the activity and decrease the risk of side effects.

1.7. Future and Challenges of Kinase Drug Discovery.

Since the launch of imatinib (Gleevec®) in 2001, a torrent of new small molecule kinase inhibitors have appeared. The development of these novel drugs has opened new routes and designs to novel therapies for the treatment of cancer. Several clinical studies have demonstrated that kinase inhibitors have improved the survival and quality of life for cancer patients.

Several kinase inhibitors are now approved by the FDA and regulatory agencies across the world. However, considering the numerous potential kinase inhibitors which have appeared, and all the research work and investment, only a limited number of small molecules have received marketing approval. This fact reveals that there are still some challenges in the design of PKIs that need to be addressed in the future. To this end, some relevant topics have been selected to illustrate current challenges, opportunities and emerging strategies in the development of small molecule kinase inhibitors.

Firstly, new drugs need to be prepared to ***selectively*** bind the target without causing off-target effects, while at the same time targeting all the abnormal pathways that originate in cancer. The selectivity problem can be solved by using irreversible kinase inhibitors that bind covalently to amino acids (Cys, Lys, Thr, Ser or Tyr) in the ATP binding cleft, or suitable allosteric pockets. Nowadays, there are several irreversible inhibitors, including afatinib and neratinib, both EGFR inhibitors and PCI-32765, a Bruton's TK (Btk) inhibitor.[49] All of these are currently undergoing clinical testing. As can be observed in Figure 1.29, all their structures contain electrophilic functionalities based on acrylamides or derivatives, which can be attacked by the nucleophilic moieties of the amino acids mentioned before.

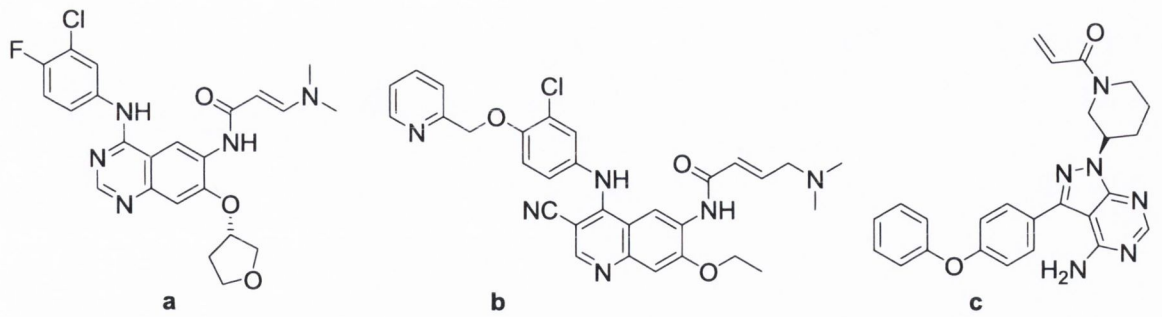


Figure 1.293. Irreversible inhibitors (a) afatinib, (b) neratinib and (c) PCI-32765.

Therefore, irreversible kinase inhibitors are a possible route to minimising promiscuous binding and achieving a high level of selectivity over related pharmacological targets. The selectivity challenge might also be addressed using proteomic strategies, which measure the affinity of each drug for a panel of kinases using immobilised kinase inhibitors, followed by sensitive mass spectrometry for subsequent protein identification.[90] In this manner, all the targets of each inhibitor can be identified to avoid unexpected off-target activity.

Additionally, selectivity can be improved by increasing the life time of the drug-target complex, in order to allow a complete blockade of the enzymatic activity to take place. This issue was already discussed in section 1.4.3. Basically, minimising the dissociation rate of the drug-target complex is the most practical way to approach this medicinal chemistry objective. Irreversible inhibitors can be an example of this approach.

Since more information has emerged about the causes of cancer and which type of mutations occur in the different diseases, another way to improve selectivity is by targeting particularly the mutated kinase.

Secondly, intrinsic or acquired drug ***resistance*** is also emerging as a major challenge. However, the identification of several resistance mechanisms has opened the door to alternatives to overcoming this problem. To defeat such liabilities and considering that some inhibitors cannot be utilised as monotherapy, some solutions have been proposed: (i) improvements of the pharmacological properties of the inhibitors, or (ii) using combination therapies with other chemotherapeutic agents.

To improve the pharmacological properties of these drugs, plasma exposure levels, dosing and schedule of the target agent need to be considered, so that individual responses can be maximised, while treatment-associated adverse effects are minimised.

Moreover, the principle behind the combination strategy consists of targeting tumour cells at multiple points so cancer cells that escape from a single cytotoxicity treatment might be less likely to simultaneously escape from two chemotherapeutic agents with different mechanisms of action. As an example, preclinical findings have provided evidence that dual PI3K/mTOR inhibitors may be less susceptible to PI3K drug resistance than selective PI3K targeted agents.[145] In order to select the right chemotherapeutic agents for the combination, preclinical data and clinical considerations need to be considered. One of the key criteria for selecting a combination is that any synergy should be selective for tumour cells and not exacerbate systemic toxicities towards normal cells. The dose level and schedule and the timing of the administration for each component of the combination therapy are important factors that may need to be optimised.

The experience gained over the past few years with molecularly targeted kinase agents suggests that interrelation of pathways, negative feedback loops and other resistance mechanisms are likely to require that kinase inhibitors will need to be administrated in combination with other agents. [146] To this end, a combination of kinase inhibitors in the clinic is a growing trend and could become commonplace in the near future.

Additionally, to advance in the discovery of new kinase inhibitors we need to move away from the "one-size-fits-all" ATP-competitive inhibitory approach or the "me too drugs" to a certain extend. A tailored strategy in which mechanism of action, activity and selectivity profile are adjusted to the target to be modulated and the disease to be treated should be utilised.

A further challenge to bear in mind for future drugs is keeping a correlation between exciting preclinical data *in vitro* and *in vivo* and application to humans, *i.e.* translational medicine. This strategy is based on the combination of the efforts of basic research, clinical research and patient-orientated research[147] (Figure 1.30). For many, the term refers to the "bench-to-bedside" enterprise of harnessing knowledge from basic sciences to produce new drugs and treatment options for patients.[148]

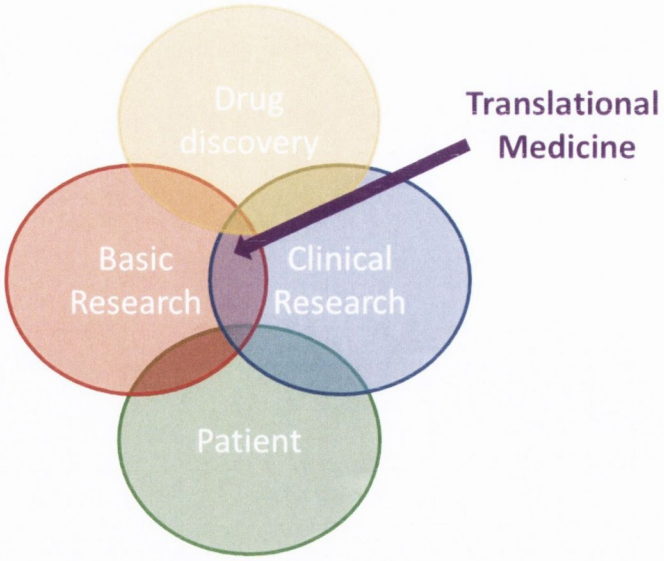


Figure 1.30. Translational medicine is based on a combination of basic, clinical and patient-orientated research, allowing for improvements in drug discovery.

One of the key challenges for translational research will be to use preclinical data in planning clinical studies, and then to capture positive and negative data from clinical samples in order to gain more insight into mechanisms of resistance or other failings. Then, returning to the laboratory to try to improve those drugs that did not work in real patients is essential, so the development of future drugs is superior. This can only be achieved by improving communication among clinical and basic scientists, so the public can benefit from modern, interactive and educated approaches to personalised health care.[147]

Innovative techniques would be crucial to overcome some of the discussed challenges in kinase lead generation. For example the 3D structure-based drug design techniques will be a key strategy to generate and/or optimise promising lead compounds. It is expected that future PK drug discovery will exploit a variety of high-powered *in silico* tools to computationally visualise, analyse and probe the molecular interactions of novel small molecule inhibitors with their desired therapeutic targets.[90] Integration of such *in silico* tools (*e.g.* molecular modelling, docking approaches and virtual screening) with structural biology (*e.g.* X-ray, and/or NMR studies) and sophisticated medicinal chemistry, chemoinformatics and bioinformatics provides great possibility for developing novel PK inhibitory drugs. The development of more advanced techniques, coupled with a better understanding of the molecular basis of cancer, protein structural data and compound

activity data, will allow for the discovery of new and improved therapeutics, providing scientists effectively use this new knowledge.

All the advances that have been made, and will be made in the future, will help the design of the next generation of PKIs. The hope is that these efforts will translate into increased survival rates for cancer patients, offering improved outcomes for patients and their families.

1.8. Previous Work Leading Up to the Present Research

In the past, research in the Rozas group was focused on the design, synthesis and biological evaluation of novel DNA minor groove binders (MGBs). It was found that several diphenyl 4,4'-*bis*-guanidinium and -2-iminoimidazolinidium derivatives (see Figure 1.31 top) showed correlation between their DNA binding affinity and their *in vitro* and *in vivo* antitrypanosomal activity, suggesting that DNA minor groove binding could be part of the mechanism of action of these compounds.[149,19]

Due to the success obtained in terms of DNA targeting with the aforementioned families of compounds, new families of 4,4-asymmetrical diaromatic guanidinium/2-iminoimidazolilidinum, isouronium and hydroxyguanidinium derivatives were synthesised (see Figure 1.31 bottom).[150,151] The affinity of these substrates for DNA was evaluated by means of DNA thermal denaturation measurements and other biophysical methods (circular and linear dichroism, isothermal calorimetry, UV titration methods, surface plasmon resonance), finding that and the some of these compounds bind to the DNA.[150]

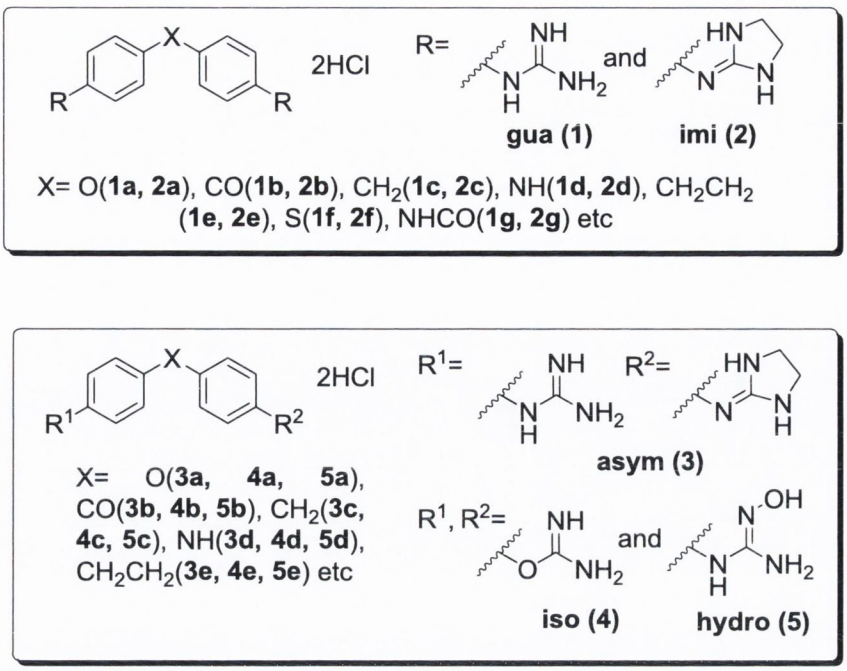


Figure 1.31. Previous families of compounds already developed in the Rozas group (*top*) 4,4'-*bis*-guanidinium and -2-aminoimidazolinium derivatives and (*bottom*) 4,4-asymmetrical diaromatic guanidinium/2-aminoimidazolinium, isouronim and hydroxyguanidinium derivatives.

These families produced encouraging results not only in terms of binding ability and affinity towards the minor groove of DNA but also in the initiation of anticancer activity through the induction of cellular apoptosis.[152, 153] However, some contradictions were encountered due to the fact that the most cytotoxic and apoptosis-inducing compounds (**1a** and **2a)** were the weakest minor groove binders and also some of the least biologically potent compounds bind the strongest to DNA (**1b** and **2b**).[154] This, suggests that the mechanism of action in inducing apoptosis of the former could take place by a different mechanism than by targeting DNA.

Looking for alternative possible mechanisms of action for those 4,4'-*bis*-guanidinium and -2-aminoimidazolinium derivatives, we noticed several structural similarities between these compounds and sorafenib, a protein kinase inhibitor: (i) both molecules have two aryl ring moieties connected by a linker (mostly heteroatomic), (ii) both have similar structural molecule shape, (iii) similar HB profile between urea and guanidine moieties and (iv) both have two HB donors (NH groups). Additionally, the guanidine group present in Rozas

derivatives is protonated at physiological pH; therefore, these compounds could form electrostatic interactions with negatively charged residues from the active site and also π-cation interactions with aromatic amino acids (interactions which have been extensively studied in the Rozas group).[155-158] These forces could substitute the HB acceptors from the urea or *N*-methylcarbamoyl groups in sorafenib and form stronger interactions with the active site (see Figure 1.32)

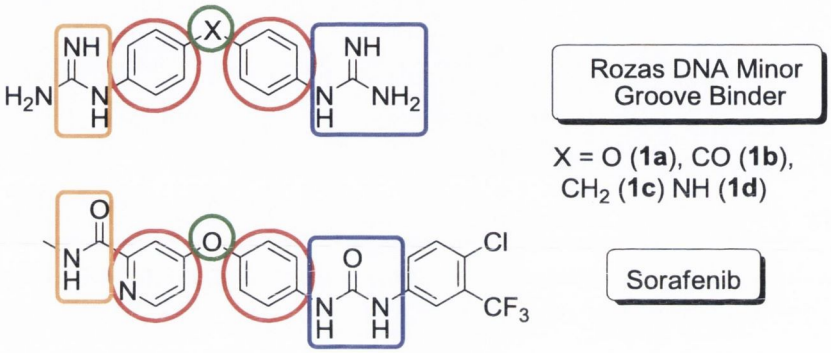


Figure 1.32. Structures of some of Rozas' DNA minor groove binders (*top*) and sorafenib (*bottom*) indicating the similar and related features.

Considering structural similarities of Rozas compounds and sorafenib, it was thought that these compounds could act as PKIs. For this reason, the present work involves the evaluation of these 4,4'-*bis*-guanidinium compounds as potential PKIs and the design of novel PKIs, guanidine-like derivatives as new PKIs.

1.9. References

(1) Pecorino, L. *Molecular biology of cancer*; Second ed.; Oxford university press, 2008.

(2) http://globocan.iarc.fr/.

(3) Saab, M.-B.; Estephan, E.; Martin, M.; Bec, N.; Larroque, C.; Cloitre, T.; Cuisinier, F.; Gergely, C. *Proc. SPIE* **2010**, *7715*, 771518.

(4) Hanahan, D.; Weinberg, R. A. *Cell* **2000**, *100*, 57-70.

(5) http://www.cancer.gov/cancertopics/understandingcancer/cancer/page4.

(6) Jacobson, M. D.; Carthy, N. M. *Apoptosis the molecular biology of programmed cel death*; Oxford, 2002.

(7) Vaux, D. L.; Strasser, A. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 2239-2244.

(8) Saikumar, P.; Dong, Z.; Mikhailov, V.; Denton, M.; Weinberg, J. M.; Venkatachalam, M. A. *Am. J. Med.* **1999**, *107*, 489-506.

(9) Popov, S. G.; R.Villasmil; Bernardi, J. *Biochem. Biophys. Res. Commun.* **2012**, *293*, 349-55.

(10) Brüne, B. *Cell Death Differ.* **2003**, *10*, 864-9.

(11) Miyajima, A.; Ito, Y.; Kinoshita, T. *Int. J. Hematol.* **1999**, *69*, 137-146.

(12) Youle, R. J.; Strasser, A. *Nat Rev Mol Cell Biol* **2008**, *9*, 47-59.

(13) Miyashita, T.; Reed, J. C. *Cell* **1995**, *80*, 293-299.

(14) Verheij, M.; Blitterswijk, W. J. v.; Bartelink, H. *Acta Oncol.* **1998**, *37*, 575-581.

(15) Thurston, D. E. *Chemistry and pharmacology of anticancer drugs*; CRC Press, 2006.

(16) Persil, O.; Hud, N. V. *Trends Biotechnol.* **2007**, *25*, 433-436.

(17) http://www.chem.strath.ac.uk/people/academic/colin_j_suckling/research/minor_groove_binders.

(18) Neidle, S. *Nat .Prod. Rep* **2001**, *18*, 291-309.

(19) Rodríguez, F.; Rozas, I.; Kaiser, M.; Brun, R.; Nguyen, B.; Wilson, W. D.; García, R. N.; Dardonville, C. *J. Med. Chem.* **2008**, *51*, 909-923.

(20) Allan, J. M. *Hematologica Rep.* **2006**, *2*, 28-29.

(21) www.pharmacorama.com.

(22) Hurley, L. *Nat. Rev. Cancer* **2002**, *2*, 188-198.

(23) Povirk, L. F. *Mutat. Res. Fund. Mol. Mech. Mut.* **1996**, *355*, 71-89.

(24) Dumontet, C.; Jordan, M. A. *Nat. Rev. Drug Discov.* **2010**, *9*, 790-803.

(25) Schiff, P. B.; Fant, J.; Horwitz, S. B. *Nature* **1979**, *277*, 665-667.

(26) Himes, R. H. *Pharmacol. Ther.* **1991**, *51*, 257-267.

(27) http://cancer.stanford.edu/information/cancerTreatment/methods/hormone.html.

(28) Gervaz, P.; Fontolliet, C. *Int. J. Exp. Pathol.* **1998**, *79*, 359-362.

(29) Kling, J. *Nat. Biotechnol.* **2012**, *30*, 381-381.

(30) Barry, B. W. *Eur. J. Pharm. Sci.* **2001**, *14*, 101-114.

(31) http://www.cancer.org/Treatment/TreatmentsandSideEffects/TreatmentTypes/photodynamic-therapy.

(32) Capella, M. A.; Capella, L. S. *J. Biomed. Sci.* **2003**, *10*, 361-366.

(33) http://www.cancer.gov/cancertopics/factsheet/prevention/HPV-vaccine.

(34) Oettgen, H. F. *J. Cancer Res. Clin.* **1990**, *116*, 116-119.

(35) Blum, R.; Kloog, Y. *Drug Resist. Updates* **2005**, *8*, 369-380.

(36) Shivakumar, L.; Lancet, J. *Clin. Leukemia* **2006**, *1*, 80-83.
(37) Canavese, M.; Santo, L.; Raje, N. *Cancer Biol. Ther.* **2012**, *13*, 451-457.
(38) Wang, L. M.; Ren, D. M. *Mini Rev. Med. Chem.* **2010**, *10*, 1058-1070.
(39) Kandel, E. S. *Cell Cycle* **2009**, *8*, 1819-1820.
(40) Nabhan, C.; Villines, D.; Dalal, N.; Tolzien, K.; Kozolff, M.; Starr, A. *Clin. Lymphoma Myeloma Leuk.* **2012**, *12*, 26-31.
(41) Richard, D. J.; Verheijen, J. C.; Zask, A. *Curr. Opin. Drug Disc.* **2010**, *13*, 428-440.
(42) Furic, L.; Livingstone, M.; Dowling, R. J. O.; Sonenberg, N. *Curr. Oncol.* **2009**, *16*, 68.
(43) Wander, S. A.; Hennessy, B. T.; Slingerland, J. M. *J Clin Invest.* **2011**, *121*, 1231-1241.
(44) Cowan-Jacob, S. W.; Möbitz, H.; Fabbro, D. *Curr. Opin. Cell Biol.* **2009**, *21*, 280-287.
(45) Neidle, S. *Cancer drug design and discovery*; Academic Press, 2008.
(46) Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sundarsanam, S. *Science* **2002**, *298*, 1912.
(47) Weinberg, R. A. *The biology of cancer*; Garland Science, 2007.
(48) www.cellsignaling.com.
(49) Ward, R. A.; Goldberg, F. W. *Kinase drug discovery*; RSC, 2012.
(50) Carmi, C.; Mor, M.; Petronini, P. G.; Alfieri, R. R. *Biochem. Pharmacol.* **2012**, *in press*.
(51) Madhusudan, S.; Ganesan, T. S. *Clin. Biochem.* **2004**, *37*, 618-635.
(52) Blume-Jensen, P.; Hunter, T. *Nature* **2001**, *411*, 355-365.
(53) Miettinen, M.; Lasota, J. *Arch. Pathol. Lab. Med.* **2006**, *130*, 1466-1478.
(54) Hehlmann, R.; Hochhaus, A.; Baccarani, M. *Lancet* **2007**, *370*, 342-350.
(55) Nicholson, R. I.; Gee, J. M. W.; Harper, M. E. *Eur. J. Cancer* **2001**, *37, Supplement 4*, 9-15.
(56) Santin, A. D.; Bellone, S.; Roman, J. J.; McKenney, J. K.; Pecorelli, S. *Int. J. Gynecol. Obstet.* **2008**, *102*, 128-131.
(57) www.willey.com.
(58) Matthews, D. J.; Gerritsen, M. E. *Targeting protein kinases for cancer therapy*; Wiley, 2010.
(59) Lin, C.; Chen, W.; Kruiger, W.; Stolarsky, L. S.; Weber, W.; Evans, R. M.; Verma, I. M.; Gill, G. N.; Rosenfeld, M. G. *Science* **1984**, *224*, 843-848.
(60) Folkman, J. *N. Engl. J. Med.* **1971**, *285*, 1182-1186.
(61) Shibuya, M. *J. Biochem. Mol. Biol.* **2006**, *39*, 469-478.
(62) Homsi, J.; Daud, A. *Cancer Control.* **2007**, *14*, 285-294.
(63) Heldin, C. H. *Upsala J. Med. Sci.* **2012**, *117*, 83-91.
(64) Lasota, J.; Wozniak, A.; Sarlomo-Rikala, M.; Rys, J.; Kordek, R.; Nassar, A.; Sobin, L. H.; Miettinen, M. *Am. J. Pathol.* **2000**, *157*, 1091-1095.
(65) Georgiou, G.; Karali, V.; Zouvelou, C.; Kyriakou, E.; Dimou, M.; Chrisochoou, S.; Greka, P.; Dufexis, D.; Vervesou, E.; Dimitriadou, E.; Efthymiou, A.; Petrikkos, L.; Dima, K.; Lilakos, K.; Panayiotidis, P. *Br. J. Haematol.* **2006**, *134*, 302-306.
(66) Hilberg, F.; Roth, G. J.; Krssak, M.; Kautschitsch, S.; Sommergruber, W.; Tontsch-Grunt, U.; Garin-Chesa, P.; Bader, G.; Zoephel, A.; Quant, J.; Heckel, A.; Rettig, W. J. *Cancer Res.* **2008**, *68*, 4774-4782.
(67) Levinson, A. D.; Oppermann, H.; Levintow, L.; Varmus, H. E.; Bishop, J. M. *Cell* **1978**, *15*, 561-572.
(68) Tatosyan, A. G.; Mizenina, O. A. *Biochem.* **2000**, *65*, 57-67.

(69) Chissoe, S. L.; Bodenteich, A.; Wang, Y.-F.; Wang, Y.-P.; Burian, D.; Clifton, S. W.; Crabtree, J.; Freeman, A.; Iyer, K.; Jian, L.; Ma, Y.; McLaury, H.-J.; Pan, H.-Q.; Sarhan, O. H.; Toth, S.; Wang, Z.; Zhang, G.; Heisterkamp, N.; Groffen, J.; Roe, B. A. *Genomics* **1995**, *27*, 67-82.

(70) Harrison, D. A. *Cold Spring Harb Perspect Biol* **2012**, *4*, a011205.

(71) Cosimo, S. D.; Scaltriti, M.; Val, D.; Rojo, F.; Guzman, M.; Jimenez, J.; Seoane, J.; Arribas, J.; Baselga, J. *J. Clin. Oncol.* **2007**, *25*, 3511.

(72) Alvarez, M.; Roman, E.; Santos, E. S.; Raez, L. E. *Expert Rev. Anticanc.* **2007**, *7*, 1423-1437.

(73) Thomas, G. V.; Tran, C.; Mellinghoff, I. K.; Welsbie, D. S.; Chan, E.; Fueger, B.; Czernin, J.; Sawyers, C. L. *Nat. Med.* **2006**, *12*, 122-127.

(74) Choi, M. E.; Ding, Y.; Kim, S. I. *Semin. Nephrol.* **2012**, *32*, 244-252.

(75) Huang, P.; Han, J.; Hui, L. *Protein Cell* **2010**, *1*, 218-226.

(76) Downward, J. *Nat. Rev. Cancer* **2003**, *3*, 11-22.

(77) Hoshino, R.; Chatani, Y.; Yamori, T.; Tsuruo, T.; Oka, H.; Yoshida, O.; Shimada, Y.; Ari-i, S.; Wada, H.; Fujimoto, J.; Kohno, M. *Oncogene* **1999**, *18*, 813-822.

(78) Yoon, S.; Seger, R. *Growth Factors* **2006**, *24*, 21-44.

(79) Chong, H.; Vikis, H. G.; Guan, K.-L. *Cell. Signal.* **2003**, *15*, 463-469.

(80) Beeram, M.; Patnaik, A.; Rowinsky, E. K. *J. Clin. Oncol.* **2005**, *23*, 6771-6790.

(81) Cargnello, M.; Roux, P. P. *Microbiol. Mol. Biol. Rev.* **2011**, *75*, 50-83.

(82) Mittelstadt, P. R.; Salvador, J. M.; Fornace, A. J.; Ashwell, J. D. *Cell Cycle* **2005**, *4*, 1189-1192.

(83) del Valle-Pérez, B.; Arqués, O.; Vinyoles, M.; de Herreros, A. G.; Duñach, M. *Mol. Cell. Biol.* **2011**, *31*, 2877-2888.

(84) Ali, A.; Hoeflich, K. P.; Woodgett, J. R. *Chem. Rev.* **2001**, *101*, 2527-2540.

(85) Luo, J. *Cancer Lett.* **2009**, *273*, 194-200.

(86) Arora, A.; Scholar, E. M. *J. Pharmacol.Exp. Ther.* **2005**, *315*, 971-979.

(87) Kettle, J. G.; Ward, R. A. *J. Chem. Info. Model.* **2010**, *50*, 525-533.

(88) Zuccotto, F.; Ardini, E.; Casale, E.; Angiolini, M. *J. Med. Chem.* **2009**, *53*, 2681-2694.

(89) Ranjitkar, P.; Brock, A. M.; Maly, D. J. *Chem Biol. .* **2010**, *17*, 195-206.

(90) Klebl, B.; Müller, G.; Hamacher, M. *Protein Kinases as Drug Targets*; Willey-vch, 2011.

(91) Dar, A. C.; Shokat, K. M. *Annu. Rev. Biochem.* **2011**, *80*, 769-795.

(92) Imai, K.; Takaoka, A. *Nat Rev Cancer* **2006**, *6*, 714-727.

(93) Gleave, M. E.; Monia, B. P. *Nat. Rev. Cancer* **2005**, *5*, 468-479.

(94) Neckers, L.; Workman, P. *Clin. Cancer Res.* **2012**, *18*, 64-76.

(95) http://nens.yellowcouch.org/articles/pki07/ProteinKinaseInhibitors2007.html.

(96) Weinmann, H.; Metternich, R. *ChemBioChem* **2005**, *6*, 455.

(97) www.cancer.gow/cancertopics/druginfo/ziv-aflibercept.

(98) Van Erp, N. P.; Gelderblom, H.; Guchelaar, H. J. *Cancer Treat. Rev.* **2009**, *35*, 692-706.

(99) http://www.centerwatch.com/drug-information/fda-approvals/.

(100) Carpiuc, K. T.; Rosti, G.; Castagnetti, F.; Treur, M.; Stephens, J. *Onco.Targets Ther.* **2010**, *3*, 205-210.

(101) Manley, P. W.; Stiefl, N.; Cowan-Jacob, S. W.; Kaufman, S.; Mestan, J.; Wartmann, M.; Wiesmann, M.; Woodman, R.; Gallagher, N. *Bioorg. Med. Chem.* **2010**, *18*, 6977-6986.

(102) Levinson, N. M.; Boxer, S. G. *PLoS One.* **2012**, *7*, e29828.

(103) Comis, R. L. *Oncologist* **2005**, *10*, 467-470.

(104) http://www.asco.org/ASCOv2/Meetings/Abstracts?&vmview=abst_detail_view&confID=55&abstractID=34685.

(105) Wood, E. R.; Truesdale, A. T.; McDonald, O. B.; Yuan, D.; Hassell, A.; Dickerson, S. H.; Ellis, B.; Pennisi, C.; Horne, E.; Lackey, K.; Alligood, K. J.; Rusnak, D. W.; Gilmer, T. M.; Shewchuk, L. *Cancer Res.* **2004**, *64*, 6652-6659.

(106) Calvo, E.; Ravaud, A.; Bellmunt, J. *Cancer Treat. Rev.* **2012**, *in press*.

(107) http://www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm314438.htm.

(108) Lebakken, C. S.; Reichling, L. J.; Ellefson, J. M.; Riddle, S. M. *J. Biomol. Screen.* **2012**, *17*, 813-821.

(109) Knight, Z. A.; Shokat, K. M. *Chem. Biol.* **2005**, *12*, 621-637.

(110) Jianming, Z.; Adrián, F. J.; Jahnke, W.; Cowan-Jacob, S. W.; Li, A. G.; Iacob, R. E.; Taebo, S.; Powers, J.; Dierks, C.; Fangxian, S.; Gui-Rong, G.; Ding, Q.; Okram, B.; Yongmun, C.; Wojciechowski, A.; Xianming, D.; Guoxun, L.; Fendrich, G.; Strauss, A.; Vajpai, N. *Nature* **2010**, *463*, 501-506.

(111) Karaman, M. W.; Herrgard, S.; Treiber, D. K.; Gallant, P.; Atteridge, C. E.; Campbell, B. T.; Chan, K. W.; Ciceri, P.; Davis, M. I.; Edeen, P. T.; Faraoni, R.; Floyd, M.; Hunt, J. P.; Lockhart, D. J.; Milanov, Z. V.; Morrison, M. J.; Pallares, G.; Patel, H. K.; Pritchard, S.; Wodicka, L. M. *Nat. Biotechnol.* **2008**, *26*, 127-132.

(112) Copeland, R. A. *Future Med. Chem.* **2011**, *3*, 1491-1501.

(113) Graczyk, P. P. *J. Med. Chem.* **2007**, *50*, 5773-5779.

(114) Maitland, M. L.; Ratain, M. J. *Ann. Intern. Med.* **2006**, *145*, 702-703.

(115) Roberts, P. J.; Der, C. J. *Oncogene* **2007**, *26*, 3291-3310.

(116) Wagner, E. F.; Nebreda, Á. R. *Nat. Rev. Cancer* **2009**, *9*, 537-549.

(117) Bamborough, p.; Drewry, D.; Harper, G.; Smith, G. K.; Smith, K.; Schneider, K. *J. Med. Chem.* **2008**, *51*, 7898.

(118) Wilhelm, S. M.; Carter, C.; Lynch, M.; Lowinger, T.; Dumas, J.; Smith, R. A.; Schwartz, B.; Simantov, R.; Kelley, S. *Nature Rev. Drug Discov.* **2006**, *5*, 835-844.

(119) Godley, P. A.; Ataga, K. I. *Curr Opin Oncol* **2000**, *12*, 260-264.

(120) McLaughlin, J. J.; Lipworth, L. *Semin. Oncol.* **2000**, *27*, 115-123.

(121) Chan, S. L.; Mok, T.; Ma, B. B. Y. *Curr Oncol Rep* **2012**.

(122) http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Scientific_Discussion/human/000690/WC500027707.pdf.

(123) Calvani, N.; Morelli, F.; Leo, S.; Orlando, L.; Lombardi, L.; Gnoni, A.; Cinefra, M.; Maiello, E.; Lorusso, V.; Cinieri, S. *Med. Oncol.* **2012**, *29*, 1908-1913.

(124) Ratain, M. J.; Eisen, T.; Stadler, W. M.; Flaherty, K. T.; Kaye, S. B.; Rosner, G. L.; Gore, M.; Desai, A. A.; Patnaik, A.; Xiong, H. Q.; Rowinsky, E.; Abbruzzese, J. L.; Xia, C.; Simantov, R.; Schwartz, B.; O'Dwyer, P. J. *J. Clin. Oncol.* **2006**, *24*, 2505-2512.

(125) Wilhelm, S. M.; Adnane, L.; Newell, P.; Villanueva, A.; Llovet, J. M.; Lynch, M. *Mol. Cancer. Ther.* **2008**, *7*, 3129-3138.

(126) Kummar, S.; Gutierrez, M.; Doroshow, J. H.; Murgo, A. J. *Brit. J. Clin. Pharmaco.* **2006**, *62*, 15-26.

(127) Wilhelm, S. M.; Carter, C.; Tang, L.; Wilkie, D.; McNabola, A.; Rong, H.; Chen, C.; Zhang, X.; Vincent, P.; McHugh, M.; Cao, Y.; Shujath, J.; Gawlak, S.; Eveleigh, D.; Rowley, B.; Liu, L.; Adnane, L.; Lynch, M.; Auclair, D.; Taylor, I.; Gedrich, R.;

Voznesensky, A.; Riedl, B.; Post, L. E.; Bollag, G.; Trail, P. A. *Cancer Res.* **2004**, *64*, 7099.

(128) Wan, P. T.; Garnett, M. J.; Roe, S. M.; Lee, S.; Niculescu-Duvaz, D.; Good, V. M.; Jones, C. M.; Marshall, C. J.; Springer, C. J.; Barford, D.; Marais, R. *Cell* **2004**, *116*, 855-867.

(129) Whittaker, S.; Kirk, R.; Hayward, R.; Zambon, A.; Viros, A.; Cantarino, N.; Affolter, A.; Nourry, A.; Niculescu-Duvaz, D.; Springer, C.; Marais, R. *Sci Transl. Med.* **2010**, *2*, 35ra41.

(130) Pao, W.; Miller, V. A.; Politi, K. A.; Riely, G. J.; Somwar, R.; Zakowski, M. F.; Kris, M. G.; Varmus, H. *PLoS Med.* **2005**, *2*, 225-235.

(131) Oh, S. J.; Erb, H. H. H.; Hobisch, A.; Santer, F. R.; Culig, Z. *Endocr. Relat. Cancer* **2012**, *19*, 305-319.

(132) Quan, M.; Zhong, Z.; Yang, J. *Zhongliu Yaoxue* **2012**, *2*, 169-172.

(133) Huang, S.; Sinicrope, F. A. *Mol. Cancer Ther.* **2010**, *9*, 742-750.

(134) Zhaoa, W.; Zhanga, T.; Qua, B.; Wua, X.; Zhua, X.; Menga, F.; Gua, Y.; Shub, Y.; Shena, Y.; Suna, Y.; Xua, Q. *Anti-Cancer Drug* **2010**, *22*, 79-88.

(135) Bankston, D.; Dumas, J.; Natero, R.; Riedl, B.; Monahan, M.-K.; Sibley, R. *Org. Process Res. Dev.* **2002**, *6*, 777-781.

(136) Riedl, B.; Dumas, J.; Khire, U.; Lowinger, T. B.; Scott, W. J.; Smith, R. A.; Wood, J. E.; Natero, R.; U.S. Patent 8124630, 2012.

(137) Breitler, S.; Oldenhuis, N. J.; Fors, B. P.; Buchwald, S. L. *Org. Lett.* **2011**, *13*, 3262-3265.

(138) Rawling, T.; McDonagh, A. M.; Tattam, B.; Murray, M. *Tetrahedron* **2012**, *68*, 6065-6070.

(139) Wang, B.; Luo, Y.; Patent CN 101671299 (A), 2010.

(140) Sun, M.; Wei, H.; Cai, J.; Ji, M. *Zhongguo Yaoxue Zazhi* **2009**, *44*, 394-396.

(141) Zhao, C.; Chen, L.; Xu, X.; Luo, X.; Ji, Y. *Zhongguo Yiyao Gongye Zazhi* **2007**, *38*, 614-616.

(142) Najjar, Y. G.; Rini, B. I. *Ther. Adv. Med. Oncol.* **2012**, *4*, 183-194.

(143) Ibrahim, N.; Yu, Y.; Walsh, W. R.; Yang, J. L. *Oncol Rep.* **2012**, *27*, 1303-1311.

(144) Roodhart, J. M.; Langenberg, M. H.; Witteveen, E.; Voest, E. E. *Curr. Clin. Pharmacol.* **2008**, *3*, 132-143.

(145) Zunder, E. R.; Knight, Z. A.; Houseman, B. T.; Apsel, B.; Shokat, K. M. *Cancer Cell* **2008**, *14*, 180-192.

(146) Carracedo, A.; Li, M.; Teruya-Feldstein, J.; Rojo, F.; Salmena, L.; Alimonti, A.; Egia, A.; Sasaki, A. T.; Thomas, G.; Kozma, S. C.; Papa, A.; Nardella, C.; Cantley, L. C.; Baselga, J.; Pandolfi, P. P. *J. Clin. Invest.* **2008**, *118*, 3065-3074.

(147) Hörig, H.; Marincola, E.; Marincola, F. M. *Nat. Med.* **2005**, *11*, 705-708.

(148) Woolf, S. H. *JAMA - J. Am. Med. Assoc.* **2008**, *299*, 211-213.

(149) Dardonville, C.; Barrett, M. P.; Brun, R.; Kaiser, M.; Tanious, F.; Wilson, W. D. *J. Med. Chem.* **2006**, *49*, 3748-3752.

(150) Nagle, P. S.; Rodriguez, F.; Kahvedžić, A.; Quinn, S. J.; Rozas, I. *J. Med. Chem.* **2009**, *52*, 7113-7121.

(151) Goonan, Á.; Kahvedžić, A.; Rodriguez, F.; Nagle, P. S.; McCabe, T.; Rozas, I.; Erdozain, A. M.; Javier Meana, J.; Callado, L. F. *Bioorg. Med. Chem.* **2008**, *16*, 8210-8217.

(152) Kahvedžić, A.; Nathwani, S.-M.; Zisterer, D.; Rozas, I. *J. Med. Chem* **2013**, *56*, 451-459.

(153) Kahvedžić A., Ph.D thesis, 2011.

(154) Kahvedžić A., final year thesis, 2007.
(155) Rozas, I.; P. E. Kruger *J. Chem. Theory Comput.* **2005**, *1*, 1055-1062.
(156) Rozas, I.; Alkorta, I.; Elguero, J. *Struct. Chem.* **2008**, *19*, 923-933.
(157) Blanco, F.; Kelly, B.; Alkorta, I.; Rozas, I.; Elguero, J. *Chem. Phys. Lett.* **2011**, *511*, 129-134.
(158) Kelly, B.; Sánchez-Sanz, G.; Blanco, F.; Rozas, I. *Comp. Theor. Chem.* **2012**, *998*, 64-73.

Chapter 2 – Objectives

Some of the previous work performed in Rozas' laboratory consisted of the synthesis of 4,4'-*bis*-(2-aminoimidazolino) and *bis*-guanidino derivatives as DNA minor groove binders. Some of these derivatives, with high affinity for the minor groove of DNA, possessed potent *in vitro* antitrypanosomal and antiplasmodial activity.[1-3] However, some others, which do not strongly target DNA – as evidenced by DNA thermal melt experiments – show cytotoxicity in cancer cells such as human promyelocytic leukemia HL-60, breast cancer MCF-7 and neuroblastoma Kelly cell lines.[4] Therefore, it was concluded that the mechanism of action by which these compounds induce apoptosis occurs in a different manner to targeting the DNA.

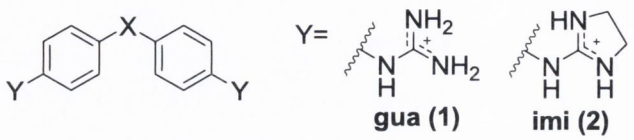
Taking into account the structural similarities between these compounds and some protein kinase inhibitors such as sorafenib, the inhibition of protein kinases was considered as the potential mechanism of action responsible for the cytotoxicity. Hence, several objectives were planned to be achieved in the present work:

(i) Chemical synthesis of new guanidine-like derivatives to be achieved through a combination of established reactions and new methodologies which are necessary to facilitate the preparation of these derivatives.

(ii) Biological evaluation of these compounds to help clarify the mode of action. This will be accomplished through cell assays to assess cytotoxicity in a range of cell lines, apoptosis evaluation and specific kinase assays to elucidate the target of the compounds.

2.1. Chemical Objectives

One of the aims of this project is to re-synthesise some of the 4,4'-*bis*-(2-aminoimidazolino) and *bis*-guanidino diaromatic derivatives that gave good cytotoxic results (Figure 2.1) and test them as possible protein kinase inhibitors.



X= O (**1a, 2a**), CO (**1b, 2b**), CH_2 (**1c, 2c**), NH(**1d,2d**)

Figure 2.1. Diaromatic 4,4'-*bis*-(2-aminoimidazoline) and *bis*-guanidine DNA minor groove binders synthesised in Rozas' laboratory.

Considering the structural similarities of these compounds with sorafenib, we aim to introduce some of the characteristic moieties of sorafenib into our *bis*-guanidine-like structures, through iterative modifications, leading to five distinct families as synthetic targets (Figure 2.2).

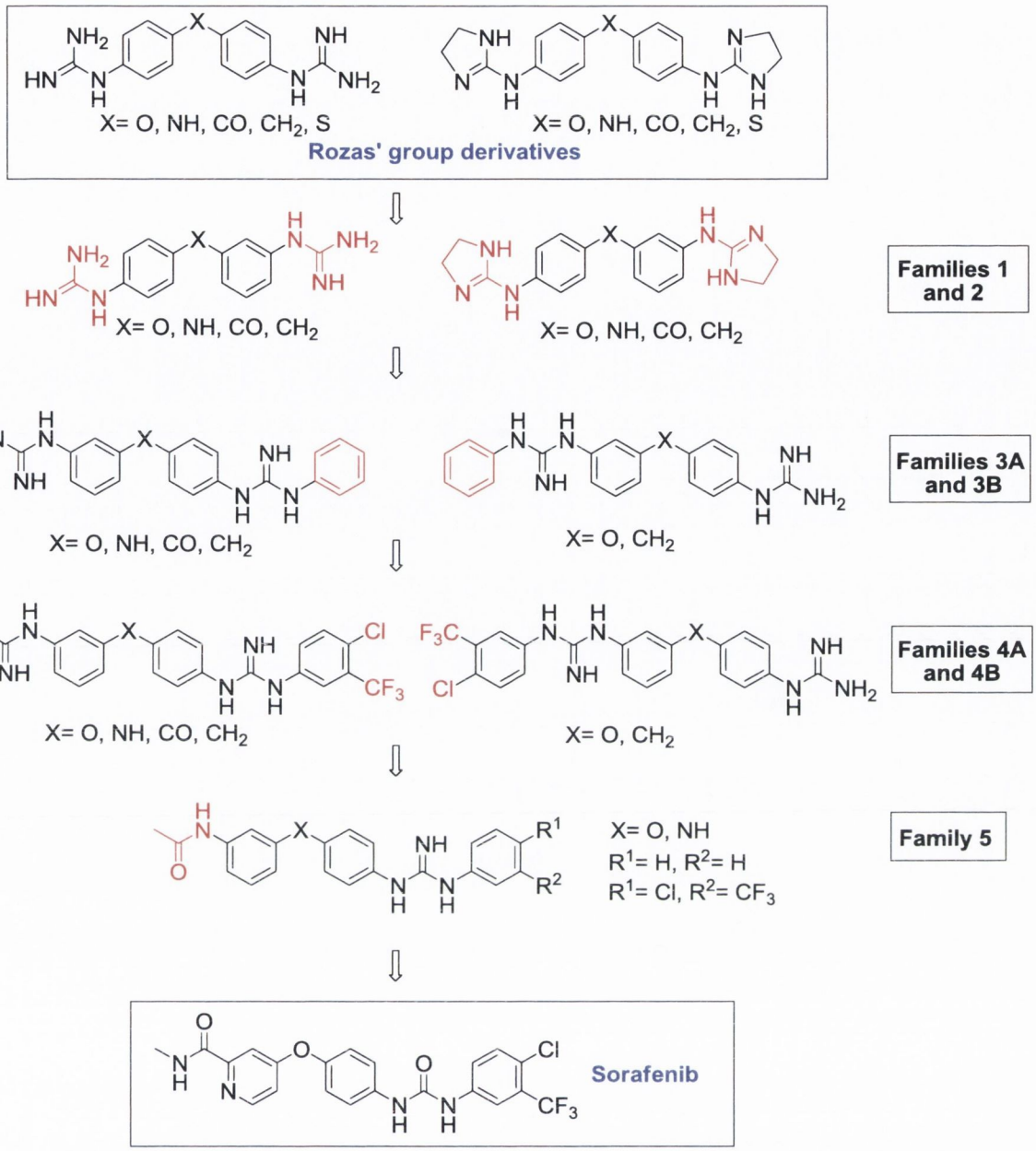


Figure 2.2. Rational behind the five families of 3,4'-guanidine-like derivatives proposed. All guanidine and 2-aminoimidazolines are prepared as hydrochloride salts.

In the first two families proposed, the 3,4'-substitution pattern, present in sorafenib, will be introduced to form the 3,4'-di-guanidinium (Family 1) and 3,4'-di-(2-aminoimidazolinium) (Family 2) derivatives. This modification aims to investigate the effect of the 3,4'- *vs.* 4,4'-substitution patterns in the kinase inhibitory activity.

The next change proposed is the addition of an aromatic moiety to the guanidine in the 4'-position (phenyl – Ph – in Family 3A and 4-chloro-3-trichloromethylphenyl (4-Cl-3-CF_3)-Ph in Family 4A) as present in sorafenib. The purpose of these two families is to check the effects on kinase inhibitory activity of aromatic substituted guanidines. Also, the phenyl derivative – substituted or not – was introduced in the guanidine in 3- position (Family 3B and 4B), opposite to sorafenib. With this change it was hoped to increase the affinity and to obtain information about the structure activity relationships (SARs) between these compounds and their targets. The (4-Cl-3-CF_3)-Ph group is a very characteristic moiety of sorafenib and its introduction will indicate if the inhibitory activity improves when the Cl and CF_3 substituents are present on the phenyl ring. This substituted phenyl group is pharmacokinetically favourable, due to its high degree of metabolic stability and lipophilic nature.

Finally, Family 5 incorporates an acetamide group in the 3- position to the diaromatic system, instead of the guanidine moiety present in all other families. Both groups differ substantially, the guanidinium is positively charged, with hydrogen bond donors and the ability to form π-cation complexes, while the acetamide is neutral and has both hydrogen bond donors and acceptors. This will allow an insight into the requirements for favourable binding of these molecules to the active site.

The aim of all these structural changes is to obtain SAR information which will help to understand the characteristics required for the binding between the compounds and their target kinase. As a result, this will lead to more efficient and improved protein kinase inhibitors.

According to the synthetic procedures established in Rozas' laboratory for the preparation of 4,4'-*bis*-guanidinium derivatives,[5] the corresponding 3,4'-diaryl diamines must be first prepared to synthesise these new five families. These precursors will serve as the core of all five families and the route to their preparation is dependent on the linker involved. A range of linkers with differing electronic properties and hydrogen bond profiles was used (O, NH, CH_2, CO) to study their influence on the biochemical activity.

Additionally, for the synthesis of Families 3 and 4, the corresponding di-substituted thiourea derivatives have to be prepared, the *N*-Boc-*N'*-phenylthiourea for Family 3 and the *N*-Boc-*N'*-(3-trifluoromethyl-4-chlorophenyl)-thiourea for Family 4. Their synthesis will follow methodology developed in Rozas' laboratory.[6]

After the relevant diamines and thioureas are synthesised, they will be coupled using mercury(II) chloride in a reaction that has been extensively employed in Rozas' group for the preparation of guanidine-like compounds. This results in the Boc-protected guanidine/2-aminoimidazoline, which are then converted to the guanidine/2-aminoimidazoline hydrochlorides of biological interest by treatment with solutions of hydrochloric acid.

2.2. Biological Objectives

Once the synthesis of the five families is completed, they will be evaluated through a variety of biological techniques to assess their role as potential anticancer drugs and in particular as protein kinase inhibitors.

The cytotoxicity of these derivatives and their effects on cell proliferation will be evaluated using the AlamarBlue® viability assay in different cancer cell lines. This assay is technically simple, rapid, non-toxic, versatile and reproducible. It is a valuable technique in terms of measuring percentage viability and proliferation; however, the information it provides is limited. This viability assay only gives information about whether or not the compound stops cell proliferation in cancer cells, but gives no information on the mechanism by which cellular death has occurred. In addition, an appropriate concentration range of each drug will be used allowing for the determination of IC_{50} values for all compounds.

To obtain as much information as possible about the clinical profile of these synthesised compounds, they will be tested in a wide range of cancer cell types including HL-60 (leukemia), MCF-7 (breast cancer), HUH-7 (liver carcinoma) and RKO (colorectal cancer) cell lines. The testing in HL-60 and HUH-7 cell lines will be performed in collaboration with Prof. D. Zisterer in the School of Biochemistry and Immunology at TCD. Subsequently, evaluations in MCF-7 and RKO cell lines will be performed by our collaborators Dr. J. Gee (Cardiff School of Pharmacy, UK) and Dr. S. Van Schaeybroeck (Queens University Belfast, UK), respectively.

The AlamarBlue® viability assays in HL-60 cells, which are very sensitive suspension cell line, will be first performed as a general screening for all the compounds synthesised to identify the most potent candidates. Then, the most cytotoxic derivatives will be tested on the rest of the more adherent and resistant cell lines (HUH-7, MCF-7 and RKO cells).

To obtain information about the type of cellular death occurring, the most active derivatives will be subjected to cell morphological studies and flow cytometry, which will allow to identify whether or not apoptosis has taken place. This will provide definitive evidence on whether the compounds in question play an active role in apoptotic cell death and have potential as anticancer agents.

To test the activity of the compounds as protein kinase inhibitors, specific kinase testing will be performed. The kinases of interest to this work are those targeted by sorafenib, to allow comparison of the mode of action with our compounds. These targets are: B-RAF, RAF-1 or C-RAF, MEK, ERK, p38 MAPK and VEGFR, which mediates the cytotoxic and antiangiogenesis activity of sorafenib. Some of the testing will be performed by Cerep which is a company specialised in biological testing. This screening will provide the percentage inhibition of the aforementioned kinases, obtained following specific procedures.[7-10]

Additionally, western blots will be carried out to analyse the expression levels of kinases involved in the RAF/MEK/ERK pathway and other related pathways, such as the expression of AKT and STAT-3. Also, a confirmation of the induction of apoptosis will be performed by analysing PARP cleavage. These assays will be carried out by our collaborators in Queen's University in Belfast. Other related kinase pathways such as CK-1δ and GSK-3 will be evaluated as possible targets of our compounds by our collaborators Dr. C. Perez in Consejo Superior de Investigaciones Cientificas (CSIC) in Madrid (Spain).

2.3. Computational Objectives

Computational modelling techniques such as docking and molecular dynamics will be utilised to understand the interactions between our novel compounds and their potential targets (B-RAF, C-RAF and MEK-1) as well as to analyse their modes of binding. This study will also aim to elucidate the specific molecular target of our compounds. This will be investigated in collaboration with Dr. Brendan Kelly and Prof. Jaime Rubio Martinez at the University of Barcelona (Spain).

2.4. References

(1) Dardonville, C.; Barrett, M. P.; Brun, R.; Kaiser, M.; Tanious, F.; Wilson, W. D. *J. Med. Chem.* **2006**, *49*, 3748-3752.

(2) Rodríguez, F.; Rozas, I.; Kaiser, M.; Brun, R.; Nguyen, B.; Wilson, W. D.; García, R. N.; Dardonville, C. *J. Med. Chem.* **2008**, *51*, 909-923.

(3) Nagle, P. S.; Rodriguez, F.; Kahvedžić, A.; Quinn, S. J.; Rozas, I. *J. Med. Chem.* **2009**, *52*, 7113-7121.

(4) Kahvedžić, A.; Nathwani, S.-M.; Zisterer, D.; Rozas, I. *J. Med. Chem.* **2013,** *in press*.

(5) Dardonville, C.; Goya, P.; Rozas, I.; Alsasua, A.; Martin, M. I.; Borrego, M. J. *Bioorg. Med. Chem.* **2000**, *8*, 1567-1577.

(6) O'Donovan, D. H.; Rozas, I. *Tetrahedron Lett.* **2011**, *52*, 4117-4119.

(7) Zhang, Y. G.; Q. Du; Fang, W. G.; Jin, M. L.; Tian, X. X. *Int. J. Oncol.* **2008**, *33*, 595-602.

(8) Li, W.; Nadelman, C.; Henry, G.; Fan, J.; Muellenhoff, M.; Medina, E.; Gratch, N. S.; Chen, M.; Han, J.; Woodley, D. *J. Invest. Dermatol.* **2001**, *117*, 1601-1611.

(9) Furet, P.; Bold, G.; Hofmann, F.; Manley, P.; Meyer, T.; Altmann, K.-H. *Bioorg. Med. Chem. Lett* **2003**, *13*, 2967-2971.

(10) Force, T.; Bonventre, J. V.; Heidecker, G.; Rapp, U.; Avruch, J.; Kyriakis, J. M. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 1270-1274.

Chapter 3 – Chemical Synthesis

3.1. Introduction

A growing number of biologically and pharmaceutically relevant compounds incorporate the guanidine functionality and their synthesis is a field of intense investigation. Many of these compounds are pharmaceutical agents possessing a broad spectrum of activity that includes, for example, anticancer drugs, histamine-2 (H-2) receptor antagonists, antiseptics and cholesterol biosynthesis inhibitors (Figure 3.1).[1,2,3,4]

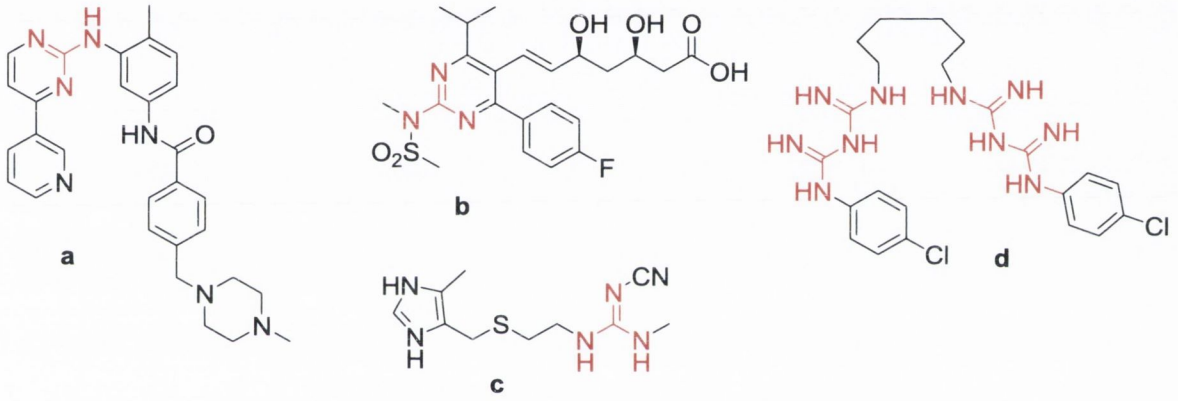
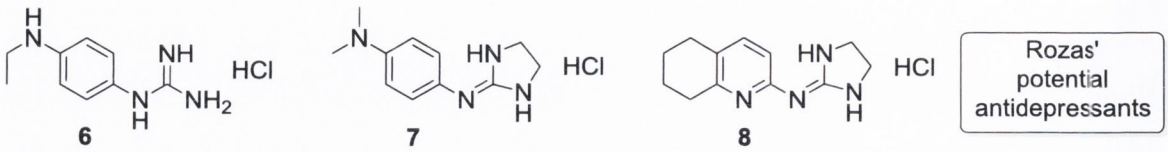


Figure 3.1. Biological substances containing guanidine-like moieties: (a) imatinib, an anti-cancer drug, (b) rosuvastatin, a cholesterol biosynthesis inhibitor, (c) cimetidine, an H2-receptor antagonist, (d) chlorhexidine, an antiseptic.

Due to the biological interest in guanidine-containing compounds, the Rozas group has been working for many years on the design, synthesis and biological evaluation of novel guanidine based derivatives. One of the biological applications of these molecules has been their use as antidepressants, inhibiting the α_2 adrenoceptor receptor (α_2-AR), causing an increase in noradrenaline concentrations in the synaptic cleft; hence alleviating the symptoms of depression. Research within our group has described the synthesis and pharmacological evaluation of over 80 molecules to date in a series of papers describing structure-activity relationships at the α_2-AR.[5,6,7] These molecules have had the general

form of a substituted phenyl, diphenyl or pyridine systems directly attached to a guanidine or 2-aminoimidazoline group (Figure 3.2). Some of these compounds show high antagonistic affinity at the α_2-AR, and also have antidepressant effects in animal models of the disease.



6 **7** **8**

Figure 3.2. Examples of Rozas' guanidine-containing compounds with potential as antidepressants

Another application of the guanidine-like compounds synthesised in Rozas' group has been their use as DNA minor groove binders with potential as anti-cancer and anti-plasmodial drugs, as previously described in section 1.7. Previous work in the Rozas laboratory[8,9,10] involving the preparation of symmetric and asymmetric guanidine and 2-aminoimidazoline derivatives (see Figure 1.30), produced encouraging results, not only in terms of binding ability and affinity towards the minor groove of DNA but also in the initiation of anticancer activity through the induction of cellular apoptosis.[11, 12]

The guanidine and 2-aminoimidazoline groups will be cationic at physiological pH due to their basicity; hence, both the antidepressant and the DNA minor groove binder compounds have been prepared as the corresponding hydrochlorides due to the favourable water solubility and stability of these salts, both of which are desirable for pharmacological testing.

These encouraging results and some observations such as the highly structural similarities of these derivatives and sorafenib (see Figure 1.31) led to the design of new families of guanidine based compounds to act as potential protein kinases' inhibitors (PKIs).

In total, five different families have been prepared in this work (Figure 3.3). In the first two families, the 3,4'-substitution pattern (as present in sorafenib) has been investigated and, thus, 3,4'-*bis*-guanidinium (Family 1) and 3,4'-*bis*-(2-aminoimidazolinium) derivatives

(Family 2) have been prepared. Different linkers connecting the aromatic rings (O, NH, CH_2, CO) were explored to investigate its influence on pharmacological activity. Next, we introduced aromatic rings in the guanidinium group at the 4'-position, a phenyl group and the 4-chloro-3-trifluoromethylphenyl group characteristic of sorafenib (Families 3 and 4, respectively). The corresponding analogues with an aromatic group in the guanidine in the 3-position have also been prepared for the sake of comparison. Finally, in Family 5, an acetamide group was inserted in the 3-position instead of the *N*-methylcarbamoyl found in sorafenib, or the guanidinium which is present in each of Families 1-4.

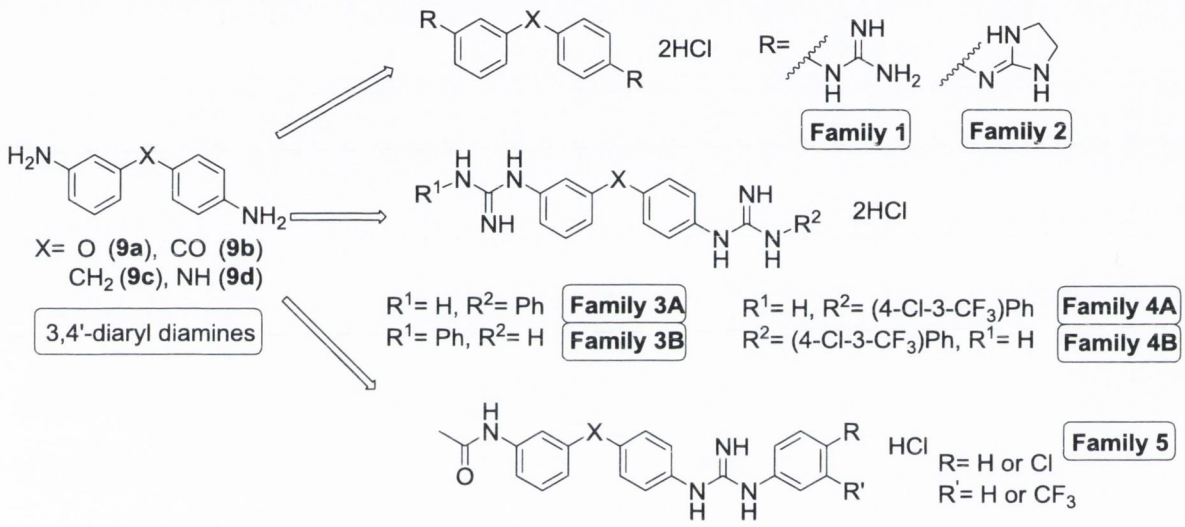


Figure 3.3. Novel families of guanidine-based potential protein kinase inhibitors proposed.

With these successive changes to the structure of our lead compounds we intend to understand the moieties responsible for PK inhibition, obtain structure activity relationships (SARs) and develop a more efficient family of PK inhibitors. The core of the structure of these five families – and also the starting point of their synthesis – is the 3,4'-diphenyl diamines; hence, they need to be synthesised in large amounts and with high yielding methods to allow preparation of Families 1-5.

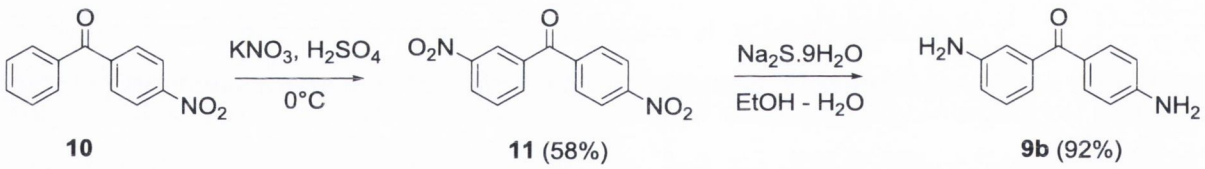
3.2. Synthesis of 3,4'-Diphenyl Diamines

To synthesise all of the proposed families of 3,4'-guanidine-like compounds in this project, firstly the 3,4'-diphenyl diamines have to be accessed. With the exception of the 3-(4-aminophenoxy)aniline (**9a**), these diamines were not commercially available and, thereby, had to be prepared. The synthesis of these diamines varies depending on the nature of the linker connecting the phenyl groups and this will be described in the following subsections.

3.2.1. Synthesis of 3-(4-aminobenzoyl)aniline (9b)

The first step in the synthesis of compound **9b** was nitration of commercially available 4-nitrobenzophenone (**10**) using H_2SO_4 and KNO_3 at 0 °C (Scheme 3.1). The nitrating mixture forms the electrophile nitronium ion (NO_2^+), which attacks the benzene ring in an electrophilic aromatic substitution (EAS) manner. Due to the electron-withdrawing effects of the carbonyl linker in compound **10**, the expected product was the *meta*-nitrated compound, 3,4'-nitrobenzophenone (**11**). Experimentally, this proved to be the major product which was purified by recrystallisation from acetone.

Scheme 3.1.

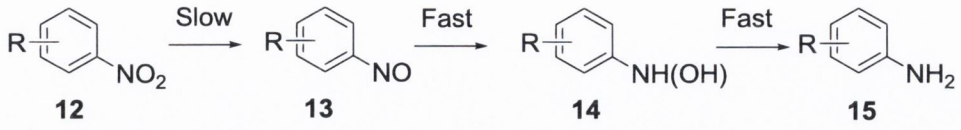


10 **11** (58%) **9b** (92%)

Once compound **11** was obtained, the reduction of the nitro groups was carried out using sodium sulphide *nona*-hydrate ($Na_2S.9H_2O$) as reducing agent (Scheme 3.1). The mechanism of the reduction of aromatic nitro groups using Na_2S is complex due to the fact that a solution of Na_2S could form several ionic species including sulphide, hydrosulphide, disulphide, hydrosulphide and thiosulphate, which are capable of effecting the reduction.[13] Haber *et al.* suggest that the course of the reduction of a general nitro group (**12**) using Na_2S could involve several steps; an initial low rate step in which compound **12** is reduced to the corresponding nitroso group (**13**), followed by a fast step involving conversion of **13**

to a hydroxylamine (**14**), which finally is reduced to obtain the aromatic amine (**15**), in a fast step (Scheme 3.2).[14]

Scheme 3.2.



12 **13** **14** **15**

There are other methodologies described for the reduction of aromatic nitro groups such as catalytic hydrogenation with Pd/C or Raney-Nickel, iron in acidic media, tin(II) chloride, titanium(III) chloride or zinc(II) chloride, among others. However, the $Na_2S.9H_2O$ method was selected because it had previously been successfully used in our group for similar substrates. Hence, this reduction afforded the desired product **9b** in 92% yield after a simple purification process, which involved removal of the solvent under vacuum and wash of the product with water.

The X-ray crystal data which was obtained for compound **11** reveals a conformation that avoids repulsion between protons on adjacent aromatic rings (Figure 3.4). The rings adopt a torsionate conformation with a dihedral angle of approximately 62° between the planes. This arrangement of the phenyl rings can be predicted to exist for the entire series of carbonyl-linked diphenyl molecules synthesised in this study. Interestingly, for **11**, a packing arrangement with alternate stacking of nitro and aromatic groups is observed, being induced by hydrogen bond interactions between the nitro oxygen atoms and the aromatic protons. At the same time, the packing for the dihydrochloride salts is based on π-cation interactions, in which the positive charge of the guanidiniums interacts with the π-electrons of the phenyl ring as can be seen in the X-ray crystal structure of *N*-(5-methylpyridin-2-yl)guanidinium chloride (Figure 3.4, right).[15]

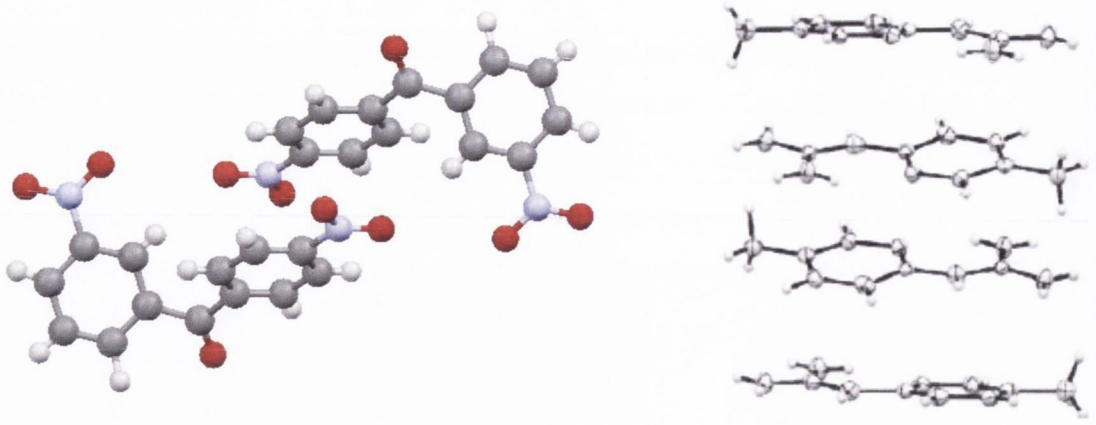


Figure 3.4. Packing arrangement in the crystal structure of compound **11** (*left*) and packing arrangement in the X-ray crystal structure of N-(5-methylpyridin-2-yl)guanidinium chloride (*right*).[15]

3.2.2. Synthesis of 3-(4-aminobenzyl)aniline (9c)

For the synthesis of **9c,** the reduction of **9b** was proposed. Although many methods exist for the reduction of aromatic nitro derivatives to anilines, and of carbonyl groups to the corresponding alkyl derivatives, their simultaneous reduction had not previously been reported and it would be of significant synthetic utility. The methods most commonly used for reduction of carbonyl groups to alkanes are the Clemmensen,[16] Wolff-Kishner,[17,18] and Mozingo reactions.[19] Although both the Clemmensen and Wolff-Kishner conversions work well, they are carried out under strongly acidic and basic conditions, respectively. The Mozingo reduction of dithioketals using Raney-Nickel overcomes some of these problems but requires formation of the dithioketal and the undesirable use of a heavy metal.

Raney-Nickel can also be used as a catalyst for the hydrogenation of aryl-nitro compounds to aryl-amines.[20] Similarly, Zn can be used for the reduction of aromatic nitro compounds to anilines, but still involves the use of a heavy metal.[21] Moreover, any attempt to obtain simultaneous reduction of carbonyl and nitro groups with Zn, using Clemmensen conditions,[21] was unsuccessful.

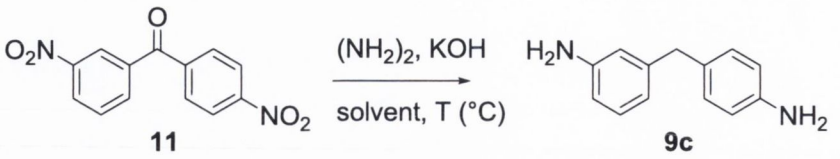
Aromatic nitro groups can be readily reduced to anilines by several methods, such as catalytic hydrogenation with Pd/C,[22] Pt(IV) oxide[23] or Raney-Nickel.[24] Reductions using metals such as Fe or Zn powder in acidic media are also widely used.[25] Among other nitro

reducing agents, sodium sulfide has been applied successfully in our laboratory.[13] However, none of these methods are capable of reducing a carbonyl group to an alkane.

The Wolff-Kishner reaction has long been used for the full reduction of carbonyls to alkanes. This reaction – which proceeds *via* the formation of a hydrazone – generally requires hydrazine hydrate and a strong base, such as KOH, in a polar solvent like propane-1,3-diol or triethylene glycol at high temperatures (> 200 °C).[26]

When trying to obtain diamine **9c** from **9b**, the reduction of the carbonyl group under Wolff-Kishner conditions was attempted with the intention of subsequently reducing the aromatic nitro groups *via* a different method. The polar solvents usually required in the Wolff-Kishner reduction limited the utility of this transformation, as the products obtained are highly polar and prove difficult to extract from the solvent. Thus, conversion to the product as adjudged by ^{1}H NMR spectroscopy was low (0-60%) in a variety of solvents (Table 3.1) and overall yields were less than 10%. Although conversion was improved by increasing the equivalents of hydrazine and base and the reaction time, the problem of product isolation persisted, and, to circumvent it, hydrazine was used in excess, not only as a reagent but also as solvent, in a pressure tube at 135 °C. It was found that, using a ratio of 1:2 of compound **11** and hydrazine, only the nitro groups were reduced; however, when using an excess of hydrazine in a pressure tube without any additional solvent, a simultaneous reduction of the carbonyl and nitro groups was observed. This reaction was not described before and, we optimized this transformation by modifying the reaction conditions and extending it to other nitro and carbonyl systems.[27] For our test compound (**11**, see Table 3.1), conversion into the totally reduced product was increased to over 90%, and an overall yield of 88% was obtained for the purified compound.

Table 3.1. Conditions and yields for the double reduction of compound **11** with and without solvent.



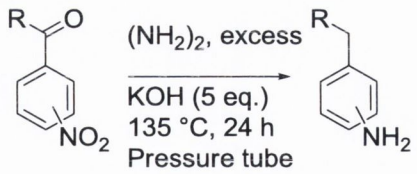
11 **9c**

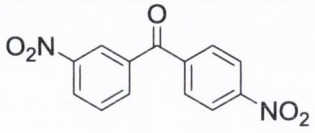
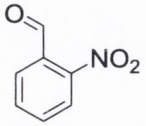
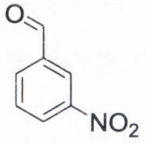
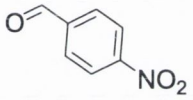
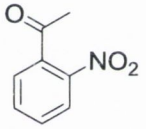
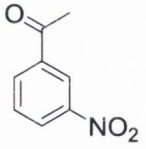
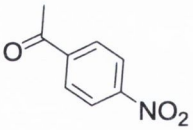
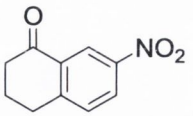
Solvent (mL)	$(NH_2)_2$ (eq.)	KOH (eq.)	T (°C)	Time (h)	Conv.[a] (%)	Yield[b] (%)
Propane-1,3-diol (5)	1.1	1.2	220 reflux	4	0	0
Triethylen glycol (1)	2	2.2	220 reflux	4	0	0
Triethylen glycol (1)	10	10.2	220 reflux	24	>60	<10
None	41	8.3	135[c]	24	>90	88

[a] Determined by ^{1}H NMR spectroscopy, [b] Pure isolated compound, [c] Reaction was run in a pressure tube.

Using these optimised conditions, this method was subsequently investigated for a wide range of aryl nitro carbonyl substrates and proved to be applicable to different systems including differentially substituted systems and heterocyclic derivatives. Only slight variations in yield, associated with the relative stability of the intermediate benzylic anion species in the carbonyl reduction, were observed (see Table 3.2).

Table 3.2. Double carbonyl/nitro reduction using hydrazine.



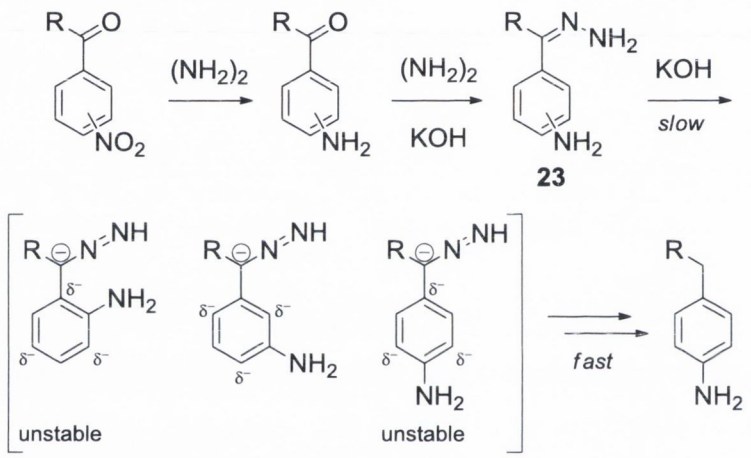
Entry	Substrate	Conversion[a] (%)	Yield [b] (%) (Fully reduced product)
1		>90	88 **(9c)**
2		>40	30 **(16)**
3		>90	83 **(17)**
4		>70	65 **(18)**
5		>40	33 **(19)**
6		>90	97 **(20)**
7		>80	76 **(21)**
8		>90	87 **(22)**

[a] Determined by ^{1}H NMR spectroscopy, [b] Pure isolated compound.

The method was not as effective for the reduction of aliphatic, heterocyclic (*i.e.* furane and thiophenes), or α,β-unsaturated nitro-carbonyl compounds; However, as can be seen in Table 3.2, this reduction strategy is an excellent method to reduce aromatic nitro-ketones or nitro-alkyl derivatives without any solvent and a yield ranging between 30 and 97%.

From these reactions it was possible to deduce a number of mechanistic details. First, in the case of compound **11**, the NO_2 group was reduced before the CO since the *bis*-anilino carbonyl derivative was isolated when only 2 eq. of hydrazine were used. Further evidence for this was obtained from the isolation of dimeric imines when aryl aldehyde substrates were used. Formation of these products could only be explained by the simultaneous presence of the amine and aldehyde groups. Differences in yield between the *ortho, meta*, and *para*-substituted substrates can be rationalised by the relative stability of the intermediate benzylic anion (Scheme 3.3).

Scheme 3.3.



The formation of this anion is thought to be the rate determining step of the reaction. The first deprotonation of the anilino-hydrazone (**23**) intermediate is more difficult for the *ortho* and *para*-substituted substrates, since conjugative electron-donation by the *N* lone-pair makes the *ipso*-carbon more electron-rich, destabilizing the benzylic anion (Scheme 3.3). It is assumed that re-protonation of this *C* atom by a water molecule and subsequent deprotonation of the azine with loss of nitrogen gas are fast processes, giving the alkane product.

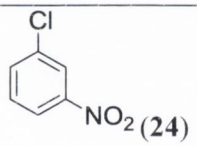
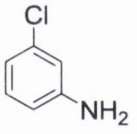
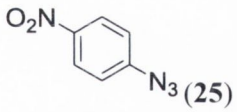
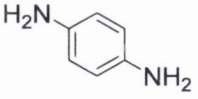
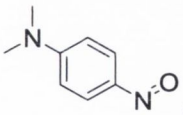
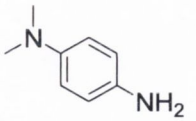
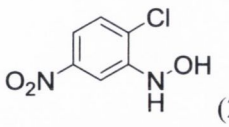
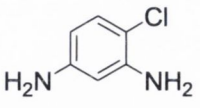
Reduction of nitro groups to amines following this method had not been reported; hence, we further explored a possible mechanism for this transformation. First, to check if the reduction of the aromatic nitro was influenced by the presence of the carbonyl group, the reduction of 1-chloro-3-nitrobenzene (**24**) was attempted. This reduction yielded the

corresponding amine, showing that the nitro reduction was independent of the presence of a carbonyl group.

Second, we explored the effect of the base, KOH, on the double reduction. Substrate **24** was reduced in the absence of base, suggesting that other species such as water or OH groups formed during the course of the reaction are sufficient to complete the reduction. Furthermore, when **11** was reacted in the absence of base, reduction of the nitro group was observed, but not that of the carbonyl group.

Considering the intermediates reported in the literature for alternative nitro reductions,[28] and to explore if the present conditions could also work on these intermediates, representative aromatic azide (**25**), nitroso (**26**), and hydroxylamine (**27**) substrates were synthesised according to literature procedures,[26,29,30] and were then subjected to our reduction conditions (Table 3.3).

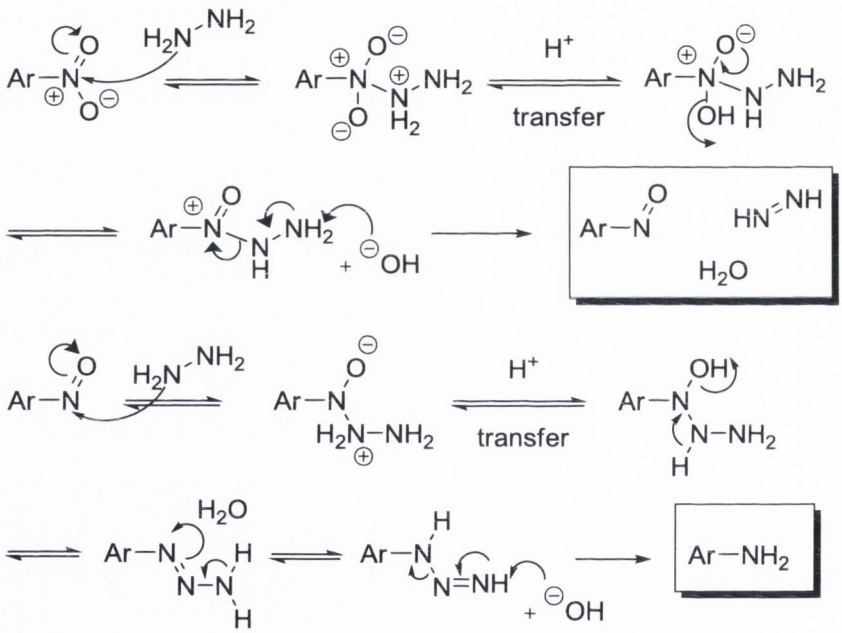
Table 3.3. Application of our reduction conditions to intermediates of the nitro reduction.

Substrate	$(NH_2)_2$ (mL)	KOH (eq.)	Temp. (°C)	Time (h)	Product	Yield[a] (%)
 **(24)**	2 Excess	5	135 pressure tube	24		98
 **(25)**	2 Excess	5	135 pressure tube	24		100
 **(26)**	2 Excess	5	135 pressure tube	24		93
 **(27)**	2 Excess	5	135 pressure tube	24		0

[a] Isolated yield of purified compound

The corresponding amines were obtained from compounds **25** and **26** but not from **27**. Furthermore, to confirm that hydrazine was involved in the reductive process, derivatives **24**, **25**, and **26** were reacted in its absence under the same conditions and no reduction was observed. In addition, reactions were carried out with hydrazine but without base, and reduction of the nitro **24** and nitroso **26** compounds was observed. Importantly, the azide derivative **25** was not reduced under these conditions. These results seem to suggest that the reduction of the nitro group proceeds *via* a nitroso intermediate and from these experiments a mechanism for the reduction of the NO_2 group through reaction with hydrazine can be proposed (Scheme 3.4).

Scheme 3.4.



The first step of this mechanism involves nucleophilic attack of hydrazine on the nitro N atom to form a 1-hydroxy-1-aryltriazine-1-oxide. Attack of a hydroxyl anion results in the formation of a nitroso intermediate and the release of diazene and water. A second hydrazine molecule can attack the nitroso derivative to yield a 1-hydroxy-1-aryltriazane which would undergo thermodynamically favoured transformation into the aniline derivative, water, and nitrogen. This hypothesis is in agreement with the lack of reactivity of the hydroxylamine derivative **27** and the positive results obtained with the nitroso

compound **26**. In addition, this mechanism would explain why the reduction of derivatives **25** and **26** could be performed in the absence of KOH.

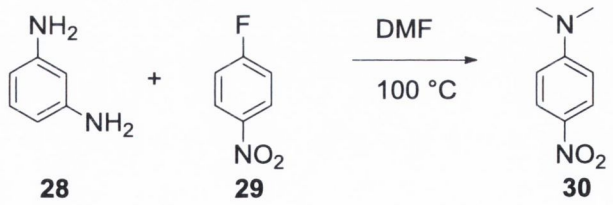
After the encouraging results obtained using the pressure tube, the reaction was attempted in a pressurised microwave to try to reduce the reaction time, unsuccessfully. The yields obtained ranged from 5-30%, revealing that the use of the microwave for this kind of reaction is not favourable.

The first objective of this reaction was to develop a methodology for the reduction of a problematic diphenyl carbonyl **11** to the corresponding diphenyl alkyl **9c**. In the process we achieved not only this reduction but also the simultaneous reduction of a variety of aryl-nitro groups to the corresponding anilines. This method of double reduction proved to be applicable to a variety of molecules and is a valuable route to alkyl-substituted anilines.

3.2.3. Synthesis of 3-(4-aminoanilino)aniline (9d)

For the synthesis of the NH-linked diamine (**9d**), a nucleophilic aromatic substitution (S_NAr or NAS) was first attempted. This involved the reaction of the nucleophilic benzene-1,3-diamine (**28**) with 1-fluoro-4-nitrobenzene (**29**), as the electrophile, in DMF (Scheme 3.5).

Scheme 3.5.



28 **29** **30**

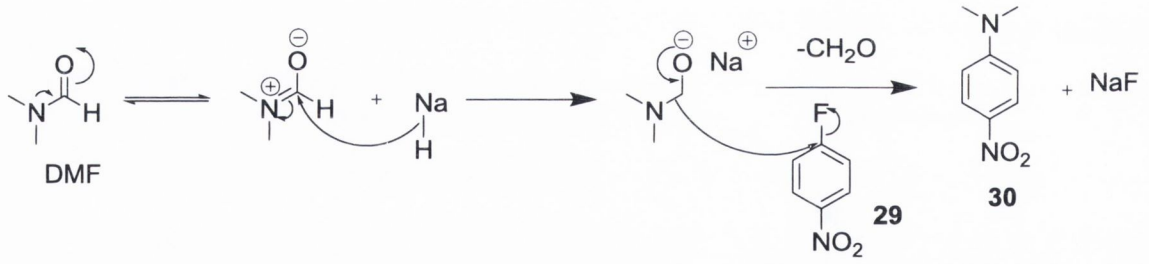
Using DMF, a polar-aprotic solvent typical of NAS reactions, a heterogeneous mixture was observed. After work-up, isolation and analysis of both the organic and aqueous layers was carried out by TLC, NMR and mass spectrometry. The NMR and mass spectrometry spectra clearly identified *N,N*-dimethyl-4-nitroaniline (**30**) as the major product originating from the attack of DMF on the fluoride. The formation of a salt was also identified. Searching in the literature, other examples of this type of side-product resulting from DMF

were found. It has been proposed that it decomposes into dimethylamine and carbon monoxide, and that the amine thus formed could be quaternised by another molecule of **29** generating the fluoride salt.[31]

Several conditions were attempted to overcome this problem with variation of solvent, time, temperature and concentration of the starting materials. Different amounts of DMF and other solvents with the same properties such as CH_3CN, THF and pyridine were tested; however, the NAS approach proved unsuccessful, with only trace amounts of the desired product **9d** ever being obtained.

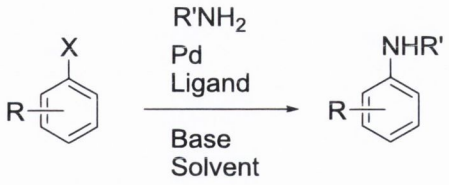
Use of sodium hydride (NaH) in dry THF to deprotonate the amine and favour the nucleophilic attack, followed by the addition of **29** directly or dissolved in DMF were also attempted. In the reaction without DMF the starting materials were always recovered, while in any reactions using DMF, the same side product **30**, was obtained. This result could be justified by considering the dual role that NaH displays in the reaction. It acts as a base and as a source of hydride (Scheme 3.6).[32] Hydride reduction of DMF could result in the formation of sodium (dimethylamino)methanolate, followed by a nucleophilic attack from the dimethylamide to the fluoride to afford **30**. As no product had been obtained in other solvents either, this approach was abandoned.

Scheme 3.6.



Finally, to overcome the problems encountered during the synthesis of **9d**, a palladium-catalysed C-N cross-coupling reaction was attempted, which successfully afforded the required compound after optimisation of the reaction conditions. The transition-metal catalysed coupling of aromatic amines is a useful strategy due to their attenuated nucleophilicity, relative to aliphatic amines, which limits the utility of NAS reactions in the formation of the diaryl amines. A general Pd-catalysed C-N cross-coupling reaction is shown in Scheme 3.7, where X is generally a halide.

Scheme 3.7.



The first example of a palladium-catalysed C-N cross-coupling reaction was published in 1983 by Migita and co-workers and described a reaction between several aryl bromides and *N*,*N*-diethylaminotributyl tin using 1 mol% $PdCl_2[P(o\text{-tolyl})_3]_2$.[33] But the credit for the development of this type of reaction is typically assigned to Stephen L. Buchwald and John F. Hartwig, whose publications between 1994 and the late 2000s established the scope of the transformation.[34-36] Consequently, this Pd-catalysed C-N bond formation reaction is known as the Buchwald-Hartwig amination reaction and it has rapidly emerged as a valuable tool in the synthesis and streamlining of small molecules as pharmaceutical agents, natural products and novel industrial materials.[37] The prevalence of aryl amines in biologically active molecules such as kinase inhibitors, antibiotics and CNS-active agents means that this reaction is of paramount importance.

The mechanism of the Buchwald-Hartwig Pd-catalysed C-N cross-coupling reaction is based on the Pd catalytic cycle, illustrated in Figure 3.5. This cycle starts by the formation of a Pd(0) complex or a Pd(II) species with the ligand as the active catalyst. After the formation of the L_n-Pd complex the oxidative addition of the aryl halide takes place. This involves the addition of a molecule Ar-X to Pd(0) with cleavage of its covalent bond, forming two new bonds to the metal. Since the two previously nonbonding electrons of Pd are now involved in bonding, the Pd increases its formal oxidation state by two units, namely Pd(0) is oxidised to Pd(II). Now, the coordination number of Pd is four and, hence, to form a stable 18-electron complex it requires four ligands, each donating two electrons, and the use of the d^{10} ($4d^8$ and $5s^2$) electrons of Pd. Once this relatively stable complex has formed, coordination of the nucleophilic amine to the metal centre occurs, followed by its deprotonation due to the base. Reduced electron-density of the metal centre and the bulkiness of the ligand then facilitate reductive elimination to ensure regeneration of the L_n.Pd(0) complex, with expulsion of the new C-N bond containing compound. Reductive elimination involves loss of two ligands with *cis* configuration from the Pd centre, and

their combination gives rise to the elimination product, being the last step of the catalytic cycle.[38] By reductive elimination, both the coordination number and the formal oxidation state of Pd(II) are also reduced by two units to regenerate Pd(0), which can restart the catalytic cycle.

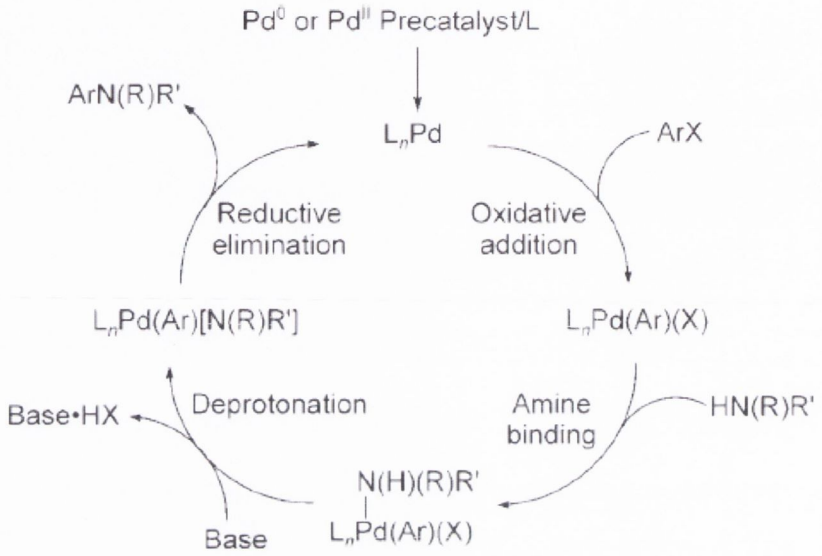


Figure 3.5. Palladium catalytic cycle of a C-N cross-coupling reaction.[39]

A β-hydride elimination side reaction can compete with the reductive elimination, leading to a reduced arene by-product (Ar-H).[40] This occurs when the amine has hydrogens in the carbon which is in the α-position relative to the nitrogen (*e.g.* acyclic secondary amines) and when electron-donating aryl halides are employed, since electron-deficient palladium species more readily undergo reductive elimination.[41]

Several factors affect the efficacy of this reaction. The choice of ligand, base and solvent are all critical to the success of the reaction, but this is also dependent on the choice of substrates to be coupled.[42] There is a plethora of reaction conditions described for each type of substrate; however, this discussion will focus on analysing the best conditions for our experiment involving reaction between an aryl halide and an aryl amine. It has been published that aryl units with electron-withdrawing groups accelerate the reductive elimination step. For this reason, we incorporated nitro groups in both the aryl halide and

amine; this was also useful as the nitro group could easily reveal the amine needed for the addition of the guanidine functionality to the molecules, on reduction.[38]

A broad range of highly complex and advanced phosphine ligands has been commercialised (Figure 3.6) such as BrettPhos, XPhos and RuPhos, but simple phosphines such as JohnPhos and (2-diphenyl)dicyclohexylphosphine are also excellent ligands for C-N bond formation with aryl chloride substrates.[43] The effectiveness of the latter ligand is believed to be due to a combination of its unique steric and electronic properties. Its electron-rich nature promotes oxidative addition and its steric bulk facilitates reductive elimination and promotes reactivity *via* monophosphine intermediates, which presumably promote all steps of the catalytic cycle.[43] The rest of the ligands have several features in common with (2-diphenyl)dicyclohexylphosphine as they also have bulky and electron-rich substituents attached to the phosphine, which accelerate the reductive elimination step and enhance the rate of the oxidative addition step. However, these ligands are very expensive. For these reasons, and because it was more affordable than the rest, (2-diphenyl)dicyclohexylphosphine was selected as the first ligand.

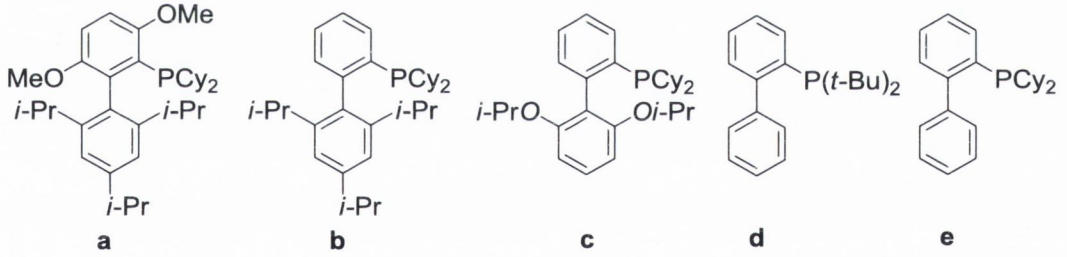


Figure 3.6. Ligands used in Pd-catalysted reactions (a) BrettPhos (b) XPhos, (c) RuPhos, (d) JohnPhos, (e) (2-diphenyl)dicyclohexylphosphine (**31**).

The efficiency by which the catalytically active *mono*-ligated Pd(0) complex is formed before entry into the catalytic cycle is a key factor in the selection of reaction conditions for Pd-catalysed amination reactions. In the case of $Pd(OAc)_2$, reduction of Pd(II) to Pd(0) must occur before the cross-coupling reduction can take place. However, this reduction step can be circumvented by the use of stable Pd(0) sources such as $Pd_2(dba)_3$, in which the oxidation state of Pd is already Pd(0), making the complex formation between Pd and the ligand more favoured.[36]

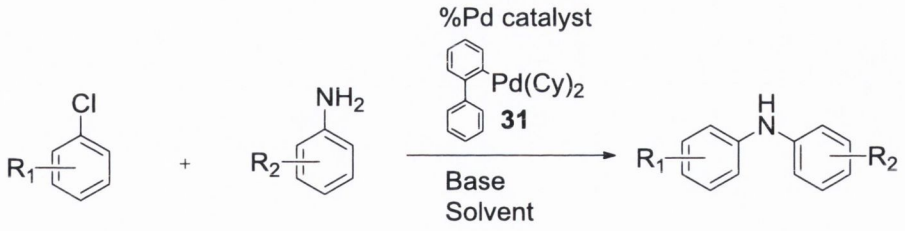
Another factor to take into account is that preheating the Pd catalyst and the ligand in the solvent, prior to the introduction of substrates, can be beneficial as it enhances the efficient formation of the Pd-ligand complex.[42] Additionally, the ideal ratio of metal to ligand in aryl chloride cross-coupling reactions is 1:1, but with other types of substrates, or in cases with long reaction times, an extra equivalent of ligand might be needed in order to stabilise the catalyst.[36]

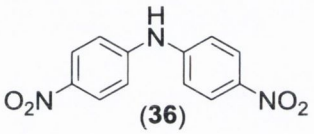
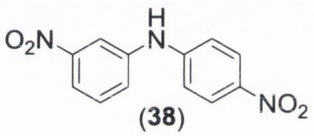
The C-N cross-coupling reaction can be performed in a wide range of solvents. Toluene and 1,4-dioxane are most commonly employed, although the latter has an unfavourable toxicity profile and can be replaced by THF or DME. The selection of the solvent mainly depends on the solubility profile of the substrates. It is worth noting that there is often an interaction between the solvent and the base used; hence, the best base for a reaction may change depending on the solvent employed, meaning higher reaction rates and lower catalyst loading can be attained with the correct combination of solvent and base. Some studies suggest the use of NaO^tBu as a base in toluene for coupling amines with aryl halides.[44] However, NaO^tBu is a relatively strong base, and it can participate leading to undesired side reactions.

Alternatively, weak inorganic bases such as Cs_2CO_3, K_3PO_4 or K_2CO_3 can bring significant benefits in terms of the functional group tolerance of Pd-catalysed amination reactions.[45] These bases provide more general conditions for substrates containing electrophilic functional groups such as nitro aromatics. These inorganic bases have low solubility in non-polar solvents such as toluene and, thus, are commonly used with DME.[43]

Taking into consideration the preferred conditions for the reaction of aryl chlorides and aryl amines described in the literature, a screen of conditions was carried out for our particular reaction (Scheme 3.8 and Table 3.4).

Scheme 3.8.



R_1	R_2	Product
m- NBoc, NHBoc (**32**)	*p*- NBoc, NHBoc (**33**)	-
p- NO_2 (**34**)	*p*- NO_2 (**35**)	 (**36**)
m- NO_2 (**37**)	*p*- NO_2 (**35**)	 (**38**)

As previously mentioned, the choice of the substrates to be coupled is crucial for the reaction to succeed. Firstly, the two substrates used as a test of the coupling reaction using $Pd(OAc)_2$ and $Pd_2(dba)_3$ were **32** and **33**,but not success was achieved (Table 3.4, entry 1 and 2). The idea behind this reaction was to effect the coupling with the guanidine groups already in place, but due to steric hindrance of the amine caused by the Boc groups, the catalytic reaction could not occur. It is reported that the cross-coupling of hindered amines remains a challenge for the further development of even more efficient and generally applicable catalyst systems.[42] As a result, less complex and bulky reactants were used in the following attempts.

Table 3.4. Optimisation conditions of the Pd-catalysed C-N cross-coupling reaction.

Entry	Aryl Chloride	Aryl Amine	Product	Pd/Lig 31 (%)	Temp/time (°C/h)	Solvent (mL)	Base	Yield (%)
1	**32**	**33**	-	$Pd(OAc)_2$ 5.0/5.0	110/22	Toluene	KO^tBu	0 [a]
2	**32**	**33**	-	$Pd_2(dba)_3$ 5.0/5.0	110/22	Toluene	KO^tBu	0 [a]
3	**34**	**35**	**36**	$Pd_2(dba)_3$ 1.0/1.0	100/24	DME	K_3PO_4	80 [a]
4	**34**	**35**	**36**	$Pd_2(dba)_3$ 1.0/1.0	100/24	Toluene	NaO^tBu	40 [a]
5	**34**	**35**	**36**	$Pd(OAc)_2$ 1.0/1.0	100/24	DME	K_3PO_4	60 [a]
6	**37**	**35**	**38**	$Pd_2(dba)_3$ 5.0/5.0	100/24	DME	K_3PO_4	34 [a]
7	**37**	**35**	**38**	$Pd_2(dba)_3$ 1.0/1.0	100/24	Toluene	NaO^tBu	<5 [a]
8	**37**	**35**	**38**	$Pd_2(dba)_3$ 1.0/1.0	100/24	DME	K_3PO_4	43 [b]
9	**37**	**35**	**38**	$Pd_2(dba)_3$ 1.0/1.0	100/24	Toluene	NaO^tBu	10 [b]
10	**37**	**35**	**38**	$Pd_2(dba)_3$ 1.0/1.0	100/24	DME	K_3PO_4	48 [b,c]
11	**37**	**35**	**38**	**$Pd_2(dba)_3$ 3.0/3.0**	**100/24**	**DME**	**K_3PO_4**	**80** [b,c]
12	**37**	**35**	**38**	$Pd_2(dba)_3$ 3.0/3.0	100/24	DME	K_3PO_4	75 [c,d]
13	**37**	**35**	**38**	$Pd_2(dba)_3$ 5.0/5.0	100/24	DME	K_3PO_4	31 [c,d]
14	**37**	**35**	**38**	$Pd_2(dba)_3$ 3.0/3.0	100/72	DME	K_3PO_4	24 [c,d]
15	**37**	**35**	**38**	$Pd_2(dba)_3$ 3.0/3.0	100/24	DME	K_3PO_4	31 [b,e]
16	**37**	**35**	**38**	$Pd_2(dba)_3$ 3.0/3.0	100/24	DME	K_3PO_4	0 [b,f]
17	**37**	**35**	**38**	$Pd_2(dba)_3$ 3.0/3.0	100/24	DME	K_3PO_4	0 [b,c,f]

[a] Reaction conditions: 1.0 eq. chloride, 1.2 eq. amine, 1.4 eq. base, 2 mL of solvent/mol of chloride. [b] Reaction conditions: 1.0 eq. chloride, 1.0 eq. amine, 1.4 eq. base, 2 mL of solvent/mol of chloride. [c] $Pd_2(dba)_3$ and ligand pre-react for 5 min for the formation of the complex prior to the addition of the rest reagents. [d] Reaction conditions: 1.2 eq. chloride, 1.0 eq. amine, 1.4 eq. base, 2 mL of

solvent/mol of chloride. [e] $Pd_2(dba)_3$ and ligand pre-react for 20 min for the formation of the complex prior to the addition of the rest reagents. [f] DME was degassed prior to use.

There are several reasons why we chose aryl chlorides such as **34** and **37**. Even though the rate of oxidative addition of organic halides to sp^2 carbons decreases in the order C-I > C-Br >>> C-Cl >>> C-F, the formation of C-N bonds from aryl chlorides is more efficient and they are used more frequently than aryl bromides.[46] One study found in the literature demonstrates that, while the overall process of amination is faster for the aryl chloride than for the aryl bromide substrates, as expected, the oxidative addition of the aryl bromide susbtrate is faster than that of the aryl chloride. The reductive elimination process should proceed at the same rate for both aryl chloride and aryl bromide; thus, the Pd-N bond-forming step must be slower for the LPd-(Ar)Br complex than for the LPd-(Ar)Cl complex, and this is presumably the rate-limiting step.[46]

Intriguingly, aryl iodides, which are typically the most successful class of electrophiles for C-C cross-coupling reactions, are relatively challenging substrates in Pd-catalysed amination reactions. Currently, there is no definitive explanation for the differences observed between the reactions which employ aryl iodides and those which use aryl bromides or chlorides. Continuing mechanistic experiments indicate that the coordination chemistry involving the oxidative addition products of aryl iodides, with respect to their stoichiometry and reactivity, is substantially different.[47] Another study suggests that the reason for the unsuccessful reactivity of aryl iodides in Pd-catalysed amination reactions could be due to formation of an alkali-metal iodide product, instead of the iodoarene itself, slowing the rate of the reaction.[48]

For this reason we decided to incorporate nitro groups into both the aryl halide and aryl amine.[38] This had the added benefits of introducing a masked amine into the product, which is needed for subsequent steps in our synthetic route.

We first tried the Pd-catalysed C-N cross-coupling reaction with 1 eq. of *p*-chloronitro benzene **34** and 1.2 eq. of *p*-nitroaniline **35**, using 1.0% of $Pd_2(dba)_3$, 1.0% of ligand **31**, DME as solvent and K_3PO_4 as base, at 100 °C for 24 h, and 80% conversion was attained (Table 3.4, entry 3). This result was very encouraging as it was the first time that C-N coupling was successfully obtained. The solvent and base were then changed to toluene

and NaO^tBu to improve conversion (Table 3.4, entry 4). The reason for changing the solvent and base simultaneously was that specific bases only work in certain solvents; however, in this case the conversion was halved.

Then, keeping the same conditions as the successful 80% conversion, the source of Pd was varied to $Pd(OAc)_2$ (Table 3.4, entry 5). However, the conversion obtained was 60% and therefore, the Pd source chosen for the rest of the experiments was $Pd_2(dba)_3$. This outcome was to be expected because, as explained before, the L_n-Pd complex only forms when Palladium is in oxidation state Pd(0). In $Pd(OAc)_2$, the oxidation state of palladium is Pd(II), so reduction of Pd(II) to Pd(0) must occur before the cross-coupling reaction can take place. However, with $Pd_2(dba)_3$, the reduction step can be avoided as palladium is already in the stable Pd(0) state; therefore, the complex between Pd and the ligand is more favoured.[36] It has also been suggested that this type of reaction was inefficient for primary anilines if $Pd(OAc)_2$ was employed as a precatalyst; however, use of $Pd_2(dba)_3$ for this reaction provided excellent results.[43]
So far, the substituents used afforded the 4-nitro-*N*-(4-nitrophenyl)aniline (**36**); however, the *meta, para* substitution pattern was desired, to form 3-nitro-*N*-(4-nitrophenyl)aniline (**38**). For this reason, *m*-chloro-nitrobenzene (**37**) and *p*-nitroaniline **35** were subsequently coupled.

In the first attempt with the new substrates, the previous stoichiometric conditions were maintained (1 eq. of the aryl chloride and 1.2 eq. for the aryl amine, use of $Pd_2(dba)_3$, ligand **31**, DME and K_3PO_4 at 100 °C for 24 h), but in this case the catalytic loading of the Pd source and ligand was increased to 5.0% each to make sure that the reaction had enough active catalytic species to work and in case some of the Pd was not regenerated (Table 3.4, entry 6). These conditions afforded the product **38** in 34% yield. To improve upon this, the catalyst and ligand loading were reduced to 1.0% as before; however, the solvent and base were changed to toluene and NaO^tBu, respectively (Table 3.4, entry 7). This had an adverse effect on conversion, reducing the yield to less than 5%.

The literature was revisited to find ways of optimising the reaction. One of the points observed was that the stoichiometric ratio used in a lot of C-N cross-coupling reactions is 1:1 (aryl halide:aryl amine). These reaction conditions were applied in the next experiment (Table 3.4, entry 8), keeping the remaining variables constant as in entry 3. This change was successful, with a yield of 43% being obtained, the highest conversion obtained so far

for substrate **38**. In the next experiment, toluene and NaO^tBu were used to try to achieve higher conversion but this drastically dropped to 10% (Table 3.4, entry 9). Therefore, toluene and NaO^tBu was avoided for subsequent experiments.

Another aspect observed in the literature was that pre-reacting step of the Pd catalyst and the ligand in solution, prior to the introduction of substrates, which can have beneficial effects because it enhances the efficient formation of the Pd-ligand complex.[42] For this reason, $Pd_2(dba)_3$ and ligand were pre-reacted for 5 min prior to the addition of the remaining substrates and reagents. This change made a significant improvement to the conversion of **38** to the coupled product. Maintaining the previous reaction conditions (1:1 ratio of aryl halide and aryl amine, 1.0% loading of $Pd_2(dba)_3$ and ligand, DME and K_3PO_4 at 100 °C for 24 h), but leaving the catalyst and ligand to pre-react for 5 min in DME, the conversion increased to 48% (Table 3.4, entry 10).

In the next experiment, the loading of catalyst and ligand was increased to 3.0% to check if the rate of the catalytic cycle could be increased by increasing the amount of active catalytic sources (Table 3.4, entry 11). The results obtained after this change were very encouraging, 80% conversion being obtained. These conditions led to the best results so far obtained and were very close to the yields obtained in the literature for similar reactions. The reaction was analysed by TLC and ^1H NMR spectroscopy, and it was observed that all of the aryl chloride was consumed; however, some aryl amine remained unreacted.

An excess of amine **35** after the reaction is a problem because its separation from product **38** is difficult due to their similar Rf values, and hence elution times in silica gel chromatography. To avoid this problem, the stoichiometry of the reaction was modified to 1.2 eq. of the aryl chloride and 1.0 eq. of the aryl amine (Table 3.4, entry 12). The conversion was very similar to that obtained when using a 1:1 ratio of the reactants, but the purification problem was still present. In an attempt to overcome this problem, the catalyst loading was increased to 5.0% (Table 3.4, entry 13) and – in separate experiments – the reaction time was increased to 72 h (Table 3.4, entry 14), but the conversion observed was less effective in all cases.

Lower yields were also observed when the catalyst and the ligand pre-reacting time was increase to 20 min instead of 5 min (Table 3.4, entry 15). It transpired that not only does longer pre-reacting time not help to stabilise the Ln-Pd complex, but it also has a negative

effect on the conversion to the product. Similarly, degasification of the solvent using the freeze-pump-thaw procedure (Table 3.4, entry 16 and 17) led to no improvements.

In conclusion, the best reaction conditions for the synthesis of compound **38** (80% yield) were the use of a Pd-catalysed cross-coupling reaction using a 1:1 ratio of aryl halide **37** and aryl amine **35**, pre-reaction for 5 min of 3.0% $Pd_2(dba)_3$ and 3.0% ligand **31**, DME as solvent, K_3PO_4 as base, at 100 °C for 24 h (Table 3.4, entry 11).

Once the synthesis of **38** was completed, reduction of the nitro groups to the corresponding amines was carried out using a catalytic hydrogenation reduction, with 10% Pd/C, to generate diamine **9d** in 88% yield. The reaction conditions of the hydrogenation are ideal because no additional extraction or purification was needed. Owing to the fact that the product has high polarity due to its two primary amines and NH linker, workup with water was avoided. This was one of the reasons other attempted reduction methods did not work. For example, the use of $Na_2S.9H_2O$ in EtOH as reducting agent led to difficult purifications and a small percentage of the mono reduced substrate (N^1-(4-nitrophenyl)-3-aminoaniline)as well. Thus, diamine **9d** was obtained *via* a Pd-catalysed cross-coupling reaction and a catalytic hydrogenation in an overall yield of 70%.

3.3. Synthesis of 4,4'- and 3,4'-*Bis*-Guanidine-like Derivatives – Families 1 and 2

3.3.1. General Aspects of Families 1 and 2

The 4,4'-*bis*-guanidine-like derivatives were previously synthesised in Rozas' laboratory as DNA minor groove binders. However, the 3,4'-*bis*-guanidine-like derivatives were the first two families prepared as protein kinase inhibitors. The 3,4'-substitution pattern (as present in sorafenib) has been introduced to produce 3,4'-*bis*-guanidinium and 3,4'-*bis*-(2-aminoimidazolinium) derivatives (Families 1 and 2, respectively, Figure 3.7).

Some of the already known 4,4'-*bis*-guanidinium-like derivatives were synthesised to evaluate them as potential protein kinase inhibitors, since some of them did not bind to DNA but still exerted cytotoxicity in cancer cells. Furthermore, these molecules would allow us to study the effects of the substitution pattern in relation to the induction of kinase inhibition.

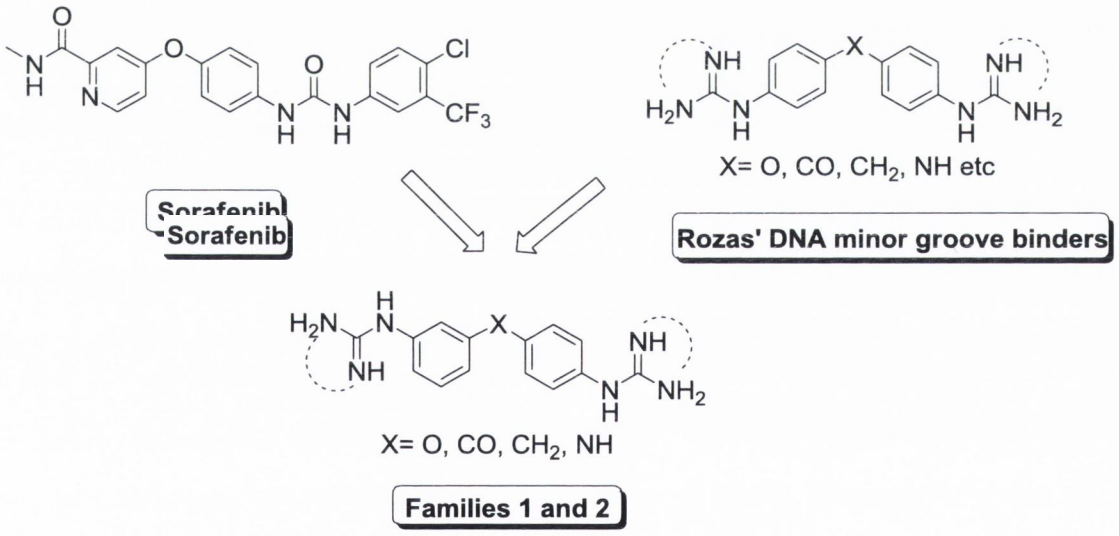


Figure 3.7. Rational behind Families 1 and 2.

The retrosynthesis of the desired families of compounds can be proposed as described in Figure 3.8.

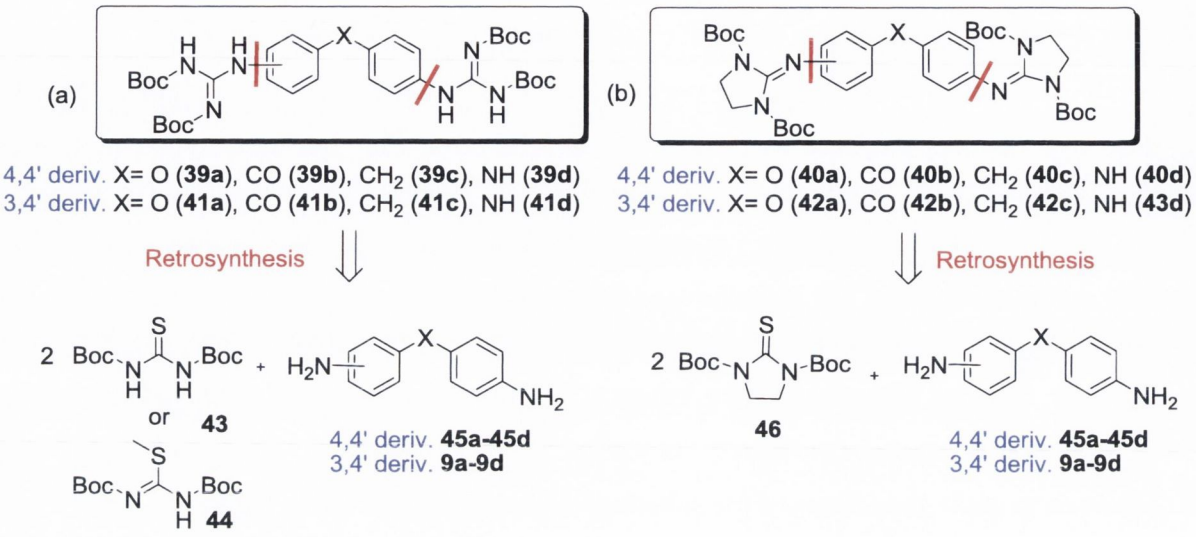


Figure 3.8. Retrosynthesis of (a) 4,4'- and 3,4-*bis*-guanidine and (b) 4,4'- and 3,4'-*bis*-2-aminoimidazoline derivatives.

Even though at first sight the differences between the 4,4'- and 3,4'-*bis*-guanidine like derivatives are scarce, their corresponding syntheses are highly different. All of the 4,4'-diphenyl diamines (**45a-d**) are commercially available; however, as was previously discussed, the 3,4'-diphenyl diamines **9b-9d** are not, with the exception of 3-(4-aminophenoxy)aniline **9a**, and thus, they had to be synthesised. Once all the diphenyl diamines were obtained, the guanidylation process is the same for the two families of substrates. The Kim and Qian[49] and the Dardonville and Rozas[50] methods were used, which involve mercury(II) chloride-promoted coupling of a primary aryl amine with the guanidylating agents **43** or **44** to form guanidines or the 2-iminoimidazolidylating agent **46** to form 2-aminoimidazolines.

The method works extremely well and despite the undesirable use of mercury salts – and the mercury-containing waste that ensues – it is particularly useful where unreactive aryl amines are involved. It is the method of choice in the Rozas group, and also in late stage natural products syntheses where guanidines need to be introduced to sensitive molecules,

such as in tetrodotoxin (TTX), (+)-saxitoxin [(+)-STX] and two enduracididin stereoisomers among others (Figure 3.9).[1,2]

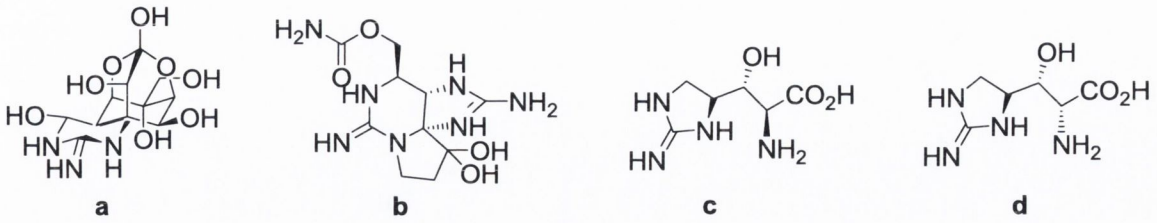


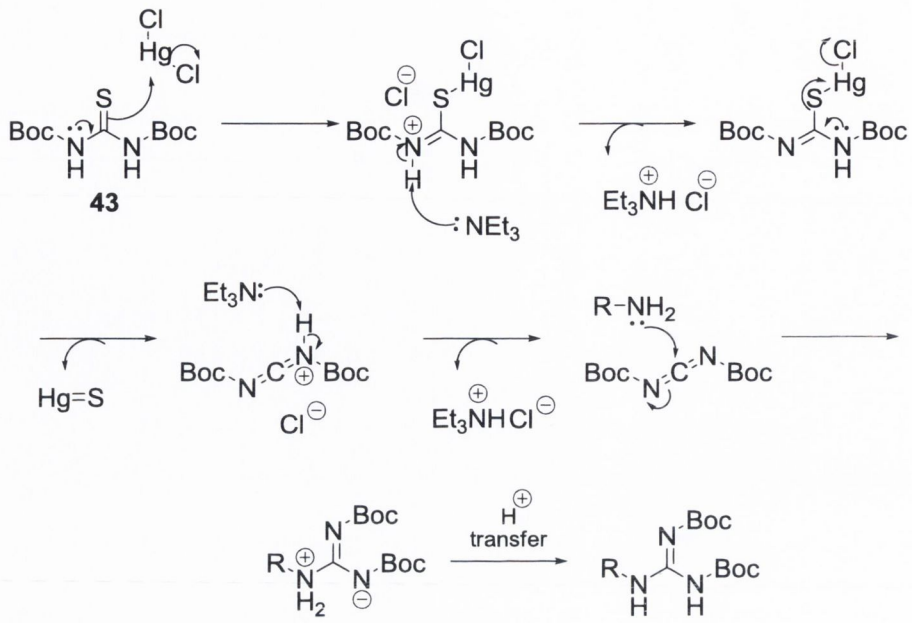
Figure 3.9. Guanidine-containing natural products synthesised with the $HgCl_2$ guanidylation method (a) tetrodotoxin, (b) (+)- saxitoxin, (c) and (d) enduracididin stereoisomers.

The mercury(II) chloride mechanism of guanidylation used for the synthesis of the 4,4'- and 3,4'-*bis*-guanidine like derivatives will be discussed in more detail in the following section.

3.3.2. Mercury(II) Chloride Guanidylation and 2-Iminoimidazolidylation Method.

The guanidylation method was developed by Kim and Qian[49] for the conversion of amines to guanidines using thiourea derivatives and mercury(II) chloride. In this procedure, the amine is reacted with *N,N'-bis-*(*tert*-butoxycarbonyl)thiourea (**43**) in the presence of mercury(II) chloride and triethylamine to generate *bis*-Boc-protected guanidines. The Boc groups on the thiourea are necessary to increase the electrophilicity of the thiourea carbon, making this method very suitable for poorly reactive aryl amines. In the original article describing this method, it was suggested that the reaction proceeds through a mercury sulphur coordination, followed by deprotonation by triethylamine, leading to the electrophilic intermediate carbodiimide species, which is formed *in situ* after desulfurisation of the thiourea derivative, forming a mercury by-product (Scheme 3.9). The carbodiimide is an excellent electrophile, which undergoes nucleophilic attack even by deactivated amines, leading to the *bis*-Boc-protected guanidine after proton exchange.

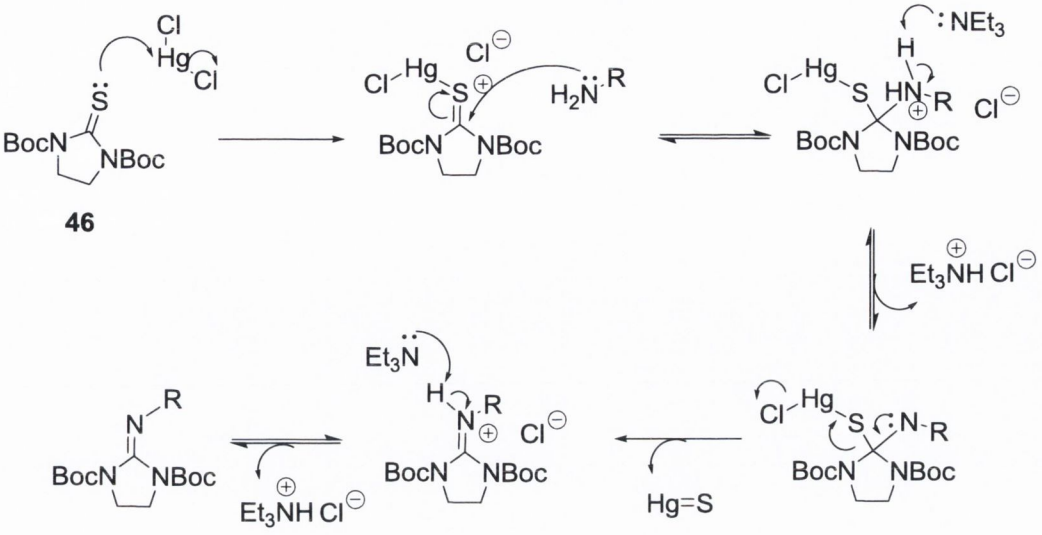
Scheme 3.9.



The authors note that they were unable to isolate the *bis*-Boc carbodiimide intermediate; however, they observed spectroscopic data *in situ* which agrees with the presence of a mercury complex of *bis*-Boc carbodiimide.[49] This finding provides insight into the structure of the reactive intermediate and into the complexation chemistry of the mercury(II) cation with sulphur.

The 2-iminoimidazolidylation process, adapted by Dardonville and Rozas,[50] consisted in the reaction of the arylamine with *N,N'-bis*-(*tert*-butoxycarbonyl)-imidazolidine-2-thione (**46**) in the presence of mercury(II) chloride and triethylamine to generate the corresponding *bis*-Boc-protected 2-imidazolidines. In the formation of these derivatives, the mechanism of guanidylation cannot be the same as described for **43** (Scheme 3.10). The reason is that the required intermediate would incorporate a carbodiimide into a five membered ring, which seems impossible given the degree of strain that this system would possess. For this reason, the mechanism of insertion of 2-aminoimidazolines is thought to be through an addition-elimination mechanism to avoid this problem.

Scheme 3.10.



Similarly, there are several papers which suggest similar addition-elimination mechanisms for related thiourea-based compounds, such as triazinium-imidothioate[51] and the guanidylation reaction of *N*-iminopyridinium ylide with *N*-benzoylthioureas,[52] providing reasonable precedent for this mechanism.

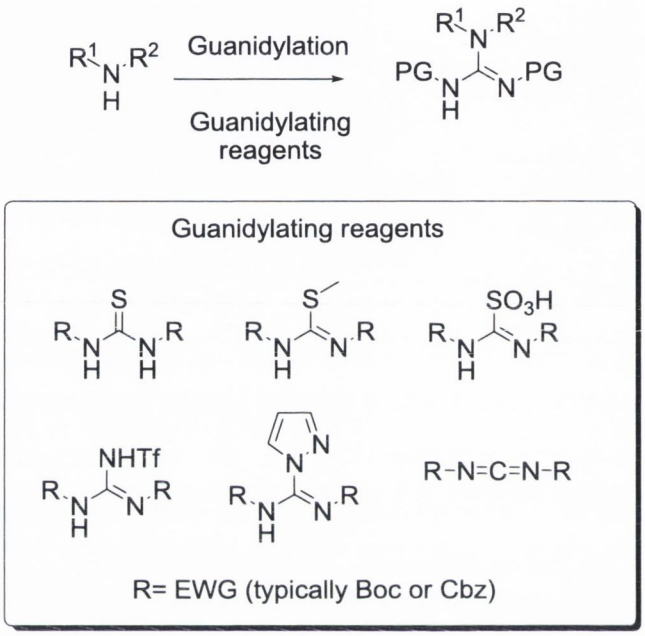
Despite the efficiency of the mercury(II) chloride method for guanidylation, it is not employed for the synthesis of commercially available drugs due to its toxicity and because it is not very practical for use on a large scale. Some other alternatives to this method are described in the next section; however, it should be kept in mind that most of them only work for alkyl amines, and not for poorly reactive aryl amines.

3.3.3. Alternative Guanidylation-like Methods

In general, the guanidylation reaction formally involves the attachment of a -C(=NH)NH_2 group (or substituted derivative), traditionally referred to as a amidine group. Most of these methods involve the reaction of a primary or secondary amine with guanidylating reagents (Scheme 3.11). Significant advances in guanidine synthesis were achieved with the introduction of more activated carbamate-protected (typically Boc or Cbz) reagents that allow for the direct synthesis of protected guanidines. Unlike free guanidines, protected guanidines are less polar and basic and can be easily purified by chromatography, and

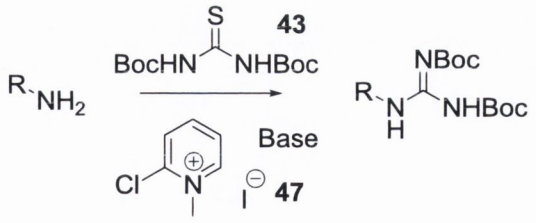
utilised in subsequent reactions. Improvements to this process have focused on the use of more electrophilic guanidine species and better leaving groups that can be displaced under milder conditions.

Scheme 3.11.



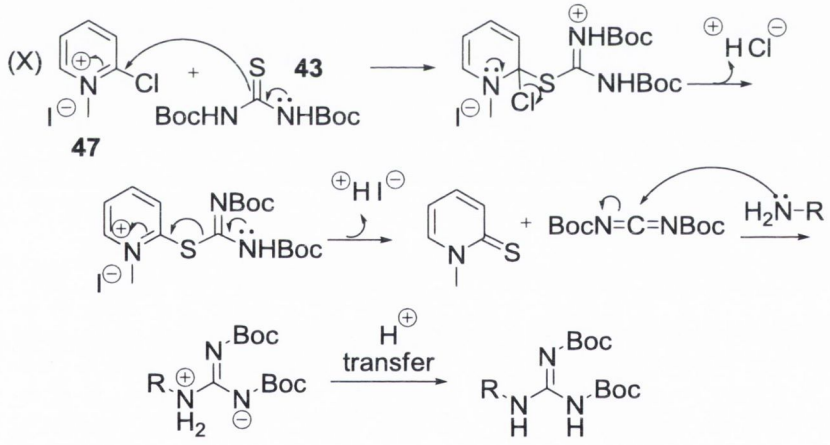
There is a wide variety of guanidylation methods available to introduce the strong organobase, guanidine, in synthetic procedures. However, most of them have the same limitation; they only work for highly nucleophilic amines. One alternative to the mercury(II) chloride guanidylation method is the use of Mukaiyama's reagent (**47**, Scheme 3.12) which, like the former, is based on the formation of the very electrophilic carbodiimide intermediate.[53]

Scheme 3.12.



The Boc-protected thiourea **43** binds to **47** through its sulphur atom (Scheme 3.13). The chlorine atom is then ejected to leave a positively charged intermediate, which, after some electron rearrangements, will form a double bond with the sulphur of the thiourea that is a good leaving group. This affords as well the carbodiimide intermediate, which will be attacked by the amine to finally form the Boc-protected guanidine.

Scheme 3.13.

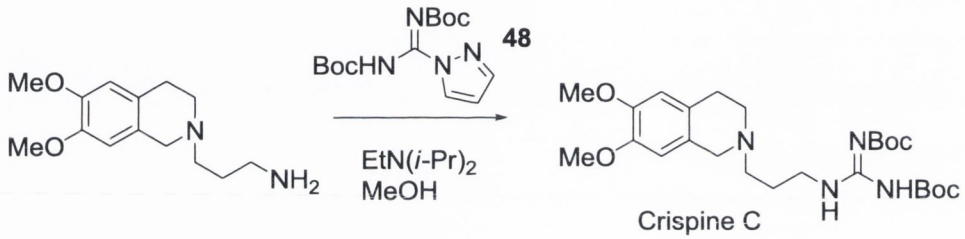


The method is effective for aliphatic amines, with very high yields; however, it is not as efficient for aromatic amines due to their low nucleophilicity. The problem with aryl amines is that their nucleophilic attack is slow, and then a competitive decomposition of the carbodiimide occurs leading to loss of reagent and thus lower yield.

Another example of a guanidylation reaction using one of the reagents shown in Scheme 3.11, is that used in the synthesis of crispine C (Scheme 3.14). It involves the use of *N,N'*-

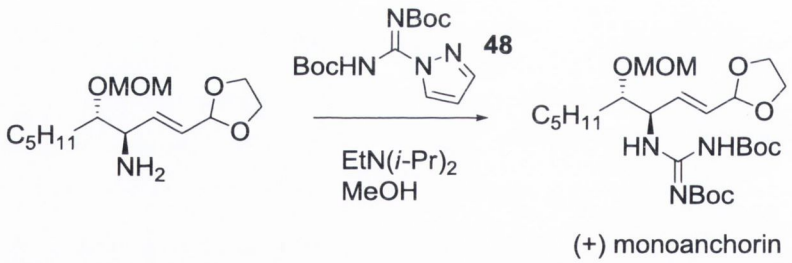
bis-(tert-butoxycarbonyl)-1H-pyrazole-1-carboxamidine (**48**) as the electrophilic source and Hünig's base as base to afford the Boc-protected guanidine in 82% yield.[54] This method is based on attaching a good leaving group to one of the nitrogens of the guanidine (pK_a of pyrazole is 19.8), so the nucleophilic attack of the amine is more favoured; however, yields for the guanidylation with aryl amines are again low.

Scheme 3.14.



These same methodology was employed in one of the steps of the synthesis of the alkaloid (+)-monoanchorin (Scheme 3.15).[2]

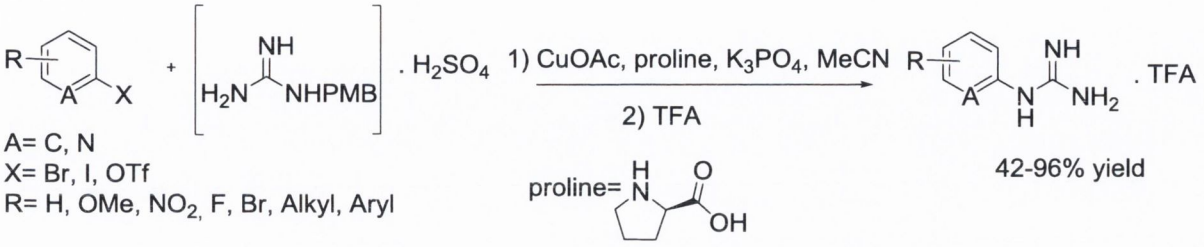
Scheme 3.15.



Additionally, the use of palladium-catalysed or copper-catalysed guanidylation reactions is another alternative to the methods previously mentioned. An example of Cu-catalysed guanidylation is the use of CuOAc and *N-p*-methoxybenzyl (PMB) guanidine (Scheme 3.16)[55] This method is particularly efficient as it dramatically suppresses double substitution, which generally occurred in methods using Cu-catalyzed guanidylation

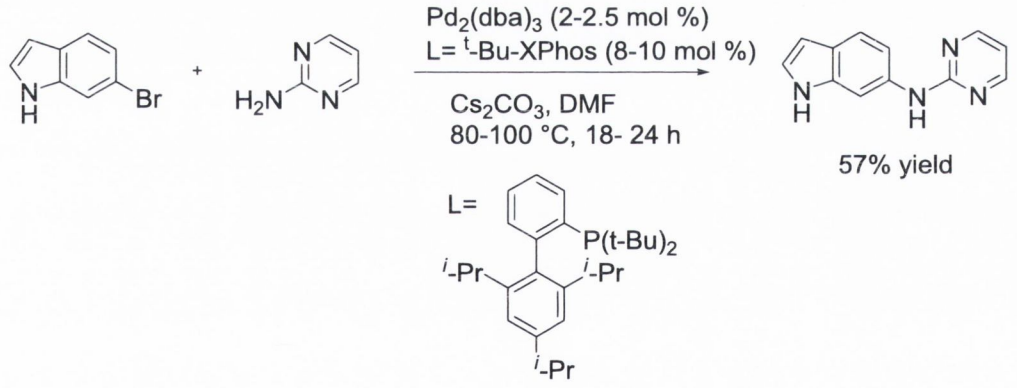
reactions with free guanidine. Easy removal of the PMB protecting group in TFA subsequently led to the guanidine salts.

Scheme 3.16.



An example of a palladium-catalysed guanidylation reaction is the *N*-arylation of heteroaromatic guanidines using Pd/^{t}Bu-XPhos as catalytic system (Scheme 3.17).[4] Many other palladium-catalysed (and other transition metal-catalysed) *N*-arylations of guanidines, are described in the referenced publication.[4]

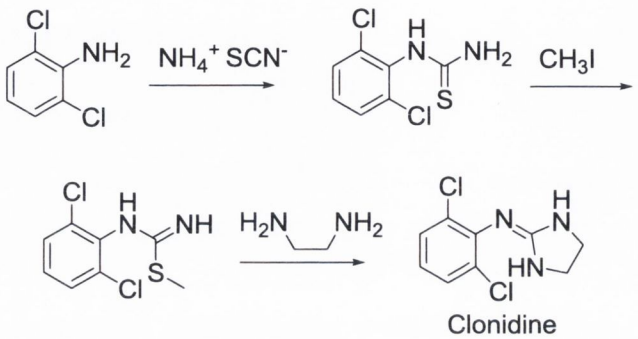
Scheme 3.17.



There are also alternatives for the synthesis of 2-aminoimidazoline compounds, avoiding the use of mercury(II) chloride. For example, the 2-aminoimidazoline moiety of clonidine (Scheme 3.18), an α_2-AR agonist used in the treatment of high blood pressure, attention deficit-hyperactivity disorder (ADHD), anxiety/panic, and certain pain conditions, is introduced from the corresponding aryl amine by first converting it to the thiourea using ammonium thiocyanate. This is then *S*-methylated by iodomethane and attacked by the

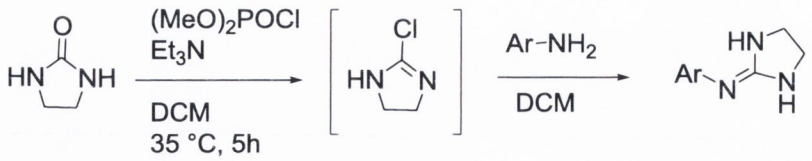
powerfully nucleophilic ethylenediamine.[56] However, this method also requires the use of light-sensitive and highly toxic iodomethane.

Scheme 3.18.



Another method to synthesise 2-aminoimidazolines involves using aniline derivatives with 2-chloro-2-imidazolidine, generated *in situ* from inexpensive imidazolidin-2-one, to afford good to excellent yields of the desired 2-aryliminoimidazolidines (Scheme 3.19).[57] The use of aniline for the synthesis of 2-aryliminoimidazolidines is widely used due to the broad range of anilines which are commercially available; however, in light of the poor nucleophilicity of the aromatic amines this method might not be the most appropriate in all cases.

Scheme 3.19.



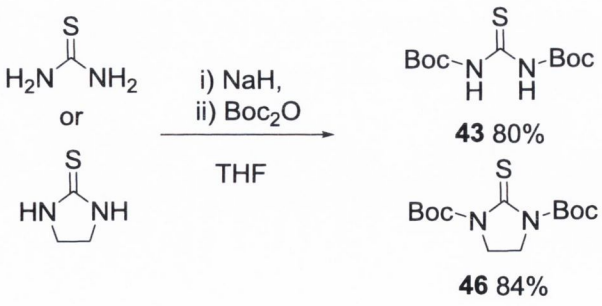
It is worth noting that the synthesis of guanidines is an area of active research, as these functionalities represent important structural motifs. They turn up as a structural feature in various drugs, agrochemicals, and in a variety of natural products. Guanidines are also important precursors used as reagents in the synthesis of new chemical entities with

promising biological activities[58] such as benzimidazoles, quinazolines and the *N*-arylation of guanidines.[4,58] For this reason, there is still a need for new, more environmentally friendly and efficient synthetic routes to guanidines.

3.3.4. Synthesis of 4,4'- and 3,4'-*bis*-Boc-Protected Guanidine-like Derivatives

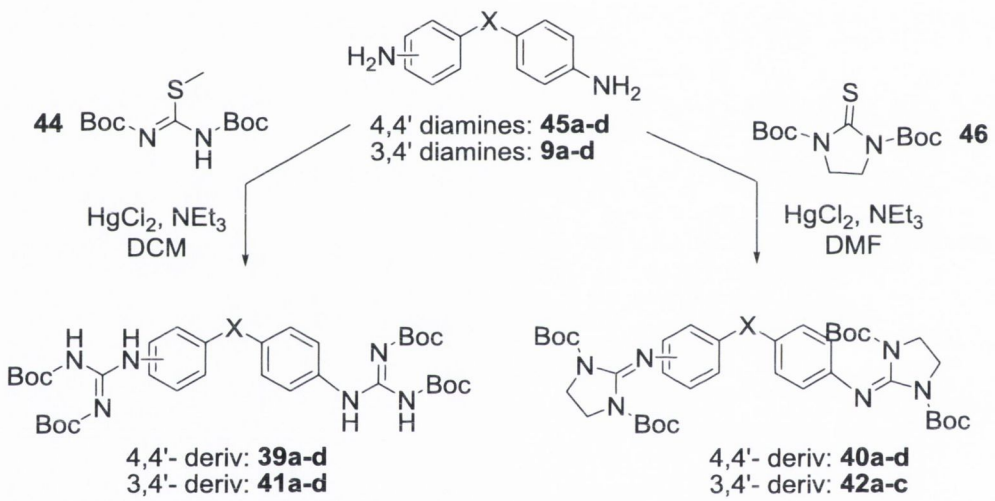
The synthetic approach to generate the 4,4'-*bis*-guanidinium-like derivatives has been previously described in a number of publications by Rozas and co-workers.[50,59,5] The same approach was used for the synthesis of the 3,4'-*bis*-guanidinium-like derivatives in this study. Taking into account that the preparation of these compounds involves the introduction of two guanidine or 2-aminoimidazoline moieties, the first reaction consisted of the preparation of such "guanidylating" agents. As mentioned in the previous chapter, in the guanidylation processes, either *N,N'-bis*-(*tert*-butoxycarbonyl)thiourea (**43**) or the commercially available *N,N'-bis*-(*tert*-butoxycarbonyl)-*S*-methylisothiourea (**44**) were used with no detrimental effect to the guanidylation, and thus, the commercial source was employed in our laboratory where possible. Similarly, *N,N'-bis*-(*tert*-butoxycarbonyl)imidazoline-2-thione (**46**) was used in the 2-iminoimidazolidylation reactions. Compounds **43** and **46**, which are not commercially available, were synthesised by treatment of thiourea or imidazolidine-2-thione with di-*tert*-butyldicarbonate (Boc_2O) and NaH in dry THF to form the Boc-activated agents.[50] Following work up and purification, a yield of 80% (**43**) and 84% (**46**) was achieved (Scheme 3.20).

Scheme 3.20.



The 2-aminoimidazolinium cation is structurally (bond distances, angles and atomic charges) and electronically (molecular electrostatic potential, electron-density) very similar to the guanidinium group[60] and, hence, many of the methods used to introduce the guanidinium moiety in a molecule have also been exploited in the introduction of the 2-aminoimidazolinium group. Using the Kim and Qian[49] guanidylation method and the Dardonville and Rozas 2-iminoimidazolidylation procedure previously described in section 3.3.2, the 4,4'- and 3,4'-*bis*-guanidinium and -2-aminoimidazolinium derivatives were prepared by coupling **44** or **46** with the appropriate commercial (**45a-d**) or previously synthesised (**9a-d**) diamines (Scheme 3.21) .

Scheme 3.21.



The coupling of the diamines with the guanidylating agents is promoted by mercury(II) chloride. The stoichiometry used for the reaction was 1:2 diamine:guanidylating agent due to the symmetric nature of these compounds. Table 3.5 describes the yields obtained for the guanidylation of the different 4,4'- and 3,4'-amines to yield the corresponding Boc-protected guanidines or 2-aminoimidazolines.

Table 3.5. Yields of the 4,4'- and 3,4'-Boc-protected guanidines (**39a-d** and **41a-d**) and 2-aminoimidazolines (**40a-d** and **42a-c**).

Derivative	X linker	Yield	Compound
4,4'-guanidines	O	79%	**39a**
	CO	52%	**39b**
	CH_2	80%	**39c**
	NH	66%	**39d**
4,4'-2-aminoimidazolines	O	72%	**40a**
	CO	40%	**40b**
	CH_2	78%	**40c**
	NH	61%	**40d**
3,4'- guanidines	O	87%	**41a**
	CO	60%	**41b**
	CH_2	86%	**41c**
	NH	83%	**41d**
3,4'-2-aminoimidazolines	O	68%	**42a**
	CO	53%	**42b**
	CH_2	88%	**42c**

It is worth mentioning that, in all cases, the CO-linked diphenyl derivatives showed the lowest yield. This is due to the electron-withdrawing effect of the carbonyl group, which reduces the electron-density of the amines in the *para* position. Consequently, the amine is less nucleophilic and less prone to attack the electrophilic carbon of the carbodiimide intermediate. A slightly higher yield was obtained for the guanidylation of the CO-linked 3,4'-diamine **9b** than for that of the 4,4'diamine **45b**, which might be due to the fact that one of the amines is in the *meta* position not in conjugation with the CO group, so no deactivation occurs.

When comparing the coupling reaction for the rest of the diamines, a coherent pattern to justify the yields obtained cannot be observed. In terms of effects to the electron density of

the amines, the rest of the linkers exert either an electron-donating effect due to free loan pairs of electrons in conjugation with the π-system (such as O and NH), or at least no deactivating effect (CH_2). In terms of inductive effects, the CH_2 will be more favourable due to its inductive donation of electron density, while O and NH inductively withdraw electron density from the ring, exerting deactivating effects on the amine in the *meta* position in particular. For these reasons, balancing the conjugative and inductive effects, the nucleophilicity of the amines in CH_2-linked diamine will be similar to the more electron-donating, but more electronegative O and NH linked diamines. As a result, no major difference in the yield of the guanidylation of the aforementioned amines was observed.

However, a decrease in the yield in the 2-iminoimidazolidylation reaction is observed in comparison with the guanidylation of the amines due to the purification process. Previously in our group, *bis*-Boc-protected phenyl-2-aminoimidazolines have degraded when purify by chromatography on silica and, hence, high speed alumina chromatography was necessary to purify them, with poor separation being observed and some product loss still incurring. This instability is likely to stem from two sources: (a) the increased basicity of the imine lone pair in *bis*-Boc-protected 2-aminoimidazoline derivatives compared to their guanidine analogues, and (b) the ring strain present in the 2-aminoimidazoline compounds due to the sp^2 hybridised nitrogens and the sp^3 hybridised carbons in the ethylene bridge, making a rearrangement of the hydrogens necessary to avoid clashing. For this reason, this ring is very sensible to weak acids or even unreacted starting amine because the ring strain is relieved by nucleophilic attack and ring opening. Therefore, protonation of the nitrogen of the imine can lead to its hydrolysis, resulting in the starting aryl amine being regenerated.

Additionally, it could be predicted that, in general, the 3,4'-diphenyl diamines would show lower yields than the 4,4'-diphenyl diamines since σ-electron-withdrawing linkers render the amine in the *meta*-position to be more deactivated compared to the amine in the *para*-position of 4,4'-diphenyl diamines, but that problem is overcome when 2 equivalents of the guanidylating agent are added and the reaction time is long enough.

All the 3,4'-*bis*-Boc-protected-guanidine-like compounds (**41a-d** and **42a-c**, Scheme 3.21) were fully characterised by ^1H and ^{13}C NMR spectroscopy, melting point, IR and mass spectroscopy. For example, characteristic peaks in the ^1H NMR spectrum of compound

41d (X = NH, Figure 3.10) such as the 4 singlet $(CH_3)_3$ peaks from the Boc groups at 1.52, 1.53, 1.56 and 1.57 ppm showed that formation of the Boc-protected derivative was achieved. Two peaks are also observed at 10.2 and 11.6 ppm indicative of the NH protons of the guanidine group, which form HBs with the carbonyl group of Boc (Figure 3.10), shifting them downfield. The reason for the different shift of the NHs is that the NH attached to the ring is more acidic because the negative charge that would result after deprotonation would be in resonance with the phenyl ring π-system. This NH is further downfield, appearing at 11.6 ppm in the ^{1}H NMR spectrum (Figure 3.10), whereas the other NH appears at lower ppm (around 10.2 ppm).[61]

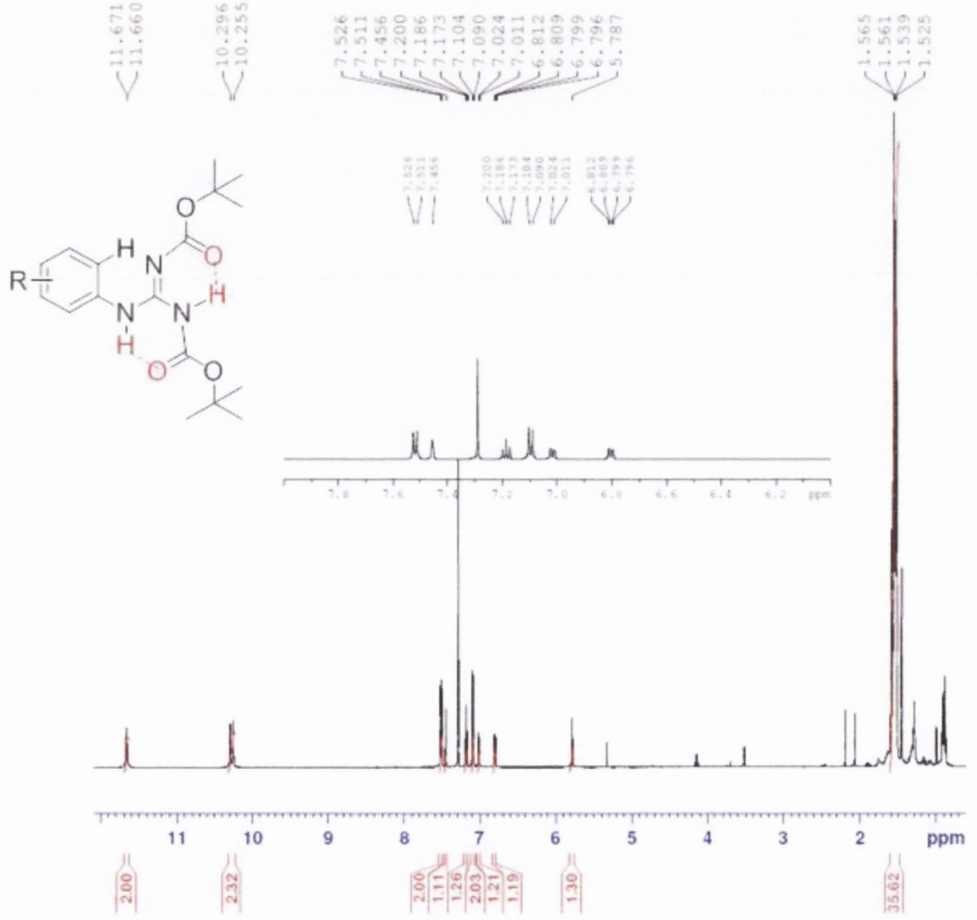
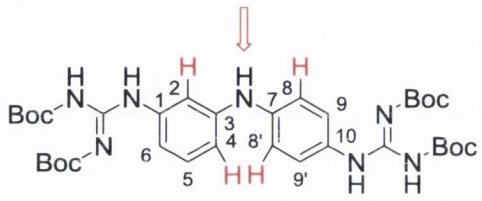


Figure 3.10. ^{1}H-NMR ($CDCl_3$) spectrum of **41d** and general structure showing the HB formation between the NHs and the carbonyls of the Boc groups.

Also, to overcome the difficulties in identifying the protons and carbons at the *meta* and *para* substituted rings, one dimension total correlation spectroscopy (TOCSY) experiments were carried out. In this experiment, the proton signal in the ^{1}H NMR spectrum from the NH linker of **41d** is irradiated continuously, revealing the protons which exist in the same spin system. The signals obtained in the TOCSY spectrum helped us to confirm the precise molecular conformation of **41d** (Figure 3.11), observing that only the doublet at 6.80 ppm, the singlet at 7.45 ppm and the doublet at 7.51 ppm are related to the NH linker (H-4, H-2, H-8 and H-8', respectively).

(a)



(b)

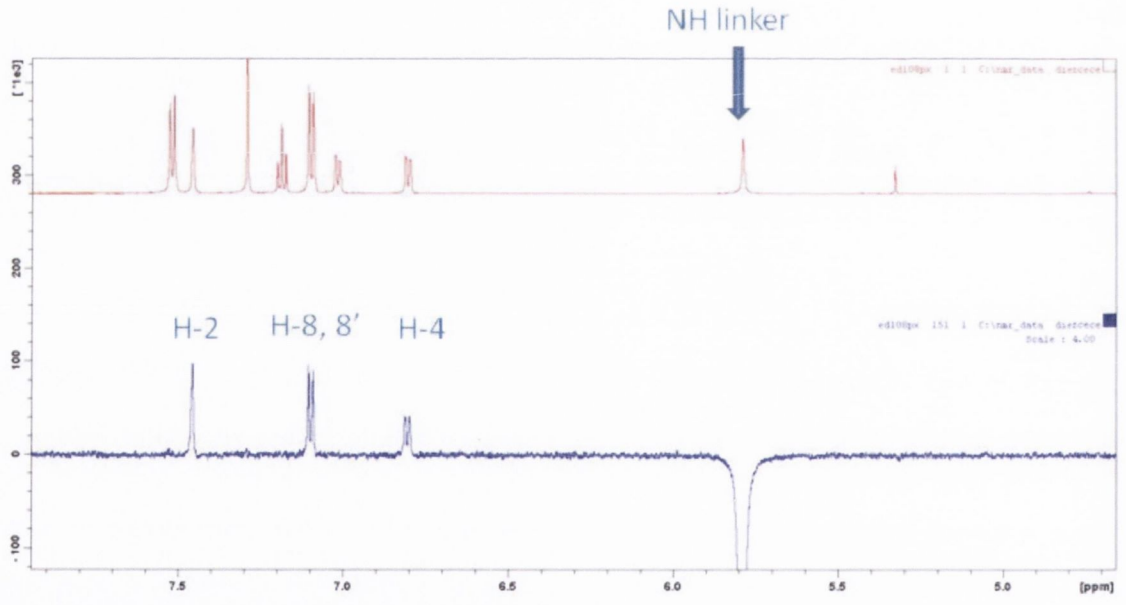


Figure 3.11. (a) Structure of compound **41d** with the proton of the NH linker indicated by an arrow; this proton was irradiated in the TOCSY spectroscopy experiment. (b) ^{1}H NMR ($CDCl_3$) spectrum of **41d** (*top* in red) and TOCSY spectrum (*bottom* in blue), where only the spin system of the NH proton were revealed.

Also, two dimensional NMR spectroscopy methods – heteronuclear single-quantum correlation spectroscopy (HSQC) and heteronuclear multiple-bond correlation spectroscopy (HMBC) – were used to completely characterise the structure of the molecule. HSQC generally detects correlations between nuclei of two different types which are separated by one bond, whereas HMBC detects heteronuclear correlations over longer ranges of about 2-4 bonds. In Figure 3.12 HSQC in red and HMBC in blue are superposed, allowing the identification of the aromatic carbons and protons related to the NH linker. Focusing on the singlet of the NH at 5.78 ppm (arrow), we can see that it is related to three peaks. Looking at the X axis (1H NMR), the signals related to the NH are the doublet at 6.80 ppm, the singlet at 7.45 ppm and the doublet at 7.51 ppm. From their integration we can notice that the doublet at 6.80 ppm integrates for 1H, therefore it must be H-4, then the singlet at 7.45 ppm integrates for 1H and corresponds to H-2, while the other doublet at 7.51 ppm integrates for 2H, meaning it must be H-8 and H-8', which are chemically equivalent. Looking at the Y axis (^{13}C NMR spectrum), the carbon peaks of those protons can be identified: 110.7 ppm is C-2, 113.1 ppm is C-4 and 119.0 ppm is C-8 and 8'.

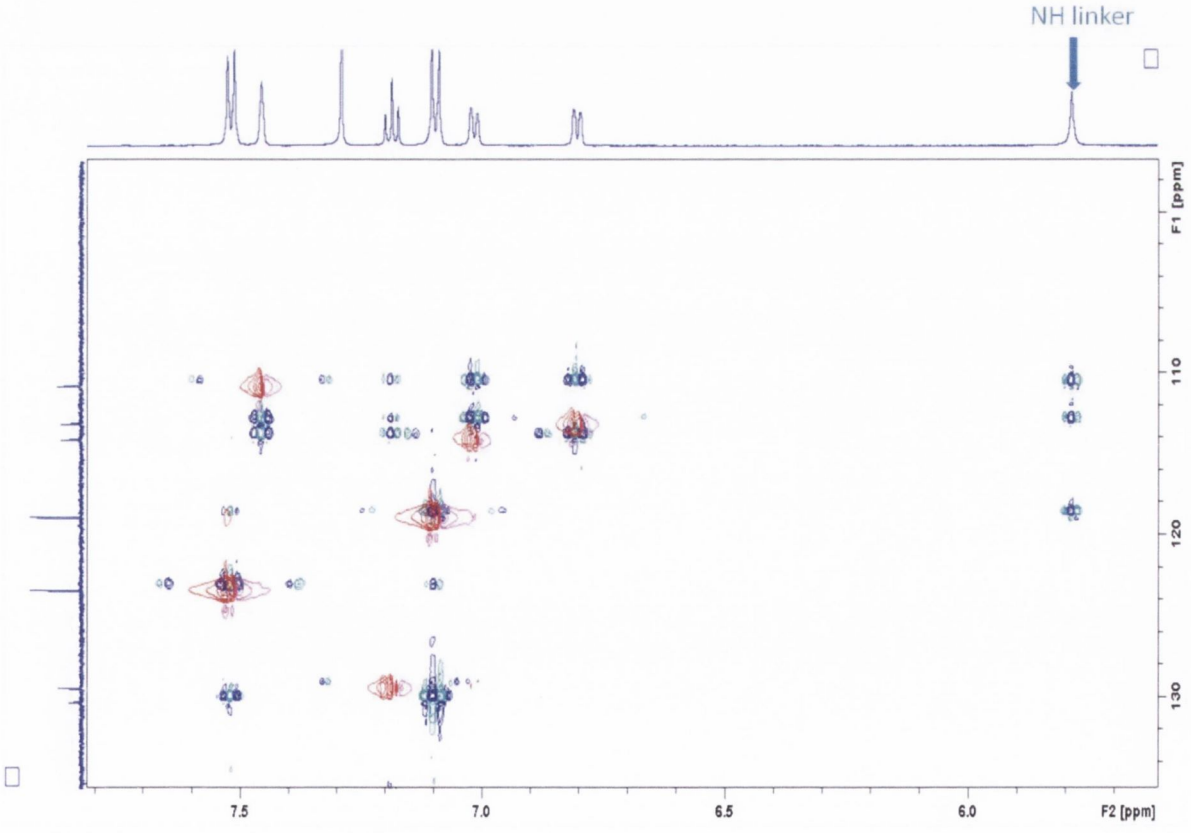


Figure 3.12. HSQC (red) and HMBC (blue) of compound **41d** in $CDCl_3$.

In this manner, H-4, 6, H-8, 8', and H-9, 9' were all identified. The rest of the signals are easily identifiable because H-5 has a characteristic multiplicity of a triplet, meaning the remaining signals must be H-6 and H-9, 9'. This full identification of signals will be useful when different groups are placed on the *meta* and *para* amines, to prove the position of the substitution.

The presence of the CN and CO can be identified in the ^{13}C NMR spectra with two CN peaks at 153.3 and 153.4 and four CO peaks, confirming the successful synthesis of the di-Boc protected compounds (Figure 3.13). The two CO, which are in conjugation with the CN, appear further downfield at 163.4 and 163.5 ppm and the other two CO appear at 153.1 and 153.2 ppm.

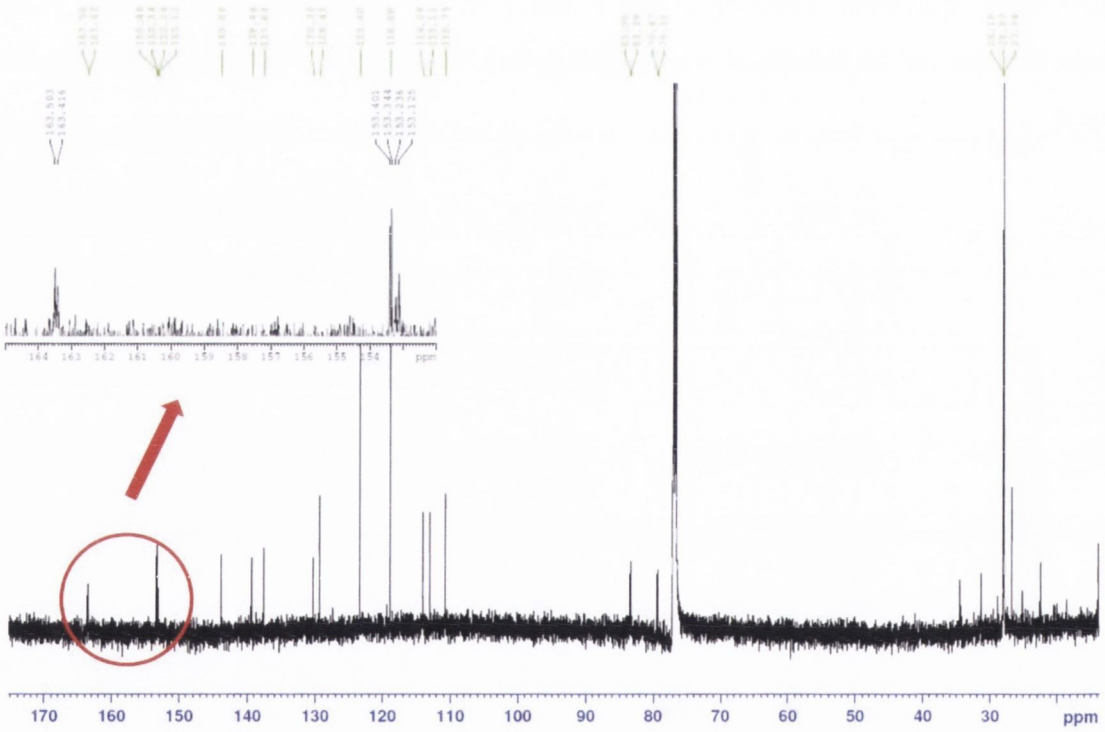


Figure 3.13. ^{13}C-NMR ($CDCl_3$) of compound **41d**, where two CN and four CO peaks are observed.

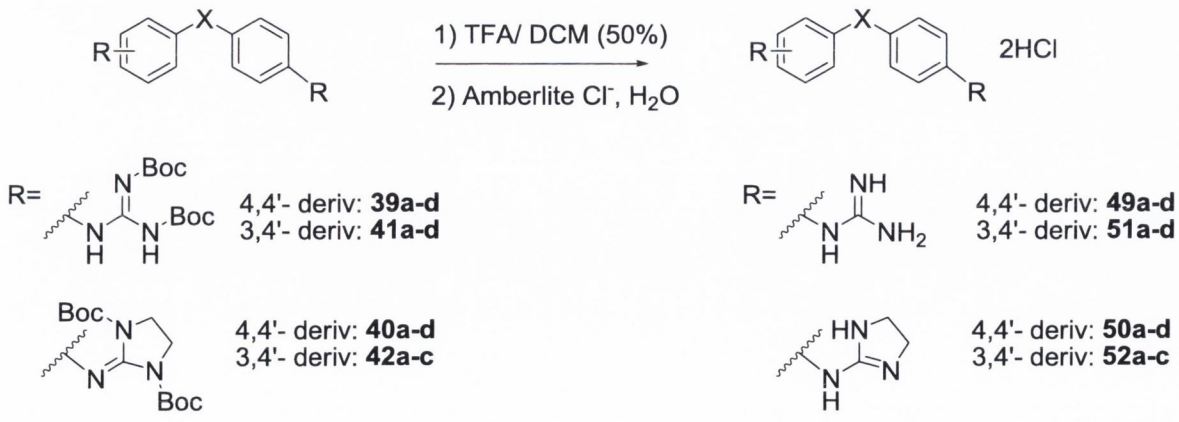
A similar spectral analysis was performed for each compound, indicating the successful preparation of this family of derivatives. In the case of the CH_2 and CO linked compounds,

the HMBC spectra were crucial for the complete identification of the molecule, because these experiment tell us which protons are related to the CO and CH_2 linker within 2-4 chemical bonds, helping us to distinguish between H-4 and H-6, and between H-8, 8' and H-9, 9'. However, the full characterisation of the 3,4'-guanidine-like derivatives of the O linker was not as trivial because in this case the heteroatomic linker precluded identification by HMBC. In this case the conjugative and inductive effects of the linker were considered.

3.3.5. Synthesis of the Hydrochloride Salts of the 4,4'- and 3,4'-*bis*-Guanidine-like Derivatives

The removal of the Boc groups of 4,4'- (**39a-d** and **40a-d**) and 3,4'- (**41a-d** and **42 a-c**) guanidine-like derivatives, was performed using a 50% (*v/v*) solution of trifluoroacetic acid (TFA) in DCM (Scheme 3.22). This yielded the corresponding trifluoroacetate salts, which are generally not water-soluble and, thus, ion exchange was necessary to obtain the chloride salt, a more suitable compound for biological testing. This was completed by stirring the salt in water and Amberlite® IRA-400 resin in its chloride activated form. After 24 h, removal of the resin by filtration yielded the guanidinium chloride salt. This was checked by using ^{19}F NMR spectroscopy, the absence of any peaks in this spectrum confirming full conversion.

Scheme 3.22.



Then, the crude hydrochloride salt was purified using small scale reverse-phase chromatography with C-8 silica affording the pure hydrochloride salt in excellent yields (Table 3.6). Using 100% water as mobile phase was normally sufficient to elute the hydrochloride salt; but where necessary, solutions of water and acetonitrile (85:15) were used to elute less polar derivatives.

Table 3.6. Yields of hydrochloride salts after Boc-deprotection and ion exchange.

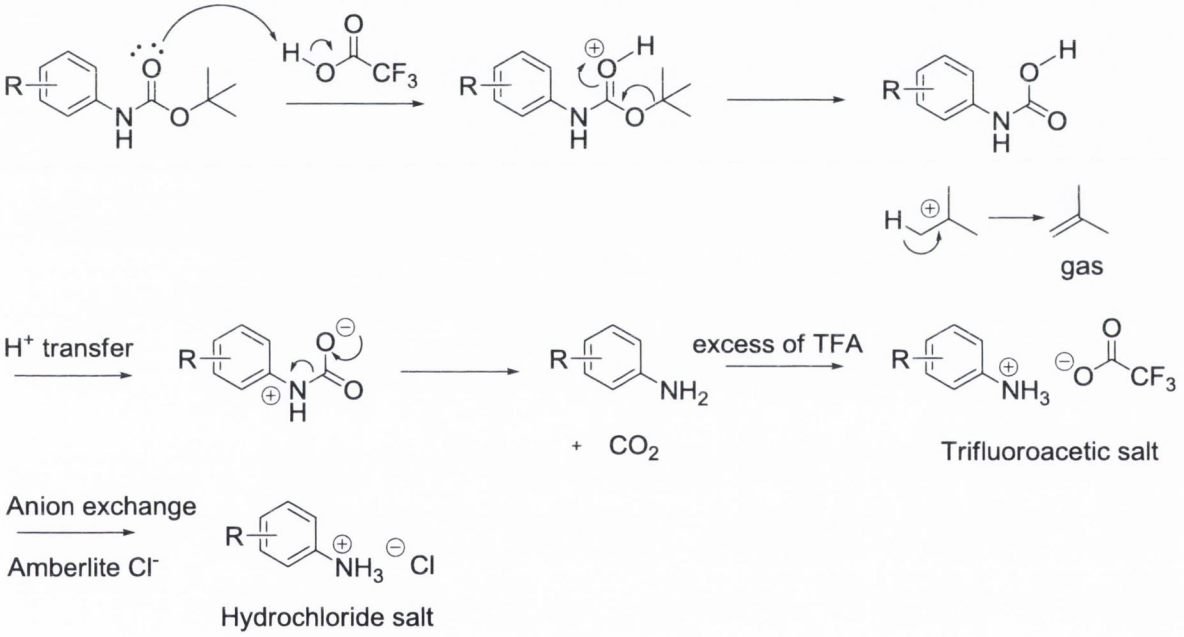
Boc derivative		**X linker**	**Yield**	**Salt**
4,4'-guanidines	**39a**	O	91%	**49a**
	39b	CO	82%	**49b**
	39c	CH_2	76%	**49c**
	39d	NH	82%	**49d**
4,4'-2-aminoimidazolines	**40a**	O	90%	**50a**
	40b	CO	78%	**50b**
	40c	CH_2	78%	**50c**
	40d	NH	68%	**50d**
3,4'- guanidines	**41a**	O	76%	**51a**
	41b	CO	53%	**51b**
	41c	CH_2	70%	**51c**
	41d	NH	62%	**51d**
3,4'-2-aminoimidazolines	**42a**	O	76%	**52a**
	42b	CO	55%	**52b**
	42c	CH_2	70%	**52c**

No significant differences were encountered in the efficacy of the Boc-deprotection, This reaction works well in most cases, due to the fact that the nature of the linker should not interfere with the progression of the reaction. It is worth noting that the purification of the salts by reverse-phase chromatography was sometimes troublesome, as their small scale is

not always enough to separate the salt from other impurities; hence, the yield of the reaction sometimes was affected.

The deprotection mechanism of this reaction (Scheme 3.23) involves protonation of the carbonyl group of the Boc group by TFA; then, electrons migrate so that the positively charged carbonyl is quenched, making the *tert*-Butyl (^{t}Bu) group leave in a S_N1 manner. This is followed by a proton transfer and the movement of the lone pair of electrons from the O to reform the carbonyl bond and CO_2 is expelled to form the deprotected amine. An excess of TFA generates the trifluoroacetate, which will be exchanged for hydrochloride salt in contact with Amberlite in its activated Cl^- form in a second step.

Scheme 3.23.



The purity of this library of hydrochloride salts (Families 1 and 2) was assessed using reverse-phase HPLC chromatography. For more information on this procedure see experimental section 7.1.2.

3.3.6. Conformational Aspects of *bis*-Boc-Protected and Hydrochloride Salts of Aryl Guanidine Derivatives.

In a previous work in Rozas' group[61] it was found that an intra-molecular hydrogen bond (IMHB) is formed between the imine nitrogen (N3') of a Boc-protected guanidine and an aromatic hydrogen, rendering these molecules planar (Figure 3.14, top, in blue). Additionally, in the *bis*-Boc-protected guanidines, two further IMHBs exist between the carbonyl groups of the Bocs and the NHs from the guanidine (Figure 3.14, top, in red). These conformational aspects have been proven by the resolution of several crystal structures using X-ray crystallography (Figure 3.14, bottom), as well as solution phase NMR spectroscopy and Density Functional Theory (DFT) theoretical calculations. This co-planarity stands in contrast with the final hydrochloride salts in which the guanidine group typically lies out-of-plane with the aromatic ring to avoid repulsion between aromatic protons and those of the protonated guanidine.

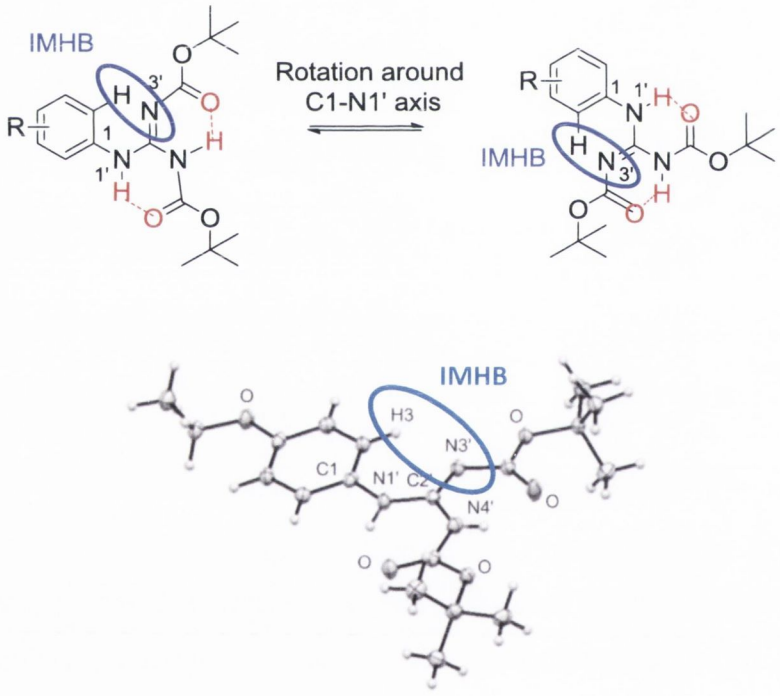


Figure 3.14. IMHB formation in di-Boc-protected guanidines (*top*), crystal structure of 1-[2,3-di-(*tert*-butoxycarbonyl)guanidino]-4-ethoxybenzene,[61] which shows the IMHB that makes the aryl and the guanidine moiety co-planar (*bottom*).

These IMHB are also observed in the 3,4'-diphenyl derivatives here prepared and a good example is 3,4'-*bis*-[2,3-di-(*tert*-butoxycarbonyl)guanidino]benzylbenzene (**41c**). The influence of the IMHB is clearly visible in its ^{1}H NMR spectrum; as the hydrogens forming the IMHB appear at higher ppm (Figure 3.15, top). In this compound, H-2, H-6, H-10 and H-10' are further downfield than the rest of the aromatic protons, their shifts being 7.32 (H-2), 7.53 (H-10 and 10') and 7.58 (H-6) (Figure 3.15, bottom). This is a result of the formation of IMHBs between the imine nitrogen of the guanidine and the aforementioned aromatic protons. Due to the free rotation around the C-N axis, this effect is felt by H-2 and H-6, and H-10 and H-10' (Figure 3.15, top).

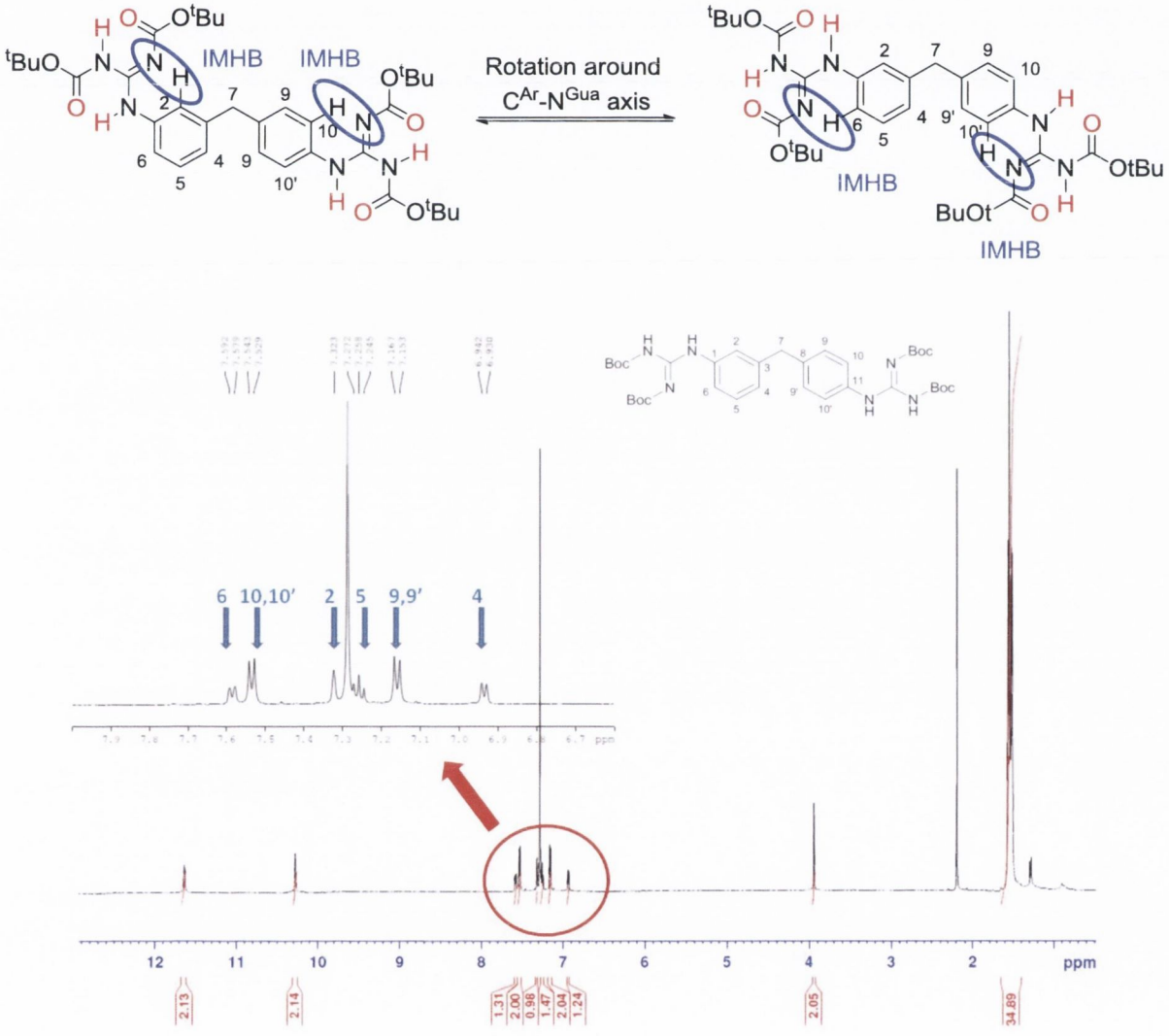


Figure 3.15. Effect of the rotation of the C^{Ar}-N^{Gua} in the aromatic hydrogens (*top*) and ^{1}H NMR spectrum of *bis*-Boc-protected guanidine **41c** (*bottom*) in $CDCl_3$.

The opposite trend is observed once the Boc groups are removed and the guanidinium chloride (**51c**) is formed. In this case H-2, H-6, H-10 and H-10' are more upfield than the rest of the aromatic protons, their shifts being 7.14 (H-6), 7.17 (H-2) and 7.22 (H-10 and H-10') (Figure 3.16). This is as a result of adopting a conformation which avoids repulsion between aromatic protons H-2, H-6, H-10 and H-10' and those of the now protonated guanidinium.

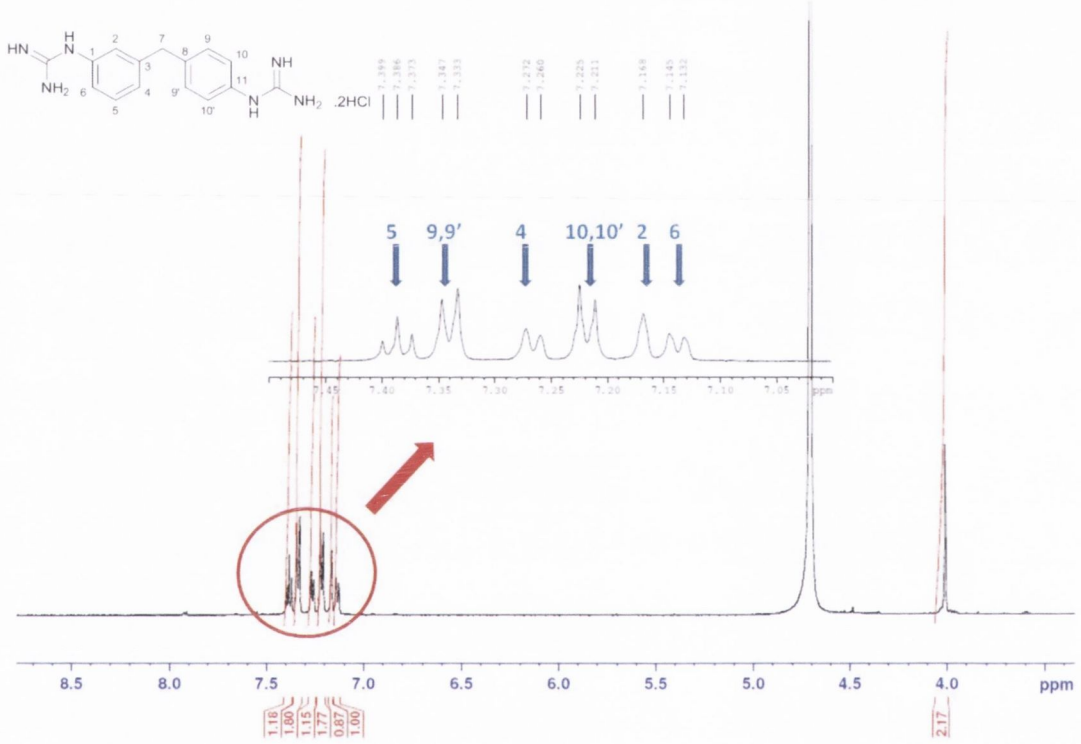


Figure 3.16. ^{1}H NMR spectrum of the hydrochloride salt **51c** in D_2O.

However, the influence of the IMHB is not present in the 2-iminoimidazoline derivatives because the imine nitrogen cannot form an IMHB with any aromatic hydrogens. Similarly, the influence of the IMHB in the rest of the guanidine derivatives with different linkers is not as clear because they are exposed to other effects such as the inductive or conjugative effects of the linker. For example, in those derivatives with a CO linker, the aromatic hydrogens *meta* to the CO linker appear more upfield, because they do not suffer the electron-withdrawing effects of the carbonyl. Then, in those derivatives with O and NH linkers, the electron-donating effects and the electonegativity of the linkers affect the shifts

of the protons; hence a systematic rule for signals in the ^{1}H NMR spectra cannot be established.

3.4. Synthesis of Phenyl Guanidine Derivatives – Families 3A, 3B, 4A and 4B

3.4.1. General Aspects of Families 3A, 3B, 4A and 4B

As previously mentioned in the objectives of this project (Chapter 2), the synthesis of different families of guanidine-like derivatives with specific structural moieties, related to those found in sorafenib, was pursued. In Families 3 and 4, a phenyl derivative – substituted or not – was introduced to the guanidine group which is *para* to the linker (Family 3A and 4A), as this is the orientation found in sorafenib. In addition, these phenyl groups were also introduced in the guanidine in *meta* (Family 3B and 4B, Figure 3.17) for the sake of comparison. These structural modifications will provide information about structure activity relationship.

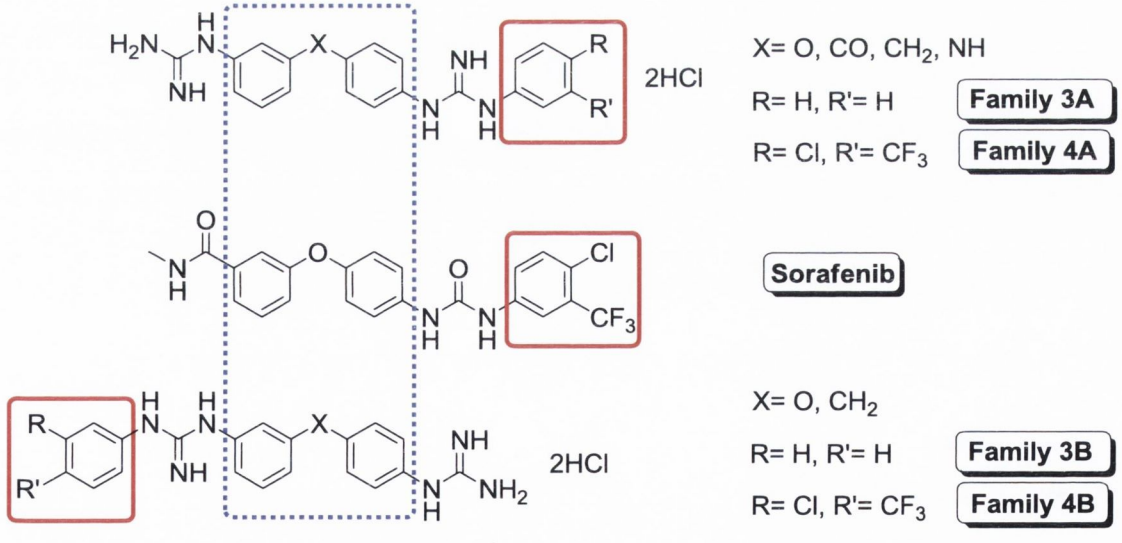


Figure 3.17. Structures of derivatives from Families 3A, 4A, 3B, 4B, and sorafenib.

The retrosynthesis of these compounds was considered (Figure 3.18) keeping in mind that their core was the 3,4'-diphenyl diamines (**9a-d**) previously described in section 3.2. Using the guanidylation method described in section 3.3.2, Families 3A and 4A would be synthesised by coupling the amine in the *meta* position relative to the linker with **43** and the amine in the *para* position with *N*-(*tert*-butoxycarbonyl)-*N'*-arylthiourea derivatives (**57** and **58**). The opposite synthetic route would be followed for Families 3B and 4B.

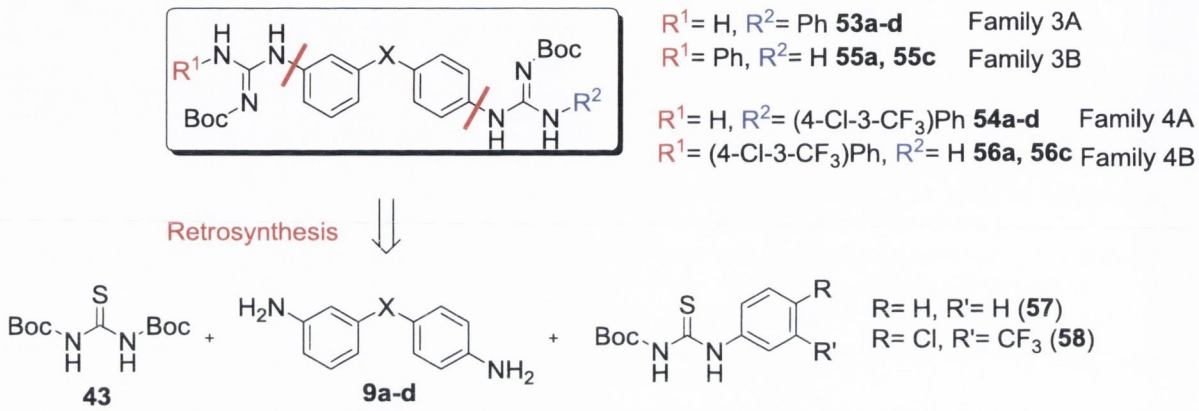
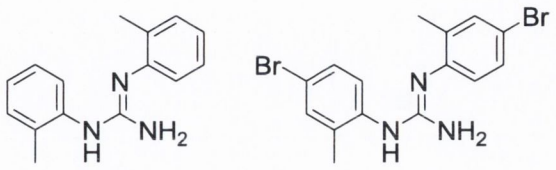


Figure 3.18. Retrosynthetic plan for the preparation of Families 3 and 4.

Obviously, before proceeding with the synthesis of these derivatives the corresponding *N, N'*-disubstituted thioureas **57** and **58**, needed to be prepared. However, there are many methods for synthesising *N,N'*-substituted guanidines due to their important biological applications in medicinal chemistry over the years and a brief discussion is included in the following Section 3.4.2.

3.4.2. Literature Methods for the Preparation of *N,N'*-Disubstituted Guanidines

The *N,N'*-disubstituted guanidines such as 1,3-di-*o*-tolylguanidine and *N,N'*-di-*p*-bromophenylguanidine (**59** and **60** in Figure 3.19) have importance as potential therapeutics for their neuroprotective and antidepressant effect as sigma receptor antagonists.[62]

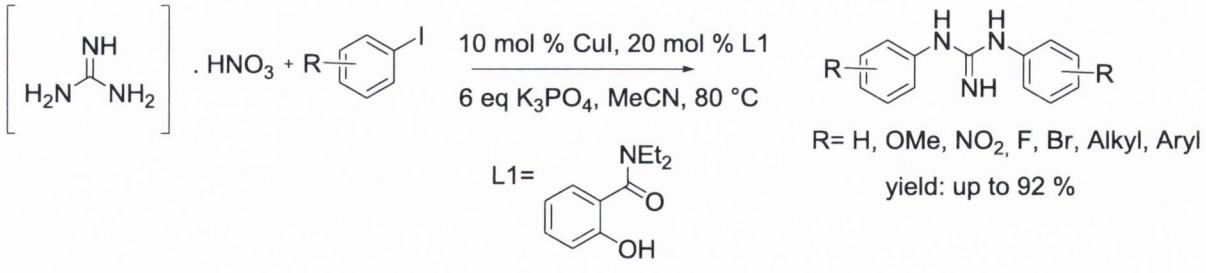


59 **60**

Figure 3.19. Chemical structures of *N,N'*-disubstituted guanidines with biological applications; (a) 1,3-di-*o*-tolylguanidine, **59** and (b) *N,N'*-di-*p*-bromophenylguanidine, **60**.

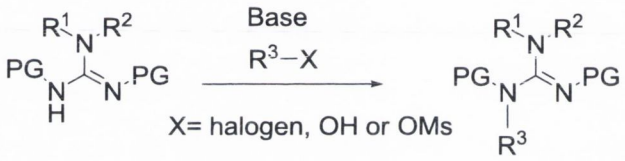
The synthesis of these *N,N'*-disubstituted guanidines involves the use of a copper-catalyzed cross-coupling reaction of guanidine nitrate with aryl iodides. A relatively inexpensive, commercially available, guanidine salt and a series of aryl iodides together with copper iodide and *N,N*-diethylsalicylamide as an efficient catalyst/ligand system provided a simple diarylation procedure (Scheme 3.24).[63] However, this method is only useful for the synthesis of symmetrically substituted *bis*-aryl derivatives and has not been extended to *N*-alkyl substituted guanidines.

Scheme 3.24.



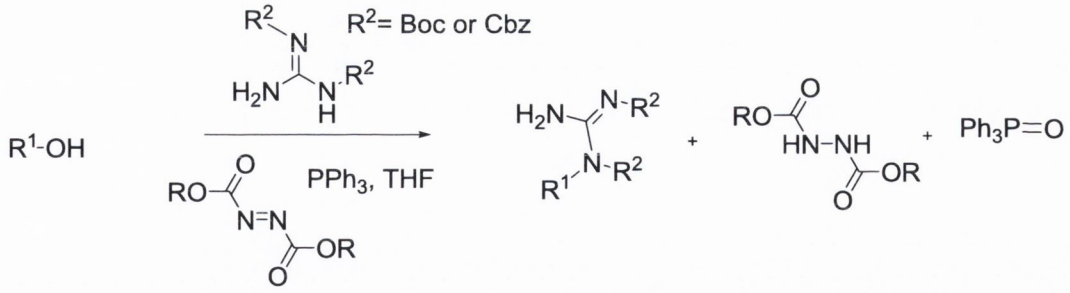
Other methods to obtain a polysubstituted guanidine by modification/funtionalisation of a pre-formed guanidine have been reported, for instance, through *N*-alkylation, *N*-acylation or transamination. In these methods, the guanidylation reaction involves the reaction of an electrophile with a nucleophilic guanidine with acidic N-H protons, to give a more highly substituted guanidine, as this moiety -NH(C=NH)-NH_2 is introduced onto an alkyl halide/alcohol (Scheme 3.25). One example of these methods is the deprotonation of guanidine with NaH followed by reaction with an alkyl halide (*e.g.* methyl iodide and benzyl bromide) to give the functionalised guanidine.[64]

Scheme 3.25.



Another approach to obtain polisubstited guanidines involves the reaction of guanidine with a primary or secondary alcohol (benzylic or allylic) under Mitsunobu conditions (Scheme 3.26).[65] The mechanism of this reaction involves the formation of an azodicarboxylate-PPh_3 complex, followed by deprotonation of the acidic hydrogen from the protected guanidine. The alcohol, then, coordinates with the positively charged phosphorous, liberating the doubly protonated azodicarboxylate product. Simultaneously, the deprotonated guanidine attacks the alcohol-PPh_3 complex, introducing the new R^1 group into the guanidine.

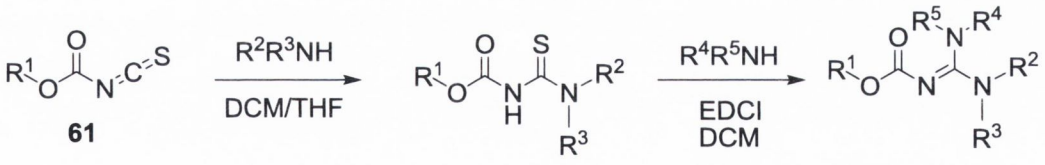
Scheme 3.26.



This method is more general and works well for a wide variety of alcohols; however, it requires expensive coupling reagents, which often complicate the purification procedure.

A different method was proposed by Hamilton *et al.* allowing for the possibility of synthesising multi-substituted guanidines.[66] The authors first prepare a carbamoyl isothiocyanate reagent (**61**) which is subjected to an appropriately substituted amine, followed by a second amine and 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDCI), in a one-pot procedure (Scheme 3.27). Successful preparation of several *N*-alkyl-*N'*-aryl and asymmetrical *N*-aryl^1-*N'*-aryl^2 disubstituted guanidines was reported, incorporating a wide variety of electron-donating and electron-withdrawing substituents.

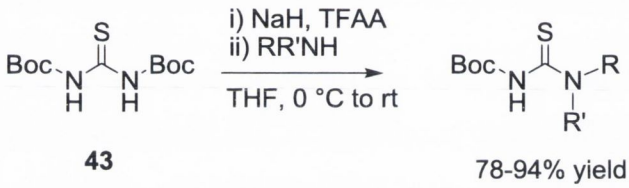
Scheme 3.27.



Although Hamilton's procedure is excellent, it has a few drawbacks such as the necessity to prepare in advance the key carbamoyl isothiocyanate reagent **61**, which requires a two-day reaction and the special preparation of dry KNCS. Also, the reaction failed to proceed when secondary amines or electron-poor aromatic amines were utilised in the second step of the reaction.

In a recent publication, Yin *et al.* described the preparation of *N*-Boc-*N'*-substituted thioureas by treatment of *N,N'-bis*-(*tert*-butoxycarbonyl)thiourea **43** with sodium hydride (NaH) and trifluoroacetic anhydride (TFAA) in the presence of an amine (Scheme 3.28).[67] The reaction proceeds with both aryl and alkyl amines in good to excellent yield. The thiourea derivatives are ideally functionalised for the synthesis of *N,N'*-disubstituted guanidines, as suggested by methods previously developed by Kim and Qian,[49] and by Hamilton.[66]

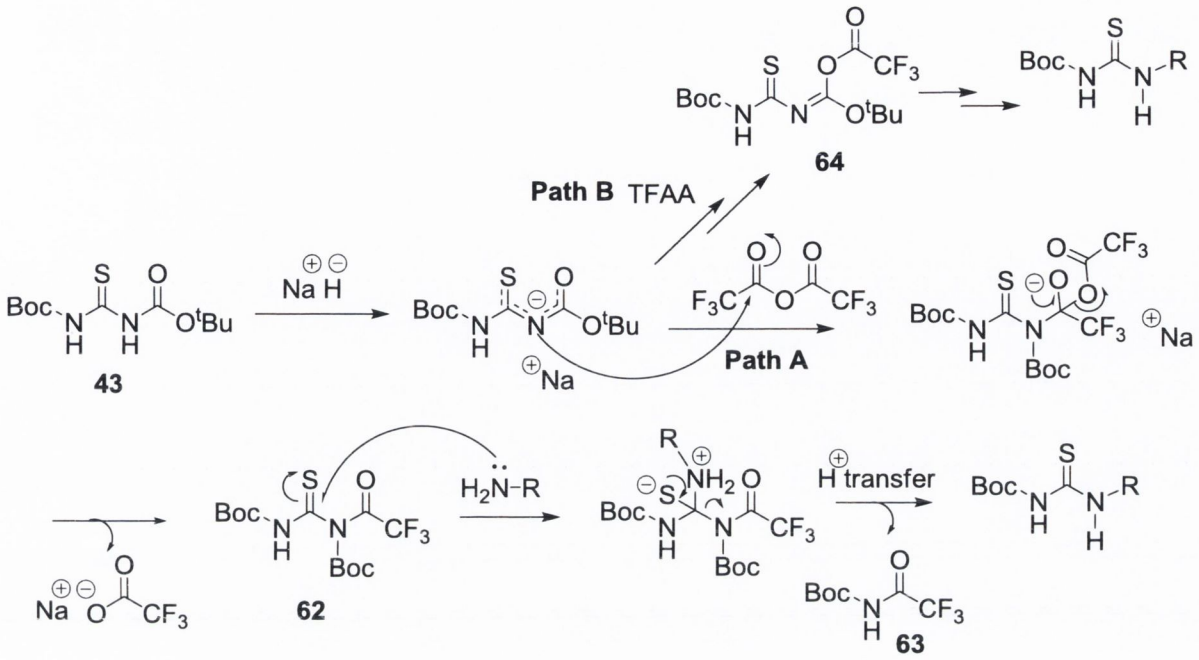
Scheme 3.28.



43

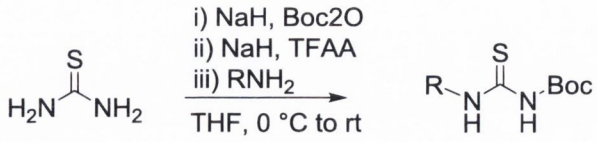
In a subsequent publication, Yin describes two possible mechanisms for this transformation.[68] In the first of these, the anion formed by deprotonation of **43** is *N*-acylated to produce the *N*-Boc-*N*-trifluoroacetyl derivative (**62**), which undergoes nucleophilic attack by the amine, with elimination of the *N*-Boc-*N*-trifluoroacetyl anion (**63**), during the second step (Scheme 3.29, path A). The alternative mechanism suggested by the authors involves acylation at the carbonyl oxygen of the Boc group to produce the *O*-trifluoroacetyl derivative (**64**) (Scheme 3.29, path B). Based on our experience and considering that intermediate **63** has been isolated from our reactions, the mechanism of Yin's reaction is likely to be the former, in which the reaction proceeds through the *N*-Boc-*N*-trifluoroacetyl derivative.

Scheme 3.29.



The possibility of generating *N,N'*-di-Boc-substituted thioureas *in situ* using inexpensive reagents such as thiourea, Boc_2O and NaH following Yin's procedure would permit the preparation of the desired *N*-Boc-*N'*-substituted thioureas from simple starting materials. Furthermore, in a one-pot procedure described by our group (Scheme 3.30), access to these compounds would be relatively quick.[69] Yields in this reaction are dependent on the nucleophilicity of the amine reacted. Considering the simplicity and low cost of the substrates involved in the last method, it was selected for the synthesis of our *N*-Boc-*N'*-substituted thioureas.

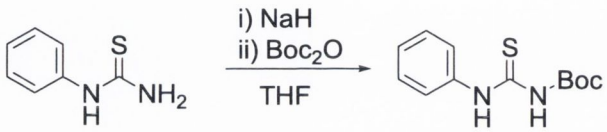
Scheme 3.30.



3.4.3. Synthesis of *N*-(*tert*-butoxycarbonyl)-*N'*-phenylthiourea (57) and *N*-(*tert*-butoxycarbonyl)-*N'*-(4-chloro-3-trifluoromethylphenyl) thiourea (58)

In the particular case of **57**, since phenylthiourea was commercially available, it was thought that its Boc-protection would be a better synthetic route than the attack of the poor nucleophilic aniline to the *N*-Boc-*N*-trifluoroacetyl derivative (Scheme 3.31).

Scheme 3.31.

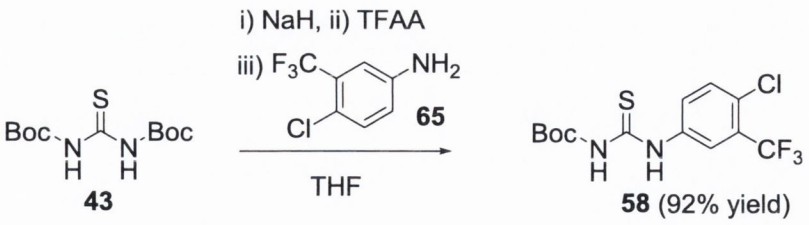


57 (58% yield)

Firstly, only one equivalent of NaH was used but this resulted in a 25% yield of the desired product. Possibly, NaH was deprotonating the secondary NH, instead of the primary one, since it is more acidic. For this reason, two equivalents of NaH were used, because, if the two NHs were deprotonated, the more reactive anion – that on the unsubstituted nitrogen – would react with one equivalent of Boc_2O. Thus, after the addition of two equivalents of NaH, the reaction was stirred at 0 °C for 1 h and then at room temperature for 3 h, to complete formation of the anion. The reaction was cooled again to 0 °C prior to addition of Boc_2O and stirred for 1 h at 0 °C, after which it was brought to room temperature and stirred overnight. The desired product was obtained in 58% yield.

In the case of compound **58**, the Yin method was used. Firstly, the reaction was attempted in one pot but the yield of the reaction was very poor. Isolating the *N,N'-bis*-(*tert*-butoxycarbonyl)thiourea **43** and then making it react with NaH and TFAA in THF, followed by addition of the 4-chloro-3-(trifluoromethyl)aniline (**65**), led to **58** in better yields (Scheme 3.32).

Scheme 3.32.



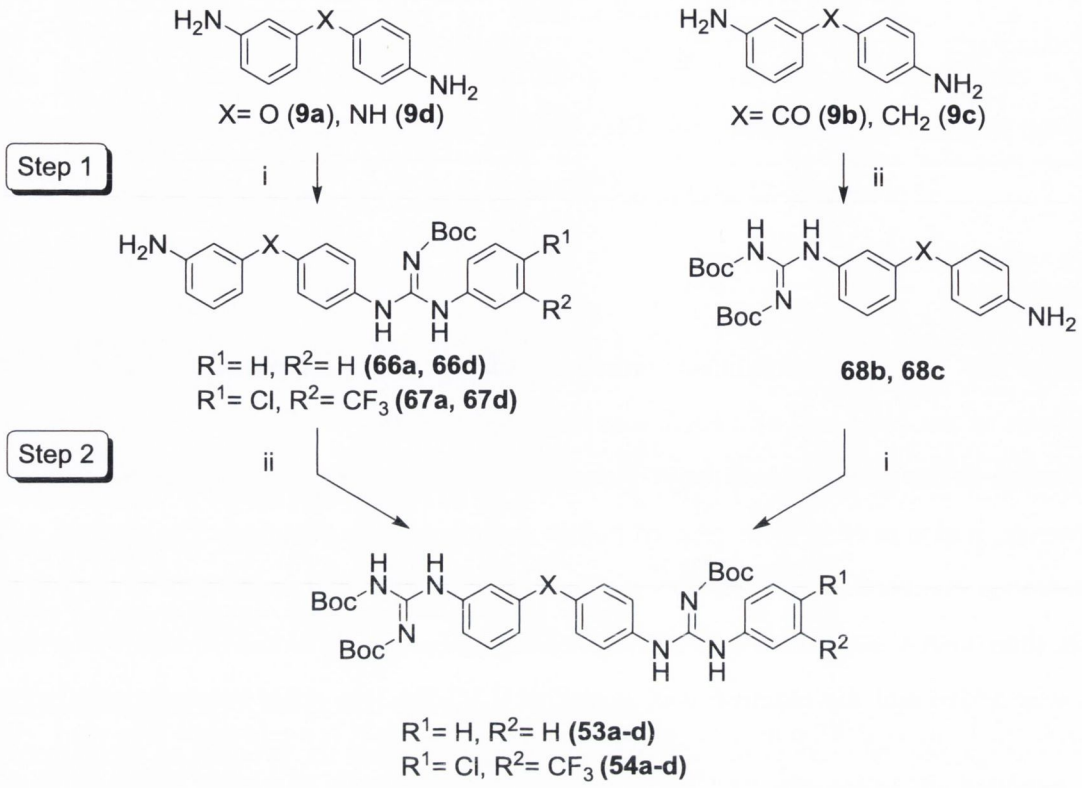
Several sets of reaction conditions were tested to optimise this process. The time after the addition of the NaH and of TFAA was modified in order to allow complete formation of the anion in the first step and the *N*-Boc-*N*-trifluoroacetyl intermediate in the second step; however, it also needed to be reacted before decomposition occurred. The optimal reaction conditions involved reaction of **43** with NaH in dry THF under argon at 0 °C for 1 h and 20 min, then TFAA was added and stirring continued for an additional 30 min. After this, the **65** was added and the reaction was stirred at 0 °C for 5 h. After workup, purification by chromatography and recrystallisation from hexane afforded the product as white needles in a yield of 92%. This was surprisingly high due to the low nucleophilicity of **65**.

3.4.4. Synthesis of Asymmetrically Substituted *para*-Phenyl Guanidine Derivatives (Family 3A and 4A)

Once the *N*-(*tert*-butoxycarbonyl)-*N'*-phenylthiourea derivatives **57** and **58**, the *N,N'-bis*-(*tert*-butoxycarbonyl) substituted thiourea **43,** and the starting 3,4'-diphenyldiamines (**9a-d**) were synthesised, the asymmetric *para*-substituted phenyl guanidine derivatives, which constitute Families 3 and 4 could be prepared.

Depending on the nature of the linker of the diamine, different synthetic routes were followed. If the linker has electron-donating effects due to a lone-pair of electrons – such as O and NH – the amine in the *para*-position is more nucleophilic than the amine in the *meta*-position and reaction with electrophiles occurs here. For this reason, the first step in the synthetic route would be the coupling of the amine in the *para*-position with compounds **57** or **58** and the second step would be the coupling of the amine in the *meta*-position with the thiourea **43** (Scheme 3.33).

Scheme 3.33.



i) **57** or **58**, $HgCl_2$, NEt_3, DCM; ii) **43**, $HgCl_2$, NEt_3, DCM

However, for linkers with electron-withdrawing effects on the phenyl rings – such as CO – or no conjugative effects on the ring – such as CH_2 – the opposite synthetic route was followed to afford the Boc-protected *para*-substituted phenyl guanidine derivatives (**53a-d** and **54a-d**) in good to moderate yields (Table 3.7).

Table 3.7. Yields of the mono guanidylation of di-aryl diamines; step 1 in the synthesis of compounds in Families 3A and 4A.

Step 1: mono substituted	Comp. no.	Linker	% Yield
para- Ph, Boc guanidine	**66a**	O	76
	66d	NH	83
para- (4-Cl-3-CF_3)Ph, Boc guanidine	**67a**	O	68
	67d	NH	86
meta- Boc, Boc guanidine	**68b**	CO	86
	68c	CH_2	47

Looking at the yields in Table 3.7, it can be concluded that compounds with O and NH linkers show selectivity to guanidylate the amine in the *para*-position, as predicted, giving yields fairly high (68-86%). Similarly, the compound with the CO linker shows high selectivity but in this case for substitution at the *meta*-position resulting in a 86% yield.

Conversely, the CH_2 linker does not display selectivity for substitution of either the *meta*- or *para*-amine, and nearly 50% of each was obtained. The similarity of these two amines was justified by the results obtained in a theoretical proton affinity (PA) study on the 3,4'-diaminodiphenylmethane (**9c**) (see experimental section 7.3). The difference between the PA of the two amines is only 1.3 kJ mol^{-1} (PA= 956.9 kJ mol^{-1} –*meta*-amine- and 955.6 kJ mol^{-1} –*para*-amine-) making them virtually indistinguishable. The PAs of the amines were calculated as the difference in energy between the protonated ammonium and the neutral amine in each case. Therefore, adding an electrophile selectively to one of these amines is not possible, as both react similarly to give a mixture of substituted products. This is also evident in the Molecular Electrostatic Potential (MEP) map of the molecule (Figure 3.20) (see experimental section 7.3), where it can be seen that the electron-density on both amines is similar (red indicates higher electron-density, which is placed on the amino groups).

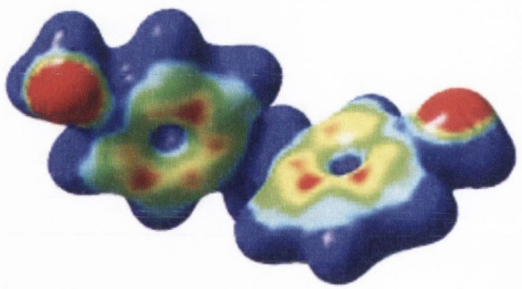


Figure 3.20. Molecular Electrostatic Potential (MEP) map of **9c**, plotted on an isosurface of 0.02 a.u., calculated at the B3LYP/6-31G* level of computation.

In the second step, the addition of the second thiourea to the free amine was carried out, affording the Boc-protected aryl substituted guanidine intermediates (**53a-d** and **54a-d,** Table 3.8).

Table 3.8. Yields of the asymmetric guanidylation, step 2 of the synthesis of compounds in Families 3A and 4A.

Step 2: asymmetric guanidylation	Comp. no.	Linker	% Yield
	53a	O	76
Ph	**53b**	CO	83
Family 3A	**53c**	CH_2	72
	53d	NH	33
	54a	O	76
(4-Cl-3-CF_3)Ph	**54b**	CO	93
Family 4A	**54c**	CH_2	70
	54d	NH	76

In the case of compounds with the O linker, 76% yield was obtained in the reaction with both the Ph and (4-Cl-3-CF_3) Ph thiourea derivatives. This was expected since the C=S carbon in the thiourea **43** is quite electron-deficient due to the electron-withdrawing effect

of the Boc and aryl groups, making it a good electrophile. The same was expected for compounds with the NH linker; however, quite low yield was obtained in the reaction with the phenyl thiourea. This could be explained by the amine in the *meta*-position which is not a very good nucleophile, or by competition from the NH linker, albeit a deactivated amine as well.

Regarding compounds with the CO linker, 83 and 93% yields were obtained when reacting the *mono*-guanidylated derivative **68b** with **57** and **58**, respectively. The C=S carbon of thiourea **58** should be more electrophilic than that of **57** because of the electron-withdrawing effect of the CF_3 group, and this leads to a higher yield. This difference was not visible in the reactions of the *mono*-guanidylated CH_2 linker derivative **68c**; both the Ph and (4-Cl-3-CF_3) Ph derivatives were obtained in yields of 72 and 70%, respectively.

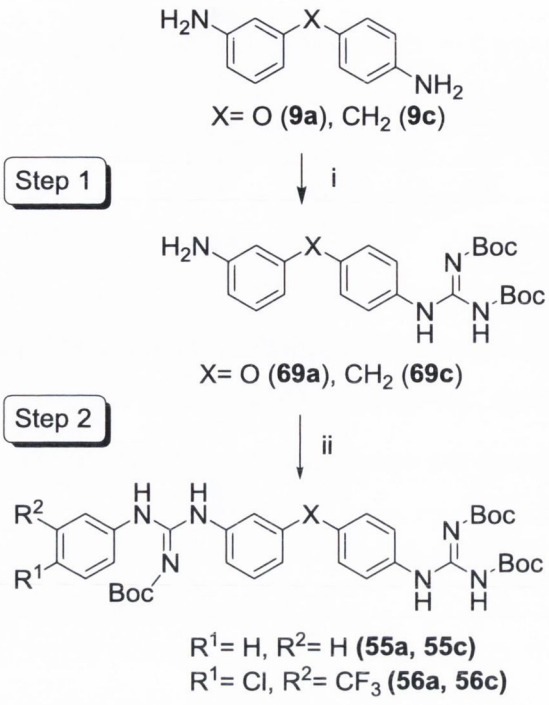
Once the Boc-protected asymmetric intermediates were synthesised, they were deprotected to obtain the hydrochloride salts, which are the biologically active compounds.

3.4.5. Synthesis of Asymmetrically Substituted *meta*-Phenyl Guanidine Derivatives (Families 3B and 4B)

Due to the similar reactivity of the two amines in the *meta* and *para* positions of the CH_2-linked diamine (**9c**), both the *meta* and the *para* *mono*-guanidylated products were obtained during reaction with **43**. Hence, both of these products were isolated and the corresponding hydrochloride salts were obtained for each (one with the substituted guanidine *meta* to the linker and the other with this group *para* to the linker), so a more complete structure-activity relationship study could be made.

The rest of the diamines with different linkers show selectivity for one of the amines over the other, so only the synthesis of the O linked inverse aryl substituted guanidine derivative was carried out, to obtain proof of concept. The reason for this was the commercial availability of the O-linked diamine, which helped us to attain these derivatives with a quick and uncomplicated synthetic route. Due to the higher nucleophilicity of the amine in the *para* position (as had been demonstrated previously), both diamines were reacted with thiourea **43** to obtain the *mono*-guanidylated product in *para* in 77% yield (Scheme 3.34, step 1 and Table 3.9).

Scheme 3.34.



i) **43**, $HgCl_2$, NEt_3, DCM; ii) **57** or **58**, $HgCl_2$, NEt_3, DCM

This was followed by reaction with both thioureas **57** and **58**, respectively to obtain the Boc-protected intermediates in 34 and 70% yield (Scheme 3.34, step 2 and Table 3.9). For those derivatives with an O-linker, this difference in yields could be due to the previously mentioned higher electrophilicity of the C=S carbon of **58** in comparison with **57**. However, this difference is not visible in the CH_2 linked diamines and both the phenyl and (4-Cl-3-CF_3)Ph derivative were obtained in excellent yields around 70% (Table 3.9). In this case the two amines have similar reactivity as they are not influenced by the conjugative effects of a linker group.

Table 3.9. Yield of the generation of compounds in Families 3B and 4B.

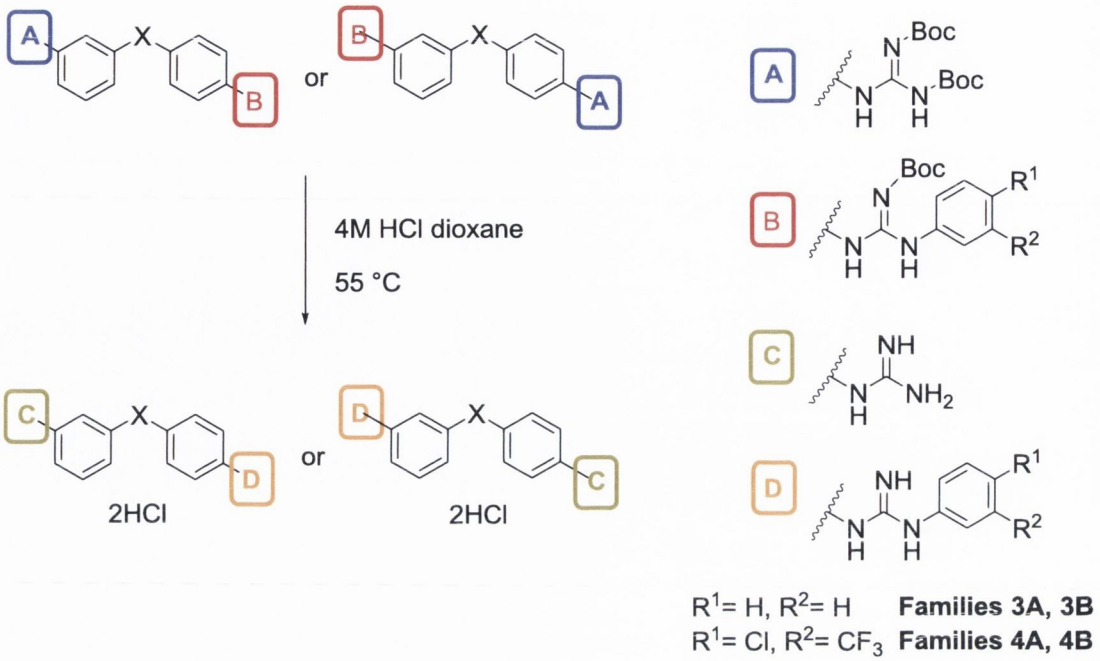
	Derivative	Comp.	Linker	% Yield
Step 1	Boc, Boc guanidine	**69a**	O	77
		69c	CH_2	52
Step 2	Ph guanidine Family 3B	**55a**	O	34
		55c	CH_2	71
Step 2	(4-Cl-3-CF_3)Ph guanidine, Family 4B	**56a**	O	70
		56c	CH_2	68

3.4.6. Synthesis of the Hydrochloride Salts of the Asymmetrically Substituted Phenyl Guanidine Derivatives from Families 3A, 3B, 4A and 4B

In the synthesis of Families 1 and 2, *N,N'-bis*-Boc-protected guanidine intermediates were generally deprotected by treatment with a 1:1 solution of trifluoroacetic acid in DCM at room temperature, obtaining the final hydrochloride salts by treatment with a basic anion-exchange resin (Amberlite IRA-400) in its chloride form. However, this method requires two steps to reach the hydrochloride salt. Taking all these drawbacks into consideration, a new method to generate the hydrochloride salt was explored.

Thus, the Boc groups of the *N,N'-bis*-Boc-protected guanidine intermediates were removed by treatment with 4M HCl in dioxane (6 eq. per Boc group) at 55 °C, until the reaction was adjudged complete (typically 8 h), by TLC (Scheme 3.35). The reason this reaction needed to be heated to 55 °C is that, in the *bis*-Boc-protected guanidines, when one Boc is deprotected and the amine is positively charged, the other Boc group in the molecule is harder to remove.

Scheme 3.35.



At the endpoint of the reaction, solvent and HCl were evaporated under vacuum and the crude salt was dissolved in a minimum volume of water. It was washed with DCM (2 × 10 mL) and then purified using reverse-phase chromatography with C-8 silica and a typical elution gradient of 100% water to 85:15 water:acetonitrile. The purified fractions were evaporated until dry to afford the pure hydrochloride salt. The yields of this reaction are excellent, regardless of the nature of the linker or the structure of the Boc-protected intermediate (Table 3.10).

Table 3.10. Yields obtained in the deprotection of compounds from Family 3A, 3B, 4A and 4B.

Family	Boc-deriv.	X linker	Yield	Salt
3A	**53a**	O	88%	**70a**
	53b	CO	61%	**70b**
	53c	CH_2	94%	**70c**
	53d	NH	58%	**70d**
4A	**54a**	O	88%	**71a**
	54b	CO	54%	**71b**
	54c	CH_2	92%	**71c**
	54d	NH	59%	**71d**
3B	**55a**	O	70%	**72a**
	55c	CH_2	80%	**72c**
4B	**56a**	O	98%	**73a**
	56c	CH_2	80%	**73c**

3.4.7. Isomerism in *N*-(*tert*-Butoxycarbonyl)-*N'*-phenyl Guanidine Compounds and their Hydrochloride Salts

As part of a larger structural study of compounds with similar features taken place in our group, three compounds (Figure 3.21) were studied by DFT computational calculations and low temperature NMR experiments to understand the unstructured signal observed in NMR spectra for aryl di-substituted guanidines. Thus, *N,N'-bis*-phenyl-*N''*-(*tert*-butoxycarbonyl)guanidine (**74**), *N*-phenyl-*N'*-(pyridin-2-yl)-*N''*-(*tert*-butoxycarbonyl) guanidine (**75**), 3-[2,3-di(*tert*-butoxycarbonyl)guanidino]-4'-[2-(*tert*-butoxycarbonyl)-3-(4-chloro-3-trifluoromethyl)phenylguanidino]-*N*-phenylaniline (**54d**).

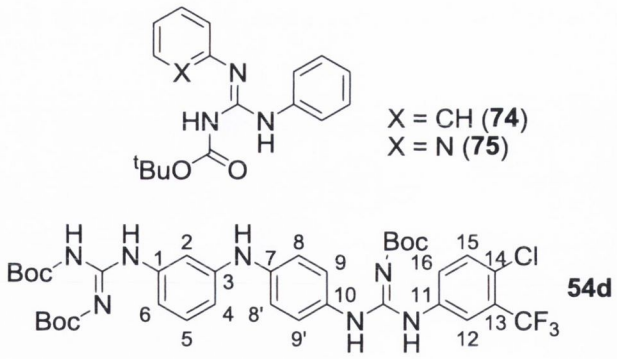


Figure 3.21. Chemical structure of compounds **74**, **75** and **54d**.

As with the *N,N'-bis-(tert*-butoxycarbonyl) analogues of Families 1 and 2 (Section 3.3.6), *N*-(*tert*-butoxycarbonyl)-*N'*-phenylguanidine derivatives from Families 3 and 4 exhibit *E/Z* isomerism with reference to a localised imine. Additionally, in the *N*-(*tert*-butoxycarbonyl)-N'-phenylguanidines, the three substituents of the nitrogens are different. Thus, the localised double bond can be conjugated with any of the three substituents on the guanidine. Considering *E* and *Z* isomerism for each of these tautomers, a total of six isomers are possible (Figure 3.22). This is reflected in the corresponding NMR signals of the aromatic moieties which appear as unstructured multiplets. To explain this, molecule **54d** was selected as a representative example of all the (4-Cl-3-CF_3)-Ph and Ph derivatives synthesised in this project. The CF_3 and Cl groups introduce particular characteristics to the aromatic ring, but the same amount of isomers was observed for the Ph derivatives. The notation used for this study indicates the group with which the double bond is in conjugation with (ac, ar1 or ar2), followed by the configuration of the double bond (E or Z).

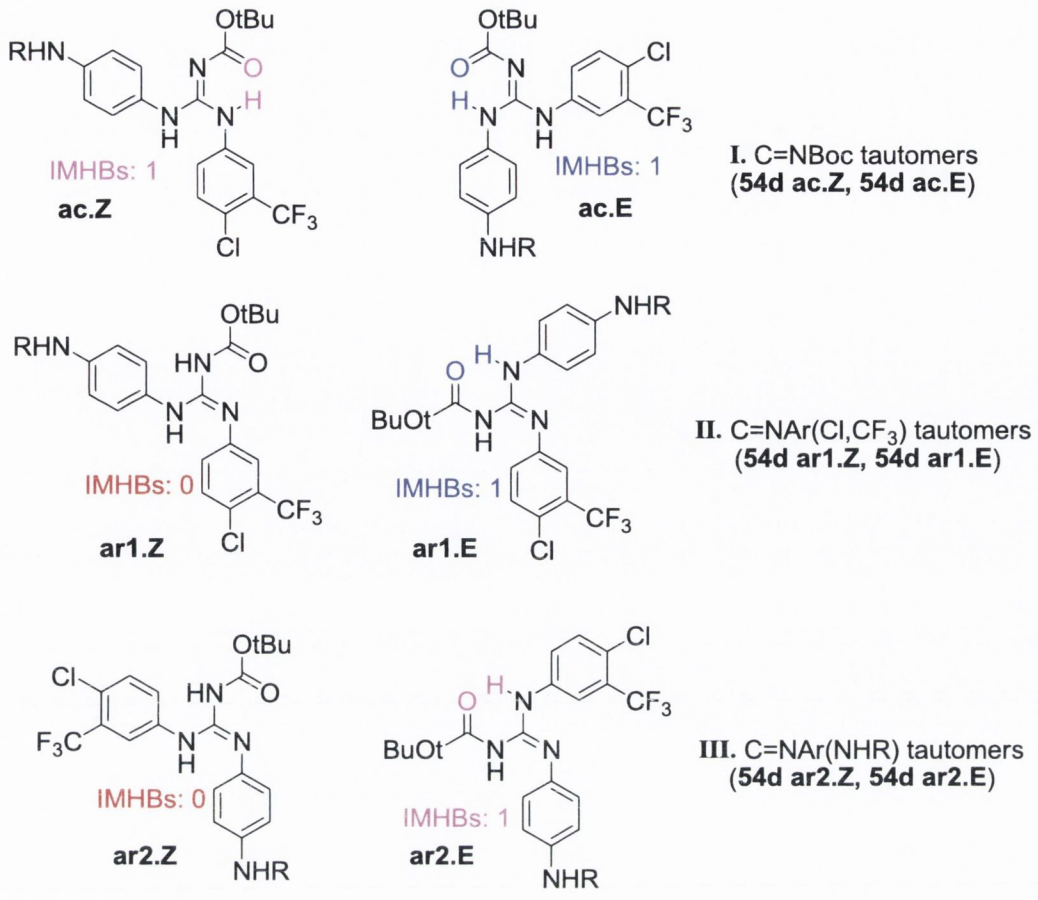


Figure 3.22. Six possible isomers of compound **54d**.

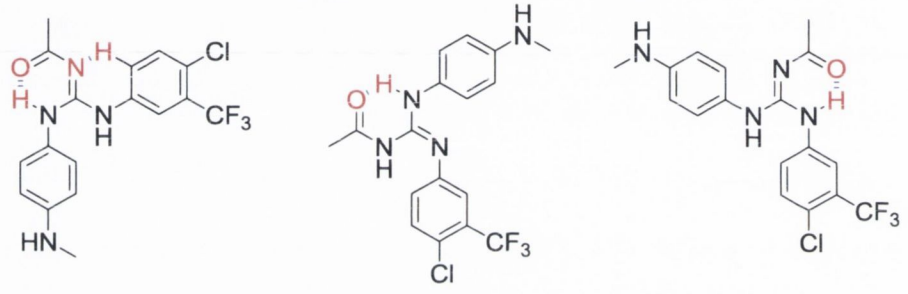
Two tautomers (**54d:ar1.Z** and **54d:ar2.Z**, see Figure 3.22) have no IMHBs and were discarded from the study as they are not stable enough relative to the other isomers. Tautomer **54d:ar2.E** (see Figure 3.22) has one IMHB but the double bond is conjugated with an electron rich aryl group Ar-NHR (R=Ar guanidine) which is not very favorable, as has been shown by Katritzky *et al.*[70] Normally the double bond will be placed closer to the most EW part of the molecule to help to stabilize the double bond by resonance; hence, this isomer is not favoured and was discarded from the study. That these three isomers – **54d**:**ar1.Z**, **54d:ar2.Z** and **54d**:**ar2.E** – were energetically disfavoured was confirmed using a low computational level (B3LYP/6-31G) (see experimental section 7.3).

The remaining three isomers are predicted to be formed in a mixture, and this is supported by the experimental observations. These isomers are those two that have IMHBs and the double bond conjugated with a Boc group (**54d**:**ac.Z**, and **54d**:**ac.E**), and the isomer which places the imine double bond in conjugation with the (4-Cl-3-CF_3)Ph ring (**54d**:**ar2.E**). In

principle, it could be expected that the major isomer would be **54d**:**ar2.E** because the electron-withdrawing effect of the (4-Cl-3-CF_3)Ph ring is large, possibly larger than that of the Boc group. However, this does not correspond with the results obtained in the DFT theoretical study (see Table 3.11). Owing to the complexity of the molecule, a simplified version using an acetyl group as a mimic of the Boc group was used for the computational studies (see structures in Table 3.11), due to the large computational cost of examining the complete system, and the fact that it would unlikely provide any further information.

Table 3.11. DFT calculations of isomers **54d**:**ac.E**, **54d**:**ar1.E** and **54d**:**ac.Z**

ac.*E* **ar1.*E*** **ac.*Z***



Isomer	E_{total} (kJ mol^{-1})	ΔE (kJ mol^{-1})
54d:**ac.E**	-4493568.8	0.0
54d:**ar1.E**	-4493557.7	+11.1
54d:**ac.Z**	-4493552.1	+16.7

The DFT calculations indicate that the most stable isomer is **54d**:**ac.*E***. The optimised geometry shows that the guanidine moiety is co-planar with the (4-Cl-3-CF_3)Ph ring, suggesting a second IMHB between the imine lone-pair and the acidic aromatic CH (Figure 3.23). This type of interaction has been previously observed by us in *N*-phenyl-*N',N''-bis*-Boc-guanidines[61] and it is likely to stabilise the molecule significantly. Isomer **54d**:**ar2.E** also shows the formation of two IMHBs; however, it involves conjugation of the guanidine double bond with a highly electron-rich aromatic ring so the second IMHB will not be as favourable as the first.

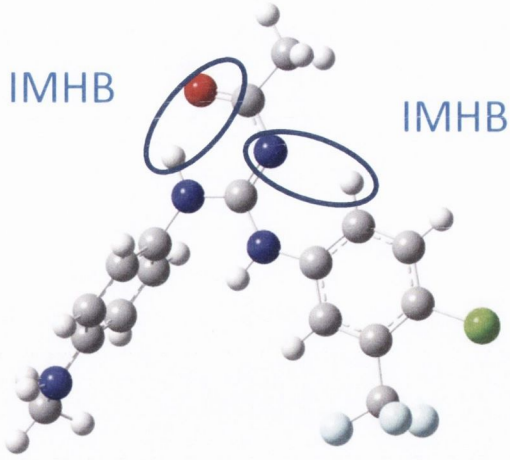


Figure 3.23. Optimised structure of isomer **54d**:**ac.E** where two IMHB are indicated.

The optimised structure of isomer **54d:ar1.E** incorporates conjugation of the guanidine double bond with the electron-deficient (4-Cl-3-CF_3)Ph ring. This, coupled with an IMHB between the acetyl C=O lone-pair and guanidine N-H, makes this isomer stable. However, it lacks the second IMHB found in **54d**:**ac.E** and the (*p*-$NHCH_3$)Ph ring appears to be out of plane with the guanidine subunit (Figure 3.24).

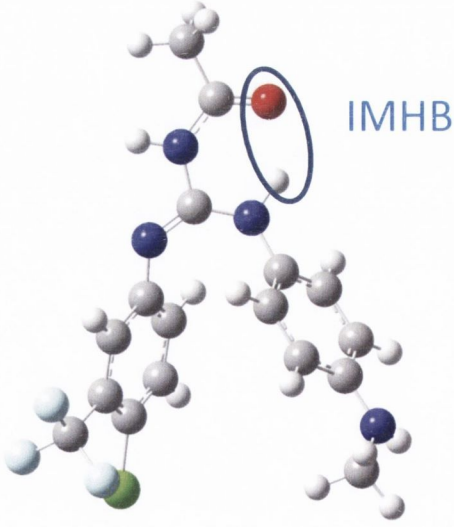


Figure 3.24. Optimised structure of isomer **54d**:**ar1.E.**

Similarly, in the optimised structure of isomer **54d:ac.Z**, this second IMHB interaction is not found. As a result both aromatic rings are twisted to avoid steric repulsion (Figure 3.25).

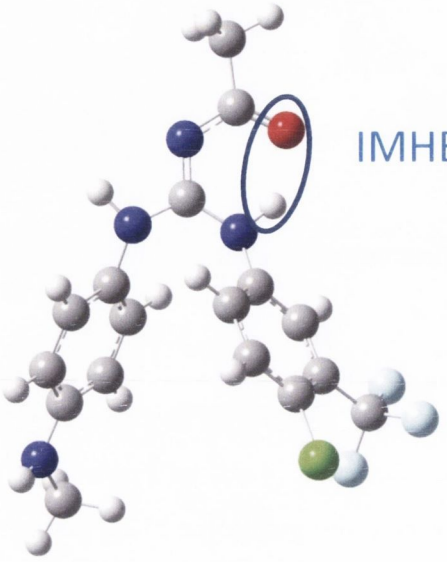


Figure 3.25. Optimised structure of isomer **54d**:**ac.Z.**

The experimental evidence for the existence of these three isomers is based in the NMR spectroscopic results. The ^{1}H NMR spectrum shows three different peaks for NH linking the two aryl rings (Figure 3.26). Also, three H-12 signals can be identified (numbering from Figure 3.21). Additionally, three peaks in the ^{19}F-NMR were observed, which would correspond to the co-existence of three isomers.

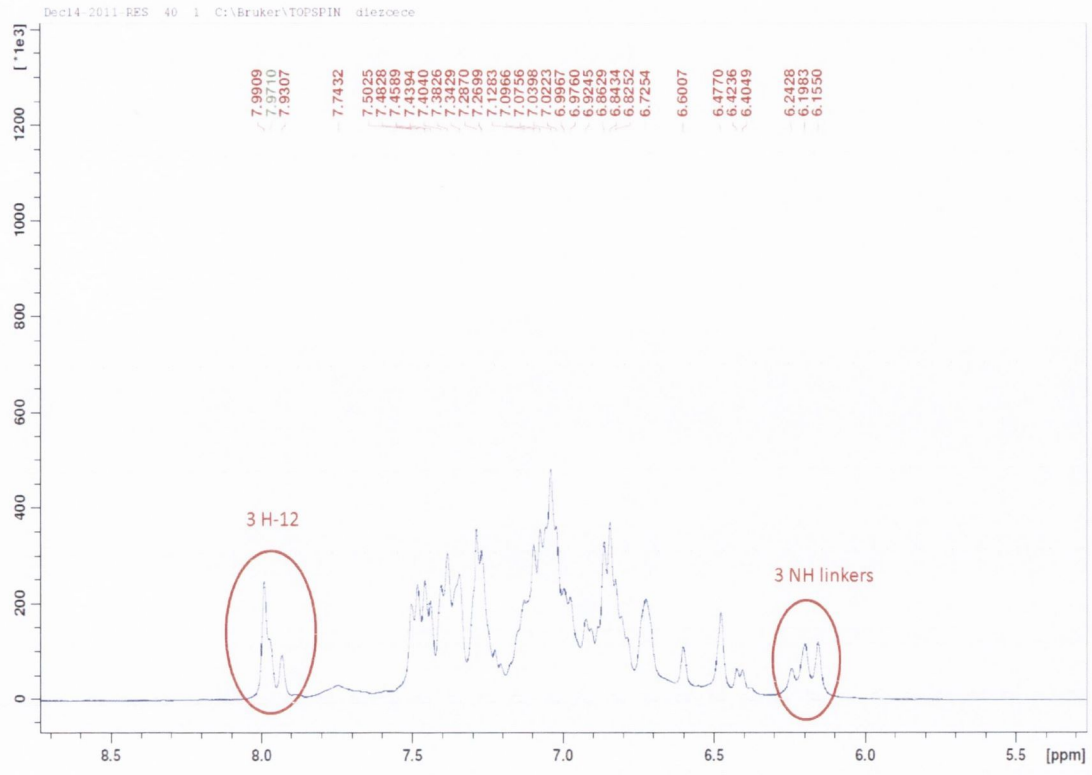


Figure 3.26. ^{1}H NMR ($CDCl_3$) spectrum of **54d**, where evidence of three different isomers was observed.

Due to the mixture of isomers and the steric crowding between the two aromatic rings and the Boc group, shorter relaxation times of the signals result, making the ^{1}H NMR and ^{13}C NMR signals severely broad and amorphous in their corresponding spectra (Figure 3.27). Thus, it was nearly impossible to definitively assign the peaks and their corresponding integrals. The ^{13}C NMR spectra are also amorphous, but sometimes the two dimensional (2D) NMR experiments such as HSQC helped to more clearly identify proton and carbon signals in the molecule. Figure 3.27 shows some examples of spectra obtained for this tri-substituted guanidine **54d**.

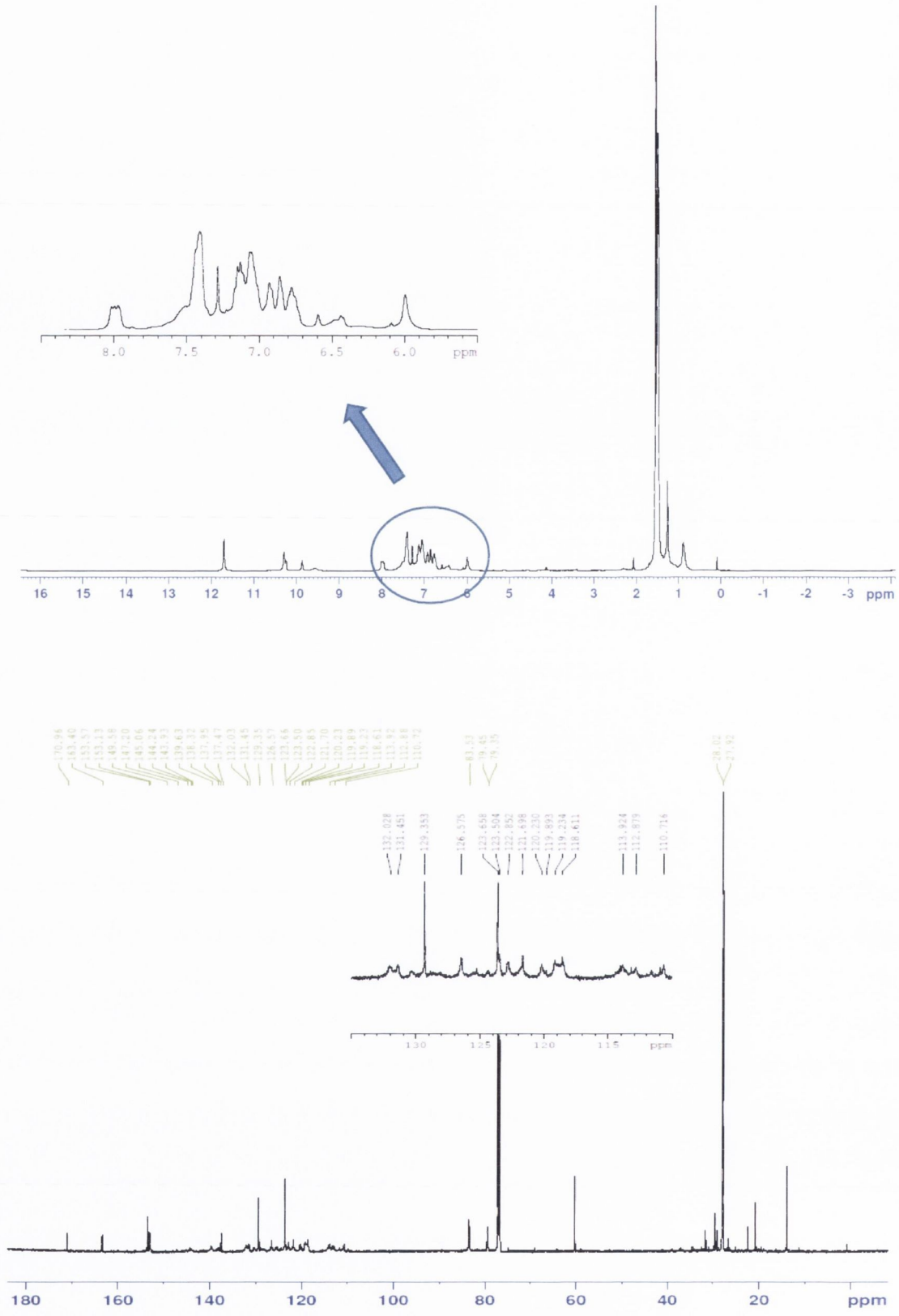


Figure 3.27. 600 MHz ($CDCl_3$) ^{1}H NMR (*top*) and 150 MHz ^{13}C NMR spectra of compound **54d** (*bottom*).

By analysing the NMR spectra, it was not possible to determine the actual isomer ratio, as it is possible that the isomers rapidly interconvert in solution, not being fully detected by the NMR timescale. In an attempt to observe interconversion of isomers and to improve the resolution of the spectra, variable temperature ^{1}H NMR experiments were carried out on compound **54d** (Figure 3.28).

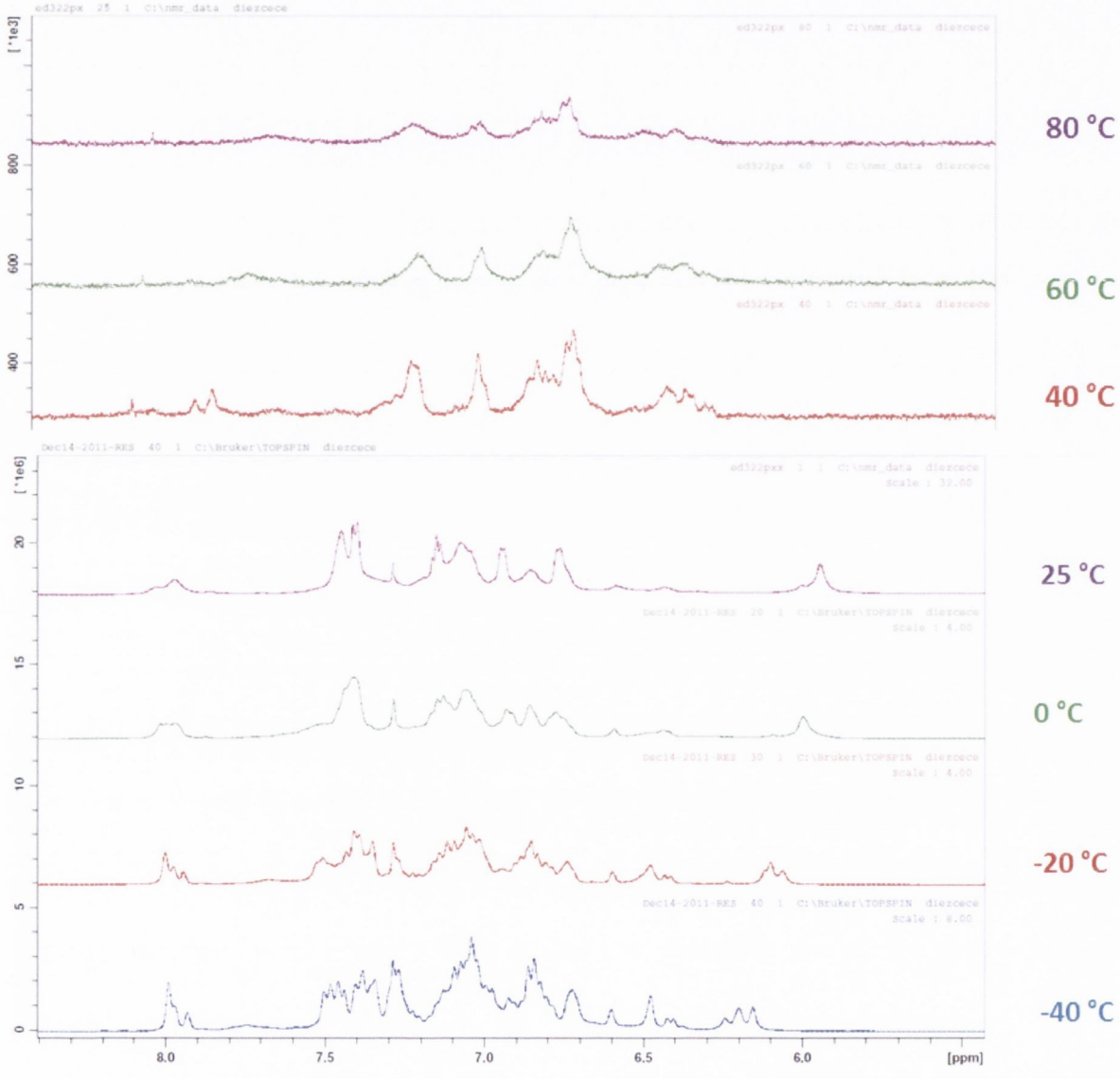


Figure 3.28. Variable temperature ^{1}H NMR ($CDCl_3$) spectra of compound **54d**.

We can see that increasing the temperature further broadens the peaks in the ^{1}H NMR spectrum, while decreasing the temperature significantly sharpens the peaks. The reason for this is that at low temperature, the inter-conversion between the isomers is slower, as the energy of the system is lower, allowing a more defined spectrum to be recorded. These low temperature studies were conducted in collaboration with Martin Kitson in DIT College in Dublin.

Thankfully, deprotection of the Boc group alleviates the steric congestion experienced by the aromatic rings. The relaxation times return to ordinary levels and the ^{1}H-NMR and ^{13}C-NMR aromatic NMR signals recorded for the corresponding deprotected *N,N'*-disubstituted guanidine hydrochloride salt (**71d**) are sharpened as can be seen in Figure 3.29.

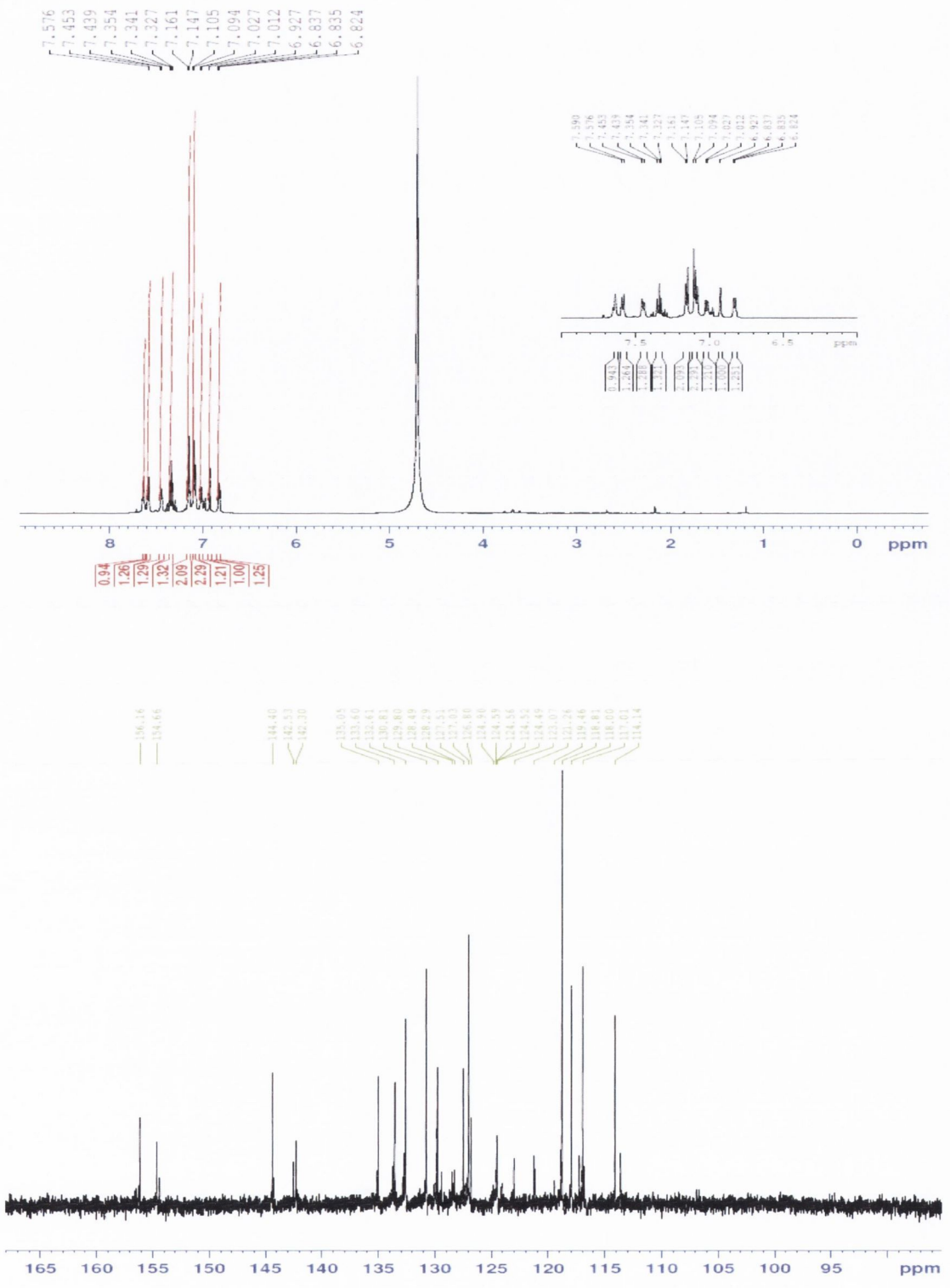


Figure 3.29. 600 MHz (D_2O) ^{1}H NMR (*top*) and 150 MHz ^{13}C NMR spectra of 3-{4'-[(4-chloro-3-trifluoromethylphenyl)guanidino]phenylamino}phenylguanidine dihydrochloride **71d** (*bottom*).

Therefore, it was possible to fully assign all the ^{1}H-NMR and ^{13}C-NMR signals of the hydrochloride salts of compounds from Families 3 and 4.

3.5. Synthesis of Acetamide Derivatives – Family 5

3.5.1. General Aspects of Family 5

As was previously mentioned in the objectives of this project, the final family of guanidine-based compounds synthesised was a series of acetamide derivatives (Figure 3.30). This family represents the closest analogue to the sorafenib structure.

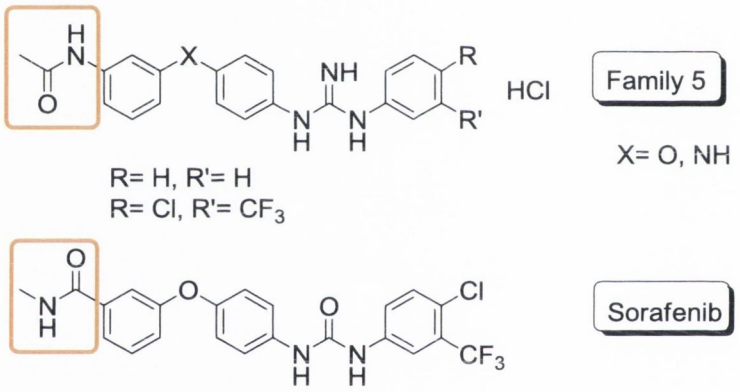


Figure 3.30. General structure of compounds from Family 5 (*top*) and sorafenib structure (*bottom*).

Regarding the retrosynthesis of these compounds, their core is based on the 3,4'-diphenyl diamines previously described in section 3.2. Then, as in Families 3 and 4, the amine in the *para* position was coupled with *N*-(*tert*-butoxycarbonyl)-*N'*-phenylthiourea derivatives **57** and **58**, and then, the amine in the *meta* position relative to the linker was acetylated, instead of guanidylated (Figure 3.31).

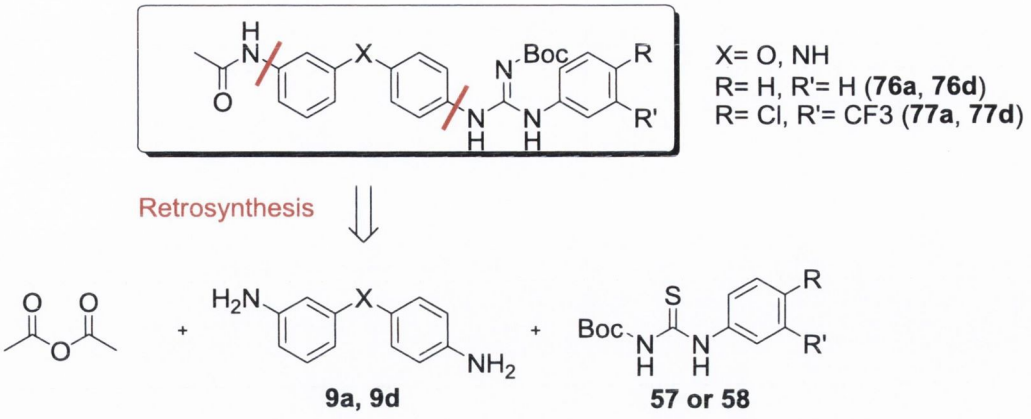
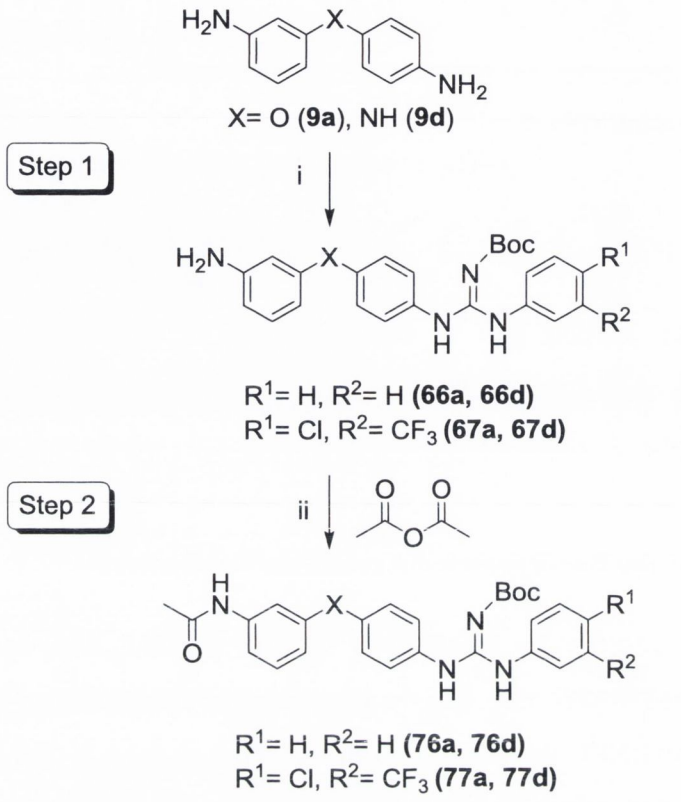


Figure 3.31. Retrosynthesis plan for the synthesis of compounds in Family 5.

3.5.2. Synthesis of the Boc-Protected Acetamide Derivatives

Derivatives of Family 5 were synthesised from compounds **66a, 66d, 67a** and **67d** (as previously described in section 3.4.4) by acetylation of the remaining primary amine using acetic anhydride and NEt_3 in DCM at room temperature for 6 h (Scheme 3.36), affording the Boc-protected derivatives **76a, 76d, 77a** and **77d** in good yields (Table 3.12).

Scheme 3.36.



i) **57** or **58**, $HgCl_2$, NEt_3, DCM; ii) NEt_3, DCM, rt, 6h

Table 3.12. Yields of the acetylation reaction, step 2 in the synthesis of compounds in Family 5.

	Derivative	Comp. no.	Linker	% Yield
Step 2	Ph	**76a**	O	53
	Ph	**76d**	NH	69
	(4-Cl-3-CF_3)Ph	**77a**	O	89
	(4-Cl-3-CF_3)Ph	**77d**	NH	61

Considering that both of these linkers enhance the reactivity of the amine in the *para* position and not that in the *meta* position, the yields observed for the acetylation of this amine are very satisfactory. Intriguingly, in the O linked derivatives, a better yield was

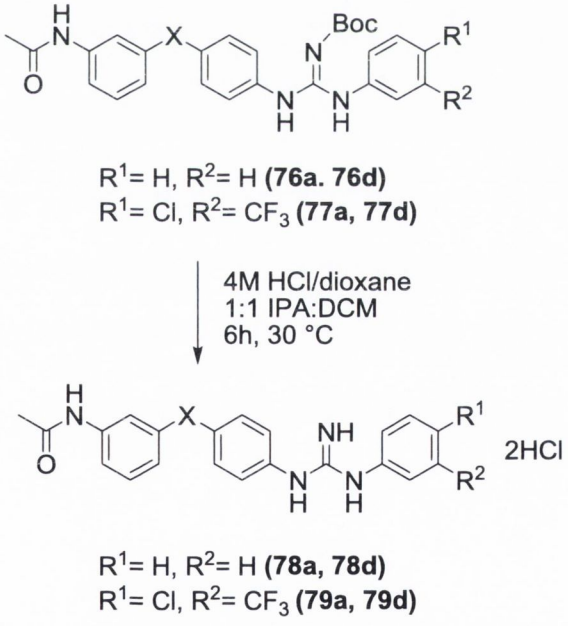
observed when acetylating the (4-Cl-3-CF_3)Ph derivative than the Ph one. However, the NH linker derivatives were obtained in similar proportions.

The isomerism observed in the Boc-protected structures of Families 3 and 4 was also present in these derivatives in Family 5 due to the tri-substitution of the guanidine; hence, the ^{1}H NMR and ^{13}C NMR spectra were also amorphous, making it difficult to completely assign all of the signals.

3.5.3. Synthesis of the Hydrochloride Salts of the Acetamide Derivatives

Boc deprotection was carried out using 4 M HCl in dioxane, diluted in a 1:1 solution of IPA:DCM, such that a final concentration of 0.2 M was reached, at 30 °C for 6 h, affording the final hydrochloride salts **78a, 78d, 79a** and **79d** (Scheme 3.37).

Scheme 3.37.



This deprotection requires milder conditions than for the derivatives form Families 3 and 4 due to the liability of the acetamide group in acidic conditions. Thus, dry conditions were imperative, and the concentration of the reaction was reduced accordingly. A mixture of

IPA:DCM was preferred as it had the best combination of being dry, and aiding the solubility of starting materials. The yields obtained in this step are shown in Table 3.13.

Table 3.13. Yields of the deprotection step to form the final hydrochloride salts of compounds in Family 5.

Boc-intermediate	X linker	% Yield	Salt
76a	O	61	**78a**
76d	NH	56	**78d**
77a	O	52	**79a**
77d	NH	43	**79d**

The yields obtained in the generation of the hydrochloride salts are moderate, in comparison to those obtained in the rest of the families. Even though milder conditions were used for this step to avoid the cleavage of the acetyl group, it is possible that the moderate yields observed were due to this fact. Prior to biological evaluation of the salts, purity was assessed by HPLC, confirming that no hydrolysed side-products remained after purification by reverse-phase chromatography.

3.6. Conclusions

This chapter presents the synthetic strategies explored in the synthesis of a novel type of potential PKI containing the guanidine functionality. Based on methodology previously described by Kim and Qian, and subseuently explored by Rozas and co-workers, the synthesis of five novel families of 3,4-*bis*-guanidininium and -2-aminoimidazolinium derivatives was successfully achieved. Most of the starting 3,4'-diphenyl diamines were not commercially available and needed to be synthesised. To achieve this, several synthetic routes were followed. Preparation of Families 1 and 2 follows the synthetic route of guanidylation previously used in the synthesis of 4,4'-*bis*-guanidines and -2-aminoimidazolines, in which the guanidylation process of the di-phenyl diamines involved the use of *N,N'*-*bis*-(*tert*-butoxycarbonyl)thiourea (**43**) in presence of $HgCl_2$ and NEt_3. This was then followed by Boc-deprotection using TFA/DCM to generate trifluoroacetate salts, followed by treatment with Amberlite resin to generate the required hydrochloride salts. However, for the synthesis of the rest of the families in this project, new synthetic procedures were introduced.

For the synthesis of Families 1 and 2, the 3,4'-diphenyl diamines were coupled with **43** and *N,N'*-bis-(*tert*-butoxycarbonyl)-imidazoline-2-thione (**46**) respectively, with $HgCl_2$ and NEt_3 to afford the Boc-protected 3,4-*bis*-guanidinium and -2-aminoimidazolinium derivatives (**41a-d** and **42a-d**). This was then followed by Boc-deprotection using TFA/DCM to generate trifluoroacetate salts, which were treated with Amberlite resin in its chloride form to generate the hydrochloride salts in excellent yields (**51a-d** and **52a-d**).

For Families 3 and 4, *N*-(*tert*-butoxycarbonyl)-*N'*-phenyl substituted thiourea derivatives such as *N*-(*tert*-butoxycarbonyl)-*N'*-phenylthiourea (**57**) and *N*-(*tert*-butoxycarbonyl)-*N'*-(4-chloro-3-trifluoromethylphenyl)-thiourea (**58**) needed to be prepared prior to the synthesis of these new derivatives. Using the Kim and Qian method, members of Families 3A and 4A were synthesised by coupling **57** and **58** with the amine in the *para* position relative to the linker and the amine in the *meta* position with **43**. The opposite substitution pattern was incorporated for members of the Families 3B and 4B. This will give us a more complete structure-activity relationship infomation. These steps afford the Boc-protected derivatives (**53-54 a-d**, **55-56a**, **c**) in good yields. The Boc groups were deprotected using 4M HCl in dioxane to afford the final hydrochloride salts in excellent yields (**70-71a-d**, **72-73a**, **c**). This deprotection method is a better route to obtaining the hydrochloride salts in

just one step from the Boc derivatives, in comparison with the TFA/DCM and exchange with Amberlite resin methodology used in Families 1 and 2.

A conformational aspect of these di-Boc-protected guanidine derivatives was discussed. They form IMHBs which define their structural conformation, making the phenyl ring and the Boc-protected guanidine co-planar. However, this phenomenon does not occur in the guanidinium salts, where this moiety and the aromatic ring are placed out of plane relative to each other to avoid repulsion of aryl and guanidinium protons.

The signals in the ^{1}H and ^{13}C NMR spectra of the Boc- protected intermediates in Families 3 and 4 were highly broad and amorphous due to the presence of isomerism of these tri-substituted guanidine derivatives. This phenomenon was studied through low-temperature NMR spectroscopy and DFT computational studies. However, when the Boc groups were removed, this isomerism disappeared and the ^{1}H and ^{13}C NMR spectra of the hydrochloride salts were normal and sharply defined.

Finally, for the synthesis of Family 5, the acetylation of the *mono*-guanidylated products obtained in the first step of the synthesis of Families 3 and 4 was carried out. This acetylation involved the use of acetic anhydride and NEt_3 in DCM at room temperature for 6 h, affording the Boc-protected derivatives (**76-77a,d**) in good yields. Then, the hydrochloride salts were obtained using the 4M HCl/dioxane method to afford the final hydrochloride salts (**78-79a,d**) in moderate yields due to the sensitivity of the acetamide group to acidic conditions.

The purity of all of the final hydrochloride salts was determined by HPLC, where a minimum purity of 95% was required before proceeding to biological testing. Thus, five novel families of potential protein kinase inhibitors were successfully synthesised and fully characterized.

3.7. References

(1) Berlinck, R. G. S.; Burtoloso, A. C. B.; Kossuga, M. H. *Nat. Prod. Rep.* **2008**, *25*, 919-954.

(2) Berlinck, R. G. S.; Trindade-Silva, A. E.; Santos, M. F. C. *Nat. Prod. Rep.* **2012**, 10.1039/c2np20071f.

(3) Berlinck, R. G. S.; Burtoloso, A. C. B.; Trindade-Silva, A. E.; Romminger, S.; Morais, R. P.; Bandeira, K.; Mizuno, C. M. *Nat. Prod. Rep.* **2010**, *27*, 1871-1907.

(4) Rauws, T. R. M.; Maes, B. U. W. *Chem. Soc. Rev.* **2012**, *41*, 2463-2497.

(5) Rodriguez, F.; Rozas, I.; Ortega, J. E.; Meana, J. J.; Callado, L. F. *J. Med. Chem.* **2007**, *50*, 4516-4527.

(6) Rodriguez, F.; Rozas, I.; Ortega, J. E.; Erdozain, A. M.; Meana, J. J.; Callado, L. F. *J. Med. Chem.* **2008**, *51*, 3304-3312.

(7) Rodriguez, F.; Rozas, I.; Ortega, J. E.; Erdozain, A. M.; Meana, J. J.; Callado, L. F. *J. Med. Chem.* **2009**, *52*, 601-609.

(8) Dardonville, C.; Barrett, M. P.; Brun, R.; Kaiser, M.; Tanious, F.; Wilson, W. D. *J. Med. Chem.* **2006**, *49*, 3748-3752.

(9) Rodríguez, F.; Rozas, I.; Kaiser, M.; Brun, R.; Nguyen, B.; Wilson, W. D.; García, R. N.; Dardonville, C. *J. Med. Chem.* **2008**, *51*, 909-923.

(10) Nagle, P. S.; Rodriguez, F.; Kahvedžić, A.; Quinn, S. J.; Rozas, I. *J. Med. Chem.* **2009**, *52*, 7113-7121.

(11) Kahvedžić A., Ph.D thesis, 2011.

(12) Kahvedžić, A.; Nathwani, S.-M.; Zisterer, D.; Rozas, I. *J. Med. Chem* **2013**, *56*, 451-459

(13) Cope, O. J.; Brown, R. K. *Can. J. Chem.* **1961**, *39*, 1695-1710.

(14) Haber, F. *Z. Physik. Chem.* **1900**, *32*, 193.

(15) Kelly, B.; Sánchez-Sanz, G.; Blanco, F.; Rozas, I. *Comp. Theor. Chem.* **2012**, *998*, 64-73.

(16) Clemmensen, E. *Chem. Ber.* **1913**, *46*, 1837.

(17) Kishner, N. *J. Russ. Chem. Soc.* **1911**, *43*, 582.

(18) Todd, D. *Org. React.* **1948**, *4*, 378.

(19) Mozingo, R.; Wolf, D. E.; Harris, S. A.; Folkers, K. J. *J. Am. Chem. Soc.* **1946**, *65*, 1013.

(20) Allen, C. F. H.; VanAllan, J. *Org. Synth. Coll. Vol.* **1955**, *3*, 63.

(21) Burdon, J.; Price, R. C. *J. Chem. Soc., Chem. Commun.* **1986**, 893.

(22) Dewar, M. J. S.; Mole, T. *J. Chem. Soc.* **1956**, 2556.

(23) Kuhn, L. P. *J. Am. Chem. Soc.* **1951**, *73*, 1510.

(24) Raney, M.; U.S. Patent 1,628,190, 1927.

(25) Binz, A.; Schickh, O. V. *Chem. Ber.* **1935**, *68*, 315.

(26) Durham, L. J.; McLeod, D. J.; Cason, J. *Org. Synth. Coll. Vol.* **1963**, *4*, 510.

(27) Diez-Cecilia, E.; Kelly, B.; Rozas, I. *Tetrahedron Lett.* **2011**, *52*, 6702-6704.

(28) Ung, S.; Falguieres, A.; Guy, A.; Ferroud, C. *Tetrahedron Lett.* **2005**, *46*, 5913.

(29) Wu, G.; Zhu, J.; Mo, X.; Wang, R.; Terskikh, V. *J. Am. Chem. Soc.* **2010**, *132*, 5143.

(30) Samet, A. V.; Kislyi, K. A.; Marshalkin, V. N.; Strelenko, Y. A.; Nelyubina, Y. V.; Lyssenko, K. A.; Semenova, V. V. *Russ. Chem. Bull. Int. Ed.* **2007**, *56*, 2089.

(31) Neumeyer, J. L.; Cannon, J. C. *J. Org. Chem.* **1961**, *26*, 4681-4682.

(32) Hesek, D.; Lee, M.; Fisher, J. F.; Mobashery, S. *J. Org. Chem.* **2009**, *74*, 2567-2570.

(33) Kosugi, M.; Kameyama, M.; Migita, T. *Chem. Lett.* **1983**, *12*, 927-928.

(34) Hartwig, J. F.; Paul, F. *J. Am. Chem. Soc.* **1995**, *117*, 5373-5374. Refer to respective group websites for comprenhensive representation of this work.

(35) Driver, M. S.; Hartwig, J. F. *J. Am. Chem. Soc.* **1997**, *119*, 8232-8245.

(36) Fors, B. P.; Buchwald, S. L. *J. Am. Chem. Soc.* **2009**, *131*, 12898-12899.

(37) King, A. O.; Yasuda, N. *Top. Organomet. Chem.* **2004**, *6*, 205-245.

(38) Tsuji, J. *Palladium reagents and catalysts*; WILEY, 2004.

(39) Surry, D. S.; Buchwald, S. L. *Angew. Chem. Int. Ed. Engl.* **2008**, *47*, 6338-6361.

(40) Littke, A. F.; Fu, G. C. *Angew. Chem. Int.* **2002**, *41*, 4176-4211.

(41) Yang, B. H.; Buchwald, S. L. *J. Organome. Chem.* **1999**, *576*, 125-146.

(42) Surry, D. S.; Buchwald, S. L. *Chem. Sci.* **2011**, *2*, 27-50.

(43) Wolfe, J. P.; Tomori, H.; Yin, J.; Buchwald, S. L. *J. Org. Chem.* **2000**, *65*, 1158-1174.

(44) Guram, A. S.; Rennels, R. A.; Buchwald, S. L. *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 1348-1350.

(45) Wolfe, J. P.; Buchwald, S. L. *Tetrahedron Lett.* **1997**, *38*, 6359-6362.

(46) Tasler, S.; Mies, J.; Lang, M. *Adv. Synth. Catal.* **2007**, *349*, 2286-2300.

(47) Shen, Q.; Ogata, T.; Hartwig, J. F. *J. Am. Chem. Soc.* **2008**, *130*, 6586-6596.

(48) Wolfe, J. P.; Buchwald, S. L. *J. Org. Chem.* **1996**, *61*, 1133-1135.

(49) Kim, K. S.; Qian, L. *Tetrahedron Lett.* **1993**, *34*, 7677-7680.

(50) Dardonville, C.; Goya, P.; Rozas, I.; Alsasua, A.; Martin, M. I.; Borrego, M. J. *Bioorg. Med. Chem.* **2000**, *8*, 1567-1577.

(51) Wermann, K.; Walther, M.; Günther, W.; Görls, H.; Anders, E. *Tetrahedron* **2005**, *61*, 673-685.

(52) Cunha, S.; Rodrigues Jr, M. T.; da Silva, C. C.; Napolitano, H. B.; Vencato, I.; Lariucci, C. *Tetrahedron* **2005**, *61*, 10536-10540.

(53) Yong, Y. F.; Kowalski, J. A.; Lipton, M. A. *J. Org. Chem.* **1997**, *62*, 1540-1542.

(54) Blair, A.; Stevenson, L.; Sutherland, A. *Tetrahedron Lett.* **2012**, *53*, 4084-4086.

(55) Hammoud, H.; Schmitt, M.; Bihel, F.; Antheaume, C.; Bourguignon, J. J. *J. Org. Chem.* **2011**, *77*, 417-423.

(56) Stahle, H.; Koppe, H.; Kummer, W.; Stockhaus, K.; U.S. Patent 3,937,717, 1976.

(57) Wai Ming, K.; Shih-Hsun, L.; Ching-Yuh, C. *Synthetic Commun.* **2005**, *35*, 2633-2639.

(58) Connolly, D. J.; Cusack, D.; O'Sullivan, T. P.; Guiry, P. J. *Tetrahedron* **2005**, *61*, 10153-10202.

(59) Dardonville, C.; Brun, R. *J. Med. Chem.* **2004**, *47*, 2296-2307.

(60) Dardonville, C.; Rozas, I.; Alkorta, I. *J. Mol. Graphics Mod.* **1999**, *16*, 150.

(61) Kelly, B.; O'Donovan, D. H.; O'Brien, J.; McCabe, T.; Blanco, F.; Rozas, I. *J. Org. Chem.* **2011**, *76*, 9216-9227.

(62) Pennypacker, K.; Cuevas, J.; Antilla, J.; Cortes-Salva, M.; U.S. Patent WO2010048164 (A2), 2010.

(63) Cortes-Salva, M.; Nguyen, B.-L.; Cuevas, J.; Pennypacker, K. R.; Antilla, J. C. *Org. Lett.* **2010**, *12*, 1316-1319.

(64) Powell, D. A.; Ramsden, P. D.; Batey, R. A. *J. Org. Chem.* **2003**, *68*, 2300-2309.

(65) Dodd, D. S.; Kozikowski, A. P. *Tetrahedron Lett.* **1994**, *35*, 977-980.

(66) Linton, B. R.; Carr, A. J.; Orner, B. P.; Hamilton, A. D. *J. Org. Chem.* **2000**, *65*, 1566-1568.

(67) Yin, B.; Liu, Z.; Yi, M.; Zhang, J. *Tetrahedron Lett.* **2008**, *49*, 3687-3690.

(68) Yin, B.; Liu, Z.; Zhang, J.; Li, Z. *Synthesis* **2010**, *6*, 991-999.

(69) O'Donovan, D. H.; Rozas, I. *Tetrahedron Lett.* **2011**, *52*, 4117-4119.

(70) Ghiviriga, I.; El-Gendy, B. E. D. M.; P.J. Steel; Katritzky, A. R. *Org. Biomol. Chem.* **2009**, *7*, 4110-4119.

Chapter 4 – Biological Evaluation of Potential Protein Kinase Inhibitors

4.1. Introduction

This chapter describes the biological evaluation of the five synthesised families of potential protein kinase inhibitors. This includes assessment of their cytotoxic activity in a wide range of hematopoietic cancer cells such as human promyelocytic leukemia HL-60 and solid tumour cell lines such as liver carcinoma HUH-7, colorectal cancer RKO and breast cancer MCF-7. These cancer cell lines were chosen as there are examples in the literature that verify that sorafenib is active in all of these cancers and thus, our novel guanidine analogues could have a similarly broad biological profile.

Cancer is the second leading cause of death following heart disease, accounting for 23% of all deaths. Each year, more than 11 million new cases are diagnosed, and more than seven million people die from cancer – over 70% of them in developing countries.[1] Leukaemia, liver carcinoma, colorectal carcinoma and breast cancer are amongst the top ten most common cancer types according to the latest statistics in 2012.[2] The percentages of estimated deaths in 2012 due to breast, colon, leukaemia and liver cancer are 14% in women, 9% in both sexes, 4% in both sexes, and 5% in men and 2% in women, respectively.[2] Additionally, leukaemia is the leading cause of cancer death among children and adolescents, breast cancer ranks first amongst women ages 20 to 59 years, whilst lung cancer ranks first amongst men aged 40 years and older.[2] As a result, the four cancers investigated in this work are of paramount importance to society at present.

In most cases, the mode of action of sorafenib in the different types of cancer cells used in this study is already characterised. It is known that sorafenib induces apoptosis in HL-60 cells by inhibiting STAT3 phosphorylation, and by decreasing the expression of the anti-apoptotic proteins MCL-1 and BCL-2, (associated with increased apoptosis in HL-60 cells) thus providing a rationale for the treatment of human AML using sorafenib.[3]

Relevant to HCC and breast carcinoma, sorafenib has been shown to down-regulate the RAS/RAF/MAPK pathway in HUH-7 and MCF-7 cell lines, as a mono-therapy and, in both cancers, combination with other drugs enhances their cytotoxic effect. Thus in HCC, sorafenib has been used in combination with PI3K/AKT/mTOR inhibitors identifying a

more successful treatment as both pathways are interrelated.[4,5] Additionally, sorafenib administration was shown to confer a survival advantage to hepatoma cells, by inducing autophagosome formation and enhanced autophagic activity. Similarly, for the treatment of breast cancer, sorafenib is used in combination with epirubicin or desatinib, exerting very satisfactory results.[6,7,8]

Mutations of the K-RAS gene have been described in colorectal cancer (CRC) patients as one of the causes of constitutive proliferation of cancer cells. EGFR inhibitors such as cetuximab, panitumumab, erlotinib, gefitinib and vandetanib show a poor ability to inhibit this anomaly when a resistance to EGFR inhibitors has been developed. However, sorafenib can inhibit cell proliferation as RAS and RAF are both downstream signals of both the EGFR and VEGFR pathways; hence, treatment of the resistant cells with sorafenib, which targets the RAS/RAF/MAPK pathway, stops cell proliferation.[9] Additionally sorafenib has been used with 5-fluorouracil (an anti-metabolite) or rapamycin (mTor inhibitor) for the treatment of colorectal cancer.[10,11]

Another aspect to take into consideration is how sorafenib enters the cell in order to exert its cytotoxic effect in all the cancer cells mentioned. Studies have shown that tyrosine kinase inhibitors are substrates for various ATP-binding cassette (ABC) transporters, and these interactions may play an important role in modulating systemic pharmacokinetics of drugs, tissue and brain distribution, and cellular accumulation and resistance.[12] Unlike other tyrosine kinase inhibitors, sorafenib does not appear to rely much on active transport to enter the cell. In particular, sorafenib uptake in human hepatocytes occurs via passive diffusion and only partly by active transport. Human organic cationic transporter 1 (hOCT1) might contribute to this active uptake as well as organic anion transporting polypeptides (OATPs) transporters.[13] Despite the similarities of our compounds with sorafenib, they contain two guanidinium cations and dicationic drugs have very slow rates of diffusion across biological membranes, and their uptake is dependent on different transporters.[14] For this reason, we suggest that our drugs enter the cell by hOCT1. This hypothesis is supported by the fact that furamidine (see structure in Figure 1.3 a), a minor groove binder that contains two amidine cations, (similar to guanidiniums), also enters the cell by hOCT1 transporter.[15]

Based on this information, the cytotoxic evaluation of our compounds was undertaken, followed by the assessment of their apoptosis inducing ability, where morphological and

flow cytometry studies were carried out. Finally, their ability to inhibit specific tyrosine kinases was examined.

4.2. Cytotoxicity Evaluation of the Novel Compounds

In vitro cell viability studies are crucial to identify the efficacy of a library of compounds in an uncomplicated and fast manner and allow selection of the most potent ones for further drug development studies. Thus, in this project viability studies were carried out on all compounds to identify their effects on cell viability and proliferation, while apoptosis evaluation and specific kinase studies were performed on the most successful derivatives from these tests.

There are various methods for the evaluation of cell viability. These are based on various cell functions such as enzyme activity, cell membrane permeability, cell adherence, ATP production, co-enzyme production and nucleotide uptake activity. Additionally, there are several established methods used for counting the number of living cells (Figure 4.1), such as the colony formation method, the crystal violet method, the tritium-labelled thymidine uptake method, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazoliumbromid (MTT), water soluble tetrazolium salts (WST) and AlamarBlue®.[16]

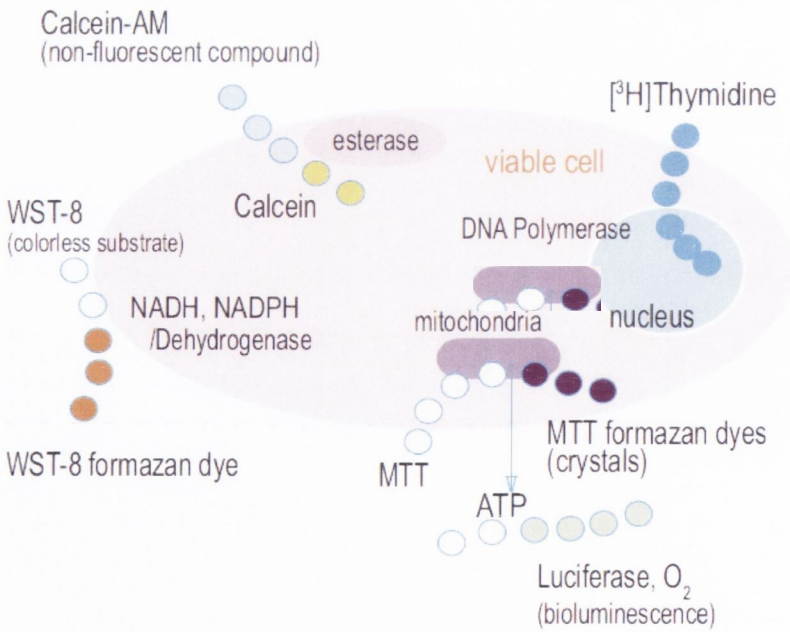


Figure 4.1. Reagents for cell viability detection.[16]

The first three methods mentioned for detection of cell viability are not very sensitive and are quite difficult to handle. However, the last three are far superior since they are easy to use, safe, highly reproducible and widely used. The MTT and WST methods are based on a colorimetric method to determine cell viability, which involves the use of a colouring reagent and a dehydrogenase. Dehydrogenase-based assays reflect cell conditions with more sensitivity than the other assays as they depend on several elements including dehydrogenase, NAD(H), NADP(H) and mitochondrial activity. In the method, MTT is reduced by NADH to a purple formazan which is insoluble in water and precipitates forming crystals. Thus, prior to measuring the absorbance, an organic solvent is required to solubilise the crystals. However, the WST method is based on water soluble triazonium salts, producing water-soluble formazans, which avoid this inconvenience.

Similarly, AlamarBlue® viability assays do not have solubility issues. This method is based on a metabolic-indicator dye which is sensitive, non-radioactive, water-soluble and non-toxic with respect to cells. In its normal form, the fluorogenic redox indicator is in an oxidized state (resazurin, Figure 4.2) and is dark blue in colour, with little intrinsic fluorescence. When exposed to the reducing environment of the proliferating cells, it is reduced to resorufin, which is pink and highly fluorescent at 590 nm wavelength. The extent of this conversion, which may be spectrophotometrically quantified by absorbance or fluorescence, corresponds to cell viability as it is only metabolised by enzymatic reactions of living cells.

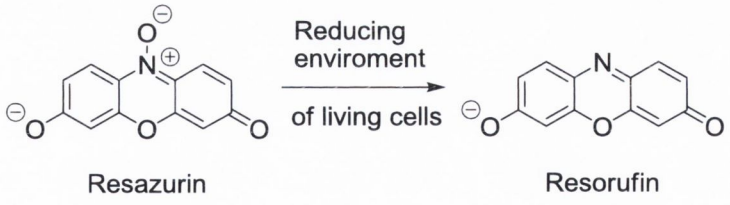


Figure 4.2. Resazurin reducing to resorufin is the principle of AlamarBlue® viability assays.

Additionally, AlamarBlue® does not require additional reagents or manipulations and it is non-toxic; hence, it does not require any special handling or disposal methods. Also, due to its homogenous nature, it is applicable to many biological systems.

Each cancer cell line has its own metabolic properties and therefore, cell density, incubation time and volume of reagent must be considered before setting up the assay. Nevertheless, a general protocol, which is illustrated in Figure 4.3, can be followed.

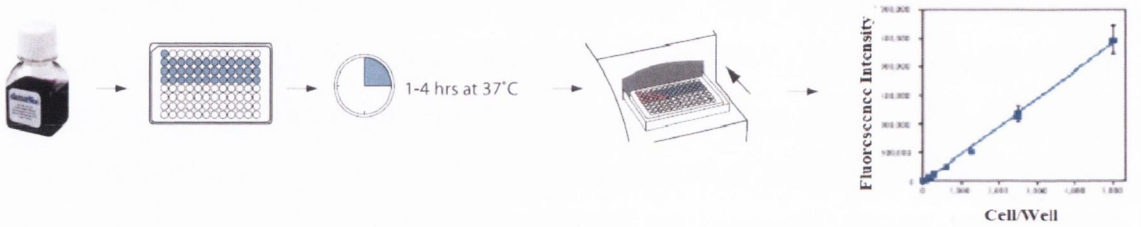


Figure 4.3. AlamarBlue® viability assay protocol.[17]

Cells are seeded at a specific density on 96-well plates – the day before the experiments for adherent cells and the same day for suspension cells. They are then treated with the required concentration of each compound and left to incubate at 37 °C. This incubation time varies depending on the experiment, but normally lasts for 72 h. After this, AlamarBlue® is directly added to each well and left to incubate in darkness for ~ 4-5 h. During this time, living cells metabolise the dye and convert resazurin to resorufin. This conversion, and the extent of the resulting fluorescence, is dependent on compound activity, so lower fluorescence means lower numbers of living cells and, thus, the high cytotoxicity, and *vice versa*. When there is a high number of living cells, the colour of the dye is pink because the cells have metabolised resazurin; however, when there are not many living cells left due to the cytotoxic action of a compound, the dye is blue, corresponding to large amounts of resazurin. This effect can be seen in the plate presented in Figure 4.4, where there is an increase of cytotoxic compound concentration from bottom to top.

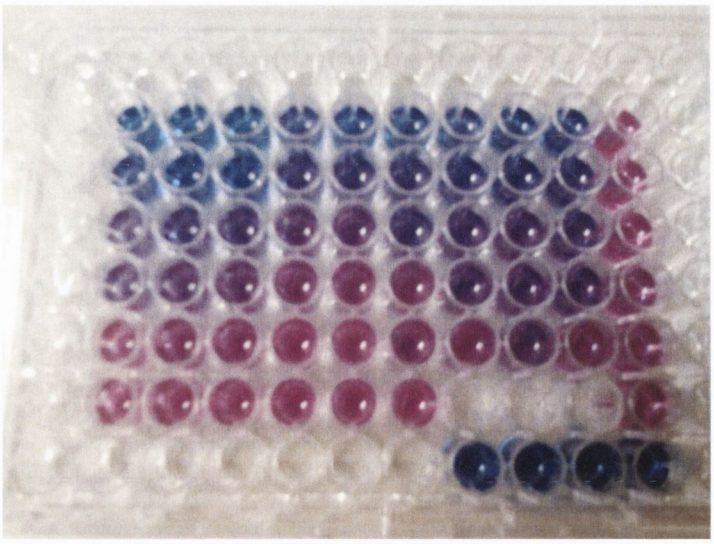


Figure 4.4. A 96 well plate where an increase of compound concentration/activity from bottom to top is observed due to the colour change of the AlamarBlue® dye.

The fluorescence is measured at a specific wavelength (excitation wavelength of 544 nm, emission wavelength of 590 nm). Results are then analysed by plotting obtained fluorescence values against compound concentration. Further analysis of results also allows for the determination of the IC_{50} (the concentration required to inhibit 50% of the cell population) of each compound in the particular cell line.

Thus, AlamarBlue® and MTT viability assays were selected to analyse the effect of our compounds on the cell viability and proliferation of a wide range of cancerous cell lines such as leukaemia, liver carcinoma, colorectal cancer and breast cancer cell lines. Consequently, IC_{50} values were calculated and, hence, a system of identifying active derivatives which could then be analysed by more advanced techniques for apoptosis evaluation and specific protein kinase inhibition was established.

4.2.1. Evaluation of the Cytotoxicity of Compounds from Families 1-5 in the HL-60 Cell Line.

The human caucasian promyelocytic leukaemia (HL-60) cell line was originally derived in 1977 from a leukemic blood cell population of a 36-year old female diagnosed with acute promyelocytic leukaemia (AML).

The study of a promyelocytic leukaemia cell line is quite relevant in this work as there are plenty of examples in the literature of the effect of sorafenib on this cell line.[3, 18]

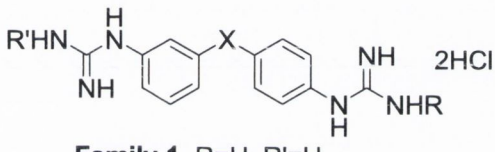
Some of the properties of HL-60 cells have made this cell line an attractive model for a range of studies. For instance, they are ideal for the assessment of cytotoxicity of anticancer drugs due to their non-adherent nature – which makes them easier to handle, they proliferate rapidly, and they are very robust – and because they are very sensitive to apoptosis inducers. For this reason, we chose them to screen the entire library of compounds synthesised in this project, allowing for comparison of different families to identify active structural features and for the selection of derivatives to be evaluated in more resistant and advanced cancer cell lines.

AlamarBlue® viability assays were carried out in the five synthesised families of 3,4'-guanidine-like derivatives (see experimental section 7.2.3 for protocol). The results obtained for Family 1 (Table 4.1) indicate that these compounds have moderate activity in HL-60 cancer cells. Both the -O- and -CH_2- linker derivatives (compounds **51a** and **51c**) show the highest cytotoxicity of the family, followed by the -NH- linker (**51d**) with nearly half of their activity and finally the -CO- linked compound (**51b**). This seems to indicate that the π-electron withdrawing effect of this linker is not favorable for the cytotoxicity of the compound. Cytotoxic evaluation of the 4,4'-guanidinium (**49a-d**) and 2-aminoimidazolinium (**50a-d**) had already been performed in Rozas' group and the IC_{50} values obtained for all of them were > 100 μM. Therefore, the 3,4'- substitution exhibit better cytotoxic results than the 4,4'- substitution of the aryl rings.

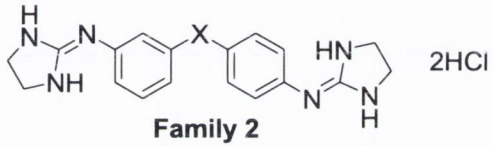
The next group to be evaluated was the 2-aminoimidazoline derivatives (Family 2), whose IC_{50} values were all > 100 μM, indicating that this family was inactive (Table 4.1); hence, we can conclude that the guanidinium moiety is more effective than the 2-aminoimidazolininium group in terms of inducing cytotoxicity in cells. Additionally, introducing a phenyl group into the guanidinium functionality in *para* or in *meta* position

(Family 3A and 3B) does not improve the activity relative to the unsubstituted guanidines of Family 1. In fact, the IC_{50} values were much larger for Family 3 (Table 4.1), showing that this phenyl substituted guanidine moiety is less effective than the unsubstituted guanidine. Furthermore, the variations to the linker do not confer any marked improvement to the activity of compounds from Families 2 and 3.

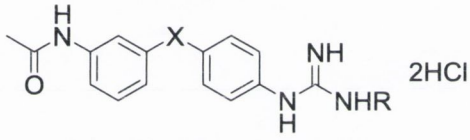
Table 4.1. The effect of compounds from Family 1-5 and sorafenib on the AlamarBlue® viability assay of HL-60 cells.



Family 1, R=H, R'=H
Family 3A, R=Ph, R'=H,
Family 3B, R=H, R'=Ph
Family 4A, R=(4-Cl-3-CF_3)Ph, R'=H
Family 4B, R=H, R'=(4-Cl-3-CF_3)Ph



Family 2



Family 5, R=(4-Cl-3-CF_3)Ph

Family	Compound	X linker	IC_{50} ± SEM μM (72h)
1	**51a**	O	**36.2** ± 1.1
1	**51b**	CO	**141.4** ± 2.5
1	**51c**	CH_2	**35.1** ± 1.2
1	**51d**	NH	**67.7** ± 1.8
2	**52a**	O	>100
2	**52b**	CO	>100
2	**52c**	CH_2	>100
3A	**70a**	O	>100
3A	**70b**	CO	>100
3A	**70c**	CH_2	>100
3A	**70d**	NH	>100
3B	**72a**	O	>100
3B	**72c**	CH_2	>100

4A	**71a**	O	**24.5** ± 1.2
4A	**71b**	CO	**22.3** ± 0.6
4A	**71c**	CH_2	**15.9** ± 0.5
4A	**71d**	NH	**36.6** ± 2.1
4B	**73a**	O	**9.72** ± 0.9
4B	**73c**	CH_2	**16.6** ± 1.6
5	**79a**	O	**9.11** ± 0.1
5	**79d**	NH	**3.96** ± 0.2
		Sorafenib	**1.68** ± 0.2

Note: Cells were seeded at a density of 2 x 10^5 cells/ mL in a 96-well plate and treated with the compounds at 0.5 μM, 1 μM, 5 μM, 10 μM, 25 μM, 50 μM and 100 μM. Sorafenib was used as a reference and tested in the same manner. Once treated, cells were incubated for 72 hours at 37 °C after which they were treated with AlamarBlue® and left in darkness in an incubator for 5 h. The resulting fluorescence was read using a plate reader from which percentage viability was calculated. IC_{50} values were calculated using Prism GraphPad 5 software from at least three independent experiments performed in triplicate.

However, when a 4-chloro-3-trifluoromethyl-phenyl group is introduced to one of the guanidium cations (Family 4), the cytotoxic activity of these compounds sharply increases (Table 4.1). The IC_{50} values of this family are ~16-37 μM, probably due to the fact that there is a lipophilic pocket in the active site of the kinase, to which the 4-chloro-3-trifluoromethylphenyl group can bind. The different X linkers seem not to influence the cytotoxic activity of these compounds significantly.

When the 4-chloro-3-trifluoromethyl phenyl moiety is incorporated in the *meta* position, instead of the *para* position, the activity of the drugs improves further. In the case of the -O- linked compound with this characteristic moiety placed in the *meta* position, the IC_{50} drops from 24.5 (**71a**) to 9.72 (**73a**) μM. This may be due to the fact that the inverse compounds place themselves in a different area of the binding site and that the interactions with these amino acids in the active site are more favoured. Conversely, in the case of the -CH_2- linked derivative, no significant difference in the cytotoxicity between the two analogues **71c** (15.9 μM) and **73c** (16.6 μM) was observed.

Compounds in Family 5, which are similar to Family 4 but with an acetamide group in the 3-position of the diaromatic system, showed an increase in activity relative to Family 4 (Table 4.1), with much lower IC_{50} values: 9.11 and 3.96 μM, compare to 24.3 and 36.6 μM for Family 4, (-O- and -NH- linked derivatives, respectively). This structural change seems

to be essential to enhance the cytotoxicity of these compounds in HL-60 cells. One explanation is that the formation of a HB between a HB donating amino acid of the active site and the carbonyl group of the acetamide could be important for activity. Additionally, it is worth mentioning that in Family 5 the -NH- linked derivative resulted in lower IC_{50} values than the -O- linker, the opposite have been observed in Family 4.

As a positive control, the AlamarBlue® viability assay of sorafenib was also carried out. The IC_{50} was quite low (1.68 μM) as expected; however, it was encouraging to see that some of our compounds have IC_{50} values within one order of magnitude of that of sorafenib, a compound that has been through several structural optimisations during development.

A graph representing the viability results of the most active compounds of all families and sorafenib is shown in Figure 4.5.

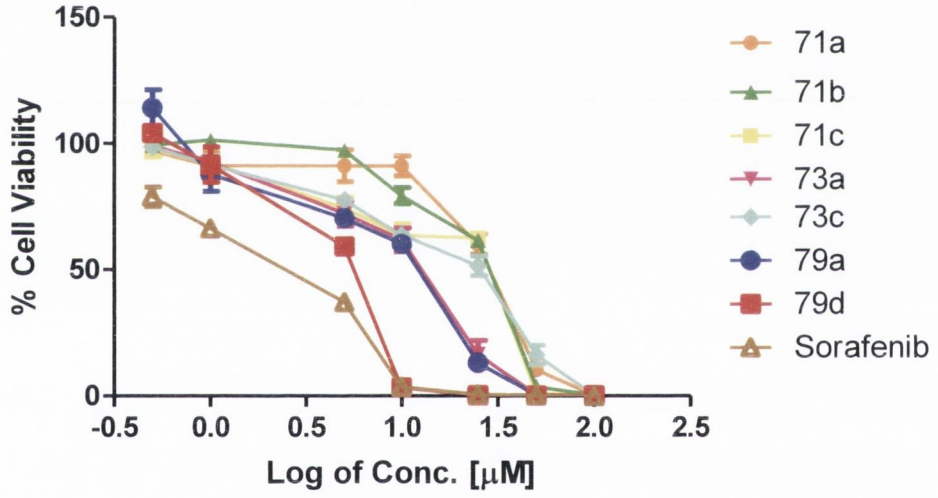


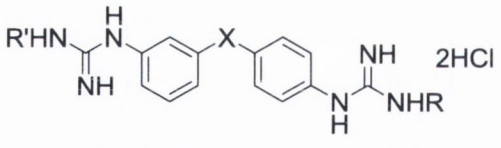
Figure 4.5. Compounds of Families 4A, 4B, 5 and sorafenib reduce the viability of HL-60 cells in a dose-dependent manner (same experimental details as described in legend from Table 4.1).

Furthermore, several SAR conclusions can be drawn from the evaluation of the cytotoxicity of all families in the HL-60 cancerous cell line: (i) 3,4'- substitution of the aryl rings exhibited better cytotoxic results than the 4,4'- substitution, (ii) the guanidinium moiety as in Family 1 showed better results than the 2-aminoimidazolinium as present in

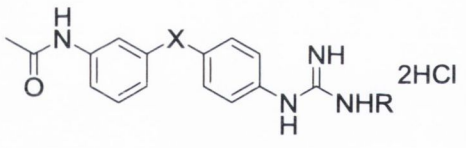
Family 2, (iii) the (4-Cl-3-CF_3)-Ph substituent present in Family 4 was crucial to achieve high cytotoxic activity; as evidenced by the unsubstituted phenyl guanidiniums in Family 3, in which the cytotoxic activity dramatically decreased, (iv) the unsubstituted guanidinium derivatives from Family 1 showed better cytotoxicity than the phenyl-substituted derivatives of Family 3, (v) placement of the (4-Cl-3-CF_3)-Ph moiety in the *meta* instead of the *para* guanidine increased the cytotoxic activity considerably, (vi) hence, the most favourable situation of guanidium substitution and position to increase cytotoxicity was *meta*-(4-Cl-3-CF_3)-Ph > *para*-(4-Cl-3-CF_3)-Ph > unsubstituted > phenyl, (vii) the acetamide group in Family 5 enhanced the cytotoxicity relative to the guanidinium moiety of Family 4, (viii) the linker between the two aryl groups appeared to be unimportant for the cytotoxic activity of these derivatives.

Due to the promising cytotoxicity results obtained with some of the derivatives at 72 h, such as **71c**, **73a**, **79a** and **79d**, we decided to investigate their effect at different time points, such as 24 h. Results are shown in Table 4.2.

Table 4.2. The effect of compounds **71c**, **73a**, **79a**, **79d** and sorafenib on the viability of HL-60 cells after 24 h and 72 h treatment.



71c: X=CH_2, R=(4-Cl-3-CF_3)Ph, R'=H
73a: X=O, R=H, R'=(4-Cl-3-CF_3)Ph



79a: X=O, R=(4-Cl-3-CF_3)Ph
79d: X=NH, R=(4-Cl-3-CF_3)Ph

Compound	IC_{50} ± SEM μM (24 h)	IC_{50} ± SEM μM (72 h)
71c	**37.7** ± 0.2	**15.9** ± 0.5
73a	**31.2** ± 1.3	**9.72** ± 0.9
79a	**10.5** ± 0.2	**9.11** ± 0.1
79d	**4.50** ± 0.2	**3.96** ± 0.2
Sorafenib	**1.70** ± 0.2	**1.69** ± 0.2

Note: Same experimental details as described in legend from Table 4.1.

The IC_{50} values obtained for **71c** and **73a** were much higher at 24 h than at 72 h, which shows that they need at least 72 h to exert their full cytotoxic activity. However, **79a** and **79d** showed similar activity at both time points, which is a significant result. We can see that in the case of **71c** and **73a** the IC_{50} value is nearly three times higher than at 72 h treatment. However, the difference was not as extensive in the **79a** and **79d** derivatives; obtaining nearly the same IC_{50} value at 72 h and at 24 h. These results were even more encouraging as they show that these compounds kill cancer cells very efficiently, even after 24 h treatment. Sorafenib was also tested at the 24 h time point, with nearly the same IC_{50} value being measured as had been at 72 h.

As previously mentioned at the beginning of this section, the HL-60 cell line was chosen to screen the entire library of compounds and identify the most potent candidates for further analysis in other more robust solid tumour cancer cell lines. The most potent families of compounds were Families 1, 4 and 5, therefore compounds **51a**, **71c**, **73a** and **79d** were selected for further analysis, as they were the most active members of these families. In this manner, a broad spectrum of structurally diverse compounds was analysed, in order to obtain as much relevant information as possible.

4.2.2. Evaluation of the Cytotoxicity of Selected Compounds in the HUH-7 Cell Line.

Considering that one of the main indications of sorafenib is the treatment of HCC, the cytotoxicity of the best compounds from the HL-60 cell screening were evaluated in the hepatocellular carcinoma cell line HUH-7.[19]

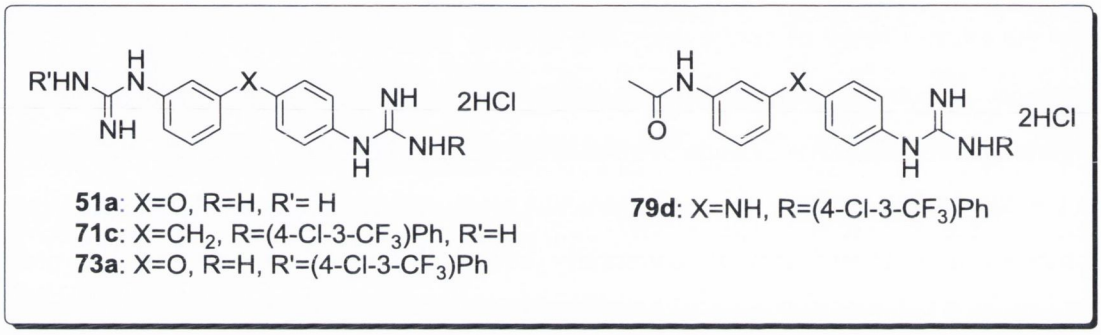
The HUH-7 cell line was originally taken from a liver tumour in a 57-year-old Japanese male in 1982. It is an adherent cell line containing the typical human plasma proteins of a liver cell line. As mentioned earlier, this cell type has been used previously to evaluate the cytotoxicity of sorafenib,[4, 5, 20] which is now used for the treatment of liver carcinoma. Then we have evaluated the cytotoxic activity in the more robust HUH-7 cell line of the compounds which were more potent in the HL-60 cell line testing.

AlamarBlue® viability assays of adherent cells require few modifications in comparison with the suspension cells. As described in the experimental section 7.2.3, HUH-7 cells were firstly trypsinised in order to dissociate them from their vessel and seeded in the 96

well plate the day before treatment with the drugs to allow attachment to the plate. The rest of the procedure was the same for both cell lines.

The results obtained are shown in Table 4.3, where the IC_{50} values of **51a**, **71c**, **73a** and **79d** were 4.43, 9.26, 10.8, 1.64 μM, respectively. As a positive control the testing of sorafenib was also carried out and an IC_{50} value of 1.26 μM was determined.

Table 4.3. The effect of compounds **51a**, **71c**, **73a** and **79d** and sorafenib on the AlamarBlue® viability assay of HUH-7 cells.



Compound	IC_{50} ± SEM μM (72h)
51a	**4.43** ± 0.5
71c	**9.26** ± 0.6
73a	**10.8** ± 1.1
79d	**1.64** ± 0.2
Sorafenib	**1.26** ± 0.1

Note: Cells were seeded at a density of 5 x 10^4 cells/ mL in a 96-well plate and incubated at 37 °C for 24 h to led them attach. The cells were then treated with the compounds at 0.25 μM, 0.5 μM, 1 μM, 5 μM, 10 μM, 25 μM, 50 μM and 100 μM. Sorafenib was used as a reference and tested in the same manner. Once treated, cells were incubated for 72 hours at 37 °C after which they were treated with AlamarBlue® and left in darkness in an incubator for 5 h. The resulting fluorescence was read using a plate reader from which percentage viability was calculated. IC_{50} values were calculated using Prism GraphPad 5 software from at least three independent experiments performed in triplicate.

Analysis of the above results allows a number of observations to be made. Compound **79d** continues to be the most potent analogue of all, suggesting that the acetamide group in the *meta* position relative to the linker is more favourable than a guanidinium group. The cytotoxic potency of **79d** is very similar to that of sorafenib, and both compounds share a

number of similar features: (i) both have a HB acceptor substituent in the aryl ring in *meta* position, (ii) regarding the linker the -NH- linker of **79d** and the -O- linker of sorafenib both have the ability to form HBs, (iii) both have the (3-Cl-4-CF_3)-Ph subunit, which has been proven to be of paramount importance. All of these similarities might lead these two molecules to orientate in a similar fashion in the binding site of the target.

Also, in the HUH-7 cell line compound **51a** – whose structure is based on two unsubstituted guanidinium groups – exhibit lower IC_{50} values than compounds **71c** and **73a**, which have the (3-Cl-4-CF_3)-Ph moiety. The opposite result had been obtained in the HL-60 cell line; hence compound **51a** displays a considerably different spectrum of activity in the HUH-7 cell line. Overall, the activity of **51a**, **71c**, **73a** and **79d** was still very satisfactory, their IC_{50} values being around 10 μM in HUH-7 cells.

The dose-response curve of all compounds is illustrated in Figure 4.6.

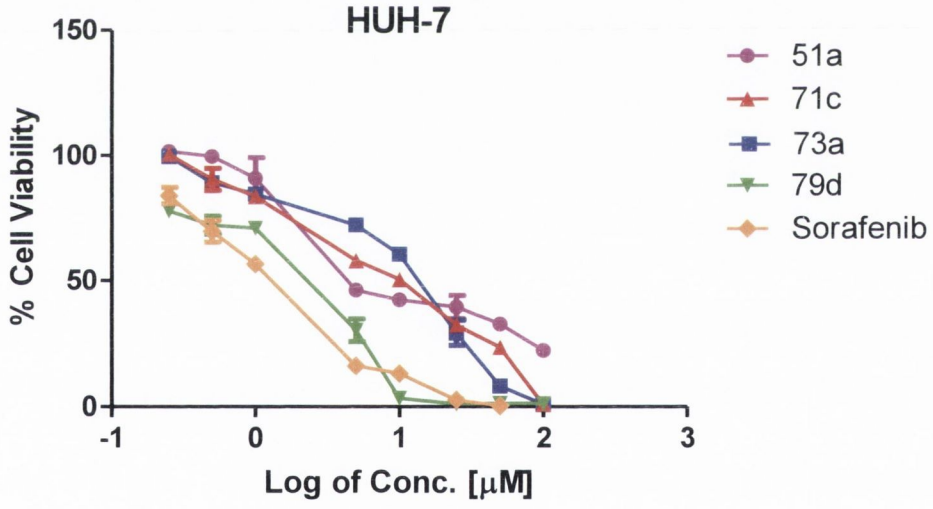


Figure 4.6. Compounds **51a**, **71c**, **73a** and **79d** and sorafenib reduce the viability of HUH-7 cells in a dose-dependent manner (same experimental details as described in legend from Table 4.3).

4.2.3. Evaluation of the Cytotoxicity of Selected Compounds in the RKO Cell Line.

In addition to evaluating the activity of the most potent derivatives from the HL-60 screening on the HUH-7 cell line, the potential anticancer effects on another adherent cell line, such as the colorectal cancer RKO cell line was evaluated.

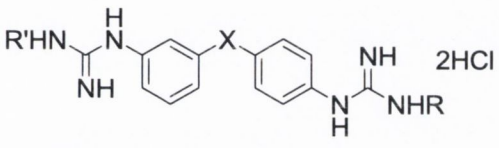
This cell line consists of a poorly differentiated colon carcinoma cell line developed by Michael Brattain.[21] RKO cells are adherent cells and contain wild-type p53 but lack the endogenous human thyroid nuclear receptor.[22] Sorafenib is successfully used for the treatment of colorectal cancer in combination with other drugs such us 5-fluorouracil or rapamycin, and also in combination with radiation therapy;[10,11,23] therefore, we tested the efficacy of our compounds (**51a**, **71c**, **73a** and **79d**) in the RKO colorectal cancer cell line, to compare the biological profiles.

Mutations in the gene encoding the B-RAF kinase are present in 70% of primary melanomas, in 10% of all colon cancers and in 30-70% of papillary thyroid carcinomas.[24] Among the B-RAF mutations observed in melanoma, over 90% are at codon 600, and among these, over 90% are a single nucleotide mutation resulting in substitution of glutamic acid for valine, ^{V600E} B-RAF mutation.[25] These mutations result in constitutive activation of downstream signalling *via* the mitogen activated protein kinase (MAPK) pathway, resulting in uncontrolled proliferation of the mutated cancer cells.[26]

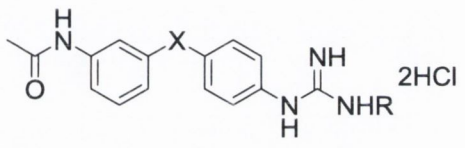
The identification of a frequent kinase mutation in melanoma provided a promising therapeutic target in a disease associated with poor survival.[26] A kinase inhibitor targeted specifically towards the ^{V600E} B-RAF gene product would potentially suppress tumour growth, while sparing the important biological functions mediated by wild-type B-RAF kinase. For this reason, the cytotoxicity of our compounds was evaluated in three types of RKO colorectal cancer cell lines: the dominant mutant F-6-8 cell line (which have the ^{V600E} B-RAF mutant allele and a wild-type allele), the A19 isogenic clone (containing the mutant allele only), and the T29 isogenic clone (containing the wild type allele only). The isogenic clones were formed by homologous recombination. There are examples in the literature which claim that sorafenib induces apoptosis in NPA melanoma cells expressing oncogenic ^{V600E} B-RAF;[27] hence, it would be interesting to investigate if sorafenib and our compounds reduce the cell viability in these three different RKO cell lines.

The viability assays in the RKO cell lines were carried out using MTT assays (this protocol is described in section 7.2.3) by the group of Dr. S. Van Schaeybroeck in the Centre for Cancer Research and Cell Biology (CCRCB) at Queen's University Belfast (UK) and the results obtained are shown in Table 4.4.

Table 4.4. The effect of compounds **51a**, **71c**, **73a**, **79d** and sorafenib on viability assay of RKO cells.



51a: X=O, R=H, R'= H
71c: X=CH_2, R=(4-Cl-3-CF_3)Ph, R'=H
73a: X=O, R=H, R'=(4-Cl-3-CF_3)Ph



79d: X=NH, R=(4-Cl-3-CF_3)Ph,

Compound	IC_{50} T-29 (WT/null)	IC_{50} F-6-8 (WT/MT)	IC_{50} A-19 (MT/null)
51a	**5.1** ± 0.4	**9.7** ± 0.5	**7.6** ± 0.5
71c	**4.4** ± 0.6	**7.6** ± 0.6	**8.8** ± 0.7
73a	**0.9** ± 0.1	**2.6** ± 0.1	**1.9** ± 0.2
79d	**4.1** ± 0.1	**6.6** ± 0.5	**5.6** ± 0.5
Sorafenib	**0.9** ± 0.0	**1.3** ± 0.1	**1.2** ± 0.1

Note: Cells were seeded at 3,000 cells/well in a 96-well plate and incubated at 37 °C for 24 h to led them attach. The cells were then treated with the compounds at 0.1 μM, 0.5 μM, 1 μM, 2.5 μM, 5 μM, 10 μM, 25 μM, 50 μM and 100 μM. Sorafenib was used as a reference and tested in the same manner. After 72 hours MTT was added to each well and cells were re-incubated for 2 hours. Following this all media was aspirated from each well and the cells were re-solubilised in MTT. Plates were read using a microplate reader. IC_{50} values were calculated using Prism GraphPad 5 software from at least three independent experiments performed in triplicate.

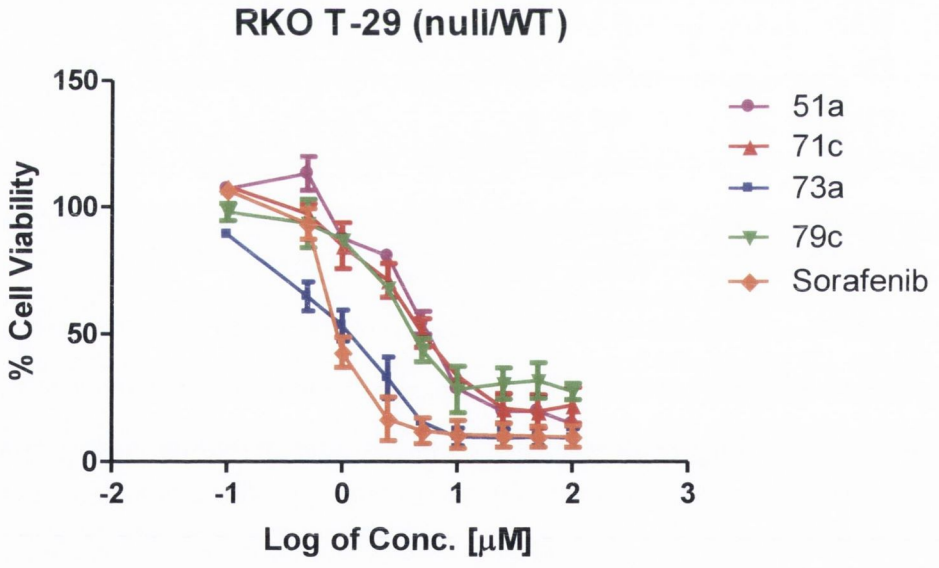
Regarding the IC_{50} values obtained in the wild type cell line T-29, the cytotoxic potency of **73a** was very similar to that of sorafenib; hence, its capacity to reduce cell viability in the wild type RKO colorectal cancer cell line was almost identical. Then, the next most active compounds were **71c** and **79d**, which have comparable cytotoxic activity, followed by compound **51a**.

When analysing the cytotoxicity of compounds in the F-6-8 cell line (Table 4.4), which has the ^{V600E} B-RAF mutant allele and a wild-type allele (dominant mutant), it was observed that the IC_{50} value of sorafenib differs more from **73a** and also all the IC_{50} values were higher than for the wild type T-29 cell line. However, the results obtained for our compounds were very satisfactory.

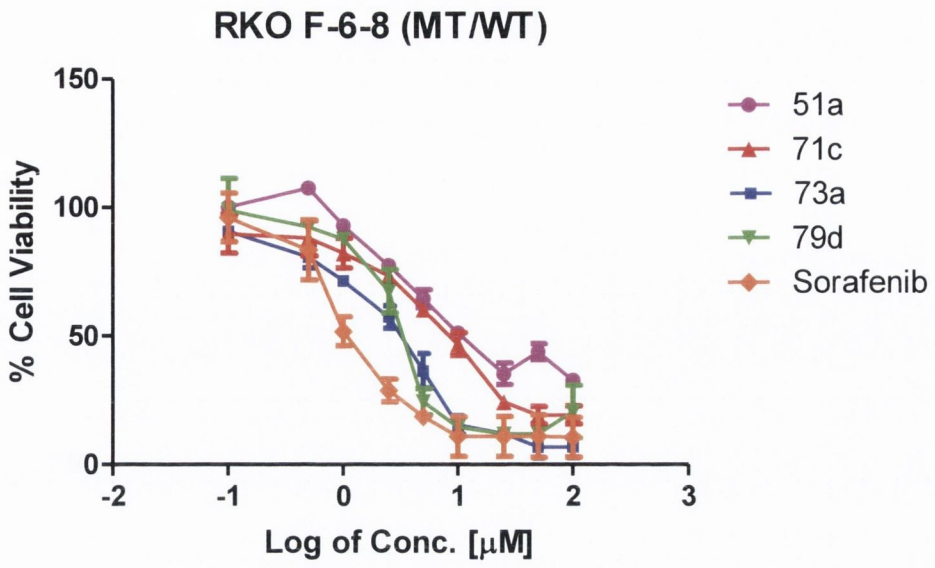
For the A-19 cells, which contain the ^{V600E} B-RAF mutant allele only, the results (Table 4.4) are comparable to the results from the other mutated cell line (F-6-8); however, greater cytotoxic activity was obtained in the wild type T-29 cell line. This pattern is also observed with sorafenib, where the IC_{50} value in this cell line was in the same range as in the other mutated cell line (F-6-8), but the differences with the wild type cell line (T-29) are not as extensive as with our compounds.

The dose-response curves of the compounds in all the colorectal cancer cell lines T-29, F-6-8 and A-19 are illustrated in Figure 4.7.

A



B



C

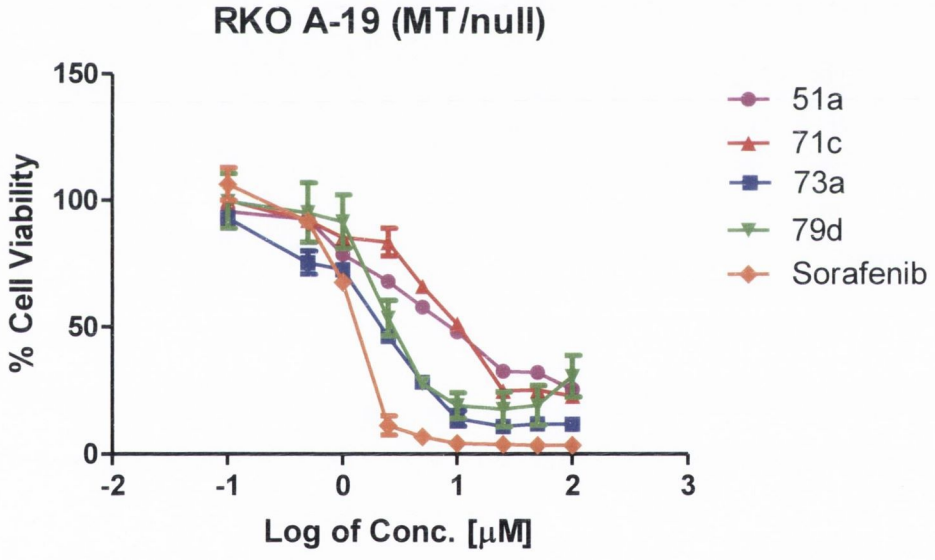


Figure 4.7. Compounds **51a**, **71c**, **73a**, **79d** and sorafenib reduce the viability of T-29 (A), F-6-8 (B) and A-19 (C) RKO cells respectively, in a dose-response manner (same experimental details as described in legend from Table 4.4).

Several conclusions can be drawn from the results obtained for the three RKO cell lines. The compounds inhibit cell proliferation in the T-29 wild type, in the A-19 mutant cell line and in the F-6-8 cell line, which have a mutant allele and a wild-type allele. This would suggest that the novel compounds would exert anti-proliferative activity in cancer cells exhibiting a mutation of ^{V600E} B-RAF, or exhibiting other types of mutation inherent in the wild type colorectal cancer cell line. However, from the results obtained we can see that our compounds do not have high selectivity for ^{V600E} B-RAF mutated cancer cells, as the IC_{50} values are higher than they are for the wild type cell line.

Analysing the IC_{50} values obtained for all the cancer cell lines, several patterns emerge. A slight selectivity for inhibition of the wild type T-29 was observed, in comparison to the other two mutant cell lines; this was the case for all synthesised derivatives, as well as sorafenib. Studies found in the literature can possibly justify these observations since the IC_{50} value of sorafenib for the inhibition of mutant ^{V599E} B-RAF was nearly 2-fold higher than the B-RAF wild type, showing higher selectivity for the wild type over the mutated B-RAF.[28]

Additionally, the potency of the compounds in the different cell lines followed a trend, where **73a** was always the most potent compound of all, with IC_{50} values comparable with – and sometimes slightly better than – sorafenib. Looking at the chemical structure of these two compounds, both have an oxygen linker between the two aryl groups; however the (4-Cl-3-CF_3)-Ph moiety of sorafenib is attached to the urea in the *para* position, while in **73a** it is attached to the guanidine in the *meta* position. Also, sorafenib has an amide group in the *meta* position, which will not be protonated at physiological pH (pK_{aH} -0.51), and **73a** has a guanidinium group in the *para* position, which is protonated at physiological pH (pK_{aH} 10.8). However, despite these structural differences, both molecules exert similar cytotoxicity in RKO cancer cells, and have broadly similar structural features.

The compound which displays the next highest cytotoxic activity after **73a** was always **79d**, followed by **71c** and **51a**, similar to what had been observed in the HL-60 cell line. Conversely, this trend was not followed in the HUH-7 cell line, and the least potent compound was actually **73a**. This is a demonstration of the fact that each cancer cell line has different genotypic and phenotypic characteristics; therefore, the activity of the compounds will not necessarily be maintained in different cell lines.

After the promising results obtained in all evaluated cell lines (HL-60, HUH-7 and RKO), we continue investigating different cancer cell lines to broaden our knowledge of the applicability of our compounds to different cancers.

4.2.4. Evaluation of the Cytotoxicity of Selected Compounds in the MCF-7 Cell Line.

We investigated next the effect of our compounds on cell proliferation in the most commonly occurring cancer in women at present, breast cancer; hence, the cytotoxicity evaluation of our derivatives was carried out in the MCF-7 cell line.

The MCF-7 (Michigan Cancer Foundation) cell line was originally derived from a pleural effusion of a 69-year old patient with metastatic breast cancer.[29] The cell line retains several characteristics of differentiated mammary epithelium, such as the ability to process estradiol. Consequently, a number of cytoplasmic receptors for estrogens, androgens, glucocorticoids and progestins have been identified in these cells.[30] As a result, the MCF-7 cell line is an excellent *in vitro* model of breast cancer and is useful for the study of tumour response to endocrine therapy, and of the nature of binding and biological functioning of these hormones. Modulation of the estrogen and progestin receptors, by the cytokine TNF, has a prominent effect on estradiol stimulation of MCF-7 cells, blocking its proliferative response and enhancing certain metabolic effects.[31]

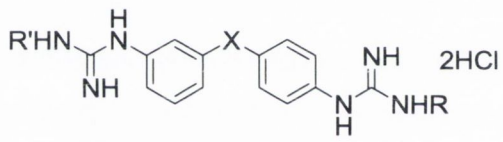
There are examples in the literature which demonstrate the cytotoxic effect of sorafenib on the MCF-7 cell line on its own or in combination with other therapies such as epirubicin or desatinib.[6,7,8] Considering the reported IC_{50} values of sorafenib in the MCF-7 cell line at 24, 48 and 72 h are 7.85 ± 0.86 μM, 3.45 ± 0.16 μM and 2.10 ± 0.64 μM, respectively,[6] we evaluated the cytotoxic activity of our novel analogues in the MCF-7 cell line in collaboration with the group of Dr. J. Gee at the Cardiff School of Pharmacy and Pharmaceutical Sciences (UK).

One of the most renowned treatments for breast cancer is tamoxifen, which has provided benefit for millions of estrogen receptor alpha (ER^+) positive breast cancer patients worldwide. However, almost all patients with metastatic disease, and as many as 40% of patients receiving tamoxifen, will acquire resistance to the drug's inhibitory effects on breast cancer cell proliferation.[32] Initially it was thought that mutation of the ER was the cause of the resistance, but later the functionality of the ER was shown to be normal.[33]

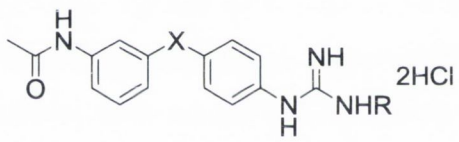
Subsequently, it was considered that a deregulation of growth factor pathways facilitates resistance proliferation, but this was disproven by the disappointing action of growth factor pathway blockade gefitinib, suggesting further key mechanisms underlie anti-hormone resistant growth.[34] Additionally, there is evidence to suggest that long term tamoxifen exposure induces genotypic changes which contribute to anti-hormone resistant cell proliferation.[35]

For this reason, drugs that induce cytotoxic effects in tamoxifen resistant MCF-7 cells provide a potential therapeutic avenue for the management of tamoxifen-resistant breast cancer and, hence, we have evaluated our compounds in the MCF-7 wild type cell line and in tamoxifen resistant MCF-7 cells (Tam-MCF-7). Following the protocol explained in the experimental section 7.2.3, the MTT viability assays were carried out in both cell lines, the results of which are shown in Table 4.5.

Table 4.5. The effect of compounds **51a**, **71c**, **73a**, **79d** and sorafenib on viability assay of MCF-7 and tamoxifen resistant MCF-7 (Tam-MCF-7) cells.



51a: X=O, R=H, R'= H
71c: X=CH_2, R=(3-Cl-4-CF_3)Ph, R'=H
73a: X=O, R=H, R'=(3-Cl-4-CF_3)Ph



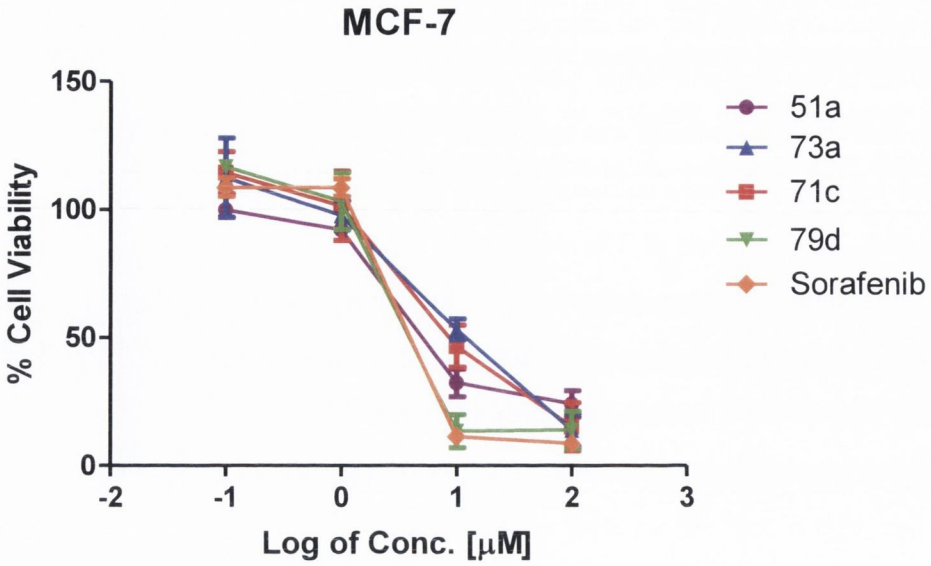
79d: X=NH, R=(3-Cl-4-CF_3)Ph,

Compound	IC_{50} MCF-7	IC_{50} Tam- MCF-7
51a	**5.20** ± 0.7	**5.48** ± 1.9
71c	**10.2** ± 4.3	**8.03** ± 0.9
73a	**11.8** ± 2.2	**28.7** ± 11.1
79d	**3.97** ± 0.8	**3.76** ± 0.4
Sorafenib	**4.02** ± 0.2	**4.64** ± 0.1

Note: Cells were seeded at 1 x 10^4 cells/well in a 96-well plate and incubated at 37 °C for 24 h to led them attach. The cells were then treated with the compounds at 0.1 μM, 1 μM, 10 μM, and 100 μM. Sorafenib was used as a reference and tested in the same manner. After 96 hours MTT was added to each well and cells were re-incubated for 4 hours. Following this all media was aspirated from each well and the cells were re-solubilised in MTT. Plates were read using a microplate reader. IC_{50} values were calculated using Prism GraphPad 5 software from at least three independent experiments performed in triplicate.

The IC_{50} values obtained for the assayed compounds (Table 5.5) were very satisfactory as they were all in the low μM range, very similar to, and in the case of **79d** even better than, sorafenib. Another observation was that nearly all of the compounds affect wild type MCF-7 and Tam-MCF-7 cell lines equipotently, as their IC_{50} values in both cell lines are almost identical with the exception of **73a** which showed greater activity in the wild type MCF-7 cell line. The same behavior was observed for sorafenib, whose IC_{50} values in both cell lines are nearly indistinguishable. The dose-response curves of the aforementioned compounds in both cell lines are presented in Figure 4.8.

A



B

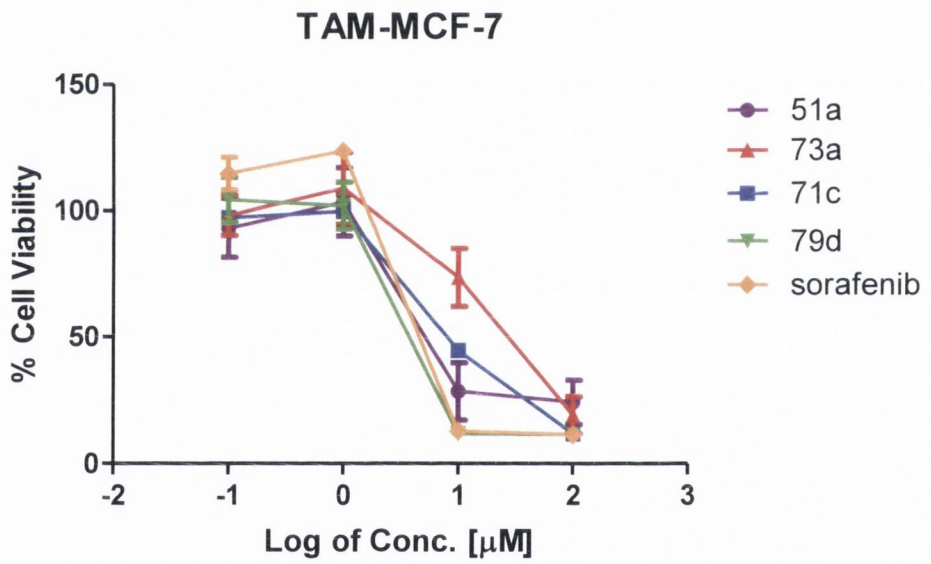


Figure 4.8. Compounds **51a**, **71c**, **73a**, **79d** and sorafenib reduce the cell viability in both MCF-7 (A) and tamoxifen-resistant MCF-7 cells (B) in a dose-dependent manner (same experimental details as described in legend from Table 4.5).

The results obtained show that our compounds are able to inhibit the cell viability of wild type and tamoxifen resistant MCF-7 cells. As a result, we can claim that the molecular target of our compounds is different to that of tamoxifen. Therefore, it is likely that our

compounds do not target the estrogen receptor (ER) directly; however, they may target another pathway which is independent of, but at the same time modifies, the activation of the ER. Considerable evidence has emerged to suggest that there are multiple regulatory interactions between the ER, growth factors and other kinase signalling pathways such as MAPK, p38 MAPK, PI3K and AKT. Hence, growth factor receptor pathways may up-regulate and stimulate growth independently of the ER, or could communicate *via* cross-talk with the ER and thereby affect cell growth and patterns of resistance.[36] This interrelation is illustrated in Figure 4.9.

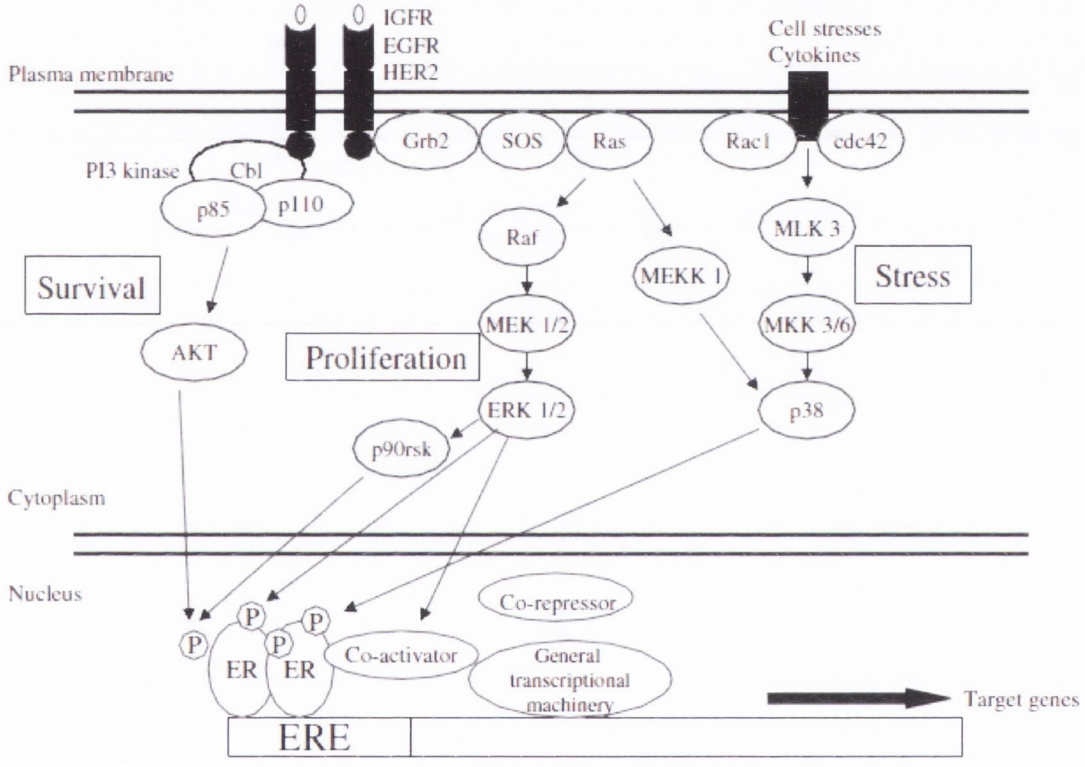


Figure 4.9. The cellular kinase pathways interacting with ER function *via* cross-talk.[36]

Numerous studies suggest that cross-talk occurs between the ER and growth factor receptor pathways, such as the EGFR/HER2 family and insulin-like growth factor receptor (IGFR) family. The ER can be phosphorylated by the MAPKs, ERK1 and ERK2, which are downstream components of the HER2 signalling pathway (Figure 4.9). This enhances the sensitivity of the ER to ligands and potentially leads to ligand-independent activation. Therefore, increased ERK activity could potentially contribute to resistance to endocrine

therapy. Indeed, ERK 1/2 expression and activity are increased in several breast cancer cell-line models of endocrine resistance, and increased ERK 1/2 activity correlates with shorter duration of response to endocrine therapy in clinical breast cancer.[37] It has been demonstrated that 72% of breast tumours have evidence that the ERK 1/2 MAPK plays a role in circumvention of anti-hormonal response in breast cancer. Therefore, it is possible that up-regulation of tyrosine-kinase growth factor receptor pathways, in particular up-regulation of ERK 1/2 during tamoxifen treatment, may lead to loss of estrogen-dependence and tamoxifen resistance.

Taking the above into consideration, and knowing that our compounds down-regulate the MAPK pathway and decrease the expression of ERK 1/2 (see western blot in section 4.4.4), the reason our compounds induce cell death in the Tam-MCF-7 could be postulated upon. Our compounds could block activation of the ER by inhibiting the expression of ERK 1/2, which is one of the pathways that induces a loss in estrogen-dependant activation of the ER, leading to tamoxifen resistance; hence, our compounds could induce a down-regulation of the ER and, therefore, a reduction in cancer cell proliferation, overcoming the resistance. For this reason, due to the fact that our compounds induce cytotoxic effects in tamoxifen resistant MCF-7 cells, we could consider compounds of this type as potential therapeutic alternatives for the management of tamoxifen-resistant breast cancer.

In conclusion, evaluating the data obtained in all analysed cell lines, we can say that our compounds inhibit cell proliferation of leukaemia, liver carcinoma, colorectal cancer, ^{V599E} B-RAF mutated colorectal cancer, breast cancer and tamoxifen-resistant breast cancer cell lines, all in the low μM range, while in some cases their cytotoxic activity is comparable or even better than that of commercialised sorafenib.

4.3. Apoptosis Evaluation of the Novel Compounds

The evaluation of potential anticancer agents on particular cell lines using viability assays such as AlamarBlue® or MTT is a suitable means of determining the anti-proliferative effects of a particular compound. However, these assays do not provide information regarding cell-pathologic mechanisms and how these biochemical changes have been induced. The principal mode of inhibition of cell viability caused by anticancer therapies is *via* the induction of apoptosis,[38] but other possible routes are necrosis, autophagy and cell cycle arrest. The mode of cell death can depend upon the intensity and duration of the injurious stimuli. Thus, further investigation of the most active derivatives was required to determine whether the drugs induce cell cycle arrest, cell death or a combination of both. This may be achieved either by analysing cell morphology, by quantifying DNA content with respect to cell cycle stages using flow cytometry or by studying PARP cleavage (a hallmark of apoptosis) .

4.3.1. Cell Morphology Studies

As previously mentioned, biological processes have unique and distinct morphological characteristics; hence, cell morphology studies supply information to identify the different cell processes. For example, the geometry of healthy cells is normally round, the membrane is elastic and the distribution of the organelles is normal. This contrasts with necrotic cells, in which the organelles are characteristically swollen, there is early membrane damage with eventual loss of plasma membrane integrity and leakage of cytosol into the extracellular space. Despite early clumping, the nuclear chromatin undergoes lysis. In contrast, apoptotic cells are shrunken, develop blebs containing dense regions of cytoplasm, membrane integrity is not lost until cell death[39] and nuclear chromatin undergoes significant condensation and fragmentation, while the cytoplasm becomes divided to form apoptotic bodies containing organelles and/or nuclear debris. These morphological characteristics are illustrated in Figure 4.10.

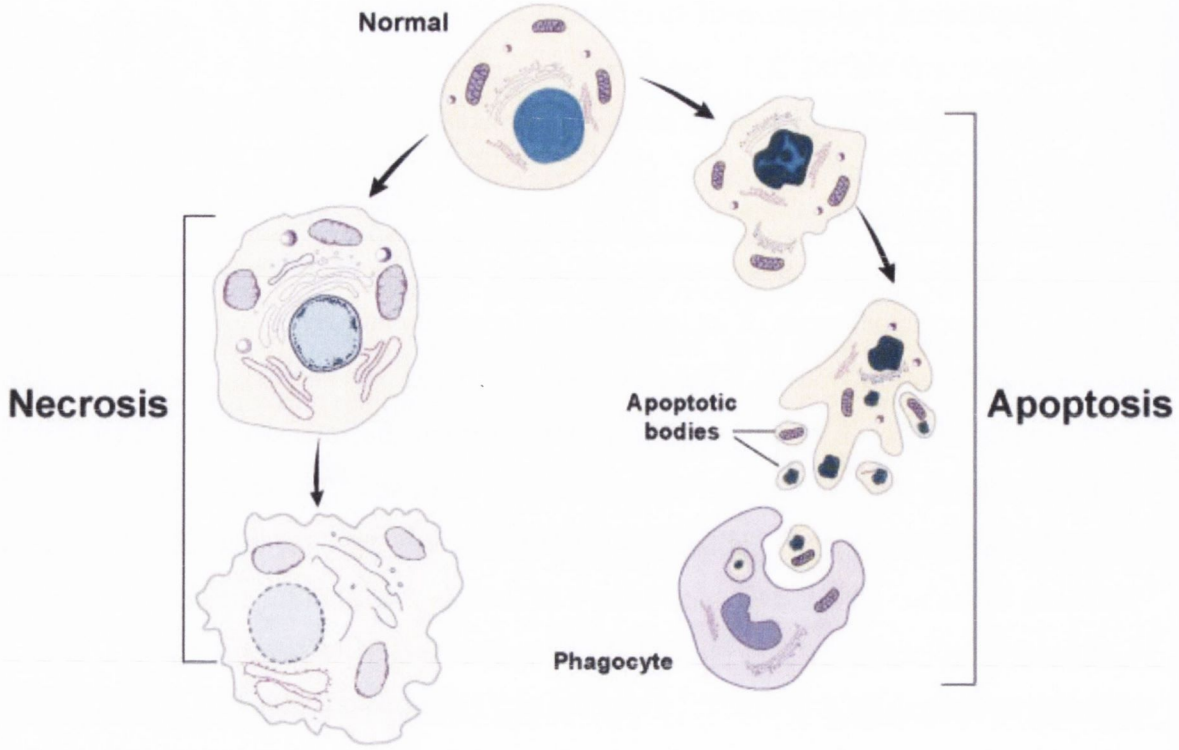


Figure 4.10. Morphological differences between healthy, apoptotic and necrotic cells.[39]

For many years, the study of apoptosis solely depended on the morphological features of the cell, observed using a light microscope. Therefore, although in modern times this technique may seldom be used for quantitative purposes, it remains highly useful to observe cell size, shape and texture which can identify morphological features characteristic of apoptotic processes.

Having investigated the library of compounds using viability assays, and having evaluated their effects on cell growth and proliferation, it was decided to confirm and substantiate these findings by consideration of their effects on cell morphology. A select group of active derivatives was chosen for investigation in the HL-60 suspension cell line, as well as in a representative of the adherent cell lines used, the HUH-7 cell line. These compounds were tested at specific concentrations and time-points as described below.

4.3.1.1. Morphological Studies of the HL-60 Cell Line

HL-60 cells are generally round or ovoid, having a modal cell diameter of 13 μm; however, there are considerable variations in cell diameter, ranging from 9 μm to 25 μm. The larger cells are normally binucleate.[40] Light and electron microscopy have established the cell morphology of the HL-60 cell line to be predominantly promyelocytic. This includes large, rounded, blast-like cells with large nuclei containing 2-4 nucleoli, a basophilic cytoplasm and prominent azurophilic granules.[40]

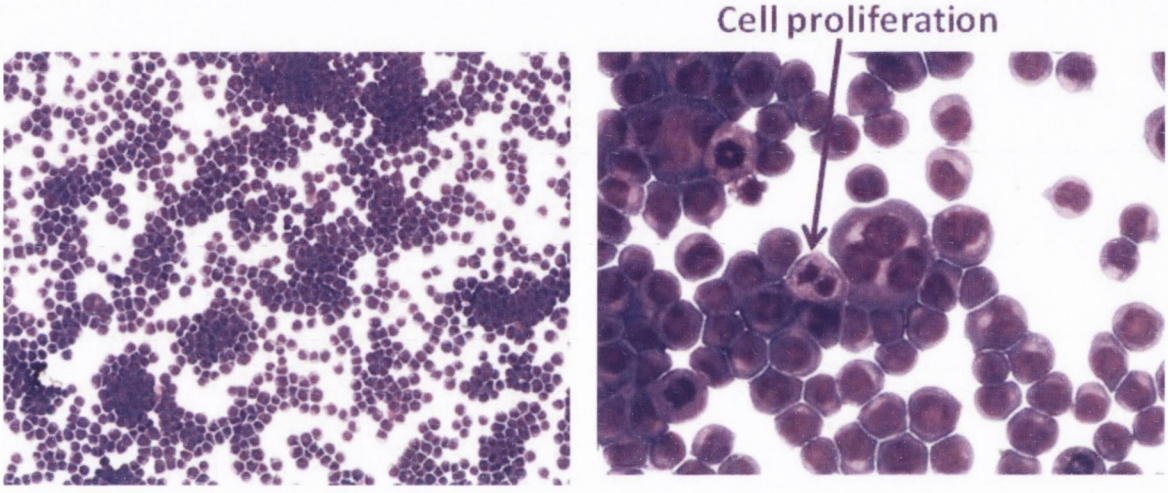


Figure 4.11. Typical HL-60 cell morphology (cells were seeded at a density of 2×10^5 cells/mL and they were fixed onto slides by cytocentrifugation using a Shandon Cytospin. This was followed by staining using the Rapidiff Kitt according to manufacturer's instructions. Images were taken at magnifications of 40 × (*right*) and 20 × (*left*) using an Olympus 1X81 microscope coupled to a CCD DP71 camera).

The cells continuously proliferate in suspension with a doubling time of approximately 36-48 hours, in comparison to other myeloid leukaemia cells which undergo a limited number of cell divisions prior to cell death.[41] An example of cell division of these cells can be seen in Figure 4.11 (right), where a cell has two nuclei because it is undergoing cell proliferation.

Based on previous cell viability results, **73a** and **79d** were chosen as suitable derivatives to study their effects on cell proliferation at 50 μM, over 72 h, as described in the experimental section 7.2.4.

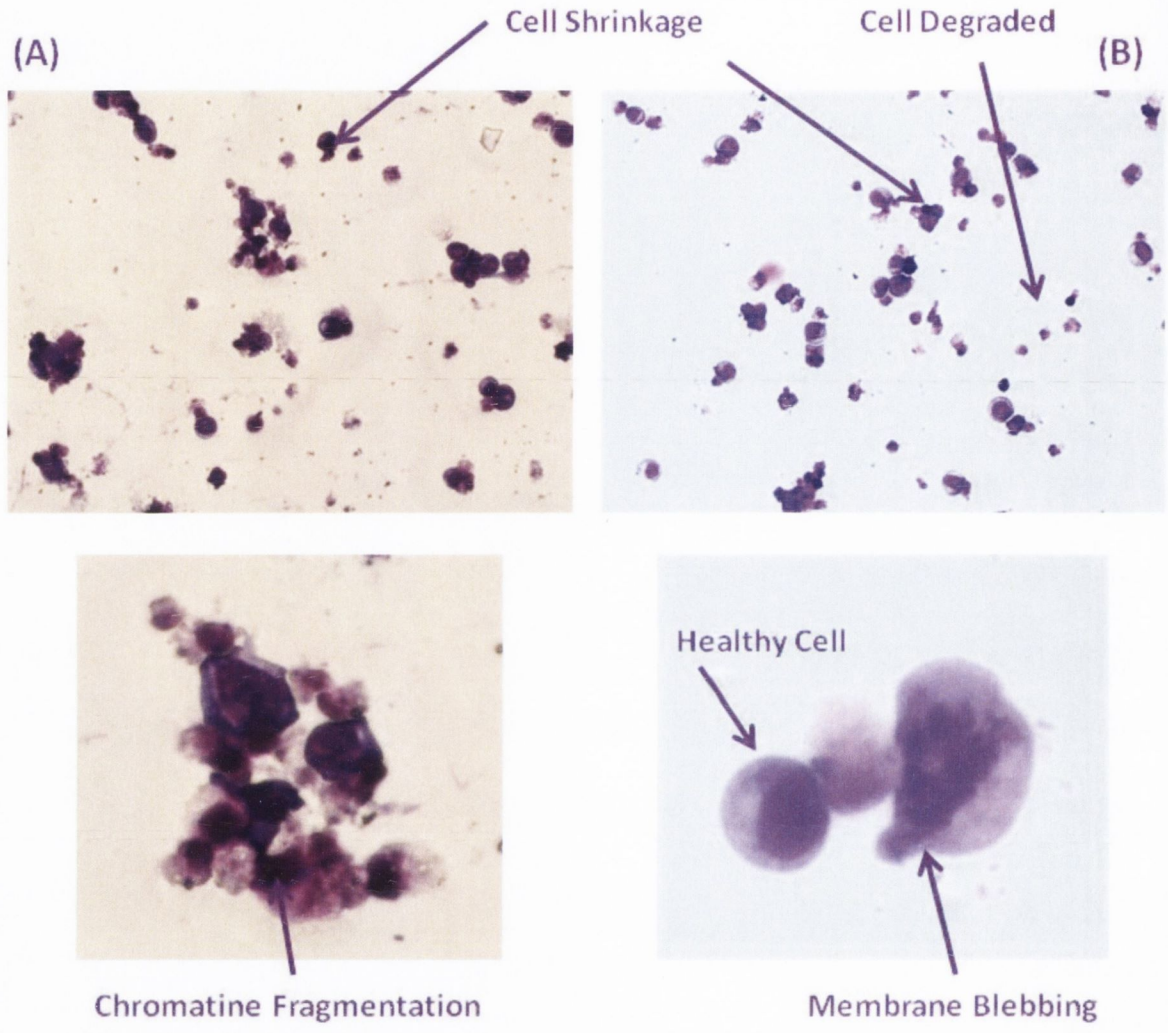


Figure 4.12. Effect of (A) **79d** and (B) **73a** on the morphology of HL-60 cells. The experimental details were the same as in Figure 4.11 but in this case the cells were treated with **73a** and **79d** (50 μM) and incubated at 37 °C for 72 h. Images taken at 40 × (*top*) and 20× (*bottom*) magnification.

We can see an obvious change in the morphology of the control vehicle (Figure 4.11) and the ones treated with **73a** and **79d** (Figure 4.12). The control cells are round with intact membrane, whereas the cells have evidently shrunk and degraded after treatment with our compounds, and they have undergone membrane blebbing. This is epitomized in the image on the bottom left of Figure 4.12, where the difference in morphology between a healthy cell and an apoptotic cell can be seen in detail.

Also, an overnight morphology study of compound **73a** in the HL-60 cell line was carried out. Cell proliferation and cell death were monitored by taking pictures every minute for 18 hours. The results are illustrated in Figure 4.13.

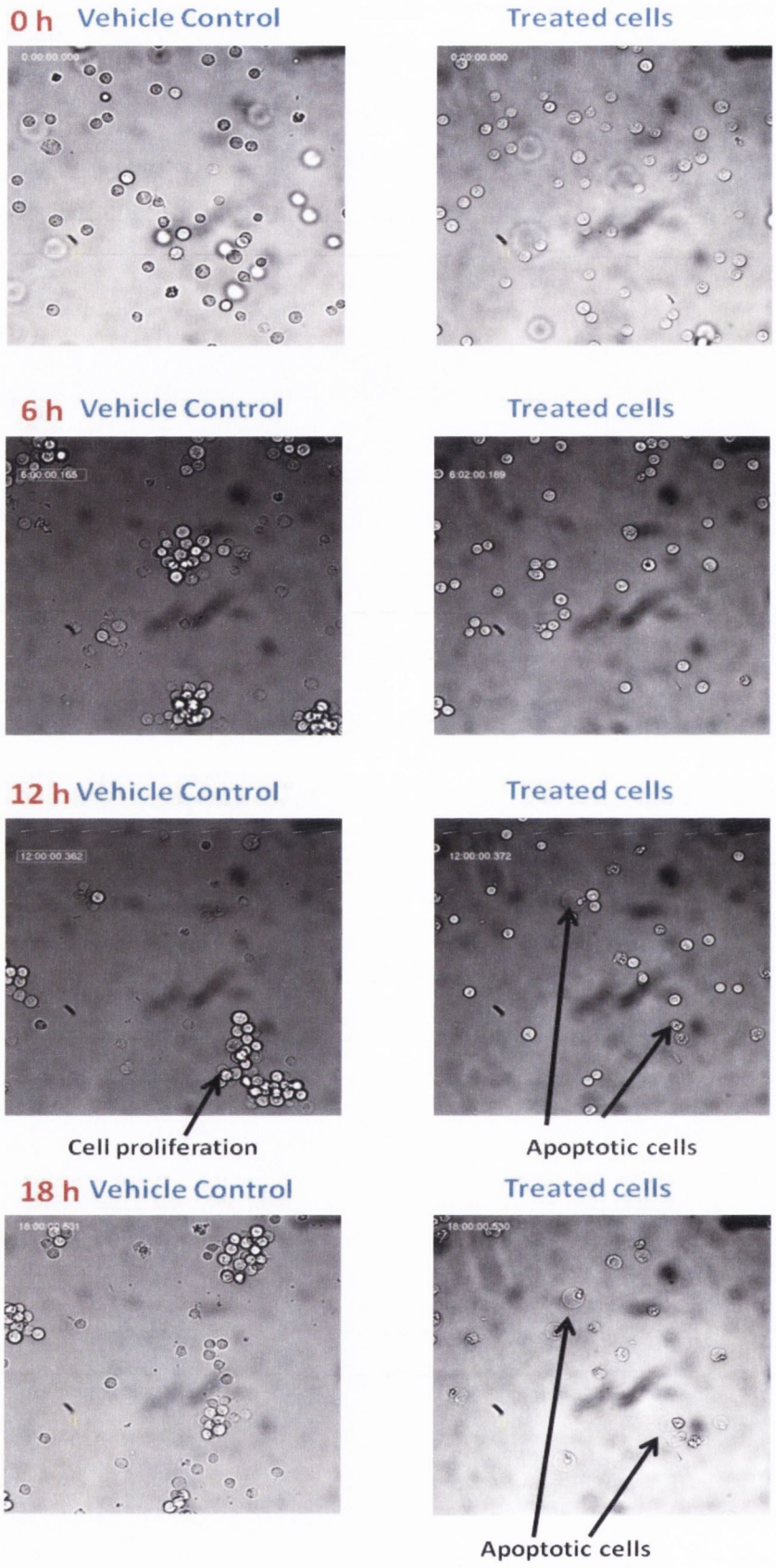


Figure 4.13. Comparison of the effect of no drug (vehicle control) and **73a** at 50 μM on the morphology of HL-60 cells.

Note: Cells were seeded at 5×10^5 cell/mL in two different flasks. One was treated with **73a** 50 μM and the other with no drug. Images of (*left*) vehicle control and (*right*) treated cells were taken at 0, 6, 12, 18 h after exposure of the drug using Olympus IX81 microscope with an Andor iXon CCD DP71 camera, running IQ software, in a humidified chamber at 37 ^{o}C and 5% CO_2 with 40 x oil objective.

We can see how, after 6 h, there is a reduction in the number of the cells treated with **73a** in comparison to the vehicle control whose cells are proliferating and forming groups (Figure 4.13), whereas the treated cells are not. A more significant change can be observed after 12 and 18 h of treatment, when apoptotic cells begin to appear. These cells undergo membrane blebbing and cellular shrinkage; however, the control cells are continuously proliferating at the same time points. This study confirms the ability of our compounds to induce cell death in the HL-60 cell line, in agreement with the previous cell viability studies of the derivatives in question. However, further biochemical assays such as flow cytometry would have to be carried out to confirm the induction of apoptosis by our compounds.

4.3.1.2. Morphological Studies of the HUH-7 Cell Line

After the promising results obtained for our synthesised compounds regarding effects on cell viability in solid tumour cell lines, their effect on the morphology of cells was studied in the HUH-7 cell line, as a representative of all of the adherent cell lines.

The HUH-7 cell line is a well differentiated hepatocyte derived cellular carcinoma cell line. It is an adherent cell line, which displays the typical epithelial feature of pavement-like cell arrangement and confluent monolayer.[42] The cells are distinguished from hepatocytes by their smaller size and architectural organization as irregular trabecular or pseudoglandular patterns (with pseudo granules in the cytoplasm).[43] These morphological characteristics can be seen in Figure 4.14.

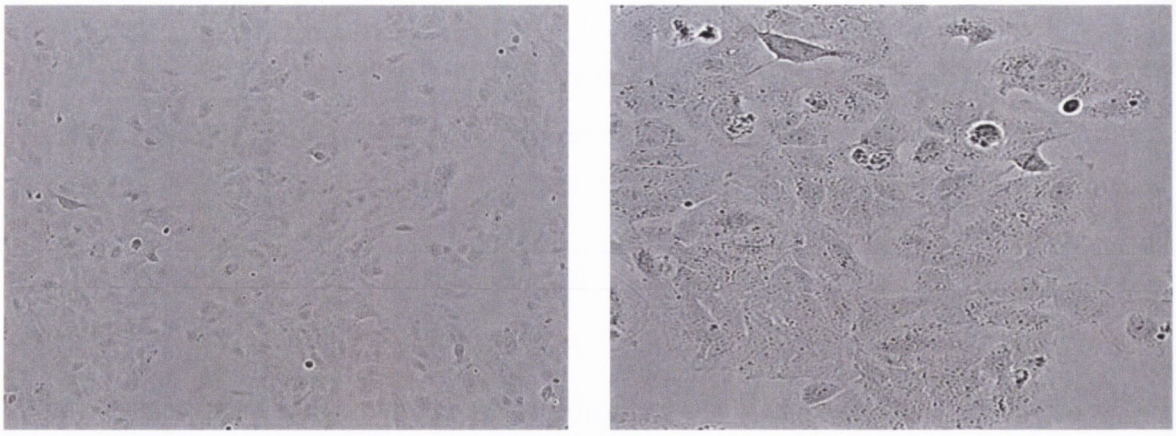


Figure 4.14. Typical HUH-7 cell morphology (Cells were seeded at 5 x 10^4 cells/mL at 37 ^oC for 24 h to let them attach). The cells were then treated with 2 μL of ddH_20 for vehicle control. After 72 h incubation, the cells were brought to the microscopy area to take images at 60 × magnification (*left*) and more detailed 20 × (*right*) in an Olympus 1X81 microscope coupled to a CCD DP71 camera using different magnifications).

However, when these cells were treated with compounds **73a** or **79d**, at 50 μM for 72 h, their morphology changed completely (Figure 4.15). Apoptotic cells detached from the bottom of the flask became round in shape, in comparison to the flat shape of the control cells. Additionally they suffer membrane blebbing, cell shrinkage and nuclear fragmentation was observed. The features are similar to those observed in the HL-60 cell line following treatment with **73a** and **79d**.

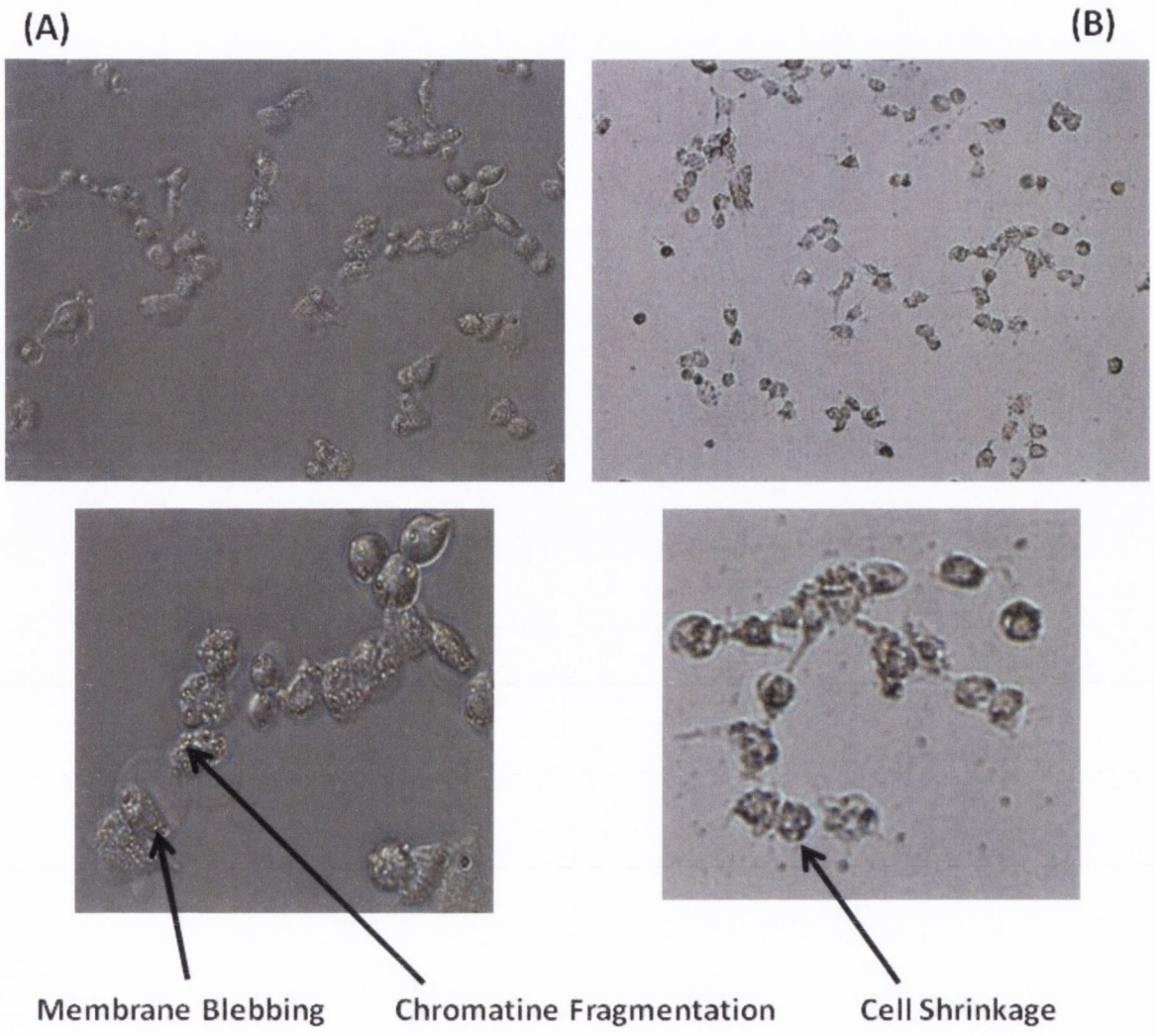


Figure 4.15. Effect of (A) **79d** and (B) **73a** on the morphology of HUH-7 cells (same experimental details as in Figure 4.14 but in this case the cells were treated with **79d** and **73a** (50 μM) and incubated at 37 °C for 72 h). Images were taken at 40 × (*top*) and 20 × (*bottom*) magnification.

These images suggest that **73a** and **79d** induce cell death in the hepatocellular carcinoma cell line, which correlates with the cytotoxicity observed in the viability assays. However, as before, further biochemical studies are necessary to confirm induction of apoptosis by these compounds.

4.3.1.3. Morphological Studies of the MCF-7 and RKO Cell Lines

A brief description of the morphology of the other two cell lines used in the cytotoxic evaluation of our compounds is presented. MCF-7 cells have round shape when they are detached but they flatten when they adhere to the vessel (Figure 4.16). At confluence, MCF-7 cells grow in continuous monolayers on plastic surfaces, but cluster formation has also been reported in dense confluent cultures. Their doubling time is approximately 29 hours.[22]

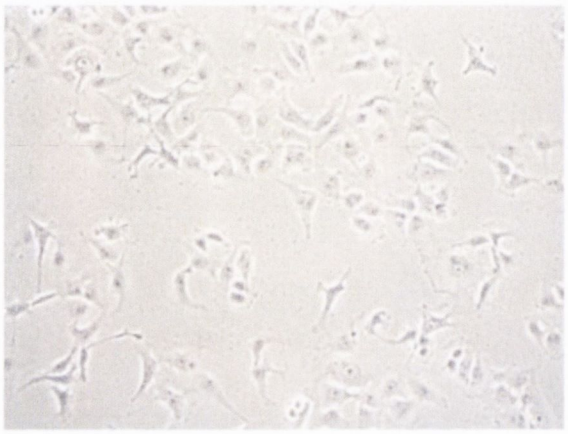


Figure 4.16. Morphology of MCF-7 breast cancer cell line.[44]

As previously mentioned, the RKO cell line shows epithelial morphology. There are three principal morphologies associated with epithelial cells. Squamos epithelium has cells which are wider than they are tall (flat and scale-like). Cuboidal epithelium has cells whose height and width is approximately the same (cube shaped). Columnar epithelium has cells which are taller than they are wide (column shaped).[45] The RKO cell line is columnar as their shape is elongated and their nucleus is longer than it is wide (Figure 4.17). They do not have cilia, which are small arm-like projections from the cells that are functionally important for transport of material along the epithelial surface. Additionally, they only form a monolayer; therefore, they are classified as simple epithelium. Three different types of RKO colorectal cancer cell lines used in this study have similar morphology: the F-6-8 cell line which have the ^{V600E} B-RAF mutant allele and a wild-type allele (dominant mutant), the A19 isogenic clone contained the mutant allele only, and the T29 isogenic

clone containing the wild type allele only. The isogenic clones were formed by homologous recombination.[46] This study was carried out by the group of Dr. S. Van Schaeybroeck in the Centre for Cancer Research and Cell Biology (CCRCB) at Queen's University Belfast (UK).

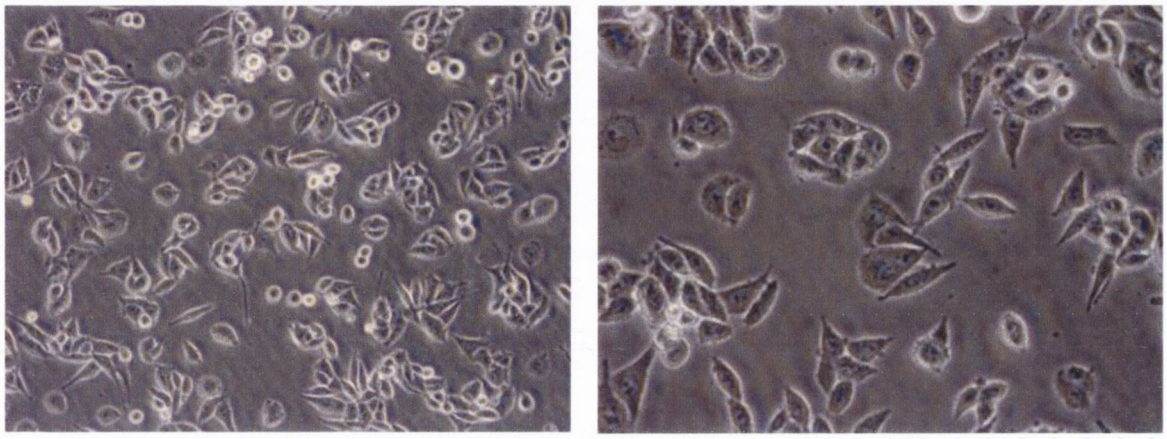


Figure 4.17. Typical morphology of RKO colorectal cancer cells at 40 × (*left*) and 20 × (*right*) (same experimental details as in legend from Figure 4.14).

4.3.2. Flow Cytometry

Flow cytometric methods are being employed at all stages of the drug development process: (i) in the early stages of drug discovery to support target identification and also to identify potential hits when screening large libraries of compounds, (ii) during lead compound characterisation to investigate interrelated mechanisms of action, (iii) in toxicology studies to assist in identifying and characterising off-target effects at the cellular level, (iv) in clinical testing to assess the safety, pharmacodynamic and pharmacokinetic effects of a drug.[47]

Since their introduction in the 1970s, the design and applications of flow cytometers have undergone tremendous change.[48] Current flow cytometers are rapid, use multiple lasers and can detect many different fluorochrome tags. The basic components of a flow cytometer consist of: (i) a flow cell that forces single cells into the middle of a fluidic sheath, (ii) a laser source, (iii) optical components to focus light of different wavelengths onto a detector, (iv) a photomultiplier to amplify the signal and (v) a computer.[47]

In a basic flow cytometer, the sample (containing the cells tagged with fluorochromes in a liquid) is drawn up and pumped into the flow cell through tubing. The cells flow through the flow chamber rapidly and are passed through one or more laser light beams. As the laser beam hits the cells (excitation), the light beam is scattered in a forward direction and a side direction. Fluorescence emission can also be detected. Scatter or fluorescence is captured, filtered and directed to the appropriate photodetectors for conversion to electronic signals. The electronics in the flow cytometer amplify the signal and convert the analogue data to digital data, which can be analysed by computer software programs. As a result of this process, the computer then produces a graphical representation of the data obtained, which is in essence the cell cycle profile.

A cell cycle profile is a histogram representation of the distribution of DNA content of individual cells in a growing cell population. Figure 4.18 shows that three major subpopulations can be found in a standard DNA profile picture consisting of two peaks and a valley, as determinated by the DNA content.

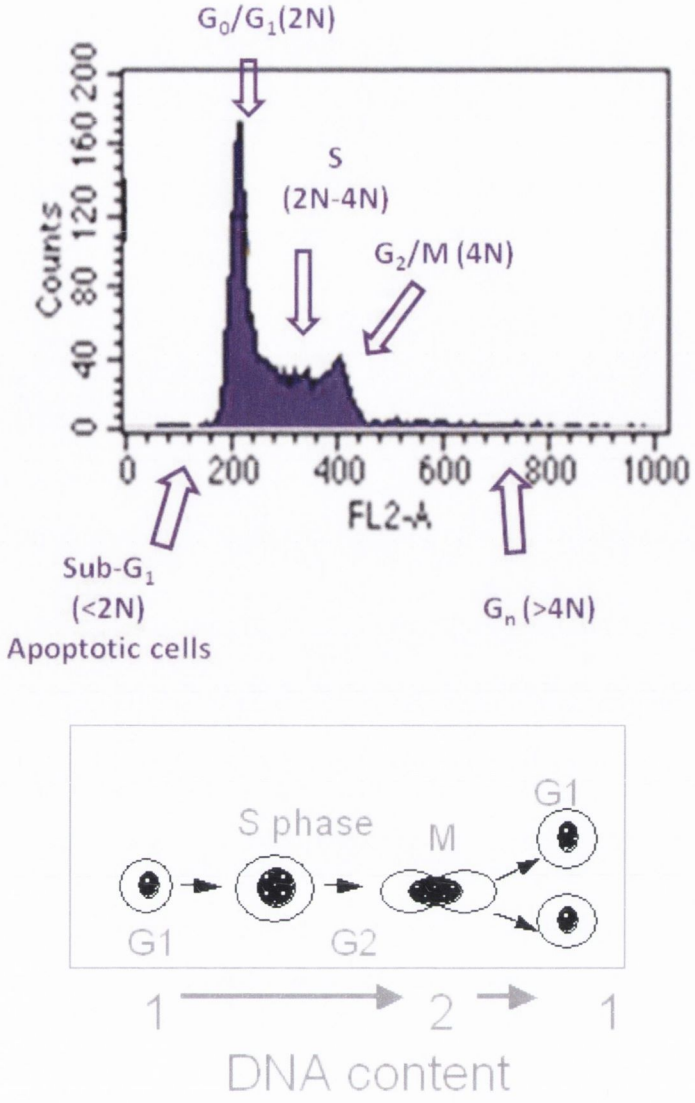


Figure 4.18. Histogram representation of a cell cycle profile (*top*), cell cycle phases with their corresponding DNA content (*bottom*).

The first peak represents cells at rest (G_0 phase) and/or cells preparing to engage in division (G_1 phase), both of which have the same DNA content (2N, where N represents the haploid chromosome number). The second population (the valley, S phase) represents cells that have engaged in DNA replication and thus have a higher DNA content than the G_0/G_1 peak (higher than 2N and lower than 4N). Finally, the third population includes cells that have doubled the DNA content but not engaged in mitosis (4N, G_2 phase), with the cells already engaged in the preliminary steps of mitosis (4N, M phase).

This overall profile shows some trailing cells at lower DNA content than that of G_0/G_1 cells (sub-G1), which are the apoptotic cells. This profile also shows the presence of cells with a higher DNA content than the G_2/M peak which identifies a cell population with a higher content of DNA per cell than tetraploid cells ($G_n > 4N$).

Cells normally move from one phase to the next during the cell cycle proliferation. Sometimes cells can get arrested in one phase prior to apoptosis. Normally, arrest is induced by drug-treatment or by other damaging agents, where a slowdown in proliferation will allow cells the time to cope with such a stress, or repair any damage by activating repair mechanisms that address the damage. Escape from phase-specific arrest can be the result of either recovery or continuation down the path towards death by activation of apoptotic cascades. In flow cytometry, arrest is usually evaluated by looking at the population distribution changes obtained from cell profiles, but also by analysing the proteins which control the cell cycle like cyclins. Drugs capable of inducing specific phase arrest include sorafenib, cisplatin, and taxol, producing G_1, S and G_2/M phase arrest, respectively.[49, 50, 51]

4.3.2.1. Flow Cytometry Evaluation of 73a and Sorafenib in the HL-60 Cell Line

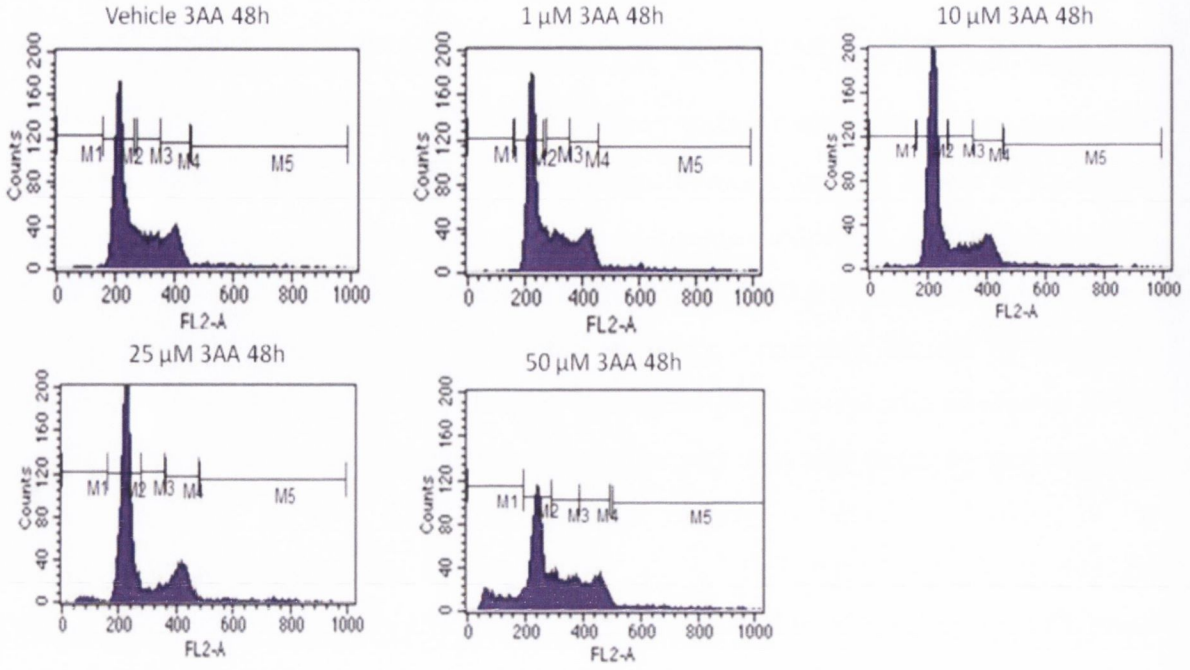
As mentioned before, it was necessary to confirm and evaluate the extent of apoptosis being induced by our compounds. Thus, taking the IC_{50} values into consideration, one of the most potent and structurally relevant compounds, **73a**, was chosen to determine the DNA content analysed by flow cytometry on the HL-60 cell line following the protocol described in the experimental section 7.2.5.

Flow cytometry results are recorded as plots of peaks for each sample of treated and untreated cells, with the horizontal axis representing increasing amounts of DNA and the vertical axis corresponding to cell numbers. In this case, 10,000 cells were monitored per experiment. Accordingly, each peak signifies cells consisting of a particular DNA content and, hence, is indicative of the proportion of cells in a particular phase of the cell cycle. As previously mentioned apoptotic cells are subdiploid, given that characteristically programmed cell death is associated with nuclear fragmentation and, as a result, a low (> 2N) DNA content. Therefore, each cell phase is assigned a section in flow cytometry: M1

(pre-G_1, apoptotic cells), M2 (G_0/G_1 phase), M3 (S phase), M4 (G_2/M) or M5 (G_n, polyploidy).

Different time points (48 and 72 h), and a range of concentrations of **73a** (1, 10, 25, 50 μM) were analysed to obtain a clear profile of the effect of **73a** on the different cell cycle phases. The results obtained from at least three independent experiments are presented in Table 4.6 as mean ± S.E.M. values and the corresponding histograms that illustrate this data are shown in Figure 4.19 – for 48 h and 72 h treatment.

A



B

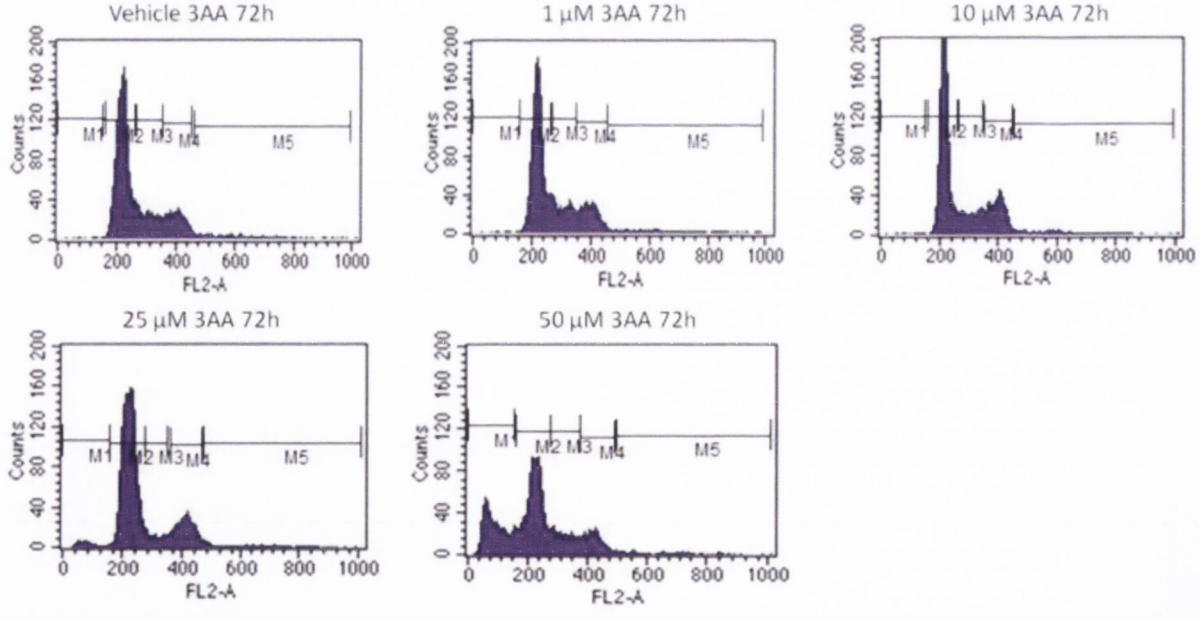


Figure 4.19. Compound **73a** induces G_0/G_1 arrest of the cell cycle prior to induction of apoptosis in HL-60 cells at 48 h (A) and 72 h (B) as demonstrated by flow cytometry analysis. Cells were seeded at a density of 2×10^5 cells/mL and treated with 1, 10, 25 and 50 μM of compound **73a** or water for the vehicle control for 48 and 72h. The cells were

then fixed by a drop-wise addition of ice-cold 70% EtOH/PBS at -20 °C overnight. Then, PBS, RNase A and propidium iodide (PI) was added. The PI fluorescence was measured on a linear scale using a FACs Calibur flow cytometer. The amount of fluorescence is proportional to the amount of DNA present and hence the population of cells in each phase of the cell cycle can be determined. Cells are gated as follows: M1 = pre-G_1 (<2N DNA), M2 = G_0/G_1 (2N DNA), M3 = S (2N – 4N DNA), M4 = G_2/M (4N DNA), M5 = G_n (>4N DNA).

Table 4.6. DNA content of each cell phase after treatment with **73a** for 48 and 72 h.

Conc. 73a (μM)	M1 (pre-G_1)	M2 (G_0/G_1)	M3 (S)	M4 (G_2/M)	M5 (G_n)
			48h		
Vehicle	**0.17** ± 0.0	**57.0** ± 1.0	**21.5** ± 1.0	**18.4** ± 1.3	**2.49** ± 0.4
1	**0.13** ± 0.0	**56.6** ± 0.2	**20.6** ± 0.3	**20.4** ± 0.1	**2.17** ± 0.1
10	**0.32** ± 0.0	**68.9** ± 0.1	**14.7** ± 0.1	**14.6** ± 0.1	**2.17** ± 0.1
25	**1.75** ± 0.4	**69.2** ± 0.4	**7.15** ± 0.0	**20.2** ± 0.4	**2.17** ± 0.1
50	**12.5** ± 0.3	**44.6** ± 0.1	**15.7** ± 0.2	**18.6** ± 0.2	**3.14** ± 0.2
			72h		
Vehicle	**0.14** ± 0.0	**65.7** ± 0.1	**16.4** ± 0.8	**14.8** ± 0.7	**2.49** ± 0.3
1	**0.13** ± 0.0	**61.7** ± 0.9	**17.1** ± 0.4	**16.7** ± 0.1	**3.23** ± 0.0
10	**0.23** ± 0.0	**61.3** ± 0.6	**14.1** ± 0.4	**18.9** ± 0.3	**3.36** ± 0.1
25	**2.22** ± 0.1	**73.4** ± 1.7	**4.19** ± 0.1	**12.9** ± 0.3	**2.56** ± 0.1
50	**25.1** ± 0.9	**43.7** ± 0.7	**14.4** ± 0.2	**13.6** ± 0.4	**1.75** ± 0.2

Generally, after 1 μM treatment with **73a** there was no difference compared to the vehicle control at 48 or 72h. However, at higher concentrations of the compound an increase in G_0/G_1 population was observed with a concomitant decrease in the rest of the cell cycle phases, such as S phase (M3) and G_2/M (M4). Cell arrest in G_0/G_1 phase was observed prior to the appearance of apoptosis in HL-60 cells after treatment with **73a**. This effect can be seen at 48 h through an increase in G_0/G_1 phase population from 1 μM (56.6%) to 10 μM (68.9%), and a stronger effect is observed at 25 μM when G_0/G_1 population increased up to 69.2% at 48 h and 73.4% at 72h. Interestingly, G_0/G_1 arrest at 72 h was

elicited only at 25 μM and at 10 μM a slight increase of 2.2% in G_2/M phase was observed. Also, at 25 μM a small increase in apoptosis was observed after 48 h (1.75%) and after 72h (2.22%). The increase in apoptosis was more dramatic when the cells are treated at 50 μM compound concentration. A percentage of 12.5% of apoptotic cells was observed at 48 h, and 25.1% at 72h. Simultaneously, while the cells are experiencing programmed cell death, the cell population is moving put of G_0/G_1; hence, a drop of 24.6% at 48h and 29.7% at 72h in G_0/G_1 cell population was observed. There is no significant change in the polyploidy DNA content (M5) during the dose-response and time course of **73a** carried out in the HL-60 cell line. These changes are summarised in Figure 4.20.

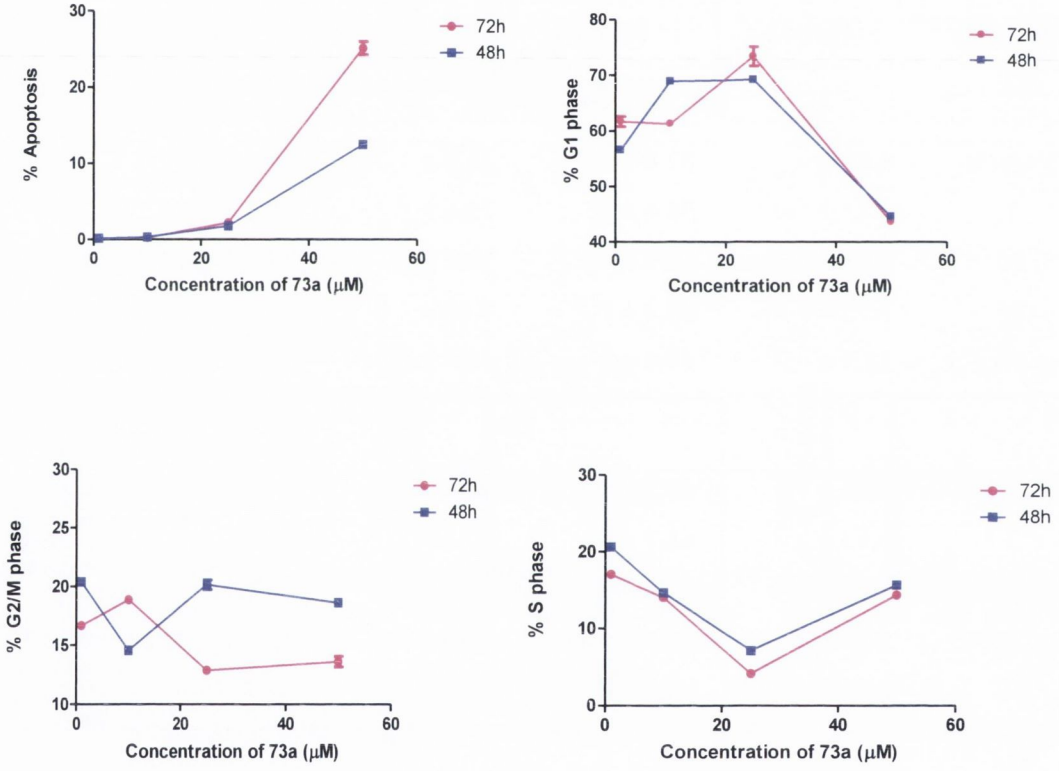


Figure 4.20. Dose-respose curves of the flow cytometry evaluation of **73a** in HL-60 cells (same experimental details as in legend from Figure 4.19. Graphs were created by using Prism GraphPad 5 software).

Additionally, the induction of apoptosis by sorafenib in the HL-60 cell line at different concentrations (1, 5, 10, 15, 25 μM) and different time points (48 h and 72 h) was also analysed (Table 4.7 and Figure 4.21) and effects similar to **73a** were observed. Both of

them induced G_0/G_1 arrest prior to apoptosis; however, sorafenib just elicited G_0/G_1 arrest at 5 μM.

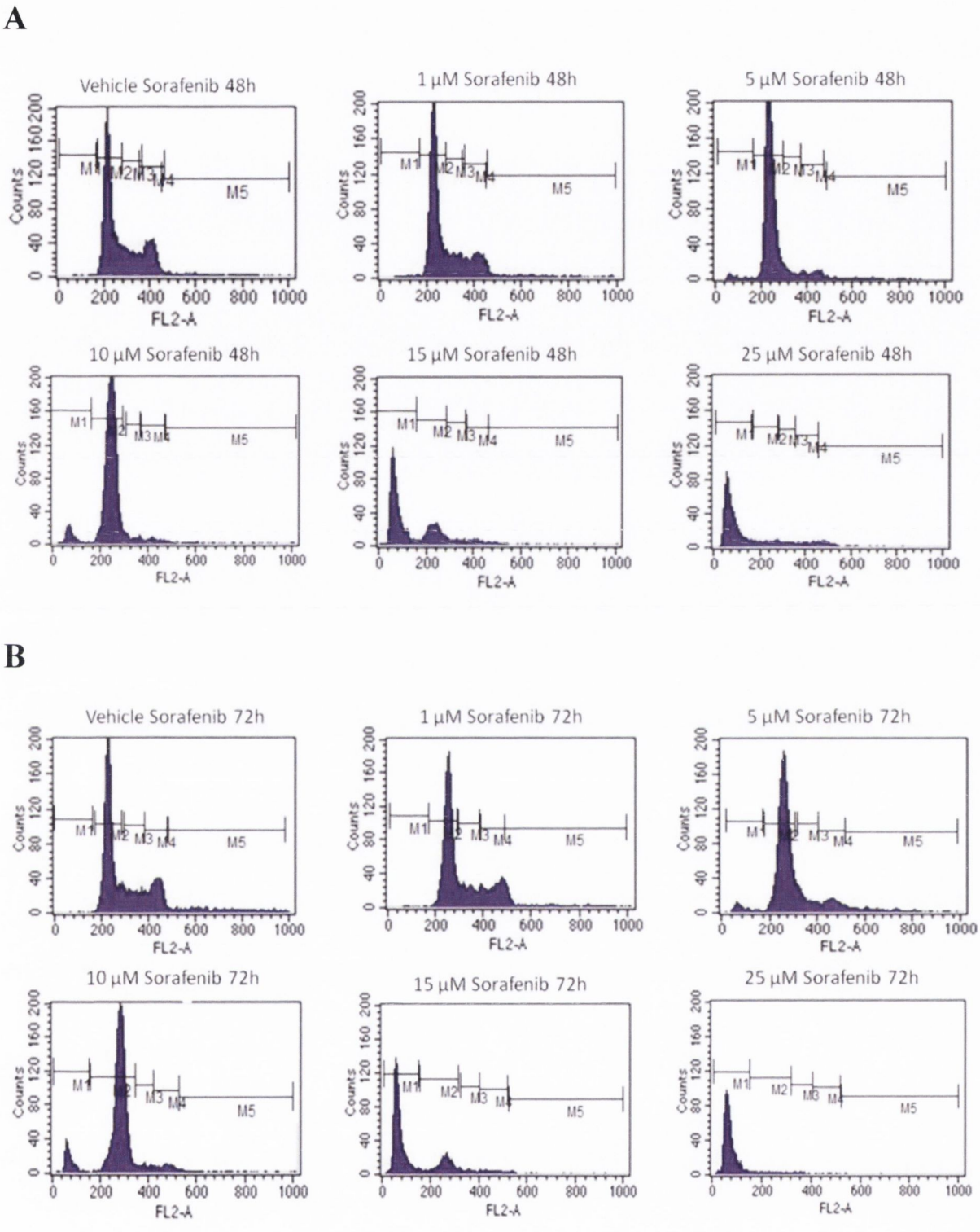


Figure 4.21. Sorafenib induces G_0/G_1 arrest of the cell cycle prior to induction of apoptosis in HL-60 cells at 48 h (A) and 72 h (B) as demonstrated by flow cytometry analysis (same experimental details as in legend from Figure 4.19).

Table 4.7. DNA content of each cell phase after treatment with sorafenib for 48 and 72h

Conc. Sorafenib(μM)	M1 (pre-G_1)	M2 (G_0/G_1)	M3 (S)	M4 (G_2/M)	M5 (G_n)
			48h		
Vehicle	**0.24** ± 0.0	**62.8** ± 0.2	**16.2** ± 0.1	**17.9** ± 0.9	**1.64** ± 0.3
1	**0.36** ± 0.0	**64.3** ± 2.5	**14.2** ± 0.4	**16.4** ± 0.6	**1.78** ± 0.3
5	**1.20** ± 0.5	**80.5** ± 2.4	**11.5** ± 0.5	**6.64** ± 1.2	**0.75** ± 0.1
10	**3.61** ± 1.0	**87.9** ± 0.4	**3.81** ± 0.6	**3.70** ± 1.0	**0.67** ± 0.1
15	**60.1** ± 1.8	**27.5** ± 1.3	**4.73** ± 0.1	**5.22** ± 0.6	**0.51** ± 0.2
25	**76.6** ± 1.0	**14.7** ± 1.8	**3.53** ± 0.2	**4.36** ± 0.7	**0.62** ± 0.2
			72h		
Vehicle	**0.10** ± 0.0	**63.8** ± 2.9	**14.9** ± 0.7	**17.8** ± 1.9	**3.11** ± 1.1
1	**0.15** ± 0.1	**59.3** ± 1.1	**15.2** ± 0.8	**20.0** ± 1.0	**3.09** ± 0.4
5	**0.17** ± 0.1	**68.4** ± 2.2	**14.9** ± 1.3	**15.3** ± 1.1	**1.42** ± 0.3
10	**10.1** ± 3.4	**79.7** ± 3.7	**5.54** ± 0.8	**3.66** ± 0.3	**1.09** ± 0.1
15	**70.9** ± 1.6	**23.0** ± 1.3	**3.45** ± 0.7	**2.46** ± 0.1	**0.75** ± 0.2
25	**93.6** ± 1.3	**3.78** ± 0.8	**2.96** ± 1.1	**0.76** ± 0.1	**0.31** ± 0.1

As previously mentioned, G_0/G_1 arrest was first elicited at 5 μM, as the cell population in this phase went from 64.3% at 1 μM to 80.5% at 5 μM, and ever higher at 10 μM (87.9% at 48 h). As a result, a decrease in the remaining cell cycle phases such as S phase (M3) and G_2/M phase (M4) was observed. A slightly lower percentage of G_0/G_1 arrest was observed at 72 h due to the fact that 10.1% of apoptosis was already occurring at 10 μM. The percentage of apoptosis increased sharply at 15 μM, being 70.9% and 93.6% at 25 μM at 72 h. A slightly more moderate increase in apoptosis was observed at 48 h, where at 10 μM the percentage of apoptosis was 3.61%, and at 15 μM and 25 μM it was 60.1% and 76.6%, respectively. This was expected, as longer time periods of cell exposure to treatment should show higher percentages of apoptosis. There is no significant change in the polyploidy DNA content (M5) during the dose-response and time course treatments of HL-60 cells with sorafenib. The changes in percentage apoptosis, G_0/G_1 phase, S phase and G_2/M phase described are summarised in Figure 4.22.

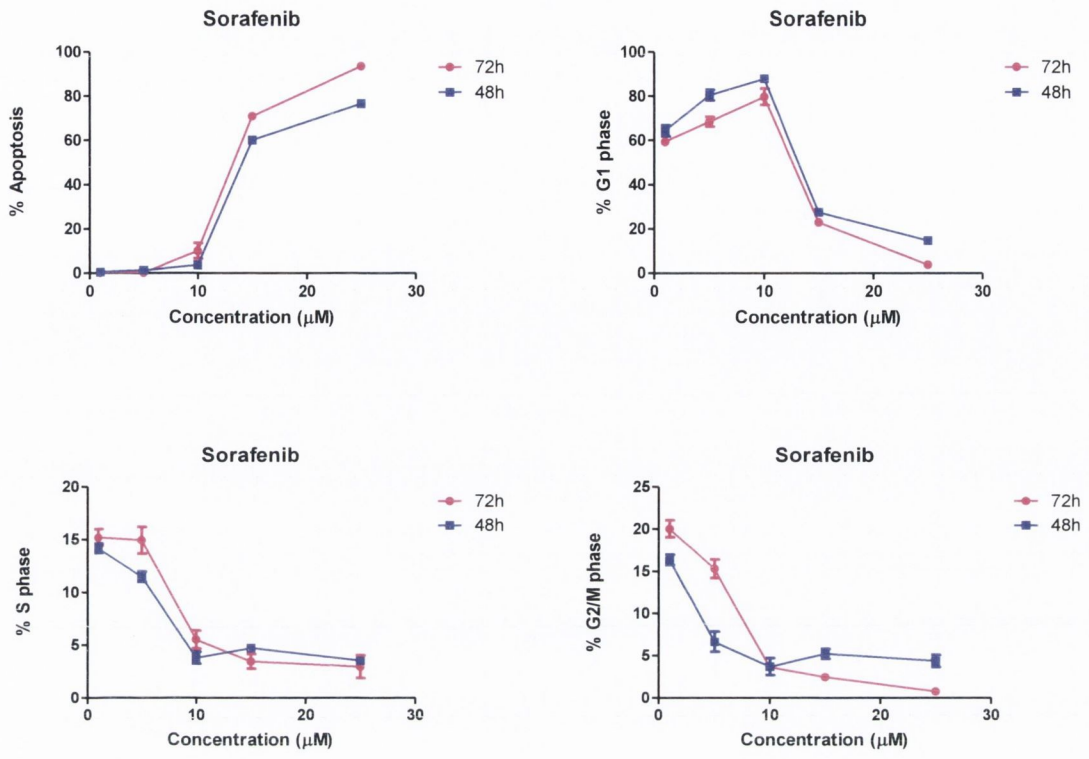


Figure 4.22. Dose-response curves of the flow cytometry evaluation of sorafenib in the HL-60 cell line (same experimental details as in legend from Figure 4.20).

Taking everything into consideration, we can conclude that compound **73a** induces apoptosis of HL-60 cells. Although, sorafenib exerted a higher percentage of apoptosis overall, both sorafenib and **73a** exert the same activity profile; both induce G_0/G_1 arrest of cells at a similar concentration range, prior to induction of apoptosis.

4.3.3. Role of Poly(ADP-ribose) Polymerase (PARP) Cleavage in Apoptosis

The enzyme poly(ADP-ribose) polymerase (PARP), whose expression is triggered by DNA-strand breaks, is also identified as a substrate for caspases. PARP is involved in DNA damage repair and functions both by catalysing the synthesis of poly(ADP-ribose) and by binding to DNA-strand breaks and modifying nuclear proteins. In cells undergoing apoptosis, PARP is cleaved from a full length 116 kD peptide to 89 kD and 24 kD polypeptides by caspase-3 during the degradation of cellular DNA, thus preventing DNA

damage repair.[52] Such cleavage essentially inactivates the enzyme by destroying its ability to respond to DNA strand breaks. The sequence at which caspase-3 cleaves PARP is very well conserved in the PARP protein from very distinct species, indicating the potential importance of PARP cleavage in apoptosis.[53]

Therefore, we studied if our compounds induce PARP cleavage, to fully confirm the mode of cell death induced by our drugs. This study was carried out by the group of Dr. S. Van Schaeybroeck at Queen's University Belfast (UK). We selected compound **73a** as a representative of the families of synthesised compounds, the same one used in our flow cytometry studies. We had already determined that **73a** induces apoptosis in a hematopoietic cancer such as leukaemia, using flow cytometric analysis of HL-60 cells. However, for the study of PARP cleavage we used one of the solid tumour cell lines (RKO), examined in this study to confirm that the mechanism of cell death caused by our compounds is by induction of apoptosis in both hematopoietic and solid tumours.

The biochemical technique used to identify the PARP cleavage was western blotting, which is one of the most common methods for protein detection in molecular biology. The western blot utilises the high affinity relationship between antibodies and antigens to detect proteins from cellular lysates, tissues, or other extracts. It is an assay that detects specific proteins within a mixture by a multistep process consisting of protein extraction, quantification, separation, western blot analysis and protein detection. Eukaryotic cells are packed with proteins which are folded into three-dimensional shapes and to perform a western blot, these proteins have to be denatured. Thus, proteins were solublised with detergents such as sodium dodecyl sulfate (SDS) and, also, intramolecular disulfide bonds were broken to reduce proteins into long chains of amino acids by the addition of a reducing agent, usually β-mercaptoethanol.[54] Afterwards, protein concentration for each sample was determined by interpolation from a standard curve generated using serial dilutions of known concentrations of bovine serum albumin (BSA). After proteins are denatured, they were resolved on a polyacrylamide gel by electrophoresis, where they were separated depending on their molecular weight. Next, the proteins were transferred to a membrane (nitrocellulose or PVDF) with a high protein binding capacity. Membranes were then blocked with one of a variety of protein blocking buffers. Blocked membranes were then incubated with a primary antibody that recognises a specific cellular protein. An enzyme conjugated secondary antibody was then used to detect the primary antibody. The membrane was then exposed to a film in the presence of a chemiluminescent substrate to

detect the amount of proteins on the membrane. β-Actin is a loading control to make sure that all the wells have the same amount of protein so the intensity changes in the blot bands are due to a biological change and not by any differences in the amount of protein loaded.

The results obtained in the western blots show that **73a** induces PARP cleavage in both the wild type (T-29) and the allele mutated and allele wild type (F6-8) colorectal cell lines (see Figure 4.23). Background cell death could be observed in the control of T-29 as the second band is already present, but the intensity of this band is stronger at 2.5 and 5 μM drug treatment, confirming the induction of apoptosis of this drug. Comparing both cell lines, it can be seen that the two bands are more defined in the T-29 cell line, we could postulate that the cleavage occurred with more intensity in the T-29 cell line. This observation is justified by the fact that the IC_{50} value of **73a** in F6-8 cells is 3-fold higher than that in T-29 cells; therefore higher concentrations of **73a** would be required to obtain the same effect in F6-8 than in T-29.

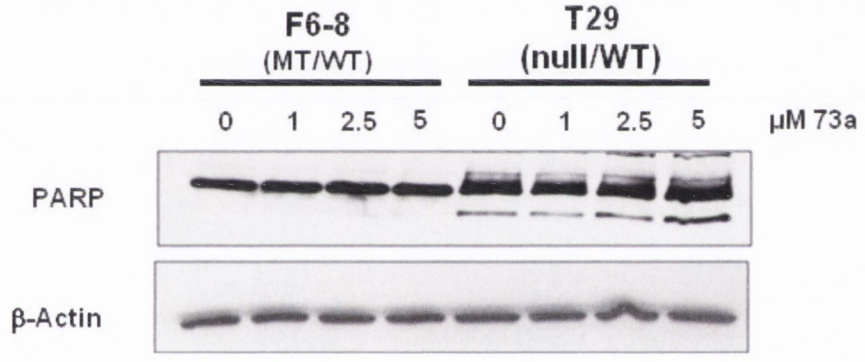


Figure 4.23. Compound **73a** induces PARP cleavage in F6-8 mutated and T29 wild type RKO cell lines. Western blot assay were performed as described in experimental section 7.2.8. Aliquots of cell lysate were separated on SDS-polyacrylamide gel by electrophoresis, followed by electro-transferred to nitrocellulose membrane. This was blocked with blocking buffer and then incubated overnight at 4 °C using the primary antibody followed by the secondary antibody against PARP. The membrane was then subjected to chemiluminescence detection.

In conclusion, we can say that **73a** induces apoptosis in both hematopoietic tumours (identified by flow cytometry) and solid tumours (identified by PARP cleavage). Consequently, **73a** as a representative of all synthesised compounds, led us to the

conclusion that the cytotoxicity induced by these derivatives in all studied cell lines (hematopoietic and solid tumour cell lines) is very likely to be caused by the induction of apoptosis.

4.4. Protein Kinase Evaluation

After evaluating the cytotoxicity of our compounds in a wide range of cancer cell lines and having determined that they reduced cell viability and proliferation in these cancer cells by the induction of apoptosis in both hematopoietic and solid, we had still not confirmed the pathway or enzymes which were being targeted *en route* to inducing cell death. Due to the structural similarities of our molecules to sorafenib, we hypothesised that cell death may be caused by an inhibition of certain protein kinases. To test this hypothesis we performed assay in a number of kinases CK-1δ, GSK-3, the RAF/MEK/ERK pathway and p38 MAPK. Additionally, we indirectly studied the potential anti-angiogenic activity of our compounds by testing their inhibition of VEGFR. Finally, we examined the effect of our compounds on the protein expression levels of B-RAF, C-RAF, MEK, ERK, STAT-3 and AKT in two RKO colorectal cancer cell lines using western blotting.

4.4.1. Casein Kinase 1δ (CK-1δ) and Glycogen Synthase Kinase-3 (GSK-3) Studies

The CK-1 family are serine/threonine kinases which play an essential role in a diverse range of physiological processes, such as DNA repair, cell cycle proliferation, cytokinesis, differentiation and apoptosis.[55,56] On the one hand, members of CK-1δ are key regulatory proteins (*e.g.* p53 and β-catenin), which act as signal integration molecules in apoptosis.[55] On the other hand, the GSK-3 family of kinases plays an important role in signal transduction systems (*e.g.* Wnt/wingless and PI-3 kinases) which influence proliferation and cell survival.[57] (see Chaper 1, section 1.4.3.2.) Therefore, study of the inhibition of these two kinases reveals whether or not the apoptosis previously observed by our compounds is caused by their particular modification. This testing was carried out in collaboration with Dr. C. Perez in the Consejo Superior de Investigaciones Cientificas (CSIC), Madrid (Spain). As mentioned in section 1.4.3.2, although they are independent pathways, the same transcription factor – activator protein 1 (AP-1)[58]– is activated in the

nucleus by GSK-3, CK-1δ and the RAF/MEK/ERK pathways which plays a critical role in neoplastic transformation and tumour development. In light of this connection between RAF/MEK/ERK and GSK-3 and CK-1, we decided to evaluate the ability of our compounds to target CK-1δ or GSK-3 kinases (see experimental details in section 7.2.6)

Unfortunately, the results obtained for these two kinases showed that none of the compounds of the 5 families (**51a-d**, **52a-c**, **70a-d**, **72a**, **c**, **71a-d**, **73a**, **c** and **79a**, **d**) targets these enzymatic systems (all the IC_{50} values >10 μM).

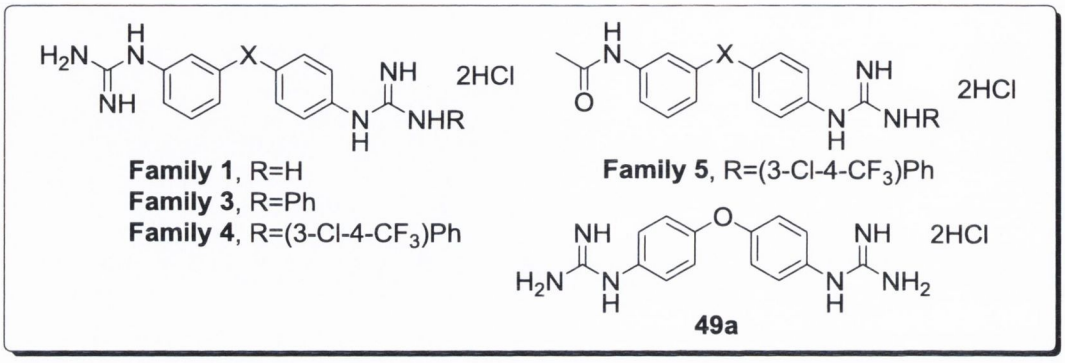
4.4.2. RAF/MEK/ERK pathway, p38 MAPK and VEGFR Study

Considering the array of kinases targeted by sorafenib, we explored if our molecules show similar activity towards several enzymatic pathways targeted by this compound. This analysis was performed by the kinase screening company Cerep.[59]

The screening involved measuring the percentage of inhibition of the enzymatic activity of ERK 1/ 2, p38 MAPK, VEGFR and RAF-1/MEK-1. A compound representative of each family was evaluated including one of the 4,4'-*bis*-guanidiniums (compound **49a**, see structure in Table 4.8) and the results obtained are shown in Table 4.8.

Compounds in Family 1, **51a** (X= O) and **51d** (X= NH), selectively inhibit the RAF-1/MEK-1 pathway (Table 4.8), showing no inhibition of the other kinases tested. In particular, compound **51d** showed 86% inhibition of this pathway in comparison to a moderate 22% inhibition in the case of **51a**, suggesting that the -NH- linker could confer more selectivity for this pathway than the -O- linker. No example from Family 2 was chosen as they showed very low anticancer activity when tested in the HL-60 cell line. The example chosen from Family 3A, compound **70a**, exhibited no inhibitory activity towards any of the kinases evaluated (Table 4.8). This correlates with the cytotoxicity results obtained in HL-60 cells, where the IC_{50} value obtained was > 100 μM.

Table 4.8. Inhibition of kinase activity (%) for ERK1/2, p38 MAPK, VEGFR and RAF-1/MEK-1 pathways by compounds **51a-d** (Family 1), **70a** (Family 3A), **71a-d** (Family 4A), **79a** and **79d** (Family 5) and **49a** at 10 μM.



Family	Compound	X	% Inh ERK1/2	% Inh p38 MAPK	% Inh VEGFR	% Inh C-RAF/ MEK-1
1	**51a**	O	-4	-1	-1	22
1	**51d**	NH	-37	4	-6	86
3A	**70a**	O	-15	1	-1	4
4A	**71a**	O	-7	21	0	96
4A	**71b**	CO	21	20	0	96
4A	**71c**	CH_2	17	-7	17	99
4A	**71d**	NH	-27	12	-3	74
5	**79a**	O	8	0	-15	35
5	**79d**	NH	12	12	1	38
	49a	O	6	-1	7	1.9

Note: Low to moderate negative values have no real meaning and are attributable to variability of the signal around the control level. High negative values (≥ 50%), that are sometimes obtained with high concentrations of test compounds, are generally attributable to non-specific effects of the test compounds in the assays. Note: Each kinase testing was carried out following a different protocol, for experimental details see section 7.2.6.

Compounds from Family 4A (**71a-d**) seemed to be the most successful in terms of kinase inhibition. All four compounds tested were highly active against kinases, especially the RAF-1/MEK-1 pathway (Table 4.8). This confirms the need for the 4-chloro-3-trifluoromethylphenyl moiety for the inhibition of these kinases. The percentage values

obtained for inhibition of the C-RAF/MEK-1 pathway were all excellent, ranging from 74% for **71d** to 96%, 96% and 99% inhibition for **71a**, **71b** and **71c**, respectively. This supported the fact that the nature of the linker in the diaromatic moiety was not critical for achieving outstanding inhibition of the C-RAF/MEK-1 pathway. Regarding the other kinases tested, **71c** displayed moderate anti-angiogenic activity by inhibiting VEGFR by 17%, and also inhibited ERK 1/2 to a similar percentage. This compound showed the best activity towards VEGFR, ERK 1/2 and C-RAF/MEK-1 of all compounds tested.

Compounds **71a**, **71b** and **71d** (X = O, CO and NH, respectively) displayed a modest 12-21% inhibition of p38 MAPK. In this case, the more polar linkers seemed to induce better activity than the lipophilic linker, allowing us to suggest that, for targeting p38 MAPK, the formation of HBs could be favorable.

Acetamide derivatives **79a** and **79d** displayed poor inhibition of all tested kinases, with only 34-35% inhibition of the C-RAF/MEK-1 pathway being observed. This seemed to indicate that the good cytotoxicity results obtained for these compounds with all tested cell lines was not a result of inhibition of these kinases and that other mechanisms should be explored.

Finally, the results achieved for compound **49a** indicate that it did not inhibit any of the kinases tested, suggesting that this type of 4,4'-*bis*-guanidinium did not undergo their cytotoxicity by targeting protein kinases tested in this study.

4.4.3. LanthaScreen® Europium (Eu) Kinase Binding Assay for B-RAF

Due to the fact that some of our molecules from Family 4A (**71a-d**) show 74-99% inhibition of RAF-1/MEK-1, and since sorafenib competitively binds to and inhibits B-RAF, we decided to evaluate if our synthesised compounds selectively bind to B-RAF by performing a LanthaScreen® Eu Kinase Binding Assay (in collaboration with Dr. Lloyd's laboratory, School of Biochemistry and Immunology, TCD).

The rapidly growing interest in kinases as drug targets has prompted the development of many kinase assay technologies. These technologies can be grouped into three categories: radiometric assays, phospho-antibody-independent fluorescence/luminescence assays, and

phospho-antibody-dependent fluorescence/luminescence assays.[60] The latter technique was the one employed in this work.

The assay used was the LanthaScreen® Eu Kinase Binding Assay. This assay is based on the binding and displacement of a fluorophore Alexa Fluor® 647-labeled, ATP-competitive kinase inhibitor scaffold (kinase tracer) to the kinase of interest, in our case the B-RAF kinase (see Figure 4.24 a).[61-62] Tracers based on a variety of scaffolds have been developed in order to address a wide range of kinase targets. Binding of the tracer to the kinase of interest, in this case the B-RAF kinase, was detected using a europium-labelled anti-tag antibody, which binds to the kinase. Simultaneous binding of both the tracer and the antibody to the kinase results in a high degree of FRET (fluorescence resonance energy transfer) from the Eu donor fluorophore to the Alexa Fluor® 647 acceptor fluorophore on the kinase tracer (see Figure 4.24 b). However, when a kinase inhibitor (theoretically our compound) competes for binding to the ATP binding site of the kinase with the tracer, a displacement of the tracer results, meaning there is a loss of FRET and no emission is detected (see Figure 4.24 c and d).

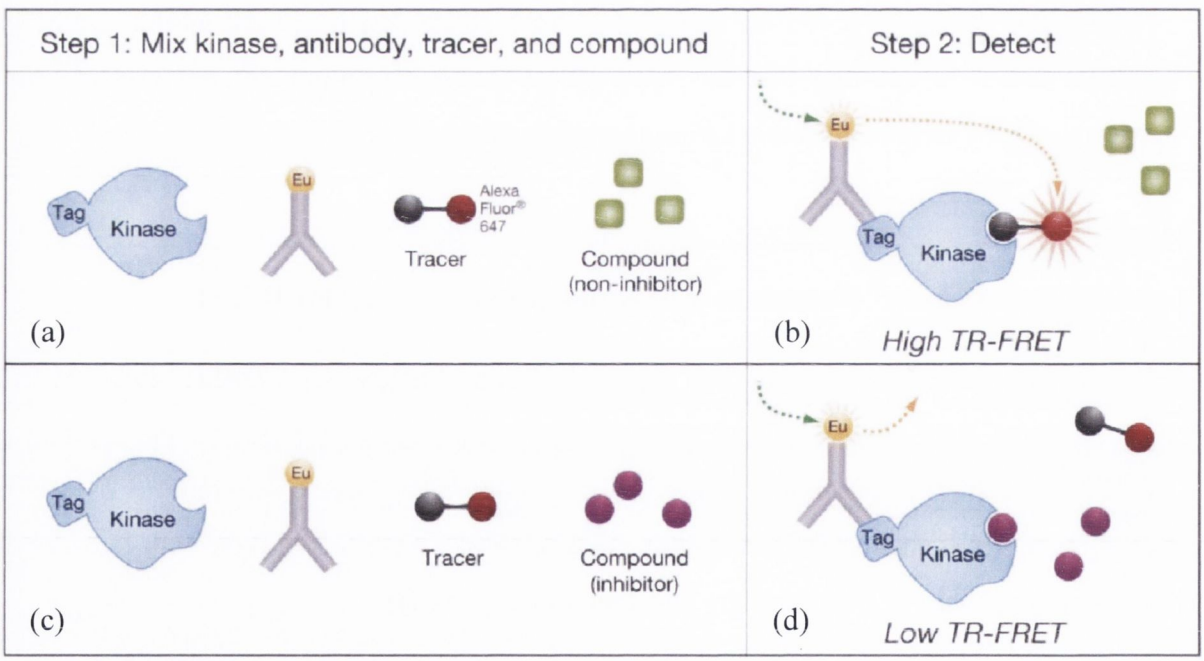


Figure 4.24. Schematic representaion of the LanthaScreen Eu-kinase binding assay for B-RAF based on fluorescence resonance energy transfer (FRET).[61]

Because the kinase tracers are based on ATP-competitive kinase inhibitors, they are suitable for detection of any compounds that bind to the ATP site. This includes Type II inhibitors (*e.g.* Gleevec®/imatinib, sorafenib, BIRB-796), which bind to both the ATP site and to an allosteric site. In contrast to most fluorescence-based kinase activity assays, the LanthaScreen® Eu Kinase Binding Assay can be read continuously, which facilitates evaluation of compounds with slow binding kinetics.[62] A further benefit of this assay is that it can be performed using either activated or non-activated kinase preparations, which enables characterisation of compounds which bind preferentially to non-activated kinases (*e.g.* Gleevec®/imatinib).

Firstly, before analysing the activity of our compounds, the evaluation of the binding of sorafenib to B-RAF was performed in order to verify the suitability of the assay. Following the protocol explained in the experimental section 7.2.7, three independent experiments evaluating the competitive binding of sorafenib to the ATP binding site of B-RAF were carried out. As expected, in the three experiments carried out, sorafenib competitively bound to B-RAF, demonstrating the robustness of the assay. Higher concentrations of sorafenib generate less FRET emission, as sorafenib displaces the tracer more strongly. This effect is illustrated in Figure 4.25.

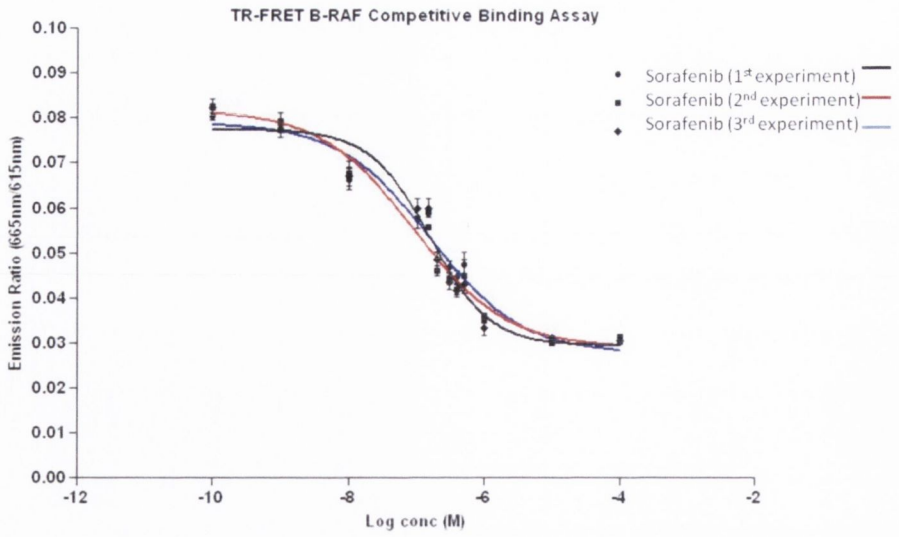


Figure 4.25. Sorafenib exhibits dose-responsive competitive binding to B-RAF. 5 μL of each concentration of serial diluted compound (test or control), kinase/antibody solution and tracer solution were added to each well in a 384-well plate, in triplicate. Each plate also contains a blank, a vehicle control and a positive control. The plate was incubated at room temperature for 60-120 min and was read in the plate reader. Error bars represent SEM. Data is representative of at least three independent experiments.

The IC_{50} value obtained in this experiment with sorafenib was 136 ± 23 nM, very similar to the reported value in the literature (120 nM).[63] The data obtained was reproducible; dose-responsive and data points fit to applied statistical models (as per R^2 and hill slope). It was therefore concluded that the assay employed for exploration of the potential competitive binding of our compounds to B-RAF was suitable.

Subsequently, the binding assay was performed on a selection of our compounds, with the best IC_{50} values obtained during the screening performed in the HL-60 cell line. Therefore, **71b-d**, **73a**, **73c** and **79d** were selected. Dose- response data showed that none of the assayed compounds displayed dose-responsive competitive binding to B-RAF at concentrations ranging from 100 nM to 100 μM, as no decrease in FRET emission was observed. Only a slight decrease in emission ratio was found in compound **71c** and **71d**, which could indicate their weak binding to B-RAF (Figure 4.26).

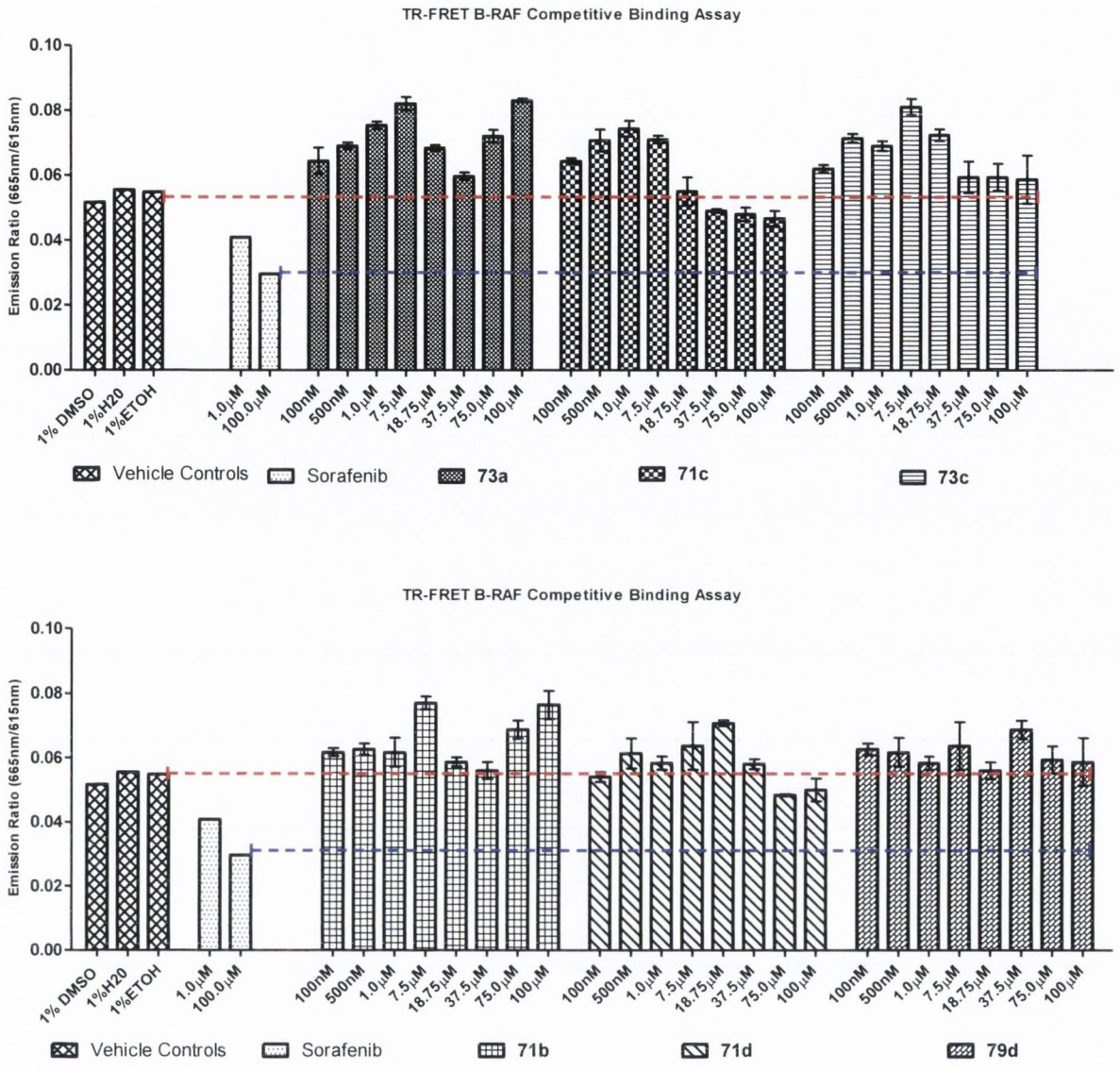


Figure 4.26. 71a-d, 73a, c, and **79d** do not exhibit competitive binding at B-RAF in a dose-dependent manner. Emission ratio values for compounds (experimental and control) compared to vehicle control(s) using two-way ANOVA (Bonferroni test). Sorafenib (1.0 and 100 μM) emission ratio values shown to be statistically significantly lower than vehicle control ($P < 0.001$). Error bars represent SEM. Data is representative of at least three independent experiments.

This study shows that active compounds do not competitively bind to the ATP binding site of B-RAF. However, assayed compounds may bind to other parts of B-RAF (*e.g.* allosteric sites) or they could exert inhibitory activity against different targets such as MEK.

4.4.4. Effect of Selected Compounds in the Expression Levels of Various Signalling Proteins in RKO Cells.

Western blot analysis of two RKO colorectal cancer cell lines (wild type T-29 and one of the ^{V600E} B-RAF mutated cell lines, F-6-8) was performed by the group of Dr. S. Van Schaeybroeck at Queen's University Belfast (UK) to study the expression of relevant signalling proteins following treatment with selected compounds. This would provide further detail of the molecular target of our compounds. We analysed the expression levels of components of the RAF/MEK/ERK pathway as it was targeted by sorafenib, and other important molecular targets in cellular processes such as STAT-3 and AKT using **73a** as lead representative of all the synthesised compounds. Results of the RAF/MEK/ERK pathway are shown in Figure 4.27.

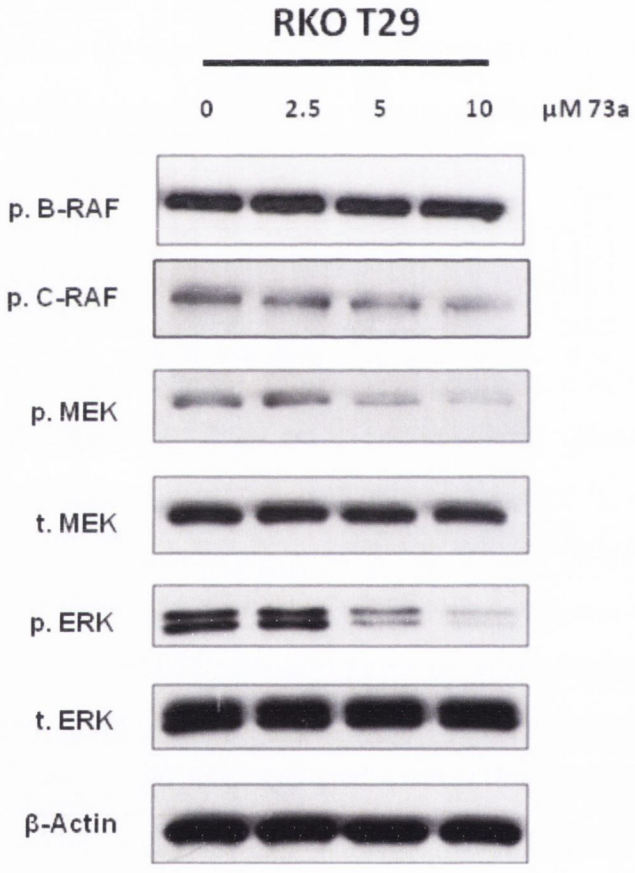


Figure 4.27. The effect of compound **73a** at 0, 2.5, 5 and 10 μM on expression levels of phosphorylated B-RAF, C-RAF, MEK, ERK and total expression of MEK and ERK in wild type T-29 RKO cell line (same experimental details as in Figure 4.23).

The results obtained showed that **73a** did not inhibit the expression of B-RAF in the RKO T-29 cell lines, as no change in the intensity of the bands when increasing the concentration of the compound was observed (Figure 4.27). This result correlates with our findings in the LanthaScreen® Eu Kinase Binding Assay for B-RAF; hence, we confirmed that the cytotoxic effect of our compounds is not the result of B-RAF inhibition.

Conversely, compound **73a** seemed to slightly inhibit C-RAF at 10 μM. This result was not expected as the ATP binding site (kinase domain) of the two isoforms B-RAF and C-RAF are very similar as is demonstrated in the literature.[64] However, it could be possible that **73a** binds to other domains to C-RAF, such as allosteric sites. This result need to be evaluated more in detail to identify the kind of binding between the drug and the protein and also the strength of this binding.

However, the reduction of the levels of p-MEK by compound **73a** at 5 and 10 μM was very obvious, as the decrease in the intensity of the bands was unambiguous. This suggests that the likely main target of our compounds is MEK, what was supported by computational chemistry studies which analysed the binding between the compounds and MEK. This will be described in more detail in Chapter 5.

The results obtained by the company Cerep also suggested that compounds from Family 4 exert 74-99% inhibition of C-RAF/MEK-1 pathway, confirming the results obtained in the western blots.

In addition, a decrease in the expression of ERK levels was observed as a decrease in the intensity of the bands at 5 and 10 μM was found. This is another proof of the fact that the RAF/MEK/ERK pathway (or MAPK pathway) is down-regulated by compound **73a**. ERK is the terminating kinase of the RAF/MEK/ERK pathway; as a result, the decrease in the expression of ERK observed is likely to be related to an inhibition of the upstream kinases of the pathway (*e.g.* RAF or MEK). This may suggest the RAF/MEK/ERK pathway is the molecular target of our compounds.

The same was expected for the mutated F6-8 cell lines; however in light of the fact that the IC_{50} of **73a** is 3-fold higher in F6-8 than in T-29 cell lines, higher concentrations of **73a** would be required to obtain the same effect in F6-8 than in T-29, so further experiments should be required to confirmed this fact.

Other important molecular targets in cellular processes, such as AKT and STAT-3, were investigated. Of these, STAT-3, plays an important role in numerous cellular events such as differentiation, proliferation and apoptosis. STAT-3 is constitutively activated in diverse tumours which induce dysregulation of growth, survival, angiogenesis and immune modulation; hence, it contributes to tumour progression. For instance, STAT-3 is constitutively activated in 50% of leukaemia cases. Sorafenib induces apoptosis in the leukaemia HL-60 cells by inhibiting Src kinase-mediated STAT-3 phosphorylation.[3] However, **73a** does not inhibit STAT-3 in any of the cell lines tested as its expression was not modified at any of the **73a** concentrations tested (Figure 4.28). As a result, the mode of action of our compounds does not appear to be due to the inhibition of STAT-3.

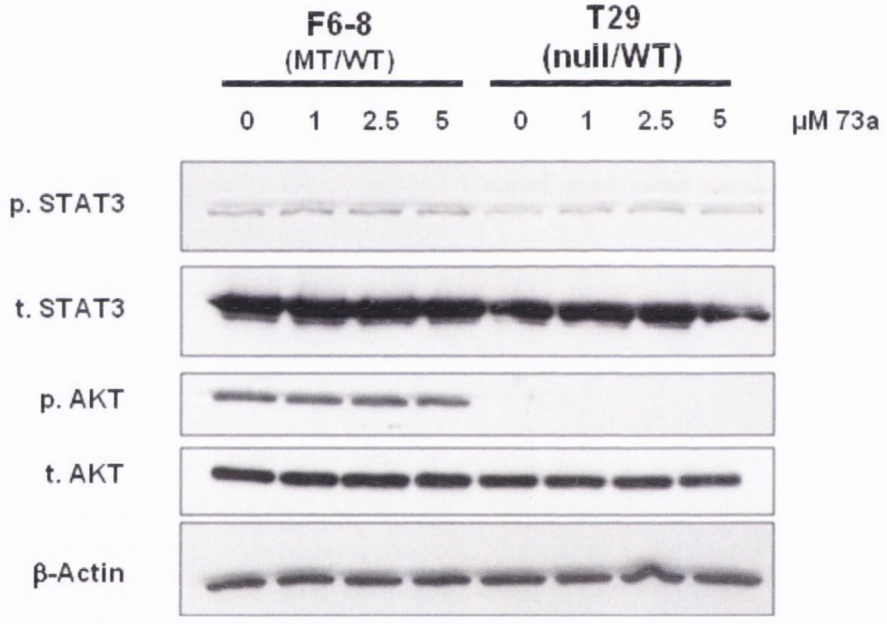


Figure 4.28. The effect of compound **73a** at 0, 1, 2.5 and 5 μM on expression levels of phosphorylated and total expression of STAT-3 and AKT in mutated F6-8 and wild type T-29 RKO cell lines (same experimental details as in legend from Figure 4.27).

Additionally, AKT – also called Protein Kinase B (PKB) – plays a key role in cell proliferation, transcription, migration and apoptosis.[65] AKT could promote growth factor-mediated cell survival by inhibiting the pro-apoptotic effect proteins of the Bcl-2 family, or by activating pro-survival genes via NF-κB.[66] AKT is also involved in the

PI3K/AKT/mTOR signalling pathway and Wnt signalling cascade. In the latter, AKT modulates the expression of GSK-3. The western blot carried out on AKT revealed that no significant change in the expression of AKT is observed after treatment with **73a** at different concentrations. This finding correlates with the results obtained in the testing of GSK-3 carried out in section 4.4.1, where it was shown that some of our compounds do not exhibit any effect on the expression of GSK-3.

4.5. Conclusions

In this chapter we have described the biological evaluation of the synthesised compounds as anticancer agents. This evaluation involved the study of their effects on cell viability and proliferation in a wide variety of cancer cell lines, their apoptosis-inducing abilities, their cell cycle effects and the study of their molecular targets as potential protein kinase inhibitors.

Evaluating the data obtained in the cytotoxicity assays in all analysed cell lines, it can be concluded that our compounds inhibit cell proliferation of leukaemia, liver carcinoma, colorectal cancer, ^{V600E} B-RAF mutated colorectal cancer, breast cancer and tamoxifen-resistant breast cancer cell lines. Their IC_{50} values were rated in the low μM range and in some cases their cytotoxic activity was comparable or even more satisfactory to that of commercial sorafenib. The antiproliferative results obtained are very promising for compounds in such an early stage of development.

Several conclusions can be achieved regarding SAR information from the evaluation of the cytotoxicity of all families of compounds in HL-60 cell line. For example that the guanidinium moiety showed higher cytotoxicity than the 2-aminoimidazolinium one or that the (4-Cl-3-CF_3)-Ph subunit was crucial to achieve high cytotoxic activity (see more in section 4.2.1). Additionally, the fact that the IC_{50} values of our compounds in MCF-7 and Tam-MCF-7 cell lines were nearly identical suggests that our compounds overcome the resistance to tamoxifen in this cell line; hence, they could be a potential therapeutic alternative for patients with tamoxifen-resistant breast cancer. However, high selectivity for colorectal cancer cell line with ^{V600E} BRAF mutated over the wild type was not observed.

Having performed cell viability studies and determined corresponding IC_{50} values (section 4.2), the potential of these compounds inducing apoptotic cell death was explored (section 4.3). Microscopy studies of the effect of **73a** and **79d** in several cancer cells were carried out where marked changes in cellular morphology such as chromatin condensation, membrane blebbing, nuclear breakdown, the appearance of membrane-associated apoptotic bodies and DNA fragmentation were observed, suggesting the induction of apoptosis by our compounds.

Flow cytometry analysis in HL-60 cell line and PARP cleavage induction in the RKO cell line were also performed to quantify and confirm the pro-apoptotic ability of our compounds. These studies were carried out by using compound **73a** which is one of the most active derivatives in the cell viability studies. Flow cytometry studies demonstrated that our compounds induce G_0/G_1 phase arrest of the cell cycle prior to apoptosis, in a similar manner to sorafenib. Also PARP cleavage confirmed that apoptosis is the type of cell death caused by **73a**.

Overall, taking into consideration the results from the viability and the apoptosis assays, it can be concluded that in addition to inhibiting cell viability and proliferation, these derivatives are significant pro-apoptotic inducers.

Due to the structural similarities of our molecules with sorafenib, we postulated that cell death was being caused thought an inhibition of protein kinases. Therefore, several kinases were examined in order to identify the molecular mechanism of action of these derivatives. B-RAF, C-RAF (also called RAF-1), MEK-1, ERK 1/2, p38 MAPK, and VEGFR were evaluated. The results showed that compounds from Family 4A displayed 74-99% inhibition of the RAF-1/MEK-1 kinase pathway. Additionally, compound **51d** – a member of Family 1 – showed 86% inhibition of the RAF-1/MEK-1 kinase pathway. However, other members of Family 1 did not show as strong inhibition of this pathway suggesting that the -NH- linker could confer more selectivity for this pathway. Compounds from Family 2 were not tested in kinase inhibition due to their poor cytotoxic activity. Families 3A and 5 did not show inhibition of the C-RAF/MEK-1, ERK-1/2, p38 MAPK or VEGFR pathway. Members of Family 5 displayed high cytotoxicity in all studied cancer cell types but their mode of action seems to be different from the compounds in Family 4A as they do not target the C-RAF/MEK-1 kinase pathway.

The effect of compound **73a** in the expression levels of various signalling proteins in RKO cells was analysed. Western blot studies showed that **73a** do not inhibit B-RAF. This result was confirmed by LanthaScreen® Eu Kinase Binding Assay, which indicated that several compounds from Families 4 and 5 do not competitively bind to B-RAF. Conversely, a slight reduction of the levels of C-RAF was observed. This interaction needs to be analysed in more detail as the ATP-binding site of C-RAF and B-RAF is nearly identical; hence these compounds could be allosteric inhibitors of C-RAF. However, a significant reduction of the expression levels of p-MEK was obtained after the treatment with the compound **73a**, suggesting that the likely target of our compounds is MEK.

Western blot studies displayed a reduction of expression levels of p-ERK when treated with **73a**, confirming the down-regulation of the RAF/MEK/ERK pathway by our compounds from Family 4. The decrease in the expression of ERK observed is likely to be related to an inhibition of the upstream kinases of the pathway (*e.g.* RAF or MEK) as results obtained by the company Cerep showed that our compounds do not target ERK 1/2 directly. This may suggest the RAF/MEK/ERK pathway is the molecular target of our compounds.

The biological results obtained are very encouraging, consequently these compounds could be considered as a basis for future biologically active compounds as improvement in their chemical structure could lead to the design of more potent and active derivatives.

4.6. References

(1) http://www.who.int/en/.

(2) Siegel, R.; Naishadham, D.; Jemal, A. *CA-Cancer J. Clin.* **2012**, *62*, 10-29.

(3) Zhaoa, W.; Zhanga, T.; Qua, B.; Wua, X.; Zhua, X.; Menga, F.; Gua, Y.; Shub, Y.; Shena, Y.; Suna, Y.; Xua, Q. *Anti-Cancer Drug* **2010**, *22*, 79-88.

(4) Gedaly, R.; Angulo, P.; Hundley, J.; Daily, M. F.; Chen, C.; Evers, B. M. *J. Surg. Res.* **2012**, *176*, 542-548.

(5) Gedaly, R.; Angulo, P.; Hundley, J.; Daily, M. F.; Chen, C.; Koch, A.; Evers, B. M. *Anticancer Res.* **2010**, *30*, 4951-4958.

(6) Hu, Y.; Huang, Y.; Chen, X.; Luo, R. *Shandong Yiyao* **2010**, *50*, 37-39.

(7) Hu, Y.; Yao, Y.; Wu, B.; Ma, G.; Wang, L.; Li, C. *Zhongliuxue Zazhi* **2011**, *17*, 521-523.

(8) Park, B. J.; Whichard, Z. L.; Corey, S. J. *Cancer Lett.* **2012**, *320*, 104-110.

(9) Morgillo, F.; Martinelli, E.; Troiani, T.; Orditura, M.; De Vita, F.; Ciardiello, F. *PLoS ONE* **2011**, *6*, e28841.

(10) Shacham-Shmueli, E.; Geva, R.; Figer, A.; Bulocinic, S.; Nalbandyan, K.; Shpigel, S.; Atsmon, J.; Brendel, E. *J. Clin. Pharmacol.* **2012**, *52*, 656-669.

(11) Gulhati, P.; Zaytseva, Y. Y.; Valentino, J. D.; Stevens, P. D.; Kim, J. T.; Sasazuki, T.; Shirasawa, S.; Lee, E. Y.; Weiss, H. L.; Dong, J.; Gao, T.; Evers, B. M. *Carcinogenesis* **2012**, *33*, 1782-1790.

(12) Hu, S.; Chen, Z.; Franke, R.; Orwick, S.; Zhao, M.; Rudek, M. A.; Sparreboom, A.; Baker, S. D. *Clin. Cancer Res.* **2009**, *15*, 6062-6069.

(13) Swift, B.; Nebot, N.; Proctor, W.; Thakker, D.; Lang, D.; Radtke, M.; Gnoth, M. J.; Brouwer, K. L. *AAPS J.* **2010**, *12*, S2.

(14) Dardonville, C.; Barrett, M. P.; Brun, R.; Kaiser, M.; Tanious, F.; Wilson, W. D. *J. Med. Chem.* **2006**, *49*, 3748-3752.

(15) Ming, X.; Ju, W.; Wu, H.; Tidwell, R. R.; Hall, J. E.; Thakker, D. R. *Drug Metab. Dispos.* **2009**, *37*, 424-430.

(16) www.dojindo.com.

(17) www.invitrogen.com.

(18) Huang, S.; Sinicrope, F. A. *Mol. Cancer Ther.* **2010**, *9*, 742-750.

(19) Kane, R. C.; Farrell, A. T.; Saber, H.; Tang, S.; Williams, G.; Jee, J. M.; Liang, C.; Booth, B.; Chidambaram, N.; Morse, D.; Sridhara, R.; Garvey, P.; Justice, R.; Pazdur, R. *Clin. Cancer Res.* **2006**, *12*, 7271-7278

(20) Shimizu, S.; Takehara, T.; Hikita, H.; Kodama, T.; Tsunematsu, H.; Miyagi, T.; Hosui, A.; Ishida, H.; Tatsumi, T.; Kanto, T.; Hiramatsu, N.; Fujita, N.; Yoshimori, T.; Hayashi, N. *Int. J. Cancer* **2012**, *131*, 548-557.

(21) Ruan, W.; Zhu, S.; Wang, H.; Xu, F.; Deng, H.; Ma, Y.; Lai, M. *Mol. Cancer* **2010**, *9*, 281.

(22) www.atcc.com.

(23) Suen, A. W.; Galoforo, S.; Marples, B.; McGonagle, M.; Downing, L.; Martinez, A. A.; Robertson, J. M.; Wilson, G. D. *Int. J. Radiat. Oncol.* **2010**, *78*, 213-220.

(24) Prahallad, A.; Sun, C.; Huang, S.; Di Nicolantonio, F.; Salazar, R.; Zecchin, D.; Beijersbergen, R. L.; Bardelli, A.; Bernards, R. *Nature* **2012**, *483*, 100-103.

(25) Cheng, S.; Koch, W. H.; Wu, L. *New Biotechnol.* **2012**, *29*, 682-688.

(26) Wan, P. T.; Garnett, M. J.; Roe, S. M.; Lee, S.; Niculescu-Duvaz, D.; Good, V. M.; Jones, C. M.; Marshall, C. J.; Springer, C. J.; Barford, D.; Marais, R. *Cell* **2004**, *116*, 855-867.

(27) Sánchez-Hernández, I.; Baquero, P.; Calleros, L.; Chiloeches, A. *Cancer Letters* **2012**, *314*, 244-255.

(28) Wilhelm, S. M.; Carter, C.; Tang, L.; Wilkie, D.; McNabola, A.; Rong, H.; Chen, C.; Zhang, X.; Vincent, P.; McHugh, M.; Cao, Y.; Shujath, J.; Gawlak, S.; Eveleigh, D.; Rowley, B.; Liu, L.; Adnane, L.; Lynch, M.; Auclair, D.; Taylor, I.; Gedrich, R.; Voznesensky, A.; Riedl, B.; Post, L. E.; Bollag, G.; Trail, P. A. *Cancer Res.* **2004**, *64*, 7099.

(29) Soule, H. D.; Vazquez, J.; Long, A.; Albert, S.; Brennan, M. *J. NatI. Cancer Inst.* **1973**, *51*, 1409-1416.

(30) Levenson, A. S.; Jordan, V. C. *Cancer Res.* **1997**, *57*, 3071-3078.

(31) Danforth, D. N.; Sgagias, M. K. *J. Endocrinol.* **1993**, *138*, 517-528.

(32) Musgrove, E. A.; Sutherland, R. L. *Nat Rev Cancer* **2009**, *9*, 631-643.

(33) Johnston, S. R.; Saccani-Jotti, G.; Smith, I. E.; Salter, J.; Newby, J. *Cancer Res.* **1995**, *55*, 3331-3338.

(34) Gutteridge, E.; Agrawal, A.; Nicholson, R.; Cheung, K. L.; Robertson, J. *Int. J. Cancer* **2010**, *126*, 1806-1816.

(35) Stone, A.; Valdés-Mora, F.; Gee, J. M. W.; Farrow, L.; McClelland, R. A.; Fiegl, H.; Dutkowski, C.; McCloy, R. A.; Sutherland, R. L.; Musgrove, E. A.; Nicholson, R. I. *PLoS ONE* **2012**, *7*, e40466.

(36) Ring, A.; Dowsett, M. *Endocr. Relat. Cancer* **2004**, *11*, 643-658.

(37) Gee, J. M. W.; Robertson, J. F. R.; Ellis, I. O.; Nicholson, R. I. *Int. J. Cancer* **2001**, *95*, 247-254.

(38) Fisher, D. E. *Cell* **1994**, *78*, 539-542.

(39) Saikumar, P.; Dong, Z.; Mikhailov, V.; Denton, M.; Weinberg, J. M.; Venkatachalam, M. A. *Am. J. Med.* **1999**, *107*, 489-506.

(40) Gallagher, R.; Collins, S.; Trujillo, J.; McCredie, K.; Ahearn, M.; S. Tsai, R. M.; Aulakh, G.; Ting, R.; Ruscetti, F.; Gallo, R. *Blood* **1979**, *54*, 713-733.

(41) Collins, S. J. *Blood* **1987**, *70*, 1233-1244.

(42) Nakabayanshi, H.; Taketa, K.; Miyano, K.; Yamane, T.; Sato, J. *Cancer Res.* **1982**, *42*, 3858-3863.

(43) Yuzugullu, H.; Benhaj, K.; Ozturk, N.; Senturk, S.; Celik, E.; Toylu, A.; Tasdemir, N.; Yilmaz, M.; Erdal, E.; Akcali, K.; Atabey, N.; Ozturk, M. *Mol. Cancer* **2009**, *8*, 90.

(44) www.cellbiolabs.com.

(45) http://www.anatomyone.com/a/epithelial-cells/.

(46) Porteus, M. *Biotechn. Genet. Eng.* **2007**, *24*, 195-212.

(47) Litwin, V.; Marder, P. *Flow cytometry in drug discovery and development*; Wiley, 2011.

(48) Givan, A. L. *Flow cytometry: first principles*; 2nd ed.; Wiley-Liss, 2001.

(49) Pawaskar, D.; Straubinger, R.; Fetterly, G.; Ma, W.; Jusko, W. *Pharm Res* **2012**, 1-7.

(50) Cruet-Hennequart, S.; Villalan, S.; Kaczmarczyk, A.; O'Meara, E.; Sokol, A. M.; Carty, M. P. *Cell Cycle* **2009**, *8*, 3043 - 3054.

(51) Ling, X.; Bernacki, R. J.; Brattain, M. G.; Li, F. *J. Biol. Chem.* **2004**, *279*, 15196-15203.

(52) Jacobson, M. D.; Carthy, N. M. *Apoptosis the molecular biology of programmed cel death*; Oxford, 2002.

(53) Boulares, A. H.; Yakovlev, A. G.; Ivanova, V.; Stoica, B. A.; Wang, G.; Iyer, S.; Smulson, M. *J. Biol. Chem.* **1999**, *274*, 22932-22940.

(54) http://howtowesternblot.net/welcome/western-blotting-101/.

(55) Knippschild, U.; Gocht, A.; Wolff, S.; Huber, N.; Löhler, J.; Stöter, M. *Cellular Signalling* **2005**, *17*, 675-689.

(56) Price, M. A. *Genes Dev.* **2006**, *20*, 399-410.

(57) Ali, A.; Hoeflich, K. P.; Woodgett, J. R. *Chem. Rev.* **2001**, *101*, 2527-2540.

(58) Luo, J. *Cancer Lett.* **2009**, *273*, 194-200.

(59) http://www.cerep.fr/Cerep/Users/index.asp.

(60) Jia, Y.; Quinn, C. M.; Kwak, S.; Talanian, R. V. *Curr. Drug Discov. Tech.* **2008**, *5*, 59-69.

(61) Lebakken, C. S.; Riddle, S. M.; Singh, U.; Frazee, W. J.; Eliason, H. C.; Gao, Y.; Reichling, L. J.; Marks, B. D.; Vogel, K. W. *J. Biomol. Screen.* **2009**, *14*, 924-935.

(62) http://tools.invitrogen.com/content/sfs/manuals/BRAF_LanthaScreen_Binding.pdf.

(63) Karaman, M. W.; Herrgard, S.; Treiber, D. K.; Gallant, P.; Atteridge, C. E.; Campbell, B. T.; Chan, K. W.; Ciceri, P.; Davis, M. I.; Edeen, P. T.; Faraoni, R.; Floyd, M.; Hunt, J. P.; Lockhart, D. J.; Milanov, Z. V.; Morrison, M. J.; Pallares, G.; Patel, H. K.; Pritchard, S.; Wodicka, L. M.; Zarrinkar, P. P. *Nat. Biotechnol.* **2008**, *26*, 127-132.

(64) Chong, H.; Vikis, H. G.; Guan, K.-L. *Cell. Signal.* **2003**, *15*, 463-469.

(65) Davies, M. A. *J. Clin. Oncol.* **2011**, *29*, 4715-4717.

(66) Faissner, A.; Heck, N.; Dobbertin, A.; Garwood, J. *Adv. Exp. Med. Biol.* **2006**, *557*.

Chapter 5 – Computational Modelling

5.1. Introduction

With the knowledge that many of our compounds exhibit low micromolar cytotoxicity in a variety of cell types, it became essential to investigate their precise target, to fully understand their mechanism of action. It is notoriously difficult to elucidate the principal target of a ligand acting as a kinase inhibitor given the large number of distinct, yet highly homologous and structurally similar, kinases which exist. The most desirable scenario is that experimental evidence can be found to point to the kinase being targeted by the specific derivative; however, screening of a compound against many different specific kinases involves a large cost, so a combination of experimental (described in Chapter 4) and theoretical data was used to attempt to elucidate the kinase being targeted by our compounds.

Computational modelling techniques play an important role in drug design, as well as in studies of the interaction between a ligand and its target. Particularly in conjunction with experimental results, *molecular docking* has been widely used to characterise ligand-target interactions and to suggest ligands which could have increased affinity for that target. Further to this, *molecular dynamics simulations* allow ligand-target complexes to be studied over time and to give detailed information on the most important interactions (pairwise analysis) and the stability of the complex (free energy of binding calculations). In this study, done in collaboration with Dr. Brendan Kelly (TCD) and Prof. Jaime Rubio Martinez at the Department of Physical Chemistry (Universidad de Barcelona, Spain), the MOE[1] (Molecular Operating Environment) software was used for docking and visualization of systems and the AMBER 12.0[2] software was used for optimization of docked poses and molecular dynamic simulations.

In our case, several pieces of experimental data helped to narrow down the potential targets of the compounds and allowed a directed molecular modelling study. Initially, the compounds prepared in this project were predicted to act in a similar fashion to sorafenib, given their structural similarity. Therefore, several kinases targeted by sorafenib such as B-RAF, C-RAF, MEK-1, ERK, p38 MAPK and VEGFR were experimentally studied (see

section 4.4.2). From all the results obtained, the most significant was the 74-99% inhibition of the C-RAF/MEK-1 pathway by our compounds from Family 4A. This showed that the potential target of our compounds could be C-RAF, MEK-1 or both. Other results obtained in this study showed that these compounds did not strongly inhibit ERK, p38 MAPK or VEGFR. Additionally, a FRET study to evaluate the inhibitory effect of our compounds on B-RAF was carried out (see section 4.4.3) showing that the compounds did not competitively bind to this kinase. Finally, the effect of a representative compound (**73a**) on the expression levels of various signaling proteins from the RAF/MEK/ERK pathway was studied. The results confirmed that this compound did not target B-RAF and that it decreased the expression levels of C-RAF, MEK and ERK. The latter was decreased as a consequence of the upstream inhibition of the pathway. Thus, the possible targets for these compounds – which are known to inhibit the RAF/MEK/ERK pathway – were narrowed down to C-RAF and MEK-1. At this point a theoretical study was proposed to help to elucidate the targeted kinase and, also, to propose future families of compounds with enhanced affinity, bearing in mind that the target may turn out to be different to those of sorafenib.

5.2. Comparison of B-RAF, C-RAF and MEK-1

Several ligand-protein crystal structures exist for B-RAF, C-RAF and MEK-1. It is instructive to view these systems together to gain an insight into both their similarities and differences, particularly in the ATP-binding domain and neighbouring regions. The choice of crystal structure for each system was made based on a variety of criteria. Firstly, crystals with structurally similar ligands to our compounds were given precedence. Secondly, all kinases of this type have a flexible loop, which is not resolved in their crystal structure and needs to be modeled using homology modelling (see below); thus, those crystal structures with the shortest gap in tertiary structure were preferred. Thirdly, the resolution of the crystal structure was used as the final criterion to decide the system used. Resulting from this, PDB entry 1UWH (B-RAF/sorafenib)[3] was chosen to represent B-RAF, 3OMV (C-RAF/SM5)[4] was selected for C-RAF, and 3PP1 (MEK-1/IZG/ATP/Mg^{2+})[5] was selected for the study of MEK-1. Each system will be discussed in turn; however, Figure 5.1 shows an overlap of 1UWH, 3OMV and 3PP1, including ATP but without the corresponding ligands to show the high degree of similarity among these three enzymes.

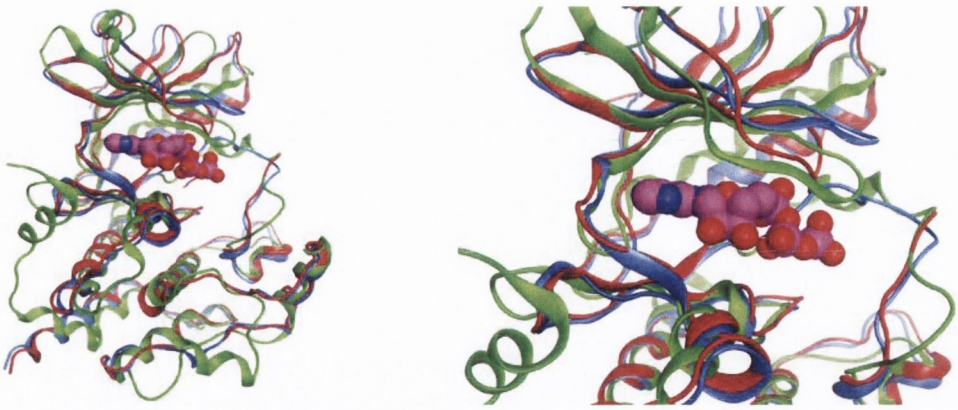


Figure 5.1. Overlaid crystal structures of B-RAF (blue), C-RAF (red) and MEK-1 (green), with ATP (pink) bound to the ATP-site (*left*), and zoomed view of ATP site (*right*).

These three protein kinases are highly conserved as they belong to the same branch of the kinome; in fact, their overall tertiary structures are mapped almost identically and the ATP binding site is located in the same place for each system. However, when the amino acid sequences – and the sites where ligands bind – are compared, differences emerge. A sequence alignment of MEK-1, B-RAF and C-RAF performed using the ClustalW2 multiple sequence alignment server,[6] in which amino acids involved in ligand and ATP binding regions are highlighted, is shown in Figure 5.2.

As expected since B-RAF and C-RAF are two isoforms of RAF, they maintain sequence similarities throughout, even in the amino acids involved in binding to ATP and ligands. However, MEK-1 diverges significantly from both B-RAF and C-RAF. The first major difference is that there is a large, flexible loop present in MEK-1 which is absent in both B-RAF and C-RAF. As will be seen later in the modeled proteins, this loop is free to extend into the space near the ATP binding site. However, the most significant differences occur in the amino acids involved in ATP and ligand binding. A visual inspection reveals that, while the sequences of B-RAF and C-RAF are closely correlated, that of MEK-1 is significantly different and these differences are borne out in the position within the protein that ligands bind.

```
MEK-1 ----MELKDDDFEKISELGAGNGGVVFKVSHKPSGLVMARKLIHLE
CRAF  DDWEIPDGQITVGQRIGS----GSFGTVYKGKWHGDVAVKMLNVTA
BRAF  YYWEIEASEVMLSTRIGS----GSFGTVYKGKWHGDVAVKILKVVD

MEK-1 IKPAIRNQIIRELQVLHECNSPYIVGFYGAFYSDGEISICMEHMDG
CRAF  PTPQQLQAFKNEVGVLRKTRHVNILLFMGYSTAP-QLAIVTQWCEG
BRAF  PTPEQFQAFRNEVAVLRKTRHVNILLFMGYMTKD-NLAIVTQWCEG

MEK-1 GSLDQVLK-KAGRIPEQILGKVSIAVIKGLTYLREKHKIMHRDVKP
CRAF  SSLYHHLHIIETKFEMIKLIDIARQTAQGMDYLHA-KSIIHRDLKS
BRAF  SSLYKHLHVQETKFQMFQLIDIARQTAQGMDYLHA-KNIIHRDMKS

MEK-1 SNILVNSRGEIKLCDFGVSGQL----IDSMANSFVGTRSYMSPERL
CRAF  NNIFLHEDLTVKIGDFGLATEKSRWSGSHQFEQLSGSILWMAPEVI
BRAF  NNIFLHEGLTVKIGDFGLATVKSRWSGSQQVEQPTGSVLWMAPEVI

MEK-1 QG---THYSVQSDIWSMGLSLVEMAVGRYPIPPPDAKELELMFMFG
CRAF  RMQDKNPYSFQSDVYAFGIVLYELMTGQLPYS--------------
BRAF  RMQDNNPFSFQSDVYSYGIVLYELMTGELPYS--------------

MEK-1 CQVEGDAAETPPRPRTPGRPLSSYGMDSRPPMPMAIFELLDYIVN-
CRAF  -------------------------------NINNRDQIIFMVGRG
BRAF  -------------------------------HINNRDQIIFMVGRG

MEK-1 EPPPKLPS—GVFSLEFQDFVNKCLIKNPAERADLKQLMVHAFI-KR
CRAF  YLSPDLSKVRSNCPKAMKRLMAECLKKKRDERPLFPQILAS--IEL
BRAF  YASPDLSKLYKNCPKAMKRLVADCVKKVKEERPLFPQILSS--IEL

MEK-1 SDAEEVD
CRAF  LARSLPK
BRAF  LQHSLPK
```

Figure 5.2. Sequence alignment of B-RAF, C-RAF and MEK-1. Highlighted amino acids are those involved in the active site regions.

Both B-RAF and C-RAF share a similar binding site, as is demonstrated by the fact that sorafenib binds to both proteins with similar affinity (B-RAF: IC_{50} = 22 ± 6 nM; C-RAF:IC_{50} = 6 ± 3 nM).[7] The ligands that bind to B-RAF and C-RAF are normally ATP-competitive, so they prevent the ATP from binding to its specific site in the protein by binding themselves. Figure 5.3 illustrates the regions in which ligands bind to these

proteins as molecular surfaces, where pink represents areas of ligand hydrophilicity and green represents areas of ligand lipophilicity – in each case ATP is maintained in the same coordinates for reference.

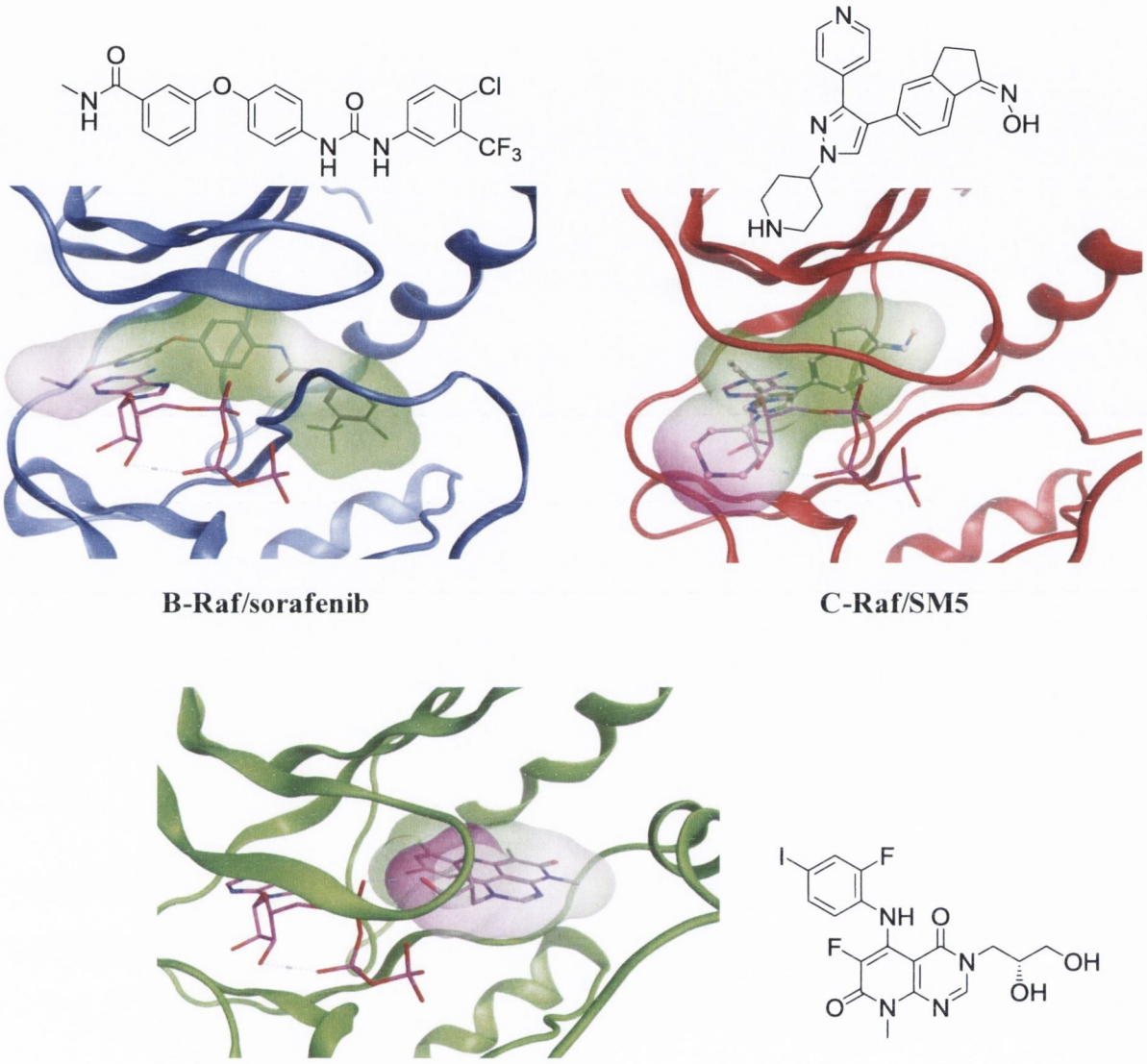


Figure 5.3. Molecular surfaces showing binding sites of ligands with their structures in B-RAF, C-RAF and MEK-1, respectively. Pink represents areas of ligand hydrophilicity and green represents areas of ligand lipophilicity. ATP is shown for reference.

As can be seen from Figure 5.3, in B-RAF, sorafenib places its acetamide group and the pyridyl ring into the space where ATP binds, preventing this from occurring and, hence, inhibiting the kinase. Sorafenib also extends into a hydrophobic pocket in the protein with its (4-Cl-3-CF_3)-Ph ring (see section 1.6.3). No crystal structure exists for sorafenib in complex with C-RAF; however, the ligand in this case (SM5) occupies the same space as sorafenib, preventing ATP binding to the protein and inhibiting its kinase activity. As SM5 is a smaller ligand it does not extend into the hydrophobic pocket. A clear difference is observed in the crystal structure of MEK-1 with IZG since ATP is present in the crystal structure and it is necessary for the binding of IZG to MEK-1.[8] This type of MEK inhibitor is called an allosteric inhibitor, as opposed to an ATP-competitive inhibitor. The ligand IZG places a hydrophobic aryl group into a hydrophobic region which is near, but not exactly the same as, that occupied by sorafenib. It does not occupy the same binding pocket as sorafenib, which extends into the region of the ATP-binding site occupied by the adenosine moiety. Instead, IZG interacts with a phosphate group of ATP through a hydroxyl group (Figure 5.3). It is assumed that this prevents the kinase substrate, ERK, from becoming phosphorylated, also inhibiting the kinase activity of MEK-1.

Given the structural and functional similarities of B-RAF and C-RAF, and our experimental data showing that our compounds do not inhibit B-RAF, we may suggest that they do not inhibit C-RAF either, as ATP-competitive inhibitors. Our compounds could possibly target other domains of C-RAF, as in western blot studies **73a** appeared to decrease C-RAF expression levels slightly (see section 4.4.4), but this fact needs to be further investigated. However, it is clear that our compounds inhibit the C-RAF/MEK-1 pathway and it seems reasonable to predict that the likely target of our compounds is MEK-1. Further evidence to suggest this came from the interaction observed between a hydroxyl group of IZG and a phosphate group of ATP, since the guanidinium cation present in our compounds would be able to undergo strong electrostatic interactions with this phosphate group. Additionally, western blot studies showed a significant decrease in the expression levels of MEK-1 after exposure to increasing concentrations of compound **73a** (see section 4.4.4).

5.3. Docking of Synthesised Compounds to Systems of Interest

With these experimental data in hand, and an indication of the potential targets of our ligands, docking studies were carried out for some of our compounds in models of B-RAF, C-RAF and MEK-1.

Similar induced-fit docking protocol (see Experimental section 7.3) was used to dock a series of ligands into the crystal structures of B-RAF, C-RAF and MEK-1, as an initial screen of the complexes found. The set of ligands decided on (Figure 5.4) were those present in the crystal structures (sorafenib, SM5 and IZG), those for which data on the inhibition of the C-RAF/MEK-1 pathway were available (**49a**, **51a**, **51d**, **70a**, **71a-d**, **79a** and **79d**) and two further molecules (**73a** and **73c**), studied to predict their binding affinities.

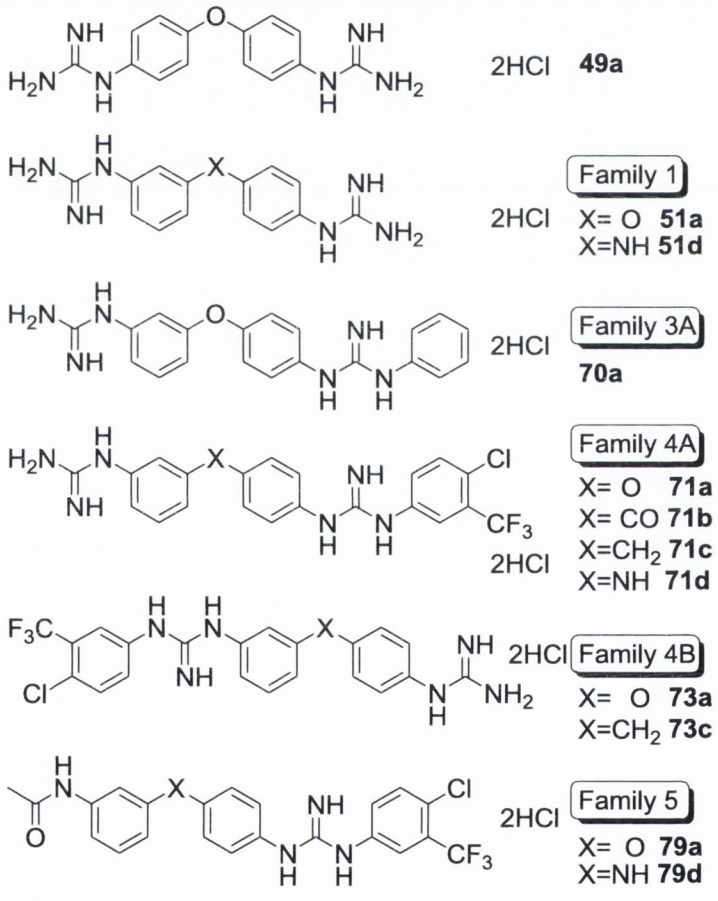


Figure 5.4. Structures of the synthesised ligands chosen for docking study: **49a**, **51a**, **51d**, **70a**, **71a-d**, **73a**, **73c**, **79a** and **79d**.

Given that we had already proven experimentally that our compounds did not bind to B-RAF, and having observed the similarities between B-RAF and C-RAF in the sequence analysis, it was expected that the results of docking compounds into these proteins would corroborate these facts. This turned out to be the case, with no observable correlations between the ranking of docking poses (orientation of the ligand in the protein after docking and refinement) and the inhibition of the C-RAF/MEK-1 pathway. In fact, the docking scores (energy values assigned by scoring functions, representative of ligand binding affinity) for all ligands were indicative of unfavourable interactions and the orientation of the ligands in the protein site was not consistent. As a result, these systems were not suitable candidates to be studied further.

On the other hand, the behaviour of our ligands in the MEK-1 system (3PP1) was much more consistent; a broad correlation existed between experimental inhibition of the C-RAF/MEK-1 pathway and the ranking of ligands after docking. As well as this, the top five poses were highly similar in orientation, an indication of a stable complex being formed. Possibly of most significance, the stable poses incorporate a highly stabilising electrostatic interaction between a phosphate group of ATP and a guanidinium from the ligand, as well as placing the lipophilic (4-Cl-3-CF_3)-Ph group in the lipophilic pocket of the protein. These two interactions suggest a reason for our compounds targeting MEK-1. A detailed description of this docking study, and of subsequent minimisation simulations, is presented in the following section.

5.4. Docking and Minimisation of Synthesised Compounds in MEK-1 (3PP1)

5.4.1. Protein Model Preparation

Prior to the docking study, the flexible loop in the MEK-1 protein, which was unresolvable by X-ray crystallography, had to be modelled by homology modelling to have a complete protein, which is required for optimization of docked poses and molecular dynamic simulations. This was carried out using the Modeller program,[9] with the 3PP1 X-ray crystal structure as template and the human sequence of MEK-1 (Q02750), obtained from the Uniprot server.[10] During the modelling, only the loop region was modelled, the rest of the protein being fixed in the crystal orientation. Five models were generated, with the best one chosen based on its DOPE (discreet optimised protein geometry) score.[11] Then, the completed protein structure, including ATP, the ligand (IZG) and Mg^{2+}, was solvated, neutralised and subjected to a seven step minimisation in AMBER 12.0 to remove any bad contacts which may have existed between atoms or residues (see experimental section 7.3). The resulting system was used for all subsequent docking and minimisations.

5.4.2. Docking of Crystal Structure Ligand and Novel Compounds into MEK-1 Minimised Structure

The docking procedure, as described in the experimental section 7.3, was applied using the previously generated conformers of the 12 synthesised compounds and the ligand from crystal structure 3PP1, IZG (see experimental section 7.3). Both the ATP and Mg^{2+} present in the crystal structure were included as part of the protein and the ligand binding site was defined by IZG.

The ranking of molecules from this docking (higher negative energy docking scores mean a stronger complex between the ligand and the protein) corresponds relatively well with the experimental inhibition data of the C-RAF/MEK-1 pathway (Table 5.1). However, in many of the top-ranked poses the ligand has been placed outside of the expected binding site due to size restrictions, or to an electrostatic interaction between the guanidinium and a phosphate group from ATP, which prevents proper orientation of the ligand into the pocket. There is also a very small spread of scores across the entire series of compounds,

where it was expected that larger differences would be observed given the broad range of experimental inhibition of the C-RAF/MEK-1 pathway obtained (2-99%).

Table 5.1. Docking of novel compounds and IZG into minimised MEK-1 structure.

Compound	Experimental C-RAF/MEK-1 % Inhibition	Docking Score (kcal mol^{-1})	Docking Rank
49a	2%	-33.9	9
51a	22%	-31.4	11
51d	86%	-31.8	10
70a	4%	-36.2	6
71a	96%	-35.3	7
73a	-	-39.3	2
71c	99%	-38.0	4
73c	-	-39.4	1
71b	96%	-31.2	12
71d	74%	-30.5	13
79a	35%	-34.6	8
79d	38%	-39.2	3
IZG	-	-36.8	5

A major problem with the docking scores obtained is that the corresponding poses are not consistent and it seems that the docking site did not accommodate the ligands as well as expected. In an attempt to overcome the known limitations of the docking protocol – namely that poses are heavily dependent on the initial placement of the ligand and that not enough protein flexibility exists to allow for ligands which differ significantly from the crystal structure ligand – the top 20 poses for each ligand were taken forward to be minimised in AMBER 12.0. The reason for this approach was that unfavoured contacts or conformations generated by docking would be removed and that, should different poses evolve to a common final orientation, we would have a good indication of the most favoured ligand-protein complexes. Furthermore, the protein would have more freedom to adapt to the ligands and give complexes with more sensible orientations.

Another reason for performing this minimisation is that it would allow the MMPBSA (Molecular Mechanic Poisson-Boltzmann Surface Area) method to be applied, which is a

better indication of the free energy of binding than docking scoring functions, as it considers solvation as well as ligand-protein interactions.

5.4.3. MMPBSA Analysis of Ligand-Protein Complexes

The top 20 poses for each ligand were solvated using the Leap tool in AMBER 12.0 and subjected to a three step minimisation as follows: 1) initially, the protein and ligand were fixed with a force constant of 5.0, the newly introduced water molecules being relaxed, 2) the main chain of the protein and the structure of the ligand were fixed with a force constant of 5.0, while the side chains and water molecules were minimised, and 3) the entire system was minimised with no restraints in place, allowing the ligand to adopt favoured orientations.

An MMPBSA analysis was then carried out to assess the binding free energy of the complexes.[12] The aim of MMPBSA is to calculate the free energy difference between the solvated unbound protein and ligand, and the solvated complex of the two. In practise, this is carried out by calculating the solvation energy of each component individually (the ligand, protein and complex), and then the binding energy of the ligand and protein in vacuum (Figure 5.5). The solvated binding free energy is not calculated directly as the largest contribution would be from solvent-solvent interactions and fluctuations in total energy would be an order of magnitude higher than binding energy.

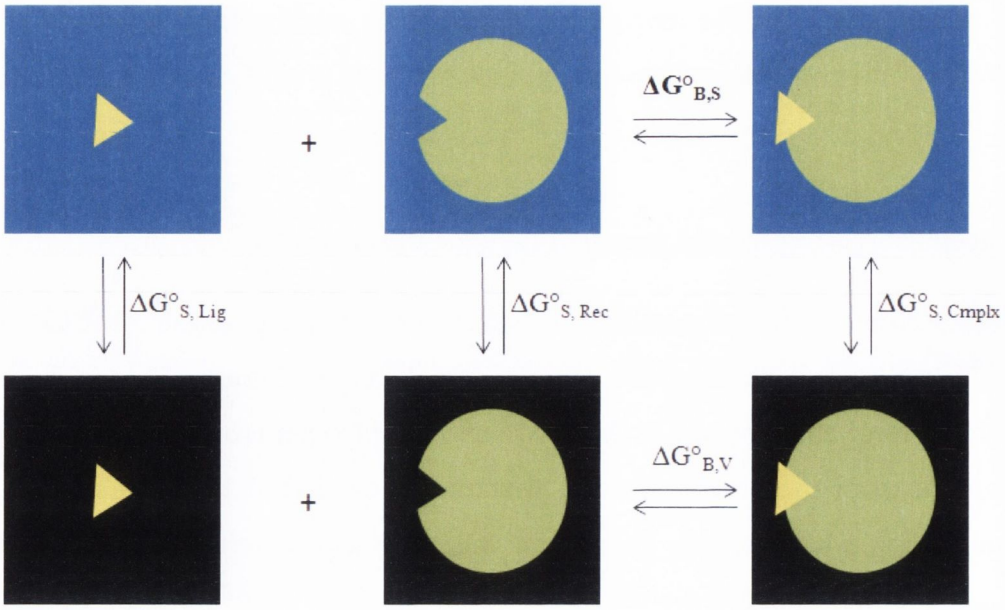


Figure 5.5. The MMPBSA method of solvated binding free energy calculates the solvation energy of each component and the binding energy of the ligand and protein in vacuum: S = solvation, B = Binding, V = Vacuum.

From this scheme, it is clear that the solvated binding free energy for the complex can be calculated from the following equation:

$$\mathbf{\Delta G^{\circ}_{B,S}} = \Delta G^{\circ}_{B,V} + \Delta G^{\circ}_{S,Cmplx} - (\Delta G^{\circ}_{S,Lig} + \Delta G^{\circ}_{S,Rec})$$

Compounds within a given family are highly similar (differing only in the linker group between the two aryl rings) and, thus, differences in binding free energies are small, making comparisons of ligands unreliable. However, a comparison across families reveals large differences in binding free energies and replicates well the experimental inhibition data for the C-RAF/MEK-1 pathway. The binding free energies (Table 5.2) also indicate that strong complexes are being formed and that some ligands fit very well into the MEK-1 binding site. Furthermore, of the 20 docking poses examined for each ligand, the majority evolved to the same orientation during the minimisation process, giving a strong indication that the resulting complexes are highly favoured.

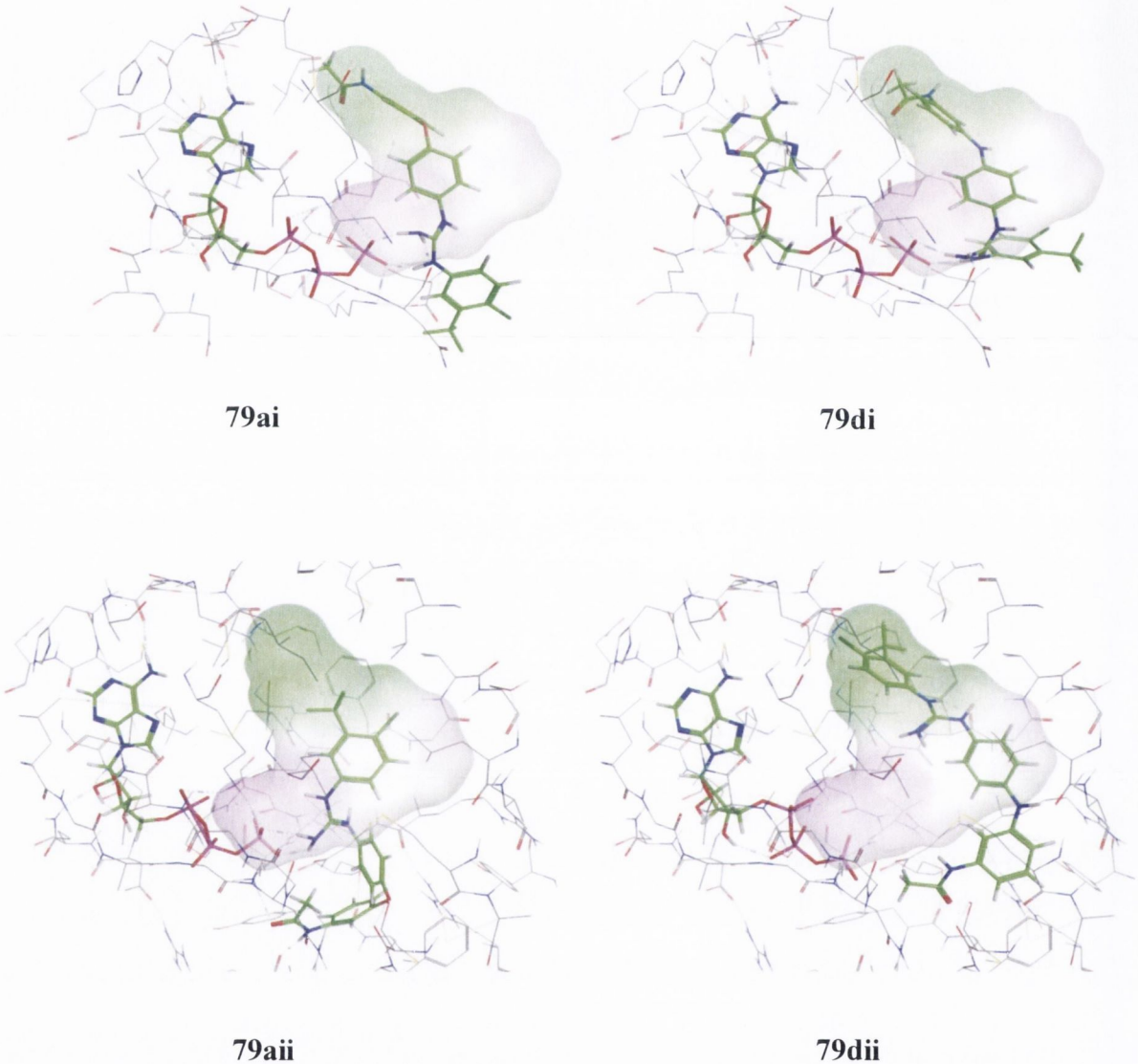
Table 5.2. Binding free energies calculated by the MMPBSA method.

Compound	Experimental C-RAF/MEK-1 % Inhibition	$\Delta G°_{B,S}$ (kcal mol^{-1})	Rank
49a	2%	-63.9	13
51a	22%	-72.4	9
51d	86%	-69.0	11
70a	4%	-73.7	8
71a	96%	-71.4	10
73a	-	-76.8	6
71c	99%	-67.4	12
73c	-	-77.4	5
71b	96%	-75.3	7
71d	74%	-80.3	4
79a	35%	-81.9	2
79d	38%	-81.1	3
IZG	-	-96.5	1

All the ligands studied can be classified into several groups. First of all, the crystal ligand IZG has the highest binding affinity. This is expected as the binding site of the protein is adapted to this ligand in the crystal structure. Even though the series of minimisations aims to diminish this effect, it is always likely to occur and is a good indication that the analysis is reliable.

The first group of ligands to discuss present an interesting case as they give an insight into the likely biological mode of binding. Family 5 (compounds **79a** and **79d**) led to complexes which had among the highest binding energies, yet they had poor experimental inhibition of the C-RAF/MEK-1 pathway. However, a closer inspection of the resulting poses suggests a reason for their poor experimental performance. All of the top 10 poses for each ligand (and the vast majority of the 20 poses) adopt an orientation driven by the electrostatic interaction between the guanidinium and the phosphate of ATP (Figure 5.6, **79ai** and **79di**). This necessarily causes two unfavourable events: 1) the (4-Cl-3-CF_3)-Ph group cannot now be placed in the lipophilic pocket, and 2) the polar amide group is now in this pocket. It is known that the lipophilic interaction between the receptor and the (4-Cl-3-CF_3)-Ph group is a driving force for binding; however, any poses in which such an interaction is present have poor connection to ATP. In these poses, the amide is not

interacting with the phosphate groups, not even by the available HB through its N-H, which in any case would be weaker than the electrostatic interaction provided by a guanidinium (Figure 5.6, **79aii** and **79dii**).



79ai **79di**

79aii **79dii**

Figure 5.6. Principal pose of ligands **79a** and **79d** after minimisation in MEK-1 (i) as well as lowest energy poses with (4-Cl-3-CF_3)-Ph group in pocket (ii). The surface is that generated by IZG; pink represents areas of ligand polarity and green represents areas of ligand hydrophobicity. The ligand and ATP are shown in bold.

When the poses, which place the lipophilic (4-Cl-3-CF_3)-Ph group in the lipophilic pocket are examined, a more likely energy of binding for these compounds is obtained. The energy of **79aii**, shown in Figure 5.6, is -55.3 kcal mol^{-1}, while that of **79dii** is -51.9 kcal mol^{-1}. This would rank these molecules as the lowest studied and indicates that they are not suited to inhibit MEK-1 at this site. Given their partial experimental inhibition of the C-RAF/MEK-1 pathway, and their structural differences to the rest of our molecules, it is possible that they are binding elsewhere to either C-RAF or MEK-1. Also, it is likely that their main target could be another kinase or protein as they possess high cytotoxic activity in a wide range of cancer cell lines. This will require further investigation.

It is important to note that the binding of molecules to this site of MEK-1 is driven and directed by the lipophilic interaction inside the pocket. This requires the molecules to enter the site *via* a lipophilic group and, for this reason, those poses which incorporate this interaction are given precedence in this study. With that in mind, all energies given in Table 5.2 are for the highest binding energy pose which incorporated this lipophilic interaction. For almost all molecules, this was either the most stable complex, or was within 2.0 $kcalmol^{-1}$ of the most stable complex. The only exceptions to this were molecules **79a** and **79d**, already discussed; thus, for these molecules, the energy provided in Table 5.2 is that of the most stable complex.

The next group of ligands in the ranking comprises **49a**, **51a**, **51d** and **70a** (Families 1 and 3), which showed generally poor to moderate inhibition of the C-RAF/MEK-1 pathway. This can be rationalised by the fact that these compounds only maintain the electrostatic interaction between the guanidinium and the phosphate of the ATP (Figure 5.7), whereas, Family 1 does not have the lipophilic interaction as they have another charged guanidinium at the end of the molecule occupying the space of the lipophilic aryl group. In terms of binding energy, these compounds have higher energy than the compounds discussed from Family 5, as they have an electrostatic interaction, which is absent in the molecules of Family 5.

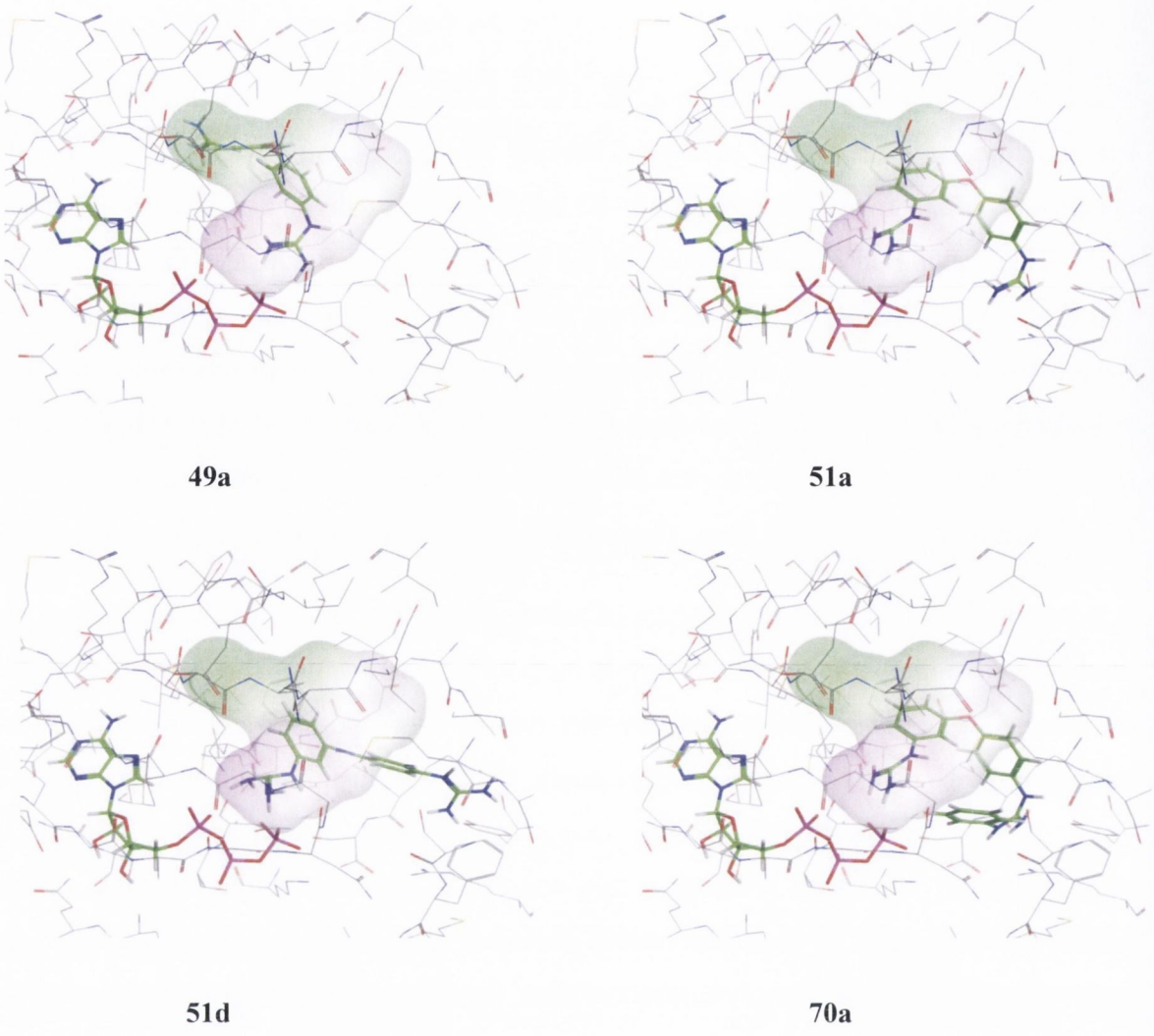


Figure 5.7. Principal pose of ligands **49a**, **51a**, **51d** and **70a** after minimisation in MEK-1. The surface is that generated by IZG; pink represents areas of ligand polarity and green represents areas of ligand hydrophobicity. The ligand and ATP are shown in green.

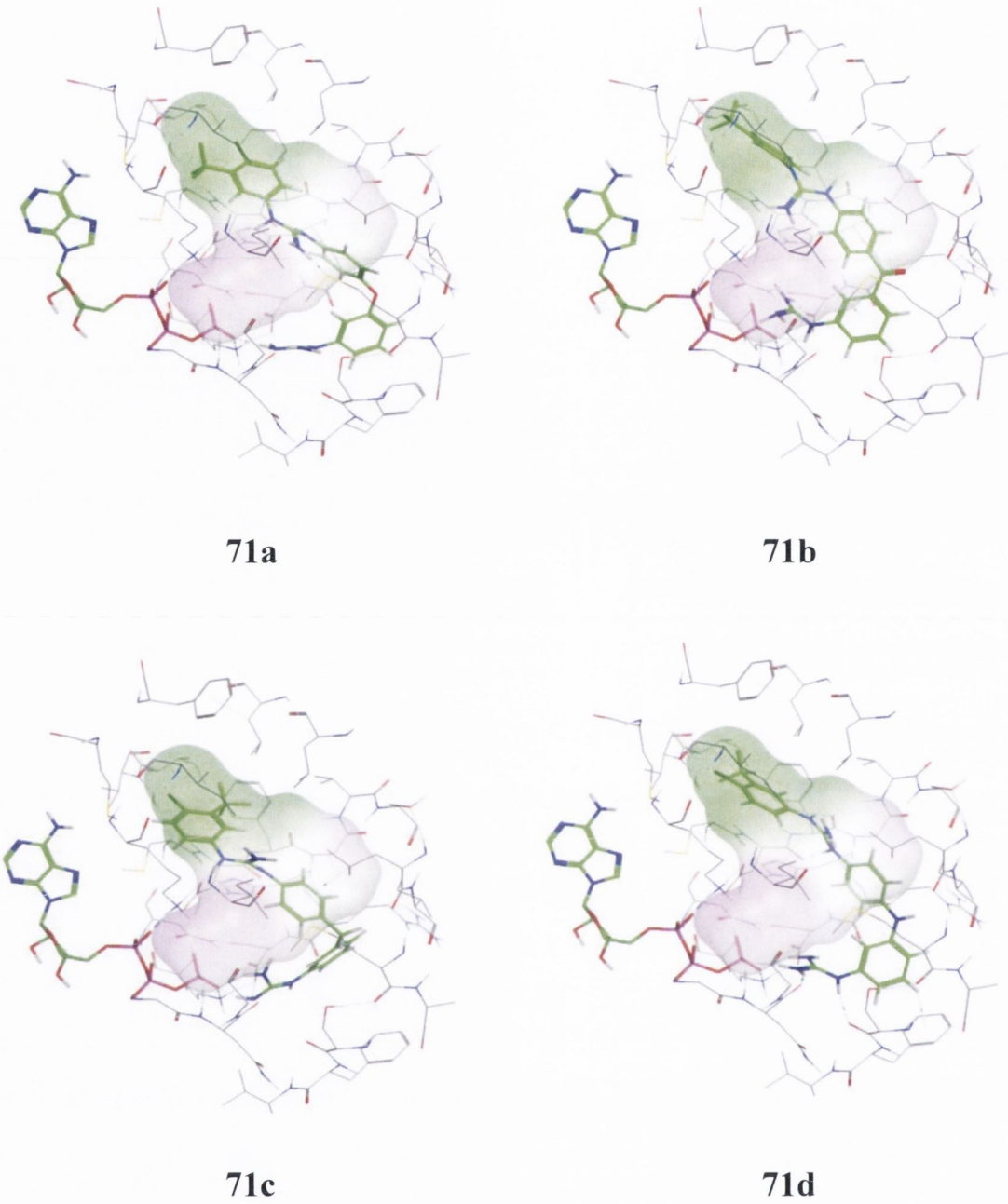
In these poses – which were replicated by the majority of the top 10 ranked poses – only ligand **49a** places the guanidinium group within the binding pocket, which will not undergo favourable interactions. The geometry of the remaining molecules is not suited to both incorporate a guanidinium-phosphate interaction and occupy the pocket; even ligand **70a** is unable to put its lipophilic phenyl ring in the pocket.

The remaining molecules belong to Family 4A and 4B (**71a-d**, **73a** and **73c**) which have the highest binding energies of the study. As seen in Table 5.2, they follow the same trends as those observed experimentally and, in the case of untested **73a** and **73c**, provide

encouraging evidence that higher affinity molecules can be obtained. All molecules are extremely similar; each contains the (4-Cl-3-CF_3)-Ph substituted guanidinium group at one end of the molecule and an unsubstituted guanidinium at the other end. Molecules with four different linkers were examined and due to the electronic and conformational changes brought about by these groups, small differences are observed in free energies of binding. Furthermore, compounds **73a** and **73c** provide useful information on the preferred geometry of this type of molecule for binding to the MEK-1 pocket as they have the (4-Cl-3-CF_3)-Ph guanidinium on the 1,3-disubstituted phenyl ring instead of at the 1,4-disubstituted one as in **71a** and **71c**. This 1,3-substitution could be the reason because **73a** and **73c** showed higher cytotoxic effects than the 1,4-subtituted **71a** and **71c** in several biochemical studies.

Such structural differences are so small that it is not possible to replicate their experimental inhibition data using molecular modelling techniques; hence, a detailed analysis of the free energies of binding would not provide any insights. However, useful information can be obtained as to the preferred orientation of the molecules, particularly when a large number of docking poses (20) are examined. Thus, in this group of compounds, **71a-d** can be compared among themelves, while both **73a** and **73c** can be compared to their analogous **71a** and **71c**.

Taking the molecules with the same substitution patterns, differing only in the linker group between the phenyl rings, the effects on their geometry are the most important, as the linker group itself does not become involved in any non-bonding interaction. At least half of the most favoured poses for all of these molecules place the substituted guanidinium interacting with the phosphate group of ATP, necessarily positioning the unsubstituted guanidinium in the lipophilic pocket. Given the structural similarities of our molecules to the crystal ligand, IZG, it must be assumed that this is not the actual mode of binding. However, the remaining poses are as expected and each compound has a desirable pose which has a free energy of binding that is within 2.0 kcal mol^{-1} of the top-ranked pose (Figure 5.8). Thus, these poses were chosen for the ranking of molecules presented in Table 5.2.



71a **71b**

71c **71d**

Figure 5.8. Principal pose of ligands **71a-d** after minimisation in MEK-1. The surface is that generated by IZG; pink represents areas of ligand polarity and green represents areas of ligand hydrophobicity. The ligand and ATP are shown in green.

All four ligands occupy the lipophilic pocket with the (4-Cl-3-CF_3)-Ph ring and incorporate the guanidinium-phosphate electrostatic interaction. For these reasons, they are the top-ranked compounds, with the highest binding energies of the ligand-protein complex. The tetrahedral linkers (O, CH_2) and the pyramidal trigonal one (NH) have highly similar poses

which appear relatively strained and are too large to occupy the same space as the crystal ligand, IZG. The carbonyl linker most effectively locates the (4-Cl-3-CF_3)-Ph ring into the lipophilic pocket and also maintains the guanidinium-phosphate electrostatic interaction; however, in order to relieve repulsion between aryl hydrogens, the 1,3-substituted phenyl ring must be placed outside the space occupied by IZG.

Further interesting comparisons can be drawn when ligands **71a** and **71c** are compared with ligands **73a** and **73c** (Figure 5.9). Both **73a** and **73c** adopt a very similar geometry that occupies the crystal structure ligand binding pocket more effectively than **71a** and **71c**. Encouragingly, the majority of the top poses of **73a** and **73c** adopt this arrangement and a large difference in binding free energy exists between these poses and those placing the unsubstituted guanidinium in the lipophilic pocket. It seems that having the (4-Cl-3-CF_3)-Ph substituted guanidinium on the 1,3-substituted phenyl ring is favourable for binding to MEK-1 and it will be interesting to see how they perform in specific assays against the C-RAF/MEK-1 pathway.

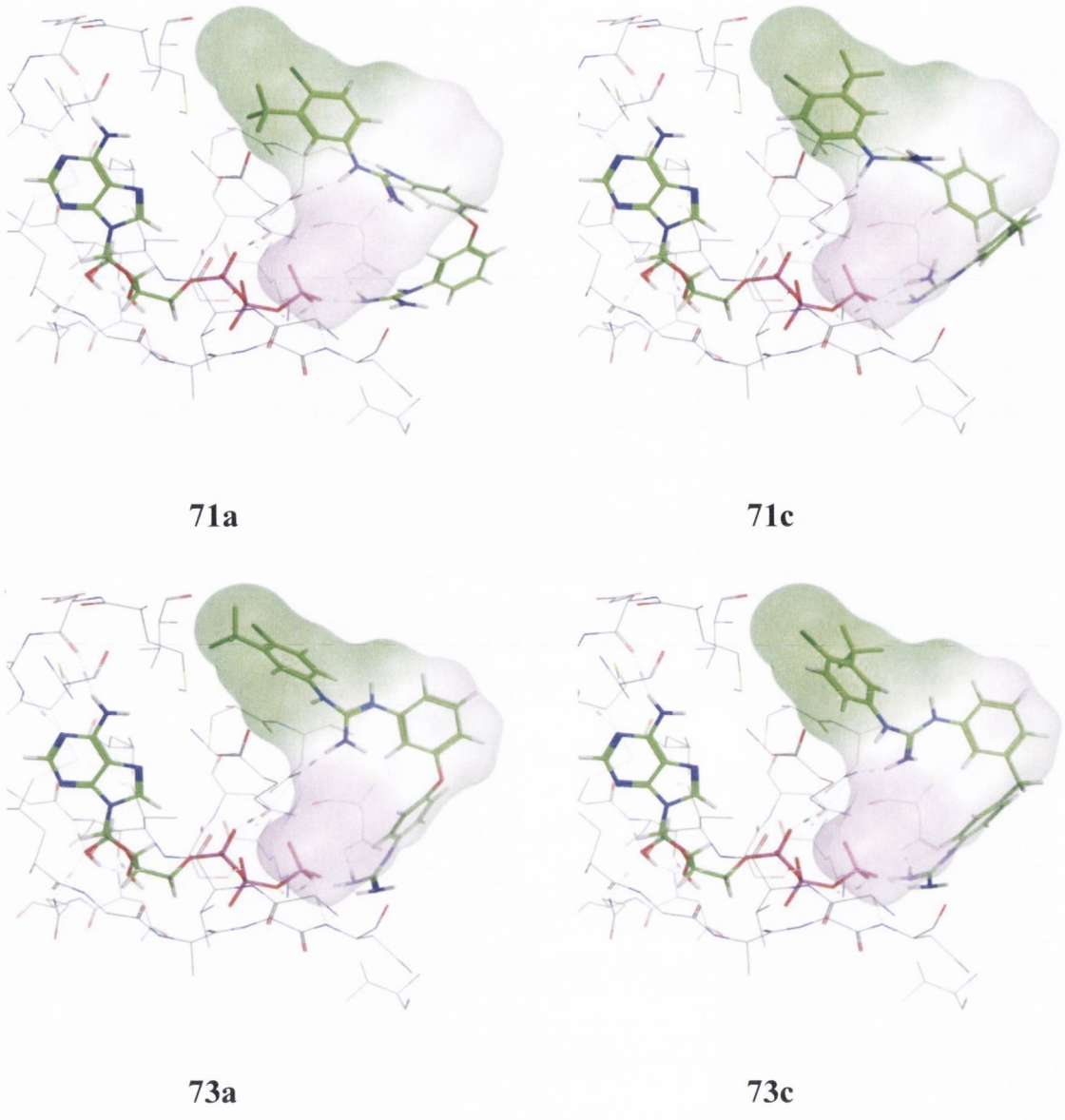


Figure 5.9. Principal pose of ligands **71a**, **71c**, **73a** and **73c** after minimisation in MEK-1. The surface is that generated by IZG; pink represents areas of ligand polarity and green represents areas of ligand hydrophobicity. The ligand and ATP are shown in green.

It is worth mentioning that the same analysis was carried out for these ligands in the absence of ATP and Mg^{2+}. The same patterns were reproduced for this system but with lower binding free energies observed for most of the ligands, as expected from the loss of the electrostatic interaction between the phosphate group of ATP and either the guanidinium of our compounds or the hydroxyl group of IZG. Considering that the most

successful MEK inhibitors are generally allosteric (do not displace ATP completely), the study with ATP present is more relevant.

5.5. Conclusions

Several conclusions can be drawn from the results obtained in this chapter. It is worth mentioning the similarities between B-RAF and C-RAF not only in structure, but also in the amino acid sequence involved in the active site region. For this reason, and considering our experimental data which showed that our compounds do not inhibit B-RAF, we propose that they will not inhibit C-RAF either, as ATP-competitive inhibitors. It is possible that they could bind to other domains of the protein (allosteric sites), which could explain the slight C-RAF expression decrease obtained in the western blot studies with compound **73a**. However, this fact needs to be studied in more detail. Experimental data confirmed that our compounds inhibit the C-RAF/MEK-1 pathway, meaning that the likely principal target of our compounds is MEK-1. This fact was backed up by the favoured electrostatic interaction between the guanidinium group in our compounds and a phosphate group of ATP found during the computational study (which is only available in the MEK-1 system), and also the significant decrease in the expression levels of MEK-1 after treatment with compound **73a**.

Taking the above into consideration, an exhaustive study of the interactions of MEK-1 and ATP with the crystal ligand IZG and compounds **49a**, **51a**, **51d**, **70a**, **71a-d**, **73a**, **73c**, **79a**, and **79d** was carried out, involving docking, minimisation techniques and calculation of solvated binding free energies. The crystal ligand IZG had the highest binding free energy, as expected. The orientation of the crystal ligand was also faithfully reproduced after both docking and minimisiation. This serves as an indication that the docking and minimisation procedures are leading to reliable poses.

Two main interactions can be highlighted as being the responsible for the binding to MEK-1: (i) an electrostatic interaction between the guanidinium group of the ligand and one phosphate group of ATP and (ii) a lipophilic interaction between a lipophilic moiety of the ligand and the lipophilic pocket of the protein near the ATP site. Considering that these interactions are crucial for the strong binding of the ligands to the enzyme, and in

particular that the lipophilic interaction directs the binding, several differences can be observed between the different compounds studied.

Compounds from Family 5 (**79a** and **79d**) are situated at the bottom-ranked compounds. They maintain the lipophilic interaction between the (4-Cl-3-CF_3)-Ph group and the lipophilic pocket from MEK-1, but not an interaction with ATP. Given their partial inhibition of the C-RAF/MEK-1 pathway, and their high cytotoxic effect in a wide range of cancer cells, it is possible that their main target is a different protein. This will require further investigation.

Higher binding free energies were found for compounds **49a**, **51a**, **51d** and **70a**. Due to the chemical structure of compounds **49a**, **51a** and **51d**, they maintain the electrostatic interaction with ATP; however, they cannot form interact with the lipophilic pocket because they only have an unsubtituted guanidinium group to place in this part of the site. This justifies the low to moderate inhibition of the C-RAF/MEK-1 pathway observed experimentally. Similar energy was encountered for compound **70a**, even though this compound has a lipophilic phenyl group; this group, which is less lipophilic than (4-Cl-3-CF_3)-Ph, was not located in the lipophilic pocket.

The top-ranked compounds were those from Families 4A and 4B (**71a-d**, **73a** and **73c**) due to the fact that they maintain the two crucial interactions: electrostatic and lipophilic. This justifies the high inhibition experimentally observed for these compounds in the C-RAF/MEK-1 pathway (up to 99%). These derivatives also provided some useful suggestions as to how to obtain higher affinity ligands for MEK-1 in the future. For example, it was evident that these molecules were under conformational strain to incorporate both the guanidinium-phosphate and lipophilic interactions. Thus, slightly shorter molecules would relieve this strain and fit better into the binding site of MEK-1. Additionally, the (4-Cl-3-CF_3)-Ph guanidinium group should preferably be on the 1,3-substituted ring (like **73a** and **73c**) instead of the 1,4-substituted ring (like **71a** and **71c**), as this small change seems to allow the molecule to interact more strongly with the protein. This justifies that compounds **73a** and **73c** have a higher cytotoxic effect than **71a** and **71c** as experimentally observed in a wide range of cancer cells.

5.6. References

(1) Molecular Operating Environment (MOE) 2011.10. Chemical Computing Group. Montreal, Canada.

(2) Case, D. A.; Cheatham Iii, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr, K. M.; Onufriev, A.; Simmerling, C.; Bing, W.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668-1688.

(3) Wan, P. T.; Garnett, M. J.; Roe, S. M.; Lee, S.; Niculescu-Duvaz, D.; Good, V. M.; Jones, C. M.; Marshall, C. J.; Springer, C. J.; Barford, D.; Marais, R. *Cell* **2004**, *116*, 855-867.

(4) Hatzivassiliou G, S. K., Yen I, Brandhuber BJ, Anderson DJ, Alvarado R, Ludlam MJ, Stokoe D, Gloor SL, Vigers G, Morales T, Aliagas I, Liu B, Sideris S, Hoeflich KP, Jaiswal BS, Seshagiri S, Koeppen H, Belvin M, Friedman LS, Malek S. *Nature* **2010**, *464*, 431-435.

(5) Dong, Q.; Dougan, D. R.; Gong, X.; Halkowycz, P.; Jin, B.; Kanouni, T.; O'Connell, S. M.; Scorah, N.; Shi, L.; Wallace, M. B.; Zhou, F. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 1315-1319.

(6) http://www.ebi.ac.uk/Tools/msa/clustalw2/.

(7) Wilhelm, S. M.; Carter, C.; Tang, L.; Wilkie, D.; McNabola, A.; Rong, H.; Chen, C.; Zhang, X.; Vincent, P.; McHugh, M.; Cao, Y.; Shujath, J.; Gawlak, S.; Eveleigh, D.; Rowley, B.; Liu, L.; Adnane, L.; Lynch, M.; Auclair, D.; Taylor, I.; Gedrich, R.; Voznesensky, A.; Riedl, B.; Post, L. E.; Bollag, G.; Trail, P. A. *Cancer Res.* **2004**, *64*, 7099.

(8) Hatzivassiliou, G.; Song, K.; Yen, I.; Brandhuber, B. J.; Anderson, D. J.; Alvarado, R.; Ludlam, M. J. C.; Stokoe, D.; Gloor, S. L.; Vigers, G.; Morales, T.; Aliagas, I.; Liu, B.; Sideris, S.; Hoeflich, K. P.; Jaiswal, B. S.; Seshagiri, S.; Koeppen, H.; Belvin, M.; Friedman, L. S.; Malek, S. *Nature* **2010**, *464*, 431-435.

(9) Šali, A.; Potterton, L.; Yuan, F.; van Vlijmen, H.; Karplus, M. *Proteins* **1995**, *23*, 318-326.

(10) http://www.uniprot.org/uniprot/Q02750.

(11) Eramian, D.; Shen, M.; Devos, D.; Melo, F.; Sali, A.; Marti-Renom, M. A. *Protein Sci.* **2006**, *15*, 1653-1666.

(12) Fogolari, F.; Brigo, A.; Molinari, H. *Biophys. J.* **2003**, *85*, 159-166.

Chapter 6 – Conclusions and Future Work

6.1. Conclusions

During this work the preparation of five novel families of protein kinase inhibitors was successfully carried out. Several different biochemical assays were performed to investigate their effects on biological systems, as potential anticancer drugs. Thus, viability assays in a wide range of cancer cell lines, apoptosis evaluation and different specific kinase inhibition tests were performed to assess their cytotoxicity and to evaluate their potential as protein kinase inhibitors. These biochemical outcomes allowed for a molecular modelling study to better understand the more likely mechanism of action of these kinase inhibitors. Based on the results obtained, some conclusions can be made.

6.1.1. Synthesis

Based on methodology previously described by Kim and Qian,[1] and posteriorly explored by Rozas and co-workers,[2, 3, 4] the synthesis of five families of novel 3,4-di-guanidininium and -2-aminoimidazolinium derivatives was successfully achieved. Additionally, some of the previously synthesised 4,4'-*bis*-guanidinium derivatives were prepared again (**49a-d** and **50a-d**) to be evaluated as kinase inhibitors, given that they had previously displayed cell cytotoxicity through an unknown mechanism. For the preparation of all five families of derivatives, the starting 3,4'-diaryl diamines had first to be synthesised (**9a-d**), as they are the starting material for each of the respective families. To achieve this, several synthetic routes were followed depending of the nature of the group linking the two aniline rings. During this synthesis a new simultaneous reduction of nitro and carbonyl groups was discovered. Since this double reduction had not been observed previously, an exhaustive study of several types of nitro and carbonyl groups with varying reactivity was performed, to establish a general new route of reduction.[5] Furthermore, a mechanism for this unreported nitro reduction was proposed, suggesting that it proceeds *via* the formation of a nitroso intermediate.

Once the starting diaryl diamines were synthesised, they were guanidylated/2-aminoimidazolydated using mercury(II) chloride promoted coupling with a thiourea or an imidazolidine-2-thione derivative to afford the Boc-protected guanidine-like derivatives.

In the first two families, a 3,4'-substitution was considered producing 3,4'-*bis*-guanidinium (Family 1) and 3,4'-*bis*-(2-aminoimidazolinium) (Family 2) derivatives. Different linkers between the aromatic rings (O, NH, CH_2, CO) were explored to study their influence on the biochemical activity, given that they can affect geometry, hydrogen bond profile and electronics of the phenyl rings. Next, a phenyl ring was introduced to the guanidinium group in the 4'-position, as a mimic of the (4-Cl-3-CF_3)-Ph group characteristic of sorafenib (Family 3A) – this substituted phenyl ring was also incorporated to give the compounds of Family 4A. During the preparation of Families 3 and 4, those derivatives with the disubstituted guanidinium placed on the 1,3-di-substituted aryl ring were also synthesised for those systems with O and CH_2 linkers. These derivatives were named as Families 3B and 4B. Finally, in Family 5, an acetamide group was introduced in the 3-position instead of the guanidinium group from Families 1-4 or the *N*-methylcarbamoyl found in sorafenib.

It was hoped that the modifications inserted in each family of compounds (Family 1-5) would generate SAR information which could be used to understand in more detail the structural requirements for a more potent binding of the compounds to the biological target. It is noteworthy that the structural changes introduced in each family of compounds improved their biological activity relative to the previous one, Families 4 and 5 being the most active sets of compounds, demonstrating that the design plan of this project was successful.

The biologically suitable hydrochloride salts from Families 1 and 2 (**51a-d** and **52a-c**) were generated from the Boc-protected intermediates (**41a-d** and **41a-c**) and using TFA/DCM followed by treatment with Amberlite resin in its Cl^- form. While, for the members of Families 3A, 3B, 4A, 4B and 5 (**70-71a-d**, **72-73a**, **72-73c**, **78-79a** and **78-79d**), these salts were generated from the Boc-protected intermediates (**53-54a-d**, **55-56a**, **55-56c**, **76-77a** and **76-77d**) using 4M HCl in dioxane in excellent yields.

The signals in the ^1H and ^{13}C NMR of the Boc-protected intermediates from Families 3, 4 and 5 were highly broad and amorphous due to the presence of isomerism of these tri-substituted guanidine derivatives. This phenomenon was studied through low temperature

NMR and DFT computational studies. Significantly, when these Boc groups were removed, this isomerism disappeared and the ^{1}H and ^{13}C NMR spectra of the hydrochloride salts were normal and could be resolved.

The purity of all hydrochloride salts from the five families was determined by HPLC analysis, the cut-off for purity being set at ≥95%. Thus, the synthesis of 24 final salts with a high degree of purity was successfully accomplished.

6.1.2. Biochemistry

Several different biochemical assays were performed on the five families of compounds synthesised. Viability assays, evaluation of apoptosis and different kinases inhibition tests were performed to assess their cytotoxicity and to evaluate their role as potential protein kinase inhibitors. A screening of the cytotoxicity of all compounds was performed in the promyelocytic leukaemia HL-60 cell line and the compounds with more potent activity in these tests were selected for further biological evaluation in several solid tumour cell lines. The most active families were Families 1, 4A, 4B and 5; therefore, a representative molecule from each family (compounds **51a**, **71c**, **73a** and **79d**) was selected to evaluate compound effects on cell proliferation in liver carcinoma HUH-7, wild type and mutated colorectal cancer RKO, breast cancer MCF-7 and tamoxifen-resistant MCF-7 cell lines. The IC_{50} values obtained in all cancer cells studied were in the low μM range – similar and sometimes better than that of sorafenib. The fact that the IC_{50} values of our compounds in MCF-7 and Tam-MCF-7 were nearly identical suggesting that our compounds may overcome the resistance to tamoxifen in this cell line; hence, they could be a potential therapeutic alternative for patients with tamoxifen-resistant breast cancer. However, high selectivity for colorectal cancer cell line with ^{V600E}BRAF mutated over the wild type was not observed, as the IC_{50} values obtained for the former were higher.

Several conclusions can be made regarding SAR information from the evaluation of the cytotoxicity of all families of compounds in HL-60 cell line: (i) Families 2, 3A and 3B showed poor cytotoxic effects with IC_{50} values >100 μM; hence, the 2-aminoimidazolinium moiety and the presence of a phenyl group attached to the guanidinium in either the 4'- or 3- position were not favourable; (ii) the (4-Cl-3-CF_3)-Ph substituent seems to be crucial for the cytotoxic activity of the compounds (Families 4A

and 4B); (iii) the un-substituted *bis*-guanidinium derivatives (Family 1) showed better cytotoxicity results than the phenyl-substituted guanidinium derivatives (Families 3A and B), but lower cytotoxicity than the (4-Cl-3-CF_3)-Ph guanidinium derivatives (Families 4A and 4B); (iv) improvements in the cytotoxic activity of the compounds were obtained when the (4-Cl-3-CF_3)-Ph group was inserted in the guanidinium moiety in the 3-position (Family 4B) instead of the 4'- position (Family 4A); (v) the acetamide group in Family 5 enhanced the cytotoxicity over the guanidinium moiety of Families 4A and 4B in all the cancer cells studied; (vi) the linker between the two aryl groups was not of significant importance for the cytotoxic activity of the derivatives of any family.

Microscopy studies of the effect of **73a** and **79d** in several cancer cells were carried out where marked changes in cellular morphology such as chromatin condensation, membrane blebbing, nuclear breakdown, the appearance of membrane-associated apoptotic bodies and DNA fragmentation were observed, suggesting that our compounds can induce apoptosis.

Flow cytometry analysis in HL-60 cell line and PARP cleavage induction in the RKO cell line were also performed to quantify and confirm the pro-apoptotic ability of our compounds, respectively. Flow cytometry studies demonstrated that our compounds induce G_0/G_1 phase arrest of the cell cycle prior to apoptosis, in a similar manner to sorafenib. Also, PARP cleavage confirmed that apoptosis is the type of cell death caused by **73a**.

Kinase testing suggested that compounds from Family 4A displayed 74-99% inhibition of the RAF-1/MEK-1 kinase pathway, as well as 20% anti-angiogenesis activity through inhibition of VEGFR, and a moderate 20% inhibition of ERK-1/2 and p-38 MAPK. Additionally, compound **51d** – a member of Family 1 – showed 86% inhibition of the RAF-1/MEK-1 kinase pathway. However, other members of Family 1 did not show as strong inhibition of this pathway and, this result is not consistent with the computational results obtained. Compounds from Families 3 and 5 did not show inhibition of the RAF-1/MEK-1, ERK-1/2, p38 MAPK or VEGFR pathway. Members of Family 5 displayed high cytotoxicity in all studied cancer cell types but their mode of action seems to be different from the compounds in Family 4 as they do not target the RAF-1/MEK-1 kinase pathway.

The effect of compound **73a** was studied in the expression levels of various signalling proteins in RKO cells. Western blot experiments displayed a reduction of p-ERK

expression levels when treated with **73a**, confirming the down-regulation of the RAF/MEK/ERK pathway by compounds from Family 4. The observed decrease in the expression of ERK is likely to be related to an inhibition of the upstream kinases of the pathway (*e.g.* RAF or MEK) because the general kinase screening carried out showed that our compounds do not target directly ERK 1/2.

A slight reduction of the levels of C-RAF was observed. This interaction needs to be analysed in more detail as the ATP-binding site of C-RAF and B-RAF is very similar as shown in the literature;[6] hence, these compounds could be allosteric inhibitors of C-RAF. However, a significant reduction of the expression levels of p-MEK was obtained after the treatment with the compound 3AA, suggesting that the likely target of our compounds is MEK.

LanthaScreen® Eu Kinase Binding Assay confirmed that our compounds do not competitively bind to B-RAF, in agreement with the results obtained in the western blot.

Other molecular targets were investigated such as CK-1δ, GSK-3, STAT-3 and AKT, which play a central role in numerous cellular events including differentiation, proliferation and apoptosis. The results obtained showed that our compounds do not modify the activity of these kinases.

The biological results obtained for these families of potential protein kinase inhibitors are very encouraging, considering that they are at the early stages of development. Consequently, these compounds have generated new opportunities for the synthesis of modified and improve anticancer compounds, specifically protein kinase inhibitors in the future which, it is hoped, will have a contribution in the "war of cancer".

6.1.3. Computational Chemistry

Computational modelling techniques such as docking and molecular dynamics have been utilised to understand the interactions between our novel compounds and their targets (B-RAF, C-RAF and MEK-1) as well as to analyse their binding.

First, based on the data obtained experimentally and the results of this computational study we confirmed that our compounds do not target B-RAF or C-RAF and their likely target of our compounds was MEK-1 (see section 5.2).

A docking study has been carried out to identify the interactions between a selection of our molecules (**49a**, **51a**, **51d**, **70a**, **71a-d**, **73a**, **73c**, **79a** and **79d**) and MEK-1, which crystal structure is known. Two main interactions could be highlighted as responsible for this binding: (i) an electrostatic interaction between the guanidinium cation and one phosphate group of ATP and (ii) a hydrophobic interaction between a lipophilic moiety of the compounds and the enzyme's lipophilic pocket near the ATP site. The later was found to be the driving force which directs the molecules inside the MEK-1 active site pocket. The presence of one or both of these interactions influences the strength of the complex ligand-enzyme. Compounds from Families 1, 3 and 5 only established one of these interactions, for this reason the calculated binding energy was less favourable than that of compounds form Families 4A and 4B, in agreement with the experimental results obtained for the inhibition of the C-RAF/MEK-1 pathway (Families 1, 3 and 5 showed poor to moderate inhibition and Family 4A and 4B showed very significant selectivity for this pathway).

6.2. Future Work

Throughout this project some interesting and very promising results were obtained and therefore several areas will require further investigation. One of them is the biological application of our compounds in breast and colorectal cancer, which will possibly open new research projects in the future. Another area to consider is the synthesis of future compounds, which computational studies suggested as high affinity inhibitors of MEK or other kinases. This could result in improved cytotoxicity and specificity for a particular target. Additionally, a more exhaustive analysis of the binding between the compounds synthesised in this work and C-RAF should be carried out to investigate the possible allosteric inhibition of this protein.

6.2.1. Future Biological Application of our Compounds

Some of our compounds, mainly those from Families 4 and 5 showed very promising results against breast cancer cells; therefore, in collaboration with Dr. J. Gee and her group (Cardiff School of Pharmacy and Pharmaceutical Sciences, UK) further investigation of these derivatives will be carried out. A broader screening of the compounds across a panel

of ER^+ and ER^- breast cancer cell lines will be carried out to gauge their applicability and selectivity. Additionally, further compounds will be prepared and investigated to improve and maximise their effects on breast cancer cells. To identify the best structural changes to be performed, computational studies of the binding of our compounds with ER will have to be first carried out. This study will open a new project in our group for which funding is being sought at the moment, in association with our collaborators at Cardiff.

Additionally, very satisfactory results were found for the treatment of colorectal cancer cells with our compounds, mainly with compounds from Families 4 and 5. Hence, further investigation of their applicability will be carried out in collaboration with Dr. S. Van Schaeybroeck in the Centre for Cancer Research and Cell Biology (CCRCB) at Queen's University Belfast (UK). In this project, evaluation of other possible molecular targets of our compounds, such as EGFR or MCL-1, will be performed using western blotting among others techniques. Also, possible combination therapies of our compounds with other anticancer drugs used in colorectal cancer such as vemurafenib (^{V600E}B-RAF inhibitor) will be investigated. Hence, targeting two different pathways ^{V600E}B-RAF and MEK with vemurafenib and our compounds respectively, more potent cancer treatments will be hopefully achieved.

Finally, it would be interesting to test the *in vivo* effects of some of the most active compounds from Families 4 and 5 on tumour models. These experiments represent a step further towards the possibility of these compounds ever reaching the clinic.

6.2.2. Future Protein Kinase Inhibitors

Considering the SAR conclusions found after the synthesis, biological evaluation and modelling of the families of compounds performed in this work, several ideas can be proposed to try to improve the potency of future protein kinase inhibitors.

During this work we identified that the presence of the (4-Cl-3-CF_3)-Ph substituent is crucial to increase the binding to the targeted protein kinases. Additionally, we saw that placing the (4-Cl-3-CF_3)-Ph substituent on the guanidinium in the 3-position (Family 4B) increases the cytotoxicity of these compounds in several cancer cell lines over the ones who have the (4-Cl-3-CF_3)-Ph substituent on the guanidinium in the 4'-position (Family 4A). Therefore, if this moiety is introduced in the guanidinium groups in both the 3- and

the 4'- position of the 3,4'-diphenyl diamine scaffold, it may increase the cytotoxic effects of these compounds in the cancer cells previously studied (Figure 6.1).

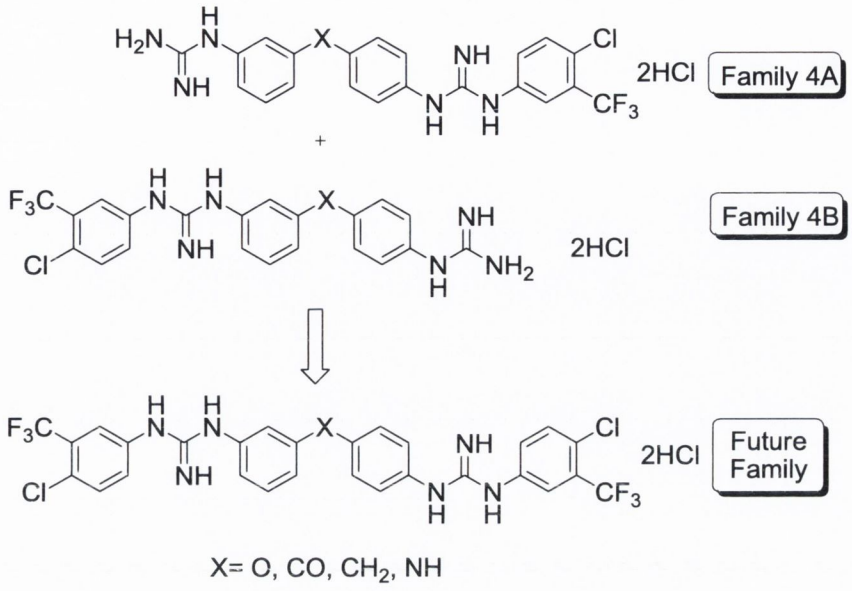


Figure 6.1. Suggested future family of compounds.

The synthesis of this 3,4'-*bis*-(4-Cl-3-CF_3)-Ph guanidinium-like family is currently being investigated by a final year student's project in our group.

Based on the results obtained in the computational study of the binding of our compounds to MEK, two families of compounds can be proposed (Figure 6.2). The most active family of compounds synthesised in this work was Family 4B, which contain (4-Cl-3-CF_3)-Ph moiety in *meta* position (see structure in Figure 6.1); hence, these two proposed families are based on them.

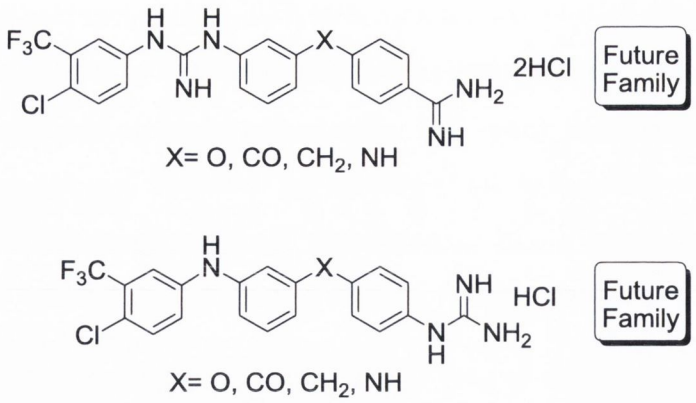


Figure 6.2. Two future families of compounds, potential MEK inhibitors.

According to our computational modelling, the main two interactions required for the binding to MEK are (i) the electrostatic interaction, between the unsubstituted guanidinium moiety at one end of the molecule and the phosphate of the ATP and (ii) the lipophilic interaction, between the (4-Cl-3-CF_3)-Ph moiety and the liphophilic pocket in the active site. These two crucial interactions are maintained in the two families proposed. The justification for the structural changes introduced is that compounds from Family 4B were slightly big for MEK active site and the molecules were under strain. Therefore, these new molecules are one and two bonds, respectively, shorter than molecules from Family 4B. These changes will allow the proposed molecules to fit better in the MEK active site and still keep the electrostatic and lipophilic interaction vital for the binding.

Another family of compounds can be proposed based on the biological results obtained for the compounds in this project. Considering that lipophilic interactions between drug and protein have been shown to be one of the main driving forces to gain selectivity in the binding to the targeted protein, compounds from Family 4B can be modified by substituting the CF_3 group with a pentafluorosulfanyl (SF_5) one. The SF_5 group has greater lipophilic, electron-withdrawing and steric properties than CF_3. Also, as CF_3, it can improve the pharmacokinetic properties of drugs. These properties make the SF_5 group an increasingly interesting structural motif for the design of bioactive compounds, to the point that it is being classified as the "substituent of the future", as is evidenced by its increasing presence in recent pharmaceutical drugs, energetic materials, polymers and pesticides.[7] However, access to SF_5-containing compounds is very limited and their chemistry remains

largely unexplored; hence, the introduction of this group in our molecules will be a challenge (Figure 6.3).

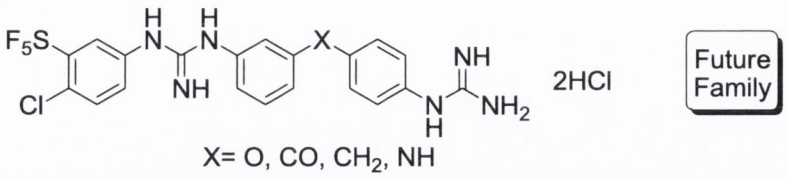


Figure 6.3. Structure of a future family of compounds containing the SF_5 group.

In conclusion, with the future biological exploration of the derivatives prepared in this work and also with the proposed families of compounds, we intend to obtain more information about the biological applicability of these derivatives and also to design further improved compounds with potential as protein kinase inhibitors in the future.

6.3. References

(1) Kim, K. S.; Qian, L. *Tetrahedron Lett.* **1993**, *34*, 7677-7680.

(2) Dardonville, C.; Goya, P.; Rozas, I.; Alsasua, A.; Martin, M. I.; Borrego, M. J. *Bioorg. Med. Chem.* **2000**, *8*, 1567-1577.

(3) Dardonville, C.; Brun, R. *J. Med. Chem.* **2004**, *47*, 2296-2307.

(4) F. Rodriguez; I.Rozas; Ortega, J. E.; Meana, J. J.; Callado, L. F. *J. Med. Chem.* **2007**, *50*, 4516-4527.

(5) Diez-Cecilia, E.; Kelly, B.; Rozas, I. *Tetrahedron Lett.* **2011**, *52*, 6702-6704.

(6) Chong, H.; Vikis, H. G.; Guan, K.-L. *Cell. Signal.* **2003**, *15*, 463-469.

(7) Jackson, D. A. *"Breaking Down the Substituent of the Future": Environmental Properties of Pentafluorosulfanyl Compounds*; University of Toronto (Canada), 2008.

Chapter 7 – Experimental

7.1. Synthetic Chemistry

7.1.1. Material and Methods

All commercial chemicals were obtained from Sigma-Aldrich or Fluka and used without further purification. Deuterated solvents for NMR use were purchased from Apollo. Dry solvents were prepared using standard procedures, according to Vogel,[1] with distillation prior to use. Chromatographic columns were run using a Biotage SP4 flash purification system with Biotage SNAP silica cartridges. Solvents for synthesis purposes were used at GPR grade. Analytical TLC was performed using Merck Kieselgel 60 F254 silica gel plates or Polygram Alox N/UV254 aluminium oxide plates. Visualisation was by UV light (254 nm). NMR spectra were recorded on Bruker DPX–400 Avance spectrometers, operating at 400.13 MHz and 600.1 MHz for 1H NMR; 100.6 MHz and 150.9 MHz for 13C-NMR. Shifts are referenced to the internal solvent signals.[2] NMR data were processed using Bruker TOPSPIN software. HRMS spectra were measured on a Micromass LCT electrospray TOF instrument with a WATERS 2690 autosampler and methanol/acetonitrile as carrier solvent. Melting points were determined using a Stuart Scientific Melting Point SMP1 apparatus and are uncorrected. Infrared spectra were recorded on a Perkin Elmer Spectrum One FT-IR Spectrometer equipped with a Universal ATR sampling accessory.

7.1.2. Purity Assessment of Hydrochloride Salts

HPLC purity analysis was carried out using a Varian ProStar system equipped with a Varian Prostar 335 diode array detector and a manual injector (20 μL). For purity assessment, UV detection was performed at 245 nm and peak purity was confirmed using a purity channel. The stationary phase consisted of an ACE 5 C18-AR column (150 × 4.6 mm), and the mobile phase used the following gradient system, eluting at 1 mL/min: aqueous formate buffer (30 mM, pH 3.0) for 10 min, linear ramp to 85% methanol buffered with the same system over 25 min, hold at 85% buffered methanol for 10 min. Minimum requirement for purity was set at 95.0%.

7.1.3. General Procedures

Method A: *General procedure for the synthesis of the Bis-Boc-protected guanidine and bis-Boc-protected 2-aminoimidazoline derivatives.*

Mercury (II) chloride (2.2 eq.) was added over a 0.25 M solution of the corresponding diamine (1 eq.), triethylamine (9 eq.), *N,N'-bis*-(*tert*-butoxycarbonyl)-*S*-methylisothiourea (**44**) (2.2 eq.) in DCM in case of the guanidinium derivatives, or *N,N'*-di(*tert*-butoxycarbonyl)imidazolidine-2-thione (**46**) (2.2 eq.) in DMF case of the 2-aminoimidazolinium derivatives, at 0 °C. The resulting mixture was stirred at 0 °C for 1 h and then overnight at room temperature. Then, the reaction mixture was diluted with EtOAc (100 mL) and filtered through a pad of Celite® in order to remove the mercury sulfide precipitate formed. The filter cake was rinsed with EtOAc (2 × 25 mL). The organic phase was extracted with water (2 × 30 mL), washed with brine (30 mL), dried over anhydrous $MgSO_4$, and concentrated under vacuum to give a residue that was purified by flash chromatography on silica gel for the guanidines and on alumina for the aminoimidazolidine, as specified.

Method B: *General procedure for the synthesis of the mono-Boc-protected guanidine derivatives.*

Mercury (II) chloride (1.2 eq.) was added over a 0.193 M solution of the corresponding diamine (3 eq.), the corresponding thiourea (1 eq.) and triethylamine (3.1 eq.) in DCM at 0 °C. The resulting mixture was stirred at 0 °C for 1 h and then at room temperature for further 2-4 h or overnight depends on the reactivity of the substrates and possible generation of side products (reaction progress adjudged by TLC). Then, it was diluted with EtOAc (100 mL) and filtered through a pad of Celite® in order to remove the mercury sulfide precipitate formed. The filter cake was rinsed with EtOAc (2 × 25 mL). The organic phase was extracted with water (2 × 30 mL), washed with brine (30 mL), dried over anhydrous $MgSO_4$, and concentrated under vacuum to give a residue. It was purified by flash chromatography on silica gel (eluting with a gradient of hexane:EtOAc) as specified.

Method C: *General procedure for the synthesis of the asymmetric 3,4-bis-Boc-protected guanidine derivatives.*

To a 0.193 M solution of the corresponding mono Boc-protected guanidine (1 eq.), the corresponding thiourea (1.2 eq.) and triethylamine (3.1 eq.) in DCM and additional isopropanol in some cases to increase solubility, mercury (II) chloride (1.2 eq.) was added. The resulting mixture was stirred at room temperature for 3-6 h depends on the reactivity of the substrates (reaction progress adjudged by TLC). Then, it was diluted with EtOAc (100 mL) and filtered through a pad of Celite® in order to remove the mercury sulfide precipitate formed. The filter cake was rinsed with EtOAc (2 × 25 mL). The organic phase was extracted with water (2 × 30 mL), washed with brine (30 mL), dried over anhydrous $MgSO_4$, and concentrated under vacuum to give a residue. It was purified by flash chromatography on silica gel (eluting with a gradient of hexane:EtOAc) as specified.

Method D: *Generic Base Mediated Acetylation of the mono-Boc-protected guanidine derivatives.*

To a 0.2 M solution of the corresponding mono-Boc-protected guanidine derivative (1 eq.) and acetic anhydride (2 eq.) in DCM, triethylamine (3 eq.) was added. The reaction was stirred at room temperature for 6 h. Afterwards, it was diluted with EtOAc, extracted with water (2 × 30 mL) and washed with brine (30 mL). The organic layer was dried over anhydrous $MgSO_4$ and concentrated under vacuum to give a residue that was purified by flash column chromatography on silica gel (eluting with a gradient of hexane:EtOAc) as specified.

Method E: *General method for the deprotection of N-Boc protected guanidines using trifluoroacetic acid and Amberlite resin.*

A solution of the corresponding Boc-protected precursors (0.5 mmol) in 20 mL of a 50% *v/v* solution of TFA and DCM was stirred at room temperature. Once disappearance of starting material was observed by TLC (3 h), the excess of TFA and DCM were removed under vacuum to generate the trifluoroacetate salt. The residue was dissolved in water (20 mL) and 1000 mg of IRA400 Amberlite resin in its chloride-activated form was added,

allowing the reaction to proceed overnight. The resin was then removed by filtration and the aqueous solution washed with DCM (2 × 10 mL). Removal of water under vacuum yielded the crude guanidine hydrochloride. Absence of the trifluoroacetate salt was confirmed by ^{19}F NMR, if necessary, stirring in Amberlite was repeated. Purification was carried out by reverse phase chromatography using C-8 silica, typical elution gradient: 100% water to 85:15 water:acetonitrile, respectively. The purified fractions were evaporated until dry to afford the pure hydrochloride salt.

Method F: *General method for the deprotection of N-Boc protected guanidines using hydrochloric acid.*

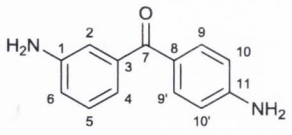
To a solution of the corresponding Boc protected guanidine (1 eq.) in 1,4-dioxane, 4 M HCl in 1,4-dioxane (6 eq. per Boc group) was added to reach a final concentration of 0.2M. The mixture was stirred at 55 °C until the reaction was adjudged complete (typically 8 h, adjudged by TLC). At the reaction endpoint, solvent and HCl were evaporated under vacuum and the crude salt was dissolved in the minimum volume of water. It was washed with DCM (2 × 10 mL) and then purified by using reverse phase chromatography using C-8 silica, typical elution gradient: 100% water to 85:15 water:acetonitrile, respectively. The purified fractions were evaporated until dry to afford the pure hydrochloride salt.

Method G: *General method for the deprotection of N-Boc protected acetylated guanidines using hydrochloric acid.*

A solution of the corresponding Boc-protected acetylated guanidine precursors (1 eq.) in 1:1 DCM:IPA was treated with 4 M HCl in dioxane (6 eq. per Boc group) to reach a final concentration of 0.2M. The mixture was stirred at 30 °C until the reaction was complete (typically 6 h, determined by TLC). At the reaction endpoint, solvent and HCl were evaporated under vacuum and the crude salt was dissolved in the minimum volume of water. It was washed with DCM (2 × 10 mL) and then purified by using reverse phase chromatography using C-8 silica, typical elution gradient: 100% water to 85:15 water:acetonitrile, respectively. The purified fractions were evaporated until dry to afford the pure hydrochloride salt.

7.1.4. Synthesis and Characterisation

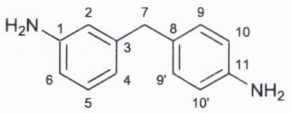
3,4'-Diaminobenzophenone (9b)



An aqueous solution (50 mL) of $Na_2S.9H_2O$ (2800 mg, 12 mmol) and NaOH (1100 mg, 28 mmol) was added to a stirred suspension of **11** (685 mg, 2.5 mmol) in ethanol (40 mL). The mixture was refluxed for 6 h and allowed to stand overnight. Ethanol was removed by rotary evaporation and the resulting precipitate was collected and washed with water and dried to yield **9b** as an orange-yellow solid (493 mg, 92%). **Mp**: 122-124 °C. (lit. 123-124 °C).[3]

δ_H (400 MHz, DMSO): 5.30 (bs, 2H, NH_2), 6.10 (bs, 2H, NH_2), 6.60 (d, 2H, J 8.5, H-10 and H-10'), 6.70 (d, 1H, J 8.0, H-4), 6.73 (d, 1H, J 8.0, H-6), 6.80 (s, 1H, H-2), 7.12 (t, 1H, J 8.0, H-5), 7.51 (d, 2H, J 8.5, H-9 and H-9').

3-(4-Aminobenzyl)-aniline (9c)

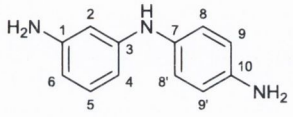


A solution of **11** (272 mg, 1.0 mmol) in 2 mL of distilled hydrazine hydrate was stirred for 3 h in a sealed pressure tube placed in an oil bath at 135 °C. The reaction was then allowed to reach room temperature and KOH (280 mg) was added. The mixture was allowed to react overnight at 135 °C. It was then extracted with a mixture 80:20 of DCM:IPA and washed with water (3 × 30 mL) and the organic solvent was evaporated to give a residue that was purified by silica gel column chromatography, eluting with hexane:EtOAc (1:1) to yield the desired product **9c** as a brown solid (239 mg, 88%). **Mp**: 87-89 °C. (lit. 89-90 °C).[4]

δ_H (600 MHz, DMSO): 3.58 (s, 2H, CH_2), 4.85 (bs, 2H, NH_2), 4.93 (bs, 2H, NH_2), 6.34 (1H, J 7.4, H-4), 6.36 (m, 2H, H-2 and H-6), 6.49 (d, 2H, J 8.0 H-10 and H-10'), 6.84 (d, 2H, J 8.0, H-9 and H-9'), 6.89 (t, 1H, J 7.4, H-5).

The same reduction conditions were followed for compounds **16 – 22.** Characterisation of these compounds is presented in the published article.[5]

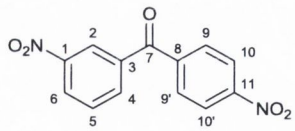
3-(4-Aminoanilino)aniline (9d)



To a hydrogenator flask, **38** (200 mg, 0.77 mmol), 10% Pd/C mixture (40 mg, 0.04 mmol, 5 mol %) and 2:1 EtOH:EtOAc (20:10 mL) were added. The mixture was vigorously stirred at room temperature under 3 atm of H_2 pressure overnight. The reaction mixture was filtered and the solvent was removed under vacuum to afford **9d** (135 mg, 88% yield) as a brown-purple solid. **Mp**: 99 °C. (lit. 97 °C).[6]

δ_H (600 MHz, DMSO): 4.71 (bs, 2H, NH_2), 4.76 (bs, 2H, NH_2), 5.87 (dd, 1H, J 7.9, 2.0, H-6), 5.98 (dd, 1H, J 7.9, 2.0, H-4), 6.05 (t, 1H, J 2.0, H-2), 6.50 (d, 2H, J 8.6, H-9 and H-9'), 6.74 (t, 1H, J 7.9, H-5), 6.79 (d, 2H, J 8.6, H-8 and H-8'), 7.13 (bs, 1H, NH).

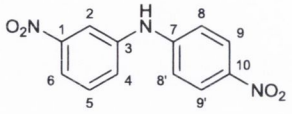
3,4'-Dinitrobenzophenone (11)



4-Nitrobenzophenone (2273 mg, 10.0 mmol) was added slowly over 15 mL of H_2SO_4 at 0 °C. KNO_3 (1010 mg, 10.0 mmol) was then added to the suspension over 10 min maintaining a constant temperature. The mixture was stirred at 0 °C for 30 min and a further 1.5 h at room temperature. Afterwards, the mixture was poured onto 50 g of ice and the precipitate that formed filtered, washed and dried until acid free and recrystallised from acetone to yield **11** as a white solid (1576 mg, 58%). **Mp**: 162-164 °C. (lit. 162-167 °C).[3]

δ_H (400 MHz,$CDCl_3$): 7.79 (t, 1H, J 8.0, H-5), 7.98 (d, 2H, J 8.8, H-9 and H-9'), 8.18 (dd, 1H, J 8.0, 2.0, H-4), 8.42 (d, 2H, J 8.8, H-10 and H-10'), 8.54 (dd, 1H, J 8.0, 2.0, H-6), 8.57 (t, 1H, J 2.0, H-2).

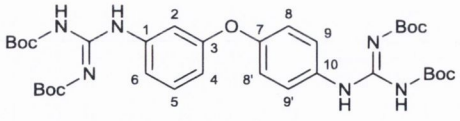
3-Nitro-*N*-(4-nitrophenyl)aniline (38)



An oven-dried Schlenk flask was evacuated and backfilled with argon. Afterwards, it was charged with $Pd_2(dba)_3$ (27.5 mg, 3.0 mol %), (2-diphenyl)dicyclohexylphosphine **31** (10.5 mg, 3.0 mol %) and 1,2-dimethoxyethane (DME) (2 mL/mmol of halide). The mixture was left to react for 5 min without heating. Then, K_3PO_4 (297.2 mg, 1.4 mmol), 1-chloro-3-nitrobenzene **37** (157.5 mg, 1.0 mmol) and 1,4-aminoaniline **35** (138.12 mg, 1.0 mmol) were added as solids maintaining the flow of argon. The mixture was stirred at 100 °C for 24 h. It was then cooled to room temperature, diluted with 1:1 ether:EtOAc (40 mL), filtered through celite, and concentrated under vacuum. The crude material was purified by chromatography on silica gel using 1:1 (hexane:EtOAc) solvent system to yield an orange solid **38** (207 mg, 80%). **Mp**: 217-218 °C. (lit. 214-218 °C).[7]

δ_H (600 MHz, DMSO): 7.23 (d, 2H, J 8.6, H-8 and H-8'), 7.65 (t, 1H, J 8.0, H-5), 7.7 (d, 1H, J 8.0, H-4), 7.85 (d, 1H, J 8.0, H-6), 8.00 (s, 1H, H-2), 8.17 (d, 2H, J 8.6, H-9 and H-9'), 9.67 (bs, 1H, NH).

3,4'-*Bis*-[2,3-di(*tert*-butoxycarbonyl)guanidino]diphenylether (41a)



Following method A, using **9a** (455 mg, 0.66 mmol) as starting material and after purification with silica gel chromatography, eluting with a gradient of hexane:EtOAc, **41a** was obtained as a white solid (397 mg, 87%). **Mp**: 120-121 °C.

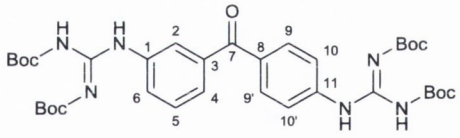
δ_H (400 MHz, $CDCl_3$): 1.51 (s, 9H, $(CH_3)_3$), 1.52 (s, 9H, $(CH_3)_3$), 1.55 (s, 9H, $(CH_3)_3$), 1.56 (s, 9H, $(CH_3)_3$), 6.75 (d, 1H, J 8.3, H-6), 7.03 (d, 2H, J 9.0, H-9 and H-9'), 7.27 (t, 1H, J 8.3, H-5), 7.30 (d, 1H, J 8.3, H-4), 7.32 (s, 1H, H-2), 7.60 (d, 2H, J 9.0, H-8 and H-8'), 10.3 (bs, 2H, NHBoc), 11.6 (bs, 1H, NH), 11.7 (bs, 1H, NH).

δ_C (100 MHz,$CDCl_3$): 26.9 ($2(CH_3)_3$), 28.1 ($2(CH_3)_3$), 79.5 (q, $\underline{C}(CH_3)_3$), 79.6 (q, $\underline{C}(CH_3)_3$), 83.7 (q, $\underline{C}(CH_3)_3$), 83.8 (q, $\underline{C}(CH_3)_3$), 112.4 (CH Ar, C-6), 114.5 (CH Ar, C-2), 116.8 (CH Ar, C-4), 119.8 (2 CH Ar, C-9 and C-9'), 123.6 (2 CH Ar, C-8 and C-8'), 129.7 (CH Ar, C-5), 132.4 (q, Ar, C-10), 138.1 (q, Ar, C-1), 153.2 (q, CO), 153.3 (q, CO), 153.4 (2q, CO), 153.5 (q, Ar, C-7), 153.6 (q, CN), 150.7 (q, CN), 158.0 (q, Ar, C-3), 163.6 (q, CO).

ν_{max} (ATR)/cm^{-1}: 3260 (NH), 3160 (NH), 2979, 2933, 1718 (CO), 1628 (CN), 1598, 1567, 1506, 1478, 1450, 1407, 1367, 1336, 1306, 1290, 1233, 1213, 1143, 1106, 1055, 1027, 879, 830, 804, 769, 686.

HRMS (*m/z* - ES): Found: 685.3560 (M^+ + H. $C_{34}H_{49}N_6O_9$ Requires: 685.3561).

3,4'-*Bis*-[2,3-di(*tert*-butoxycarbonyl)guanidino]benzophenone (41b)



Following method A, using **9b** (696 mg, 1.0 mmol) as starting material and after purification with silica gel chromatography, eluting with a gradient of hexane:EtOAc, **41b** was obtained as a pale orange solid (418 mg, 60%). **Mp**: 184-185 °C.

δ_H (600 MHz, $CDCl_3$): 1.50 (s, 9H, $(CH_3)_3$), 1.52 (s, 9H, $(CH_3)_3$), 1.54 (s, 9H, $(CH_3)_3$), 1.55 (s, 9H, $(CH_3)_3$), 7.45 (t, 1H, J 8.0, H-5), 7.53 (d, 1H, J 8.0, H-4), 7.79 (d, 2H, J 8.5, H-10 and H-10'), 7.85 (d, 2H, J 8.5, H-9 and H-9'), 7.87 (s, 1H, H-2), 7.96 (d, 1H, J 8.0, H-6), 10.45 (bs, 1H, NHBoc), 10.64 (bs, 1H, NHBoc), 11.66 (bs, 1H, NH), 11.68 (bs, 1H, NH).

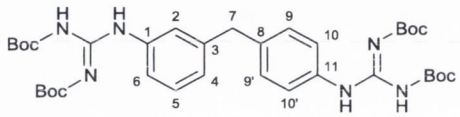
δ_C (150 MHz,$CDCl_3$): 28.0 ($(CH_3)_3$), 28.1 ($(CH_3)_3$), 28.1 ($(CH_3)_3$), 28.2 ($(CH_3)_3$), 79.5 (q, $\underline{C}(CH_3)_3$), 79.6 (q, $\underline{C}(CH_3)_3$), 83.6 (q, $\underline{C}(CH_3)_3$), 83.7 (q, $\underline{C}(CH_3)_3$), 121.1 (2 CH Ar, C-10

and C-10'), 123.6 (CH Ar, C-2), 128.2 (CH Ar, C-6), 128.3 (CH Ar, C-4), 128.7 (CH Ar, C-5), 131.4 (2 CH Ar, C-9 and C-9'), 130.8 (q, Ar, C-3), 133.0 (q, Ar, C-8), 138.5 (q, Ar, C-1), 140.9 (q, Ar, C-11), 153.2 (q, CO), 153.3 (q, CO), 153.7 (q, CN), 153.8 (q, CN), 163.2 (q, CO), 163.3 (q, CO), 194.9 (q, CO, C-7).

ν_{max} (ATR)/cm^{-1}: 3259 (NH), 3135 (NH), 2979, 2933, 1719 (CO), 1633 (CN), 1599, 1552, 1407, 1368, 1336, 1297, 1230, 1148, 1099, 1056, 1028, 941, 881, 845, 805, 757.

HRMS (*m/z* - ES): Found: 697.3564 (M^+ + H. $C_{35}H_{49}N_6O_9$ Requires: 697.3561).

3,4'-*Bis*-[2,3-di(*tert*-butoxycarbonyl)guanidino]benzylbenzene (41c)



Following method A, using **9c** (1.03 g, 0.0015 mol) as a starting material and after purification with silica gel chromatography, eluting with a gradient of hexane:EtOAc, **41c** was obtained as a white solid (885 mg, 86%). **Mp**: 160-162 °C.

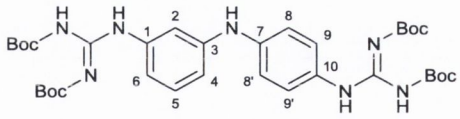
δ_H (600 MHz, $CDCl_3$): 1.52 (s, 9H, $(CH_3)_3$), 1.53 (s, 9H, $(CH_3)_3$), 1.55 (s, 9H, $(CH_3)_3$), 1.57 (s, 9H, $(CH_3)_3$), 3.94 (s, 2H, H-7), 6.93 (d, 1H, J 8.0, H-4), 7.16 (d, 2H, J 8.3, H-9 and H-9'), 7.26 (t, 1H, J 8.0, H-5), 7.32 (s, H, H-2), 7.53 (d, 2H, J 8.3, H-10 and H-10'), 7.58 (d, 1H, J 8.0, H-6), 10.3 (bs, 2H, NHBoc), 11.6 (bs, 2H, NH).

δ_C (150 MHz, $CDCl_3$): 27.9 ($2(CH_3)_3$), 28.1 ($2(CH_3)_3$), 41.1 (CH_2, C-7), 79.4 (q, 2 $\underline{C}(CH_3)_3$), 83.4 (q, 2 $\underline{C}(CH_3)_3$), 120.1 (CH Ar, C-6), 122.1 (2 CH Ar, C-10 and C-10'), 122.5 (CH Ar, C-2), 125.3 (CH Ar, C-4), 128.8 (CH Ar, C-5), 129.2 (2 CH Ar, C-9 and C-9'), 134.8 (q, Ar, C-11), 136.7 (q, Ar, C-8), 137.1 (q, Ar, C-1), 141.6 (q, Ar, C-3), 153.2 (q, CO), 153.2 (q, CO), 153.3 (q, CN), 153.4 (q, CN), 163.4 (q, CO), 163.4 (q, CO).

ν_{max} (ATR)/cm^{-1}: 3287 (NH), 3188 (NH), 2980, 2933, 1796 (CO), 1628 (CN), 1600, 1560, 1477, 1454, 1407, 1367, 1338, 1300, 1232, 1145, 1106, 1056, 1028, 804, 766, 693.

HRMS (*m/z* - ES): Found: 683.3762 (M^+ + H. $C_{35}H_{51}N_6O_8$ Requires: 683.3768).

3,4'-*Bis*-[2,3-di(*tert*-butoxycarbonyl)guanidino]-*N*-phenylaniline (41d)



Following method A, using **9d** (62 mg, 0.30 mmol) as starting material and after purification with silica gel chromatography, eluting with a gradient of hexane:EtOAc, **41d** was obtained as a pale solid (170 mg, 83%). **Mp**: 117-120 °C.

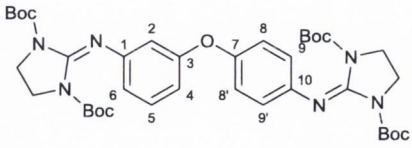
δ_H (600 MHz, $CDCl_3$): 1.52 (s, 9H, $(CH_3)_3$), 1.54 (s, 9H, $(CH_3)_3$), 1.56 (s, 9H,$(CH_3)_3$), 1.57 (s, 9H, $(CH_3)_3$), 5.78 (bs, 1H, NH linker), 6.80 (d, 1H, J 8.0, H-4), 7.01 (d, 1H, J 8.0, H-6), 7.09 (d, 2H, J 8.6, H-8 and H-8'), 7.20 (t, 1H, J 8.0, H-5), 7.45 (s, 1H, H-2), 7.51 (d, 2H, J 8.6, H-9 and H-9'), 10.2 (bs, 1H, NHBoc), 10.3 (bs, 1H, NHBoc), 11.6 (bs, 1H, NH), 11.7 (bs, 1H, NH).

δ_C (150 MHz,$CDCl_3$): 27.9 ($(CH_3)_3$), 27.9 ($(CH_3)_3$), 28.0 ($(CH_3)_3$), 28.1 ($(CH_3)_3$), 79.3 (q, $\underline{C}(CH_3)_3$), 79.4 (q, $\underline{C}(CH_3)_3$), 83.4 (q, $\underline{C}(CH_3)_3$), 83.5 (q, $\underline{C}(CH_3)_3$), 110.7 (CH Ar, C-2), 113.1 (CH Ar, C-4), 114.1 (CH Ar, C-6), 119.0 (CH Ar, C-8 and C-8'), 123.4 (CH Ar, C-9 and C-9'), 129.4 (CH Ar, C-5), 130.3 (q, Ar, C-10), 137.6 (q, Ar, C-1), 139.4 (q, Ar, C-7), 143.8 (q, Ar, C-3), 153.1 (q, CN), 153.2 (q, CN), 153.3 (q, CO), 153.4 (q, CO), 163.4 (q, CO), 163.5 (q, CO).

ν_{max} (ATR)/cm^{-1}: 3283 (NH), 1720 (CO), 1625 (CN), 1587, 1570, 1500, 1510, 1473, 1336, 1288, 1211, 1140, 1117, 1055, 879, 823, 767, 685.

HRMS (*m/z* - ES): Found: 684.3719 (M^+ + H. $C_{34}H_{50}N_7O_8$ Requires: 684.3721).

3,4'-*Bis*-[1,3-di(*tert*-butoxycarbonyl)-2-imidazolidinylimino]diphenylether (42a)



Following method A, **9a** (113 mg, 0.565 mmol) as starting material and after purification with neutral alumina chromatography, eluting with a gradient of hexane:EtOAc, **42a** was obtained as a white solid (283 mg, 68%). **Mp**: 88-89 °C.

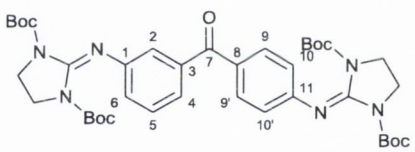
δ_H (400 MHz, $CDCl_3$): 1.37 (s, 18H, $2(CH_3)_3$), 1.38 (s, 18H, $2(CH_3)_3$), 3.81 (s, 4H, 2 CH_2), 3.84 (s, 4H, 2 CH_2), 6.52 (dd, 1H, J 8.3, 2.0, H-4), 6.60 (app. t, 1H, J 2.0, H-2), 6.72 (dd, 1H, J 8.3, 2.0, H-6), 6.91 (d, 2H, J 8.7, H-9 and H-9'), 6.98 (d, 2H, J 8.7, H-8 and H-8'), 7.13 (t, 1H, J 8.3, H-5).

δ_C (100 MHz, $CDCl_3$): 27.9 ($2(CH_3)_3$), 28.0 ($2(CH_3)_3$), 43.1 (2 CH_2), 43.2 (2 CH_2), 82.8 (q, 2 $\underline{C}(CH_3)_3$), 82.9 (q, 2 $\underline{C}(CH_3)_3$), 110.9 (CH Ar, C-6), 111.8 (CH Ar, C-2), 116.3 (CH Ar, C-4), 120.1 (CH Ar, C-9 and C-9'), 122.6 (CH Ar, C-8 and C-8'), 129.5 (CH Ar, C-5), 139.2 (q, CN), 139.5 (q, CN), 144.2 (q, Ar, C-10), 149.8 (q, Ar, C-1), 150.2 (q, CO), 150.3 (q, CO), 150.3 (q, CO), 150.7 (q, CO), 152.1 (q, Ar, C-7), 158.8 (q, Ar, C-3).

ν_{max} (ATR)/cm^{-1}: 2977, 1756 (CO), 1700 (CN), 1594, 1499, 1477, 1366, 1298, 1248, 1217, 1144, 976, 848, 766, 699.

HRMS (*m/z* - ES): Found: 736.3802 (M^+ + H. $C_{38}H_{53}N_6O_9$ Requires: 736.3796).

3,4'-*Bis*-[1,3-di(*tert*-butoxycarbonyl)-2-imidazolidinylimino]benzophenone (42b)



Following method A, using **9b** (212 mg, 1.0 mmol) as starting material and after purification with neutral alumina chromatography, eluting with a gradient of hexane:EtOAc **42b** was obtained as a pale orange solid (397 mg, 53%). **Mp**: 84-85 °C.

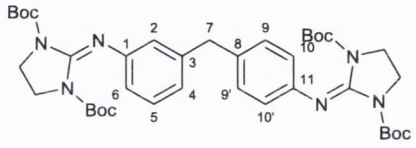
δ_H (600 MHz, $CDCl_3$): 1.29 (s, 18H, $(CH_3)_3$), 1.30 (s, 18H, $(CH_3)_3$), 3.76 (s, 4H, 2 CH_2), 3.80 (s, 4H, 2 CH_2), 6.95 (d, 2H, J 8.5, H-10 and H-10'), 7.15 (d, 1H, J 7.5, H-6), 7.22 (s, 1H, H-2), 7.26 (m, 1H, H-4), 7.28 (t, 1H, J 7.5, H-5), 7.69 (d, 2H, J 8.5, H-9 and H-9').

δ_C (150 MHz, $CDCl_3$): 27.7 ($4(CH_3)_3$), 43.0 (2 CH_2), 43.1 (2 CH_2), 82.6 (q, 2 $\underline{C}(CH_3)_3$), 82.9 (q, 2 $\underline{C}(CH_3)_3$), 120.6 (2 CH Ar, C-10 and C-10'), 121.6 (CH Ar, C-2), 123.5 (CH Ar, C-4), 125.4 (CH Ar, C-6), 128.2 (CH Ar, C-5), 131.2 (q, Ar, C-11), 131.3 (2 CH Ar, C-9 and C-9'), 138.6 (q, Ar, C-1), 139.8 (q, CN), 139.9 (q, CN), 148.4 (q, Ar, C-3), 149.7 (q, 2 CO), 150.1 (q, 2 CO), 152.7 (q, Ar, C-8), 196.9 (q, CO, C-7).

ν_{max} (ATR)/cm^{-1}: 2980, 2932, 1757, 1702 (CO), 1648 (CN), 1588, 1477, 1457, 1366, 1291, 1251, 1144, 974, 848, 766.

HRMS (*m/z* - ES): Found: 749.3871 (M^+ + H. $C_{39}H_{53}N_6O_9$ Requires: 749.3874).

3,4'-*Bis*-[1,3-di(*tert*-butoxycarbonyl)-2-imidazolidinylimino]benzylbenzene (42c)



Following method A, using **9c** (1105 mg, 1.5 mmol) as starting material and after purification with neutral alumina chromatography, eluting with a gradient of hexane:EtOAc, **42c** was obtained as a white solid (973 mg, 88%). **Mp**: 90-92 °C.

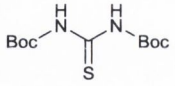
δ_H (600 MHz, $CDCl_3$): 1.33 (s, 36H, 4$(CH_3)_3$), 3.83 (s, 8H, 4 CH_2), 3.84 (s, 2H, H-7), 6.73 (d, 1H, J 7.8, H-4), 6.82 (d, 1H, J 7.8, H-6), 6.83 (s, 1H, H-2), 6.91 (d, 2H, J 8.3, H-10 and H-10'), 7.06 (d, 2H, J 8.3, H-9 and H-9'), 7.09 (t, 1H, J 8.3, H-5).

δ_C (150 MHz, $CDCl_3$): 27.7 (4$(CH_3)_3$), 41.2(CH_2, C-7), 42.9 (4 CH_2), 82.5 (q, 2 $\underline{C}(CH_3)_3$), 82.6 (q, 2 $\underline{C}(CH_3)_3$), 118.9 (CH Ar, C-6), 121.2 (CH Ar, C-10 and C-10'), 122.9 (CH Ar, C-2), 123.2 (CH Ar, C-4), 128.4 (CH Ar, C-5), 129.2 (CH Ar, C-9 and C-9'), 135.2(q, Ar, C-8), 138.6 (q, CN), 138.7 (q, CN), 141.8 (q, Ar, C-3), 146.(q, Ar, C-11), 148.1 (q, Ar, C-1), 150.2 (q, 2 CO), 150.3 (q, 2 CO).

ν_{max} (ATR)/cm^{-1}: 2977, 2929, 1754 (CO), 1695 (CN), 1596, 1477, 1457, 1366, 1297, 1247, 1143, 974, 848, 765, 695.

HRMS (*m/z* Maldi): Found: 757.3864 (M^+ + Na. $C_{39}H_{54}N_6O_8Na$ Requires: 757.3901).

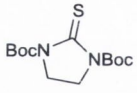
N,N'-Bis-(_tert_-butoxycarbonyl)thiourea (43)



To a cooled solution of thiourea (3.32 g, 43.56 mmol) in dry tetrahydrofuran (600 mL) under argon, sodium hydride as a 60% suspension in mineral oil (7.49 g, 312.2 mmol) was added. After 5 min, the ice-bath was removed and the reaction was stirred for 10 min at room temperature. The reaction mixture was again cooled to 0 °C and di-*tert*-butyldicarbonate (20.08 g, 92.04 mmol) was added. After 30 min, the ice-bath was removed and the reaction mixture was stirred overnight at room temperature. The reaction was quenched by the dropwise addition of saturated $NaHCO_3$ solution (5 mL) and extracted three times with EtOAc (3 × 40 mL). The organic phase was washed with brine (25 mL), dried over MgSO4, filtered, and concentrated under vacuum. The crude product was purified by washing with hexane followed by ether which afforded **43** a white solid (9.61 g, 80% yield). **Mp**: 136-138 °C (lit. 136-138 °C).[8]

δ_H (400 MHz, $CDCl_3$): 1.55 (s, 18H, $(CH_3)_3$), 3.93 (bs, 2H, NH).

1,3-Di(_tert_-butoxycarbonyl)imidazolidine-2-thione (46)

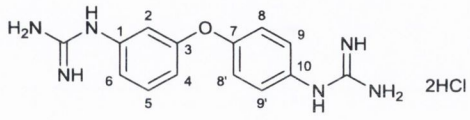


To a cooled solution of the imidazolidine-2-thione (500 mg, 4.92 mmol) in dry tetrahydrofuran (40 mL) under argon, NaH as a 60% suspension in mineral oil (882 mg, 22.1 mmol) was added. After 5 min, the ice-bath was removed and the reaction was stirred for 10 min at room temperature. The reaction mixture was again cooled to 0 °C and di-*tert*-butyldicarbonate (2.38 g, 10.78 mmol) was added neat. After 30 min, the ice-bath was removed and the reaction mixture was stirred overnight h at room temperature (reaction progress adjudged by TLC). The reaction was quenched by the dropwise addition of saturated $NaHCO_3$ solution (5 mL) and extracted with EtOAc (3 × 40 mL). The organic phase was washed with brine (25 mL), dried over $MgSO_4$, filtered, and concentrated under vacuum. The product was purified by flash chromatography using alumina column (eluting with hexane:EtOAc) and recrystallised from hexane slowly (1 to 2 days) when dry, to

afford the title compound as lustrous yellow needles (1.24 g, 84%). **Mp**: 115 °C (lit.: 117-119 °C).[8]

δ_H (400 MHz, $CDCl_3$): 1.47 (s, 18H, Boc), 3.90 (s, 4H, CH_2).

3-[4'-(Guanidino)phenoxy]phenylguanidine dihydrochloride (51a)



Following method E, using **41a** (151 mg, 0.22 mmol) as starting material, the hydrochloride salt **51a** was obtained as a yellow solid (60 mg, 76%). **Mp**: 159-160 °C.

δ_H (600 MHz, D_2O): 6.90 (app. t, 1H, J 2.1, H-2), 7.00 (dd, 1H, J 8.3, 2.1, H-6), 7.06 (dd, 1H, J 8.3, 2.1, H-4), 7.08 (d, 2H, J 8.8, H-9 and H-9'), 7.27 (d, 2H, J 8.8, H-8 and H-8'), 7.43 (t, 1H, J 8.3, H-5).

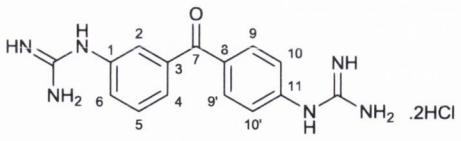
δ_C (600 MHz, D_2O): 115.7 (CH Ar, C-6), 117.9 (CH Ar, C-2), 120.2 (2 CH Ar, C-9 and C-9'), 120.7 (CH Ar, C-4), 127.9 (2 CH Ar, C-8 and C-8'), 129.6 (q, Ar, C-10), 131.3 (CH, Ar, C-5), 135.5 (q, Ar, C-1), 155.5 (q, Ar, C-7), 155.9 (q, CN), 156.3 (q, CN), 157.4 (q, Ar, C-3).

ν_{max} (ATR)/cm^{-1}: 3311 (NH), 3134 (NH), 1665 (CN), 1580, 1505, 1487, 1214, 1154, 1016, 962, 843.

HRMS (*m/z* - ES): Found: 284.1394 (M^+ + H. $C_{14}H_{17}N_6O$ Requires: 284.1386).

HPLC: 98.6% (*t*R: 19.8 min).

3-[4'-(Guanidino)benzoyl]phenylguanidine dihydrochloride (51b)



Following method E, using **41b** (116 mg, 0.55 mmol) as starting material, the dihydrochloride salt **51b** was obtained as a yellow solid (32 mg, 53%). **Mp**: 177-179 °C.

δ_H (600 MHz, D_2O): 7.44 (d, 2H, J 8.6, H-10 and H-10'), 7.60 (d, 1H, J 7.5, H-6), 7.63 (t, 1H, J 7.5, H-5), 7.69 (s, 1H, H-2), 7.73 (d, 1H, J 7.5, H-4), 7.86 (d, 2H, J 8.6, H-9 and H-9').

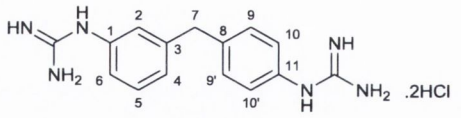
δ_C (150 MHz, D_2O): 124.2 (2 CH Ar, C-10 and C-10'), 127.1 (CH Ar, C-2), 129.6 (CH Ar, C-4), 130.3 (CH Ar, C-5), 130.6 (CH Ar, C-6), 132.1 (2 CH Ar, C-9 and C-9'), 134.6 (q, Ar, C-8), 134.7 (q, Ar, C-3), 138.3 (q, Ar, C-1), 139.3 (q, Ar, C-11), 155.8 (q, CN), 155.9 (q, CN), 198.2 (q, CO, C-7).

ν_{max} (ATR)/cm^{-1}: 3302 (NH), 3130 (NH), 2901, 1644 (CN), 1589, 1485, 1382, 1285, 1083, 1019, 933, 844.

HRMS (*m/z* - ES): Found: 297.1471 (M^+ + H. $C_{15}H_{17}N_6O$ Requires: 297.1464).

HPLC: 97.3% (*t*R: 19.6 min).

3-[4'-(Guanidino)benzyl]phenylguanidine dihydrochloride (51c)



Following method E, using **41c** (44 mg, 0.065 mmol) as starting material, the hydrochloride salt **51c** was obtained as a brown solid (16 mg, 70%). **Mp**: 97-99 °C.

δ_H (600 MHz, D_2O): 4.01 (s, 2H, H-7), 7.14 (d, 1H, J 7.8, H-6), 7.17 (s, 1H, H-2), 7.22 (d, 2H, J 8.3, H-10 and H-10'), 7.27 (d, 1H, J 7.8, H-4), 7.34 (d, 2H, J 8.3, H-9 and H-9'), 7.38 (t, 1H, J 7.8, H-5).

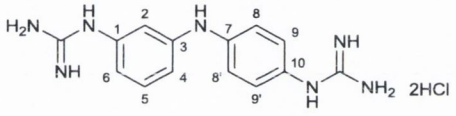
δ_C (150 MHz, D_2O): 40.1 (CH_2), 123.6 (CH Ar, C-6), 125.9 (CH Ar, C-2), 126.2 (CH Ar, C-10 and C-10'), 128.2 (CH Ar, C-4), 130.1 (CH Ar, C-5), 130.1 (CH Ar, C-9 and C-9'), 132.1 (q, Ar, C-11), 134.2 (q, Ar, C-1), 140.9 (q, Ar, C-8), 143.2 (q, Ar, C-3), 156.2 (q, CN), 156.3 (q, CN).

ν_{max} (ATR)/cm^{-1}: 3159 (NH), 2981 (NH), 2904, 2475, 1642 (CN), 1585, 1514, 1414, 1487, 1417, 1382, 1288, 1200, 1084, 1019, 839.

HRMS (*m/z* - ES): Found: 283.1673 (M^+ + H. $C_{15}H_{19}N_6$ Requires: 283.1671).

HPLC: 95.1% (*t*R: 20.4 min).

3-[4'-(Guanidino)phenylamino]phenylguanidine dihydrochloride (51d)



Following method E, using **41d** (50 mg, 0.07 mmol) as starting material, the dihydrochloride salt **51d** was obtained as a yellow-brown solid (16 mg, 62%). **Mp**: decomp. 229-235 °C.

δ_H (600 MHz, D_2O): 6.80 (d, 1H, J 8.0, H-6), 7.01 (s, 1H, H-2), 7.09 (d, 1H, J 8.0, H-4), 7.17 (d, 2H, J 8.8, H-9 and H-9'), 7.21 (d, 2H, J 8.8, H-8 and H-8'), 7.36 (t, 1H, J 8.0, H-5).

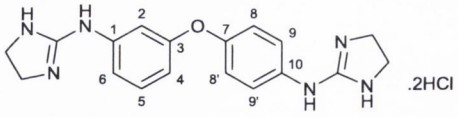
δ_C (150 MHz, D_2O): 114.5 (CH Ar, C-2), 117.1 (CH Ar, C-4), 118.1 (CH Ar, C-6), 118.8 (2 CH Ar, C-9 and C-9'), 126.8 4 (q, Ar, C-7), 127.7 (2 CH Ar, C-8 and C-8'), 130.7 (CH Ar, C-5), 135.1 (q, Ar, C-3), 142.6 (q, Ar, C-10), 144.4 (q, Ar, C-1), 156.2 (q, CN), 156.6 (q, CN).

ν_{max} (ATR)/cm^{-1}: 3289 (NH), 3127 (NH), 1659 (CN), 1582, 1508, 1505, 1331, 1238, 1167, 880.

HRMS (*m/z* - ES): Found: 284.1621 (M^+ + H. $C_{14}H_{18}N_7$ Requires: 284.1624).

HPLC: 96.1% (*t*R: 21.99 min).

3-[4'-*N*-(Amino-2-imidazolinyl)phenoxy]phenylamino-2-imidazoline dihydrochloride (52a)



Following method E, using **42a** (78 mg, 0.191 mmol) as a starting material, the dihydrochloride salt **52a** was obtained as a white solid (60 mg, 76%). **Mp**: 146-150 °C.

δ_H (400 MHz, D_2O): 3.65 (s, 4H, CH_2), 3.65 (s, 4H, CH_2), 6.87 (app. t, 1H, J 2.0, H-2), 6.92 (dd, 1H, J 8.0, 2.0, H-4), 6.99 (dd, 1H, J 8.0, 2.0, H-6), 7.05 (d, 2H, J 8.8, H-9 and H-9'), 7.22 (d, 2H, J 8.8, H-8 and H-8'), 7.36 (t, 1H, J 8.0, H-5).

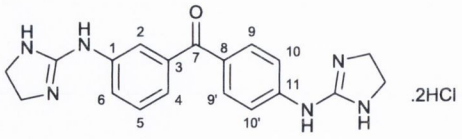
δ_C (100 MHz, D_2O): 42.7 (2 CH_2), 42.7 (2 CH_2), 114.5 (CH Ar, C-2), 117.4 (CH Ar, C-4), 119.4 (CH Ar, C-6), 120.2 (CH Ar, C-9 and C-9'), 126.7 (CH, Ar, C-8 and C-8'), 130.8 (q, Ar, C-7 or C-10), 131.2 (CH, Ar, C-5), 136.5 (q, Ar, C-1 or C-3), 155.8 (q, Ar, C-7 or C-10), 157.5 (q, Ar, C-1 or C-3), 158.6 (q, CN), 159.1 (q, CN).

ν_{max} (ATR)/cm^{-1}: 3155 (NH), 2978, 2903, 1647 (CN), 1588, 1505, 1484, 1384, 1269, 1212, 1150, 1086, 1014, 956, 847.

HRMS (*m/z* - ES): Found: 337.1763 (M^+ + H. $C_{18}H_{21}N_6O$ Requires: 337.1777).

HPLC: 98.0% (*t*R: 20.79 min).

3-[4'-*N*-(Amino-2-imidazolinyl)benzoyl]phenylamino-2-imidazoline dihydrochloride (52b)



Following method E, using **42b** (144 mg, 0.07 mmol) as starting material, the hydrochloride salt **52b** was obtained as a brown solid (44 mg, 55%). **Mp**: 155-158 °C.

δ_H (600 MHz, D_2O): 3.72 (s, 4H, 2 CH_2), 3.77 (s, 4H, 2 CH_2), 7.33 (d, 2H, J 8.6, H-10 and H-10'), 7.52 (s, 1H, H-2), 7.53 (d, 1H, J 7.6, H-6), 7.56 (t, 1H, J 7.6, H-5), 7.60 (d, 1H, J 7.6, H-4), 7.77 (d, 2H, J 8.6, H-9 and H-9').

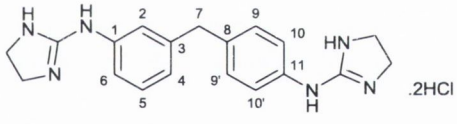
δ_C (150 MHz, D_2O): 42.6 (4 CH_2), 122.1 (2 CH Ar, C-10 and C-10'), 125.0 (CH Ar, C-2), 128.6 (CH Ar, C-6), 128.8 (CH Ar, C-4), 130.1 (CH Ar, C-5), 132.0 (2 CH Ar, C-9 and C-9'), 133.7 (q, Ar, C-8), 135.3 (q, Ar, C-3), 137.9 (q, Ar, C-1), 140.1 (q, Ar, C-11), 157.7 (q, CN), 158.4 (q, CN), 197.5 (q, CO, C-7).

ν_{max} (ATR)/cm^{-1}: 3299 (NH), 3122 (NH), 1627 (CO), 1593, 1559 (CN), 1453, 1292, 1257, 1088, 968, 845, 808, 759.

HRMS (*m/z* - ES): Found: 349.1781 (M^+ + H. $C_{19}H_{21}N_6O$ Requires: 349.1777).

HPLC: 95.9% (*t*R: 20.6 min).

3-[4'-*N*-(Amino-2-imidazolinyl)benzyl]phenylamino-2-imidazoline dihydrochloride (52c)



Following method E, using **42c** (400 mg, 0.55 mmol) as starting material, the hydrochloride salt **52c** was obtained as a yellow solid (153 mg, 70%). **Mp**: 94-96 °C.

δ_H (600 MHz, D_2O): 3.69 (s, 8H, 4 CH_2), 3.99 (s, 2H, H-7), 7.11 (d, 1H, J 7.8, H-6), 7.13 (s, 1H, H-2), 7.19 (d, 2H, J 8.3, H-10 and H-10'), 7.22 (d, 1H, J 7.8, H-4), 7.31 (d, 2H, J 8.3, H-9 and H-9'), 7.36 (t, 1H, J 7.8, H-5).

δ_C (150 MHz, D_2O): 40.1 (CH_2, C-7), 42.6 (4 CH_2), 122.1 (CH Ar, C-6), 124.5 (CH Ar, C-2), 124.7 (CH Ar, C-10 and C-10'), 127.7 (CH Ar, C-4), 129.9 (CH Ar, C-5), 130.0 (CH Ar, C-9 and C-9'), 133.1 (q, Ar, C-11), 135.2 (q, Ar, C-1), 140.4 (q, Ar, C-8), 143.1 (q, Ar, C-3), 158.7 (q, CN), 158.8 (q, CN).

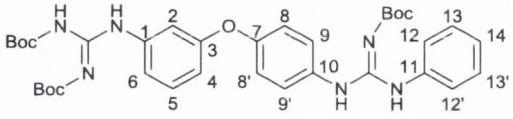
ν_{max} (ATR)/cm^{-1}: 3138 (NH), 2898, 1644 (CN), 1585, 1513, 1435, 1381, 1281, 1282, 1201, 1083, 1018, 983.

HRMS (*m/z* - ES): Found: 335.1984 (M^+ + H. $C_{19}H_{23}N_6$ Requires: 335.1984).

HPLC: 98.8% (*t*R: 21.1 min).

For compounds **53a-d, 54a-d, 66a, 66d, 67a, 67d, 76a, 76d, 77a, 77d** not all signals from ^1H and ^{13}C-NMR spectra could be assigned due to highly broad signals and extra peaks found in the spectra, since different isomeric forms were present. Integration was not possible due to the same reasons.

3-[2,3-Di-(*tert*-butoxycarbonyl)guanidino]-4'-[2-(*tert*-butoxycarbonyl)-3-(phenyl) guanidino]diphenylether (53a)



Following Method C, to a solution of **66a** (395 mg, 0.944 mmol), **43** (313 mg, 1.13 mmol) and triethylamine (407 µl, 2.93 mmol) in DCM (5 mL), $HgCl_2$ (307 mg, 1.13 mmol) was added. The reaction mixture was stirred at room temperature for 5 h and 30 min. Usual work-up followed by silica gel chromatography (eluting with a gradient of hexane:EtOAc) to afford the desired product as transparent hydroscopic solid (471 mg, 76%). **Mp**: 90-91 °C.

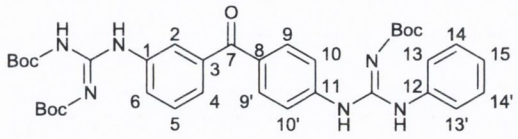
δ_H (600 MHz, $CDCl_3$): 1.52 (s, 9H, $(CH_3)_3$), 1.55 (s, 9H, $(CH_3)_3$), 6.71 (bs, NH), 6.67-6.79 (m, CH Ar), 6.86-7.15 (m, CH Ar), 7.21-7.43 (m, CH Ar), 7.53 (d, J 7.7, CH Ar), 7.65-7.80 (m, CH Ar), 9.57 (bs, NH), 9.87 (bs, NH), 10.33 (bs, NH), 11.6 (bs, NH).

δ_C (150 MHz, $CDCl_3$): 27.9 ($(CH_3)_3$), 28.0 ($2(CH_3)_3$), 79.7 (q, $\underline{C}(CH_3)_3$), 83.3 (q, $\underline{C}(CH_3)_3$), 83.7 (q, $\underline{C}(CH_3)_3$), 112.0 (CH Ar), 114.0 (CH Ar), 116.5 (CH Ar), 119.8 (2 CH Ar), 120.0 (CH Ar), 122.5 (CH Ar), 124.0 (CH Ar), 128.8 (2 CH Ar), 129.6 (CH Ar), 130 – 150 (Quaternary signals in this range could not be assigned as they were highly amorphous and there was not enough resolution), 137.8 (q), 153.1 (q), 153.2 (q), 162.9 (q), 163.2 (q).

ν_{max} (ATR)/cm^{-1}: 3407 (NH), 3285, 2978, 2932, 1718 (CO), 1596 (CN), 1555, 1490, 1408, 1367, 1331, 1289, 1233, 1143, 1105, 1056, 946, 804, 752, 690.

HRMS (*m/z* - ES): Found: 661.3339 (M^+ + H. $C_{35}H_{45}N_6O_7$ Requires: 661.3350).

3-[2,3-Di-(*tert*-butoxycarbonyl)guanidino]-4'-[2-(*tert*-butoxycarbonyl)-3-(phenyl) guanidino]benzophenone (53b)



Following Method C, to a solution of **68b** (186 mg, 0.41 mmol), **57** (124 mg, 0.49 mmol) and triethylamine (177 µl, 1.27 mmol) in DCM (2.1 mL), $HgCl_2$ (133 mg, 0.49 mmol) was added. The reaction mixture was stirred at room temperature for 5 h. Usual work-up followed by silica gel chromatography (eluting with a gradient of hexane:EtOAc) to afford the desired product as light brown solid (228 mg, 83%). **Mp**: 119-121 °C.

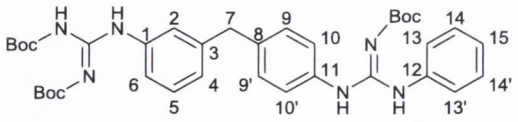
δ_H (600 MHz, $CDCl_3$): 1.52 (s, 9H, $(CH_3)_3$), 1.56 (s, 9H, $(CH_3)_3$), 6.75 (bs, NH), 6.97 (bs, CH Ar), 7.12 (bs, CH Ar), 7.32-7.42 (m, CH Ar), 7.47 (bs, CH Ar), 7.78-8.04 (m, CH Ar), 9.56-10.0 (bs, NH), 10.4-10.5 (bs, NH), 11.6 (bs, NH).

δ_C (150 MHz, $CDCl_3$): 27.9 ($(CH_3)_3$), 28.0 ($(CH_3)_3$), 79.6 (q, $\underline{C}(CH_3)_3$), 83.5 (q, $\underline{C}(CH_3)_3$), 83.7 (q, $\underline{C}(CH_3)_3$), 118.6 (CH Ar), 122.2 (CH Ar), 123.1 (CH Ar), 123.5 (CH Ar), 123.6 (CH Ar), 126.0 (CH Ar), 126.1 (CH Ar), 128.8 (CH Ar), 129.7 (CH Ar), 130.9 (CH Ar), 131.6 (CH Ar), 132.2 (CH Ar), 136.6 (q), 136.7 (q), 138.5 (q), 138.7 (q), 139.4 (q), 143.3 (q), 146.1 (q), 152.9 (q), 153.2 (q), 153.6 (q), 153.6 (q), 163.4 (q), 194.7 (q, CO linker).

ν_{max} (ATR)/cm^{-1}: 3407 (NH), 3256, 2978, 2932, 1717 (CO), 1637 (CN), 1584, 1546, 1461, 1409, 1368, 1293, 1233, 1148, 1106, 1057, 941, 848, 804, 758, 691.

HRMS (*m/z* - ES): Found: 673.3339 (M^+ + H. $C_{36}H_{45}N_6O_7$ Requires: 673.3350).

3-[2,3-Di-(*tert*-butoxycarbonyl)guanidino]-4'-[2-(*tert*-butoxycarbonyl)-3-(phenyl) guanidino]benzylbenzene (53c)



Following Method C, to a solution of **68c** (30 mg, 0.07 mmol), **57** (21 mg, 0.084 mmol) and triethylamine (30 μl, 0.22 mmol) in DCM (0.36 mL), $HgCl_2$ (23 mg, 0.084 mmol) was added. The reaction mixture was stirred at room temperature for 4 h. Usual work-up followed by silica gel chromatography (eluting with a gradient of hexane:EtOAc) to afford the desired product as off-white solid (50 mg, 72%). **Mp**: 92-94 °C.

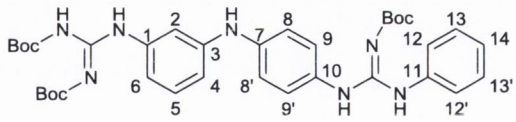
δ_H (600 MHz, $CDCl_3$): 1.50 (s, 9H, $(CH_3)_3$), 1.53 (s, 9H, $(CH_3)_3$), 1.56 (s, 9H, $(CH_3)_3$), 3.95 (s, 2H, CH_2), 6.71 (bs, NH), 6.80 – 7.44 (Aromatic signals in this range could not be assigned as they were broad and highly amorphous), 7.56 (d, 2H, J 7.8, CH Ar) 7.60-7.77 (m, CH Ar) 9.54 (bs, NH), 10.30 (bs, NH), 11.60 (bs, NH).

δ_C (150 MHz, $CDCl_3$): 28.0 (2 $CH_3)_3$), 28.2 ($(CH_3)_3$), 41.2 (CH_2), 79.5 (q, $\underline{C}(CH_3)_3$), 83.0 (q, $\underline{C}(CH_3)_3$), 83.6 (q, $\underline{C}(CH_3)_3$), 119.8 – 129.8 (Aromatic signals in this range could not be assigned as they were broad and highly amorphous) 122.6 (2 CH Ar), 129.8 (2 CH Ar), 128.9 (CH Ar), 135 – 150 (Quaternary signals in this range could not be assigned as they were highly amorphous and there was not enough resolution), 136.8 (q), 152.9 (q) 153.3 (q), 153.5 (q), 163.5 (q).

ν_{max} (film)/cm^{-1}: 3408 (NH), 3266, 2983, 2939, 1720 (CO), 1646 (CN), 1597, 1568, 1480, 1466, 1331, 1300, 1238, 1207, 1147, 1098, 1070, 838, 802, 765, 678.

HRMS (*m/z* - ES): Found: 659.3540 (M^+ + H. $C_{36}H_{47}N_6O_6$ Requires: 659.3557).

3-[2,3-Di-(*tert*-butoxycarbonyl)guanidino]-4'-[2-(*tert*-butoxycarbonyl)-3-(phenyl) guanidino]-*N*-phenylaniline (53d)



Following Method C, to a solution of **66d** (328 mg, 0.78 mmol), **43** (260 mg, 0.94 mmol) and triethylamine (337 μl, 2.42 mmol) in DCM (4 mL), $HgCl_2$ (254 mg, 0.94 mmol) was added. The reaction mixture was stirred at room temperature for 5 h. Usual work-up followed by silica gel chromatography (eluting with a gradient of hexane:EtOAc) to afford the desired product as light dark solid (168 mg, 33%). **Mp**: 95-98 °C.

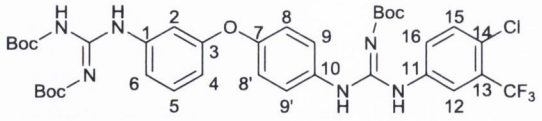
δ_H (600 MHz, $CDCl_3$): 1.52 (s, 18H, $2(CH_3)_3$), 1.56 (s, 9H, $(CH_3)_3$), 5.82 (bs, NH), 6.03 (bs, NH), 6.68-7.73 (m, 13H, CH Ar), 9.46 (bs, NH), 9.56 (bs, NH), 10.30 (bs, 1H, NH), 11.50 (bs, 1H, NH).

δ_C (150 MHz, $CDCl_3$): 27.9 (2 ($(CH_3)_3$), 28.0 ($(CH_3)_3$), 78.8 (q, $\underline{C}(CH_3)_3$), 79.5 (q, $\underline{C}(CH_3)_3$), 82.9 (q, $\underline{C}(CH_3)_3$), 83.5 (q, $\underline{C}(CH_3)_3$), 110.3 (CH Ar), 111.8 (CH Ar), 112.6 (CH Ar), 113.6 (CH Ar), 113.9 (CH Ar), 114.8 (CH Ar), 118.4 (CH Ar), 119.7 (CH Ar) , 119.9 (CH Ar), 120.7 (CH Ar), 120.7 (CH Ar), 122.6 (CH Ar), 123.1 (CH Ar), 124.9 (CH Ar), 126.1 (CH Ar), 128.6 (CH Ar), 129.0 (CH Ar), 129.4 (2 CH Ar), 133.2 (q), 136.9 (q), 137.6 (q), 139.0 (q), 140.1 (q), 140.7 (q), 141.2 (q), 143.2 (q), 144.6 (q), 147.0 (q), 152.8 (q, CO), 153.1 (q, CO), 153.5 (q, CO), 156.6 (q), 163.0 (q, CN).

ν_{max} (ATR)/cm^{-1}: 3406 (NH), 3299, 2978, 2930, 2167, 1718 (CO), 1593 (CN), 1478, 1497, 1459, 1406, 1367, 1326, 1233, 1146, 1105, 1057, 903, 804, 771, 750, 690.

HRMS (*m/z* - ES): Found: 660.3527 (M^+ + H. $C_{35}H_{46}N_7O_6$ Requires: 660.3510).

3-[2,3-Di(*tert*-butoxycarbonyl)guanidino]-4'-[2-(*tert*-butoxycarbonyl)-3-(4-chloro-3-trifluoromethylphenyl)guanidino]diphenylether (54a)



Following Method C, to a solution of **67a** (60 mg, 0.12 mmol), **43** (36 mg, 0.14 mmol) and triethylamine (50.2 μl, 0.36 mmol) in DCM (0.6 mL), $HgCl_2$ (38 mg, 0.14 mmol) was added. The reaction mixture was stirred at room temperature for 5 h and 30 min. Usual work-up followed by silica gel chromatography (eluting with a gradient of hexane:EtOAc) to afford the desired product as off-white hygroscopic solid (67 mg, 76%). **Mp**: decomp. above 114 °C.

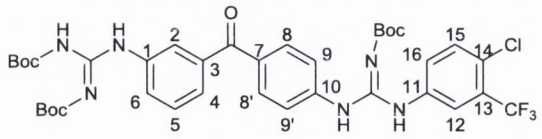
δ_H (600 MHz, $CDCl_3$): 1.52 (s, 9H, $(CH_3)_3$), 1.54 (s, 9H, $(CH_3)_3$), 1.55 (s, 9H, $(CH_3)_3$), 6.75 (bs, CH Ar), 6.92 (bs, CH Ar), 6.99-7.14 (m, CH Ar), 7.22-7.48 (m, CH Ar), 7.60 (bs, CH Ar), 7.92-8.07 (m, CH Ar), 9.64 (bs, NH), 9.84 (bs, NH), 10.30 (bs, NH), 11.60 (bs, NH).

δ_C (150 MHz, $CDCl_3$): 27.9 ($(CH_3)_3$), 28.0 ($(CH_3)_3$), 28.1 ($(CH_3)_3$), 79.6 (q, $\underline{C}(CH_3)_3$), 83.8 (q, $\underline{C}(CH_3)_3$), 112.2 (CH Ar), 114.3 (CH Ar), 116.7 (CH Ar), 118.7 (CH Ar), 119.8 (CH Ar), 120.9 (CH Ar), 122.0 (CH Ar), 123.4 (CH Ar), 123.6 (c, q, J 274, CF_3), 124.7 (q), 126.7 (CH Ar), 128.4 (c, q, J 30, C-13), 129.7 (CH Ar), 131.6 (CH Ar), 132.4 (CH Ar), 137.9 (q), 138.1 (q), 140.1 (q), 141.8 (q), 152.5 (q), 153.0 (q), 153.2 (q), 153.4 (q), 158.1 (q), 163.4 (q).

ν_{max} (ATR)/cm^{-1}: 3290, 3119, 2966, 1654 (CO), 1561, 1482, 1468, 1405, 1316, 1258, 1176, 1136, 1035, 960, 834, 808, 681.

HRMS (*m/z* - ES): Found: 763.2825 (M^+ + H. $C_{36}H_{43}N_6O_7F_3Cl$ Requires: 763.2834).

3-[2,3-Di(*tert*-butoxycarbonyl)guanidino]-4'-[2-(*tert*-butoxycarbonyl)-3-(4-chloro-3-trifluoromethylphenyl)guanidino]benzophenone (54b)



Following Method C, to a solution of **68b** (100 mg, 0.22 mmol), **58** (94 mg, 0.264 mmol) and triethylamine (95 µl, 0.68 mmol) in DCM (1.1 mL), $HgCl_2$ (72 mg, 0.264 mmol) was added. The reaction mixture was stirred at room temperature for 5 h. Usual work-up followed by silica gel chromatography (eluting with a gradient of hexane:EtOAc) to afford the desired product as light brown solid (159 mg, 94%). **Mp**: 88-90 °C.

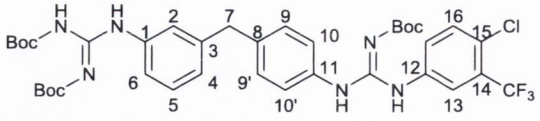
δ_H (600 MHz, $CDCl_3$): 1.52 (s, 9H, $(CH_3)_3$), 1.53 (s, 9H, $(CH_3)_3$), 1.56 (s, 9H, $(CH_3)_3$), 6.49 (bs, NH), 6.66 (bs, NH), 7.04 (bs, 1H, CH Ar) 7.28 (bs, 2H, CH Ar), 7.43-7.58 (m, 3H, CH Ar), 7.81-8.03 (m, 5H, CH Ar), 9.95 (bs, NH), 10.0 (bs, NH), 10.4 (bs, NH), 10.5 (bs, NH), 11.6 (bs, 1H, NH).

δ_C (150 MHz, $CDCl_3$): 28.0 ($(CH_3)_3$), 28.0 ($(CH_3)_3$), 28.1 ($(CH_3)_3$), 79.8 (q, $\underline{C}(CH_3)_3$), 83.9 (q, $\underline{C}(CH_3)_3$), 84.2 (q, $\underline{C}(CH_3)_3$), 118.9 (2 CH Ar), 121.9 (CH Ar), 122.1 (CH Ar), 122.7 (c, q, J 274, CF_3), 123.8 (3 CH Ar), 126.1 (2 CH Ar), 126.2 (2 CH Ar), 126.4 (CH Ar), 128.9 (CH Ar), 131.4 (2 CH Ar), 132.4 (2 CH Ar), 132.7 (CH Ar), 136.8 – 145.6 (Quaternary signals in this range could not be assigned as they were broad and highly amorphous), 152.9 (q), 153.2 (q), 153.7 (q), 163.4 (q), 195.0 (q, CO linker).

ν_{max} (ATR)/cm^{-1}: 3328 (NH), 2958, 2927, 2858, 2163, 1979, 1721 (CO), 1662 (CN), 1596, 1547, 1482, 1420, 1321, 1260, 1173, 1111, 1072, 1033, 896, 827, 797, 742, 703, 664.

HRMS (*m/z* - ES): Found: 775.2837 (M^+ + H. $C_{37}H_{43}N_6O_7F_3Cl$ Requires: 775.2834).

3-[2,3-Di-(*tert*-butoxycarbonyl)guanidino]-4'-[2-(*tert*-butoxycarbonyl)-3-(4-chloro-3-trifluoromethylphenyl)guanidino]benzylbenzene (54c)



Following Method C, to a solution of **68c** (75 mg, 0.17 mmol), **58** (72.64 mg, 0.21 mmol) and triethylamine (73.9 µl, 0.53 mmol) in DCM (1 mL) and isopropanol (0.3 mL), $HgCl_2$ (56 mg, 0.21 mmol) was added. The reaction mixture was stirred at room temperature for 5 h and 30 min. Usual work-up followed by silica gel chromatography (eluting with a gradient of hexane:EtOAc) to afford the desired product as off-white solid (91 mg, 70%). **Mp**: 92-94 °C.

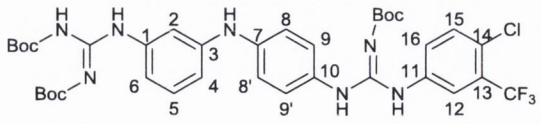
δ_H (600 MHz, $CDCl_3$): 1.53 (s, 9H, $(CH_3)_3$), 1.56 (s, 9H, $(CH_3)_3$), 3.95 (s, 2H, CH_2), 6.79 (bs, NH), 6.86 (bs, CH Ar), 6.95 (bs, CH Ar), 6.99 (bs, CH Ar), 7.17 (bs, CH Ar), 7.22 (bs, CH Ar), 7.28 (bs, CH Ar), 7.36 (bs, CH Ar), 7.42 (bs, CH Ar), 7.57 (bs, CH Ar), 7.58 (bs, CH Ar), 7.92-8.09 (m, CH Ar), 9.46-9.94 (m, NH), 10.3 (bs, NH), 11.7 (bs, NH).

δ_C (150 MHz, $CDCl_3$): 27.9 (2 $(CH_3)_3$), 28.0 ($(CH_3)_3$), 41.1 (CH_2), 79.4 (q, $\underline{C}(CH_3)_3$), 83.5 (q, $\underline{C}(CH_3)_3$), 118.5 (CH Ar), 120.2 (CH Ar), 122.1 (CH Ar), 122.4 (CH Ar), 122.5 (CH Ar), 123.5 (CH Ar), 125.2 (CH Ar), 126.6 (q), 128.9 (CH Ar), 129.3 (CH Ar), 130.1 (CH Ar), 131.5 (CH Ar), 132 – 151 (Quaternary signals in this range could not be assigned as they were highly amorphous and there was not enough resolution), 136.7 (q), 153.2 (q), 153.4 (q), 163.1 (q), 163.4 (q).

ν_{max} (ATR)/cm^{-1}: 2978, 2929, 2162, 1719 (CO), 1593 (CN), 1554, 1464, 1409, 1368, 1318, 1240, 1147, 1107, 1057, 1030, 770, 694.

HRMS (*m/z* - ES): Found: 761.3055 (M^+ + H. $C_{37}H_{45}N_6O_6F_3Cl$ Requires: 761.3041).

3-[2,3-Di-(*tert*-butoxycarbonyl)guanidino]-4'-[2-(*tert*-butoxycarbonyl)-3-(4-chloro-3-trifluoromethyl)phenylguanidino]-*N*-phenylaniline (54d)



Following Method C, to a solution of **67d** (203 mg, 0.39 mmol), **43** (130 mg, 0.47 mmol) and triethylamine (168.7 µl, 1.21 mmol) in DCM (2.0 mL), $HgCl_2$ (128 mg, 0.47 mmol) was added. The reaction mixture was stirred at room temperature for 5 h. Usual work-up followed by silica gel chromatography (eluting with a gradient of hexane:EtOAc) to afford the desired product as light dark solid (272 mg, 76%). **Mp**: 108-110 °C.

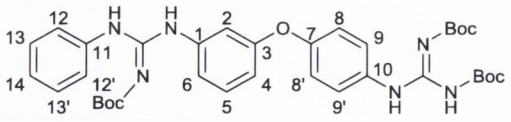
δ_H (600 MHz, $CDCl_3$): 1.49 (s, 9H, $(CH_3)_3$), 1.51 (s, 9H, $(CH_3)_3$), 1.55 (s, 9H, $(CH_3)_3$), 5.94 (bs, 1H, NH), 6.76 (app. d, 1H, J 6.5, CH Ar), 6.85 (bs, 1H, CH Ar), 6.94 (app. d, 1H, J 6.5, CH Ar), 7.00-7.23 (m, 4H, CH Ar), 7.36-7.50 (m, 4H, CH Ar), 7.90-8.08 (m, 1H, CH Ar), 9.82 (bs, NH), 9.56 (bs, NH), 10.2 (bs, NH), 10.3 (bs, NH), 11.7 (bs, 1H, NH).

δ_C (150 MHz, $CDCl_3$): 27.9 ($2(CH_3)_3$), 28.0 ($(CH_3)_3$), 79.3 (q, $\underline{C}CH_3$), 79.4 (q, $\underline{C}CH_3$), 83.5 (q, $\underline{C}CH_3$), 110.7 (CH Ar), 112.7 (CH Ar), 113.9 (CH Ar), 118.8 (CH Ar), 119.2 (CH Ar), 120.2 (CH Ar), 121.7 (c, q, J 274, CF_3), 122.8 (CH Ar), 123.5 (CH Ar), 123.6 (CH Ar), 124.4 (CH Ar), 125.3 (CH Ar), 126.6 (CH Ar), 129.3 (CH Ar), 131.4 (CH Ar), 132.0 (CH Ar), 137.5 (q), 137.9 (q), 138.3 (q), 139.6 (q), 144 – 149 (Quaternary signals in this range could not be assigned as they were broad and highly amorphous).

ν_{max} (ATR)/cm^{-1}: 3264 (NH), 2978, 2929, 2049, 1719 (CO), 1596 (CN), 1560, 1509, 1479, 1368, 1320, 1234, 1146, 1106, 1057, 1029, 806, 771, 688, 665.

HRMS (*m/z* - ES): Found: 762.2999 (M^+ + H. $C_{36}H_{44}N_7O_6F_3Cl$ Requires: 762.2994).

4'-[2,3-Di(*tert*-butoxycarbonyl)guanidino]-3-[2-(*tert*-butoxycarbonyl)-3-(phenyl) guanidino]diphenylether (55a)



Following Method C, to a solution of **69a** (212 mg, 0.48 mmol), **57** (145 mg, 0.576 mmol) and triethylamine (207 µl, 1.49 mmol) in DCM (2.5 mL), $HgCl_2$ (156 mg, 0.576 mmol) was added. The reaction mixture was stirred at room temperature for 6 h. Usual work-up followed by silica gel chromatography (eluting with a gradient of hexane:EtOAc) to afford the desired product as off-white hydroscopic solid (108 mg, 34%). **Mp**: 78-80 °C.

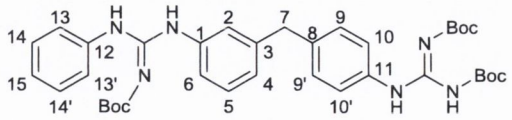
δ_H (600 MHz, $CDCl_3$): 1.53 (s, 18H, 2 $(CH_3)_3$), 1.57 (s, 9H, $(CH_3)_3$), 6.63 (bs, NH), 6.61-7.68 (m, 13H, CH Ar), 9.60 (bs, NH), 10.3 (bs, NH), 11.6 (bs, NH).

δ_C (150 MHz, $CDCl_3$): 28.1 ($(CH_3)_3$), 28.2 (2 $(CH_3)_3$), 79.6 (q, $\underline{C}(CH_3)_3$), 83.3 (q, $\underline{C}(CH_3)_3$), 83.7 (q, $\underline{C}(CH_3)_3$), 110.6 (CH Ar), 112.7 (CH Ar), 112.8 (CH Ar), 114.7 (CH Ar), 117.3 (CH Ar), 119.3 (CH Ar), 119.7 (2 CH Ar), 120.0 (CH Ar), 122.5 (CH Ar), 122.8 (CH Ar), 123.7 (2 CH Ar), 128.7 (CH Ar), 129.5 (CH Ar), 129.6 (2 CH Ar), 130.6 (CH Ar), 132.0 (q), 132.5 (q), 138.8 (q), 139.8 (q), 140.2 (q), 140.5 (q), 146.7 (q), 148.5 (q), 152.9 (q), 153.3 (q), 153.5 (q), 163.0 (q).

ν_{max} (ATR)/cm^{-1}: 3411, 3264, 3158, 2978, 2932, 1666, 1718, 1637, 1590, 1559, 1481, 1461, 1408, 1367, 1337, 1305, 1230, 1211, 1145, 1095, 1056, 963, 944, 835, 806, 771, 750, 690.

HRMS (*m/z* - ES): Found: 661.3364 (M^+ + H. $C_{35}H_{45}N_6O_7$ Requires: 661.3350).

4'-[2,3-Di(*tert*-butoxycarbonyl)guanidino]-3-[2-(*tert*-butoxycarbonyl)-3-(phenyl) guanidino]benzylbenzene (55c)



Following Method C, to a solution of **69b** (131 mg, 0.29 mmol), **57** (90 mg, 0.36 mmol) and triethylamine (153.7 μl, 1.104 mmol) in DCM (1.54 mL), $HgCl_2$ (97 mg, 0.36 mmol) was added. The reaction mixture was stirred at room temperature for 4 h. Usual work-up followed by silica gel chromatography (eluting with a gradient of hexane:EtOAc) to afford the desired product as off-white solid (164 mg, 71%). **Mp**: 99-100 °C.

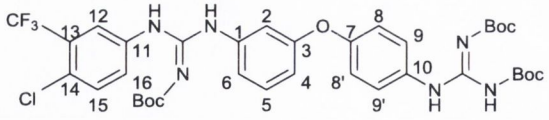
δ_H (600 MHz, $CDCl_3$): 1.52 (s, 18H, 2 $(CH_3)_3$), 1.56 (s, 9H, $(CH_3)_3$), 3.93 (s, 2H, CH_2), 6.73-7.32 (Aromatic signals in this range could not be assigned as they were broad and highly amorphous), 7.55 (d, 2H, J 7.4, CH Ar), 7.70 (bs, CH Ar), 9.54 (bs, NH), 10.3 (bs, NH), 11.6 (bs, NH).

δ_C (150 MHz, $CDCl_3$): 28.1 (2 $CH_3)_3$), 28.2 ((CH3)3), 41.3 (CH_2), 79.5 (q, $\underline{C}(CH_3)_3$), 83.6 (q, $\underline{C}(CH_3)_3$), 120.0 – 130 (Aromatic signals in this range could not be assigned as they were broad and highly amorphous) 122.3 (2 CH Ar), 129.2 (2 CH Ar), 135 – 150 (Quaternary signals in this range could not be assigned as they were highly amorphous and there was not enough resolution), 152.9 (q), 153.3 (q), 153.5 (q), 163.6 (q).

ν_{max} (ATR)/cm^{-1}: 3411(NH), 3264, 2978, 2932, 1718 (CO), 1636 (CN), 1590, 1559, 1481, 1461, 1408, 1337, 1305, 1230, 1211, 1145, 1095, 1056, 835, 806, 750, 690.

HRMS (*m/z* -ES): Found: 659.3544 (M^+ + H. $C_{36}H_{47}N_6O_6$ Requires: 659.3557).

4'-[2,3-Di-(*tert*-butoxycarbonyl)guanidino]-3-[2-(*tert*-butoxycarbonyl)-3-(4-chloro-3-trifluoromethylphenyl)guanidino]diphenylether (56a)



Following Method C, to a solution of **69a** (263 mg, 0.60 mmol), **58** (253 mg, 0.715 mmol) and triethylamine (257.2 μL, 1.85 mmol) in DCM (3.1 mL), $HgCl_2$ (194 mg, 0.715 mmol) was added. The reaction mixture was stirred at room temperature for 4 h. Usual work-up followed by silica gel chromatography (eluting with a gradient of hexane:EtOAc) to afford the desired product as off-white solid (91 mg, 70%). **Mp**: 85-87 °C.

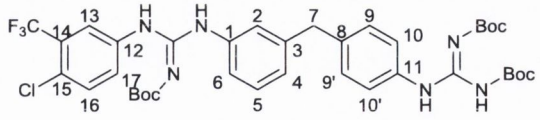
δ_H (600 MHz, $CDCl_3$): 1.52 (s, 9H, $(CH_3)_3$), 1.57 (s, 9H, $(CH_3)_3$), 6.57 (bs, CH Ar), 6.65 (bs, CH Ar), 6.69-6.80 (m, CH Ar and NH), 7.25 (bs, CH Ar), 7.43 (bs, CH Ar), 7.60 (bs, CH Ar), 7.89-8.03 (m, CH Ar), 9.60-9.90 (m, NH), 10.3 (bs, NH), 11.6 (bs, NH).

δ_C (150 MHz, $CDCl_3$): 27.9 (2 $(CH_3)_3$), 28.0 ($(CH_3)_3$), 79.5 (q, $\underline{C}(CH_3)_3$), 83.6 (q, $\underline{C}(CH_3)_3$), 111.9 (CH Ar), 113.2 (CH Ar), 116.7 (CH Ar), 118.7 (CH Ar), 119.7 (CH Ar), 123.6 (CH Ar), 130.7 (CH Ar), 131.5 (CH Ar), 132.4 (CH Ar), 139.7 (q), 140 – 151 (Quaternary signals in this range could not be assigned as they were highly amorphous and there was not enough resolution) 145.0 (q), 153.2 (q), 153.4 (q), 163.3 (q).

ν_{max} (ATR)/cm^{-1}: 3411 (NH), 3258, 2981, 2932, 2037, 1720 (CO), 1639 (CN), 1593, 1556, 1506, 1479, 1462, 1408, 1368, 1305, 1232, 1145, 1111, 1029, 957, 834, 805, 771, 686.

HRMS (*m/z* - ES): Found: 763.2825 (M^+ + H. $C_{36}H_{43}N_6O_7F_3Cl$ Requires: 763.2834).

4'-[2,3-Di(*tert*-butoxycarbonyl)guanidino]-3-[2-(*tert*-butoxycarbonyl)-3-(4-chloro-3-trifluoromethylphenyl)guanidino]benzylbenzene (56c)



Following Method C, to a solution of **69b** (107 mg, 0.24 mmol), **58** (104 mg, 0.29 mmol) and triethylamine (105.3 μl, 0.76 mmol) in DCM (1.3 mL) and isopropanol (0.3 mL), $HgCl_2$ (80 mg, 0.29 mmol) was added. The reaction mixture was stirred at room temperature for 5 h and 30 min. Usual work-up followed by silica gel chromatography (eluting with a gradient of hexane:EtOAc) to afford the desired product as off-white solid (110 mg, 68%). **Mp**: 98-100 °C.

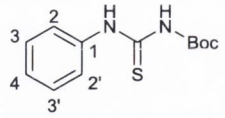
δ_H (600 MHz, $CDCl_3$): 1.52 (s, 18H, 2 $(CH_3)_3$), 1.56 (s, 9H, $(CH_3)_3$), 3.95 (s, 2H, CH_2), 6.77-6.80 (m, CH Ar and NH), 6.94 (bs, CH Ar), 7.09-7.33 (m, CH Ar), 7.38-7.44 (m, CH Ar), 7.51-7.58 (m, CH Ar, CF_3 ring), 7.95 (bs, CH Ar), 8.01 (bs, CH Ar), 9.40-9.85 (m, bs, NH), 10.3 (bs, NH), 11.6 (bs, NH).

δ_C (150 MHz, $CDCl_3$): 27.9 (2 $CH_3)_3$), 28.0 ($(CH_3)_3$), 41.1 (CH_2), 41.2 (CH_2), 79.4 (q, $\underline{C}(CH_3)_3$), 83.5 (q, $\underline{C}(CH_3)_3$), 118.5 (CH Ar), 119.6 (CH Ar), 122.1 (CH Ar), 122.2 (CH Ar), 122.7 (c, q, J 274, CF_3), 123.6 (CH Ar), 123.8 (CH Ar), 126.6 (CH Ar), 129.6 (CH Ar), 128.1 (c, q, J 30, C-14), 131.5 (CH Ar), 132.2 (CH Ar), 134.8 (q), 137.1 (q), 137.8 (q), 139.5 (q), 142.9 (q), 146.1 (q), 152.9 (q), 153.2 (q), 153.3 (q), 163.4 (q).

ν_{max} (ATR)/cm^{-1}: 3408 (NH), 3261, 3156, 2981, 2933, 2162, 1718 (CO), 1632 (CN), 1602, 1559, 1480, 1463, 1409, 1368, 1319, 1303, 1235, 1146, 1116, 1057, 957, 804, 770, 687.

HRMS (*m/z* - ES): Found: 761.3030 (M^+ + H. $C_{37}H_{45}N_6O_6F_3Cl$ Requires: 761.3041).

N-tert-Butoxycarbonyl-_N'_-phenylthiourea (57)

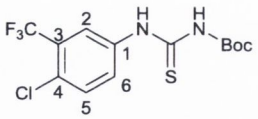


To a solution of *N*-phenylthiourea (500 mg, 3.29 mmol) in dry tetrahydrofuran (25 mL) under argon at 0 °C, sodium hydride as a 60% suspension in mineral oil (290 mg, 7.24 mmol) were added. The reaction was stirred at 0 °C for 1 h and then 3 h at room temperature to complete formation of the anion. The reaction is cooled again to 0 °C prior to addition of di-*tert*-butyl dicarbonate (862 mg, 3.95 mmol). The reaction was stirred for 1 h at 0 °C and then it was brought to room temperature and stirred overnight. The reaction was cooled again to 0 °C and dropwise H_2O (20 mL) was added to quench the reaction, followed by extraction with EtOAc (3 × 20 mL). The combined organic phases were washed with 80% brine (30 mL) and dried using anhydrous $MgSO_4$, followed by removal of solvents under vacuum. Purification proceeded by silica gel chromatography using gradient elution of hexane:EtOAc, followed recrystallisation from boiling hexane to afford the product **7** as white needles (483 mg, 58%). **Mp**: 104 °C. (lit. 106 °C).[9]

δ_H (600 MHz, $CDCl_3$): 1.56 (s, 9H, $(CH_3)_3$), 7.28 (t, 1H, J 7.9, H-4), 7.42 (t, 2H, J 7.9, H-3, H-3'), 7.66 (d, 2H, J 7.9, H-2, H-2'), 7.96 (bs, 1H, NHPh), 11.52 (bs, 1H, NHBoc).

HRMS (*m/z* - ES): Found: 275.0830 (M^+ + Na. $C_{12}H_{16}N_2O_2SNa$ Requires: 275.0823).

***N*-(*tert*-Butoxycarbonyl)-*N'*-(4-chloro-3-trifluoromethylphenyl)thiourea (58)**



To a solution of **43** (2 g, 7.25 mmol) in dry tetrahydrofuran (58.5 mL) under argon at 0 °C, 1.5 equivalents of sodium hydride as a 60% suspension in mineral oil (435 mg, 10.9 mmol) were added. The reaction mixture was stirred at the same temperature for 1 h and 20 min, then 1.1 equivalents of trifluoroacetic anhydride (1.11 mL, 8.0 mmol), and the stirring continue for an additional 30 min. Then, 1.15 equivalents of the 4-chloro-3-(trifluoromethyl)aniline (1.63 g, 8.34 mmol) were added and the reaction was stirred at 0 °C for 5 h. The reaction was cooled again to 0 °C and dropwise H_2O (20 mL) was added to

quench the reaction, followed by extraction with EtOAc (3 × 30 mL). The combined organic phases were washed with 80% brine (60 mL) and dried using anhydrous $MgSO_4$, followed by removal of solvents under vacuum. Purification proceeded by silica gel chromatography using gradient elution (hexane:EtOAc), followed by removal of solvents under vacuum. Recrystallisation from boiling hexane afforded the product as white needles (1680 mg, 92%). **Mp**: 140-142 °C.

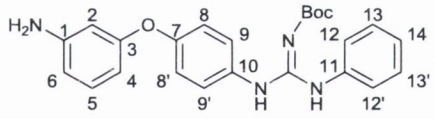
δ_H (600 MHz, $CDCl_3$): 1.56 (s, 9H, $(CH_3)_3$), 7.53 (d, 1H, J 8.5, H-5), 7.94 (d, 1H, J 8.5, H-6), 7.99 (bs, 1H, NHPh), 8.02 (s, 1H, H-2), 11.7 (bs, 1H, NHBoc).

δ_C (150 MHz, $CDCl_3$): 28.0 ($(CH_3)_3$), 84.9 (q, $\underline{C}CH_3$), 123.2 (c, J 5.2, CH Ar, C-2), 123.3 (c, q, J 274, CF_3), 128.2 (CH Ar, C-6), 128.8 (c, q, J 30, C-3), 129.5 (q, C-4), 131.7 (CH Ar, C-5), 136.5 (q, C-1), 152.0 (q, CO), 178.0 (q, CS).

ν_{max} (ATR)/cm^{-1}: 3172, 2979, 2900, 1727, 1710, 1589, 1527 (CO), 1478, 1425, 1324, 1258, 1239, 1128 (CS), 1113, 1033, 1015, 972, 838, 820, 775, 691, 663.

HRMS (*m/z* - ES): Found: 353.0332 (M^- - H. $C_{13}H_{13}N_2O_2ClF_3S$ Requires: 353.0338).

3-Amino-4'-[2-(*tert*-butoxycarbonyl)-3-(phenyl)guanidino]diphenylether (66a).



Following Method B, $HgCl_2$ (407 mg, 1.5 mmol) was added over a solution of **9a** (751 mg, 3.75 mmol), **57** (315 mg, 1.25 mmol) and triethylamine (539 µL, 3.87 mmol) in DCM (6.5 mL) at 0 °C. The resulting mixture was stirred at 0 °C for 1 h and further 3 h at room temperature. Usual work-up, followed by purification using silica gel chromatography eluting with a gradient of 100% hexane to 80:20 hexane:EtOAc respectively to afford the title product as a light brown hygroscopic solid (395 mg, 76%). **Mp**: 70-72 °C.

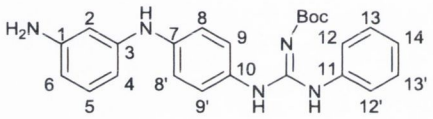
δ_H (600 MHz, $CDCl_3$): 1.51 (s, 9H, $(CH_3)_3$), 3.70 (bs, 2H, NH_2), 6.32 (bs, CH Ar), 6.36-6.50 (m, CH Ar), 6.72 (bs, NH), 6.84-7.20 (m, CH Ar), 7.30-7.41 (m, CH Ar), 7.70 (bs, CH Ar), 9.56 (bs, NH).

δ_C (150 MHz, $CDCl_3$): 28.1 ($(CH_3)_3$), 83.2 (q, $\underline{C}(CH_3)_3$), 104.8 (CH Ar), 108.2 (CH Ar), 109.7 (CH Ar), 120.0 (CH Ar), 120.9 (CH Ar), 121.4 (CH Ar), 122.6 (CH Ar), 123.6 (CH Ar), 128.8 (CH Ar), 129.6 (CH Ar), 130.3 (CH Ar), 131 – 150 (Quaternary signals in this range could not be assigned as they were highly amorphous and there was not enough resolution), 140.3 (q), 147.9 (q), 152.9 (q), 159.2 (q).

ν_{max} (ATR)/cm^{-1}: 3339, 2976, 2928, 1719 (CO), 1660 (CN), 1590, 1551, 1487, 1460, 1388, 1366, 1327, 1284, 1207, 1143, 1089, 960, 837, 804, 750, 687.

HRMS (*m/z* - ES): Found: 419.2071 (M^+ + H. $C_{24}H_{27}N_4O_3$ Requires: 419.2083).

3-Amino-4'-[2-(*tert*-butoxycarbonyl)-3-(phenyl)guanidino)]-*N*-phenylaniline (66d)



Following Method B, $HgCl_2$ (308 mg, 1.13 mmol) was added over a solution of **9d** (564 mg, 2.83 mmol), **57** (238 mg, 0.944 mmol) and triethylamine (407 μL, 2.93 mmol) in DCM (5 mL) at 0 °C. The resulting mixture was stirred at 0 °C for 1 h and further 3 h at room temperature. Usual work-up, followed by purification using silica gel chromatography eluting with a gradient of hexane:EtOAc to afford the title product as a light dark solid (328 mg, 83%). **Mp**: 69-71 °C.

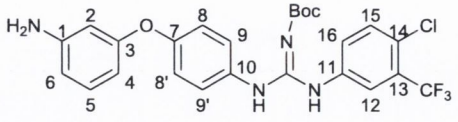
δ_H (600 MHz, $CDCl_3$): 1.53 (s, 9H, $(CH_3)_3$), 3.74 (bs, 2H, NH_2), 5.55-5.71 (bs, 1H, NH), 6.26 (d, 1H, J 7.9, H-4 or H-6), 6.36 (app. t, 1H, J 1.9, H-2), 6.43 (d, 1H, J 7.9, H-4 or H-6), 6.59-6.67 (m, 1H, CH Ar), 6.98-7.36 (m, 9H, CH Ar and 1NH), 9.52 (bs, 1H, NH).

δ_C (150 MHz, $CDCl_3$): 28.2 ($(CH_3)_3$), 82.9 (q, $\underline{C}(CH_3)_3$), 103.6 (CH Ar), 108.0 (CH Ar), 116.1 (CH Ar), 118.3 – 127.2 (Aromatic signals in this range could not be assigned as they were broad and highly amorphous), 129.3 (CH Ar), 130.2 (CH Ar), 133.3 – 163.7 (Quaternary signals in this range could not be assigned as they were broad and highly amorphous), 147.7 (q, CO).

ν_{max} (ATR)/cm^{-1}: 3341 (NH), 3320 (NH), 2977, 2930, 1981, 1719 (CO), 1589 (CN), 1509, 1494, 1366, 1336, 1289, 1237, 1151, 1090, 1044, 966, 901, 828, 803, 750, 690.

HRMS (*m/z* - ES): Found: 418.2255 (M^+ + H. $C_{24}H_{28}N_5O_2$ Requires: 418.2243).

3-Amino-4'-[2-(*tert*-butoxycarbonyl)-3-(4-chloro-3-trifluoromethyl)guanidino] diphenylether (67a)



Following Method B, $HgCl_2$ (92 mg, 0.34 mmol) was added over a solution of **9a** (170 mg, 0.85 mmol), **58** (100 mg, 0.282 mmol) and triethylamine (121.7 μL, 0.874 mmol) in DCM (1.5 mL) at 0 °C. The resulting mixture was stirred at 0 °C for 1 h and further 2 h at room temperature. Usual work-up, followed by purification using silica gel chromatography eluting with a gradient of 100% hexane to 80:20 hexane:EtOAc respectively to afford the title product as a light dark solid (100 mg, 68%). **Mp**: 48-49 °C.

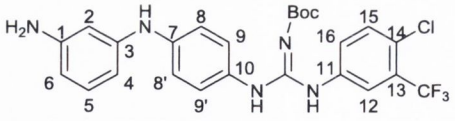
δ_H (600 MHz, $CDCl_3$): 1.52 (s, 9H, $(CH_3)_3$), 3.71 (bs, 2H, NH_2), 6.29-6.48 (m, 3H, CH Ar), 6.69-7.70 (m, 7H, CH Ar), 6.76 (bs, 1H, NH), 7.90-8.11 (m, 1H, CH Ar), 9.47-9.94 (m, 1H, NH).

δ_C (150 MHz, $CDCl_3$): 28.1 ($(CH_3)_3$), 83.8 (q, $\underline{C}(CH_3)_3$), 105.0 (CH Ar), 108.4 (CH Ar), 109.9 (CH Ar), 118.6 (CH Ar), 119.9 (CH Ar), 120.9 (CH Ar), 121.9 (CH Ar), 123.3 (CH Ar), 123.5 (CH Ar), 126.7 (CH Ar), 130.5 (2 CH Ar), 131.6 (CH Ar), (132.4 – 145.6 (Quaternary signals in this range could not be assigned as they were broad and highly amorphous), 147.9 (q, CO), 153.1 (q, CN).

ν_{max} (ATR)/cm^{-1}: 3357 (NH), 3202 (NH), 2980, 2934, 1718 (CO), 1657 (CN), 1594, 1480, 1461, 1410, 1317, 1207, 1138, 1030, 991, 960, 834, 771, 685.

HRMS (*m/z* - ES): Found: 521.1560 (M^+ + H. $C_{25}H_{25}N_4O_3F_3Cl$ Requires: 521.1567).

3-Amino-4'-[2-(*tert*-butoxycarbonyl)-3-(4-chloro-3-trifluoromethylphenylguanidino]-*N*-phenylaniline (67d)



Following Method B, $HgCl_2$ (220 mg, 0.81 mmol) was added over a solution of **9d** (404 mg, 2.03 mmol), **58** (239 mg, 0.676 mmol) and triethylamine (291.7 μL, 2.10 mmol) in DCM (3.5 mL) at 0 °C. The resulting mixture was stirred at 0 °C for 1 h and further 3 h at room temperature. Usual work-up, followed by purification using silica gel chromatography eluting with a gradient of hexane:EtOAc to afford the title product as a dark solid (263 mg, 86%). **Mp**: 78-80 °C.

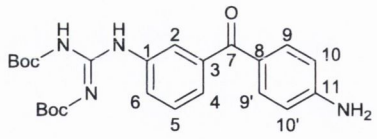
δ_H (600 MHz, $CDCl_3$): 1.52 (s, 9H, $(CH_3)_3$), 3.72 (bs, 2H, NH_2), 5.44-5.73 (bs, 1H, NH), 6.27 (d, 1H, J 7.2, H-4 or H-6), 6.39 (s, 1H, H-2), 6.43 (d, 1H, J 7.2, H-4 or H-6), 6.55-6.73 (m, 1H, CH Ar), 6.76-7.47 (m, 9H, CH Ar and 1NH), 7.89-8.13 (m, 1H, CH Ar), 9.40-10.00 (m, 1H, NH).

δ_C (150 MHz, $CDCl_3$): 28.0 ($(CH_3)_3$), 83.5 (q, $\underline{C}(CH_3)_3$), 103.2 (CH Ar, C-2), 107.6 (CH Ar, C-4 or C-6), 108.1 (CH Ar, C-4 or C-6), 116.0 (CH Ar), 120.7 (CH Ar), 121.8 (q, c, J 274, CF_3) 123.0 (CH Ar), 123.6 (CH Ar), 124.6 (q, d, J 30, C-13), 126.6 (CH Ar), 130.2 (CH Ar, C-5), 131.5 (CH Ar), 132.2 (CH Ar), 134.4 – 160.0 (Quaternary signals in this range could not be assigned as they were broad and highly amorphous), 147.5 (q, CO).

ν_{max} (ATR)/cm^{-1}: 3340 (NH), 3333 (NH), 2978, 2933, 2162, 1720 (CO), 1591 (CN), 1556, 1509, 1479, 1368, 1318, 1240, 1135, 1111, 1080, 1031, 991, 831, 771, 688, 666.

HRMS (*m/z* - ES): Found: 520.1727 (M^+ + H. $C_{25}H_{26}N_5O_2F_3Cl$ Requires: 520.1727).

4-Amino-[3'-(2,3-di-*tert*-butoxycarbonyl)guanidino]benzophenone (68b)



Following Method B, $HgCl_2$ (231 mg, 0.85 mmol) was added over a solution of **9b** (451 mg, 2.13 mmol), **43** (184 mg, 0.71 mmol) and triethylamine (306.5 μL, 2.2 mmol) in DCM (4 mL) at 0 °C. The resulting mixture was stirred at 0 °C for 1 h and further 1 h and 30 min at room temperature.Usual work-up, followed by purification using silica gel chromatography eluting with a gradient of hexane:EtOAc to afford the title product as a brown-orange solid (263 mg, 86%). **Mp**: 92-94 °C.

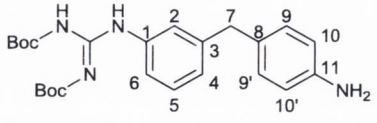
δ_H (600 MHz, $CDCl_3$): 1.53 (s, 9H, $(CH_3)_3$), 1.56 (s, 9H, $(CH_3)_3$), 4.15 (bs, 2H, NH_2), 6.70 (d, 2H, J 8.6, H-10, H-10'), 7.45 (t, 1H, J 7.8, H-5), 7.50 (d, 1H, J 7.8, H-4), 7.76 (d, 2H, J 8.6, H-9, H-9'), 7.81 (s, 1H, H-2), 7.97 (d, 1H, J 7.8, H-6), 10.4 (bs, 1H, NHBoc), 11.6 (bs, 1H, NH).

δ_C (150 MHz, $CDCl_3$): 28.1 ($(CH_3)_3$), 28.2 ($(CH_3)_3$), 79.7 (q, $\underline{C}(CH_3)_3$), 84.0 (q, $\underline{C}(CH_3)_3$), 113.7 (2 CH Ar, C-10, C-10'), 123.3 (CH Ar, C-2), 125.6 (CH Ar, C-6), 130.0 (CH Ar, C-4), 127.3 (q), 128.7 (CH Ar, C-5), 133.0 (2 CH Ar, C-9, C-9'), 136.7 (q, C-1), 139.4 (q, C-3), 150.9 (q, C-11), 153.3 (q), 153.6 (q), 153.7 (q, C-8), 163.5 (q), 194.5 (q, CO linker).

ν_{max} (ATR)/cm^{-1}: 3357 (NH), 3235, 2983, 2928, 1718 (CO), 1627 (CN), 1587, 1407, 1368, 1291, 1235, 1149, 1105, 1057, 841, 803, 759, 690.

HRMS (*m/z* - ES): Found: 477.2127 (M^+ + Na. $C_{24}H_{30}N_4O_5Na$ Requires: 477.2114).

4-Amino-[3'-(2,3-di-*tert*-butoxycarbonyl)guanidino]benzylbenzene (68c)



Following Method B, $HgCl_2$ (154 mg, 0.6 mmol) was added over a solution of **9c** (282 mg, 1.42 mmol), **43** (138 mg, 0.5 mmol) and triethylamine (204.3 μL, 1.5 mmol) in DCM (2.5 mL) at 0 °C. The resulting mixture was stirred at 0 °C for 1 h and then

overnight at room temperature.Usual work-up, followed by purification of the *meta* and *para* mono-guanidylated products by silica gel chromatography eluting with a gradient of 100% hexane to 80:20 hexane:EtOAc respectively to afford the title product as a brown solid (103 mg, 47%). **Mp**: decomp. above 145 °C.

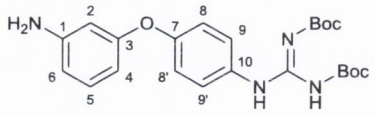
δ_H (600 MHz, $CDCl_3$): 1.53 (s, 9H, $(CH_3)_3$), 1.55 (s, 9H, $(CH_3)_3$), 3.88 (s, 2H, CH_2), 6.65 (d, 2H, J 8.0, H-10, H-10'), 6.94 (d, 1H, J 7.8, H-4), 7.00 (d, 2H, J 8.0, H-9, H-9'), 7.25 (t, 1H, J 7.8, H-5), 7.30 (s, 1H, H-2), 7.58 (d, 1H, J 7.8, H-6), 10.2 (bs, 1H, NHBoc), 11.6 (bs, 1H, NH).

δ_C (150 MHz, $CDCl_3$): 28.0 ($(CH_3)_3$), 28.1 ($(CH_3)_3$), 41.0 (CH_2), 79.5 (q, $\underline{C}(CH_3)_3$), 83.6 (q, $\underline{C}(CH_3)_3$), 115.6 (2 CH Ar, C-10, C-10'), 120.2 (CH Ar, C-6), 122.4 (CH Ar, C-2), 125.4 (CH Ar, C-4), 128.9 (CH Ar, C-6), 129.7 (2 CH Ar, C-9, C-9'), 130.7 (q, C-8), 136.3 (q, C-1), 142.1 (q, C-3), 143.5 (q, C-11), 153.2 (q, CO), 153.4 (q, CO), 163.5 (q, CN).

ν_{max} (ATR)/cm^{-1}: 3370 (NH), 3260, 3026, 2921, 2851, 2163, 1722 (CO), 1640 (CN), 1602, 1493, 1451, 1366, 1240, 1155, 1111, 1064, 1029, 905, 803, 753, 645.

HRMS (*m/z* - ES): Found: 441.2502 (M^+ + H. $C_{24}H_{33}N_4O_4$ Requires: 441.2502).

3-Amino-[4'-(2,3-di-*tert*-butoxycarbonyl)guanidino]diphenylether (69a)



Following Method B, $HgCl_2$ (561 mg, 2.1 mmol) was added over a solution of **9a** (1500 mg, 5.16 mmol), **43** (500 mg, 1.72 mmol) and triethylamine (744.3 µL, 5.34 mmol) in DCM (8.2 mL) at 0 °C. The resulting mixture was stirred at 0 °C for 1 h and then overnight at room temperature.Usual work-up, followed by purification of the *meta* and *para* mono-guanidylated products by silica gel chromatography eluting with a gradient of 100% hexane to 80:20 hexane:EtOAc respectively to afford the title product as a white solid (586 mg, 77%). **Mp**: 78-79 °C.

δ_H (600 MHz, DMSO): 1.40 (s, 9H, $(CH_3)_3$), 1.52 (s, 9H, $(CH_3)_3$), 5.21 (bs, 2H, NH_2), 6.14 (dd, 1H, J 8.0, 2.1, H-4), 6.18 (app. t, 1H, J 2.1, H-2), 6.33 (dd, 1H, J 8.0, 2.1, H-6),

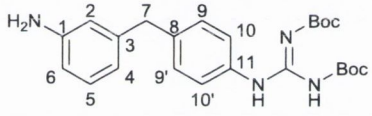
6.99 (d, 2H, J 8.8, H-8 and H-8'), 7.00 (t, 1H, J 8.0, H-5), 7.51 (d, 2H, J 8.8, H-9 and H-9'), 9.96 (bs, 1H, NHBoc), 11.5 (bs, 1H, NH).

δ_C (150 MHz, DMSO): 28.0 ($(CH_3)_3$), 28.3 ($(CH_3)_3$), 79.3 (q, $\underline{C}(CH_3)_3$), 83.7 (q, $\underline{C}(CH_3)_3$), 104.0 (CH Ar, C-6), 106.0 (CH Ar, C-4), 109.6 (CH Ar, C-2), 119.1 (2 CH Ar, C-8 and C-8'), 125.1 (2 CH Ar, C-9 and C-9'), 130.4 (CH Ar, C-5), 132.0 (q, Ar, C-10), 148.0 (q, Ar, C-1), 152.0 (q, CO), 153.0 (q, Ar, C-7), 154.0 (q, CN), 158.0 (q, Ar, C-3), 163.0 (q, CO).

ν_{max} (ATR)/cm^{-1}: 3474 (NH), 3381 (NH), 3264 (NH), 2976, 2930, 1724 (CO), 1617 (CN), 1492, 1414, 1367, 1289, 1305, 1217, 1145, 1112, 958, 802, 753.

HRMS (*m/z* - ES): Found: 443.2290 (M^+ + H. $C_{23}H_{31}N_4O_5$ Requires: 443.2294).

3-Amino-[4'-(2,3-di-*tert*-butoxycarbonyl)guanidino]benzylbenzene (69b)



Following Method B, $HgCl_2$ (154 mg, 0.6 mmol) was added over a solution of **9c** (282 mg, 1.42 mmol), **43** (138 mg, 0.5 mmol) and triethylamine (204.3 μL, 1.5 mmol) in DCM (2.5 mL) at 0 °C. The resulting mixture was stirred at 0 °C for 1 h and then overnight at room temperature.Usual work-up, followed by purification of the *meta* and *para* mono-guanidylated products by silica gel chromatography eluting with a gradient of 100% hexane to 80:20 hexane:EtOAc respectively to afford the title product as a brownish solid (107 mg, 52%). **Mp**: 118-119 °C.

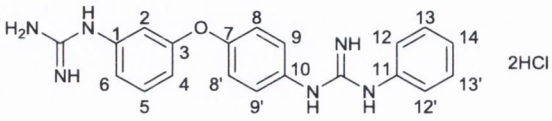
δ_H (600 MHz, $CDCl_3$): 1.51 (s, 9H, $(CH_3)_3$), 1.56 (s, 9H, $(CH_3)_3$), 1.92 (bs, 2H, NH_2), 3.89 (s, 2H, CH_2), 6.59 (s, 1H, H-2), 6.64 (d, 1H, J 7.7, H-6), 6.68 (d, 1H, J 7.7, H-4), 7.11 (t, 1H, J 7.7, H-5), 7.16 (d, 2H, J 8.1, H-9, H-9'), 7.51 (d, 2H, J 8.1, H-10, H-10'), 10.3 (bs, 1H, NHBoc), 11.6 (bs, 1H, NH).

δ_C (150 MHz, $CDCl_3$): 27.9 ($(CH_3)_3$), 28.0 ($(CH_3)_3$), 41.1 (CH_2), 79.5 (q, $\underline{C}(CH_3)_3$), 83.5 (q, $\underline{C}(CH_3)_3$), 113.2 (CH Ar, C-6), 116.0 (CH Ar, C-2), 119.7 (CH Ar, C-4), 122.2 (2 CH Ar, C-10, C-10'), 129.2 (CH Ar, C-5), 129.3 (2 CH Ar, C-9, C-9'), 134.6 (q, C-11), 137.5 (q, C-8), 142.2 (q, C-3), 145.5 (q, C-1), 153.2 (q, CO), 153.4 (q, CN), 165.1 (q, CO).

ν_{max} (ATR)/cm^{-1}: 3377(NH), 3261(NH), 2984, 2965, 2924, 1722 (CO), 1643 (CN), 1603, 1559, 1477, 1405, 1367, 1294, 1242, 1226, 1153, 1112, 1058, 1029, 855, 803, 768, 695.

HRMS (*m/z* - ES): Found: 441.2488 (M^+ + H. $C_{24}H_{33}N_4O_4$ Requires: 441.2502).

3-[4'-(Phenylguanidino)phenoxy]phenylguanidine dihydrochloride (70a)



Following Method F, **53a** (66 mg, 0.10 mmol) was dissolved in 4M HCl in dioxane (0.452 mL, 1.81 mmol) and in additional dioxane (0.05 mL) until a final concentration of 0.2M was reached. After 8 h stirring at 55 °C, the reaction was adjudged complete (TLC), solvents were evaporated and the residue was purified by reverse phase chromatography to afford the pure hydrochloride salt as an white solid (38 mg, 88%). **Mp**: decomp. above 140 °C.

δ_H (600 MHz, D_2O): 6.96 (app. t, 1H, J 2.1, H-2), 7.03 (dd, 1H, J 8.0, 2.1, H-4 or H-6), 7.08 (dd, 1H, J 8.0, 2.1, H-4 or H-6), 7.13 (d, 2H, J 8.9, H-8, H-8' or H-9, H-9'), 7.31-7.36 (m, 4H, H-8, H-8' or H-9, H-9' and H-12, H-12'), 7.38 (d, 1H, J 7.5, H-14), 7.43-7.47 (m, 3H, H-5 and H-13, H-13').

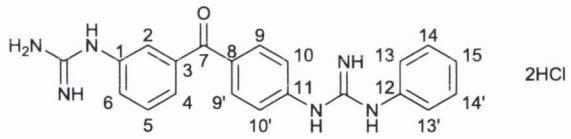
δ_C (150 MHz, D_2O): 116.0 (CH Ar, C-2), 118.0 (CH Ar, C-4 or C-6), 120.3 (2 CH Ar, C-8, C-8' or C-9, C-9'), 121.0 (CH Ar, C-4 or C-6), 125.9 (2 CH Ar, C-12, C-12'), 127.9 (CH Ar, C-14), 128.0 (2 CH Ar, C-8, C-8' or C-9, C-9'), 129.7 (q, C-7 or C-10), 129.8 (2 CH Ar, C-13, H-13'), 131.2 (CH Ar, C-5), 134.0 (q, C-11), 135.6 (q, C-1 or C-3), 155.1 (q, CN), 155.6 (q, C-7 or C-10), 156.1 (q, CN) 157.5 (q, C-1 or C-3).

ν_{max} (ATR)/cm^{-1}: 3268 (NH), 3105 (NH), 2953, 2050, 1655, 1619, 1578 (CN), 1486, 1415, 1211, 1154, 1104, 1016, 963, 840, 755, 691.

HRMS (*m/z* - ES): Found: 361.1775 (M^+ + H. $C_{20}H_{21}N_6O$ Requires: 361.1777).

HPLC: 99.1% (*t*R: 22.8 min).

3-[4'-(Phenylguanidino)benzoyl]phenylguanidine dihydrochloride (70b)



Following Method F, **53b** (228 mg, 0.34 mmol) was dissolved in 4M HCl in dioxane (1.53 mL, 6.12 mmol) and in additional dioxane (0.17 mL) until reach a final concentration of 0.2M. After 8 h stirring at 55 °C, the reaction was adjudged complete (TLC), solvents were evaporated and the residue was purified by reverse phase chromatography to afford the pure hydrochloride salt as a hygroscopic yellowish solid (92 mg, 61%). **Mp**: decomp. above 161 °C.

δ_H (600 MHz, D_2O): 7.33 (d, 2H, J 7.5, H-13, H-13'), 7.36 (d, 1H, J 7.5, H-15), 7.43-7.48 (m, 4H, H-10, H-10', H-14, H-14'), 7.59 (dd, 1H, J 7.9, 1.8, H-6), 7.62 (t, 1H, J 7.9, H-5), 7.65 (app. t, 1H, J 1.8, H-2), 7.73 (d, 1H, J 7.5, 1.8, H-4), 7.83 (d, 2H, J 8.6, H-9, H-9').

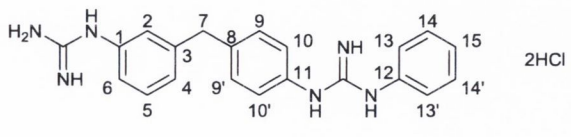
δ_C (150 MHz, D_2O): 123.7 (2 CH Ar, C-10, C-10'), 125.3 (2 CH Ar, C-13, C-13'), 127.1 (CH Ar, C-2), 127.8 (CH Ar, C-15), 129.4 (CH Ar, C-4), 129.8 (2 CH Ar, C-14, C-14'), 130.1 (CH Ar, C-5), 130.5 (CH Ar, C-6), 131.9 (2 CH Ar, C-9, C-9'), 134.2 (q), 134.3 (q), 134.5 (q), 138.2 (q), 140.1 (q), 154.3 (q, CN), 156.3 (q, CN), 198.0 (q, CO).

ν_{max} (ATR)/cm^{-1}: 3143 (NH), 2970, 2287, 1738, 1560 (CN), 1490, 1448, 1228, 1076, 1020, 753, 694.

HRMS (*m/z* - ES): Found: 373.1769 (M^+ + H. $C_{21}H_{21}N_6O$ Requires: 373.1777).

HPLC: 99.8% (*t*R: 21.7 min).

3-[4'-(Phenylguanidino)benzyl]phenylguanidine dihydrochloride (70c)



Following Method F, **53c** (50 mg, 0.076 mmol) was dissolved in 4M HCl in dioxane (0.345 mL, 1.38 mmol) and in additional dioxane (0.04 mL) until reach a final concentration of 0.2M. After 6 h stirring at 55 °C, the reaction was adjudged complete

(TLC), solvents were evaporated and the residue was purified by reverse phase chromatography to afford the pure hydrochloride salt as a hygroscopic off-white crystalline solid (31 mg, 94%). **Mp**: decomp. above 132 °C.

δ_H (600 MHz, D_2O): 4.00 (s, 2H, CH_2), 7.14 (d, 1H, J 8.0, H-6), 7.15 (s, 1H, H-2), 7.25-7.26 (m, 3H, H-10, H-10', H-4), 7.30-7.34 (m, 4H, H-13, H-13', H-9, H-9'), 7.36-7.39 (m, 2H, H-5, H-15), 7.44 (t, 2H, J 8.0, H-14, H-14') .

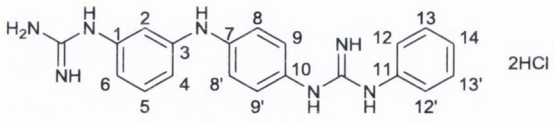
δ_C (150 MHz, D_2O): 40.2 (CH_2), 123.6 (CH Ar, C-6), 125.6 (2 CH Ar, C-13, C-13'), 125.9 (CH Ar, C-2), 126.2 (2 CH Ar, C-10, C-10'), 127.9 (CH Ar, C-15), 128.3 (CH Ar, C-4), 129.8 (2 CH Ar, C-14, C-14'), 130.1 (CH Ar, C-5), 130.2 (2 CH Ar, C-9, C-9'), 132.2 (q, C-11), 134.0 (q, C-12), 134.3 (q, C-1), 141.0 (q, C-8), 143.2 (q, C-3), 154.9 (q, CN), 156.3 (q, CN).

ν_{max} (ATR)/cm^{-1}: 3143 (NH), 2970, 2287, 1738, 1560 (CN), 1490, 1448, 1228, 1076, 1020, 753, 694.

HRMS (*m/z* - ES): Found: 359.1981 (M^+ + H. $C_{21}H_{23}N_6$ Requires: 359.1984).

HPLC: 98.2% (*t*R: 23.5min).

3-[4'-(Phenylguanidino)phenylamino]phenylguanidine dihydrochloride (70d)



Following Method F, **53d** (168 mg, 0.255 mmol) was dissolved in 4M HCl in dioxane (1.15 mL, 4.59 mmol) and in additional dioxane (0.125 mL) until reach a final concentration of 0.2M. After 8 h stirring at 55 °C, the reaction was adjudged complete (TLC), solvents were evaporated and the residue was purified by reverse phase chromatography to afford the pure hydrochloride salt as a hygroscopic light dark solid (64 mg, 58%). **Mp**: decomp. above 160 °C.

δ_H (600 MHz, D_2O): 6.83 (d, 1H, J 7.6, H-4 or H-6), 6.96 (s, 1H, H-2), 7.07 (d, 1H, J 7.6, H-4 or H-6), 7.15 (d, 2H, J 6.9, H-8, H-8' or H-9, H-9'), 7.22 (d, 2H, J 6.9, H-8, H-8' or H-9, H-9'), 7.31-7.38 (m, 4H, H-5, H-12, H-12', H-14), 7.45 (app. t, 2H, J 7.1, H-13, H-13').

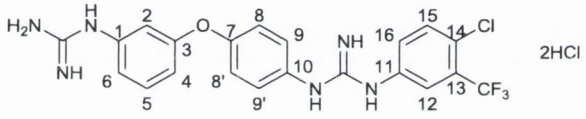
δ_C (150 MHz, D_2O): 114.3 (CH Ar, C-2), 117.0 (CH Ar, C-4 or C-6), 118.0 (CH Ar, C-4 or C-6), 188.8 (2 CH Ar, C-8, C-8' or C-9, C-9'), 125.9 (2 CH Ar, C-12, C-12'), 126.6 (q, C-7 or C-10), 127.6 (2 CH Ar, C-8, C-8' or C-9, C-9'), 127.9 (CH Ar, C-14), 129.8 (2 CH Ar, C-13, C-13'), 130.8 (CH Ar, C-5), 133.9 (q, C-11), 135.1 (q, C-1 or C-3), 142.6 (q, C-7 or C-10), 144.4 (q, C-1 or C-3), 155.2 (q, CN), 156.1 (q, CN).

ν_{max} (ATR)/cm^{-1}: 3259 (NH), 3129 (NH), 2162, 1619, 1578 (CN), 1509, 1492, 1331, 1226, 1195, 1167, 1112, 1077, 1022, 997, 819, 756, 691.

HRMS (*m/z* - ES): Found: 360.1932 (M^+ + H. $C_{20}H_{22}N_7$ Requires: 360.1937).

HPLC: 99.5% (*t*R: 22.51 min).

3-[4'-(4-Chloro-3-trifluoromethylphenyl)guanidinophenoxy]phenylguanidine dihydrochloride (71a)



Following Method F, **54a** (117 mg, 0.154 mmol) was dissolved in 4M HCl in dioxane (0.69 mL, 2.77 mmol) and in additional dioxane (0.076 mL) until a final concentration of 0.2M was reached. After 6 h stirring at 55 °C, the reaction was adjudged complete (TLC), solvents were evaporated and the residue was purified by reverse phase chromatography to afford the pure hydrochloride salt as a hygroscopic yellowish solid (72 mg, 88%). **Mp**: decomp. above 160 °C.

δ_H (600 MHz, D_2O): 6.93 (app. t, 1H, J 1.7, H-2), 6.98 (dd, 1H, J 8.5, 1.7, H-4 or H-6), 7.07 (d, 2H, J 8.9, H-8, H-8' or H-9, H-9'), 7.08 (dd, 1H, J 8.5, 1.7, H-4 or H-6), 7.27 (d, 2H, J 8.9, H-8, H-8' or H-9, H-9'), 7.44-7.47 (m, 2H, H-5 and H-16), 7.61 (d, 1H, J 8.5, H-15), 7.67 (d, 1H, J 1.9, H-12).

δ_C (150 MHz, D_2O): 116.0 (CH Ar, C-2), 118.0 (CH Ar, C-4 or C-6), 120.0 (2 CH Ar, C-8, C-8' or C-9, C-9'), 121.1 (CH Ar, C-4 or C-6), 123 (c, q, J 273, CF_3), 124.6 (c, J 5.2, CH Ar, C-12), 127.4 (2 CH Ar, C-8, C-8' or C-9, C-9'), 128.4 (c, q, J 30, C-13), 129.6 (q, C-7 or C-10), 129.9 (CH Ar, C-16), 130.0 (q, C-14), 131.2 (CH Ar, C-5), 132.7 (CH Ar, C-15),

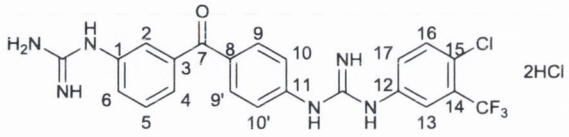
133.5 (q, C-11), 135.5 (q, C-1 or C-3), 154.7 (q, CN), 155.5 (q, C-7 or C-10), 156.1 (q, CN), 157.3 (q, C-1 or C-3).

ν_{max} (ATR)/cm^{-1}: 3323 (NH), 1656, 1626, 1579, 1484, 1505 (CN), 1484, 1322, 1259, 1214, 1174, 1130, 1113, 1035, 832, 664.

HRMS (*m/z* - ES): Found: 463.1269 (M^+ + H. $C_{21}H_{19}N_6OF_3Cl$ Requires: 463.1261).

HPLC: 99.7% (*t*R: 27.4 min).

3-[4'-(4-Chloro-3-trifluoromethylphenyl)guanidinobenzoyl]phenylguanidine dihydrochloride (71b)



Following Method F, **54b** (159 mg, 0.21 mmol) was dissolved in 4M HCl in dioxane (0.925 mL, 3.7 mmol) and in additional dioxane (0.105 mL) until reach a final concentration of 0.2M. After 6 h stirring at 55 °C, the reaction was adjudged complete (TLC), solvents were evaporated and the residue was purified by reverse phase chromatography to afford the pure hydrochloride salt as a hygroscopic yellowish crystalline solid (54 mg, 54%). **Mp**: decomp. above 170 °C.

δ_H (600 MHz, D_2O): 7.36 (d, 2H, J 8.6, H-10, H-10'), 7.45 (dd, 1H, J 7.7, 2.3, H-6), 7.56-7.60 (m, 3H, H-13, H-16, H-17), 7.63 (t, 1H, J 7.7, H-5), 7.64 (app. t, 1H, J 2.3, H-2), 7.69 (dd, 1H, J 7.7, 2.3, H-4), 7.75 (d, 2H, J 8.6, H-9, H-9').

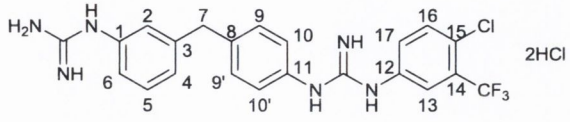
δ_C (150 MHz, D_2O): 123.0 (c, q, J 273, CF_3), 123.5 (2 CH Ar, C-10, C-10'), 123.9 (c, J 5.2, CH Ar, C-13), 127.1 (CH Ar, C-2), 128.3 (c, q, J 30, C-14), 129.2 (CH Ar, C-6), 129.4 (CH Ar, C-4), 129.6 (q, C-15), 130.2 (CH Ar, C-5), 130.7 (CH Ar, C-16 or C-17), 131.8 (2 CH Ar, C-10, C-10'), 132.6 (CH Ar, C-16 or C-17), 133.9 (q), 134.1 (q), 134.4 (q), 138.1 (q), 139.7 (q), 154.0 (q, CN), 156.3 (q, CN), 197.9 (q, CO).

ν_{max} (ATR)/cm^{-1}: 3285, 3143 (NH+), 2324, 1648, 1561 (CN), 1482, 1421, 1315, 1258, 1176, 1129, 1035, 969, 834, 759, 665.

HRMS (*m/z* - ES): Found: 475.1246 (M^+ + H. $C_{22}H_{19}N_6OF_3Cl$ Requires: 475.1261).

HPLC: 97.5% (*t*R: 26.1 min).

3-[4'-(4-Chloro-3-trifluoromethylphenyl)guanidinobenzyl]phenylguanidine dihydrochloride (71c)



Following Method F, **54c** (107 mg, 0.14 mmol) was dissolved in 4 M HCl in dioxane (0.635 mL, 2.54 mmol) and in additional dioxane (0.071 mL) until reach a final concentration of 0.2M. After 6 h stirring at 55 °C, the reaction was adjudged complete (TLC), solvents were evaporated and the residue was purified by reverse phase chromatography to afford the pure hydrochloride salt as a hygroscopic white crystalline solid (69 mg, 92%). **Mp**: decomp. above 140 °C.

δ_H (600 MHz, D_2O): 3.97 (s, 2H, CH_2), 7.12-7.13 (m, 2H, H-2, H-6), 7.18 (d, 2H, J 8.0, H-10, H-10'), 7.22 (d, 1H, J 7.7, H-4), 7.27 (d, 2H, J 8.0, H-9, H-9'), 7.38 (t, 1H, J 7.7, H-5), 7.42 (d, 1H, J 8.3, H-17), 7.56 (d, 1H, J 8.3, H-16), 7.64 (s, 1H, H-13).

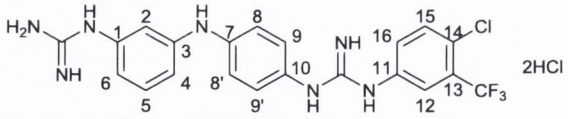
δ_C (150 MHz, D_2O): 40.1 (CH_2), 123.1 (c, q, J 273, CF_3), 123.6 (CH Ar, C-6), 124.4 (c, J 5.2, CH Ar, C-13), 125.6 (2 CH Ar, C-10, C-10'), 125.9 (CH Ar, C-2), 128.2 (CH Ar, C-4), 128.4 (c, q, J 30, C-14), 129.6 (CH Ar, C-17), 130.0 (2 CH Ar, C-9, C-9'), 130.1 (CH Ar, C-5), 132.5 (q, C-15), 132.6 (q, C-11), 132.6 (CH Ar, C-16), 134.1 (q, C-12), 134.2 (q, C-1), 140.6 (q, C-8), 143.2 (q, C-3), 154.5 (q, CN), 156.2 (q, CN).

ν_{max} (ATR)/cm^{-1}: 3025 (NH), 2970, 2949, 2162, 1738, 1655, 1625, 1571 (CN), 1511, 1433, 1366, 1322, 1257, 1229, 1217, 1128, 1034, 891, 830.

HRMS (*m/z* - ES): Found: 461.1468 (M^+ + H. $C_{22}H_{21}N_6F_3Cl$ Requires: 461.1468).

HPLC: 97.4% (*t*R: 27.6 min).

3-[4'-(4-Chloro-3-trifluoromethyl)phenylguanidino]phenylamino]phenylguanidine dihydrochloride (71d)



Following Method F, **54d** (272 mg, 0.357 mmol) was dissolved in 4M HCl in dioxane (1.6 mL, 6.43 mmol) and in additional dioxane (0.18 mL) until reach a final concentration of 0.2M. After 8 h stirring at 55 °C, the reaction was adjudged complete (TLC), solvents were evaporated and the residue was purified by reverse phase chromatography to afford the pure hydrochloride salt as a hygroscopic light dark solid (112 mg, 59%). **Mp**: decomp. above 205 °C.

δ_H (600 MHz, D_2O): 6.83 (d, 1H, J 8.1, H-4 or H-6), 6.93 (s, 1H, H-2), 7.02 (d, 1H, J 8.1, H-4 or H-6), 7.10 (d, 2H, J 8.6, H-8, H-8' or H-9, H-9'), 7.15 (d, 2H, J 8.6, H-8, H-8' or H-9, H-9'), 7.34 (t, 1H, J 8.1, H-5), 7.44 (dd, 1H, J 8.3, 2.1, H-16), 7.58 (d, 1H, J 8.3, H-15), 7.63 (d, 1H, J 2.1, H-12).

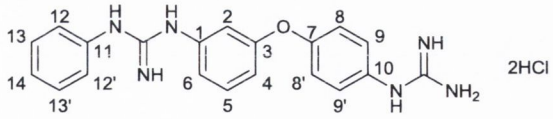
δ_C (150 MHz, D_2O): 114.1 (CH Ar, C-2), 117.0 (CH Ar, C-4 or C-6), 118.0 (CH Ar, C-4 or C-6), 118.8 (2 CH Ar, C-8, C-8' or C-9, C-9'), 122.2 (c, q, J 273, CF_3), 124.5 (c, J 5.2, CH Ar, C-12), 126.8 (q, C-7 or C-10), 127.0 (2 CH Ar, C-8, C-8' or C-9, C-9'), 128.4 (c, q, J 30, C-13), 129.8 (CH Ar, C-16), 130.8 (CH Ar, C-5), 132.5 (q, C-14), 132.6 (CH Ar, C-15), 133.6 (q, C-11), 135.0 (q, C-1 or C-3), 142.3 (q, C-7 or C-10), 144.4 (q, C-1 or C-3), 154.6 (q, CN), 156.2 (q, CN).

ν_{max} (ATR)/cm^{-1}: 3260 (NH), 3133 (NH), 2163, 1619, 1578 (CN), 1509, 1493, 1332, 1226, 1167, 1112, 1077, 1021, 997, 819, 757, 691.

HRMS (*m/z* - ES): Found: 462.1417 (M^+ + H. $C_{21}H_{20}N_7F_3Cl$ Requires: 462.1421).

HPLC: 99.9% (*t*R: 26.75 min).

4-[3'-(Phenylguanidino)phenoxy]phenylguanidine dihydrochloride (72a)



Following Method F, **55a** (108 mg, 0.164 mmol) was dissolved in 4M HCl in dioxane (0.74 mL, 2.94 mmol) and in additional dioxane (0.084 mL) until a final concentration of 0.2M was reached. After 7 h stirring at 55 °C, the reaction was adjudged complete (TLC), solvents were evaporated and the residue was purified by reverse phase chromatography to afford the pure hydrochloride salt as a hygroscopic off-white solid (50 mg, 70%). **Mp**: decomp. above 289 °C.

δ_H (600 MHz, D_2O): 7.02-7.03 (m, 2H, H-2 and H-4 or H-6), 7.11-7.12 (m, 3H, H-4 or H-6 and H-8, H-8' or H-9, H-9'), 7.27-7.29 (m, 4H, H-8, H-8'or H-9, H-9' and H-12, H-12'), 7.37 (t, 1H, J 7.4, H-14), 7.43-7.46 (m, 3H, H-13, H-13' and H-5).

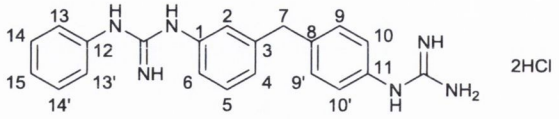
δ_C (150 MHz, D_2O): 115.6 (CH Ar, C-2), 117.8 (CH Ar, C-4 or C-6), 120.2 (2 CH Ar, C-8, C-8' or C-9, C-9'), 120.6 (CH Ar, C-4 or C-6), 125.5 (2 CH Ar, C-12, C-12'), 127.9 (CH Ar, C-14), 128.0 (2 CH Ar, C-8, C-8' or C-9, C-9'), 129.7 (q, C-7 or C-10), 129.8 (2 CH Ar, C-13, C-13'), 131.2 (CH Ar, C-5), 134.0 (q, C-11), 135.8 (q, C-1 or C-3), 154.7 (q, CN), 155.6 (q, C-7 or C-10), 156.4 (q, CN), 157.4 (q, C-1 or C-3).

ν_{max} (ATR)/cm^{-1}: 3159 (NH), 3058 (NH), 2289, 1560 (CN), 1504, 1484, 1449, 1215, 1166, 1150, 1102, 1076, 1014, 1002, 964, 842, 788, 752, 690.

HRMS (*m/z* - ES): Found: 361.1772 (M^+ + H. $C_{20}H_{21}N_6O$ Requires: 361.1777).

HPLC: 99.5% (*t*R: 23.13 min).

4-[3'-(Phenyl)guanidinobenzyl]phenylguanidine dihydrochloride (72c)



Following Method F, **55c** (164 mg, 0.25 mmol) was dissolved in 4M HCl in dioxane (1.122 mL, 4.49 mmol) and in additional dioxane (0.125 mL) until reach a final concentration of 0.2M. After 6 h stirring at 55 °C, the reaction was adjudged complete (TLC), solvents were evaporated and the residue was purified by reverse phase chromatography to afford the pure hydrochloride salt as a white crystalline solid (83 mg, 80%). **Mp**: decomp. above 170 °C.

δ_H (600 MHz, D_2O): 4.00 (s, 2H, CH_2), 7.15-7.20 (m, 4H, H-2, H-6, H-10, H-10'), 7.25 (d, 1H, J 8.0, H-4), 7.27 (d, 2H, J 7.7, H-13, H-13'), 7.32 (d, 2H, J 8.2, H-9, H-9'), 7.34-7.39 (m, 2H, H-5, H-15), 7.43 (t, 2H, J 7.7, H-14, H-14').

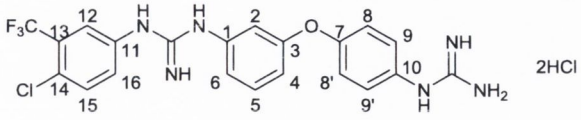
δ_C (150 MHz, D_2O): 40.1 (CH_2), 123.4 (CH Ar, C-6), 125.6 (2 CH Ar, C-13 and C-13'), 125.7 (CH Ar, C-2), 126.1 (2 CH Ar, C-10 and C-10'), 127.8 (CH Ar, C-15), 128.1 (CH Ar, C-4), 129.8 (2 CH Ar, C-14 and C-14'), 130.1 (CH Ar, C-5), 130.1 (2 CH Ar, C-9, C-9'), 132.1 (q, C-11), 134.0 (q, C-12), 134.4 (q, C-1), 140.9 (q, C-8), 143.2 (q, C-3), 154.8 (q, CN), 156.2 (q, CN).

ν_{max} (ATR)/cm^{-1}: 3029 (NH), 2970, 2162, 1738, 1571 (CN), 1492, 1440, 1366, 1230, 1217, 756, 692.

HRMS (*m/z* - ES): Found: 369.1980 (M^+ + H. $C_{21}H_{23}N_6$ Requires: 359.1984).

HPLC: 98.5% (*t*R: 23.6 min).

4-[3'-(4-Chloro-3-trifluoromethylphenyl)guanidinophenoxy]phenylguanidine dihydrochloride (73a)



Following Method F, **56a** (150 mg, 0.197 mmol) was dissolved in 4M HCl in dioxane (0.89 mL, 3.54 mmol) and in additional dioxane (0.1 mL) until a final concentration of 0.2M was reached. After 6 h stirring at 55 °C, the reaction was adjudged complete (TLC), solvents were evaporated and the residue was purified by reverse phase chromatography to afford the pure hydrochloride salt as a hygroscopic off-white solid (103 mg, 98%). **Mp**: decomp. above 125 °C.

δ_H (600 MHz, D_2O): 6.95 (app. t, 1H, J 2.2, H-2), 6.98 (dd, 1H, J 8.5, 2.2, H-4 or H-6), 7.06-7.09 (m, 3H, H-8, H-8' and H-4 or H-6), 7.24 (d, 2H, J 8.7, H-9 and H-9'), 7.41 (t, 1H, J 8.5, H-5), 7.44 (dd, 1H, J 8.5, 2.2, H-16), 7.60 (d, 1H, J 8.5, H-15), 7.64 (d, 1H, J 2.2, H-12).

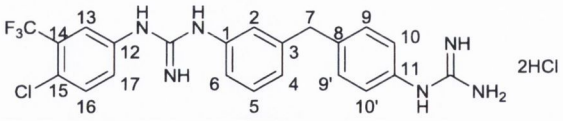
δ_C (150 MHz, D_2O): 115.0 (CH Ar, C-2), 117.6 (CH Ar, C-4 or C-6), 120.1 (CH Ar, C-4 or C-6), 121.2 (2 CH Ar, C-8, C-8' or C-9, C-9'), 123.0 (c, q, J 273, CF_3), 124.1 (c, J 6.5, CH Ar, C-12), 127.8 (2 CH Ar, C-8, C-8' or C-9, C-9'), 128.5 (c, q, J 30, C-13), 129.4 (CH Ar, C-16), 129.7 (q, C-7 or C-10), 129.8 (q, C-14), 131.2 (CH Ar, C-5), 132.7 (CH Ar, C-15), 133.9 (q, C-11), 135.8 (q, C-1 or C-3), 154.4 (q, CN), 155.5 (q, C-7 or C-10), 156.4 (q, CN), 157.4 (q, C-1 or C-3).

ν_{max} (ATR)/cm^{-1}: 3294 (NH), 3117 (NH), 1981, 1663, 1583 (CN), 1505, 1482, 1321, 1258, 1215, 1173, 1130, 1034, 878, 829, 665.

HRMS (*m/z* - ES): Found: 463.1253 (M^+ + H. $C_{21}H_{19}N_6OF_3Cl$ Requires: 463.1261).

HPLC: 99.3% (*t*R: 27.7 min).

4-[3'-(4-Chloro-3-trifluoromethylphenyl)guanidinobenzyl]phenylguanidine dihydrochloride (73c)



Following Method F, **56c** (110 mg, 0.145 mmol) was dissolved in 4M HCl in dioxane (0.651 mL, 2.6 mmol) and in additional dioxane (0.072 mL) until reach a final concentration of 0.2M. After 6 h stirring at 55 °C, the reaction was adjudged complete (TLC), solvents were evaporated and the residue was purified by reverse phase chromatography to afford the pure hydrochloride salt as a hygroscopic white crystalline solid (62 mg, 80%). **Mp**: 135-137 °C.

δ_H (600 MHz, D_2O): 3.91 (s, 2H, CH_2, H-7), 7.05-7.10 (m, 4H, H-2, H-6, H-10, H-10'), 7.15 (d, 1H, J 7.8, H-4), 7.23 (d, 2H, J 8.5, H-9, H-9'), 7.29 (d, 1H, J 7.8, H-5), 7.34 (dd, 1H, J 8.5, 2.5, H-17), 7.51 (d, 1H, J 8.5, H-16), 7.57 (d, 1H, J 2.5, H-13).

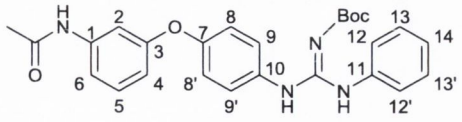
δ_C (150 MHz, D_2O): 40.0 (CH_2), 122.8 (CH Ar, C-6), 123.5 (c, q, J 273, CF_3), 124.1 (c, J 5.2, CH Ar, C-13), 125.2 (CH Ar, C-2), 125.8 (2 CH Ar, C-10 and C-10'), 127.9 (CH Ar, C-4), 128.4 (c, q, J 30, C-14), 129.4 (CH Ar, C-17), 129.8 (q, C-15), 130.0 (CH Ar, C-5), 130.2 (2 CH Ar, C-9 and C-9'), 132.1 (q, C-11), 132.7 (CH Ar, C-16), 133.7 (q, C-12), 134.4 (q, C-1), 140.8 (q, C-8), 143.1 (q, C-3), 154.5 (q, CN), 156.1 (q, CN).

ν_{max} (ATR)/cm^{-1}: 3118 (NH), 2970, 2163, 1655, 1577 (CN), 1505, 1483, 1318, 1258, 1214, 1172, 1146, 1035, 949, 833, 753, 689.

HRMS (*m/z* - ES): Found: 461.1473 (M^+ + H. $C_{22}H_{21}N_6F_3Cl$ Requires: 461.1468).

HPLC: 97.7% (*t*R: 27.41 min).

_N_-{3-[4-(_N'_-Phenyl-_N''_-_tert_-butoxycarbonylguanidino)phenoxy]phenyl}acetamide (76a)



Following method D, to a solution of **66a** (573 mg, 1.37 mmol), and acetic anhydride (280 mg, 2.74 mmol) in DCM (8.65 mL), triethylamine (416 mg, 4.11 mmol) was added. The reaction was stirred at room temperature for 6 h. Workup as described, followed by silica gel chromatography, eluting with hexane:EtOAc, afforded the title compound **76a** as a white solid (338 mg, 53%). **Mp**: 97-99 °C.

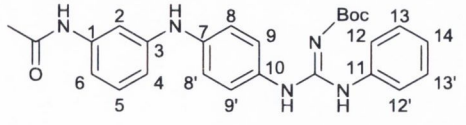
δ_H (400 MHz, $CDCl_3$): 1.50 (s, 9H, $(CH_3)_3$), 2.13 (s, 3H, CH_3), 6.70-8.00 (Aromatic signals in this range could not be assigned as they were broad and highly amorphous), 7.65 (bs, NH), 9.54 (bs, NH).

δ_C (100 MHz, $CDCl_3$): 24.5 (CH_3), 28.8 ($(CH_3)_3$), 100-180 (Signals in this range could not be assigned as they were broad and highly amorphous).

ν_{max} (ATR)/cm^{-1}: 3262 (NH), 2974, 2928, 2284, 2165, 2050, 1981, 1722 (CO), 1638 (CN), 1592, 1542, 1485, 1438, 1420, 1390, 1366, 1334, 1287, 1245, 1209, 1148, 1089, 1015, 966, 919, 853, 804, 767, 750, 689.

HRMS (*m/z* - ES): Found 461.2191 (M^+ + H. $C_{26}H_{29}N_4O_4$ Requires 461.2189).

_N_-{3-[4-(_N'_-Phenyl-_N''_-_tert_-butoxycarbonylguanidino)phenylamino]phenyl} acetamide (76d)



Following method D, to a solution of **66d** (234 mg, 0.56 mmol), and acetic anhydride (114 mg, 1.12 mmol) in DCM (2.8 mL), triethylamine (170.0 mg, 1.68 mmol) was added. The reaction was stirred at room temperature for 6 h. Workup as described, followed by silica gel chromatography, eluting with a gradient of hexane:EtOAc, afforded the title compound **76d** as a light brown solid (177 mg, 69%). **Mp**: 110-112 °C.

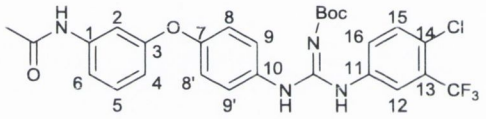
δ_H (400 MHz, $CDCl_3$): 1.49 (s, 9H, $(CH_3)_3$), 2.13 (s, 3H, CH_3), 5.75 (s, 1H, NH linker), 6.50-7.50 (Aromatic signals in this range could not be assigned as they were broad and highly amorphous), 7.78 (bs, NH), 9.54 (bs, NH).

δ_C (100 MHz, $CDCl_3$): 24.5 (CH_3), 28.2 ($(CH_3)_3$), 108.1 (CH Ar), 112.2 (CH Ar), 119.8 (CH Ar), 121.5 (CH Ar), 123.3 (CH Ar), 129.5 (CH Ar), 129.7 (CH Ar), 139.3 (q), 140-165 (Quaternary signals in this range could not be assigned as they were broad and highly amorphous), 168.8 (q).

ν_{max} (ATR)/cm^{-1}: 3288 (NH), 2975, 2938, 1980, 1722 (CO), 1641 (CN), 1591, 1510, 1497, 1453, 1389, 1367, 1311, 1239, 1154, 1092, 964, 905, 803, 773, 728, 691.

HRMS (*m/z* - ES): Found 460.2350 (M^+ + H. $C_{26}H_{30}N_5O_3$ Requires 460.2349).

_N_-(3-{4-[_N'_-(4-Chloro-3-trifluoromethylphenyl)-_N''_-_tert_-butoxycarbonylguanidino] phenoxy}phenyl)acetamide (77a)



Following method D, to a solution of **67a** (391 mg, 0.75 mmol), and acetic anhydride (153 mg, 1.5 mmol) in DCM (3.75 mL), triethylamine (228 mg, 2.25 mmol) was added. The reaction was stirred at room temperature for 6 h. Workup as described, followed by silica gel chromatography, eluting with a gradient of hexane:EtOAc, afforded the title compound **77a** as a red solid (376 mg, 89%). **Mp**: 59-61 °C.

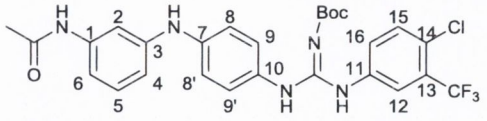
δ_H (400 MHz, $CDCl_3$): 1.52 (s, 9H, $(CH_3)_3$), 2.13 (s, 3H, CH_3), 6.50-8.00 (Aromatic signals in this range could not be assigned as they were broad and highly amorphous), 6.78 (bs, NH), 7.70 (bs, NH), 8.44-9.95 (m, bs, NH).

δ_C (100 MHz, $CDCl_3$): 24.4 (CH_3), 27.9 ($(CH_3)_3$), 83.7 (q, $\underline{C}(CH_3)_3$), 109.3 (CH Ar), 109.7 (CH Ar), 113.8 (CH Ar), 114.1 (CH Ar), 118.7 (CH Ar), 119.8 (CH Ar), 120.7 (CH Ar), 122.1 (CH Ar), 123.4 (CH Ar), 123.5 (c, q, J 273, CF_3), 126.7 (CH Ar), 128.4 (d, q, J 30, C-13), 129.3 (q, C-14), 129.9 (CH Ar), 131.5 (CH Ar), 132.4 (CH Ar), 137.8 (q), 139.3 (q), 140.1 (q), 141.8 (q), 152.4 (q), 152.9 (q), 158.2 (q), 168.3 (q).

ν_{max} (ATR)/cm^{-1}: 3264 (NH), 2980, 2933, 2163, 1722 (CO), 1655 (CN), 1597, 1547, 1502, 1481, 1438, 1316, 1246, 1208, 1089, 1031, 1015, 967, 920, 836, 772, 687, 666.

HRMS (*m/z* - ES): Found 561.1514 (M^- - H. $C_{27}H_{25}N_4O_4ClF_3$ Requires 561.1516).

_N_-(3-{4-[_N'_-(4-Chloro-3-trifluoromethylphenyl)-_N''_-*tert*-butoxycarbonylguanidino] phenylamino}phenyl)acetamide (77d)



Following method D, to a solution of **69d** (442 mg, 0.85 mmol), and acetic anhydride (174 mg, 1.70 mmol) in DCM (4.2 mL), triethylamine (258 mg, 2.55 mmol) was added. The reaction was stirred at room temperature for 6 h. Workup as described, followed by silica gel chromatography, eluting with hexane:EtOAc, afforded the title compound **77d** as a red solid (291 mg, 61%). **Mp**: 100-102 °C.

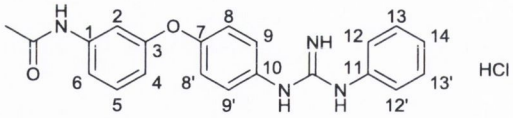
δ_H (400 MHz, $CDCl_3$): 1.51 (s, 9H, $(CH_3)_3$), 2.12 (s, 3H, CH_3), 5.82 (bs, 1H, NH linker), 5.84 (bs, 1H, NH linker), 6.50-8.50 (Aromatic signals in this range could not be assigned as they were broad and highly amorphous), 7.90 (bs, NH), 9.30 – 10.0 (m, bs, NH).

δ_C (100 MHz, $CDCl_3$): 24.4 (CH_3), 28.03 ($(CH_3)_3$), 83.6 (q, $\underline{C}(CH_3)_3$), 108.1 (CH Ar), 112.1 (CH Ar), 118.7 (CH Ar), 119.6 (CH Ar), 120.7 (CH Ar), 121.7 (CH Ar), 122.0 (CH Ar), 123.0 (CH Ar), 123.5 (c, q, J 273, CF_3), 123.6 (CH Ar), 126.9 (CH Ar), 129.6 (CH Ar), 131.6 (CH Ar), 132.2 (CH Ar), 137.9 (q), 139.1 (q), 140.1 (q), 142 – 150 (Quaternary signals in this range could not be assigned as they were broad and highly amorphous), 153 (q), 168.7 (q).

ν_{max} (ATR)/cm^{-1}: 3312 (NH), 2982, 2936, 2080, 1724 (CO), 1650 (CN), 1598, 1509, 1479, 1425, 1370, 1315, 1239, 1138, 1090, 1043, 1033, 994, 938, 833, 773, 725, 690, 666.

HRMS (*m/z* - ES): Found 560.1696 (M^- - H. $C_{27}H_{26}N_5O_3ClF_3$ Requires 560.1676).

_N_-{3-[4-(_N'_-Phenylguanidino)phenoxy]phenyl}acetamide hydrochloride (78a)



Following Method G, a solution of **76a** (338 mg, 0.73 mmol) in 1:1 DCM:IPA (2.55 mL) was treated with 4 M HCl in dioxane (1.1 mL). After 6 h. stirring at 30 °C, the reaction was adjudged complete by TLC. Solvents were evaporated and the residue was extracted with DCM and purified by reverse phase chromatography to afford the pure hydrochloride salt as a white solid (26 mg, 61%) **Mp**: 130-132 °C.

δ_H (600 MHz, D_2O): 2.08 (s, 3H, CH_3), 6.85 (d, 1H, J 8.2, H-6 or H-4), 7.09 (m, 3H, H-6 or H-4 and H-8, H-8'or H-9, H-9'), 7.19 (s, 1H, H-2), 7.28-7.40 (m, 6H, H-8, H-8' or H-9, H-9' and H-5, H-12, H-12', H-14), 7.45 (t, 2H, H-13, H-13').

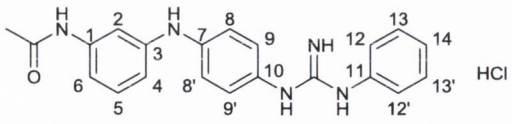
δ_C (150 MHz, D_2O): 20.4 (CH_3), 112.3 (CH Ar, C-2), 115.6 (CH Ar, C-4 or C-6), 116.9 (CH Ar, C-4 or C-6), 120.0 (2 CH Ar, C-8, C-8' or C-9, C-9'), 125.9 (2 CH Ar, C-12, C-12'), 127.9 (2 CH Ar, C-8, C-8' or C-9, C-9'), 128.0 (CH Ar, C-14), 129.4 (q, C-7 or C-10), 129.9 (2 CH Ar, C-13, C-13'), 130.4 (CH Ar, C-5), 133.9 (q, C-11), 138.4 (q, C-1 or C-3), 155.1 (q, C-1 or C-3), 156.0 (q, C-7 or C-10), 156.7 (q, CN), 172.9 (q, CO).

ν_{max} (ATR)/cm^{-1}: 3242 (NH), 2162, 1654, 1624 (CO), 1596 (CN), 1579, 1542, 1500, 1484, 1420, 1370, 1289, 1255, 1212, 1166, 1142, 1017, 966, 920, 873, 838, 754, 689, 663.

HRMS (*m/z* - ES): Found 361.1676 (M^+ + H. $C_{21}H_{21}N4O_2$ Requires 361.1665).

HPLC: 97.3% (*t*R: 27.4 min).

_N_-{3-[4-(_N'_-Phenylguanidino)phenylamino]phenyl}acetamide hydrochloride (78d)



Following Method G, a solution of **76d** (177 mg, 0.385 mmol) in 1:1 DCM:IPA (1.33 mL) was treated with 4 M HCl in dioxane (0.6 mL). After 6 h. stirring at 30 °C, the reaction was adjudged complete by TLC. Solvents were evaporated and the residue was extracted with DCM and purified by reverse phase chromatography to afford the pure hydrochloride salt as an off-white solid (85 mg, 56%). **Mp**: 137-139 °C.

δ_H (600 MHz, D_2O): 2.03 (d, 3H, CH_3), 6.78-6.84 (m, 1H, H-4, H-6), 7.03 (d, 2H, J 8.7, H-8, H-8' or H-9, H-9'), 7.90 (d, 2H, J 8.7, H-8, H-8' or H-9, H-9'), 7.14-7.19 (m, 2H, H-2, H-5), 7.24 (d, 2H, J 7.5, H-12, H-12'), 7.34 (t, 1H, J 7.5, H-14), 7.43 (d, 2H, J 7.5, H-13, H-13').

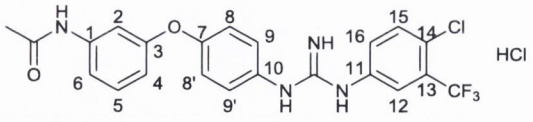
δ_C (150 MHz, D_2O): 22.0 (CH_3), 110.7 (CH Ar, C-2), 114.1 (CH Ar, C-4 or C-6), 114.7 (CH Ar, C-4 or C-6), 188.2 (2 CH Ar, C-8, C-8' or C-9, C-9'), 125.8 (2 CH Ar, C-12, C-12'), 126.6 (q, C-7 or C-10), 127.5 (2 CH Ar, C-8, C-8' or C-9, C-9'), 128.0 (CH Ar, C-14), 129.9 (2 CH Ar, C-13, C-13'), 130.5 (CH Ar, C-5), 133.8 (q, C-11), 135.0 (q, C-1 or C-3), 142.5 (q, C-7 or C-10), 144.2 (q, C-1 or C-3), 155.2 (q, CN), 156.1 (q, CO).

ν_{max} (ATR)/cm^{-1}: 3134 (NH), 2985, 2885, 2323, 1980, 1739 (CO), 1595 (CN), 1509, 1484, 1428, 1397, 1370, 1313, 1230, 1164, 1114, 997, 964, 932, 835, 757, 712, 689.

HRMS (*m/z* - ES): Found 358.1685 (M^- - H. $C_{21}H_{20}N_5O$ Requires 358.1668).

HPLC: 99.5% (*t*R: 25.2 min).

_N_-(3-{4-[_N'_-(4-Chloro-3-trifluoromethylphenyl)guanidino]phenoxy}phenyl)acetamide hydrochloride (79a)



Following Method G, a solution of **77a** (376 mg, 0.67 mmol) in 1:1 DCM:IPA (2.35 mL) was treated with 4 M HCl in dioxane (1 mL). After 6 h. stirring at 30 °C, the reaction was adjudged complete by TLC. Solvents were evaporated and the residue was extracted with DCM and purified by reverse phase chromatography to afford the pure hydrochloride salt as a white solid (174 mg, 52%) **Mp**: 140-142 °C.

δ_H (600 MHz, D_2O): 2.09 (s, 3H, CH_3), 6.81 (dd, 1H, J 8.1, 1.8, H-4 or H-6), 7.05 (d, 2H, J 8.9, H-8, H-8' or H-9, H-9'), 7.09 (dd, 1H, J 8.1, 1.8, H-4 or H-6), 7.17 (app. t, 1H, J 1.8, H-2), 7.27 (d, 2H, J 8.9, H-8, H-8' or H-9, H-9'), 7.36 (t, 1H, J 8.1, H-5), 7.47 (d, 1H, J 8.5, H-16), 7.61 (d, 1H, J 8.5, H-15), 7.68 (s, 1H, H-12).

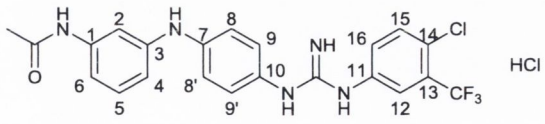
δ_C (150 MHz, D_2O): 22.6 (CH_3), 112.4 (CH Ar, C-2), 115.5 (CH Ar, C-4 or C-6), 117.1 (CH Ar, C-4 or C-6), 119.8 (2 CH Ar, C-8, C-8' or C-9, C-9'), 123.0 (c, q, J 273, CF_3), 124.6 (c, J 5.2, CH Ar, C-12), 127.5 (2 CH Ar, C-8, C-8' or C-9, C-9'), 128.5 (c, q, J 30, C-13), 129.3 (q, C-7 or C-10), 129.9 (CH Ar, C-16), 130.0 (q, C-14), 130.3 (CH Ar, C-5), 132.7 (CH Ar, C-15), 133.5 (q, C-11), 138.3 (q, C-1 or C-3), 154.7 (q, CN), 156.0 (q, C-7 or C-10), 156.6 (q, C-1 or C-3), 172.8 (q, CO).

ν_{max} (ATR)/cm^{-1}: 3126 (NH), 2971, 2878, 2035, 1659 (CO), 1596 (CN), 1544, 1505, 1483, 1418, 1257, 1216, 1173, 1139, 1113, 1036, 967, 921, 878, 832, 785, 713, 668.

HRMS (*m/z* - ES): Found 463.1162 (M^+ + H. $C_{22}H_{19}N_4O_2F_3Cl$ Requires 463.1149).

HPLC: 96.1% (*t*R: 30.51 min).

_N_-(3-[4-[_N'_-(4-Chloro-3-trifluoromethylphenyl)guanidino]phenylamino]phenyl) acetamide hydrochloride (79d)



Following Method G, a solution of **77d** (238 mg, 0.42 mmol) in 1:1 DCM:IPA (1.48 mL) was treated with 4 M HCl in dioxane (0.63 mL). After 6 h. stirring at 30 °C, the reaction was adjudged complete by TLC. Solvents were evaporated and the residue was extracted with DCM and purified by reverse phase chromatography to afford the pure hydrochloride salt as a white solid (83 mg, 43%). **Mp**: 155-157 °C.

δ_H (600 MHz, D_2O): 2.11 (s, 3H, CH_3), 6.87-6.91 (m, 2H, H-4, H-6), 7.11 (d, 2H, J 8.7, H-8, H-8' or H-9, H-9'), 7.17 (d, 1H, J 8.7, H-8, H-8' or H-9, H-9'), 7.23 (s, 1H, H-2), 7.28 (t, 1H, J 8.0, H-5), 7.47 (d, 1H, J 8.1, H-16), 7.61 (d, 1H, J 8.1, H-15), 7.67 (s, 1H, H-12).

δ_C (150 MHz, D_2O): 22.6 (CH_3), 111.3 (CH Ar, C-2), 114.7 (CH Ar, C-4 or C-6), 115.1 (CH Ar, C-4 or C-6), 118.6 (2 CH Ar, C-8, C-8' or C-9, C-9'), 123.5 (c, q, J 273, CF_3), 124.7 (c, J 5.2, CH Ar, C-12), 127.2 (2 CH Ar, C-8, C-8' or C-9, C-9'), 128.5 (c, q, J 30, C-13), 129.3 (q, C-7 or C-10), 129.9 (CH Ar, C-16), 130.0 (CH Ar, C-5), 130.3 (q, C-14), 132.7 (CH Ar, C-15), 133.5 (q, C-11), 137.8 (q, C-1 or C-3), 154.7 (q, CN), 156.0 (q, C-7 or C-10), 156.6 (q, C-1 or C-3), 172.9 (q, CO).

ν_{max} (ATR)/cm^{-1}: 3073 (NH), 3016, 2971, 2360, 2036, 1982, 1915, 1739 (CO), 1588 (CN), 1547, 1509, 1484, 1460, 1416, 1366, 1290, 1279, 1229, 1217, 1170, 1140, 1037, 1016, 973, 879, 856, 767, 728, 687.

HRMS (*m/z* - ES): Found: 460.1155 (M^- - H. $C_{22}H_{18}N_5OF_3Cl$ Requires 460.1152).

HPLC: 95.8% (*t*R: 29.92 min).

7.2. Biochemical Procedures

7.2.1. General Procedures

Cell Storage/Cryopreservation

Low passage number stocks of cells were harvested by centrifugation at 168 × g for 5 min. In the case of the suspension HL-60 cell line, the supernatant was discarded and the pellet was resuspended in 10 mL medium. Cells were then counted and dissolved in freezing medium (90% FBS, 10% DMSO) at a density of 10 million cells/mL and transferred to sterile cryotubes. For the rest of adherent cells, cells were grown to confluence in T75 flasks, washed with 10 mL PBS and trypsinised in the usual manner. Once the cells were harvested, pellets were resuspended in 4 - 5 mL freezing medium and aliquoted to the cryotubes. Cells were frozen down at a cooling rate of 1 °C/min by incubating the cyrotubes in a freezing container (Nalgene "Mr. Frosty"), filled with isopropyl alcohol overnight at -80 °C. Finally the cyrotubes were transferred to liquid nitrogen for storage at -180 °C.

When an aliquot of cells was required, the cells were quickly removed from the liquid nitrogen and thawed at 37 °C for 2 min. Just before the samples had fully thawed, their contents were gently pipetted into a sterile 20 mL tube containing medium supplemented with 20% FBS. This is done in order to prevent damage to the cells which are very fragile after cryopreservation. The cells were then centrifuged at 168 × g for 5 min. The supernatant was discarded and the pellet was resuspended in 5 mL of medium. This solution was then added to a T25 flask and the cells monitored closely over the next few days.

Use of Haemocytometer

A haemocytometer is a modified microscope slide containing an accurately sub-divided grid that enables the counting of cells in suspension. Each grid (there are two) consists of nine large squares, each measuring 1 mm^2. Each of these primary squares contains 16 medium squares, each measuring 0.04 mm^2. When a particular seeding density of cells was required, a clean cover slip was placed on the haemocytometer; 10 μL of the cell suspension was then pipetted into the groove at the end of the plane. The cell suspension

was then drawn across the grid by capillary action. The number of cells within the four outer primary squares was counted. Those that touched the top or right hand sides of the grid or were outside the four squares were not counted. As each medium square is 0.04 mm^3, the total volume per primary square is 1×10^{-4} mL. This implies that the total cell concentration in the original suspension (cells/mL) is = average cell count/primary square $\times 1 \times 10^4$.

Preparation of the Drugs

The compounds were dried and specific quantities (2 – 5 mg) weighed out. These were dissolved in sterile ddH_2O and 10 mM solutions made up. Once in solution, these were filtered using sterile syringes and 0.2 μM acrodiscs in order to ensure sterility. Required concentration ranges (10 mM – 0.1 mM) of each drug were prepared in sterile ddH_2O in eppendorf tubes and stored at –20 °C until required. The range of concentration is selected in relation to the activity of the drugs in each cell line.

7.2.2. Growth and Maintenance of the Cell Lines

HL-60 Cell Line

The HL-60 (human caucasian promyelocytic leukemia) cell line was maintained between 200,000 – 2,000,000 cells/mL in Roswell Park Memorial Institute (RPMI) 1640 medium with stable glutamate (GlutaMax I) supplemented with 10% (v/v) foetal bovine serum (FBS) and 50 μg/mL penicillin/streptomycin (pen-strep). The growth medium was stored in the fridge at 4 °C and heated to 37 °C prior to culture work. Cells were grown at 37 °C in a humidified environment maintained at 95% O_2 and 5% CO_2 and passaged at least three times weekly depending on their levels of confluency. When required for sub-culturing, cells were transferred to a sterile tube and centrifuged at 168 × g for 5 min. The supernatant was discarded and the cell pellet was resuspended in 10 mL of fresh medium. Cells were then counted using a haemocytometer slide and seeded at the required density (200,000 cells/mL).

HUH-7 Cell Line

The HUH-7 (Human Hepatoma) is a human liver cancinoma cell line grown in with Dulbecco's modified Eagle medium (DMEM) with stable glutamate (GlutaMax) supplemented with 10% (v/v) FBS and 50 μg/mL pen-strep. The growth medium was stored in the fridge at 4 °C and heated to 37 °C prior to culture work. The adherent cells were grown at 37 °C in a humidified environment maintained at 95% O_2 and 5% CO_2 and passaged at least twice weekly depending on their levels of confluency. When required for sub-culturing, the medium was removed and the cells were rinsed with 10 mL PBS. 10 mL of 1x trypsin/EDTA solution was then added and the cells incubated at 37 °C for 3 – 5 min in order to disperse the cell layer and detach the cells from the bottom of the flask. The cells were observed under a light microscope and gently agitated if necessary by hitting the flask to complete the detachment. Once dispersed, 10 mL medium was added to deactivate the trypsin/EDTA and the solution transferred to a sterile tube and centrifuged at 168 × g for 5 min. The supernatant was discarded and the pellet was resuspended in 10 mL fresh medium (resuspending first in 1mL of media with pipette P1000 and then adding 9 mL on top because adherent cells tend to form clumps). Cells were counted using a haemocytometer slide and seeded at an appropriate density (1:3 cells:media).

RKO Cell Lines

Three types of RKO colorectal cancer cell lines were used, F-6-8 cell line which have the BRAF V600E mutant allele and a wild-type allele (dominant mutant), the A19 isogenic clone contained the mutant allele only, and the T29 isogenic clone contained the wild type allele only. The isogenic clones were formed by homologous recombination. All cells were maintained in McCoy's 5A media. Cells were obtained from Bert Vogelstein (John Hopkins University, Baltimore, USA). Cells were maintained in culture media supplemented with 10% dialysed foetal calf serum, 50μg/mL penicillin streptomycin, 2mM L-glutamine and 1mM sodium pyruvate. All experiments were performed with 2% dialysed foetal calf serum, 2 mM L-glutamine and 1mM sodium pyruvate (Invitrogen). All cells were grown at 37°C in a 5% CO_2 humidified incubator. Cells were grown in a T-75 tissue culture flask until around 70% confluent. Medium was removed and the cells were washed with 4mL sterile 1X PBS. This was removed and 1mL 1X trypsin/PBS added, and the flask was then placed back into the incubator until all the

cells detached from the surface of the flask. 9 mL of the appropriate cell medium was added to each flask and the cells were gently pipetted up and down to break up any cell clumps. 2 mL of this cell suspension was then placed in a new flask and 18 mL of medium added. Cells were then returned to the incubator. Passage of cells occurred approximately every 3 days.

MCF-7 Wild Type and Tamoxifen Resistant MCF-7

The MCF-7 wild type cell line was maintained in RPMI 1640 containing 5% fetal calf serum (FCS), penicillin-streptomycin (100 IU/mL – 100 μg/mL) and fungizone (2.5 μg/mL). The TAM-MCF-7 cell line was maintained in phenol-red free RPMI containing 5% charcoal stripped (i.e. steroid-depleted) fetal calf serum (SFCS), penicillin-streptomycin (100iU/mL – 100 μg/mL), fungizone (2.5 μg/ml) and glutamine (4 mM), together with 0.1μM 4-hydroxytamoxifen (4-OH-TAM). Cells were cultured at 37°C in a humidified atmosphere of 5% CO_2. The routinely maintenance of cell cultures consisted of medium change every 3-5 days and passaging approximately every 7 days with a seeding ratio of 1:10. Passaging was done by disruption of the cell monolayer by Trypsin/EDTA (0.05%/0.02% w/v), centrifugation at 250 × g for 5 min, and resuspension of the cell pellet in the appropriate medium.All tissue culture medium and constituents were purchased from Gibco Europe Ltd. (Paisley, Scotland) and tissue culture plastic ware was obtained from Nunc (Roskilde, Denmark).

7.2.3. Cell Viability Assays

AlamarBlue® Assay of HL-60 and HUH-7 Cell Lines

HL-60 cells in the log phase of growth were seeded in 96-well plates at a density of 200,000 cells/mL (200 μL/well or 40,000 cells/well) in complete RPMI medium the day before in case of HUH-7 cells and the same day of the experiment for HL-60 cell line. The cells were then treated with 2 μL of a 1:100 dilution of stock concentrations of drugs or ddH_2O as vehicle control in triplicate. Three wells containing 200 μL RPMI with no cells were also set-up as blanks. HUH-7 cells were seeded in a 96-well plate at a density of 50,000 cells/mL (200 μL/well or 10,000/well) and incubated at 37 °C for 24 h to led them

attach. The cells were then treated with the drugs as for the HL-60 cells. After 72 h incubation, 20 µL AlamarBlue® was added to each well. The plates were incubated in darkness at 37 °C for 5 hours. Using a Molecular Devices microplate reader, the fluorescence (F) was then read at an excitation wavelength of 544 nm and an emission wavelength of 590 nm. Cell viability was then determined by subtracting the mean blank fluorescence (Fb) from the treated sample fluorescence (Fs) and expressing this as a percentage of the fluorescence of the blanked vehicle control (Fc). This is demonstrated in the equation below. The results were then plotted as nonlinear regression, sigmoidal dose-response curves on Prism GraphPad 5 software, from which the IC_{50} value for each drug was determined.

$$\frac{(Fs - Fb)}{(Fc - Fb)} \times \frac{100}{1} = \% \text{ Cell Viability}$$

MTT Assay of RKO Cell Lines

IC_{50} values were calculated from cell viability data using Thiazolyl Blue Tetrazolium Bromide (MTT) assays (Sigma-Aldrich, Dorset, UK). Cells were seeded at 3,000 cells/well in 200µl of 2% M^{c}Coy's 5A medium in a 96-well plate and left to incubate overnight at 37 °C in a humidified environment maintained at 95% O_2 and 5% CO_2. The cells were then treated with a 1:100 dilution of stock concentrations of drugs or ddH_2O as vehicle control in triplicate and left to incubate at 37°C in a humidified environment maintained at 95% O_2 and 5% CO_2 for 72 hours. After 72 hours 22µl of MTT was added to each well and cells were re-incubated for 2 hours. Following this all media was aspirated from each well and the cells were re-solubilised in 200µL MTT. Plates were placed on a plate shaker for 20 min prior to being read at 570 nm using a microplate reader (Molecular Devices, UK). IC_{50} values were calculated using Prism GraphPad 5 software.

MTT Assay of MCF-7 and TAM-MCF-7

Anti-tumour evaluation in MCF-7 wild type and tamoxifen-resistant MCF-7 (Tam-MCF-7) cell lines was performed by MTT assay. Compounds were prepared as 0.1–100 mM stock solutions dissolved in DMSO and stored at -20°C. Cells were seeded into 96-well microtitre plates at a density of 1×10^4 cells per well and allowed 24 h to adhere. Decimal compound dilutions were prepared in medium immediately prior to each assay (final concentration 0.1–100 μM). Experimental medium was phenol-red free RPMI +5% FCS for MCF7 wild type or RPMI +5% SFCS + 10^{-7} M 4-hydroxytamoxifen for TAM-MCF-7. Following 96 h compound exposure at 37°C, 5% CO_2, MTT reagent (Sigma Aldrich) was added to each well (final concentration 0.5 mg/mL). Incubation at 37°C for 4h allowed reduction of MTT by viable cells to an insoluble formazan product. MTT was removed and formazan solubilised by addition of 10% Triton X-100 in PBS. Absorbance was read on a Tecan Sunrise spectrophotometer at 540 nm as a measure of cell viability; thus inhibition relative to control was determined (IC_{50}).

7.2.4. Cell Morphology Studies

HL-60 Cell Line

Cells were seeded at 200,000/mL in a 6-well plate and treated with an appropriate volume of compound solution of particular concentration or with sterile ddH_20 for vehicle control (20 μL of drug or ddH_20 in 2 mL of cells). At 72 h, solutions were spun at 625 × g for 10 min. Old media was discarded and cell pellets resuspended in 2 mL fresh media. Pellets were then resuspended in 2 mL of fresh media. 200 μL of this was then cytocentrifuged (Shandon Cytospin) onto appropriate glass slides at 187 × g for 10 min. Slides were then air-dried for 10 min and cells stained using the RapiDiff kit according to manufacturer's instructions. Images were taken at a magnification of 40X using an Olympus 1X81 microscope coupled to a CCD DP71 camera.

HUH-7 Cell Line

HUH-7 cells were seeded in a 96-well plate at a density of 50,000 cells/mL (200 µL/well or 10,000cells/well) and incubated at 37 °C for 24 h to let them attach. The cells were then treated with 2 µL of the compound solution of particular concentration or with sterile ddH_20 for vehicle control. After 72 h incubation, the cells were brought to the microscopy area to take images in an Olympus 1X81 microscope coupled to a CCD DP71 camera using different magnification such as 60X, 40X and 20X.

Video of Cell Proliferation and Cell Death in HL-60 Cells

Cells were seeded at 5×10^5 cell/mL in two different flasks. One was treated with 3AA µM and the other with the same amount of ddH_2O (vehicle control). Images were collected on an Olympus IX81 microscope with a prior motorised stage, an Andor iXon CCD DP71 camera using IQ software in a humidified camber at 37 °C and 5% CO_2 with 40X oil objective at 1 min time intervals over night. Images were saved as multi TIFF files.

7.2.5. Flow Cytometry (Determination of DNA content)

HL-60 cells were seeded at 200,000 cells/mL in 6-well plates and treated with appropriate drug concentrations and ddH_2O as vehicle control (20 µL of drug or ddH_2O in 2mL of cells) for 48 h and 72 h time points. Cells were harvested at each time-point by centrifugation at 550 × g for 8 min. Media was drained and the pellets resuspended in 200 µL non-sterile PBS. The cells were then fixed by a drop-wise addition of 2 mL of ice-cold 70% EtOH/PBS while vortexing. Following overnight fixation at - 20 °C, 10 µL of non-sterile FBS was added to the cells and the solutions centrifuged at 550 x g for 10 min. The ethanol was then drained and the cell pellet resuspended in 0.4 mL PBS. 25 µL 10 mg/mL RNase A (which destroys RNA so that only DNA can be observed) was then added.75 µL 1 mg/mL propidium iodide (PI) (a DNA binding dye which fluoresces when activated and thus allows the amount of DNA present to be determined) was also added to each sample. Cells were vortexed and incubated in the dark at 37 °C for 30 min. The PI fluorescence was measured on a linear scale using a FACs Calibur flow cytometer (Becton Dickinson, San Jose, CA). Data collections (10,000 events per sample) were gated to exclude cell debris

and cell aggregates. The amount of fluorescence is proportional to the amount of DNA present and hence the population of cells in each phase of the cell cycle can be determined. Cells are gated as follows: M1= pre-G_1 (<2N DNA), M2= G_0/G_1 (2N DNA), M3= S (2N – 4N DNA), M4= G_2/M (4N DNA), M5= (Gn > 4N DNA). Apoptotic cells are hypoploid (< 2N DNA). Therefore, apoptosis was determined from the peak in M1. All data was recorded and analysed using the CellQuest software (Becton Dickinson, San Jose, CA).

7.2.6. Kinase Studies

Casein Kinase-1δ (CK1-δ) Assay Protocol

Human recombinant Casein Kinase-1δ was purchased from Millipore (Millipore Ibérica S.A.U.) Casein solution from bovine milk, 5%, was purchased from Sigma-Aldrich (St. Louis, MO). Kinase-Glo Luminescent Kinase Assay was obtained from Promega (Promega Biotech Ibérica, SL). ATP and all other reagents were from Sigma-Aldrich (St. Louis, MO). Assay buffer contained 50 mM HEPES, pH 7,5; 0,01% Brij-35; 10 mM Cl_2Mg; 1 mM EGTA and 0,01 % NaN_3.

The "Kinase-Glo" Kit from Promega was used to screen compounds for activity against CK-1δ Kinase-Glo assays were performed in assay buffer using black 96-well plates. In a typical assay, to 10 μL of test compound (dissolved in DMSO at 1 mM concentration and diluted in advance in assay buffer to the desired concentration), 10 μL (16 ng) of enzyme were added to each well followed by 20 μL of assay buffer containing 0,1 % casein as substrate and 4 μM ATP. The final DMSO concentration in the reaction mixture did not exceed 1%. After 60-min incubation at 30 °C the enzymatic reaction was stopped with 40 μl of Kinase-Glo reagent. Glow-type luminescence was recorded after 10 min using a FLUOstar Optima multimode reader (BMG Labtechnologies GmbH, Offenburg, Germany). The activity is proportional to the difference of the total and consumed ATP. The inhibitory activities were calculated on the basis of maximal activities measured in the absence of inhibitor. The IC_{50} was defined as the concentration of each compound that reduces a 50% the enzymatic activity with respect to that without inhibitors.

Inhibition of Glycogen Synthase-3 (GSK-3).

Human recombinant GSK-3β was purchased from Millipore (Millipore Iberica S.A.U.) The prephosphorylated polypeptide substrate was purchased from Millipore (Millipore Iberica S.A.U.). Kinase-Glo Luminescent Kinase Assay was obtained from Promega (Promega Biotech Ibérica, SL). ATP and all other reagents were from Sigma-Aldrich (St. Louis, MO). Assay buffer contained 50 mM HEPES (pH 7.5), 1 mM EDTA, 1 mM EGTA, and 15 mM magnesium acetate.

The method of Baki et al.[10] was followed for the inhibition of GSK-3β. Kinase-Glo assays were performed in assay buffer using black 96-well plates. In a typical assay, 10 μl (10 μM) of test compound (dissolved in dimethyl sulfoxide [DMSO] at 1 mM concentration and diluted in advance in assay buffer to the desired concentration) and 10 μL (20 ng) of enzyme were added to each well followed by 20 μl of assay buffer containing 50 μM substrate and 2 μM ATP. The final DMSO concentration in the reaction mixture did not exceed 1 %. After 30 min incubation at 30 °C the enzymatic reaction was stopped with 40 μL of Kinase-Glo reagent. Glow-type luminescence was recorded after 10 min using a FLUOstar Optima (BMG Labtechnologies GmbH, Offenburg, Germany) multimode reader. The activity is proportional to the difference of the total and consumed ATP. The inhibitory activities were calculated on the basis of maximal activities measured in the absence of inhibitor.

RAF-1/MEK-1 Kinase Assay

The test compound, reference compound or water (control) were preincubated for 5 min at room temperature with RAF-1 (2 ng) in a buffer containing 50 mM Hepes/NaOH (pH 7.4), 5 mM $MgCl_2$, 1 mM DTT, 40 μM Na_3VO_4 and 0.005% Tween 20. Thereafter, the reaction was initiated by adding 0.2 μM of inactive MEK-1, 100 nM of inactive ERK-2 and 15 μM ATP, and the mixture was incubated for 15 min at room temperature. For control basal measurements, RAF-1 was omitted from the reaction mixture. Following incubation, the reaction was stopped by adding 33 mM EDTA. The fluorescence acceptor (XL665-labeled anti-GST antibody) and the fluorescence donor (anti-phospho-ERK2 antibody labeled with europium cryptate) were then added.[11] After 180 min, the fluorescence transfer was measured at λ_{ex}=337 nm, λ_{em}=620 and λ_{em}=665 nm using a microplate reader (Rubystar,

BMG). The enzyme activity is determined by dividing the signal measured at 665 nm by that measured at 620 nm (ratio). The results were expressed as a percent inhibition of the control enzyme activity. The standard inhibitory reference compound was staurosporine, which was tested in each experiment at several concentrations to obtain an inhibition curve from which its IC_{50} value is calculated.

ERK 1/2 Kinase (Inhibitory Effect)

Confluent A431 cells (epidermoid carcinoma cell line) were suspended in HBSS buffer (Invitrogen) complemented with 0.1% BSA, then distributed in microplates at a density of 4.10^4 cells/well and preincubated for 5 min at 22 °C in the presence of either HBSS (basal and stimulated control) or the test compounds. Thereafter, 0.1 nM EGF is added and the mixture was incubated for 10 min at 22°C. For control basal measurements, EGF is omitted from the reaction mixture. Thereafter, cells were incubated during 15 min at room temperature with a lysis buffer. The fluorescence acceptor (alphalisa protein A coupled-beads) and the fluorescence donor (streptavidin coupled-beads) were then added. Following 120 min incubation at 22 °C, the signal is measured at λ_{ex} = 680 nm and λ_{em} = 500 and 600 nm using a microplate reader (EnVision, Perkin Elmer).[12] The results were expressed as a percent inhibition of the control response to 0.1 nM EGF. The standard reference inhibitor is AG1478, which is tested in each experiment at several concentrations to generate a concentration-response curve from which its IC_{50} value is calculated.

P38 MAPKinase Assay (Inhibitory Effect)

A431 (epidermoid carcinoma cell line) are grown to confluence for 48h at 37°C (CO_2 incubator). Then, cells are preincubated for 5 min at 22°C in the presence of either of HBSS (basal and stimulated control) or the test compounds. Thereafter, 1 nM TNFα is added and the mixture is incubated for 10 min at 22°C. For control basal measurements, TNFα is omitted from the reaction mixture. Thereafter, cells are lysed during 15 min at room temperature by adding the lyse buffer. To the cell lysat samples, the fluorescence acceptor (alphalisa protein A coupled-beads) is then added. Following 120 min incubation at 22°C, fluorescence donor (streptavidin coupled-beads) is added. Following 60 min incubation at 22°C, the signal is measured at λ_{ex} = 680 nm and λ_{em} = 500 and 600 nm using

a microplate reader (EnVision, Perkin Elmer).[13] The results are expressed as a percent inhibition of the control response to 1 nM TNFα. The standard reference inhibitor is SB202190, which is tested in each experiment at several concentrations to generate a concentration-response curve from which its IC_{50} value is calculated.

VEGFR Assay (Inhibitory Effect)

HUV-EC-C cells (Human Umbilical Vein Endothelial Cell Culture are primary cells used to measure angiogenesis activity) were seeded onto 96-well plate at 5×10^4 cells/well and allowed to grow overnight in standard growth media containing 76% F12K, 15% inactivated FCS, 5% inactivated human serum, 1% antibiotics, 1% ECGS, 1% glutamine and 1% heparin under standard culture conditions. Growth media was exchanged to HBSS buffer + 20 mM HEPES (Invitrogen) and cells are allowed to equilibrate for 30 min at 28°C before the start of the experiment. Plates were placed onto the system and measurements were made at a temperature of 28°C. HBSS (basal and stimulated controls), reference inhibitor (IC_{50} determination) or the test compounds were preincubated for 20 min before the addition of HBSS (basal control) or the reference stimulant VEGF at 0.1 nM (EC_{80}).[14] Impedance measurements are monitored for 10 min. The standard reference inhibitor was VEGF receptor tyrosine kinase inhibitor II, which is tested in each experiment at several concentrations to generate a concentration-response curve from which its IC_{50} value was calculated.

7.2.7. LanthaScreen® Eu Kinase Binding Assay for B-RAF

Prior to the set up of the experiment the reagents used in the assay were prepared. 100X concentrations of control (sorafenib) and test compounds were prepared. Sorafenib solution was made in DMSO, **79a** and **79d** in EtOH and the rest of the compound solutions were made in ddH_2O. Intermediate dilutions were prepared (3X) by transferring fixed volumes of 100X compound into fixed volumes of kinase buffer A (DMSO/EtOH concentration ratio was 3%). Additionally, tracer solution (Kinase Tracer 178) was prepared at 60 nM (3X final) in kinase buffer A. Finally, 15 nM of B-RAF kinase and 6 nM antibody (3X final) kinase/antibody solution in kinase buffer A (on ice) was prepared. Prior to kinase/antibody solution preparation, antibody was centrifuged at 10,000 × g for

10 min and solution aspirated from the top of the vial to minimise the potential effect of spurious donor emission spikes in assay. Once all the reagents were prepared 5 µL of each concentration of serial diluted compound (test or control) was added in a 384-well plate in triplicate wells. Then, 5 µL of kinase/antibody solution followed by 5 µL of tracer solution was then added to each well. Each plate also contains a blank, a vehicle control and a positive control. The blank contains kinase buffer A only, the vehicle control contains the solvent used for the drug solutions preparation which was DMSO, EtOH and ddH_2O and the positive control contains varying concentrations of sorafenib. The plate was incubated at room temperature for 60-120 min and was read in the plate reader (BMG LABTECH PHERAstar) where the general instrument settings for excitation wavelength was 340 nm, the kinase trace emission was at 665 nm and the LanthaScreen® Eu-anti-Tag antibody emission was at 615 nm.

The data analysis consists of division of the acceptor or tracer emission (665 nm) by the antibody or donor emission (615 nm) to calculate the emission ratio. Then, the results were plotted in a graph were test/control compound concentration (M) in the x axis versus emission ratio in the y axis was performed. Where applicable, data was fitted using a sigmoidal dose-response curve (variable slope) using Prism GraphPad 5.0 and the IC_{50} was determined.

7.2.8. Western Blot Preparation

Protein Extraction

Cells were trypsinised as described previously and seeded at a density of 1×10^5cells/mL for each experiment in P90 culture plates and allowed to adhere overnight. Following treatment, cells were returned to the incubator until required for protein extraction. Culture medium was removed from each P90 and transferred to a 15 mL falcon tube. Cells were detached by use of a cell scraper in ice cold PBS with sodium fluoride and sodium orthovanadate solution.

The detached cells were subsequently transferred into the 15 mL Falcon tube containing the medium. This was followed by centrifugation of the falcon tubes for 5 min at 600 × g at 4°C and the resulting supernatant was removed. The pellets were re-suspended in 70 µl RIPA buffer with protease inhibitors. The suspensions was then transferred into a 1.5 mL

eppendorf tube, carefully pipetted up and down, snap frozen into liquid nitrogen and thawed on ice. The eppendorfs were then spun by centrifugation for 20 min at 4 °C, 3300 × g. The protein lysate was transferred to a fresh eppendorf; snap frozen and stored at -80 °C until required for use.

Protein Quantification

Protein quantification was measured using BCA Protein Assay Reagent (Pierce, Rockford, USA). Absorbance of protein was measured at 570 nm using a Biotrack II Visible Plate Reader (Amersham, UK). Protein concentration for each sample was determined by interpolation from a standard curve generated using serial dilutions of known concentrations of bovine serum albumin (BSA) (Sigma-Aldrich, Dorset, UK).

Western blot Analysis

Western loading 2X dye containing 10% β-mercaptoethanol was added to the protein sample and the samples were denatured at 95% for 5 min. Equal amount, 30 μg of protein was separated on an 8-12% SDS-polyacrylamide gel by electrophoresis. The concentration of gel used was dependent on the molecular weight of the proteins to be analysed. Gels were immersed in 1X running buffer and run at 100 V for 90 min. The gels were then transferred onto nitrocellulose membrane placed between Whatman paper and sponges. The gels were transferred at 100 V for 90 min in 1X transfer buffer. The membrane was then blocked in 5% marvel/1X PBS-Tween for 1 hour at room temperature.

Protein Detection

The membrane was incubated with the appropriate primary antibody diluted in 5% marvel in 1X PBS-Tween and placed on a roller overnight at 4°C in a cold room. The next day the membrane was washed 3 times for 10 min in 1X PBS-Tween and incubated in the appropriate secondary antibody for 1 h at 4°C in a cold room. This was followed by 3X 10 min washes with 1X PBS-Tween before the membranes were subjected to ECL chemiluminescence detection (Amersham). The membrane was soaked in ECL for 5 min

at room temperature, then placed between two acetated sheets in a cassette and exposed to X-ray film (Analab) and developed using ILFORD Multigrade Paper Developer and Rapid Fixer.

7.3. Computational Chemistry Procedures

7.3.1. Proton Affinity and Molecular Electrostatic Potential

Geometries of both neutral and *mono*-protonated systems were fully optimised in gas phase at the B3LYP theoretical level with the 6-31+G(d,p) basis set as implemented in the Gaussian09 program.[15] Harmonic frequency calculations verified the nature of the stationary points as minima (all real frequencies). The proton affinity was calculated as the difference in total energy of the protonated and unprotonated states, and was not corrected for the basis set superposition error. The molecular electrostatic potential was calculated at the same level of theory and basis set (at 512,000 points) using the cubegen utility in Gaussian, and was mapped onto an isosurface with constant electron density of 0.02 a.u. The range of potential values displayed was between 1.0 e^{-2} and 5.0 e^{-2}.

7.3.2. Isomerism Study Protocol

Geometries were fully optimized at the B3LYP theoretical level with the 6-31+G** (d, p) basis set as implemented in the Gaussian09 program.[15] Harmonic frequency calculations verified the nature of the stationary points as minima (all real frequencies). NMR calculations were performed using the Gauge-Invariant Atomic Orbital (GIAO) protocol at the same level of theory. Chemical shifts were taken with reference to the tetramethylsilane standard. Where chemically equivalent signals gave different GIAO values these were reported as obtained, though this occurs as GIAO is a single point calculation on the optimised structure, which considers chemically equivalent signals as existing in distinct environments.

7.3.3. Molecular Docking Protocol

A set of conformers was generated for each ligand using the conformational search tool in MOE. The LowModeMD[16] method was used with an iteration limit of 100 and an energy window of 12 kcal mol^{-1} above the minimum, to ensure a large sample of conformers. These databases of conformers were then docked into the receptor in question using MOE. Ligands were placed using the triangle matcher method, rotatable bonds were left free to rotate and the LondondG method was used for rescoring. A refinement of poses (induced-fit) was carried out using the forcefield method with side-chains free to move and a cutoff distance of 12 Å set for interactions to be considered.

7.3.4. Minimisation of MEK-1 Structure in AMBER 12.0

Both ATP and the ligand were prepared for minimisation using the Generalised Amber Forcefield (GAFF) in antechamber, to generate topology and atom-type information. Then, the Leap tool in AMBER was used to solvate the complex in a 15 Å radius cube of TIP3P explicit water molecules with periodic boundary conditions, and neutralise in the system by adding Na^+/Cl^- ions (the amber99SB forcefield was used for the protein and GAFF was used for the ligand and ATP).

Complex minimisation was carried out using the amber99SB forcefield and the steepest-descent method of minimisation. The first step of minimisation fixed the complex with a force constant of 10.0 and relaxed the water molecules to remove close contacts. Subsequently, the main chain of the newly modelled loop region was fixed, while its amino acid side chains were minimised along with the water molecules. The same restriction was applied on the rest of the system. Next, the side chains of the entire complex were minimised along with the water molecules, while the main chain was fixed (force constant 10.0). The entire system was then minimised at reducing restraints on the main chain with force constants 5.0, 1.0 and 0.1, before the system was fully minimised with no restraints imposed. The resulting minimised structure was taken as the starting point for docking calculations.

7.3.5. Minimisation of Docking Poses and MMPBSA Analysis

The topology and coordinate files previously generated for the ligands and ATP were also applied in this study. The Leap tool in AMBER was used to solvate the complex in a 15 Å radius cube of TIP3P explicit water molecules with periodic boundary conditions, and neutralise in the system by adding Na^+/Cl^- ions (the amber99SB forcefield was used for the protein and GAFF was used for the ligand and ATP).

Complex minimisation was carried out using the amber99SB forcefield and the steepest-descent method of minimisation. A non-bonded cutoff of 10.0 Å was applied for each step of minimisation. The receptor and ligand were first fixed with a force constant of 5.0, the newly introduced water molecules and ions being relaxed. Then, the main chain of the protein and the ligand were fixed with a force constant of 5.0, while the side chains and water molecules were minimised. Finally, the entire system was minimised with no restraints in place.

The MMPBSA analysis was performed within AMBER12.0 program for each complex resulting from the minimisation of the docking poses. The solvated binding free energy for the complex was calculated by: $\Delta G°_{B,S} = \Delta G°_{B,V} + \Delta G°_{S,Cmplx} - (\Delta G°_{S,Lig} + \Delta G°_{S,Rec})$, where B = binding, S = solvation, Cmplx = complex, Lig = ligand and Rec = receptor.

7.4. References

(1) Vogel, A. I.; Tatchell, A. R.; Furnis, B. S.; Hannaford, A. J.; Smith, P. W. G. *Vogel's Textbook of Practical Organic Chemistry*; 5th ed.; Pearson Education, 1996.

(2) Gottlieb, H. E.; Kotlyar, V.; Nudelman, A. *J. Org. Chem.* **1997**, *62*, 7512-7515.

(3) Hunsberger, I. M.; Amstutz , E. D. *J. Am.Chem. Soc.* **1949**, *71*, 2635.

(4) Gattermann, L.; Rudt, H.; *Chemische Berichte* **1894**, *27*, 2296.

(5) Diez-Cecilia, E.; Kelly, B.; Rozas, I. *Tetrahedron Lett.* **2011**, *52*, 6702-6704.

(6) Albert, A. *J. Chem. Soc.* **1948**, 1225-1230.

(7) Wang, Z.; Fu, H.; Jiang, Y.; Zhao, Y. *Synlett* **2008**, *16*, 2540-2546.

(8) Dardonville, C.; Goya, P.; Rozas, I.; Alsasua, A.; Martin, M. I.; Borrego, M. J. *Bioorg. Med. Chem.* **2000**, *8*, 1567-1577.

(9) O'Donovan, D. H.; Rozas, I. *Tetrahedron Lett.* **2011**, *52*, 4117-4119.

(10) Baki, A. B., A.; Molnar, L.; Szendrei, G.; Keseru, G. M. *Assay Drug Dev. Technol.* **2007**, *5*, 75-83.

(11) Force, T.; Bonventre, J. V.; Heidecker, G.; Rapp, U.; Avruch, J.; Kyriakis, J. M. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 1270-1274.

(12) Zhang, Y. G.; Q. Du; Fang, W. G.; Jin, M. L.; Tian, X. X. *Int. J. Oncol.* **2008**, *33*, 595-602.

(13) Li, W.; Nadelman, C.; Henry, G.; Fan, J.; Muellenhoff, M.; Medina, E.; Gratch, N. S.; Chen, M.; Han, J.; Woodley, D. *J. Invest. Dermatol.* **2001**, *117*, 1601-1611.

(14) Furet, P.; Bold, G.; Hofmann, F.; Manley, P.; Meyer, T.; Altmann, K.-H. *Bioorg. Med. Chem. Lett* **2003**, *13*, 2967-2971.

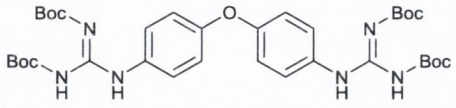
(15) Gaussian 09, R. A., M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian, Inc., Wallingford CT, 2009.

(16) Labute, P. *J. Chem. Inf. Model.* **2010**, *50*, 792-800.

Appendix

Characterisation of 4,4'-bis-guanidine-like derivatives previously synthesised in Rozas' group.

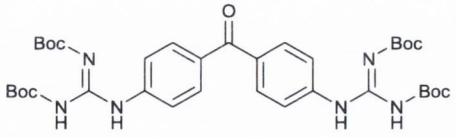
4,4'-*Bis*-[2,3-di(*tert*-butoxycarbonyl)guanidino]diphenylether (39a)



Following method A, using the commercial 4,4'-diaminodiphenyl ether (601 mg, 3.0 mmol) as starting material, after purification with silica gel chromatography, eluting with a gradient of hexane:EtOAc, **39a** was obtained as a white solid (1622 mg, 79%). **Mp**: decomp. over 270 °C. (lit. decomp. over 300 °C).[1]

δ_H (400 MHz, $CDCl_3$): 1.50 (s, 18H, $2(CH_3)_3$), 1.54 (s, 18H, $2(CH_3)_3$), 6.96 (d, 4H, J 9.0, Ar), 7.54 (d, 4H, J 9.0, Ar), 10.28 (bs, 2H, 2NH), 11.65 (bs, 2H, 2NH).

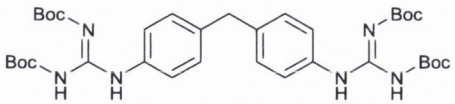
4,4'-*Bis*-[2,3-di(*tert*-butoxycarbonyl)guanidino]benzophenone (39b)



Following method A, using the commercial 3,4'-diaminobenzophenone (230 mg, 1.08 mmol) as starting material, after purification with silica gel chromatography, eluting with a gradient of hexane:EtOAc, **39b** was obtained as colourless foam (391 mg, 52%). **Mp**: 140-142 °C. (lit. 142-143 °C).[2]

δ_H (400 MHz, $CDCl_3$): 1.52 (s, 18H, $2(CH_3)_3$), 1.55 (s, 18H, $2(CH_3)_3$), 7.76 (d, 4H, J 8.8, Ar), 7.79 (d, 4H, J 8.8, Ar), 10.6 (bs, 1H, NH), 11.63 (bs, 1H, NH).

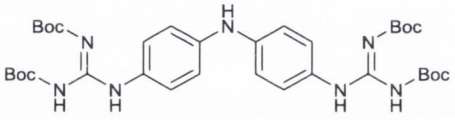
4,4'-*Bis*-[2,3-di(*tert*-butoxycarbonyl)guanidino]benzylbenzene (39c)



Following method A, using the commercial 4-(4-aminobenzyl)aniline (198 mg, 1.0 mmol) as starting material, after purification with silica gel chromatography, eluting with a gradient of hexane:EtOAc, **39c** was obtained as a colourless foam (546 mg, 80%). **Mp**: 120-122 °C. (lit. 118-122 °C).[2]

δ_H (400 MHz, $CDCl_3$): 1.49 (s, 18H, $2(CH_3)_3$), 1.52 (s, 18H, $2(CH_3)_3$), 3.9 (s, 2H, CH_2), 7.11 (d, 4H, J 8.5, Ar), 7.49 (d, 4H, J 8.5, Ar), 10.26 (bs, 1H, NH), 11.3 (bs, 1H, NH).

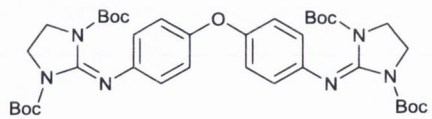
4,4'-*Bis*-[2,3-di(*tert*-butoxycarbonyl)guanidino]-*N*-phenylaniline (39d)



Following method A, using the commercial 4-(4-aminoanilino)aniline (199 mg, 1.0 mmol) as starting material and after purification with silica gel chromatography, eluting with a gradient of hexane:EtOAc, **39d** was obtained as a off-white foam (451.0 mg, 66%). **Mp**: 133-135 °C. (lit. 130-135 °C).[2]

δ_H (400 MHz, $CDCl_3$): 1.45 (s, 18H, $2(CH_3)_3$), 1.53 (s, 18H, $2(CH_3)_3$), 6.89 (d, 4H, J 8.6, Ar), 7.32 (d, 4H, J 8.6, Ar), 10.12 (bs, 1H, NH), 11.6 (bs, 1H, NH).

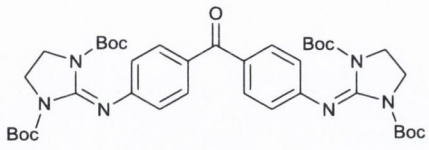
4,4'-*Bis*-[1,3-di(*tert*-butoxycarbonyl)-2-imidazolidinylimino]diphenylether (40a)



Following method A, using the commercial 4,4'-diaminodiphenyl ether (601 mg, 3.0 mmol) as starting material, after purification with neutral alumina chromatography, eluting with a gradient of hexane:EtOAc, **40a** was obtained as a colourless foam (1593 mg, 72%). **Mp**: 170-172 °C. (lit. 168-170 °C).[1]

δ_H (400 MHz, $CDCl_3$): 1.37 (s, 36H, 4$(CH_3)_3$), 3.84 (s, 8H, 4CH_2), 6.86 (d, 4H, J 7.4, Ar), 6.96 (d, 4H, J 7.4, Ar).

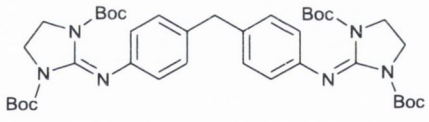
4,4'-*Bis*-[1,3-di(*tert*-butoxycarbonyl)-2-imidazolidinylimino]benzophenone (40b)



Following method A, using the commercial 3,4'-diaminobenzophenone (159 mg, 0.75 mmol) as starting material, after purification with neutral alumina chromatography, eluting with a gradient of hexane:EtOAc, **40b** was obtained as a colourless solid (224.5 mg, 40%). **Mp**: 104-106 °C. (lit. 106-108 °C).[2]

δ_H (400 MHz, $CDCl_3$): 1.27 (s, 36H, 4$(CH_3)_3$), 3.78 (s, 8H, 4 CH_2), 6.93 (d, 4H, J 8.4, Ar), 7.60 (d, 4H, J 8.4, Ar).

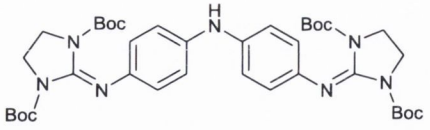
4,4'-*Bis*-[1,3-di(*tert*-butoxycarbonyl)-2-imidazolidinylimino]benzylbenzene (40c)



Following method A, using the commercial 4-(4-aminobenzyl)aniline (102 mg, 0.51 mmol) as starting material, after purification with neutral alumina chromatography, eluting with a gradient of hexane:EtOAc, **40c** was obtained as a colourless solid (294.9 mg, 78%). **Mp**: 88-90 °C. (lit. 90-93 °C).[2]

δ_H (400 MHz, $CDCl_3$): 1.31 (s, 36H, 4$(CH_3)_3$), 3.81 (bs, 10H, 5 CH_2), 6.88 (d, 4H, J 8.4, Ar), 7.00 (d, 4H, J 8.4, Ar).

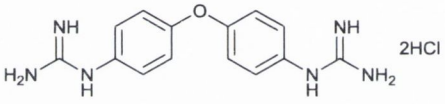
4,4'-*Bis*-[1,3-di(*tert*-butoxycarbonyl)-*N*-2-imidazolidinylimino]-*N*-phenylaniline (40d)



Following method A, using the commercial 4-(4-aminoanilino)aniline (199 mg, 1.0 mmol) as starting material and after purification with neutral alumina chromatography, eluting with a gradient of hexane:EtOAc, **40d** was obtained as a brown solid (448.6 mg, 61%). **Mp**: 165-167 °C. (lit. 165-169 °C).[2]

δ_H (400 MHz, $CDCl_3$): 1.32 (s, 36H, $4(CH_3)_3$), 3.79 (s, 8H, 4 CH_2), 6.84 (d, 4H, J 8.4, Ar), 6.88 (d, 4H, J 8.4, Ar).

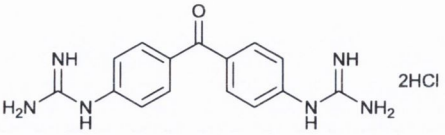
4,4'-Diguanidinediphenyl ether dihydrochloride (49a)



Following method E, using **39a** (530 mg, 0.77 mmol) as starting material, the dihydrochloride salt **49a** was obtained as a white solid (251 mg, 91%). **Mp**: decomp. over 200 °C.(lit. decomp. over 230 °C).[1]

δ_H (400 MHz, $CDCl_3$): 7.16 (d, 4H, J 8.5, Ar), 7.34 (d, 4H, J 8.5, Ar).

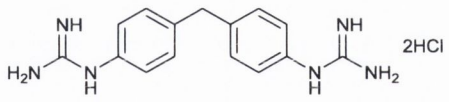
4,4'-Diguanidinediphenylketone dihydrochloride (49b)



Following method E, using **39b** (255 mg, 0.37 mmol) as starting material, the dihydrochloride salt **49b** was obtained as a white solid (110.5 mg, 82%). **Mp**: 150-152 °C.(lit. 149-152 °C).[2]

δ_H (400 MHz, D_2O): 7.25 (d, 4H, J 8.3, Ar), 7.64 (d, 4H, J 8.3, Ar).

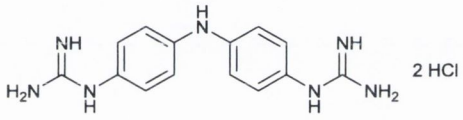
4,4'-Diguanidinediphenylmethane dihydrochloride (49c)



Following method E, using **39c** (176 mg, 0.26 mmol) as starting material, the dihydrochloride salt **49c** was obtained as a white solid (69.4 mg, 76%). **Mp**: 199-200 °C.(lit. 199-202 °C).[2]

δ_H (400 MHz, D_2O): 3.88 (s, 2H, CH_2), 7.07 (d, 4H, J 8.4, Ar), 7.21 (d, 4H, J 8.4, Ar).

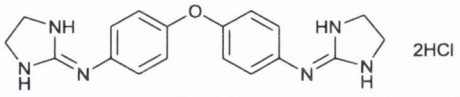
4,4'-Diguanidinediphenylamine dihydrochloride (49d)



Following method E, using **39d** (260 mg, 0.38 mmol) as starting material, the dihydrochloride salt **49d** was obtained as a grey solid (94 mg, 70%). **Mp**: decomp. over 180 °C.(lit. decomp over 170 °C).[2]

δ_H (400 MHz, D_2O): 6.96 (m, 8H, Ar).

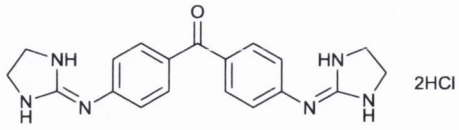
4,4'-Di(2-imidazolidinylimino)diphenylether dihydrochloride (50a)



Following method E, using **40a** (500 mg, 0.68 mmol) as starting material, the dihydrochloride salt **50a** was obtained as a white solid (249.4 mg, 90%). **Mp**: decomp. over 200 °C.(lit. decomp. over 220 °C).[1]

δ_H (400 MHz, D_2O): 3.74 (s, 8H, 4 CH_2), 7.13 (d, 4H, J 8.5, Ar), 7.30 (d, 4H, J 8.5, Ar).

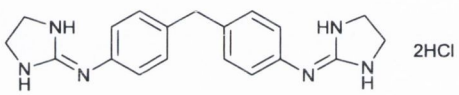
4,4'-Di(2-imidazolidinylimino)benzophenone dihydrochloride (50b)



Following method E, using **40b** (130 mg, 0.174 mmol) as starting material, the dihydrochloride salt **50b** was obtained as a white solid (58 mg, 78%). **Mp**: decomp. over 162 °C.(lit. decomp. over 159 °C).[2]

δ_H (400 MHz, D_2O): 3.62 (s, 8H, 4 CH_2), 7.20 (d, 4H, J 8.6, Ar), 7.63 (d, 4H, J 8.6, Ar).

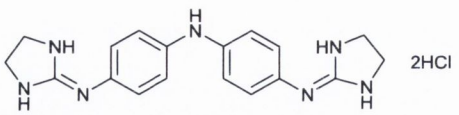
4,4'-Di(2-imidazolidinylimino)diphenylmethane dihydrochloride (50c)



Following method E, using **40c** (150 mg, 0.2 mmol) as starting material, the dihydrochloride salt **50c** was obtained as a brown solid (66.4 mg, 78%). **Mp**: 143-145 °C.(lit. 140-145 °C).[2]

δ_H (400 MHz, D_2O): 3.46 (s, 8H, 4 CH_2), 3.71 (s, 2H, CH_2), 6.92 (d, 4H, J 8.0, Ar), 7.07 (d, 4H, J 8.0, Ar).

4,4'-Di(2-imidazolidinylimino)diphenylamine dihydrochloride (50d)



Following method E, using **40d** (180 mg, 0.25 mmol) as starting material, the dihydrochloride salt **50d** was obtained as a grey solid (99.6mg, 68%). **Mp**: 166-168 °C.(lit. 165-170 °C).[2]

δ_H (400 MHz, D_2O): 3.62 (s, 8H, 4 CH_2), 7.07 (d, 4H, J 8.9, Ar), 7.11 (d, 4H, J 8.9, Ar).

References

(1) Rodriguez, F.; Rozas, I.; Ortega, J. E.; Meana, J. J.; Callado, L. F. *J. Med. Chem.* **2007**, *50*, 4516-4527.

(2) Dardonville, C.; Goya, P.; Rozas, I.; Alsasua, A.; Martin, M. I.; Borrego, M. J. *Bioorg. Med. Chem.* **2000**, *8*, 1567-1577.