

Guanidine based derivatives as potential anti-infective agents: Tackling protozoal diseases and *Mycobacterium tuberculosis*



**A thesis presented to the University of Dublin for the degree of
Doctor of Philosophy**

by

Adeyemi Rahman, B.Sc.

Under the supervision of Prof. Isabel Rozas

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.

.....

Adeyemi Rahman

September 2021

Trinity College Dublin

Acknowledgments

Firstly, I'd like to thank my supervisor Prof. Isabel Rozas. She has been an amazing mentor and the support and encouragement she has given me within the past four years has molded me into the scientist and person I am today. I would also like to thank the current (Helene and Nikolina) and past (Aaron, Jagdeep and Filippo) Rozas group members who have helped me along the way. My Ph.D. experience would have not been the same without them. I am very grateful to "extended members" of the Rozas group Dr. Cristina Trujillo, an amazing computational chemist, who taught me everything I know about computational modelling and Prof. David Grayson for his expert advice on organic synthesis. I am immensely thankful to Prof. Stephen Gordon (School of Veterinary Medicine, University College Dublin) who has been my second supervisor throughout my Ph.D. His advice, support and access to his lab have been a massive help which I will always be grateful for. Many thanks to Gordon group members (Jose-Marie, Yao, Viktor, Sophie, Joe and Damien) who taught me all the microbiology, molecular biology and bioinformatics techniques I used in this work.

I would also like to thank our collaborators overseas who assisted with some biochemical and biological experiments presented in this work; Dr. Apoorva Bhatt and Dr. Sarah Batt from the University of Birmingham, Prof. Lydia Taberner from the University of Manchester and Dr. Marcel Kaiser from the Department of Medical Parasitology and Infection Biology in the Swiss Tropical and Public Health Institute (Basel, Switzerland). Additionally, I would like to express my gratitude to the TCD School of Chemistry staff especially Dr. John O'Brien and Dr. Manuel Ruether for their assistance with NMR and Dr. Gary Hessman for mass spectroscopy.

Lastly, I would like to thank my parents, Adetola and Babatunde Rahman, who have always encouraged me to reach for the stars and chase my dreams. I am also grateful for my siblings, in particular, Yinka Martin who has been an amazing role model to me. A special thank you to my partner, Elizabeth Sainsbury for all her love and support these past couple of years.

Dedicated to my parents, Adetola and Babatunde Rahman

Abstract

Infectious diseases caused by microorganisms of the genus *Trypanosoma* (Human African Trypanomiasis-HAT), *Leishmania* (Leishmaniasis), *Plasmodium* (Malaria) and *Mycobacterium* (Tuberculosis) cause some of the deadliest human infections in the world. According to the WHO, in 2019, there were 229 million cases of malaria and 10 million cases of TB with death rates of 409,000 and 1.3 million, respectively. In addition, in sub-Saharan Africa, HAT has an overwhelming risk population of 65 million people with 700,000 - 1 million annual cases of leishmaniasis. Still no definitive cure has been found for these infectious diseases due to factors such as poor patient compliance, severe side effects and multi-drug resistance. Hence, there is a need for the development of new drug therapies.

Compounds that bind into the DNA minor groove have proven to be very promising for the treatment of parasitic diseases. Accordingly, DNA minor groove binders (MGBs) such as pentamidine and distamycin-A have been used to treat malaria and HAT. However, drug resistance and the associated side effects present major drawbacks for the use of these agents. During the last 20 years, Prof. Rozas' group has prepared a variety of guanidine-like derivatives that are strong MGBs. More recently, very good antimalarial activity was obtained from a family aminoalkyl derivatives of di-aryl mono guanidines. However, it has been hypothesised that two possible mechanisms of action could be behind this activity (i.e. as MGBs or as inhibitors of the parasite's dihydrofolate reductase—DHFR). Hence, we investigated this further and prepared new analogues which displayed improved DNA binding and anti-protozoal activity.

Secondly, based on a series of piperazine and piperidine-linked di-aryl amidines that were recently shown to be active against *Mycobacterium tuberculosis*, structurally similar MGBs previously synthesised by the Rozas group were tested and showed very good anti-mycobacterial activity. As the DNA binding affinity of the active compounds from this series did not correlate with the anti-mycobacterial activity, various computational and biological studies were performed to identify Enoyl ACP reductase (InhA) as a possible target. Furthermore, using a rational design approach novel guanidine-based derivatives

were produced that showed better anti-mycobacterial activity than previous in-house compounds and the known first-line anti-tubercular agent, rifampicin.

Abbreviations

BCG:	Bacillus Calmette–Guérin
Boc:	<i>tert</i> -Butyloxycarbonyl
Boc ₂ O:	Di- <i>tert</i> -butyldicarbonate
CHCl ₃ :	Chloroform
CH ₂ Cl ₂ :	Dichloromethane
CD:	Circular Dichroism
DFT:	Density Functional Theory
DHFR:	Dihydrofolate Reductase
DIPEA:	<i>N,N</i> -Diisopropylethylamine
DMSO:	Dimethylsulfoxide
DNA:	Deoxyribonucleic acid
DprE:	Decaprenylphosphoryl- <i>D</i> -ribose 20-epimerase
ECIT:	Etoxycarbonyl Isothiocyanate
EDC.HCl:	1-Ethyl-3-(3'-dimethylaminopropyl)carbodiimide, hydrochloride
Et ₂ O:	Diethyl Ether
EtOH:	Ethanol
EtOAc:	Ethyl Acetate
Gua:	Guanidine
FAME:	Fatty Acid Methyl Ester
FDA:	Food and Drug Administration
HAT:	Human African Trypanosomiasis
HB:	Hydrogen Bond

HCl:	Hydrochloric acid
Hex:	Hexane
HgCl ₂ :	Mercury (II) chloride
HPLC:	High-Pressure Liquid Chromatography
HRMS:	High-Resolution Mass Spectrometry
HSQC:	Heteronuclear Single Quantum Coherence Spectroscopy
IC ₅₀ :	Half Maximal Inhibitory Concentration
ICD:	Induced CD signal
INDEL:	Insertion or Deletion
Imi:	Imidazoline
InhA:	Enoyl-acyl carrier reductase
J:	Coupling Constant
L. don:	<i>Leishmania donovani</i>
LBDD:	Ligand Based Drug Design
LBPS:	Ligand Based Pharmacophore Screen
MAME:	Mycolic Acid Methyl Ester
MDR:	Multi-Drug Resistance
MeCN:	Acetonitrile
MeOH:	Methanol
MGB:	Minor Groove Binder
MIC:	Minimum Inhibitory Concentration
mmol:	Millimole
MOA:	Mechanism of Action

mp:	Melting Point
M. bovis:	Mycobacterium Bovis
M. smeg:	Mycobacterium Smegmatis
NaH:	Sodium hydride
NEt ₃ :	Triethylamine
NMR:	Nuclear Magnetic Resonance
OD:	Optical Density
<i>P. f.</i> :	<i>Plasmodium falciparum</i>
P/D:	Phosphate to Drug ratio
PD:	Piperidine
PZ:	Piperazine
REMA:	Resazurin Based Microtiter Plate Assay
rt:	Room Temperature
RT-qPCR:	Real Time Quantitative Polymerase Chain Reaction
SAR:	Structure–Activity Relationship
SBDD:	Structure Based Drug Design
SNP:	Single Nucleotide Polymorphism
SI:	Selectivity Index
stDNA:	Salmon Testes DNA
THF:	Tetrahydrofuran
TFA:	Trifluoroacetic acid
TLC:	Thin Layer Chromatography
T _M :	Melting Temperature

T. b. r: *Trypanosome brucei rhodesiense*

WHO: World Health Organisation

WGS: Whole-Genome Sequencing

XDR: Extensively-Drug Resistance

Table of Contents

Chapter 1 – Introduction

1.1. Overview of infectious diseases	1
1.1.2. Statistics	1
1.1.3. Disease pathogenesis and diagnosis	5
1.1.3.1. Malaria	5
1.1.3.2. Human African Trypanosomiasis	8
1.1.3.3. Leishmaniasis	11
1.1.3.4. Tuberculosis	13
1.1.4. Standard of care and current treatments	14
1.1.4.1. Malaria	14
1.1.4.2. HAT and Leishmaniasis	17
1.1.4.3. Tuberculosis	18
1.1.5. Drug resistance	19
1.1.6. Prospects and drug pipeline	22
1.2. Therapeutic targets	23
1.2.1. Overview of DNA	23
1.2.2. DNA targeting agents	27
1.2.3. DNA minor groove binders	31
1.2.4. Characterisation of DNA binding	37
1.2.5. Elucidation of mechanism of action – emerging drug discovery technologies	40
1.3. Previous work that constitutes the basis of this research	45
1.3.1. Guanidine based minor groove binders as anti-protozoal agents	45
1.3.2. Guanidine based derivatives as anti-mycobacterial agents	47
1.4. References	50

Chapter 2 - Overview

2. Objective overview	56
2.1. DNA Minor Groove Binders as potential anti-protozoal agents	57
2.1.1. Computational aims	57
2.1.2. Synthetic aims	60

2.1.3.	Biophysical aims	61
2.1.4.	Biological aims.....	61
2.2.	Guanidine based derivatives as potential anti-tubercular agents	62
2.2.1.	Preliminary investigation	62
2.2.2.	Computational aims.....	63
2.2.3.	Synthetic aims.....	64
2.2.4.	Biochemical and Biological aims	64
2.3.	References	65

Chapter 3 - Results and Discussion: Guanidine based minor groove binders as antiprotozoal agents

3.1.	Computational studies	66
3.1.1.	Computational Methods.....	68
3.1.2.	Molecular docking of compounds from previous Families A, B, C and D in a model of the DNA minor groove	69
3.1.3.	Molecular docking of previous Family E in the DNA minor groove and DHFR	76
3.1.4.	Molecular docking of novel Family F in the DNA minor groove and DHFR (Family F)....	88
3.2.	Synthesis.....	98
3.2.1.	Synthesis of novel amino-alkyl derivatives of di-aromatic bis- guanidines	98
3.2.2.	Synthesis of mono-guanidylated di-aromatic systems	99
3.2.3.	Synthesis of mono-Boc-protected diamines.....	101
3.2.4.	Synthesis of <i>N</i> -substituted guanidines.....	102
3.2.5.	Trans-carbamylation of substituted guanidines	113
3.2.6.	Deprotection of Boc protected substituted guanidine intermediates	117
3.3.	Biophysical studies	120
3.3.1	UV thermal denaturation of Family F	120
3.3.2.	Circular Dichroism studies of Family F.....	127
3.3.3.	UV-Vis titrations of compounds 37c and 37d	133
3.4.	Biological studies.....	136
3.4.1.	Results obtained against the <i>Trypanosoma brucei rhodesiense</i> STIB 900 cell line.....	137
3.4.2.	Results obtained against the <i>Plasmodium falciparum</i> NF54 cell line.....	139
3.4.3.	Results obtained against the <i>Leishmania donovani</i> axenic amastigotes.....	142
3.5.	Conclusions.....	145
3.6.	References	147

Chapter 4 - Results and Discussion: Guanidine based derivatives as anti-mycobacterial agents

4.1. Preliminary investigations	149
4.1.1. Resazurin based microtiter plate assay (REMA) analysis of previously made compounds against mycobacterial strains	150
4.1.2. Analysis of mycolic acid biosynthesis through TLC FAMES and MAMES assays	156
4.1.3. Biochemical evaluation of piperidine/piperazine derivatives against DprE1	159
4.1.4. Time-kill assay	161
4.2. Computer-aided investigations	165
4.2.1. Ligand-based pharmacophore screen	166
4.2.2. Structure-based molecular docking in InhA	169
4.3. Whole Genome Sequencing	176
4.3.1 Resistance generation	176
4.3.1.1. Liquid media continual passaging approach	176
4.3.1.2. Solid media single-shot approach	180
4.3.2. Characterisation of resistant mutants through DNA sequencing	183
4.3.3. InhA expression analysis	190
4.3.4. Evaluation of cross-resistance	192
4.4. Design and Molecular Modelling	195
4.4.1. Molecular docking studies of novel Families G and H in InhA	197
4.4.2. Molecular docking studies of novel Families G and H in DNA	203
4.5. Synthesis	206
4.5.1. Synthesis of previously reported piperidine/piperazine containing di-aryl guanidines/2-aminoimidazolines	207
4.5.2. Synthesis of novel families of guanidine based anti-mycobacterial agents	208
4.6. Biological evaluation	213
4.6.1. REMA analysis of Family G and H against <i>Mycobacterium smegmatis</i>	213
4.6.2. DNA binding of Family G and H	218
4.6.3. Time-kill assay	220
4.6.4. FAMES and MAMES TLC assays of <i>M. smegmatis</i> with compound 72a	222
4.6.5. Cytotoxicity analysis	224
4.7. Conclusions	228
4.8. References	230

Chapter 5 - Conclusions and Future Work

5.1. Conclusions: DNA Minor Groove Binders as potential anti-protozoal agents	233
5.1.1. Computational studies	233
5.1.2. Synthesis	234
5.1.3. Biophysical experiments	235
5.1.4. Biological experiments	236
5.2. Conclusions: Guanidine based derivatives as potential anti-tubercular agents	237
5.2.1. Computational studies	237
5.2.2. Synthesis	238
5.2.3. Biological studies	239
5.3. Future work: DNA Minor Groove Binders as potential anti-protozoal agents	242
5.3.1. Synthesis	242
5.3.2. Biophysical studies	244
5.3.3. Biochemical and biological studies	244
5.4. Future work: Guanidine based derivatives as potential anti-tubercular agents	245
5.4.1. Synthesis	245
5.4.2. Biochemical and Biological studies	246
5.5. References	247

Chapter 6 - Experimental

6.1. Synthetic Chemistry	248
6.1.1. Materials and Methods	248
6.1.2. General Procedures	248
6.1.3. Synthesis and Characterisation	251
6.1.4. HPLC purity assessment of final compounds	333
6.2. Computational methods	334
6.2.1. DFT calculations	334
6.2.2. Molecular Docking	334
6.2.3. Ligand based pharmacophore screening	335
6.3. Biophysical experiments	335
6.3.1. UV-thermal denaturation	335
6.3.2. Circular Dichroism	336
6.3.3. UV-Vis titrations	336

6.4. Biochemical & Biological methods	337
6.4.1. Resazurin Microtiter Plate Assay (REMA)	337
6.4.2. Drug resistant mutant generation	337
6.4.2.1. Liquid media continual passaging	337
6.4.2.2. Solid media single shot experiment	338
6.4.3. Whole genome sequencing	338
6.4.4. Cytotoxicity analysis against THP-1 cells	339
6.5. References	342
 Appendix	 344

Chapter 1

Introduction

1.1. Overview of infectious diseases

Infectious diseases pose a major public concern and constitute one of the main causes of mortality worldwide. These diseases have resulted in some of the most devastating pandemics such as the plague, ebola, and smallpox virus. Alarming some of these diseases are making a deadly comeback in many parts of the world, despite being preventable or treatable. Additionally, at least 30 new diseases have emerged in the last 20 years and now together threaten the health of hundreds of millions of people.¹ A more recent example of this is the emergence of the Covid-19 pandemic, which has affected the lives of people all over the world. Developing countries are at particular risk to the detrimental effects of infectious diseases due to poverty, malnutrition, over-population, lack of proper sanitation, and lack of funding. Diseases such as malaria, human African trypanosomiasis (HAT), leishmaniasis, and tuberculosis have a high prevalence in these poor regions of the world. However, increasing incidences of these diseases in developed countries have been reported.^{2,3,4,5} Increased understanding of the pathogenesis of infectious diseases has yielded drug treatments aimed to affect the metabolic integrity of parasites or aid host immune defense against infections. Issues such as drug resistance, lack of efficacy, and poor safety profiles have prevented the eradication of infectious diseases via drug-based therapy. This highlights the need for the identification of novel drug compounds as well as new therapeutic targets. The focus of this project is the design, synthesis, and pharmacological evaluation of potential drug candidates that may be used as novel treatments against infectious diseases.

1.1.2. Statistics

Malaria, HAT, leishmaniasis and tuberculosis are major contributors to the global burden of infectious diseases. Malaria cases from 2000 to 2015 (80 to 58 cases per 1000 population at risk) decreased by 27% owing to widespread malaria treatment and prevention campaigns. However, between 2015 and 2019 it declined by less than 2%, indicating a slowing of the rate of decline. According to the WHO, In 2019, there were an estimated 229 million cases and 409,000 deaths from malaria worldwide where the death rate of affected children under the age of 5 was 67%. There is a large prevalence of malaria in the WHO African Region which

consists of 47 states across the African subcontinent (Figure 1.1.2.1). This region constitutes an alarming 94% of malaria cases compared to South-east Asia (3%), Eastern Mediterranean region (2%), Western pacific region (0.74%), America (0.4 %) and Europe which has been malaria free since 2015. Consequently, the WHO African region has accumulated 384,000 deaths which is 95% of global malaria deaths.²

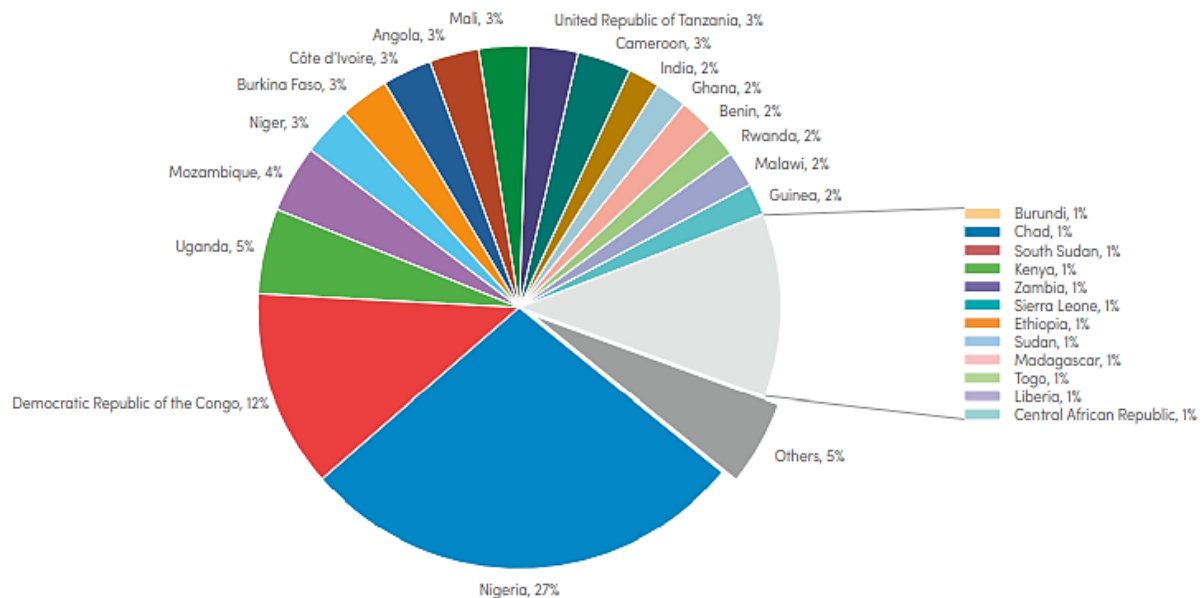


Figure 1.1.2.1 Global distribution of malaria cases by country, 2019.²

Major hurdles impeding the goal to eliminate malaria are the lack of sustainable and predictable international and domestic funding, risks posed by conflict in endemic zones and the emergence of parasite resistance to current anti-malarial agents and mosquito resistance to insecticides. Additionally, the Covid-19 pandemic has been shown to cause reductions in access to effective anti-malarial treatment in sub-Saharan Africa. Despite the presence of prevention campaigns, it is estimated that this has led to an additional 100,000 deaths in 2020.²

Human African trypanosomiasis, also known as African sleeping sickness, is another major protozoal disease caused by the parasite *Trypanosoma brucei* (*T.b.*) -*gambiense* or *T.b.* -*rhodesiense*. The *T.b.* -*rhodesiense* parasite accounts for 88% of reported cases where *T.b.* -*gambiense* represents 12% of reported cases. The disease mostly affects poor populations living in remote rural areas of Africa and if untreated can lead to death. In 2001, WHO launched an initiative to reinforce control and surveillance, and HAT decreased markedly in

the ensuing years. In 2009, the number of cases reported dropped below 10,000 for the first time in 50 years. This decline has continued to occur with only 1000 cases reported in 2019. Despite the low number of cases, there is an estimated risk population of 65 million people. The Democratic Republic of Congo (DRC) presents the highest percentage of annual cases of HAT of over 70% (Figure 1.1.2.2). However, countries such as Burkina Faso, Ghana, and Nigeria, have reported sporadic cases in the last 10 years.³

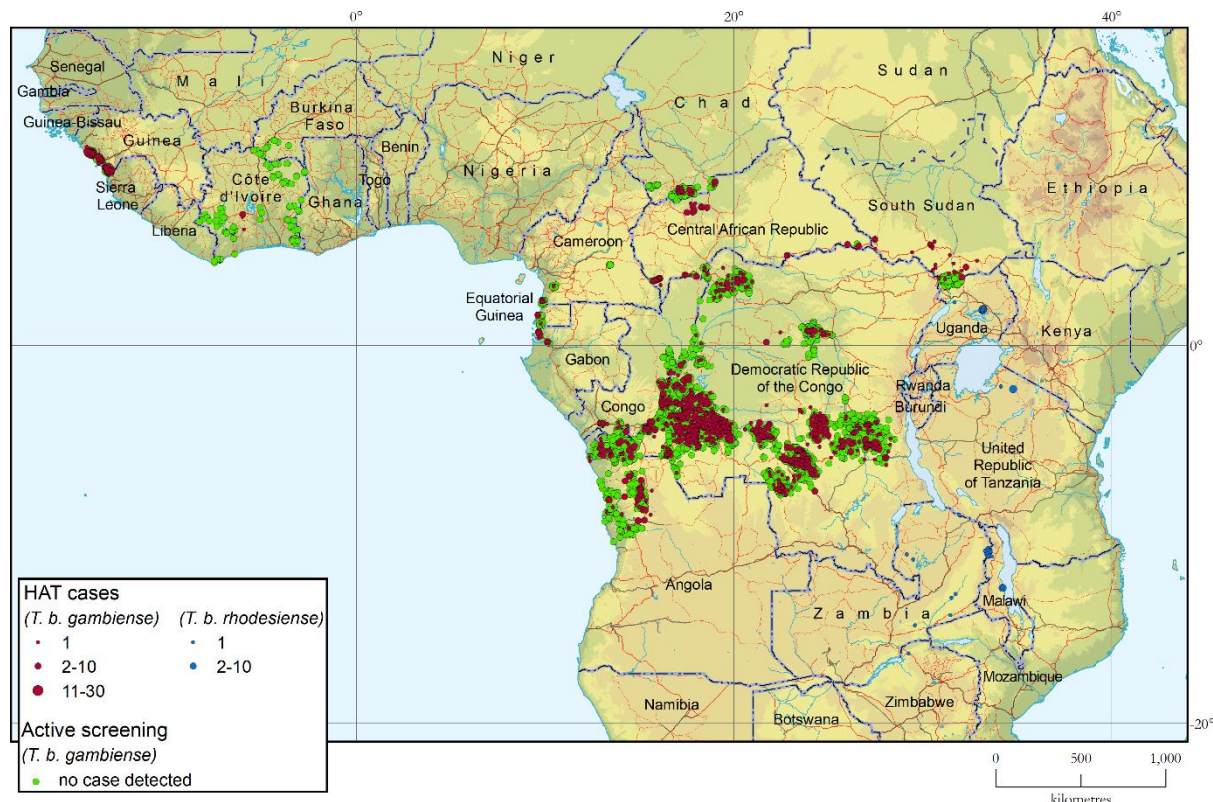


Figure 1.1.2.2. Geographical Distribution of HAT as reported from 2017-2018.⁶

Leishmaniasis has caused some major epidemics throughout history in East Africa and has resulted in high morbidity and mortality in high risk areas such as Kenya, Ethiopia and South Sudan. The disease manifests in three main forms which are visceral (VL), cutaneous (CL) and mucocutaneous (ML) leishmaniasis. Approximately 90% of global VL cases were reported in Brazil and certain areas in East Africa and Asia. Similarly, 87% of CL cases were reported in parts of Brazil, Colombia, Iraq and particularly high in Turkey and Iran. Even though it has not affected the western parts of the world, this neglected tropical disease results in over 20,000

deaths each year as a result of VL alone. Additionally, there is an overwhelming risk population of up to 1 billion people in endemic areas.⁴

Tuberculosis (TB) is ranked the most fatal infectious disease worldwide and constitutes the leading cause of death amongst HIV/AIDS patients. Being amongst the top ten leading causes of death it is no surprise that a lot of research is being conducted on sustainable treatments. In 2019, 10 million cases of TB were reported worldwide with 1.3 million deaths. As shown in Figure 1.1.2.3, the countries that are most affected are Nigeria, Indonesia, China, the Philippines, Pakistan and South Africa. It has been estimated that between 2000 and 2019, 60 million lives were saved through diagnosis and treatment. However, the emergence of Multi-drug resistant TB (MDR-TB) and Extensively-drug resistant TB (XDR-TB) has created a public health crisis and a massive health security threat. In 2019, 206,030 cases of MDR-TB were reported which is a 10% increase from 186,883 in 2018. Alarming, only 54% of MDR-TB patients and 30% of XDR-TB are currently successfully treated. Globally, TB incidence is falling at about 2% per year and between 2015 and 2019 the cumulative reduction was 9%. This was less than halfway to the End TB Strategy milestone of 20% reduction between 2015 and 2020. Given the rise of drug-resistant strains and failure to hit TB research funding targets, the United Nations goal of ending the TB epidemic by 2030 seems doubtful.⁵

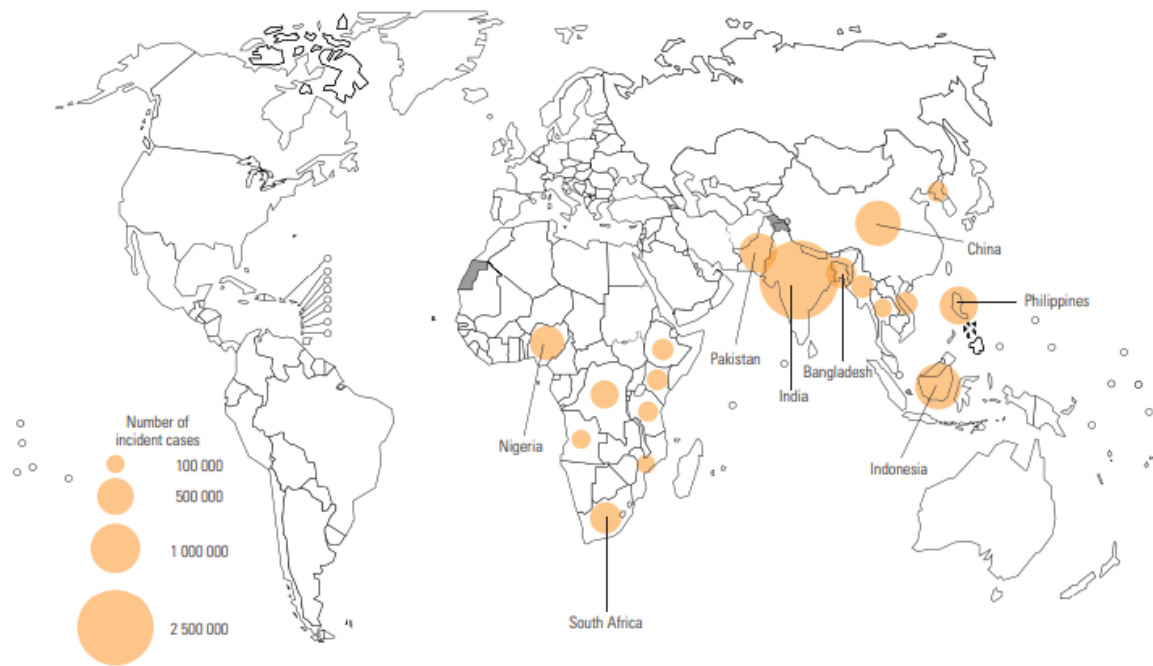


Figure 1.1.2.3. Estimated TB incidences in 2019 for countries with at least 100,000 incident cases.⁵

1.1.3. Disease pathogenesis and diagnosis

1.1.3.1. Malaria

The five main protozoans responsible for human malaria include *Plasmodium falciparum*, *P. vivax*, *P. ovale*, *P. malariae* and *P. knowlesi*. *P. falciparum* is prevalent in major parts of Africa, and parts of East Asia, whereas others such as *P. vivax* affect large portions of the Indian subcontinent. The malaria parasites undergo 10 or more morphological states, replicate from a single to more than 10,000 cells, and vary in total population of up to greater than 10^6 organisms. Following the bite from an infected female *Anopheles* mosquito (Figure 1.1.3.1.1), sporozoites are released into the human host. These sporozoites undergo a pre-erythrocytic liver-stage which typically lasts for 2 weeks. The liver merozoites produced are then released into the bloodstream where they invade erythrocytes and undergo cycles of asexual replication to produce more merozoites. As the parasites continue through their asexual life cycle over 24 to 48 hours, the level of parasitemia parallels the level of human response (i.e., fever, C-reactive protein [CRP], and tumor necrosis factor α [TNF- α]) until the patient crosses

a threshold of awareness and “feels ill”.⁷ Some erythrocytic stages develop into male and female micro- and macro- gametocytes which are ingested by the mosquito during feeding and subsequently develop into gametes in the insect gut undergoing sexual reproduction to form new sporozoites.

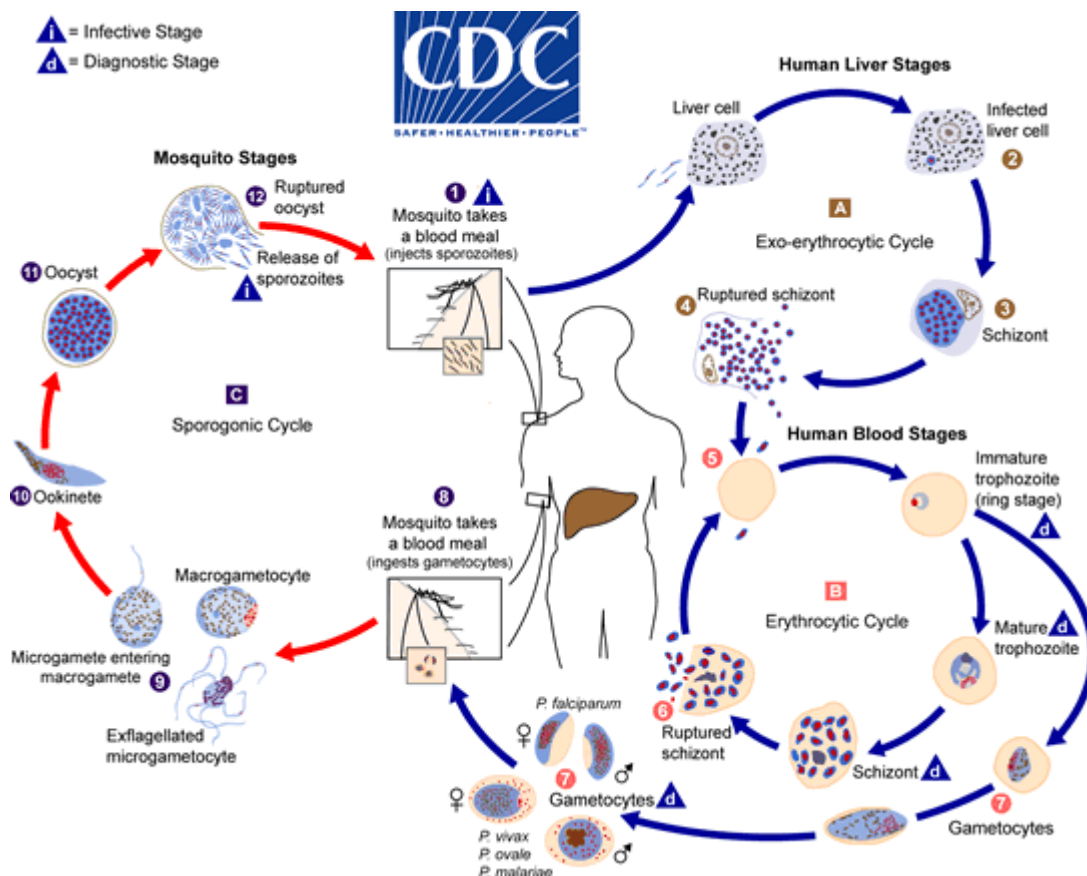


Figure 1.1.3.1. The Life cycle of *P. falciparum* in the human host.⁸

P. falciparum produces a high level of blood-stage parasites that sequester in critical organs in all age groups and cause anaemia in African children, in whom the vast majority of malaria deaths occur.⁹ Although *P. vivax* and *P. ovale* produce a milder disease, a portion of sporozoites (pre-blood stage spore) transform into dormant hypnozoites leading to relapse months or years after initial infection. *P. malariae* is the most benign form of malaria infection and has significantly lower levels of parasitemia than other malarial parasites. This is due to its long life cycle and the lower number of blood-stage parasite particles produced which leads to a more robust immune response. However, its long life cycle can manifest as chronic

malaria in humans that may last for decades. Despite the lower incidences of *P. knowlesi* infections, the disease severity and mortality rate is proportionally higher than *P. falciparum* or *P. vivax* infections. This is due to its new emergence in humans (zoonosis) and the absence of sufficient time for human adaptability.^{10, 11}

The symptoms of malaria develop once the erythrocytic cycle produces a parasitemia above a certain threshold (roughly 100 parasites per μL).¹² Clinical symptoms include prostration, respiratory distress, convulsions, anemia, vomiting, jaundice, and abnormal bleeding. *P. falciparum* is unique in that erythrocytes containing mature parasites sequester inside small vessels, avoiding parasite clearance in the spleen, and causes host endothelial cell injury and microvascular obstruction. This enhanced pathogenesis is partly caused by the parasite's ability to adhere to the host endothelial surface which is mediated through erythrocyte membrane protein 1 (PfEMP1). Sub-types of PfEMP1 are capable of binding to the intercellular adhesion molecule-1 and endothelial protein C receptor, which is associated with cerebral malaria that can lead to a coma. Cytoadherence to the placenta of infected pregnant women can also lead to complications such as maternal anemia, low birth weight, increased risk of abortion and stillbirth.

Correct diagnosis is essential to ensure species-specific anti-malarial treatment. The gold-standard method for diagnosing malaria involves light microscopy of blood-stained films (Figure 1.1.3.1.2). Most parasites are distinguishable under light microscopy except for *P. knowlesi* and *malariae* which have almost similar morphology. On the other hand, *P. falciparum* is easily detectable as only ring forms and banana-like gametocytes are usually present in the peripheral blood since mature parasites become sequestered. However, due to the large distribution of *P. falciparum* across Africa, rapid diagnostic tests (RDTs) have been created as the first-line method.¹³ The RDTs is based on the detection of antigens or enzymatic activities associated with the parasites. The most common antigen is the histidine-rich protein 2 (PfHRP2) antigen which is specific for *P. falciparum* malaria and is highly expressed. Worryingly, *P. falciparum* isolates from Amazon regions, Africa, and India have been found with HRP2 gene deletions, which threatens the ability to diagnose and appropriately treat people infected with this type of malaria. This leaves a reliance on other key markers including enzymes of the parasite glycolytic pathways such as plasmodial lactate dehydrogenase (pLDH) and aldolase which can be present in all human malarias.¹⁴

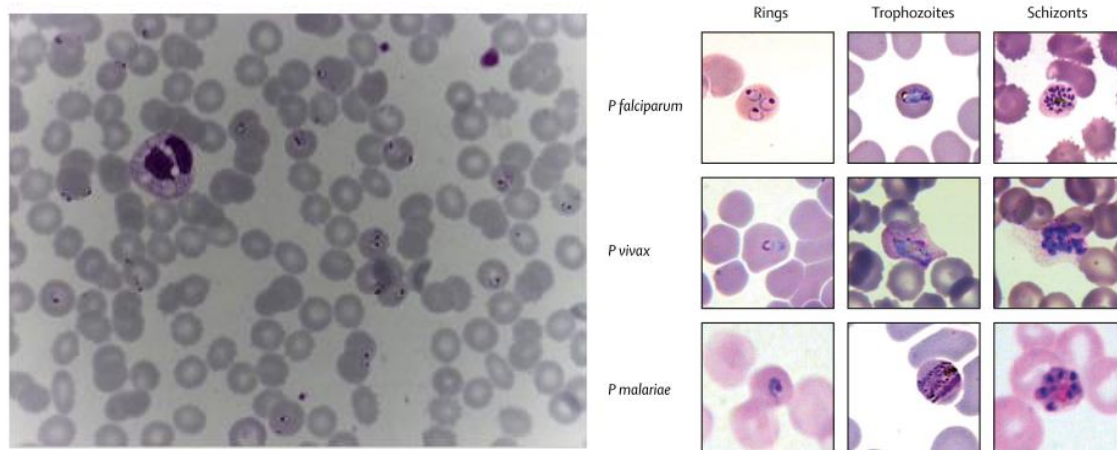


Figure 1.1.3.1.2. Malaria blood slides showing parasitised red blood cells.¹⁵

PCR or LAMP (loop-mediated isothermal amplification) methods are the most sensitive tests capable of identifying low levels of parasitemia, parasite species, or mixed infections. However, these techniques are not suitable for routine use due to time and expense. Another diagnostic method are serology tests based on the detection of antibodies against malarial parasites, using either indirect immunofluorescence (IFA) or enzyme-linked immunosorbent assays (ELISA).^{14, 16} This last method is used the least as it does not detect current infection but rather measures past exposure. Due to the decreasing efficacy of the current diagnostic tests, novel malaria diagnostic targets are being investigated including extremely conserved genes of certain proteins such as heat-shock protein (HSP-70), heme detoxification protein (HDP), dihydrofolate reductase (DHFR) and high-mobility group box 1 (HMGB1).¹⁶

1.1.3.2. Human African Trypanosomiasis

Like malaria, HAT is a neglected tropical disease that presents with similar symptoms but has severe neurological complications. There are two forms of the parasite that cause human disease; the slow-progressing form, caused by *Trypanosoma brucei gambiense* which is endemic in western and central Africa; and, the faster-progressing form, caused by *T. b. rhodesiense* found in eastern and southern Africa. The third subspecies *T. b. brucei* is a parasite that primarily infects cattle and other animals, and under normal conditions does not infect humans. The main two species are different from other family members as they are

resistant to apolipoprotein A1, a serum protein that triggers trypanosome death. This factor enables both species to infect humans with minimal immune detection.¹⁷ Additionally, the trypanosome cell membrane is covered by a dense coat of identical glycoprotein dimers that shields the underlying membrane against innate immunological attack.¹⁸ These glycoproteins induce a specific antibody response that triggers the destruction of all antibody-opsonised trypanosomes. However, antigenic variation caused by the replacement of the parasites glycoprotein coat results in irregular fluctuation in parasitemia which is reflected by intermittent fevers as a host immune response.

Following the bite from a tsetse fly that carries the parasite, several days of local trypanosome multiplication occur which results in the formation of skin inflammatory lesions at the site of injection. The trypomastigotes then spread via the lymphatic system and the bloodstream to various peripheral organs and tissues where they continue to replicate by binary fission (Figure 1.1.3.2.1). The entire life cycle of the parasite is extracellular meaning they spend their life in the blood and tissues without entering the host's cells. This makes it easy for the reuptake of bloodstream trypomastigotes during the tsetse fly's next blood meal from an infected human host. Within the infected fly, transformation from a procyclic trypomastigotes to mature epimastigotes occurs which sequester into the insect's salivary glands. Multiplication by binary fission occurs followed by transformation into metacyclic trypomastigotes which are injected into the mammalian host upon the insect's next blood meal.

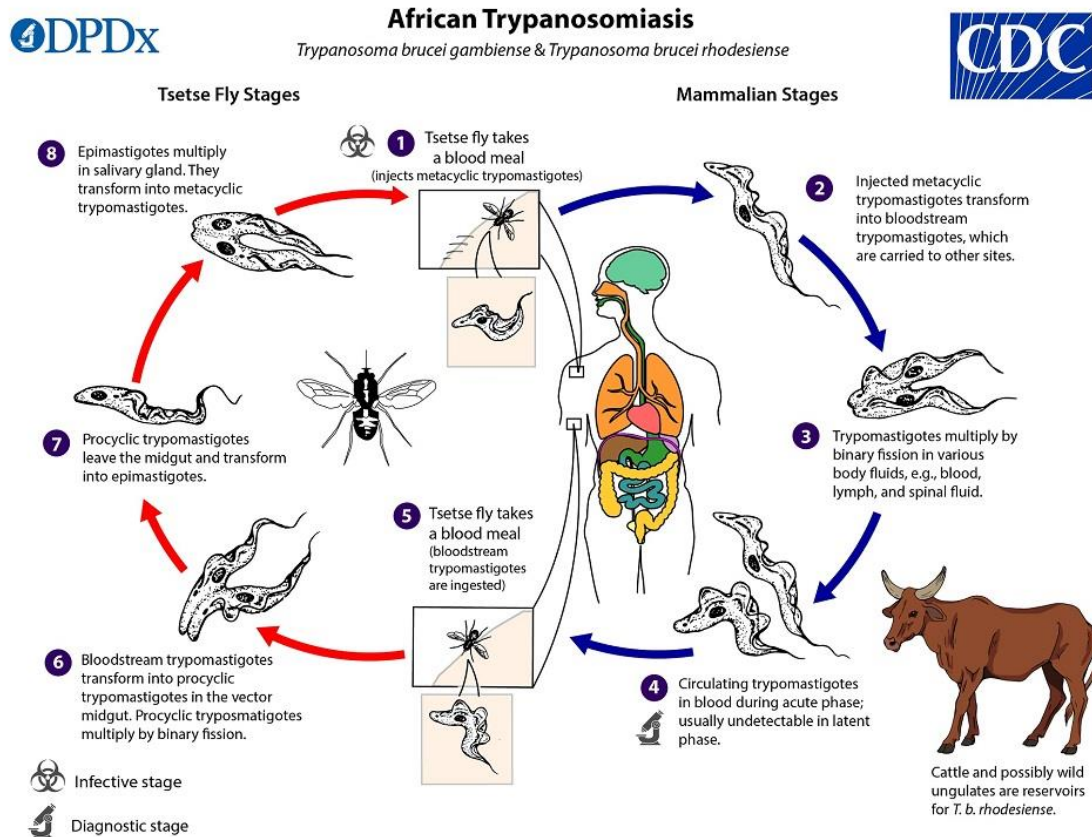


Figure 1.1.3.2.1. The Life cycle of *Trypanosoma brucei gambiense* and *rhodesiense*.¹⁹

During the later stages of the disease in an infected human host the parasites cross the blood-brain barrier, where they trigger local inflammation and neurological damage.²⁰ The hallmark alterations observed in HAT patients are polyphasic sleep-wake patterns and alterations of the internal structure of sleep characterised by SOREM (sleep onset REM). These features result in daytime somnolence, sudden overwhelming sleep urges and nocturnal insomnia. Other physical symptoms include dermal reaction at the site of infection, intermittent fever, headaches, pruritis, oedema, endocrine dysfunction, tremors, speech disorder and motor weakness.¹⁸ Clinical features of HAT are unspecific and can sometimes be mistaken for other related diseases. The traditional approach for the detection of *T. b. gambiense* is via light microscopy of lymph node extracts. As parasitemia levels for *T. b. rhodesiense* are higher the parasite is detectable in the blood and even found in chancre fluid or bone marrow aspirates. Serodiagnostic tests only exist for *T. b. gambiense* infection and are based on specific antibody detection. However, serological methods are not suitable for passive screening and

surveillance in fixed health centers in endemic countries due to the lack of electricity and laboratory infrastructure. For this reason, card agglutination test of trypanosomiasis (CATT) and RDTs have been developed to test for *T. b. gambiense*; however, they are not 100% specific. Second generation RDTs that are capable of detecting both species are currently under development. However, there is a high risk of cross reaction with non-human infective trypanosomes.²¹ Moreover, light microscopy of blood-stained films remains to be the most sensitive laboratory technique used to detect the presence *T. b. gambiense* and *T. b. rhodesiense*.

1.1.3.3. Leishmaniasis

Leishmaniasis is a vector-borne disease caused by several species of the genus *Leishmania* and is transmitted between mammalian hosts by the female sand-fly. Different species of *Leishmania* cause different clinical outcomes that can be classified as cutaneous leishmaniasis (CL) or visceral leishmaniasis (VL). VL caused by *L. donovani* and *L. infantum* is the most severe resulting in severe skin manifestations and fatality if untreated. Cutaneous leishmaniasis is usually limited to an ulcer that self-heals over 3-18 months but can progress to scarring, disfigurement and stigmatisation.²² As shown in Figure 1.1.3.3.1, infiltration of the mammalian host by the leishmanial parasite in the form of promastigotes are phagocytosed by macrophages and transform into amastigotes. These amastigotes multiply and are released into the bloodstream where they invade other tissues. Upon contact of a sand-fly with an infected host, parasitized cells are ingested where the amastigotes transform into promastigotes for further infection.²³

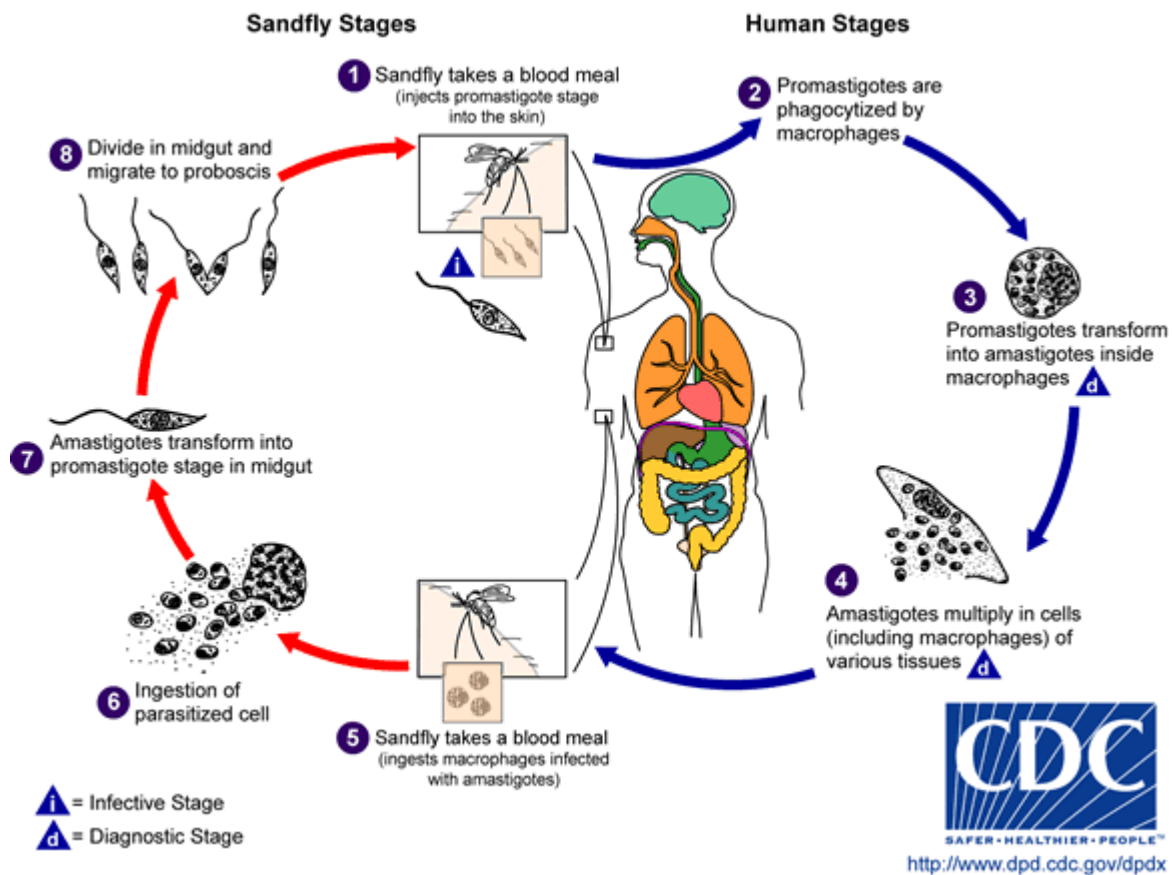


Figure 1.1.3.3.1. *Leishmania* life cycle.²⁴

The classical diagnosis for VL is through microscopic observation of the amastigote parasite stage in tissue specimens or cultures obtained from the spleen, bone marrow or lymph node aspirates.²⁵ Similar to HAT infections, blood samples tend to have low sensitivity, except in HIV-co-infected patients, who have a higher parasitemia. Various serological assays are available including direct agglutination tests, ELISA, immunofluorescence, and western blot. Although these techniques share a high sensitivity for acute visceral disease they do not identify disease stage. Additionally, these assays are too complicated for use in endemic areas. Currently, there is no single reference test for cutaneous leishmaniasis, but the observation of amastigotes in a clinical biopsy from a direct smear of needle aspirate, a skin-slit smear or cultured material establishes the diagnosis. The development of RDTs and use of PCR and LAMP methods is currently under investigation.^{26, 27}

1.1.3.4. Tuberculosis

Mycobacterium tuberculosis (*M. tuberculosis*) is the causative agent of tuberculosis (TB) which is referred to as the most fatal infectious disease worldwide.⁵ *M. tuberculosis* is a gram-positive, slow-growing bacterium with a highly complex, unique lipid cell wall which has evolved to play a specialised role in the pathogenesis of the bacterium.²⁸ While it can infect many animals, they cannot transmit the infection to others. This means that the continued survival of *M. tuberculosis* depends upon transmission among humans. TB typically affects the lungs (i.e. pulmonary TB), but can also affect other sites of the body such as lymph nodes, skin and bones (i.e. extrapulmonary TB). Upon infection of a human host, macrophages are considered the first line of defense; however, the ability of the bacterium to survive within these immune cells after phagocytosis renders the bacterium invisible to the immune system. Within the macrophage, the pathogen inhibits phagosome maturation and phagosome-lysosome fusion by causing a downregulation in effector molecules IFN- γ and tryptophan aspartate-containing protein (TACO).^{29,30} This facilitates access to required nutrients and induces apoptosis in the host cell.

If the combination of both innate and adaptive immune responses fails to eliminate the bacteria upon initial exposure, a granuloma is formed which is a hallmark of the disease (Figure 1.1.3.4.1). This acts as a barrier between the pathogen and the host immune system. The inflammatory cytokine, TNF- α , is key to the formation of the granuloma and its size has been shown to be proportional to TNF- α production.³¹ In most cases, the mycobacteria within the granuloma are killed due to the hypoxic and cytotoxic conditions created by the mass of surrounding immune cells. This creates a central necrotic core (caseum) which is shielded by activated macrophages, T cells, B cells, monocytes, neutrophils, and Natural Killer cells. If the pathogen is not eliminated by these means, the mycobacteria can remain dormant in a latent state, termed latent tuberculosis infection (LTBI). These individuals are asymptomatic and do not transmit the disease. Less than 10% of infected LTBI individuals will develop clinical disease during their lifetimes, but once disease does develop and remains untreated, it is fatal in 50% of patients.²⁹ Interestingly, most cases of tuberculosis are typically a result of a reactivation of existing foci, rather than as a direct outcome of primary infection. This is accelerated in immunocompromised individuals such as HIV patients who can also develop

the active disease after primary infection. The symptoms of active TB include severe coughing (often with blood), chest pains, weight loss, fever, night sweats and breathlessness.³²

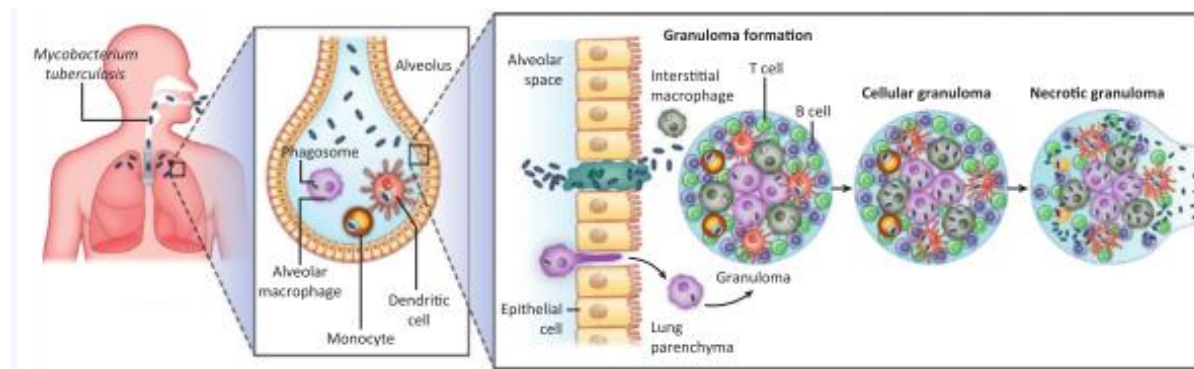


Figure 1.1.3.4.1. Infection cycle of *M. tuberculosis* showing macrophage invasion and granuloma formation.³³

The current diagnostic tests for TB are the tuberculin skin test (TST). Although TST is most suitable for TB testing in developing countries due to the minimal laboratory equipment and healthcare workers needed, TST has several inherent problems. For instance, as the antigens present in test extracts also exist in the vaccine strain *M. bovis* BCG and in several environmental mycobacteria, there is a risk of miss diagnosis. Additionally, TSTs are only designed to measure the adaptive immune response of individuals who have been infected by the pathogen six to eight weeks after exposure. Due to these pitfalls and advances in genomics and immunology, more sensitive and more specific tests for the diagnosis of TB have been developed. These include two recently FDA approved interferon-gamma release assays, QFT-G-IT and T-SPOT.³⁴

1.1.4. Standard of care and current treatments

1.1.4.1. Malaria

The current standard of care against malaria consists of four classes of compounds: quinoline derivatives, anti-folates, artemisinin derivatives and off-label anti-microbial agents. First isolated from the bark of the cinchona tree in 1820, quinine was the first effective treatment

for malaria caused by *P. falciparum*. Quinine remained the drug of choice until the late 1940s, when other compounds were introduced.³⁵ However, based on the structure of quinine a large number of compounds have been synthesised, known as the quinoline derivatives. These include chloroquine, mefloquine, primaquine and amodiaquine. Quinoline derivatives act by forming complexes with heme which prevents their crystallisation in the plasmodium food vacuole. The cytotoxic free heme molecule then accumulates and causes parasitic death. Currently, quinoline derivatives are rarely used alone as they are commonly used in combination with a complementary drug (e.g. mefloquine and artesunate, sold as Artequin™) to reduce the chance of resistance development.³⁶

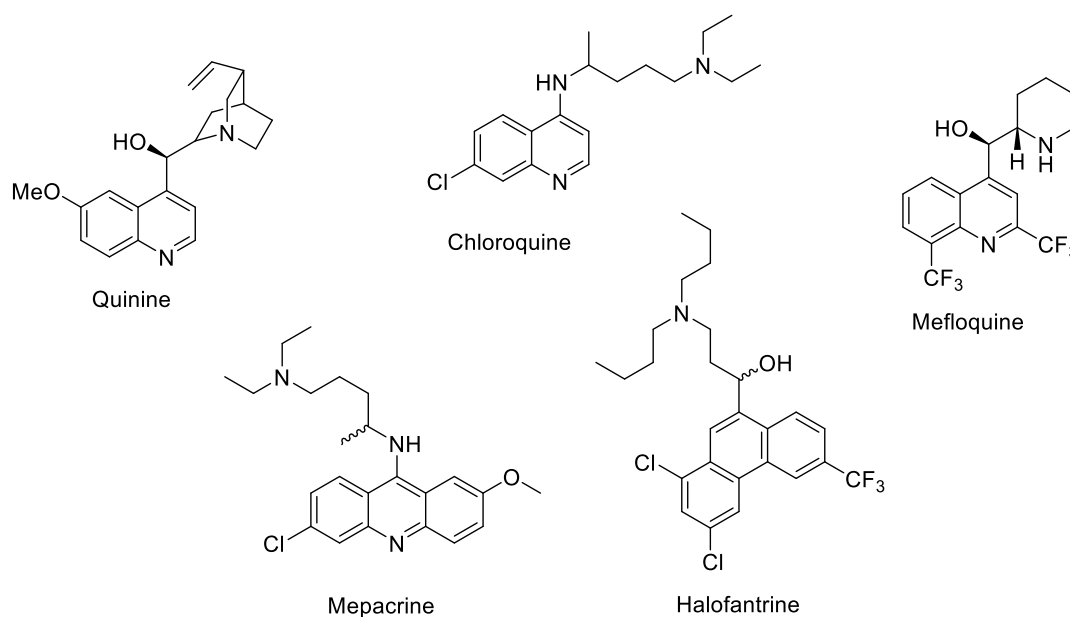


Figure 1.1.4.1.1. Structures of uinoline derivates used to treat malaria.

Another class of drugs that can either be used as a first or second-line treatment against malaria are anti-folates. These folate antagonists target the dihydrofolate reductase (DHFR) activity of the parasite leading to the inhibition of purines, thymidylic acid and certain amino acids which inevitably results in the inhibition of DNA and protein synthesis. Pyrimethamine and proguanil are two of the most prescribed anti-folates; however, they are commonly used in combination with quinolone derivatives, sulfa drugs and certain antibiotics such as azithromycin and tetracycline. The combination of mefloquine-sulfadoxine-pyrimethamine

(MSP) has been shown to have a synergistic effect and is used for prophylaxis against multi-drug resistant *falciparum* malaria.³⁷

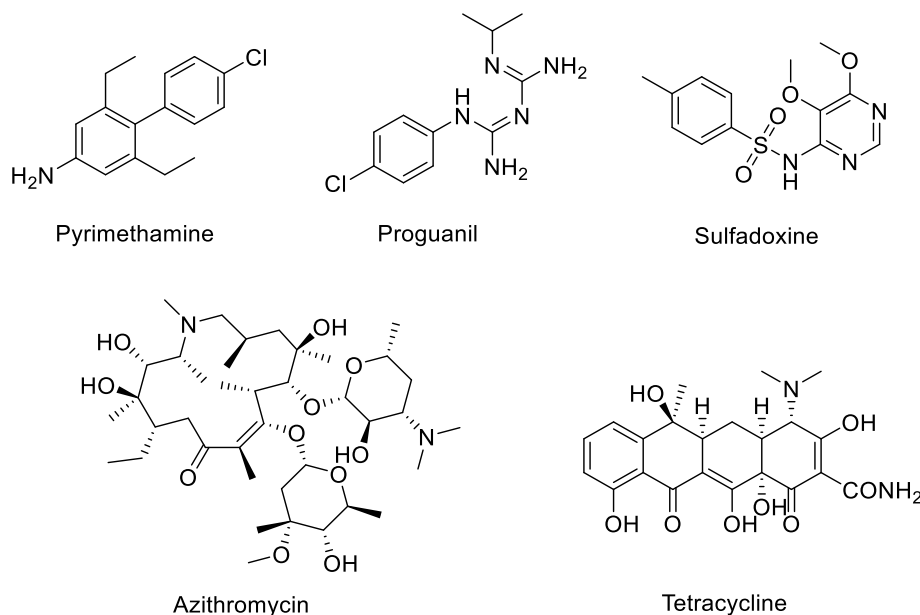


Figure 1.1.4.1.2. Structures of anti-folates: pyrimethamine and proguanil, sulfa-drug: sulfadoxine and antibiotics: azithromycin and tetracycline.

The deployment of artemisinin-based combination therapies (ACTs) has been, and continues to be, integral to reducing the number of malaria cases and deaths. A significant discovery program by Chinese chemists in the 1970s has led to the discovery of artemisinin from an herb known as *Artemisia annua*.³⁸ Artemisinin and its derivatives are effective not only against multi-resistant strains of *P. falciparum* but have broad stage specificity against the plasmodium life cycle including activity throughout the asexual blood stages. The most common derivatives of artemisinin are artemether, artesunate and arteether (Figure 1.1.4.1.3). These semi-synthetic derivatives are prodrugs that are transformed to the active metabolite, dihydroartemisinin. The emergence of these classes of compounds has benefited anti-malarial drug therapy due to their fast reduction of parasite growth and rapid clearance of clinical symptoms. The mechanism of action through which artemisinin acts has been widely debated. The most accepted theory is that the endoperoxide bridge in the structure interacts with heme iron in the parasite and generates oxygen radicals. The resulting free radical selectively binds to membrane proteins and causes lipid peroxidation, damages the endoplasmic reticulum, inhibits protein synthesis and ultimately results in lysis of the

parasite.³⁹ However, due to the very short half-life of artemisinins, they are commonly used in combination with longer half-life agents in order to reduce multi-dose regimens, increase efficacy and decrease resistance. Examples of these combinations are artesunate-amodiaquine (AS-AQ), artesunate-sulphadoxine-pyrimethamine (AS-SP), or artemether-lumefantrine. Artemisinin hybrid compounds have also been produced such as Artemisinin-Quinine hybrid which has superior activity against chloroquine-sensitive (CQS) and chloroquine-resistant (CQR) strains of *P. falciparum* compared to quinine or artemisinin alone and compared to a 1:1 mixture of the two.⁴⁰

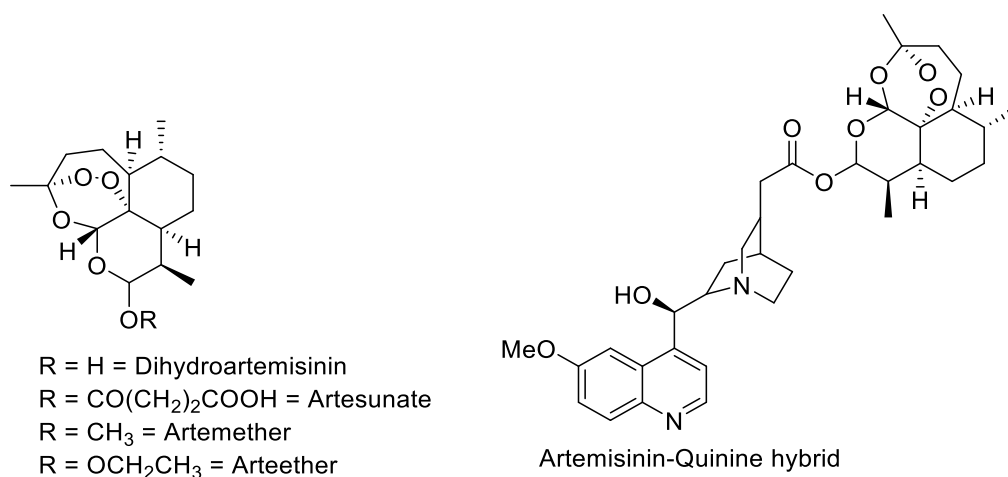


Figure 1.1.4.1.3. Structures of artemisinin derivatives and hybrid compound.

1.1.4.2. HAT and Leishmaniasis

In relation to the treatment of trypanosomal diseases such as leishmaniasis and HAT certain agents also used to treat malaria are prescribed. These include quinolone derivatives, anti-folates and off-label use of certain antibiotics such as Amphotericin B. However, certain drugs such as pentavalent antimonials (e.g. meglumine), miltefosine and melarsoprol are specifically used to treat leishmaniasis and HAT, respectively. Among the large array of compound families used to treat these tropical diseases, DNA minor groove binders (MGBs) have a track record of success as anti-protozoal agents. One of the oldest MGBs, pentamidine, has been at the forefront of anti-parasitic therapy since the 1940s, but, like many of this class of compounds, it has several toxicity issues including renal and hepatotoxicity.⁴¹ Additionally, as pentamidine cannot cross the blood-brain barrier, it has proven to be less useful for the

treatment of late-stage HAT infections. In order to improve the pharmacokinetics and safety profile of pentamidine other bis-amidine derivatives such as the orally bioavailable furamidine were developed. Prodrugs of furamidine have emerged that have been shown to cure mouse and primate models of stage 2 trypanosomiasis. Another class of MGBs that have shown to display anti-parasitic activities are polyamides such as netropsin and distamycin. Even though netropsin and distamycin have shown good activity in in-vitro parasitic studies, derivatives have not yet been produced with sufficient activity and selectivity to become candidate drugs.⁴¹

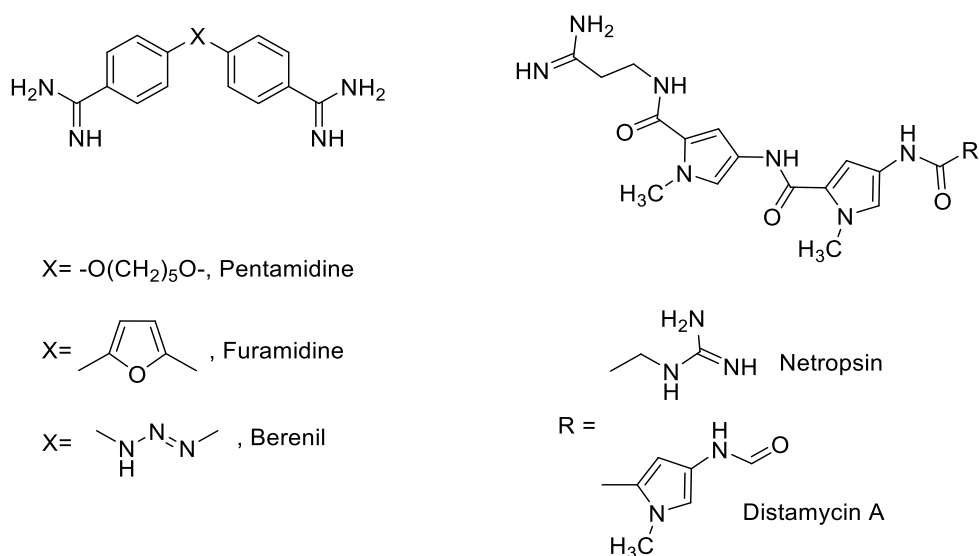


Figure 1.1.4.2.1. Minor groove binders with anti-parasitic activity.

1.1.4.3. Tuberculosis

The current standard of care for the treatment of drug-susceptible TB consists of a 6-9 month dosage regime with a combination of first-line anti-TB agents (Figure 1.1.4.3.1). This is broken down into a 2-month “intensive” phase of a four-drug cocktail containing rifampicin (RIF), isoniazid (INH), pyrazinamide (PZA), and ethambutol (EMB) followed by a longer “continuation” phase of RIF and INH to eradicate the remaining bacilli that have entered a dormant, slowly replicating latent phase.⁴² Although this regime has proved to be effective in improving the clinical outcomes of infected patients, poor compliance is a major issue due to the prolonged treatment periods as well as the observed side effects. It is estimated that

nearly 50% of patients treated for pulmonary tuberculosis develop long-term lung damage and functional disability. Even though these first-line agents have been extensively studied it is clear that there is a lack of efficacy with the sole use of this strategy. For this reason, several host-directed therapies are being developed which aims to shorten treatment duration, reduce excessive inflammation, minimise permanent lung damage and prevent drug resistance. In addition, research into biomarkers to predict response to treatment, as well as risk of relapse, is being conducted.⁴³ As well as these adjunct therapies there is a large focus on the repurposing of existing drugs used to treat other infectious diseases and the development of new agents with novel mechanisms of action. Both of these approaches have become the main defense strategies to combat the current issues of anti-tubercular therapy.

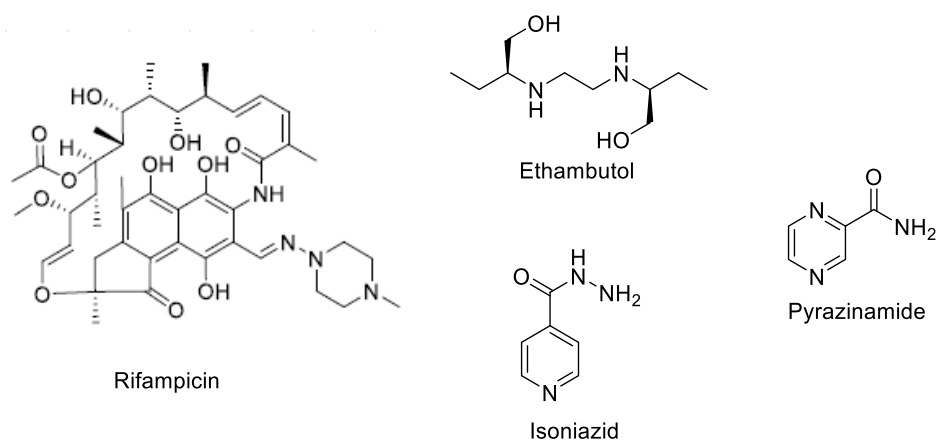


Figure 1.1.4.3.1. Structure of first-line anti-TB agents; Isoniazid, pyrazinamide, ethambutol and rifampicin.

1.1.5. Drug resistance

Drug resistance has resulted in an increase in mortality and morbidity as a result of infectious diseases. This has huge knock-on economic consequences which are exacerbated in some of the world's poorest countries. The various methods in which infectious microbes develop resistance towards drug compounds have been extensively studied. Some of these methods include drug efflux, limitation of drug uptake, target by-pass, drug inactivation and modification of the drug target.⁴⁴ The emergence of multidrug resistant *P. falciparum* has led to a loss in efficacy of current anti-malarial drug therapies. Production of plasmodium transmembrane proteins chloroquine resistance transporter (PfCRT), multidrug resistance

protein 1 (PfMDR1) and multidrug resistance-associated protein 1/2 (PfMRP1 and PfMRP2) are thought to be responsible for changes in the parasite's sensitivity to almost all current anti-malarial agents including the gold standard artemisinin.⁴⁵ Mutations in the *pfcr* gene enable the transporter to efflux chloroquine and other derivatives across the digestive vacuole membrane preventing contact with the anti-plasmodial target. However, the mechanism of resistance inferred by PfMDR1, PfMRP1 and PfMRP2 remains unclear. Resistant strains with *PfDHFR* mutations have drastically compromised the efficacy of anti-folates such as pyrimethamine and sulfadoxine. Interestingly, resistance towards DHFR inhibitors has also been reported as a result of the presence of pteridine reductase (PTR1) which is an enzyme capable of acting as a by-pass for DHFR in *Leishmania* and trypanosomes.⁴⁶ Alarming, resistance towards the gold-standard artemisinin has been detected in Cambodia and neighboring countries. The potential of widespread resistance and the fact that the molecular mechanisms responsible for artemisinin resistance is still under debate threatens future anti-malarial drug therapy.⁴⁷

Resistance to anti-trypanocidal agents is also a major issue that is hindering the effectiveness of HAT and leishmaniasis treatments. Pentamidine resistance isolates that have been identified lack the key transporters high affinity pentamidine transporter (HAPT1) and P2 aminopurine transporter (P2) which are responsible for the uptake of pentamidine into trypanosome cells.⁴¹ Mutations in the genome of *T. b. gambiense* that confirms resistance to malarsporol have emerged as early as the 1970s and became widespread by the late 1900s. Most resistant isolates identified were shown to display mutations in the solute and drug transporter, aquaglyceroporin (AQP2).⁴⁸ Similarly, overexpression of ABC efflux transporters has been shown to confer resistance to miltefosine, another common agent used to treat trypanosomal diseases. Other mutations such as those found in the pyridoxal kinase gene and MT/ROS3 transporter gene have severely hindered its anti-parasitic efficacy.⁴⁹

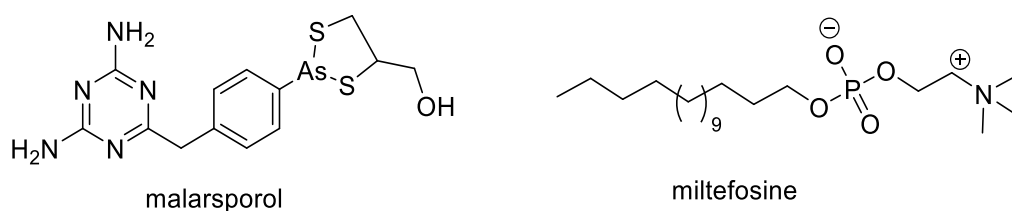


Figure 1.1.5.1. Structures of anti-HAT agent malarsporol and anti-leishmaniasis agent, miltefosine.

Multi-drug resistant TB (MDR-TB) has been characterised as resistance to at least two of the most effective anti-TB drugs, isoniazid and rifampicin. This has led to the implementation of a longer regime of 18 months of a combination of first-line and second-line agents which are more expensive and toxic. These second-line agents include aminoglycosides, fluoroquinolones, cycloserine and recently approved agents bedaquiline and delamanid (Figure 1.1.5.2).⁵⁰ Additionally, extensively drug-resistant (XDR) strains have emerged which combine MDR with resistance to at least six classes of second-line agents.^{51,52} Common mechanisms responsible for these levels of resistance include mutations in *rpoB* (rifampicin resistance), *KatG* (isoniazid resistance) and overexpression of drug efflux pumps which results in intrinsic resistance to both first-line and second-line agents.⁵³

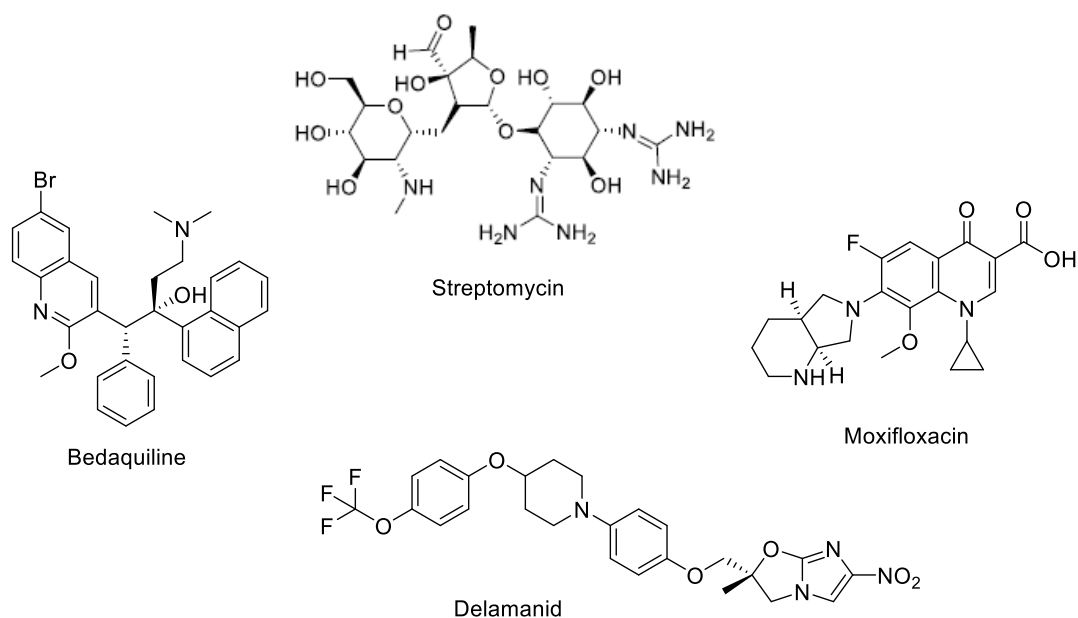


Figure 1.1.5.2. Structures of second-line agents: streptomycin, moxifloxacin, bedaquiline and delamanid.

1.1.6. Prospects and drug pipeline

There is a limited number of chemical entities in the current drug armory to combat infectious diseases and this has led to the increased use of combination therapy and drug repositioning. These approaches offer a short and fast method of treatment with the added advantage of a significantly reduced cost of development. Old drugs such as miltefosine, moxifloxacin and pentamidine as mentioned above were primarily designed for a single indication but are currently approved for the treatment of various infectious diseases.⁴⁶ Despite the success that these compounds have had on treating infectious diseases, the emergence of multidrug resistance, lack of efficacy and toxicity poses a huge threat. More recently, there has been a greater focus on the discovery of novel compounds with different mechanisms of action which may offer long-term benefits. Examples of this include the discovery of bedaquiline and delamanid and their subsequent approval in 2014, which were the first novel anti-TB drugs to be approved in over 40 years.⁵⁰ Since then, there have been ongoing trials of promising drug candidates such as AZD5847, pretomanid and SQ109 which are to be used for the treatment of MDR-TB.⁵⁴ Similarly, with anti-protozoal drug discovery, not many new chemical entities have reached the market in the past decade. The current ongoing clinical trials involve potential vaccines and hybrid compounds made from existing anti-protozoal agents as shown in Figure 1.1.6.1. However, two novel drug compounds, acoziborole and NITD609, are currently being investigated for the treatment of HAT and malaria, respectively.^{36, 55}

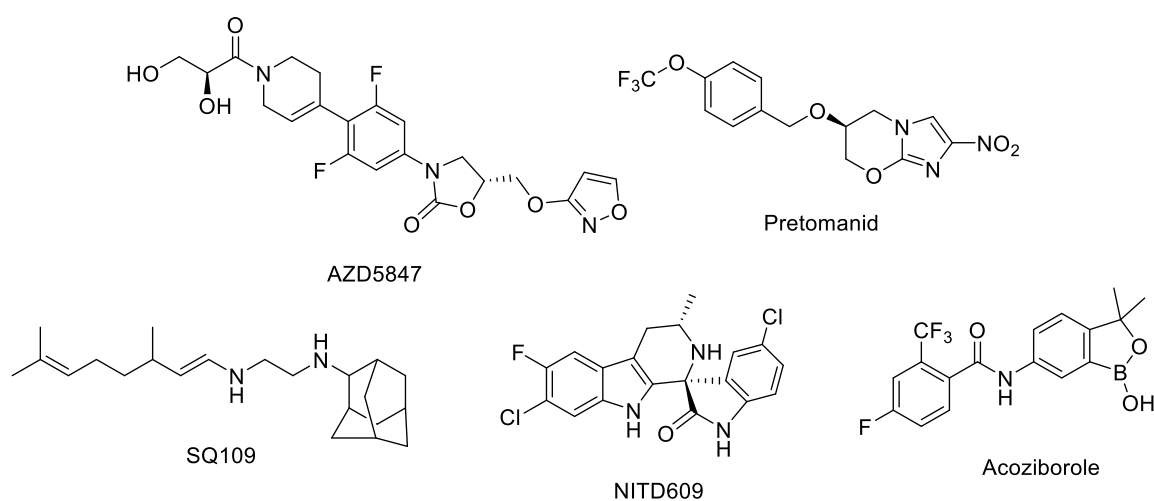


Figure 1.1.6.1. Structures of current drug candidates in the infectious disease treatment pipeline; AZD5847, pretomanid, SQ109, Acoziborole, NITD609.

1.2. Therapeutic targets

Using whole-cell based approaches, the molecular targets for the long list of current and emerging drugs used to treat infectious diseases have been elucidated. Some of these include cell wall biosynthetic enzymes, heme metabolism, folate biosynthetic enzymes, ribosomal subunits, kinases and DNA binding agents. In this report, we will mainly be focusing on DNA as a viable therapeutic target for the treatment of protozoan infectious. Additionally, the use of new technologies to elucidate the mechanism of action of repurposed drug compounds will be explored for the treatment of *M. tuberculosis* infections. Methods for target identification and target validation have revolutionised the field of drug discovery and development and we will see an example of this in the present investigation.

1.2.1. Overview of DNA

DNA has been an attractive drug target for the treatment of various diseases such as cancer, anti-parasitic and anti-bacterial infections. To appreciate the DNA-ligand interactions involved in DNA binding, one must understand the molecular architecture of DNA (deoxyribonucleic acid).

The DNA helical polymer contains the genetic information that is responsible for the function and development of all living organisms. The HB base pairing responsible for the double helix causes a difference in electrostatic potential along the DNA strand where A-T regions have a greater negative potential and G-C rich regions have greater positive potential. This is a key property responsible for the regional selectivity of certain DNA binding agents. More recently another base pairing network was discovered by Karst Hoogsteen which shows considerably different geometries from the Watson-Crick base pairs. This Hoogsteen base-pair often involves the N7 position of purine bases as an HB acceptor and the C6-exocyclic amino group of adenine as an HB donor, which then interact with the Watson-Crick surface of pyrimidine bases (Figure 1.2.1.1, b). These base pairs exist transiently in some DNA molecules in thermal equilibrium with standard Watson–Crick base pairs. Interestingly, Hoogsteen base pairing

display distinct behaviour and allow for the formation of secondary structures including DNA triplexes and quadruplexes.⁵⁶

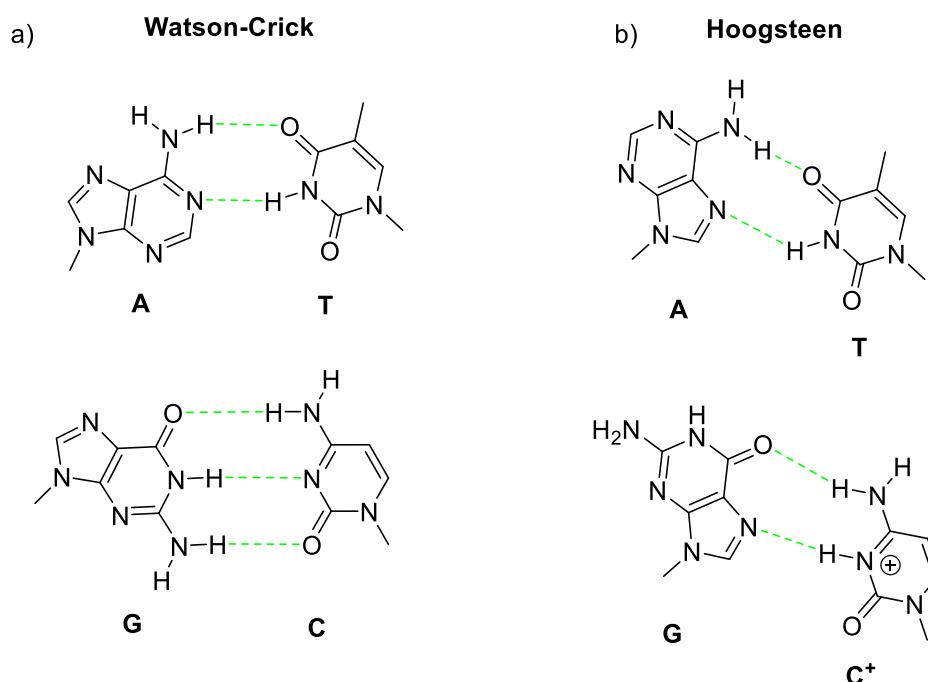


Figure 1.2.1.1. (a) Watson-Crick and (b) Hoogsteen interactions between DNA base pairs.

In both Watson-Crick and Hoogsteen base pairing cases, the molecular organisation is strengthened by π - π stacking interactions between the nucleotide bases. Due to the location of the phosphate groups, outside of the polymer, monovalent and divalent cations have easy access to them and the binding of these cations leads to the further stabilisation of DNA. Additionally, the open structure of DNA has a high content of water molecules known as the “spine of hydration”. At lower water content the bases tilt so that the structure becomes more compact. Guided by the chirality of the deoxyribose unit, these structures are generally a right-handed helix. As shown in Figure 1.2.1.2, the right-handed helix possesses grooves that can be of the same size (termed “A-DNA”) or where each groove is different (B-DNA). Poly-CG strands in very high salt concentrations (>4M NaCl) can adopt a left-handed helix with very wide grooves (Z-DNA). As DNA can exist in these various forms it is known to be polymorphic. B-DNA represents the most common form of DNA and was the first crystal structure of DNA to be determined in 1979.⁵⁷

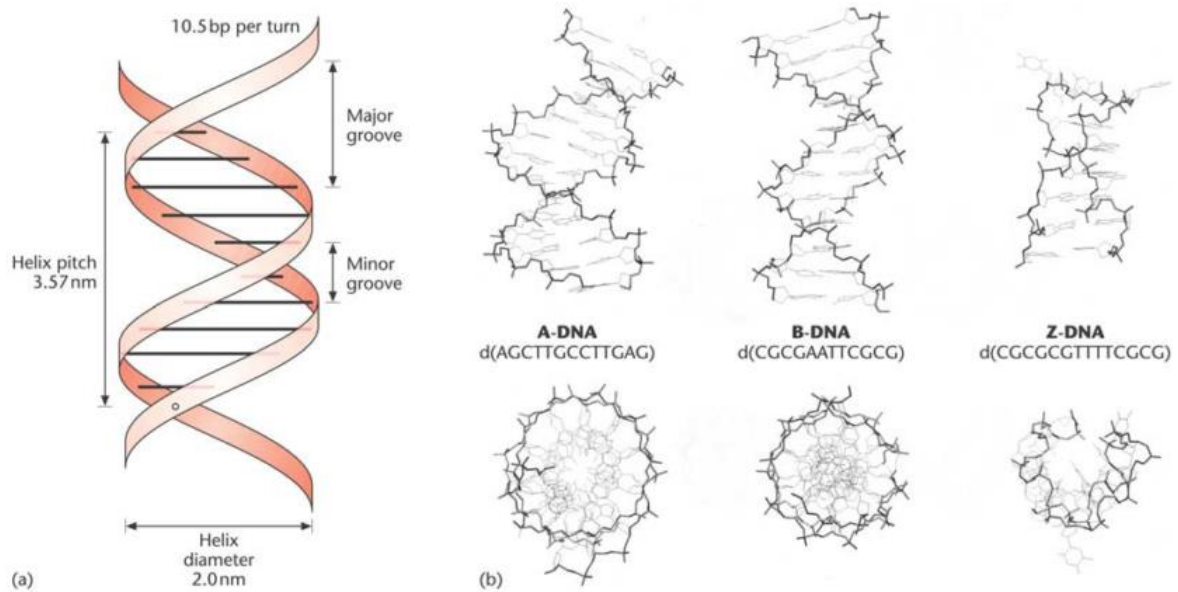


Figure 1.2.1.2. (a) Schematic of B-DNA as proposed by Watson and Crick and (b) views along the axis and down the axis of A-, B- and Z-DNA from crystal structures.⁵⁷

Within cells, DNA is mostly confined in the nucleus and is tied together by positively charged histones into a tertiary structure known as chromosomes. The histone proteins contain various epigenetic factors that are responsible for the opening and closing of the DNA-histone complex. This is key during cell division when the DNA has to replicate. Another important function of DNA is transcription and replication which are required for synthesis of proteins that are essential for cellular function. However, during the transcription of DNA in RNA, the complex only needs to open up the specific region where the gene to be transcribed is located.

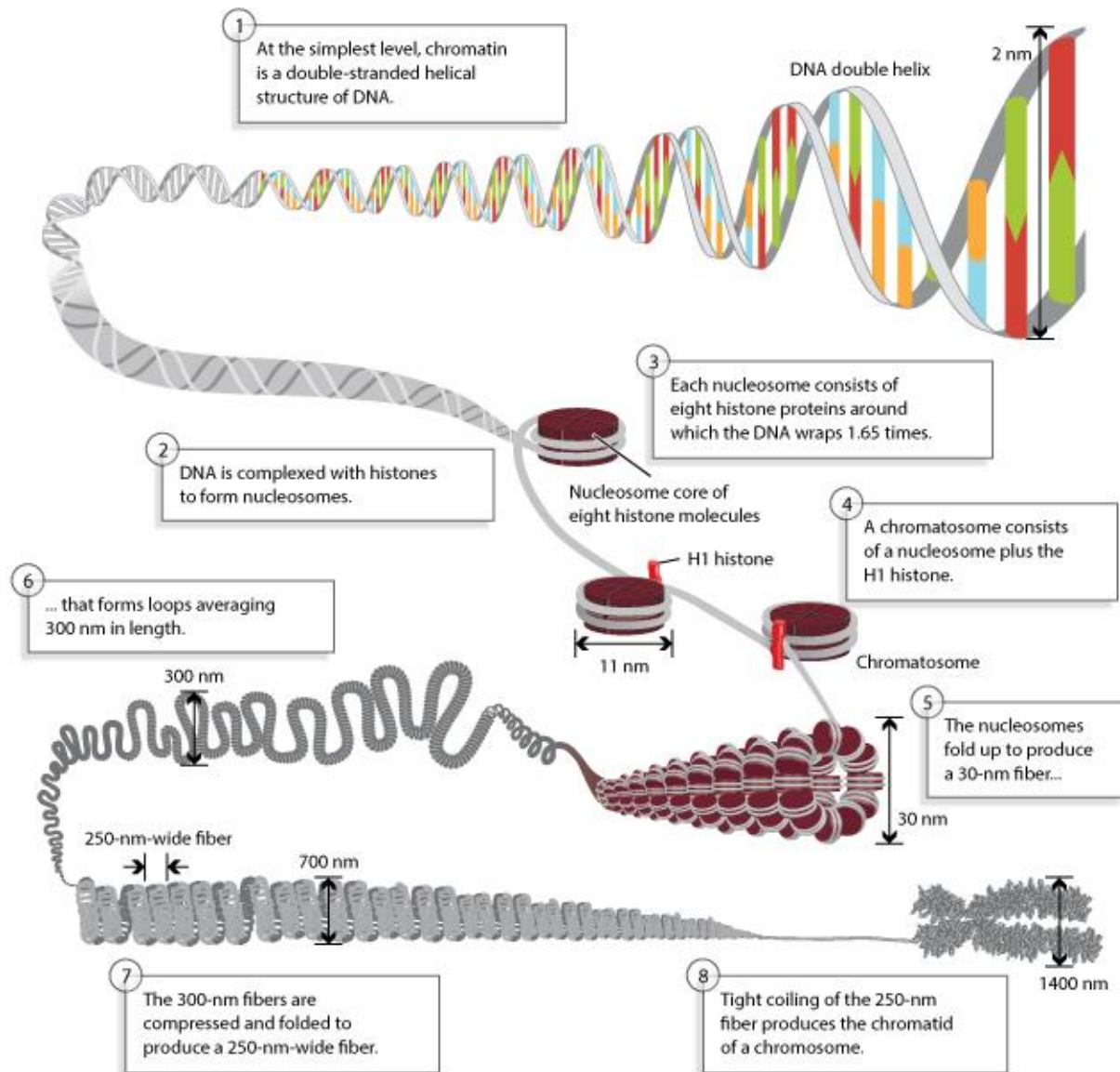
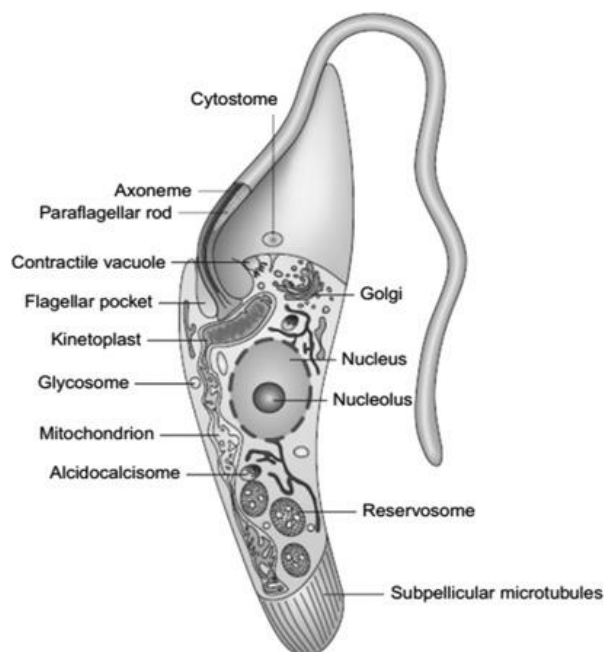


Figure 1.2.1.3. Schematic of DNA packaging, from DNA double helix to chromosomes.⁵⁸

DNA is also located in the mitochondria of a cell and is termed mitochondrial DNA (mtDNA). Within the mitochondria, mtDNA resides in nucleoids in the surrounding cytoplasm. Unlike nuclear DNA, mtDNA is a circular plasmid, encoding 33 genes responsible for mitochondrial function. In protozoal organisms, the mitochondria contain kinetoplasts which are giant networks of circular DNA in the form of minicircles and maxicircles (Figure 1.2.1.4). kDNA represents approximately 30% of the total cellular genome of certain protozoans and is responsible for encoding ribosomal RNAs (rRNA), guide RNA (gRNA) and certain proteins important for cellular function.⁵⁹ The selectivity of certain DNA binding agents towards

protozoans versus human cells is thought to be due to the selective uptake of these drugs into the kinetoplast which is absent in mammalian cells.

a)



b)

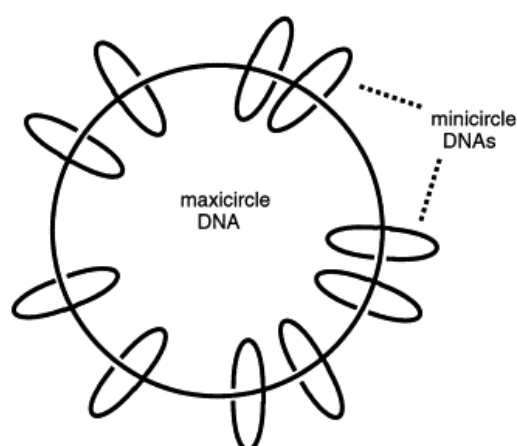


Figure 1.2.1.4. (a) Schematic representation of a longitudinal section of an epimastigote showing the main structures and organelles found in trypanosomes; and (b) illustration of the topological form of kinetoplastid mitochondrial DNA of trypanosomes.^{60, 61}

1.2.2. DNA targeting agents

Various forms of DNA targeting agents have been used as effective chemotherapeutic agents. These include intercalators, alkylators, strand-cleaving agents and minor groove binders (Figure 1.2.2.1). Depending on the specific mode of binding these agents serve to inhibit or strengthen the binding of DNA-binding proteins to the DNA duplex which is required for cellular processes such as transcription or replication. These essential cellular processes are perturbed through binding directly to the DNA-binding protein, to the protein-DNA ternary complex or to the DNA itself.

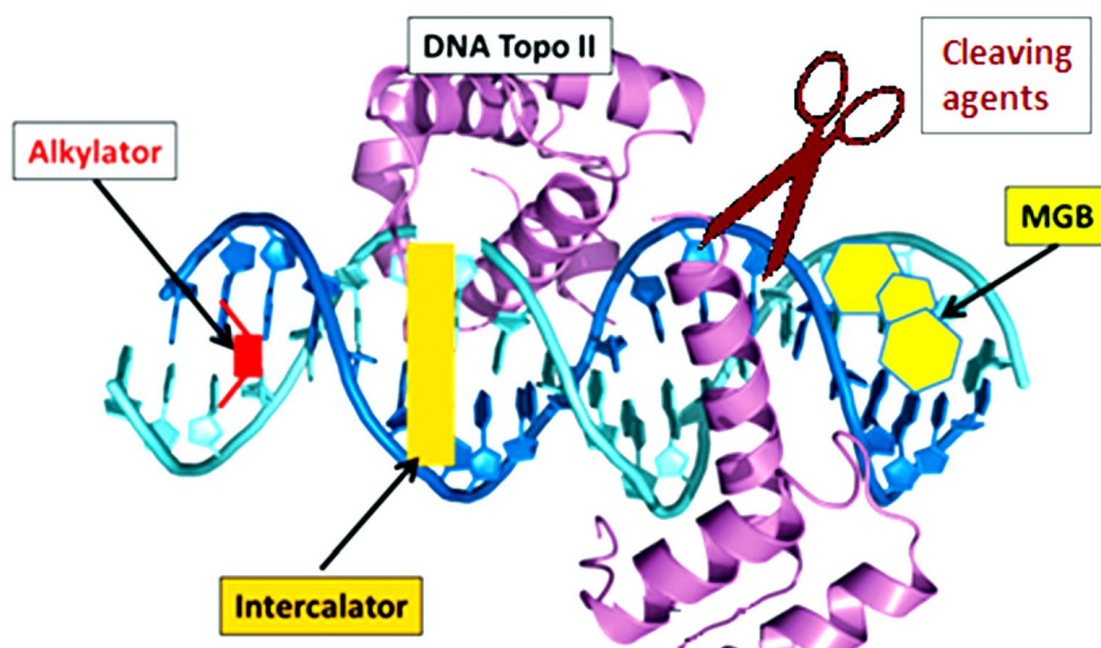


Figure 1.2.2.1. Key DNA–drug interactions: intercalators, alkylators, cleaving agents and minor groove binders (MGBs).⁶²

An essential class of DNA binding proteins that have been extensively studied are the DNA topoisomerases. These have been well characterised and have shown to have key roles in triggering, controlling, and modifying a wealth of topological DNA problems during cell proliferation, differentiation, and survival. Two subtypes of topoisomerases, namely, type I and type II have been probed for therapeutic intervention and differ either by catalytic function, reaction mechanism or amino acid sequence and structure. In the late 1900s, a plant alkaloid, camptothecin (CPT) (Figure 1.2.2.2), was discovered to be a topo I inhibitor. Isolated from a Chinese tree, *camptotheca cuminata*, CPT was shown to display remarkable anticancer properties where its antibacterial activity was later discovered.⁶³ Due to its poor water solubility, extensive structural modifications were made to produce the FDA-approved belotecan.⁶⁴ Various other water soluble derivatives are currently under clinical investigation as anticancer, antiviral and anti-parasitic agents. Where topo I inhibitors prevent the single-strand DNA cleavage, topo II inhibitors prevent the cleavage of both DNA strands. The clinically relevant topo II inhibitor, etoposide, binds to both topo II and DNA forming a stable ternary complex.⁶⁵

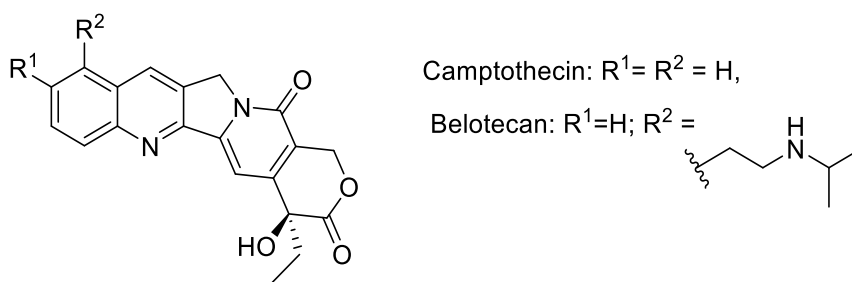


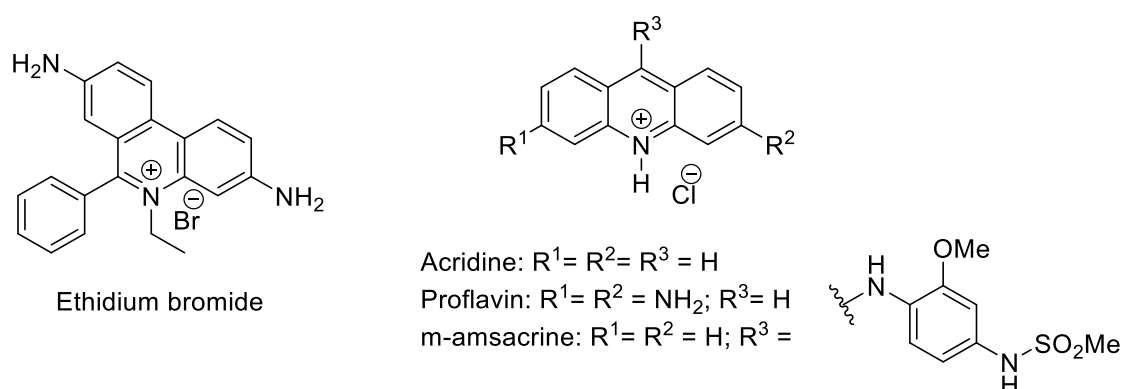
Figure 1.2.2.2. Structures of DNA topo I inhibitors camptothecin and belotecan.

Within certain microbial cells, DNA is in the form of circular or irregular shaped structures which are supercoiled. The compact aggregation of DNA needs to be relaxed or unraveled prior to replication and transcription. This requires a type of topoisomerase different from that found in mammalian cells known as DNA gyrase. In mycobacteria, DNA gyrase is the sole topoisomerase, unlike other bacteria, which utilize both DNA gyrase and topoisomerase IV.⁶⁶ This makes DNA gyrase a popular target for anti-mycobacterial drug design. The GyrA subunit of DNA gyrase is the clinically validated target of the fluoroquinolone class and has been used as a second-line treatment for various microbial infections.

Within recent years, there has been a huge focus on non-covalent DNA binding agents (i.e. intercalators and MGBs) as therapeutics. To understand the differences between these two DNA binders we will delve into their structural and functional properties.

DNA Intercalating agents were one of the first discovered DNA binding therapeutics. These important class of compounds are similar in structure bearing large, hydrophobic, planar scaffolds that can easily insert between DNA bases through a combination of π - π stacking and hydrophobic interactions. Examples of well-known intercalators include acridine, proflavin, ethidium bromide, doxorubicin, and m-amsacrine (Figure 1.2.2.3). Introduction of cationic functional groups to intercalators (e.g. proflavin) has been shown to strengthen DNA binding through the formation of electrostatic interactions and lowering of the LUMO to facilitate stronger π - π interactions.⁶⁷

a)



b)

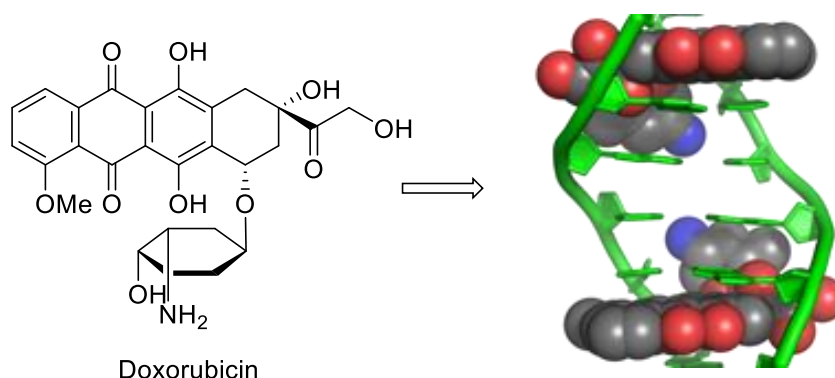


Figure 1.2.2.3. (a) Structures of intercalators: ethidium bromide, acridine, proflavin, m-amsacrine and (b) doxorubicin with a representation of its intercalation into DNA.⁶⁸

The intercalation of these flat polyaromatic compounds leads to the lengthening and unwinding of the DNA helix. It was originally thought that this binding-induced distortion of the DNA strand was their sole mode of action. It is now well understood that intercalators elicit their action through inhibition of topoisomerases. Consequently, the DNA binding of intercalators must be dynamic enough to allow the intercalator to detach from DNA and reattach to the DNA-protein binding region. Intercalators create a ternary structure with DNA and the topoisomerase after a single strand break occurs, rendering the enzyme unable to complete the double-strand breakage and sterically blocking any other enzyme from repairing the break. Once the DNA-topo complex is poisoned by the intercalator a sequence of cellular events occurs which ultimately leads to cell death.⁶⁹

Although several intercalators have been shown to be very effective non-covalent DNA binders, this research will specifically focus on the development of DNA minor groove binders (MGBs). Hence, we will delve deeper into their structural and functional properties.

1.2.3. DNA minor groove binders

Minor groove binders (MGBs) are another class of DNA binding compounds that have been exploited as chemotherapeutic agents for the treatment of cancer, parasitic infections, fungal infections and bacterial infections. The sugar-phosphate backbone of DNA is located on the outer surface of the DNA duplex. The spiraling nature of this backbone creates a helix which forms two grooves of different sizes known as the major and minor groove. The major groove is twice as wide as the minor groove measuring at 11.6 Å wide and 8.5 Å deep. Major grooves contain a high G:C base pair content whereas the minor groove is generally A:T rich. This inherently creates a greater positive potential in comparison to the negative electrostatic potential seen in the minor groove. This can be easily explained by considering the hydrogen bonding network in both cases (Figure 1.2.3.1). Within the minor groove, the hydrogen bonding between the A:T bases exposes hydrogen bond acceptors found in the N3 and O2 of adenine and thymine respectively. On the other hand, the major groove contains a lot more HB donors and acceptors with the addition of two protruding exocyclic amino groups of cytosine which changes the steric and electronic properties of the groove.⁷⁰

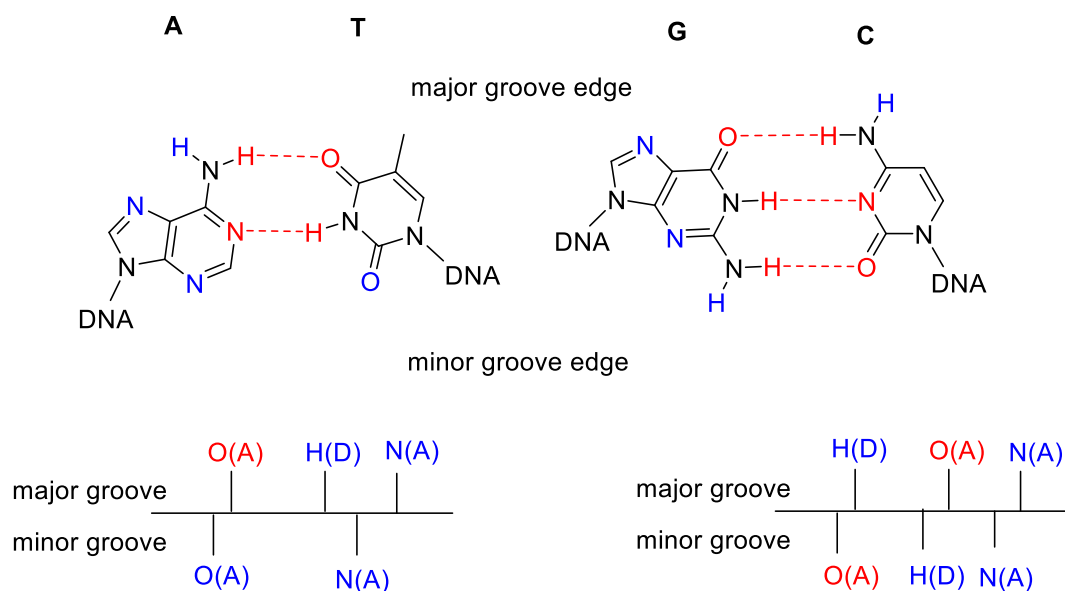


Figure 1.2.3.1. HB donors (D) and acceptors (A) within the major groove and minor groove available for ligand-DNA interactions.

Both the major and minor groove are surrounded by physiologically essential cations (e.g. Na^+ , K^+ , Ca^{2+} and Mg^{2+}) and contain water molecules that form a network of hydrogen bonds and is referred as the spine of hydration. Hydrophobic interactions that displace the spine of hydration cause a large increase in entropy. This is a strong driving force for the binding of a lot of DNA-binding proteins and enzymes. For this reason, the major and minor groove are interaction sites that play an important role in replication and transcription. It is important to note that DNA-protein interactions in the major groove do not lead to major alterations of the helical structure of DNA. Conversely, proteins that bind in the minor groove induce significant changes. For instance, the transcription promotor regions such as the TATA boxes are rich in AT base pairs which bind the TATA-box binding protein (TBT). Formation of the DNA-protein complex causes dramatic changes in the DNA structure and recruits various transcription factors leading to the transcription of certain genes.⁷¹ Hence, considering that the minor groove is the main interaction site for a number of DNA-dependent proteins and enzymes, compounds that target the minor groove have a clear therapeutic potential. Extensive research which has aimed to gain a better understanding of the characteristics of the minor groove has yielded the discovery and design a large amount of therapeutic minor groove binding agents.

MGBs have certain key characteristics that determine their affinity towards the minor groove. Some of these characteristics include a concave geometry, two or more planar aromatic rings, cationic in nature and have the ability to form hydrogen bonds. Aromatic rings are lipophilic and can therefore displace the spine of hydration upon binding which is entropically favourable. In certain cases the aromatic systems present in MGBs form a concave geometry which complements the convex shape of the minor groove. Cationic functional groups are favourable as they allow the MGBs to form electrostatic interactions with the negatively charged phosphates of the DNA backbone, as well as the intrinsic negative potential dispersed across the floor of the minor groove. Compounds that are rich in hydrogen bond donors, are beneficial due to the presence of the N3 of adenine and O2 of thymine at the bottom of the groove which is available to accept hydrogen bonds. Certain families of compounds fit these criteria. These include diaryl amidine-like derivatives such as pentamidine, furamidine, berenil and polyamide derivatives such as distamycin and netropsin (Figure 1.2.3.2).

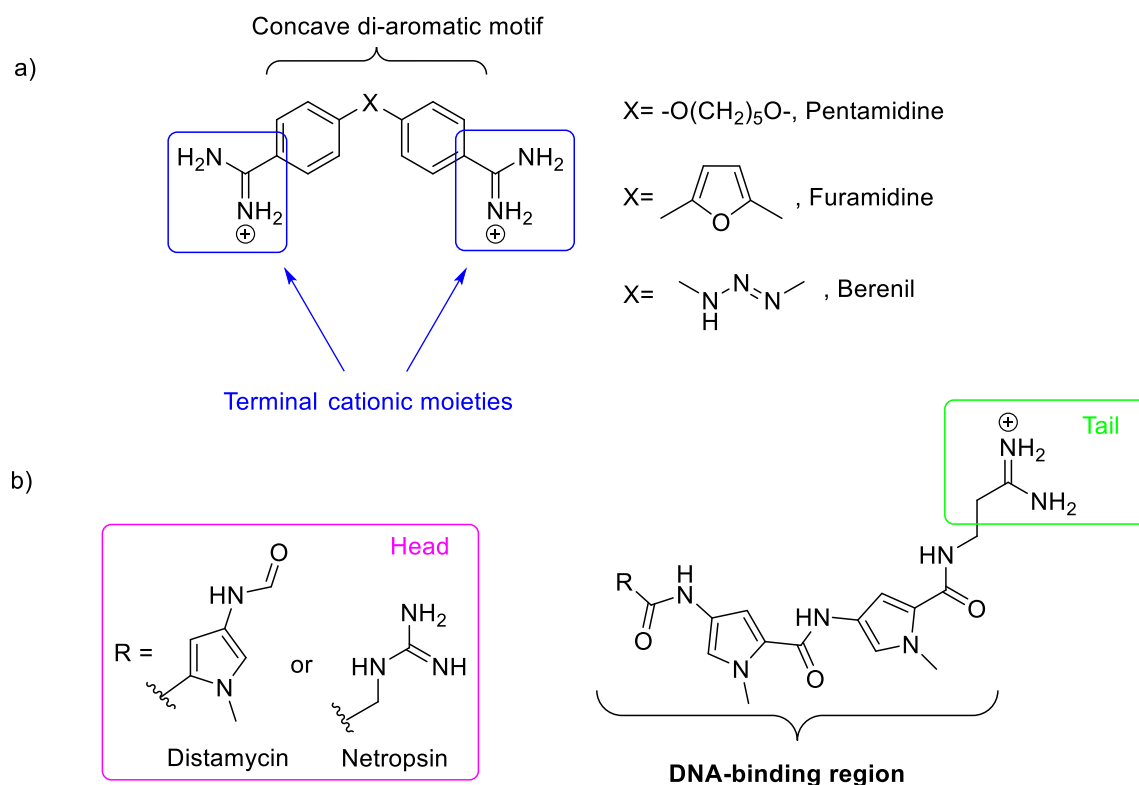


Figure 1.2.3.2. Structures of bis-amidines (pentamidine, furamidine and berenil) and polyamide (distamycin and netropsin) MGBs.

The polyamide derivatives, distamycin and netropsin, were among the first compounds found to bind the DNA minor groove. This is no surprise as they bear almost all the desirable features required for minor groove binding. Bearing multiple amide bonds, netropsin and distamycin are able to form three/four centred HBs with the O2 of thymine.⁷² Additionally, the terminal amidinium groups are capable of forming electrostatic and HB interactions within the groove. Although netropsin and distamycin have been used to treat malaria, derivatives have been designed for the treatment of bacterial and parasitic infections. Distamycin analogues containing *N*-terminal bi-aryl motifs (**1**, Figure 1.2.3.3, a) were assessed for their potential anti-mycobacterial activity where the thiophene derivative showed 10-fold higher inhibitory activity against *M. tuberculosis* than distamycin. Suckling's group has worked widely in developing polyamides as antimicrobial and antifungal agents. Among these include **2** and **3**, which exhibited MIC₉₉ of 4.09 and 419 μM against *M. tuberculosis* strain HN878, respectively, with low toxicity (Figure 1.2.3.3, b).⁷³ Further optimisation of these derivatives by replacement of a pyrrole for a thiazole unit and modification of the head group yielded **4**, **5**, **6** and **7** (Figure 1.2.3.3, c). Compounds **5** and **7** exhibited the best MIC₉₉ of 3.1 μM against *M.*

tuberculosis with limited toxicity. The authors concluded that the presence of the thiazole unit plays an important role in the anti-infective activity due to increased lipophilicity which aids cell permeability and creation of additional van der Waal interactions resulting in increased DNA affinity. Additionally, the Suckling group further studied the cytotoxicity of the aforementioned compounds against *T. b. brucei*. They found five compounds with nM activity ($IC_{50} > 40$ nM) that became new hits for optimisation towards novel treatments for HAT.⁷⁴

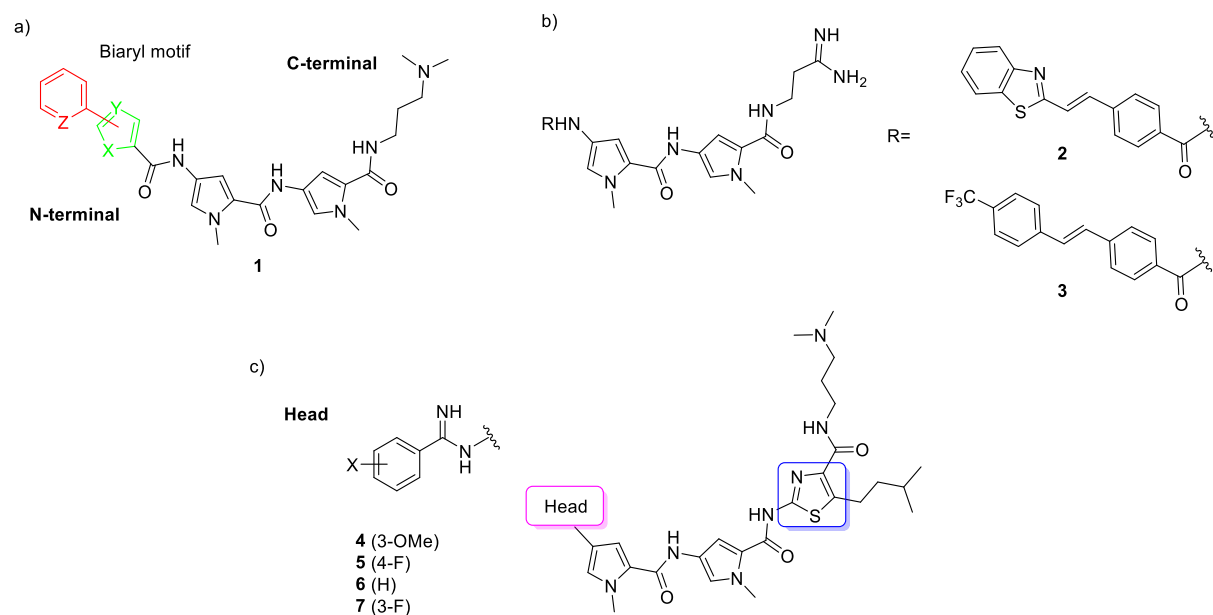


Figure 1.2.3.2. Structures of polyamides **1-7** developed as MGBs with anti-infective activity.

Bis-amidine MGBs as shown those in Figure 1.2.3.2 have a long track record as anti-parasitic agents. Pentamidine, furamidine and berenil have been used to treat malaria, trypanosomiasis and leishmaniasis. Pentamidine has specifically been used in stage-1 HAT caused by *T. brucei gambiense*, but as it does not cross the blood-brain barrier it is not as effective in the treatment of late stage disease. For this reason, a more orally bioavailable derivative, furamidine was developed. More specifically, pafuramidine, the *N*-methoxy prodrug of furamidine was analysed in clinical trials for the treatment of late-stage HAT. However, nephrotoxicity was observed in phase I and similar efficiency to injected pentamidine was observed in phase III trials; for this reason, the development of pafuramidine was halted.⁷⁵ Due to these pitfalls, a number of second-generation anti-

parasitic di-aryl amidine-like derivatives have been investigated. These include novel bis-amidines, bis-benzimidazoles, bis-guanidines and bis-2-aminoimidazolines. Sauer *et al.* has prepared compounds by taking the furamide scaffold and introducing modifications at the furan core as well as the amidine groups (**8**, Figure 1.2.3.4). These derivatives showed very good activity (low μM IC_{50} values) against drug-sensitive and drug-resistant *P. falciparum*, with low cytotoxicity against human HEK293 cells.⁷⁶ Several aza-analogues of furamide have also been reported to cure animal models of late-stage HAT. Specifically, pyridine derivative **9** was shown to display an exceptionally improved safety profile and has been considered as a potential candidate for clinical evaluation.⁷⁷ Additionally, Wilson and coworkers have prepared a family of indole and benzimidazole containing bis-amidines which have been shown to bind strongly to DNA and display very good *in vitro* and *in vivo* anti-trypanosomal activity.⁷⁸ Interestingly, these compounds bear structural similarity to fluorescent MGBs such as DAPI and Hoechst dyes.

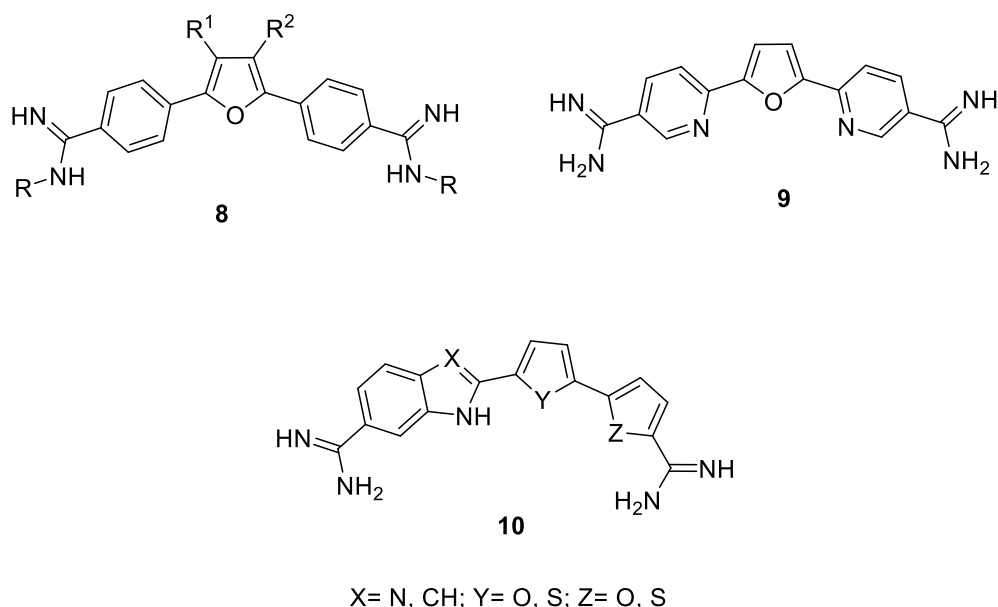


Figure 1.2.3.4. Structures of furamide derivatives **8**, **9** and **10**.

Furthermore, bis-benzimidazole derivatives such as **DB173** (Figure 1.2.3.5) have been developed which were shown to be highly active against *Naegleria fowleri*, the amoebae parasite that causes primary amoebic meningoencephalitis (PAM).⁷⁹ In another study conducted by Wilson *et al.*, **DB2120** was produced bearing a bis-benzimidazole motif with terminal amidines. Substantial lengthening of the linker by the introduction of the 2,6-(methyl

ether)pyridine core resulted in high selectivity for AAAAGTTT regions, aided by hydrogen bonding between the pyridine nitrogen and the guanine amino group.⁸⁰

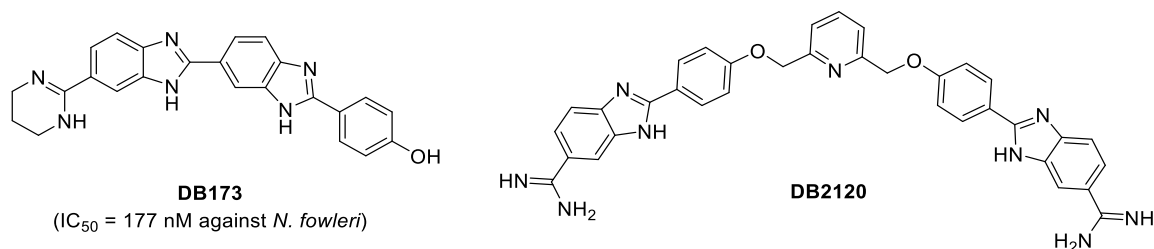


Figure 1.2.3.5. Structures of bis-benzimidazole containing MGBs **B173** and **DB2120**.

Dardonville *et al.* have described the synthesis, biophysical evaluation and anti-trypanosomal activity of guanidine/2-aminoimidazoline MGB agents. A library of alkane, diphenyl and aza-alkane bis-guanidine/bis-2-aminoimidazolines which bear structural similarity to trypanocidal agents synthalin and 4,4'-bis-guanidino-diphenylmethane showed activity against *T. b. rhodesiense* (Figure 1.2.3.6).⁸¹ These compounds displayed low μ M to high nM IC₅₀ values with high selectivity for the trypanosome versus healthy human L6 cells. Due to the greater DNA binding affinity and anti-trypanosomal activity observed for the diphenyl derivatives, a series of *N*-alkyl, *N*-alkoxy and *N*-hydroxy bis-guanidine derivatives of the *N*-phenylbenzamidines (**11**) and 1,3-diphenylurea scaffolds (**12**) were prepared. Of these derivatives, the *N*-alkyl analogues displayed sub μ M and low nM activity against *P. falciparum* and *T. b. rhodesiense*, respectively. Further *in vivo* analysis showed that the *N*, *N'*-diethyl and *N*, *N'*-diisopropylguanidino phenylbenzamidines displayed high efficacy (100% cures) in the STIB900 mouse model of acute HAT.⁸² In collaboration with Dardonville, our group prepared an in-house library of guanidine and 2-aminoimidazolines which show strong DNA binding and good anti-parasitic activity in both *in vitro* and *in vivo* experiments. These compounds will be discussed further in Section 1.3.

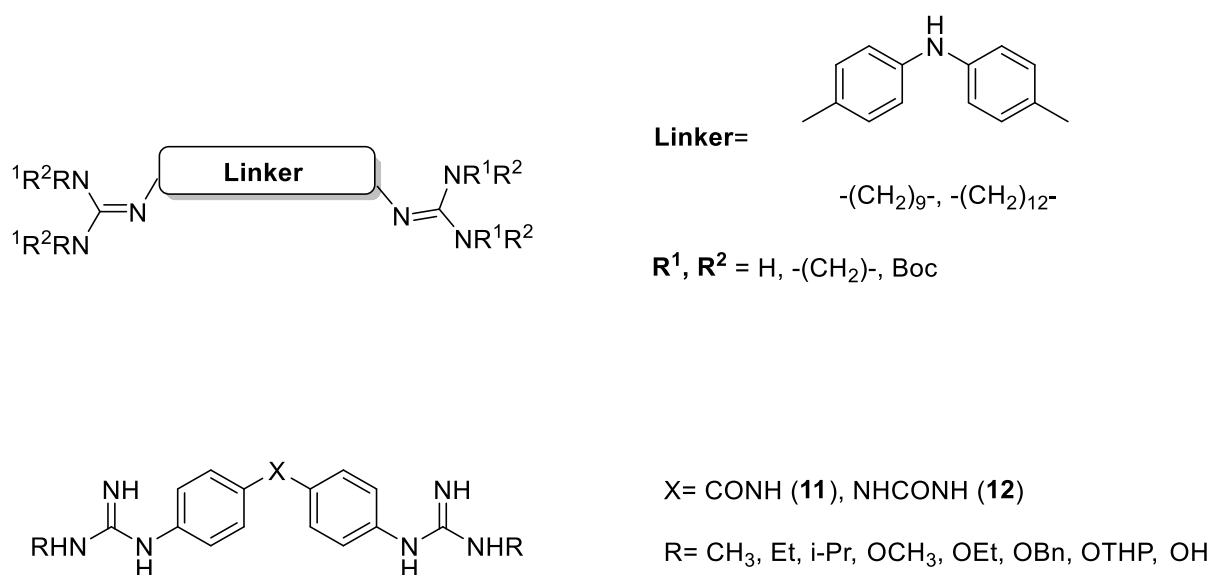


Figure 1.2.3.6. Structures of guanidine-like derivatives produced by Dardonville *et al.*⁸¹

1.2.4. Characterisation of DNA binding

Various biophysical techniques can be employed to characterise the binding of a compound to DNA. Information such as mode of binding and binding affinity can be elucidated from these methods. These techniques, among others, include DNA UV-thermal denaturation, UV-Vis titrations, Circular Dichroism (CD) and Linear Dichroism (LD) which all rely on the absorption of light.

Optical techniques such as DNA UV-thermal denaturation is often used as an initial determinant for DNA binding affinity. Rapid qualitative and quantitative data generated from this method allows for the ranking of compounds according to their DNA binding affinity. The principles of this method takes advantage of the different UV absorbance of double-stranded and single-stranded DNA. When the DNA sample is heated, the energy input breaks the intermolecular forces that hold the double helix together resulting in two single strands. This is referred to as the melting or denaturation of DNA. The melting temperature (T_M) describes the temperature at which half of the DNA is in the double-stranded form and half is in the single-stranded form. T_M can therefore be used as a metric to assess the effect of DNA binding agents. When a molecule binds to DNA, it stabilises the double helix, thus extra energy is

required to split the DNA double helix resulting in an increase in the T_M . This increment in T_M (ΔT_M) is a measure of a compound's affinity towards DNA. The value of ΔT_M is dependent on experimental conditions such as salt concentration, buffer composition, pH, rate of heating, the composition of the DNA oligomer and ratio of DNA base pair to drug (expressed as P/D = phosphate to drug ratio).⁸³

Similar to DNA UV-thermal denaturation, UV titrations can be used to identify and determine the strength on DNA binding. However, these experiments are conducted at constant temperature (room temperature) in an appropriate buffer with the only variable being DNA concentration. Upon addition of DNA aliquots to a specific concentration of ligand spectral changes occur. These changes include hypochromic shifts which represent a decrease in free ligand concentration and are commonly accompanied by a blue or red shift due to the formation of a DNA-ligand complex. Binding constants can easily be calculated by extrapolation of the resulting absorbance data and construction of the corresponding binding saturation curves.

CD and LD are routinely used to determine the mode of binding to DNA. Although the nucleotide bases are planar they are in a chiral environment due to the right-handed helix formed and their proximity to the chiral deoxyribose sugars. As a result, they absorb circularly polarised light differently. CD takes advantage of this by measuring the differential absorption of left and right circularly polarised light to produce a CD spectrum. Unlike other optical techniques, CD spectra can be used to distinguish between the various forms of DNA (Figure 1.2.4.1).⁸⁶ Generally, DNA binding ligands are achiral and are therefore optically inactive. However, upon binding to DNA it absorbs asymmetric electronic transition moments when the complex is excited resulting in the production of an induced CD (ICD) signal. Therefore, the presence of an ICD can be used to confirm a ligand-DNA interaction. Minor groove binders generally display a large positive ICD where intercalators display a small negative ICD ($<10 \text{ M}^{-1} \text{cm}^{-1}$).⁸⁴ Additionally, CD can be used to measure the binding constant of ligands towards DNA. However, the use of this technique to calculate binding constants is heavily concentration-dependent. At high concentrations, DNA binding ligands can stack on top of each other resulting in a large exciton coupling of two identical electronic transitions which gives rise to a large ICD.⁸⁵ This results in inaccurate binding constant calculations.

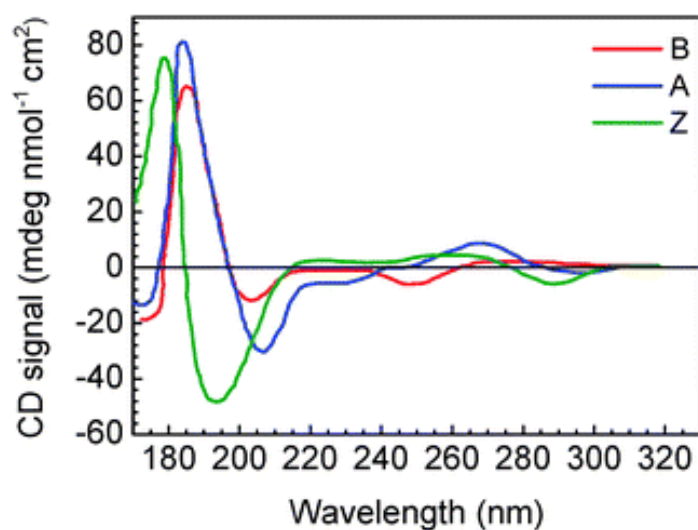


Figure 1.2.4.1. CD spectra of A-, B-, and Z-DNA.⁸⁶

Similarly, LD can be used to determine the mode of binding towards DNA. It takes advantage of the differential absorption of parallel and perpendicular polarised light in flow. A positive induced LD signal indicates minor groove binding whereas a negative ILD signal indicates intercalation (Figure 1.2.4.2). This method is particularly useful in cases of dual binding compounds that simultaneously groove bind and intercalate.^{87, 88}

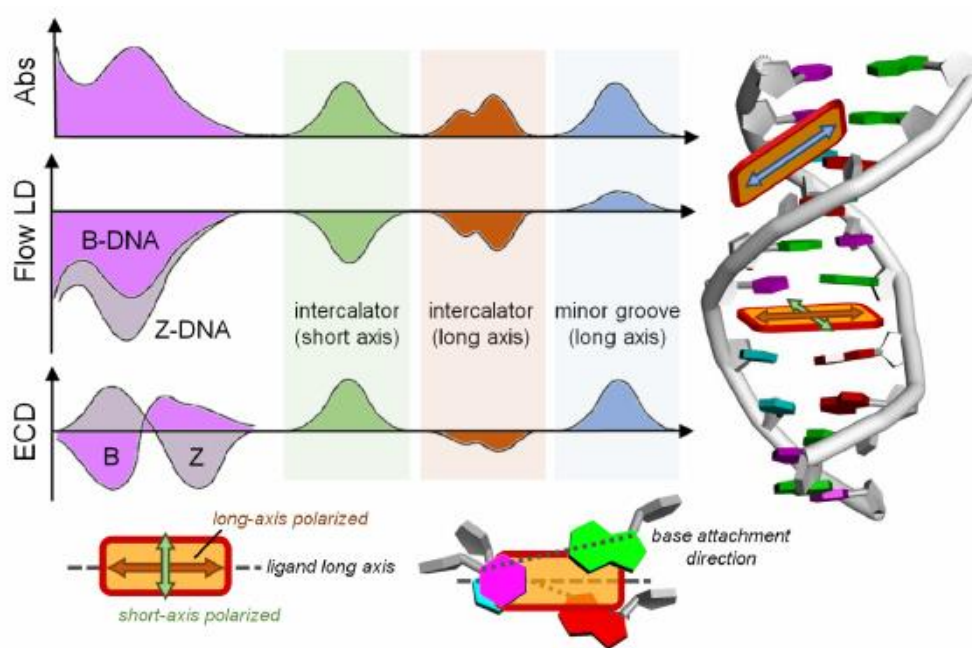


Figure 1.2.4.2. Empirical analysis of ECD and flow LD spectra to establish the preferred orientation of a ligand bound to DNA.⁸⁸

As well as the techniques mentioned above, certain non-optical techniques such as Isothermal Calorimetry (ITC) and Surface Plasmon Resonance can also be used to assess DNA binding. SPR is commonly used to measure the binding constants to specific DNA hairpin strands. When the oligonucleotide of choice is immobilised onto a sensor chip (usually gold), light is shone on the surface of the chip which is then reflected. Upon injection of a compound onto the sensor and exposure to light, a change in the refractive index occurs due to DNA binding. In simple terms, molecules with a strong binding affinity produce large signals whereas weak binders give weak signals. Steady-state binding analysis can be carried out with multiple injections of different compound concentrations over the immobilised DNA surface which generates real-time kinetics of a complex's association and dissociation.⁸⁹

ITC is a useful technique used to measure the thermodynamics of binding. This is achievable due to changes in enthalpy, entropy and Gibbs free energy when compounds bind to DNA. Based on this principle binding stoichiometry, binding constant and the enthalpy of binding can be determined. In combination with structural and spectroscopic data, a full picture of the specific interactions and geometry of binding of ligands in DNA can be elucidated.

1.2.5. Elucidation of mechanism of action – emerging drug discovery technologies

In the early days of drug discovery, some of the current anti-infective agents that are still used today were discovered through hit generation from large compound libraries with the use of whole-cell high throughput screening. This represents an irrational approach to the production of efficient drug therapeutics which is costly and time consuming. Additionally, due to minimal information given on how these hits are acting to produce their therapeutic effect certain issues arise such as the development of off-target side effects, poor efficacy and drug resistance. Within recent years, more rational approaches have been employed which include the use of structure-based drug design (SBDD) and ligand-based drug design (LBDD) (Figure 1.2.5.1). SBDD analyses the 3-dimensional macromolecular target structure to identify key sites and interactions that are important for its biological function. This information is then utilised to design drugs that can compete with essential interactions involved in the

normal function of the target, hence interrupting the biological pathways essential for the survival of the microorganism. LBDD focuses on ligands with known targets to establish a structure-activity relationship (SAR) generated from information on their physicochemical properties and biological activity. Thus, this information can be used to optimise known drug compounds or guide the design of new drugs with improved activity.⁹⁰

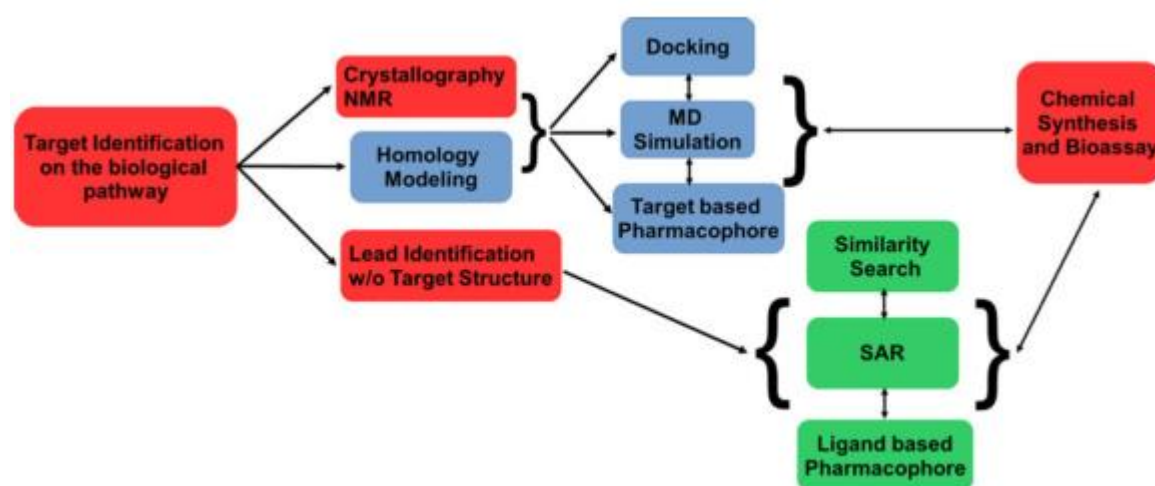


Figure 1.2.5.1. Basic workflow for computer aided methods SBDD and LBDD.⁹⁰

Modern day drug design with the use of the computer-aided methods mentioned above has reduced the cost and time taken to generate drug compounds that can eventually reach the market. However, due to the extremely large proteome and the complex architecture of microorganisms, these methods may fail to identify polypharmacological effects and potential loss in efficacy due to drug resistance mechanisms. As illustrated in Figure 1.2.5.1, the mid-19th century marks the golden age for the discovery of most of the antibiotics used today, where their molecular targets and biological activities were just defined. However, from 1970 onwards, an innovation gap in medicinal chemistry arose which clashed with the occurrences of drug resistance. This led to the search for second-line agents in a time where the rise in these infectious diseases warranted immediate action.

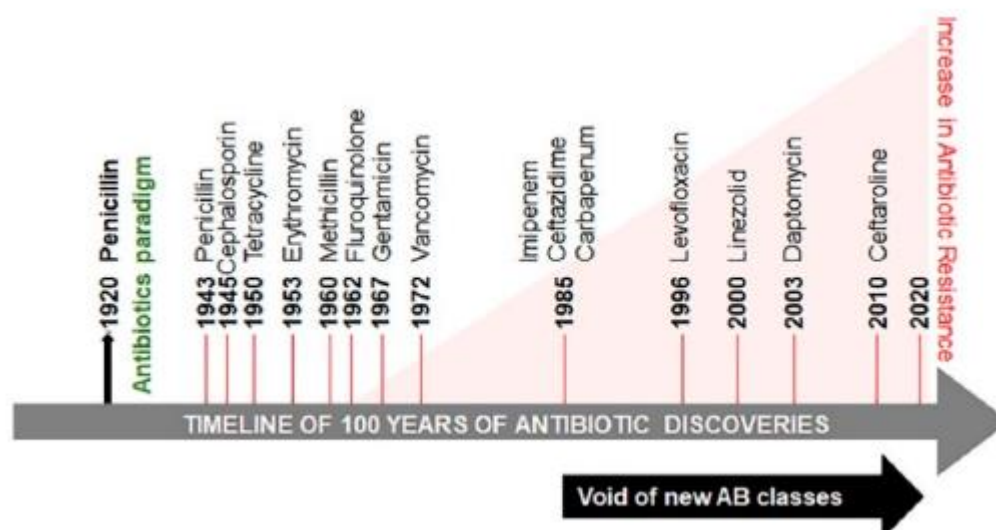


Figure 1.2.5.2. Timeline of antibiotic discovery and the emergence of drug resistance.⁹¹

To fix the mistakes of the past and change the way drug design and development is approached, a large body of work has been focusing on the use of whole genome analysis in combination with traditional drug design methodologies. Due to the tremendous amount of data generated by the rise of the genomic era, a fresh perspective in our understanding of chemical-biological interactions has been established. This fusion of chemoinformatics and bioinformatics has been termed "Chemobioinformatics".⁹²

Whole-genome sequencing (WGS) approaches have been used to study the genomes of a large array of microorganisms. More specifically, WGS technologies are currently used as a tool to guide all components of TB control including diagnosis, treatment, surveillance and source investigation. The strength of these technologies have caused its transition from research laboratories to clinical care and public health applications. As shown in Figure 1.2.5.3, the standard workflow involves culturing microbes on solid or liquid media, extraction of DNA from cells, library preparation and sequencing using short read technologies (e.g. Illumina platforms). Analysis of the resulting output involves several key steps including data validation, quality control followed by mapping to a reference genome and detection of genomic variants such as single nucleotide polymorphism (SNPs), deletions and insertions. Based on the resulting variants detected information on drug resistance and susceptibility profiles, strain typing and identification of transmission clusters can be obtained.⁹³

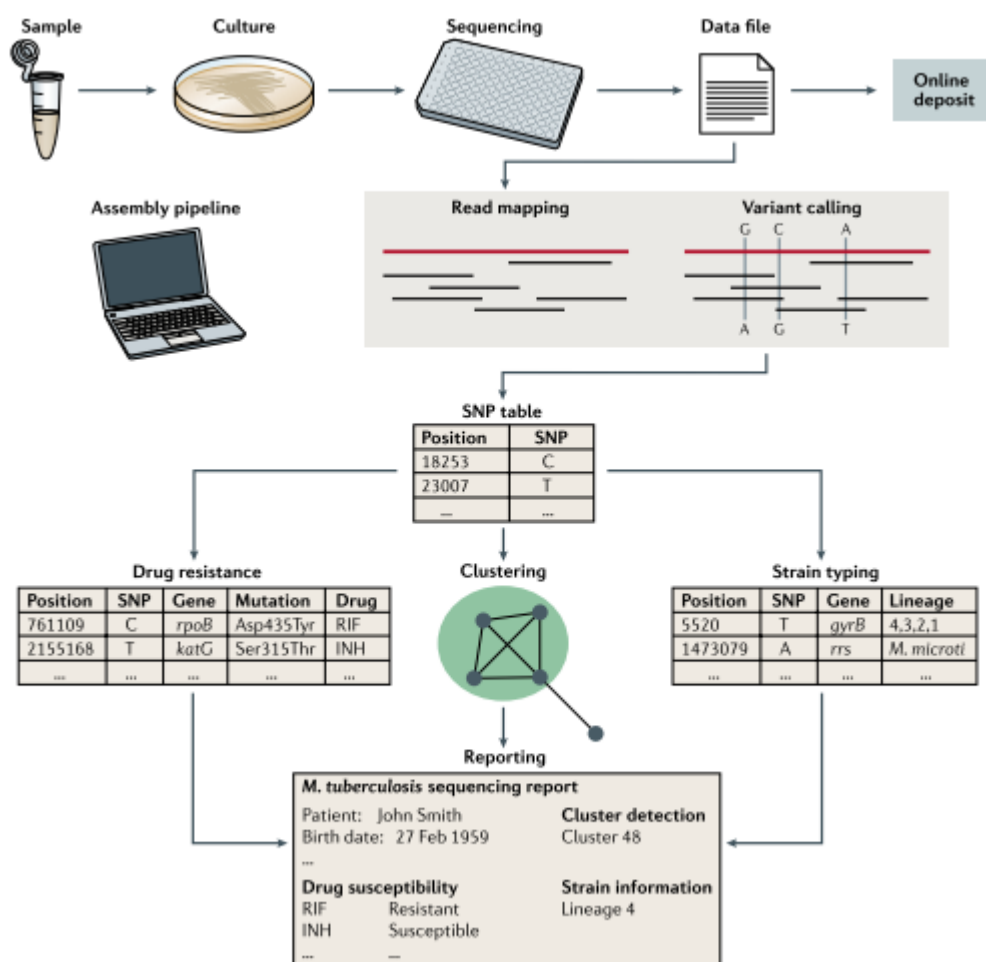


Figure 1.2.5.3. Standard workflow for whole-genome sequencing of *Mycobacterium tuberculosis* complex isolates.⁹³

As stated previously, the various drug resistance mechanisms used by microorganisms include drug efflux, limitation of drug uptake, target by-pass, drug inactivation and modification of the drug target. Therefore applying WGS at the preclinical stage of a drug design and development process has the potential of informing us about what transport proteins are being used for the uptake or efflux of the drug compound. Additionally, information on the activation of inactivating proteins, activation of alternative pathways and most importantly modifications within the target protein(s) can be obtained. It is key to note that all this information can be gained prior to clinical analysis and approval. By target confirmation and resistance mechanism elucidation through variant analysis, we have the ability to design more potent anti-infective agents and the development of second-line agents before the initial drug compounds have hit the market.

A study conducted by Pose *et al.* showed that the pyrazole containing derivative, **BM212**, targeted the mycobacterium membrane protein Large 3 (MmpL3) with the use of WGS.⁹⁴ Similarly, a recent study by Abrahams *et al.* showed that the mechanism of action of a novel anti-mycobacterial agent, **GSK3011724A**, was through inhibition of the enzyme KasA through identification of SNPs within the *KasA* gene in resistant mutants.⁹⁵ Additionally, Harrison *et al.* showed that a series of 4-amino-thieno[2,3-d]pyrimidines (e.g. **CD37**) targeted the electron transport chain enzyme cytochrome bc1 (QcrB) due to the presence of a nonsynonymous mutation in the gene *qcrB*.⁹⁶ The use of WGS in this context will be described Chapter 4.3.

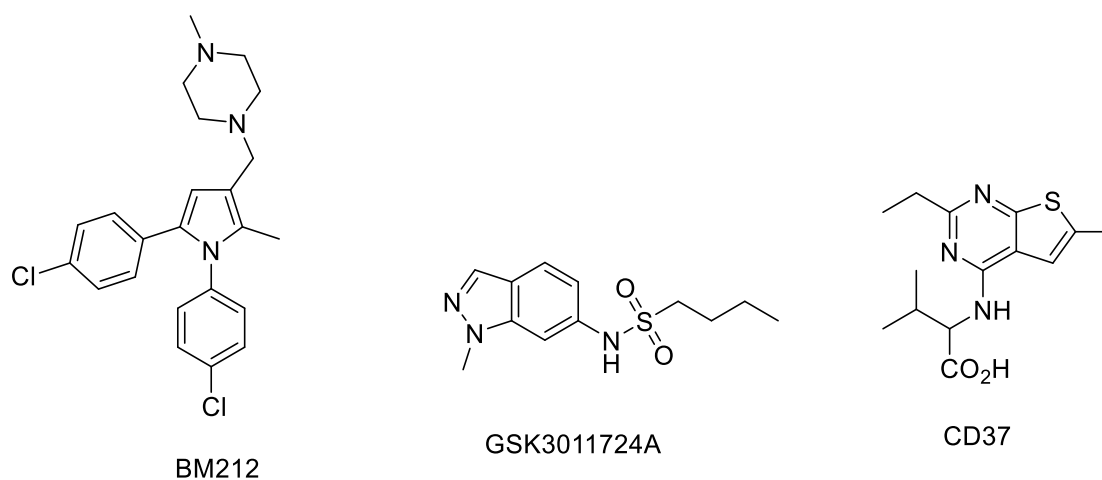


Figure 1.2.5.4. Structures of anti-mycobacterial agents; **BM212**, **GSK3011724A**, **CD37**.

1.3. Previous work that constitutes the basis of this research

As highlighted above, there is a need for novel drug compounds for the treatment of infectious diseases. This research will be split into two parts which will focus on guanidine based DNA minor groove binders as potential anti-protozoal agents and the investigation of guanidine based derivatives as anti-mycobacterial agents for the treatment of TB.

1.3.1. Guanidine based minor groove binders as anti-protozoal agents

During the last 20 years, the Rozas group has prepared a large number of di-aromatic symmetric and asymmetric guanidine/2-aminoimidazoline compounds that have been shown to be strong DNA MGBs (Figure 1.3.1.1).^{81, 97, 98, 99, 100} These compounds have a similar scaffold: two aryl rings connected by various linkers (e.g. CO, NH, O etc.) and para-substituted with guanidinium-like cations. A clear trend in DNA binding affinity was observed with the earliest set of MGBs where the bis-2-aminoimidazolinium (Family B) derivatives displayed the strongest binding followed by the asymmetric guanidinium/2-aminoimidazoliniums (C) and bis-guanidiniums (A).^{81, 98} This was attributed to the increase in lipophilicity across the three families which introduces additional van der Waal interactions upon binding to the minor groove. Additionally, the importance of the linker was apparent by the stronger DNA binding affinity observed for compounds with CO or NH linkers. This was postulated to be due to the geometry of these linkers which brings the phenyl rings close to an almost co-planar conformation. Therefore, due to this pre-organisation, the energetic penalty is lower for binding to the minor groove in the favoured planar geometry as indicated by previously published crystal structures. The superior DNA binding affinity observed for NH-linked derivatives of Family A and B led to further anti-protozoal analysis which showed low μM *in vitro* activity in *T. b. rhodesiense* and 100% cure in acute models of HAT. Due to the observed trend in lipophilicity, bis-(2-amino-1,4,5,6-dihydropyrimidinium) derivatives (Family D) were produced with the aim of producing compounds with even greater DNA binding affinity.⁹⁴ However, biophysical studies showed that this family has weaker DNA binding affinity than Family B and comparable binding affinity to Family C. Hence, it was concluded that the size and properties of the cyclic cationic moieties has an optimal limit which is that of the 2-aminoimidazolinium cation.

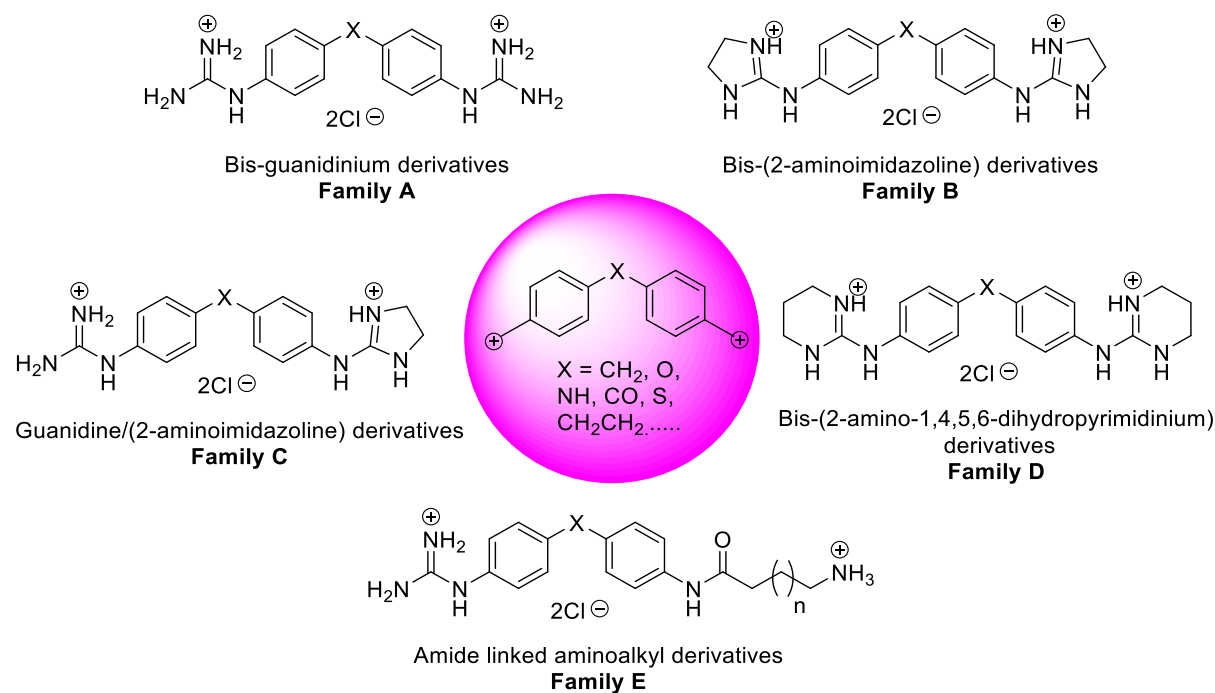


Figure 1.3.1.1. Structures of DNA MGBs produced by the Rozas group.

Incorporation of larger lipophilic groups was shown to be tolerated by the introduction of an aliphatic chain as shown for the aminoalkyl derivatives (Family E). These derivatives showed very promising anti-malarial activity with IC_{50} values as low as 100 nM against *P. falciparum*.⁹⁷ However, the structure-activity-relationships (SARs) of these compounds followed two patterns indicating different mechanisms of action: Derivates with alkyl chain legnth of four to seven methylene carbons that showed a good activity/DNA-binding correlation indicating that their activity involves DNA-binding; and shorter alkyl chain derivatives (<4 methylene carbons) that showed poor correlation between DNA binding affinity and anti-malarial activity. Based on these disparities further investigation into the mechanism of action of these amino-alkyl derivatives was conducted in this work. Additionally, a new family of compounds (Figure 1.3.1.2) were designed with the aim of creating stronger and more selective MGBs by combining structural components of previous families that gave promising results.

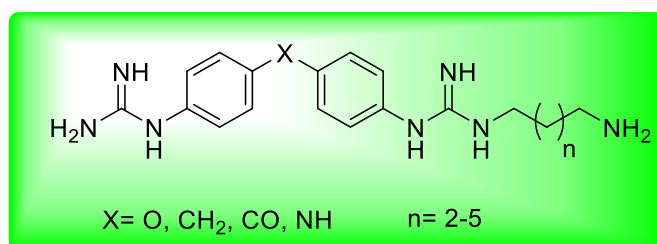


Figure 1.3.1.2. Novel aminoalkyl derivatives of di-aryl bis-guanidines.

1.3.2. Guanidine based derivatives as anti-mycobacterial agents

Nitrogen heterocycles remain an attractive topic for small molecule drug design and discovery. Among all heterocycles, piperazines (PZs) and piperidines (PDs), six membered nitrogen-containing heterocyclic rings, are certainly an established important pharmacophore in medicinal chemistry. PZ and PD derivatives are reported to elicit a broad spectrum of biological activities and can be witnessed in numerous well known drugs, belonging to diverse pharmacological classes.¹⁰¹ Various examples of anti-tubercular agents containing these rings have been shown to target essential mycobacterial enzymes such as matrix metalloproteinases, decaprenylphosphoryl-b-D-ribose-2-epimerase (DprE1/E2), protein kinases, NADH dependent enoyl-acyl carrier protein reductase (InhA) and DNA gyrase (GyrA) (Figure 1.3.2.1).

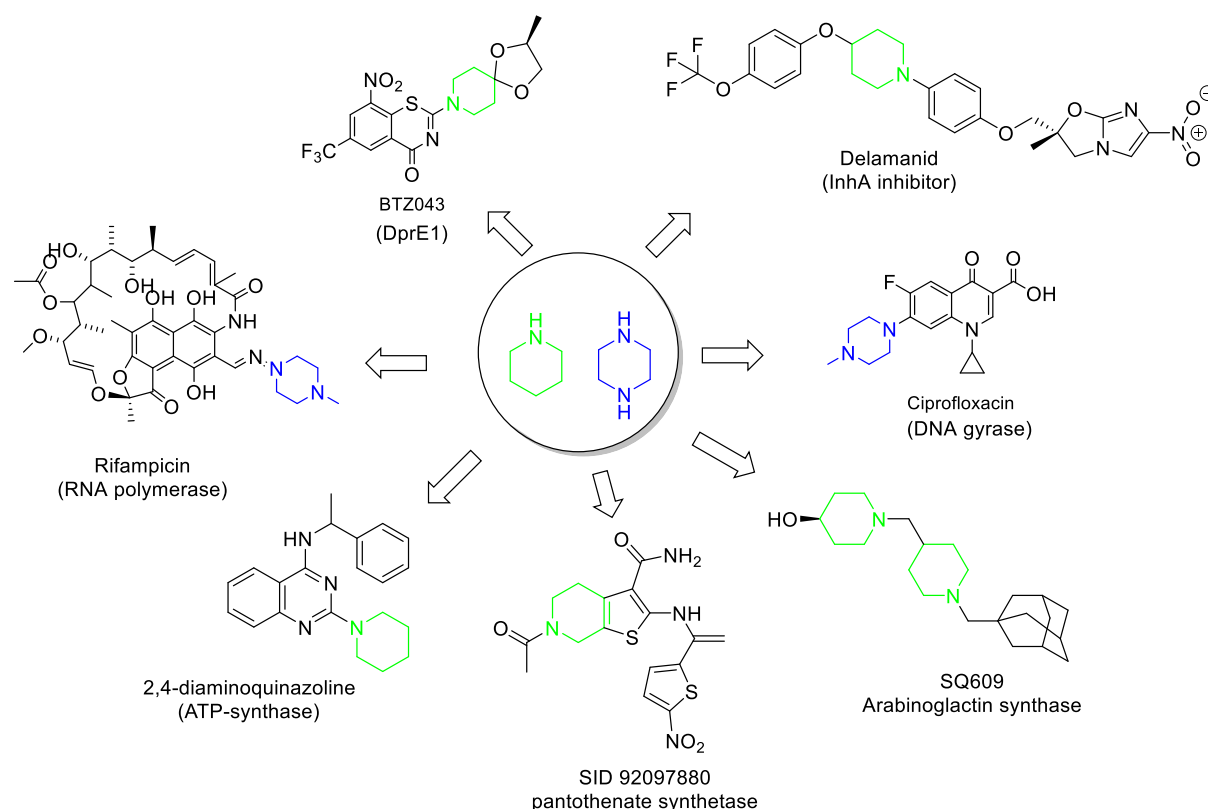


Figure 1.3.2.1. PZ and PD containing anti-tubercular agents.

Based on the apparent activities of these scaffolds, PZs and PDs have been widely used in the literature as building blocks for anti-tubercular agents. A study conducted by Forge *et al.*, reported a series of di-aryl piperazine analogues that showed good anti-mycobacterial activity against drug-sensitive ($\text{MIC}_{50} = 0.052 - 0.264 \mu\text{M}$) and drug-resistant ($\text{MIC}_{50} = 0.95 - 3.85 \mu\text{M}$) strains of *M. tuberculosis* (Figure 1.3.2.2). These compounds also displayed good selectivity indices as measured by the minimum inhibitory concentration against *M. tuberculosis* strain H37Rv and that against C3A hepatocytes. The most promising compounds **13-18**, bearing alkyl-substituted amidine functional groups were shown to kill intracellular *M. tuberculosis* residing in mouse J774A.1 macrophages.¹⁰²

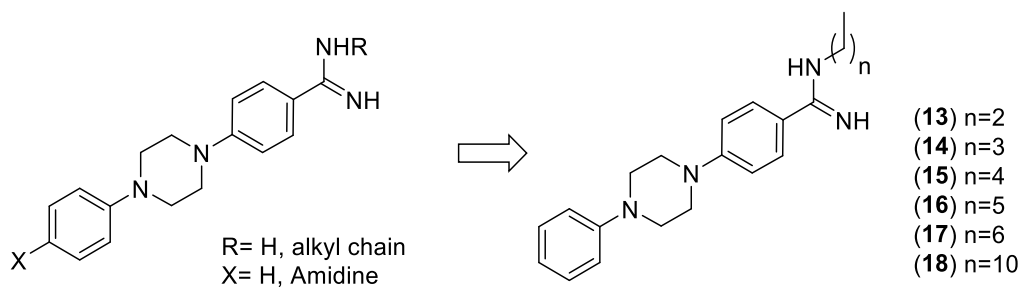


Figure 1.3.2.2. Structures of di-aryl piperazine based compounds produced by Forge *et al.* with anti-mycobacterial activity.

Even though a clear target was not mentioned in the study and has yet to be established, a structure-activity relationship study yielded potential drug candidates for the treatment of MDR-TB. Based on the structural similarity of these compounds and that of our previously prepared PZ and PD containing compounds from Family A, B and C (Figure 1.3.2.3), we decided to investigate the potential of these compounds as anti-tubercular agents. Exploration of their mechanism of action and use of the scaffold for further drug development will be explained in Chapter 4.

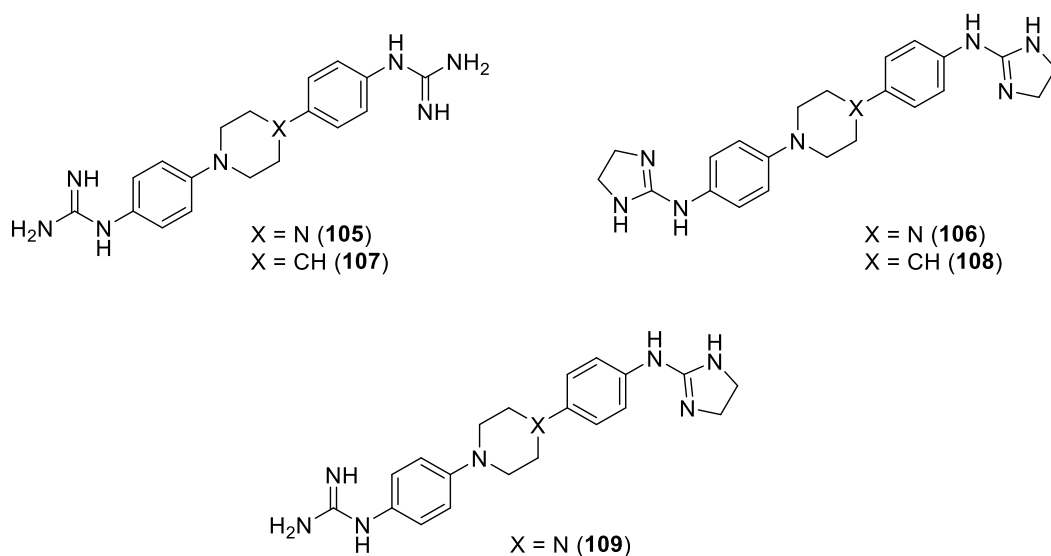


Figure 1.3.2.3. Structures of PZ and PD based MGBs previously made by the Rozas group.

1.4. References

1. Relman, D. A.; Hamburg, M. A.; Eileen, R.; Mack, A. *Microbial Evolution and Co-Adaptation: A Tribute to the Life and Scientific Legacies of Joshua Lederberg*; **2009**.
2. <https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2020>
3. <https://www.who.int/health-topics/human-african-trypanosomiasis>
4. <https://www.who.int/health-topics/leishmaniasis>
5. <https://www.who.int/teams/global-tuberculosis-programme/tb-reports>
6. Franco, J. R.; Cecchi, G.; Priotto, G.; Paone, M.; Diarra, A.; Grout, L.; Simarro, P. P.; Zhao, W.; Argaw, D.; *PLoS Negl. Trop. Dis.*, **2020**, *14*, 1–18.
7. Milner, D. A.; *Cold Spring Harb. Perspect. Med.*, **2018**, *8*, 1–11.
8. <https://www.cdc.gov/malaria/about/biology/index.html>
9. Ashley, E. A.; Pyae Phyo, A.; Woodrow, C. J.; *The Lancet*, **2018**, *391*, 1608–1621.
10. Rasti, N.; Wahlgren, M.; Chen, Q.; *FEMS Immunol. Med. Microbiol.*, **2004**, *41*, 9–26.
11. Cowman, A. F.; Tonkin, C. J.; Tham, W. H.; Duraisingh, M. T.; *Cell Host Microbe*, **2017**, *22*, 232–245
12. White, N. J.; Malaria. In: Farrar J, ed. *Manson’s tropical diseases*, 23rd edn. London: Elsevier, **2013**, 532–600.
13. Chiodini, P. L., *Parasitology*, **2014**, *141*, 1873–79.
14. Nureye, D.; Assefa, S., *Sci. World J.*, **2020**, 2020.
15. Walker, N. F.; Nadjm, B.; Whitty, C. J. M., *Medicine (United Kingdom)*, **2018**, *46*, 52–58.
16. Su, X. Z.; Lane, K. D.; Xia, L.; Sá, J. M.; Wellems, T. E., *Clin. Microbiol. Rev.*, **2019**, *32*, 1–29.
17. Pays, E.; Vanhollebeke, B.; Uzureau, P.; Lecordier, L.; Perez-Morga, D., *Nat. Rev. Microbiol.*, **2014**, *12*, 575–84
18. Büscher, P.; Cecchi, G.; Jamonneau, V.; Priotto, G., *Lancet*, **2017**, *390*, 2397–2409.
19. <https://www.cdc.gov/parasites/sleepingsickness/biology.html>
20. Kristensson, K.; Masocha, W.; Bentivoglio, M.; *Handb. Clin. Neurol.*, **2013**, *114*, 11–22

21. Sullivan, L.; Fleming, J.; Sastry, L.; Mehlert, A.; Wall, S. J.; Ferguson, M. A., *PLoS Negl. Trop. Dis.*, **2014**, *8*, e2976
22. Burza, S.; Croft, S. L.; Boelaert, M., *The Lancet*, **2018**, 6736 (Figure 2), 1-20
23. Reithinger, R.; Dujardin, J. C.; Louzir, H.; Pirmez, C.; Alexander, B.; Brooker, S., *Lancet Infect. Dis.*, **2007**, *7*, 581–96
24. <https://www.cdc.gov/parasites/leishmaniasis/biology.html>
25. WHO. Control of the leishmaniasis. Geneva: World Health Organization, **2010**.
26. Vega-Lopez, F., *Curr. Opin. Infect. Dis.*, **2003**, *16*, 97-101
27. CDC. Practical guide for specimen collection and reference diagnosis of leishmaniasis., https://www.cdc.gov/parasites/leishmaniasis/resources/pdf/cdc_diagnosis_guide_leishmaniasis.pdf (**2017**)
28. Rao, V.; Fujiwara, N.; Porcelli, S. A., *J. Exp. Med.*, **2005**, *201*, 535-543
29. Kaufmann, S. H. E., *Nat Rev. Immunol.*, **2001**, *1*, 20-30
30. Ferrari, G.; Langen, H.; Naito, M. and Pieters, J., *Cell*, **1999**, *97*, 435-447
31. Rao, V.; Fujiwara, N.; Porcelli, S. A.; Glickman, M. S., *J. Exp. Med.*, **2005**, *201*, 535-543
32. Egelund, E. F., Dupree, L., Huesgen, E.; Peloquin, C. A., *Expert Rev. Clin. Pharmacol.*, **2017**, *10*, 213-223.
33. Koch, A.; Mizrahi, V., *Trends Microbiol.*, **2018**, *26*, 555–556.
34. Ahmad, S., *Respir. Res.*, **2010**, *11*, 169.
35. Vandekerckhove, S.; D’Hooghe, M., *Bioorg. Med. Chem.*, **2015**, *23*, 5098–5119.
36. Tse, E. G.; Korsik, M.; Todd, M. H., *Malar. J.*, **2019**, *18*, 1–21.
37. Magill, A. J.; Zegarra, J.; Garcia, C.; Marqui, W., *Rev. Soc. Bras. Med. Trop.*, **2004**, *37*, 273-278.
38. O'Neill, P. M.; Barton, V. E.; Ward, S. A., *Molecules*, **2010**, *15*, 1705-1721
39. Mishra, M.; Mishra, V. K.; Kashaw, V.; Iyer, A. K.; Kashaw, S. K., *Eur. J. Med. Chem.*, **2017**, *125*, 1300–1320.
40. Walsh, J. J.; Coughlana, D.; Heneghan, N.; Gaynor, C.; Bell, A. *Bioorg. Med. Chem. Lett.*, **2007**, *17*, 3599.
41. Barrett, M. P.; Gemmell, C. G.; Suckling, C. J., *Pharmacol. Ther.*, **2013**, *139*, 12–23.
42. Hoagland, D. T.; Liu, J.; Lee, R. B.; Lee, R. E., *Adv. Drug Deliv. Rev.*, **2016**, *102*, 55–72.

43. Tiberi, S.; du Plessis, N.; Walzl, G.; Vjecha, M. J.; Rao, M.; Ntoumi, F.; Mfinanga, S.; Kapata, N.; Mwaba, P.; McHugh, T. D.; Ippolito, G.; Migliori, G. B.; Maeurer, M. J.; Zumla, A. *Lancet Infect. Dis.*, **2018**, *18*, e183–e198
44. Relman, D. A.; Hamburg, M. A.; Eileen, R.; Mack, A., *Microbial Evolution and Co-Adaptation: A Tribute to the Life and Scientific Legacies of Joshua Lederberg*; **2009**.
45. Martin, R. E.; Shafik, S. H.; Richards, S. N., *Curr. Opin. Phar.*, **2018**, *42*, 71-80.
46. Nagle, A. S.; Khare, S.; Kumar, A. B.; Supek, F.; Buchynskyy, A.; Mathison, C. J. N.; Chennamaneni, N. K.; Pendem, N.; Buckner, F. S.; Gelb, M. H.; Molteni, V., *Chem. Rev.*, **2014**, *114*, 11305–11347.
47. Rosenthal, M. R.; Ng, C. L., *ACS Infect. Dis.*, **2020**, *6*, 1599–1614.
48. Fairlamb, A. H.; Horn, D., *Trends Parasitol.*, **2018**, *34*, 481–492.
49. Kapil, S.; Singh, P. K.; Silakari, O., *Eur. J. Med. Chem.*, **2018**, *157*, 339-367.
50. Andries, K.; Verhasselt, P.; Guillemont, J.; Göhlmann, H. W. H.; Neefs, J.-M.; Winkler, H.; Van Gestel, J.; Timmerman, P.; Zhu, M.; Lee, E.; Williams, P.; de Chaffoy, D., *Science*, **2005**, *307*, 223-227
51. Veziris, N.; Bernard, C.; Guglielmetti, L.; Le Du, D.; Marigot-Outtandy, D.; Jaspard, M.; Aubry, *Eur. Resp. J.*, **2018**, *49*, 1601719.
52. Gagneux, S., *Nat. Rev. Mic.*, **2018**, *16*, 202–213
53. Papaventsis, D.; Casali, N.; Kontsevaya, I.; Drobniewski, F.; Cirillo, D. M.; Nikolayevskyy, V., *Clin. Microbiol. Infect.*, **2017**, *23*, 61–68.
54. Olaru, I. D.; Von Groote-Bidlingmaier, F.; Heyckendorf, J.; Yew, W. W.; Lange, C.; Chang, K. C., *Eur. Respir. J.*, **2015**, *45*, 1119–1131.
55. Cullia, G.; Tamborini, L.; Conti, P.; De Micheli, C.; Pinto, A., *ChemMedChem.*, **2018**, *13*, 2150–2158.
56. Al-hashimi, H. M., *Biopolymers*, **2014**, *99*, 1–21.
57. Ussery, D. W., *Encycl. Life Sci.*, **2002**, No. 1998.
58. Annunziato, A., *Nat. Sci. Educ.*, **2008**, *1*, 26.
59. Vargas-Parada, L., *Nat. Sci. Educ.*, **2010**, *3*, 63.
60. Souza, W.; Barrias, E.S., *American Trypanosomiasis Chagas Disease (Second Edition)*, **2017**, Chapter 18.
61. Kool, E., *Comprehensive Natural Products Chemistry*, **1999**, chapter 7
62. Rahman, A.; Sullivan, P. O.; Rozas, I., *Med. Chem. Commun.*, **2019**, *10*, 26-40.

63. Sriram, D.; Yogeeswari, P.; Thirumurugan, R.; Ratan Bal, T., *Nat. Prod. Res.*, **2005**, 19, 393–412
64. Lee, J. H.; Lee, J. M.; Kim, J. K.; Ahn, S. K.; Lee, S. J.; Kim, M. Y.; Jew, S. S.; Park, J. G.; Hong, C. I.; *Arch Pharm Res.*, **1998**, 21, 581–590
65. Wu, C. C.; Li, T. K.; Farh, L.; Lin, L. Y.; Lin, T. S.; Yu, Y. J.; Yen, T. J.; Chiang, C. W.; Chan, N. L., *Science*, **2011**, 333, 459–462.
66. Stokes, S. S.; Vemula, R.; Pucci, M. J., *ACS Infect. Dis.*, **2020**, 6, 1323–1331.
67. Strekowski, L.; Wilson, B., *Mutat. Res. Mol. Mech. Mutagen.*, **2007**, 623, 3–13.
68. Frederick, C. A.; Williams L. D.; Ughetto G.; van der Marel G. A.; van Boom J. H.; Rich A.; Wang A. H., *Biochemistry*, **1990**, 29, 2538–49.
69. Champoux, J. J., *Annu. Rev. Biochem.*, **2001**, 70, 369.
70. Neidle, S., *Nat .Prod. Rep.*, **2001**, 18, 291.
71. Neidle, S., *Principles of Nucleic Acid Structure*, Elsevier, 1st edn., **2008**.
72. Kopka, M. L., Yoon, C., Goodsell, D., Pjura P. and Dickerson, R. E., *J. Mol. Biol.*, **1985**, 183, 553.
73. Suckling, C. J., *Future Med. Chem.*, **2012**, 4, 971–989
74. Scott, F. J.; Nichol, R. J. O.; Khalaf, A. I.; Giordani, F.; Gillingwater, K.; Ramu, S.; Elliott, A.; Zuegg, J.; Duffy, P.; Rosslee, M. J.; Hlaka, L.; Kumar, S.; Ozturk, M.; Brombacher, F.; Barrett, M.; Guler, R.; Suckling, C. J., *Eur. J. Med. Chem.*, **2017**, 136, 561–572
75. Wang, M. Z.; Saulter, J. Y.; Usuki, E.; Cheung, Y. L.; Hall, M.; Bridges, A. S.; Loewen, G.; Parkinson, O. I.; Stephens, C. E.; Allen, J. L.; Zeldin, C.; Boykin, D. W.; Tidwell, R. R.; Parkinson, A.; Paine, M. F.; Hall, J. E., *Drug Metab. Dispos.*, **2006**, 34, 1985–1994.
76. Sauer, B.; Skinner-Adams, T. S.; Bouchut, A.; Chua, M. J.; Pierrot, C.; Erdmann, F.; Robaa, D.; Schmidt, M.; Khalife, J.; Andrews K. T.; Sippl, W., *Eur. J. Med. Chem.*, **2017**, 127, 22–40.
77. Wenzler, T.; Boykin, D. W.; Ismail, M. A.; Hall, J. E.; Tidwell, R. R.; Brun, R., *Antimicrob. Agents Chemother.*, **2009**, 53, 4185–4192.
78. Arafa, R. K.; Wenzler, T.; Brun, R.; Chai Y.; Wilson, W. D., *Eur. J. Med. Chem.*, **2011**, 46, 5852–5860.
79. Rice, C. a.; Colon, B. L.; Alp, M.; Göker, H.; Boykin, D. W.; Kyle, D. E., *Antimicrob. Agents Chemother.*, **2015**, 59, 2037–2044

80. Paul, A.; Nanjunda, R.; Kumar, A.; Laughlin, S.; Nhili, R.; Depauw, S.; Deuser, S. S.; Chai, Y.; Chaudhary, A. S.; David-Cordonnier, M.H.; Boykin, D. W.; Wilson, W. D. *Bioorg. Med. Chem. Lett.*, **2015**, 25, 4927–4932.
81. Dardonville, C. and Brun, R. *J. Med. Chem.*, **2004**, 47, 2296–2307
82. Rodríguez, F.; Rozas, I.; Kaiser, M.; Brun, R.; Nguyen, B.; Wilson, W. D.; Garcia R. N.; Dardonville, C., *J. Med. Chem.*, **2008**, 51, 909–923
83. Liu, Y.; Collar, C. J.; Kumar, A.; Stephens, C. E.; Boykin, D. W.; Wilson, W. D., *J. Phys. Chem. B.*, **2008**, 112, 11809
84. Garbett, N. C.; Ragazzon, P. A.; Chaires, J. B., *Nat. Protoc.*, **2007**, 2, 3166.
85. Nordén, B.; Kurucsev, T., *J. Mol. Recognit.*, **1994**, 7, 141–55.
86. Holm, A.; Nielsen, L.; Hoffmann, S.; Nielsend, S., *Phys. Chem. Chem. Phys.*, **2010**, 12, 9581–9596
87. Colson, P.; Bailly C.; Houssier, C., *Biophys.Chem.*, **1996**, 58, 125
88. Šmidlehner, T.; Piantanida, I.; Pescitelli, G.; *Beilstein J. Org. Chem.* **2017**, 14, 84–105.
89. Munde, M.; Ismail, M. A.; Arafa, R.; Peixoto, P.; Collar, C. J.; Liu, Y.; Hu, L.; David-Cordonnier, M. H.; Lansiaux, A.; Bailly, C.; Boykin, D. W.; Wilson, W. D., *J. Am. Chem. Soc.*, **2007**, 129, 13732
90. Methods, C. D. D. Antibiotics. *Biotechnol. Bioeng.*, **1965**, 7, 29–51.
91. Kumar, S. B.; Arnipalli, S. R.; Ziouzenkova, O., *Antibiotics*, **2020**, 9, 1–26.
92. Basak, S. C., *Curr. Comput. Aided. Drug Des.*, **2012**, 8, 1–2.
93. Meehan, C. J.; Goig, G. A.; Kohl, T. A.; Verboven, L.; Dippenaar, A.; Ezewudo, M.; Farhat, M. R.; Guthrie, J. L.; Laukens, K.; Miotto, P.; Ofori-Anyinam, B.; Dreyer, V.; Supply, P.; Suresh, A.; Utpatel, C.; van Soelingen, D.; Zhou, Y.; Ashton, P. M.; Brites, D.; Cabibbe, A. M.; de Jong, B. C.; de Vos, M.; Menardo, F.; Gagneux, S.; Gao, Q.; Heupink, T. H.; Liu, Q.; Loiseau, C.; Rigouts, L.; Rodwell, T. C.; Tagliani, E.; Walker, T. M.; Warren, R. M.; Zhao, Y.; Zignol, M.; Schito, M.; Gardy, J.; Cirillo, D. M.; Niemann, S.; Comas, I.; Van Rie, A., *Nat. Rev. Microbiol.*, **2019**, 17, 533–545.
94. Poce, G.; Consalvi, S.; Venditti, G.; Alfonso, S.; Desideri, N.; Fernandez-Menendez, R.; Bates, R. H.; Ballell, L.; Barros Aguirre, D.; Rullas, J.; De Logu, A.; Gardner, M.; Ioerger, T. R.; Rubin, E. J.; Biava, M., *ACS Med. Chem. Lett.*, **2019**, 10, 1423–1429.
95. Gurcha, S. S.; Cox, J. A. G.; Mendoza, A.; Jime, E.; Casanueva, R.; Loman, N. J.; Moynihan, P. J.; Izquierdo, C.; Kremer, L.; Bantscheff, M.; Cammack, N. C.; Drewes, G.; Ballell, L.; Barros, D.; Besra, G. S.; Bates, R. H., *Nature Comm.*, **2016**, 7, 12581.

96. Foo, C. S.; Lupien, A.; Kienle, M.; Vocat, A.; Benjak, A.; Sommer, R.; Lamprecht, D. A.; Steyn, A. J. C.; Pethe, K.; Piton, J.; Altmann, K.; Cole, T., *A. S. Micr.*, **2018**, *9*, 1–13.
97. McKeever, C.; Kaiser, M.; Rozas, I., *J. Med. Chem.*, **2013**, *56*, 700–711.
98. Nagle, P. S.; Rodriguez, F.; Kahvedžić, A.; Quinn, S. J., Rozas, I., *J. Med. Chem.*, **2009**, *52*, 7113–7121.
99. 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.
100. O’Sullivan, P.; Rozas, I., *ChemMedChem*, **2014**, *9*, 2065–2073.
101. Shaquiquzzaman, M.; Verma, G.; Marella, A.; Akhter, M.; Akhtar, W.; Khan, M. F.; Tasneem, S.; Alam, M. M., *Eur. J. Med. Chem.*, **2015**, *102*, 487–529.
102. Forge, D.; Cappoen, D.; Laurent, J.; Stanicki, D.; Mayence, A.; Huang, T. L.; Verschaeve, L.; Huygen, K.; Eynde, J. J. Vanden. *Eur. J. Med. Chem.*, **2012**, *49*, 95–101.

Chapter 2

Objectives

2. Objective overview

The first part of this thesis will focus on the development of novel DNA minor groove binders (MGBs) as potential anti-protozoal agents. Considering the good DNA binding properties and anti-parasitic activity of some of our previous guanidine-based DNA MGBs, we propose to produce a novel family of MGBs which combines key structural features of previous promising compounds. Firstly, computational studies will be conducted to re-evaluate the mechanism of action of a series of guanidine-based MGBs previously produced by our group based on miscorrelations between DNA binding affinity and anti-parasitic activity. Similarly, computational studies will be used to assess the theoretical DNA binding properties of the proposed novel compounds. Following on from this, the most promising derivatives will be synthesised with the use of organic transformations commonly utilised by our group. The DNA binding properties including binding affinity and mode of binding will be assessed with the use of appropriate biophysical techniques. Finally, their anti-parasitic activity against three different protozoans will be assessed to determine their ability to be used as effective anti-protozoal agents.

The second part of this thesis will focus on the development of guanidine-based derivatives as potential anti-tubercular agents. Based on structural similarities between some anti-mycobacterial agents within the literature and some of our in-house family of compounds, we propose to investigate the anti-mycobacterial potential of these compounds. Additionally, biochemical and computational studies will be conducted to investigate their mechanism of action. Based on these results, a novel series of guanidine-based derivatives will be designed using a structure-based approach with the aim of creating more potent anti-mycobacterial agents. Finally, their activity against surrogate strains of *Mycobacteria tuberculosis* will be determined to assess their potential as novel anti-tubercular agents.

2.1. DNA Minor Groove Binders as potential anti-protozoal agents

2.1.1. Computational aims

i) In previous work conducted by our group a series of aminoalkyl conjugates of mono-guanidine derivatives (Family E) were produced which showed moderate DNA binding affinity and very promising antimalarial activity (Figure 2.1.1.1). However, upon closer analysis of the biophysical and biological results miscorrelations between the DNA binding affinity and the anti-parasitic activity was observed.¹

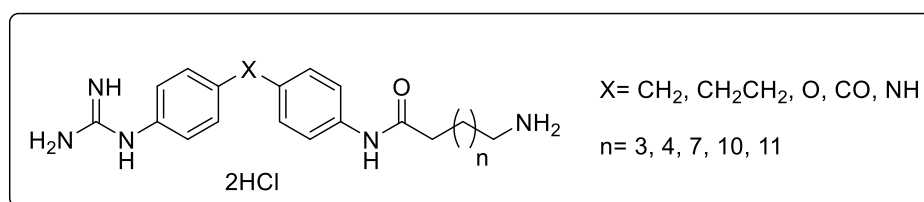
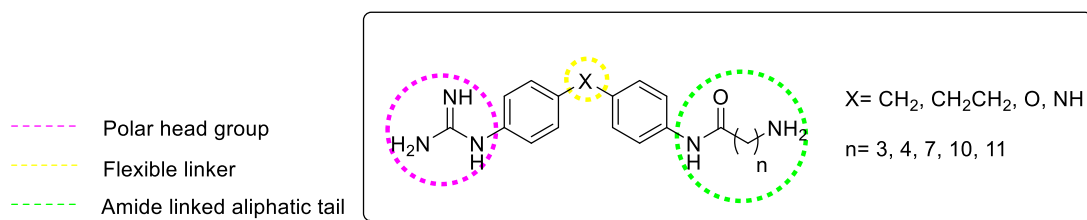


Figure 2.1.1.1. Structure of aminoalkyl conjugates of mono-guanidine derivatives (Family E) previously made by the Rozas group

In search for possible alternative targets, structural similarity between these compounds and *Plasmodium falciparum* Dihydrofolate Reductase (*pf*DHFR) substrates and inhibitors led to the hypothesis that they may also elicit their activity through *pf*DHFR inhibition (Figure 2.1.1.2). With the use of *in silico* docking, we proposed to perform a theoretical study to understand the possible interactions between these previously produced compounds in the *pf*DHFR active site.



***pf*DHFR ligands:**

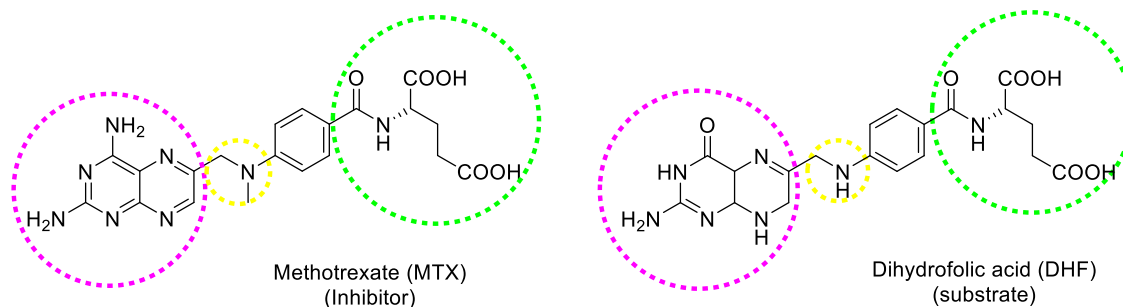


Figure 2.1.1.2. Structural similarities drawn from the comparison of Family E and *pf*DHFR ligands methotrexate (MTX) and dihydrofolic acid (DHF).

ii) Secondly, based on the attractive DNA binding properties of previously made bis-guanidine (Family A, Figure 2.1.1.3) and aminoalkyl derivatives (Family E, Figure 2.1.1.3) we propose to design a new series of aminoalkyl derivatives of di-aryl bis-guanidines (Family F, 2.1.1.3) which combines the key structural features of both families that positively contribute to DNA minor groove binding. By attaching the aminoalkyl chain to one of the guanidine-like cations instead of condensing it directly to the aromatic ring through an amide bond, we hypothesise that this will improve minor groove binding through the incorporation of an additional positive charge and reduction of the inter-cationic distance. Based on biophysical results of Family E compounds, carbon chain length between 4 and 7 methylene carbons will be explored to avoid decreases in DNA binding due to the presence of excessively large lipophilic tails.

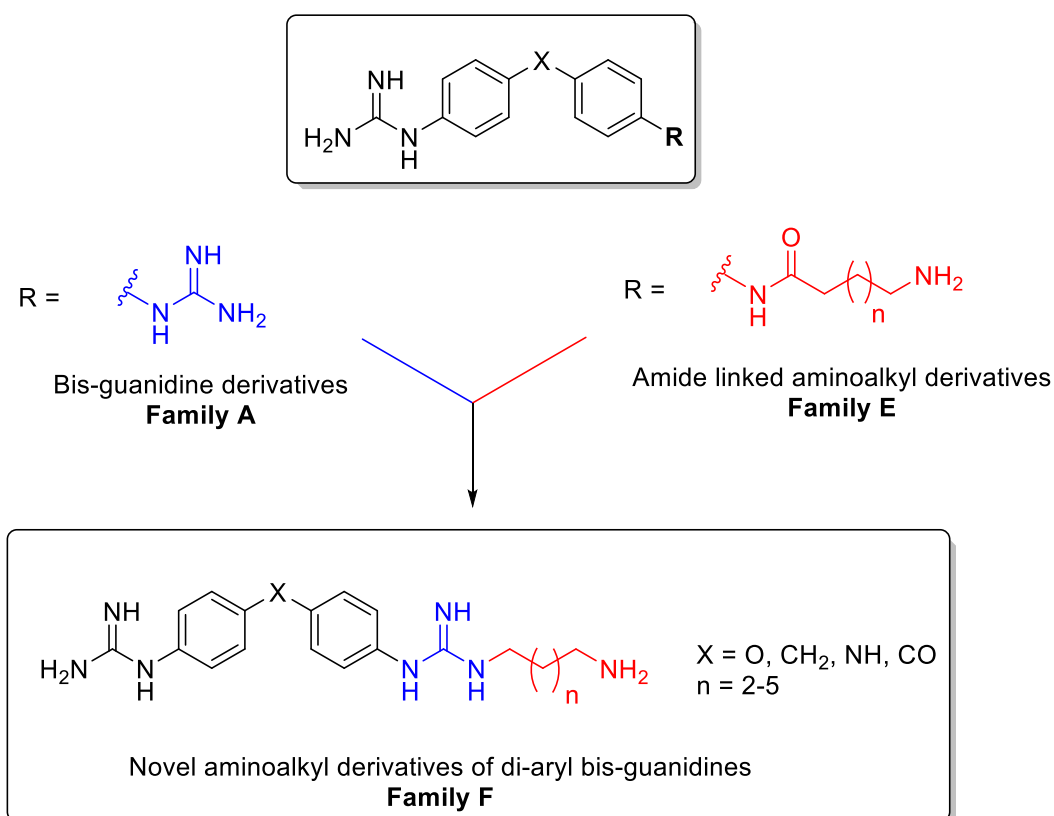


Figure 2.1.1.3. Rationale behind the design of the novel aminoalkyl derivatives (Family F).

Using *in silico* docking techniques, the theoretical binding properties of these proposed compounds will be assessed. The effect of linker type ($-X-$ in Family F general structure), carbon chain length (n in Family F general structure) and the additional positive charge on the binding will be investigated which may give explanations to the results that will be obtained from biophysical and biological studies. In addition, the theoretical binding of these compounds to *pf*DHFR will be explored to anticipate any potential off-target effects.

Optimization of the structures of previous amino-alkyl derivatives (Family E) and proposed compounds (Family F) will be carried out by using Density Functional Theory (DFT) methods (B3LYP² or M06-2X³) with the 6-311++G** basis set⁴ and SMD⁵ water solvation model using the computational package Gaussian16⁶. The Autodock Vina⁷ software will be used for the docking of these ligands to a 3D-crystal structure of DNA (an AT oligomer as a model of the minor groove) and of *Pf*DHFR obtained from the RCSB Protein Data Bank.

2.1.2. Synthetic aims

Based on the computational studies, the new amino-alkyl derivatives will be synthesised by employing similar synthetic routes used to produce previous compounds. Like families A and E, the diaromatic mono-guanidine portion of the molecule remains the same. Through the retrosynthetic analysis shown below (Figure 2.1.2.1), the desired intermediate can be obtained from commercially available starting materials. However, various methodologies will be attempted to produce the *N*-aminoalkyl substituted guanidine at the other terminal of the diaromatic core. All synthetic intermediates and final compounds will be characterised by NMR, MS, and IR. Additionally, final compounds will be assessed by HPLC before biophysical and biological assays. The synthesis of each compound will be discussed in detail in Chapter 3.2.

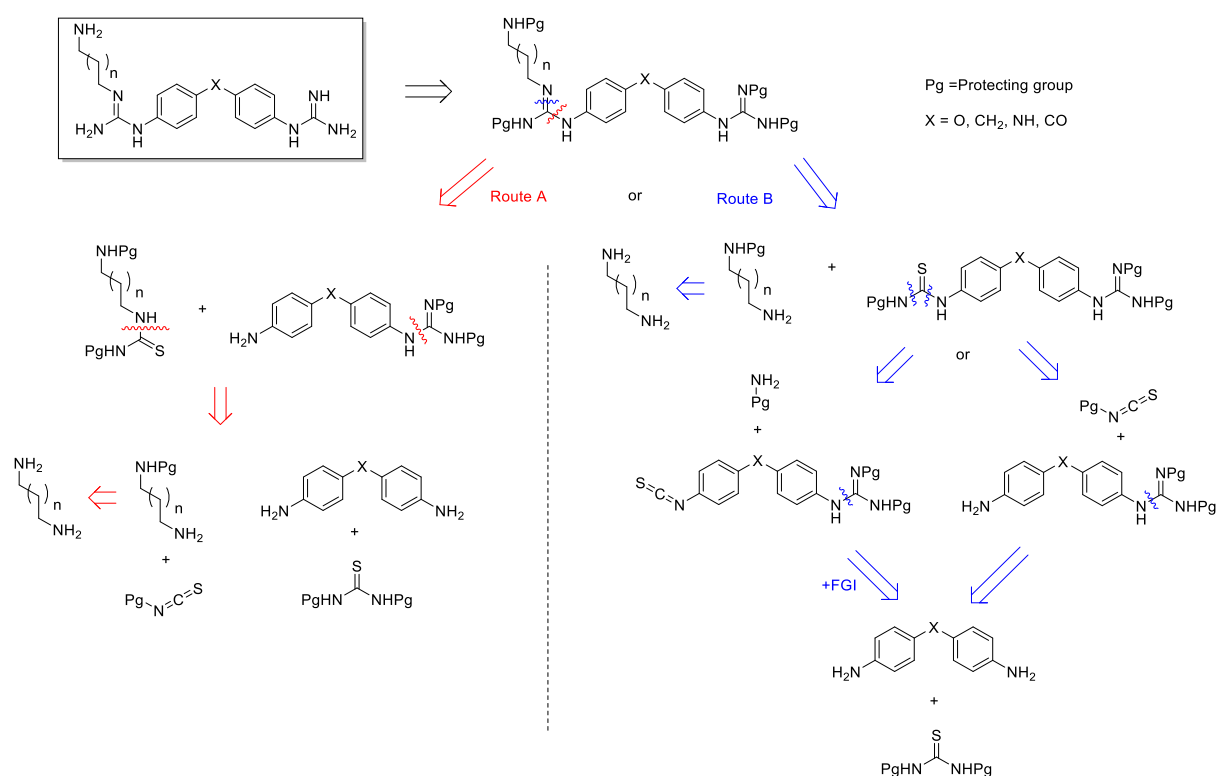


Figure 2.1.2.1. Retrosynthetic analysis for the synthesis of Family F.

2.1.3. Biophysical aims

The DNA binding properties of the proposed compounds will be assessed by UV-thermal denaturation, Circular Dichroism (CD) and UV-Vis titrations.

- i) UV-thermal denaturation studies will be carried out to measure the binding affinity of the compounds towards DNA. Salmon testes (stDNA) will be used to determine the affinity towards unspecific DNA and the most promising compounds will then be tested using the synthetic poly(dA-dT)₂ (DNA minor groove established model) to assess binding affinity towards the DNA minor groove.
- ii) To investigate the mode of binding of these compounds, CD titration experiments will be performed. Based on the identification of key spectral changes upon addition of the ligand to the DNA solution, we will be able to determine whether a compound binds either mainly through the minor groove or intercalates in between the DNA bases.
- iii) UV-Vis titration experiments with the use of the poly(dA-dT)₂ oligonucleotide will be conducted to determine the binding constants of the most promising compounds. Considering that these compounds exhibit good DNA binding properties, their anti-parasitic activity will be investigated.

2.1.4. Biological aims

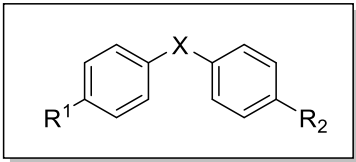
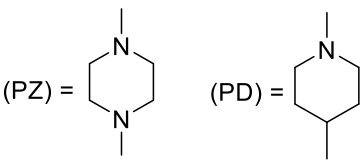
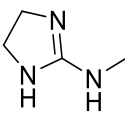
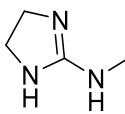
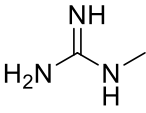
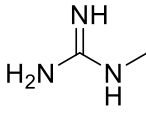
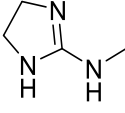
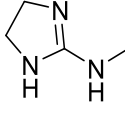
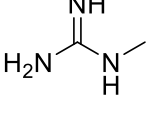
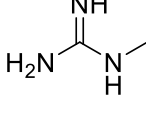
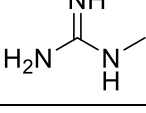
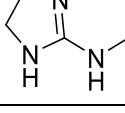
In collaboration with the Swiss Tropical and Public Health Institute, biochemical assays will be carried out with a subset of the novel compounds that exhibit good DNA binding affinity. Their anti-parasitic activity will be assessed in *Trypanosoma brucei rhodesiense* STIB 900, *Leishmania donovani* axenic amastigotes and *Plasmodium falciparum* NF54 cells. Cytotoxic effects of these compounds will also be assessed in L6 cells (rat skeletal muscle myoblast). The resulting IC₅₀ values will be used to rank the compounds in terms of anti-parasitic activity as well as determining their selectivity for the parasite. In addition, by comparing the DNA binding affinity with the anti-parasitic activity we will be able to identify whether a correlation between the two sets of data exists. Hence, if such a correlation exists, we could conclude that DNA minor groove binding is the most likely mechanism of action.

2.2. Guanidine based derivatives as potential anti-tubercular agents

2.2.1. Preliminary investigation

A study conducted by Forge *et al* reported a series of PD and PZ containing bis-amidines with good anti-mycobacterial activity.⁸ Interestingly, these compounds bear high structural similarity to a subset of PZ and PD containing MGBs within our in-house library of compounds (Table 2.2.1.1).

Table 2.2.1.1. Structures of in-house piperazine and piperidine containing MGBs

			
Cmp	X	R ¹	R ²
108	PD		
107	PD		
106	PZ		
105	PZ		
109	PZ		

We propose to investigate the anti-mycobacterial potential of these compounds through:

- i) Drug susceptibility screening against surrogate strains of *M. tuberculosis*.
- ii) Whole-cell phenotypic assays and enzyme-based biochemical assays
- iii) Time-kill assays
- iv) Whole-genome sequencing

Drug susceptibility screening against the *M. tuberculosis* surrogate strains *M. bovis* BCG and *M. smegmatis* will be conducted with the use of resazurin microtiter plate assay (REMA) analysis in collaboration with Prof. Stephen Gordon from the School of Veterinary Medicine at University College Dublin. As these compounds were previously designed as DNA MGBs, correlations between anti-mycobacterial activity and DNA binding affinity will be investigated. If no correlations exist, we aim to further analyse their mechanism of action through whole-cell phenotypic and enzyme-based assays with the help of Dr. Apoorva Bhatt and Dr. Sarah Batt from the University of Birmingham. To determine whether the compounds under investigation are killing the bacteria (bactericidal) or inhibiting their growth (bacteriostatic), time-kill experiments will be performed. Confirmation of the drug target will be attempted with the use of whole-genome sequencing (WGS) conducted at the UCD Conway Institute. Given the recurring issue of drug resistance in TB drug therapy, WGS analysis will also be used to assess the drug resistance susceptibility of the most active compound of the series.

2.2.2. Computational aims

Based on the results from the preliminary investigation of our in-house family of compounds, computer-aided techniques such as ligand-based pharmacophore screening and structure-based molecular docking will be employed. These techniques will be used to understand the key structural components essential for the compound's activity and to determine the binding interactions to the proposed target. Ligand-based pharmacophore screening will be conducted through the Schrodinger Maestro software with the use of chemical databases from the ChEMBL and ZINC database. Similar to the *in silico* studies that will be used for the potential anti-protozoal agents, Gaussian16 and Autodock vina will be employed. Based on

the results gained from both of these techniques, new derivatives will be designed with potentially more potent anti-mycobacterial activity.

2.2.3. Synthetic aims

- i) Previous in-house PD and PZ containing compounds will be resynthesized using the same synthetic methodology established within the Rozas' group. To determine the significance of the counterion associated with the guanidinium/imidazolinium, both the trifluoroacetate and hydrochloride salts were prepared for biological testing.
- ii) Based on preliminary investigations and computational studies, novel PD/PZ containing derivatives will be synthesised using well known organic transformations. All synthetic intermediates and final compounds will be characterised by NMR, MS, and IR. Additionally, final compounds will be assessed by HPLC before biophysical and biological assays. The synthesis of each compound will be discussed in detail in Chapter 4.5.

2.2.4. Biochemical and Biological aims

- i) To assess the drug susceptibility of the novel compounds produced, REMA analysis will be conducted to obtain their minimum inhibitory concentration (MIC) against *M. smegmatis* which can be used to construct a structure-activity relationship (SAR).
- ii) Biochemical analysis will be used to verify the mechanism of action of the compound from the novel series which exhibits the greatest anti-mycobacterial activity.
- iii) Time-kill assays will be conducted to determine the bacteriostatic or bactericidal nature of the most active novel compound using *M. smegmatis*. The killing kinetics will be compared to previous in-house compounds and known anti-tubercular agents. In addition, the anti-mycobacterial potency of the most promising compound in combination with known anti-tubercular agents will be investigated.
- iv) Cytotoxicity analysis of the novel series of compounds against healthy mammalian cells (i.e. THP-1 cells and HEK293 cells) will be performed to determine their safety.

2.3. References

1. McKeever, C.; Kaiser, M.; Rozas, I., *J. Med. Chem.*, **2013**, *56*, 700–711.
2. Becke, A. D., *J. Chem. Phys.*, **1993**, *98*, 1372–1377
3. Zhao, Y.; Truhlar, D. G., *Theor. Chem. Acc.*, **2008**, *120*, 215–241.
4. Hehre, W. J.; Ditchfield, R. and Pople, J. A., *J Chem. Phys.*, **1972**, *56*, 2257–2262.
5. Marenich, A. V.; Cramer, C. J.; Truhlar, D. G., *J. Phys. Chem. B.*, **2009**, *113*, 6378–6396.
6. Gaussian 16, Revision C.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian, Inc., Wallingford CT, **2016**.
7. Trott, O and Olson, A. J., *Journal of Computational Chemistry*, **2010**, *31*, 455-461
8. Forge, D.; Cappoen, D.; Laurent, J.; Stanicki, D.; Mayence, A.; Huang, T. L.; Verschaeve, L.; Huygen, K.; Eynde, J. J. Vanden. *Eur. J. Med. Chem.*, **2012**, *49*, 95–101.

Chapter 3
***DNA Minor Groove Binders as potential
anti-protozoal agents***

3.1. Computational studies

Di-aromatic bis-amidines and bis-guanidines have been shown to bind strongly to the DNA minor groove. As mentioned previously, the minor groove is mainly composed of repeating sequences of A:T base pairs. Considering the Watson - Crick Model, the hydrogen bonding network between these bases exposes the A-N3, T-O2 and deoxyribose-O4' forming the base of the minor groove while a backbone of hydrogen atoms form the hydrophobic walls. This in turn creates a negative electrostatic potential within the groove which facilitates the interaction of positively charged molecules. In addition, the convex structure of the minor groove complements the concave structure of typical MGBs allowing the formation of close contacts in the deep, narrow space formed by the two strands of DNA. Furthermore, this binding is driven by the high entropic increment due to the displacement of the spine of hydration. The crystal structure (PDB:1D64) of the known MGB pentamidine in complex with the dodecanucleotide duplex sequence d(CGCGAATTCGCG)₂ provided detailed experimental information of the interactions established with the A:T base pairs explaining how pentamidine is selective for the minor groove (Figure 3.1.1 (a)).¹ Adhering to the isohelicity principle due to its concave geometry upon binding, key hydrogen bond (HB) interactions with A-deoxyribose-O4', G-deoxyribose-O4' and C-O2 are formed as well as electrostatic interactions with the backbone phosphate moieties. This is also the case of the diamidine, netropsin, which shows high selectivity for the minor groove where each amide NH connecting the pyrroles forms HB interactions with A-N3 and T-O2 (Figure 3.1.1 (b), PDB: 1Z8V).²

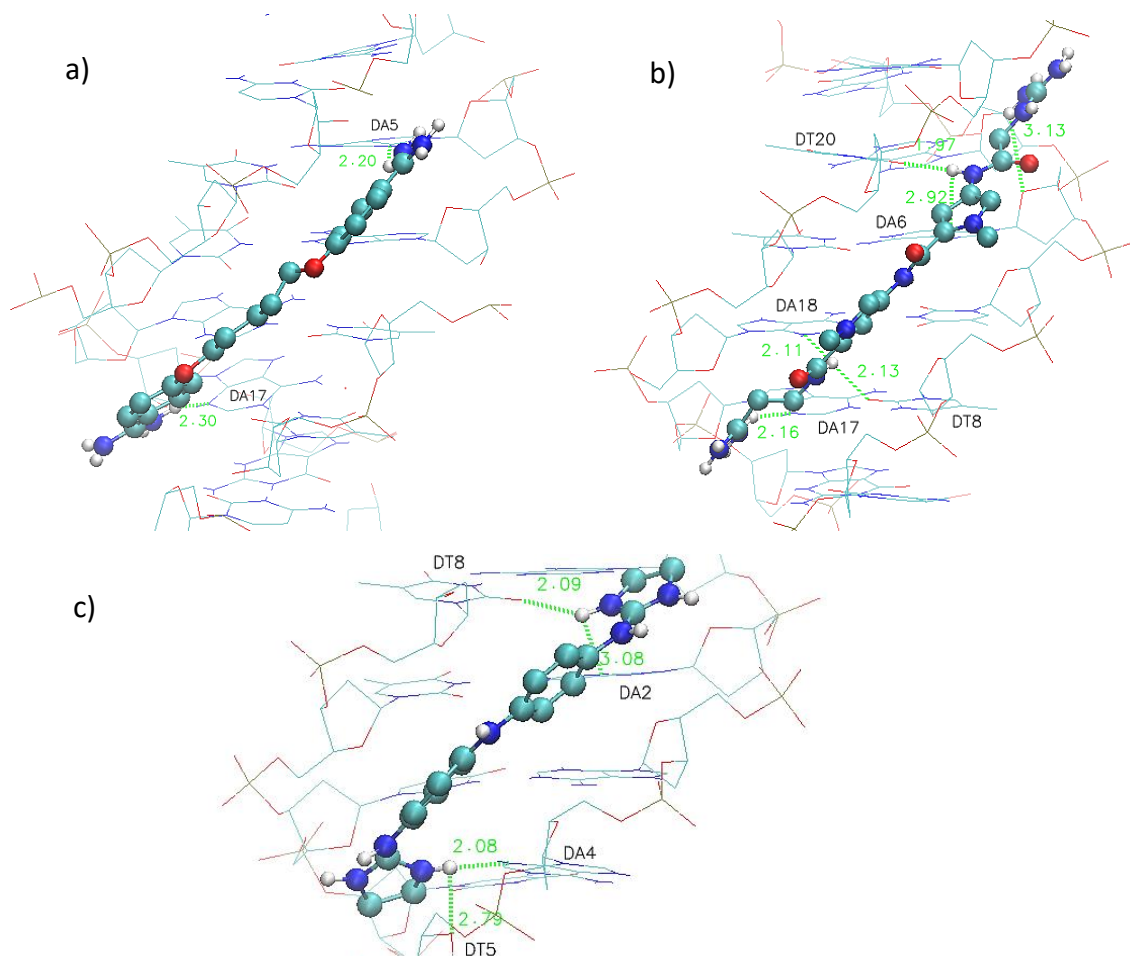


Figure 3.1.1. Crystal structures of (a) pentamidine (PDB: 1D64), (b) netropsin (PDB: 1Z8V) and (c) compound **19** (PDB: 4OCD) in the DNA minor groove.³

Previously made guanidine/2-aminoimidazoline based compounds from our in-house library have been shown experimentally to bind to the minor groove of DNA with high affinity and selectivity. Furthermore, the X-ray crystal structure of a bis-2-aminoimidazoline derivative (**19**) has been published showing a very similar binding mode to that of known MGB furamidine (Figure 3.1.1 (c), PDB: 4OCD).³ Hence, with the use of computational molecular modelling we aim to investigate and understand the mode of binding of these previously made derivatives with the objective to guide the design of new derivatives with improved minor groove binding ability.

Prior to *in silico* docking in DNA, a search for an alternative target for previous Family E was conducted based on results gained from biological studies. Structural similarity with DHFR

ligands encouraged molecular docking studies in a reference crystal structure of *pf*DHFR (PDB: 1J3I). In addition, the newly designed MGBs (Family F) will be docked in *pf*DHFR to anticipate potential dual targeting.

3.1.1. Computational Methods

Optimisation of the structures of previous di-aryl bis-guanidines/2-aminoimidazolines and amino-alkyl derivatives, as well as proposed compounds, was carried out by using Density Functional Theory (DFT) methods (B3LYP⁴ or M06-2X⁵ with the 6-311++G** basis set⁶) in a SMD⁷ water solvation model using the computational package Gaussian16⁸ (Figure 3.1.1.1).



Figure 3.1.1.1. Optimised structures of selected compounds from previous Family A, B, D, E and newly designed Family F.

Using the Autodock vina⁹ software, docking of these ligands to DNA (d(CGCGAATTCGCG)₂ oligomer as a model of the minor-groove; PDB:1D64) and *Pf*DHFR (in the case of Family E and F) was performed. Visualisation software such as pyMOL¹⁰ and VMD¹¹ were used to identify key binding interactions.

Two different dockings methods exist which include rigid-rigid docking and flexible-rigid docking. With the rigid-rigid approach, the dihedral angles within a ligand are fixed to match that of previously identified ligand-macromolecule crystal structures. The macromolecule is also used in its rigid state and is obtained by removal of the ligand from a ligand-

macromolecule crystal structure, thereby, utilising the rigid ligand-bound conformation. On the other hand, the flexible-rigid approach is where the DFT optimised ligands are used as an initial guess, with all bonds freely rotatable in the binding site of the rigid macromolecule. Moreover, the flexible-rigid approach was employed in our docking experiments to accommodate for modes of binding that may be different to that of the reference ligand-macromolecule crystal structure. When a docking experiment has been performed, various binding poses are obtained which are ranked by the Gscore. The Gscore is an empirical scoring function that represents the binding affinity (kcal mol^{-1}) where the best poses are those with lower Gscore values.

3.1.2. Molecular docking of compounds from previous Families A, B, C and D in a model of the DNA minor groove

The di-aryl scaffold has been utilised by our group to create a massive series of compounds aimed towards the development of effective DNA minor groove binding agents. By varying the linker (X) and cationic moiety (+), symmetric and asymmetric di-aryl guanidine/2-aminoimidazoline based compounds have been produced (Figure 3.1.2.1). The curvature of these compounds as well as their cationic nature allows for strong minor groove binding capability. As mentioned earlier, DNA thermal denaturation studies have shown a trend in increasing DNA binding affinity from symmetric bis-guanidines to asymmetric guanidine/2-aminoimidazolines to symmetric bis-2-aminoimidazolines, which show the strongest binding. This is attributed to the added hydrophobicity allowing for the development of van der Waals interactions with the backbone of the DNA minor groove.

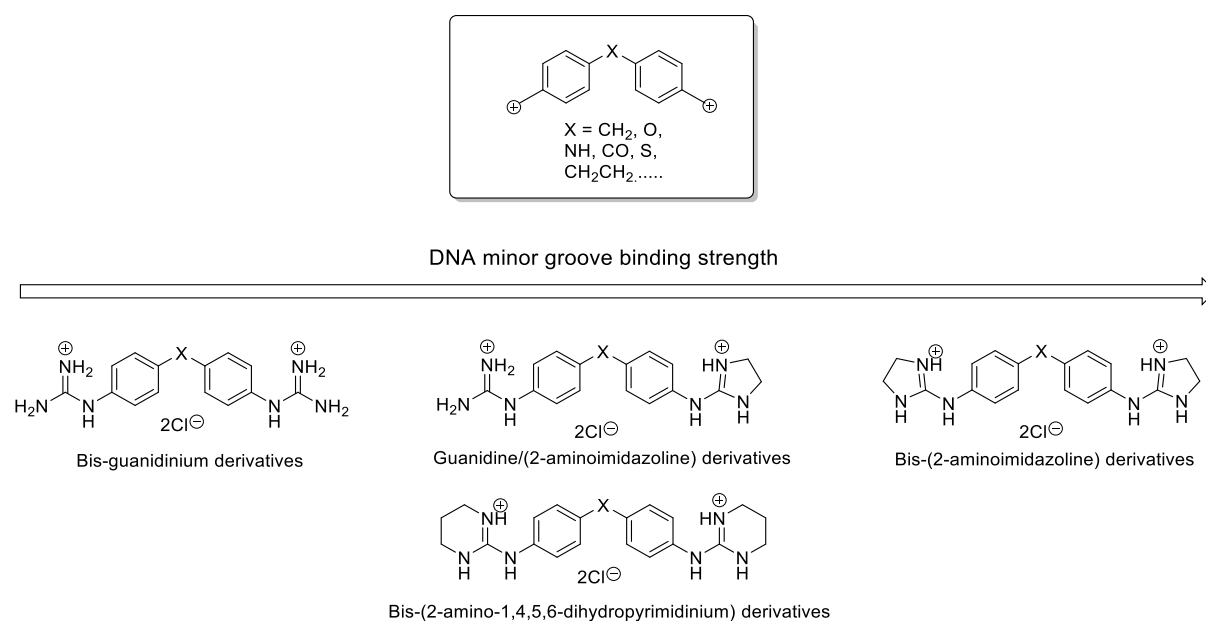
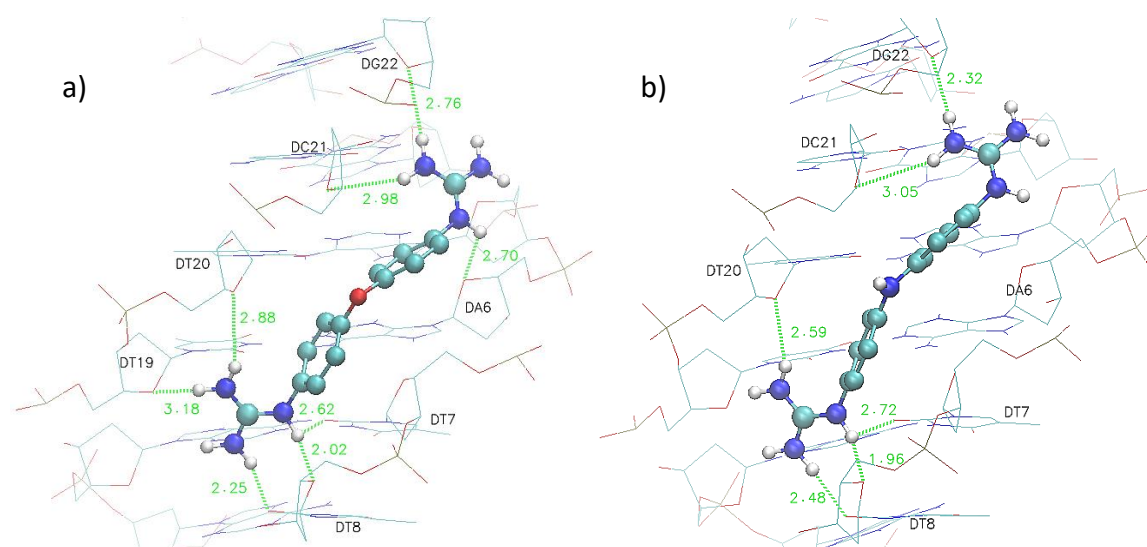


Figure 3.1.2.1. General structures of in-house families of MGBs and the trend in DNA binding strength.

Firstly, considering the DNA binding mode of the symmetric di-aryl bis-guanidines (Family **A**), our docking studies show that HBs are formed with the nucleotide bases of the groove. As shown in Figure 3.1.2.2, the guanidinium moiety of compounds **20-23** form HB interactions with C-deoxyribose-O4', A-deoxyribose-O4', T-deoxyribose-O4' and T-O2. Electrostatic interactions with the negatively charged phosphates of the DNA backbone are also apparent. Additionally, linkers of compounds **20** (X= O), **21** (X= NH) and **22** (X= CO) may be capable of forming HB interactions with surrounding water molecules.



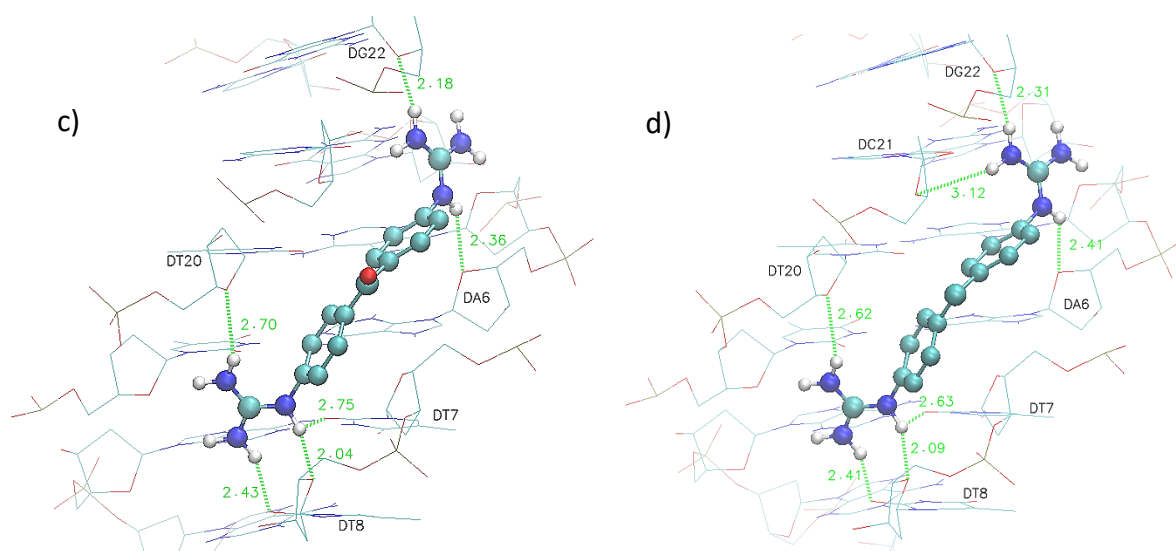
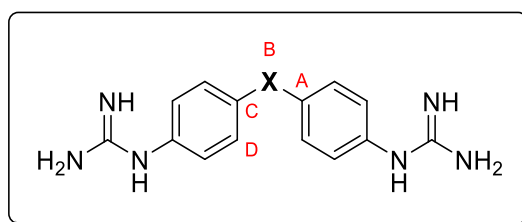


Figure 3.1.2.2. Optimal docking poses for Family A compounds (**20(a)**, **21(b)**, **22(c)** and **23(d)**) (CPK model) in the DNA minor groove (line model) (reference crystal structure - PDB: 1D64).

Considering the geometry of binding of these compounds, they adopt a curved concave shape which allows for the snug fit into the groove. Interestingly, compounds **21** and **22** bear the smallest changes in the A-B-C-D dihedral angle ($\Delta\Phi$) between the docked and isolated state which may explain the better Gscore values obtained (Table 3.1.2.1). This almost pre-organised geometry may be responsible for the strong DNA binding affinity observed in previous thermal denaturation studies, since a low energy penalty is expected upon binding. Additionally, when comparing the estimated binding affinity (Gscore) values of compounds **20**, **21**, **22** and **23**, similar values were obtained; however, compounds **21** and **22** show slightly stronger binding affinity.

Table 3.1.2.1. Dihedral angles between the phenyl rings of compounds **20**, **21**, **22** and **23** in their docked and isolated state.

Cmp	Linker	Isolated Dihedral angle A-B-C-D(Φ)	Docked Dihedral angle A-B-C-D(Φ)	$\Delta\Phi$ (unbound – bound)	Gscore (kcal mol ⁻¹)
22	CO	33.0	28.96	4.04	-9.4
21	NH	31	30.07	0.93	-9.4
23	CH ₂	65.9	30	35.09	-9.0
20	O	44.1	36.08	8.02	-9.2

Isolated dihedral angles taken from DFT optimised compounds and docked dihedral angles taken from docking outputs of the optimised ligands with the DNA minor groove model (PDB: 16D4)

As shown in Figure 3.1.2.2, each ligand only spans across four nucleotide base pairs in the AT-rich region, highlighting the possibility of potential interactions and filling of the unoccupied space within the groove. This inspired the development of the asymmetric guanidine/2-aminoimidazoline (Family C) and symmetric bis-2-aminoimidazole derivatives (Family B) which were shown to have greater DNA binding affinity than the symmetric bis-guanidine counterparts. Comparing the binding interactions established by these compounds, similar HB interactions are formed (Figure 3.1.2.3). The improved experimental DNA binding affinity of compound **25** may be attributed to the greater hydrophobic contacts of the two 2-aminoimidazoline moieties to the DNA backbone.¹²

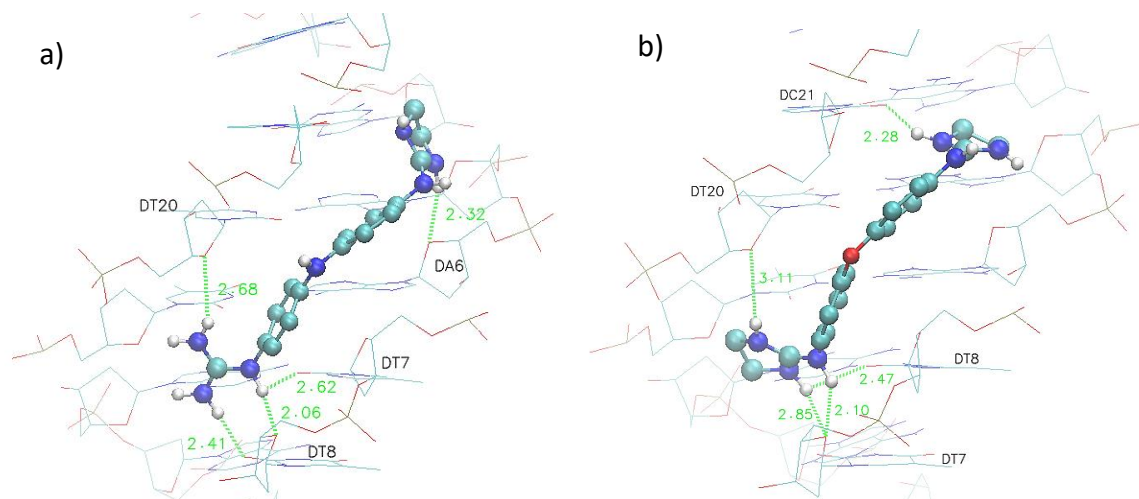


Figure 3.1.2.3 Optimal docking poses of compound **24** (Family C) and compound **25** (Family B) within a model of the DNA minor groove.

Our results are in agreement with the reported crystallographic study with compound **19**, which shows that the compound binds preferably to the DNA-binding site 5'-AATT within a self-complementary oligonucleotide composed of sequence 5'-d(CTTAATTCGAATTAAG).¹³ Additionally, the highly twisted DFT optimised structure that we obtained adopts a more planar crescent shape when bound in the minor groove of the DNA (Figure 3.1.2.4). The presence of bifurcated HBs, as well as van der Waals interactions observed in the crystallographic structure, is mirrored in our docking studies. However, it is important to note that these ligands were docked using a flexible-rigid approach. Hence, the slight differences in the binding modes observed in our computational studies compared to the above crystal structure may be a result of the reported conformational changes that occur within the DNA-ligand complex upon binding. This can be seen in the slightly different orientation of the di-aromatic system and methylene carbons of the imidazoline moiety of compound **19** as shown in Figure 3.1.2.4. Although there is a significant change in the dihedral angle between the DFT optimised form of compound **19** and the docked structure, a more coplanar geometry between the two phenyl rings is apparent within the crystal structure. This highlights the fact that there is an energetic penalty upon binding of these compounds allowing for a tighter induced fit. The superior experimental binding affinity of these compounds indicates that the energetic penalty is circumvented by the gain of additional van der Waals interactions.

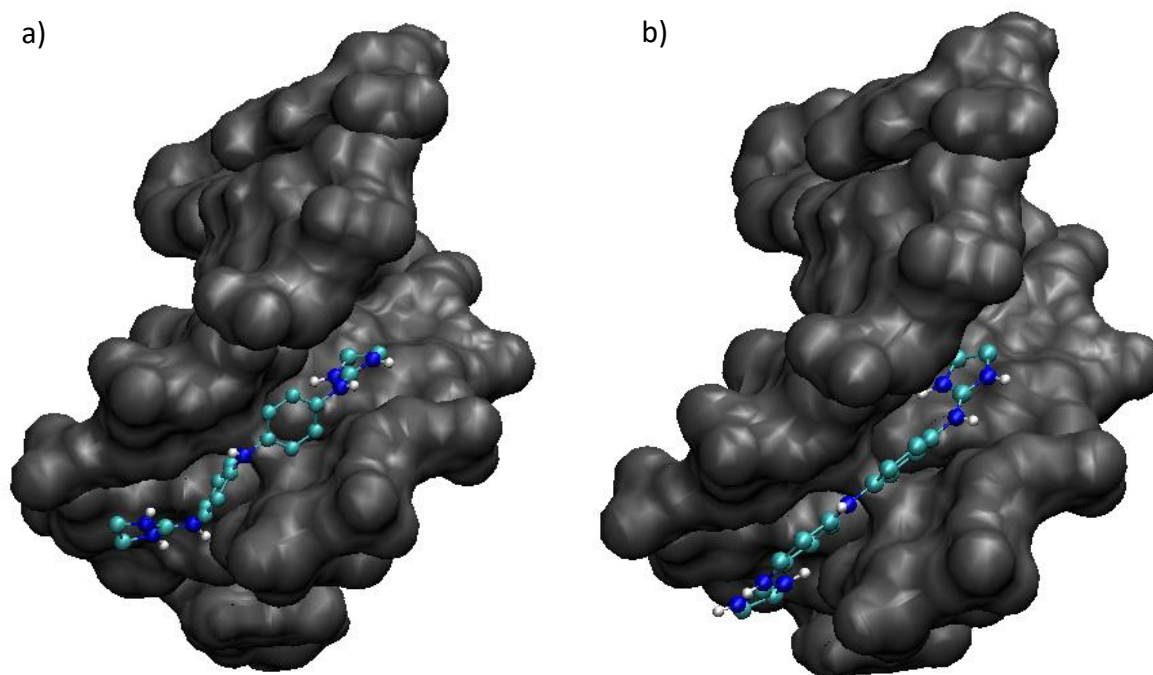
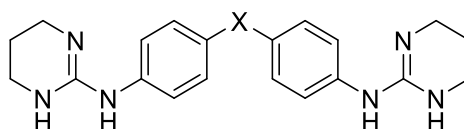


Figure 3.1.2.4. (a) Docking pose of compound **19** in a model of the DNA minor groove and (b) crystal structure of compound **19** in the DNA minor groove (surf model).³

Given the encouraging results of previous derivatives and the identification of the trend in increasing size and lipophilicity on DNA binding affinity, Family D was designed (Figure 3.1.2.5). As recently reported, prior to molecular docking of these 2-amino-1,4,5,6-tetrahydropyrimidinium based compounds theoretical physicochemical parameters were calculated using the Marvin software package to elucidate their logP, van der Waals surface area, polar surface area (PSA) and size of the guanidinium-like cation in order to access their feasibility as MGBs.¹²

X= CO, CH₂, O, NH



Bis-(2-amino-1,4,5,6-tetrahydropyrimidine) derivatives

Family D

Figure 3.1.2.5. General structure for Family D compounds.

However, upon docking of these derivatives to the DNA minor groove model in a flexible-rigid approach, a different binding pose was observed where the bulky 1,4,5,6-tetrahydropyrimidinium moieties are pointed in slightly opposite directions to accommodate the structure into the groove given its curvature (Figure 3.1.2.6 (a)). This yielded Gscores between -8.0 and -8.3 kcal mol⁻¹ which is lower than that of previous derivatives. In order to visualise a binding geometry similar to that previously determined from the crystal structure of compound **19** bound to DNA, all angles within the structure of compound **26** were set to that of DNA bound compound **19** and rigidified prior to docking (rigid-rigid approach). This resulted in the previously hypothesised binding geometry (Figure 3.1.2.6 (b)) but produced very poor Gscore values for all possible poses ($\sim$ -6.9 kcal mol⁻¹). This indicates that the energetic penalty for the torsion required for binding of these compounds is significantly large which explains the poor DNA binding affinity obtained in biophysical experiments.¹² Furthermore, it can be said that the size and properties of the cationic moieties of these guanidine-like diaromatic MGBs have an optimal limit which is that of the 2-aminoimidazolinium cation bearing a van der Waals surface area of 141.5 Å².

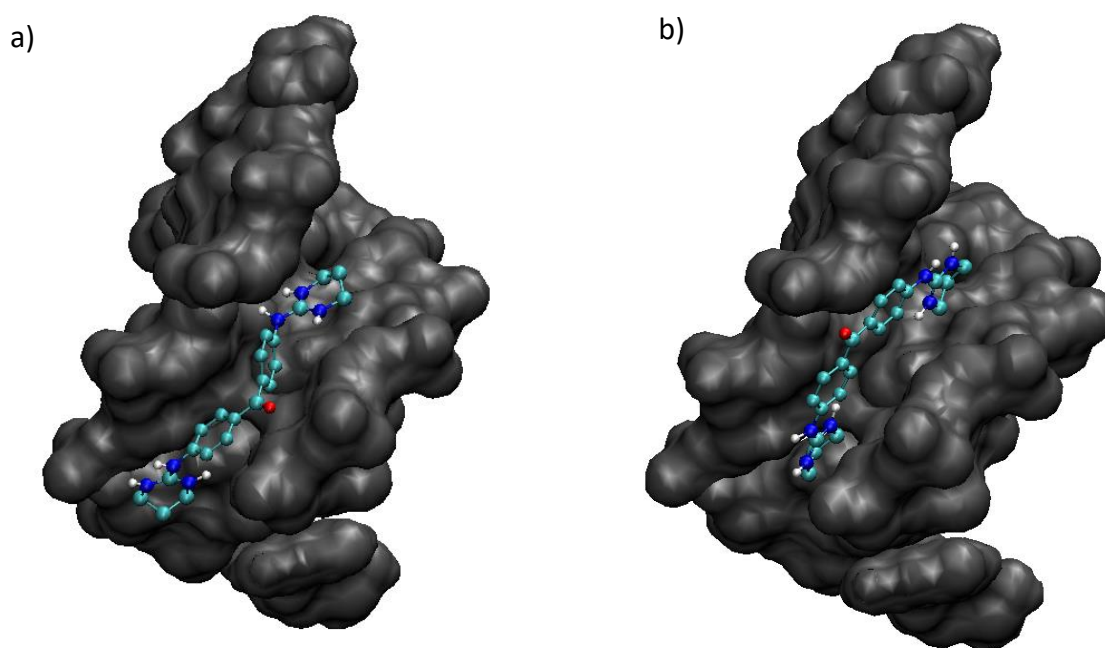


Figure 3.1.2.6. Compound **26** bound to DNA using (a) flexible-rigid and (b) rigid-rigid approaches.

3.1.3. Molecular docking of previous Family E in the DNA minor groove and DHFR

To further exploit the DNA binding activity of our in-house library of compounds another series of derivatives were designed by combining structural components of previous compounds with that of known MGBs such as netropsin and pentamidine. In early investigations the Rozas group prepared mono guanidinium-like derivatives, leaving one of the aromatic amino groups unsubstituted (Figure 3.1.3.1). These compounds showed very poor affinity toward DNA which was probably due to their mono cationic nature.¹⁴

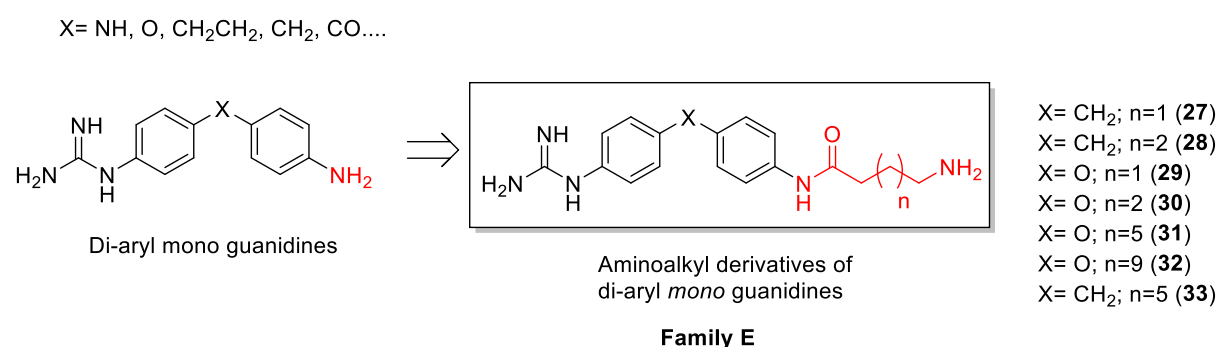


Figure 3.1.3.1. General structures for previously made di-aryl mono guanidines and aminoalkyl derivatives (Family E).

Taking this into account, Family E had been created by condensing aliphatic aminoalkyl chains as seen in pentamidine to the mono cationic di-aromatic scaffold through an amide bond, similar to that of netropsin. As the terminal amine will be protonated at physiological pH, the desired di-cationic property is retained. The added advantage of the aminoalkyl chain is the potential increased binding affinity due to the enthalpic displacement of the spine of hydration due to the hydrophobic effect. Similar to previous studies various linkers (X) were explored as well as aminoalkyl chains of different lengths ranging from 3 to 11 methylene carbons. DNA thermal denaturation experiments showed a trend in increased binding affinity with an increase in the aliphatic chain length. However, derivatives with very large aminoalkyl chains gave poor ΔT_M values due to excessive lipophilicity and flexibility.¹⁵

To gain a better idea of the distance between the guanidinium and the ammonium cations in these aminoalkyl derivatives, their structures were optimized using M062X/6-31+G* and PMC

for water solvation. Considering the high flexibility of the aminoalkyl chains, several conformations would be possible; however, only that with the aminoalkyl chain completely extended was considered since it is the most stable. Correlating the intercationic distance (guanidinium-alkylammonium) with experimental ΔT_M values it was evident that a distance of 18 – 19 Å (4 - 6 methylene carbons) is optimal for good binding. However, due to the high flexibility, it cannot be discarded that aminoalkyl chains that are slightly longer could adopt a conformation that would result in a good binding to DNA. Using the DFT optimised structures, these derivatives were docked in a flexible-rigid approach into the model of the minor groove as used previously. These compounds produced good binding modes with Gscore values ranging from -7.0 to -9.0 kcal mol⁻¹. As seen in Figure 3.1.3.2, compounds **31** and **32** bind in a curved geometry to complement the structure of the minor groove while establishing HB interactions with T-O2, T-deoxyribose-04, C-O2 and electrostatic interactions with backbone phosphate moieties. These anchoring interactions allows the aliphatic chain to be spread along the groove creating additional hydrophobic interactions. In the case of compound **32** with alkyl chain lengths of 11 methylene groups results in protrusion of the excessively long alkyl chain out from the groove resulting in weaker binding (Gscore = -7.7 kcal mol⁻¹). Comparison with compound **31** (alkyl chain of 7 methylene groups) that gives a Gscore = -8.5 kcal mol⁻¹ may explain the higher binding affinity obtained by compound **31** versus compound **32** in biophysical thermal denaturation studies.¹⁵

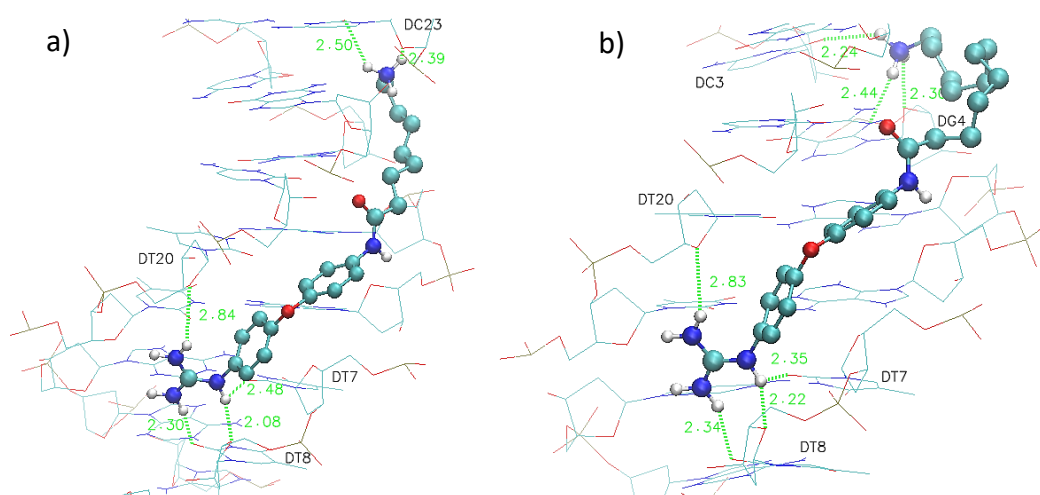


Figure 3.1.3.2. Optimal docking poses of compounds **31** and **32** from Family E in the DNA minor groove.

As mentioned previously, certain miscorrelations between DNA binding affinity and antimalarial activity were apparent for these aminoalkyl guanidine derivatives (Family E) with a general weak linear correlation ($\Delta T_M = -0.646 \times (IC_{50}) + 6.183$; $R^2 = 0.648$). This led to the hypothesis that another mechanism of action may be at play. One possible route could be the inhibition of *Pf*DHFR. Due to the high structural similarity with dihydrofolate and some inhibitors of *Pf*DHFR such as methotrexate and WR9921018 (Figure 3.1.3.3), molecular docking studies were conducted to analyse the feasibility of these aminoalkyl derivatives as potential DHFR inhibitors.

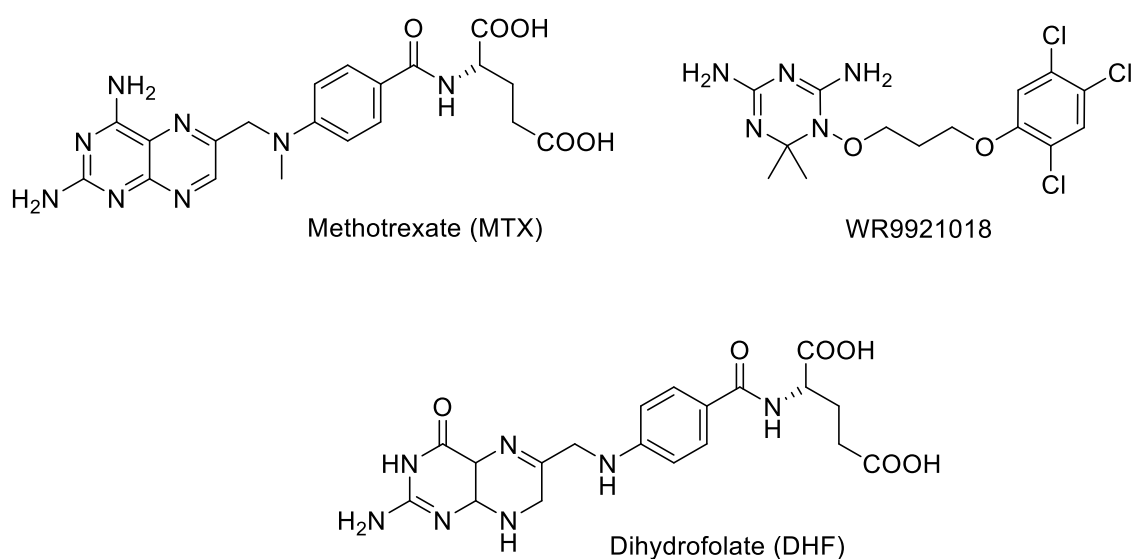
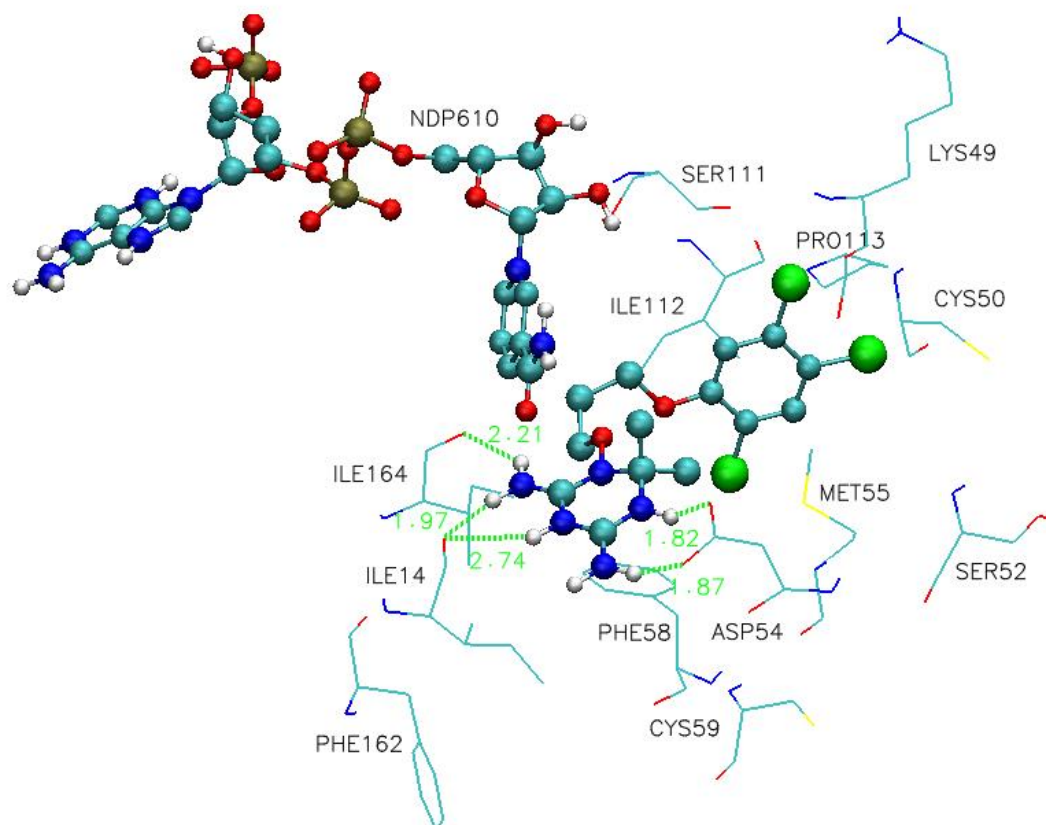


Figure 3.1.3.3. Structures of dihydrofolate (DHF) and *Pf*DHFR inhibitors methotrexate (MTX) and WR9921018.

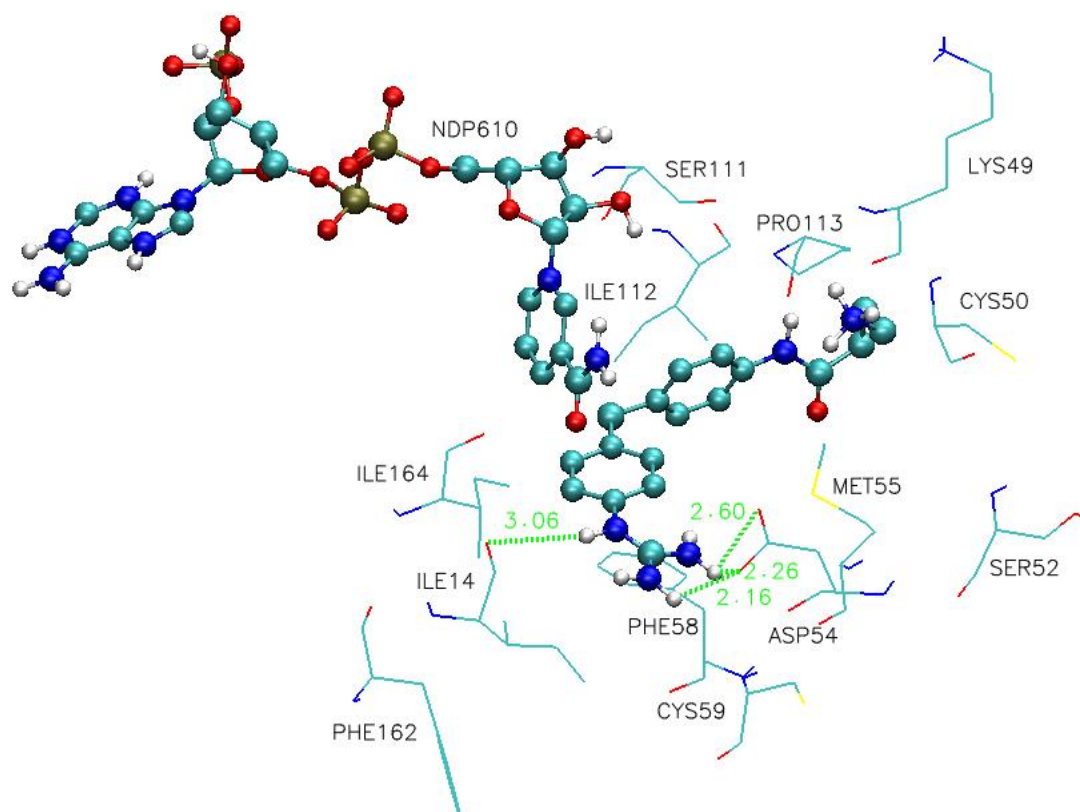
As compounds **27-30** displayed poor DNA binding but good antimalarial activity, it may be possible that these shorter chained derivatives with tetrahedral linkers elicit their activity through DHFR inhibition. For this reason, they were chosen as test ligands for molecular docking along with compound **31** to assess the binding of longer chained derivatives. Figure 3.1.3.4 (a) shows the crystal structure of a structurally related DHFR inhibitor WR9921018 bound to the *Pf*DHFR active site (PDB 1J3I). The binding pocket consists of two distinct domains being the NADPH binding domain and the dihydrofolate binding domain bearing a polar and a hydrophobic region. In the crystal structure, compound WR9921018 forms HB

with ASP54, ILE164 and ILE14A via the charged pyrimidine ring as well as van der Waals interactions with the hydrophobic region formed by CYS15, MET55, PHE58 and TYR170 via its aliphatic linker and p-chlorophenyl moiety. Using this crystal structure as a template, WR9921918 was removed and compounds **27-31** were docked following preparation and optimisation of docking parameters. Considering the binding mode of all test compounds, various interactions were observed to be formed with the binding site residues. All compounds studied present bifurcated and single HB interactions with ASP54 as well as a strong electrostatic interaction. Within that region of the binding site additional HB interactions with ILE14 are established in the case of compounds **27** and **30** (Figure 3.1.3.4 (b) and (e)) whereas compounds **28**, **29** and **31** (Figure 3.1.3.4 (c), (d) and (f)) form a HB interaction between the guanidinium NH and the carbonyl of the nicotinamide moiety of NAD instead. In all compounds, π - cation interactions between the guanidinium cation and the aromatic ring of PHE58 are also apparent. Also, π - π interactions with the nicotinamide aromatic ring of NAD and that of PHE58 are formed. The sum of these interactions anchors the compounds into the inner pocket of the binding site allowing the hydrophobic portion of the molecule to orient towards the hydrophobic region formed by ILE112, PRO113, LYS49, CYS50 and MET55. It is also important to note that the flexible nature of these linkers allows the molecule to orient in this favourable L-like shape conformation, which would differ in the case of the CO and NH linked derivatives presenting a slightly more rigid and coplanar geometry. Comparing the Gscore values now obtained (Table 3.1.3.1, entry 1-5), the shorter chained compounds (**27 – 30**) show greater binding affinity to *Pf*DHFR than compound **31**, which has a slightly longer aminoalkyl chain. This may be attributed to the fact that the aminoalkyl chain of the shorter molecules can optimally occupy the hydrophobic pocket forming van der Waals interaction whereas the aminoalkyl chain of compound **31** protrudes out from the pocket and orient towards the aqueous surface. This in turn perturbs its interactions within the inner binding pocket which is evident by the reduced number of HB interactions formed.

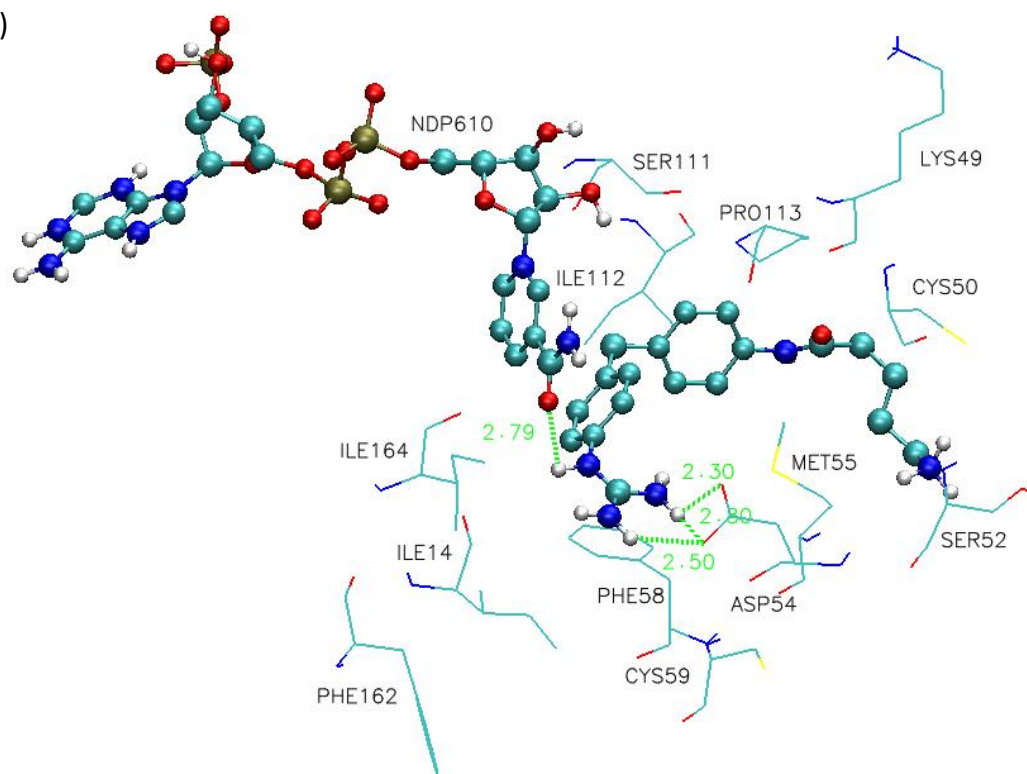
a)



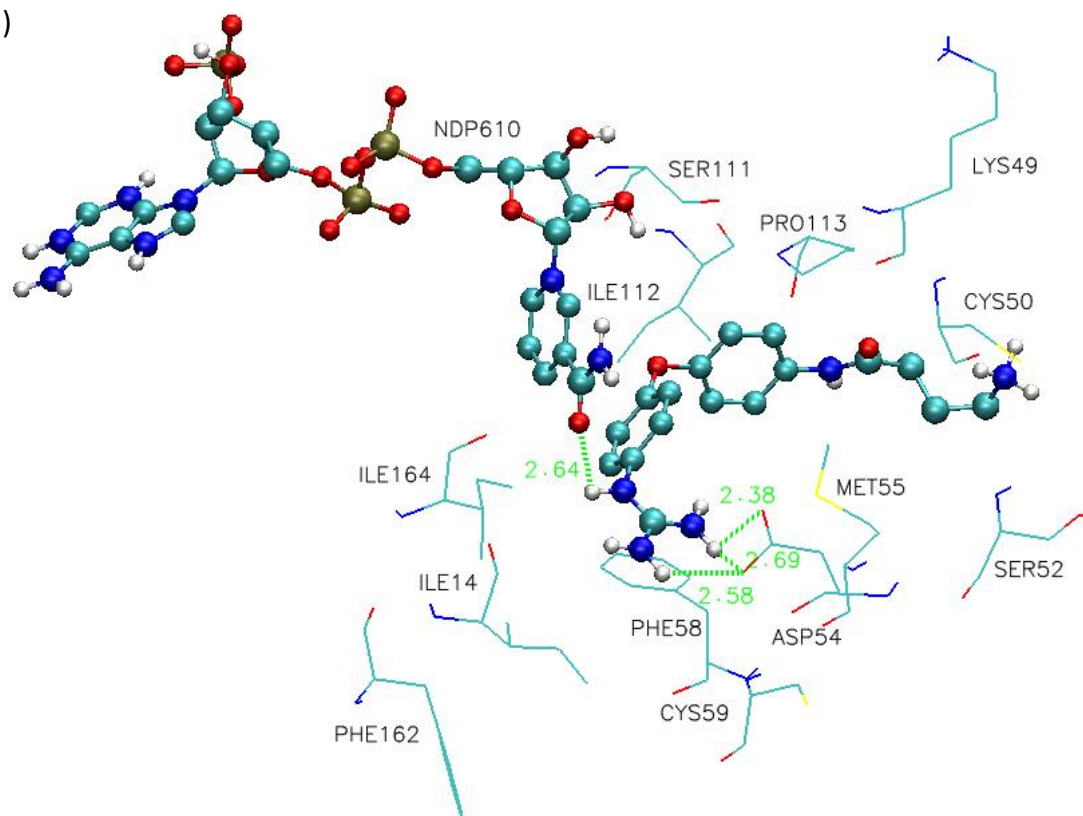
b)



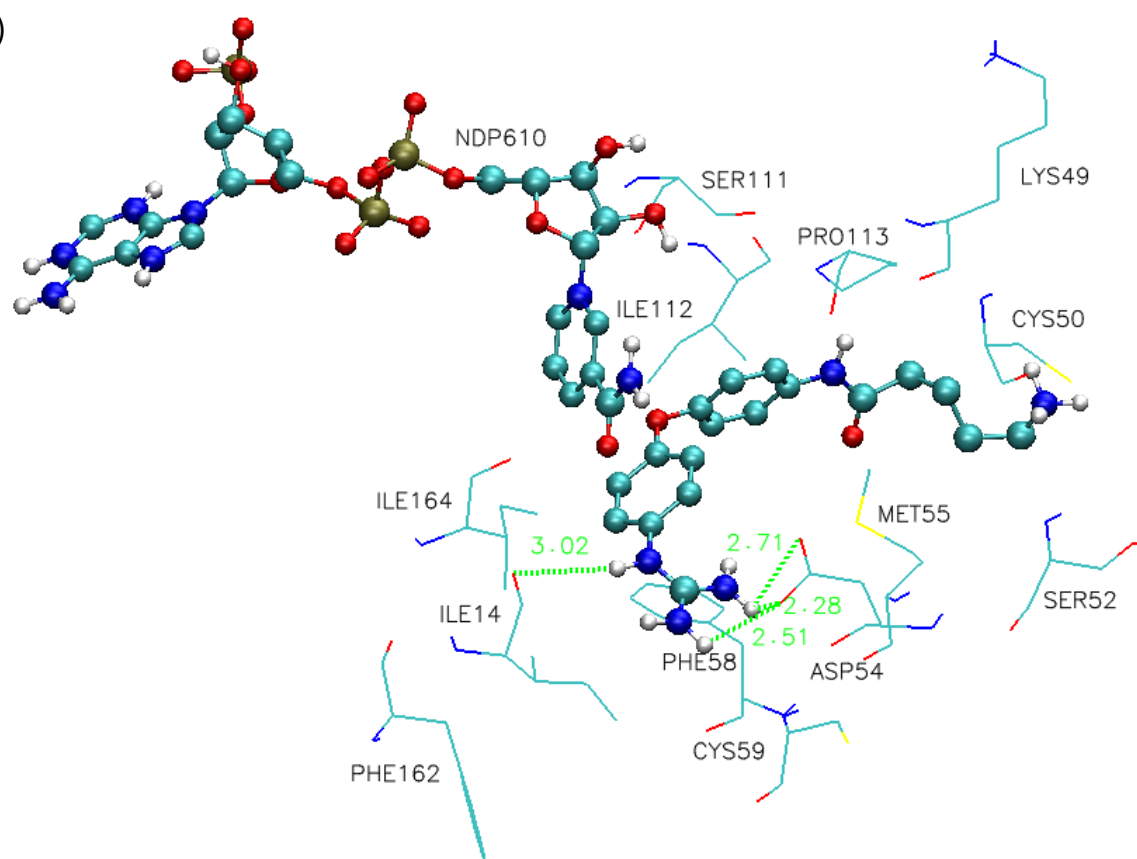
c)



d)



e)



f)

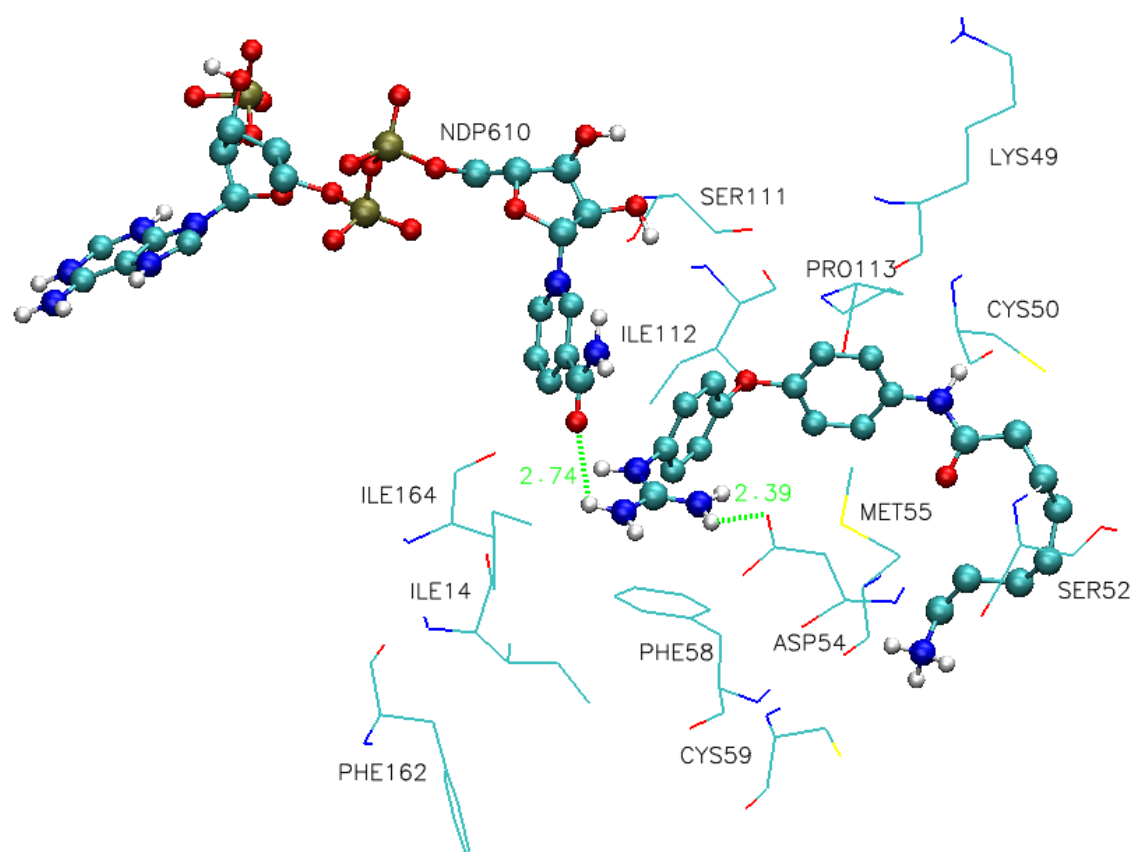
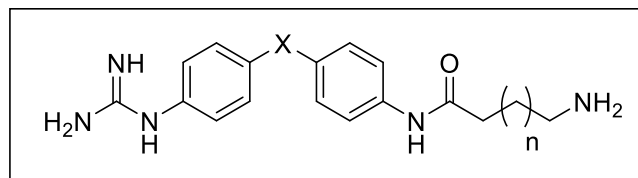


Figure 3.1.3.4. Docking poses for WR9921018 (a) and compounds **27** (b), **28** (c), **29** (d), **30** (e) and **31** (f) in *Pf*DHFR.

Table 3.1.3.1. Gscore values (kcal mol⁻¹) obtained for compounds **27-31** bound to *pf*DHFR.



Entry	Cmp	Linker (X)	Chain length (n)	Gscore (kcal mol ⁻¹)
1	27	CH ₂	1	-7.8
2	28	CH ₂	2	-7.4
3	29	O	1	-7.3
4	30	O	2	-7.5
5	31	O	5	-7.2
6	WR9921018	-	-	-9.6

Poses with appropriate binding mode and Gscore were chosen among all possible modes.

The high similarity of the human DHFR and *Pf*DHFR binding site is one of the major causes of the lack of selectivity of early DHFR inhibitors such as pyrimethamine and cycloguanil. However, in humans, DHFR is bound to its cognate RNA and binding of inhibitors causes RNA dissociation which leads to rapid translation of DHFR providing rapid intrinsic protection to humans. For this reason, compounds **27-31** were also docked into the binding site of human DHFR using a reference crystal structure of folic acid bound to human DHFR (PDB: 4M6K). The main difference between the human and plasmodium DHFR binding site is the replacement of ARG54 with GLU30, ILE14 with THR136 and MET55 with PHE31. Despite these differences, our studies showed that similar ligand binding interactions are established. As seen in Figure 3.1.3.5, HBs and electrostatic interactions are formed with GLU30 which anchor the ligand deep within the binding pocket. The “L-like” curvature allows the aliphatic tail of the molecule to occupy the hydrophobic pocket formed by PHE31, PRO61, ILE60 and the backbone of

GLN35 and ASN64. In addition, HB interactions between the carbonyl of the amide bond and ASN64 are formed as well as HB and electrostatic interactions between the adjacent carboxylate group and ARG70. The terminal carboxylate group extends towards the outer surface of the protein interacting with the aqueous solvent. Another key set of interactions to note is the network of π - π interactions formed between PHE34, PHE31, the nicotinamide moiety of NADPH and the two aromatic rings of folic acid. Given the high structural similarity between Family E compounds and folic acid as seen in the superimposition model using compound **33** ($X=CH_2$; $n=5$) (Figure 3.1.3.5 (b)), similar binding interactions are expected.

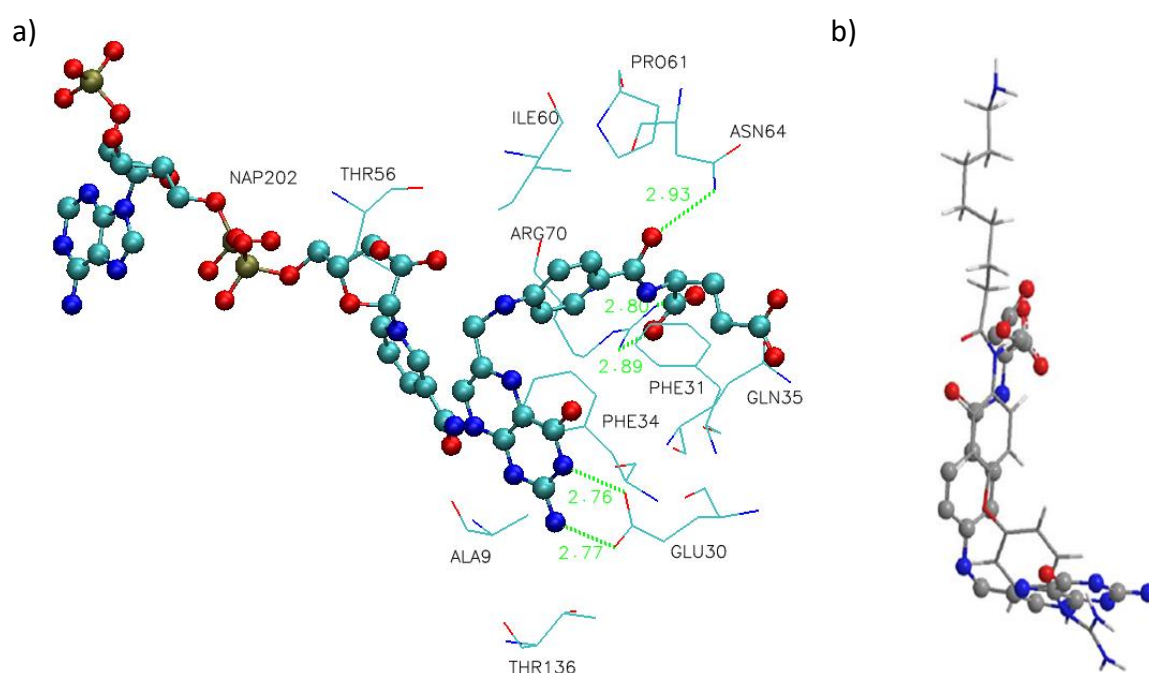
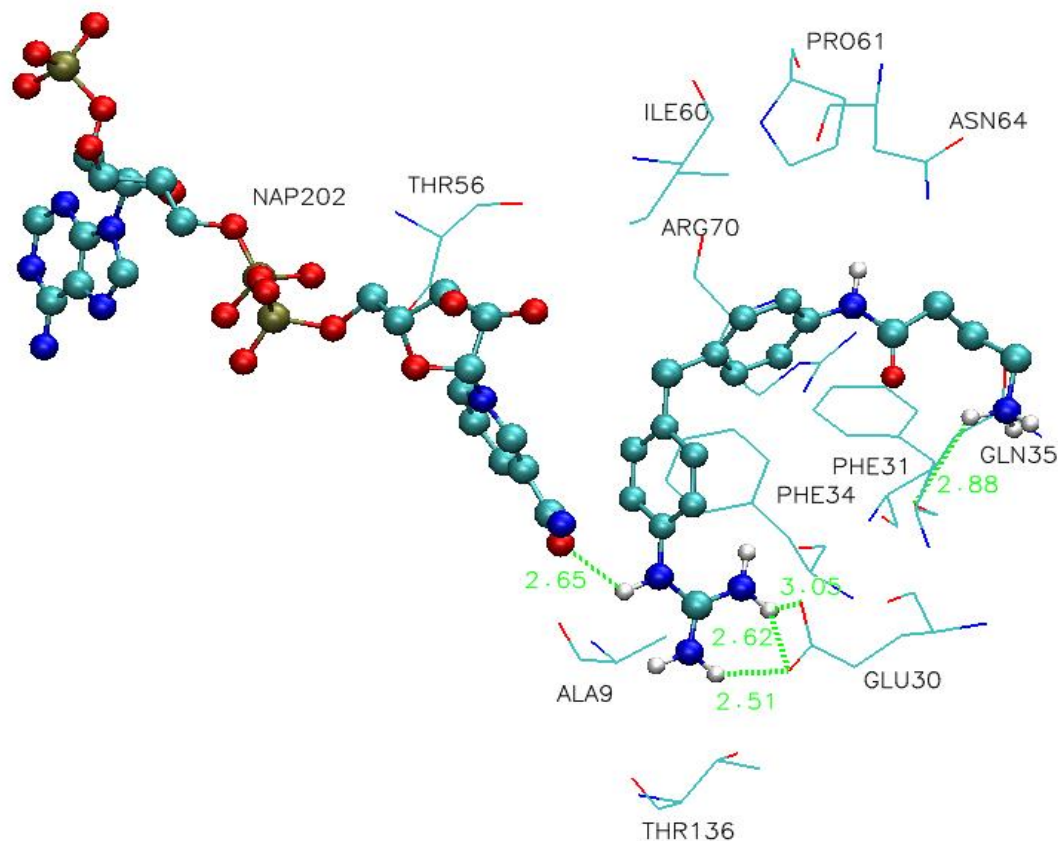


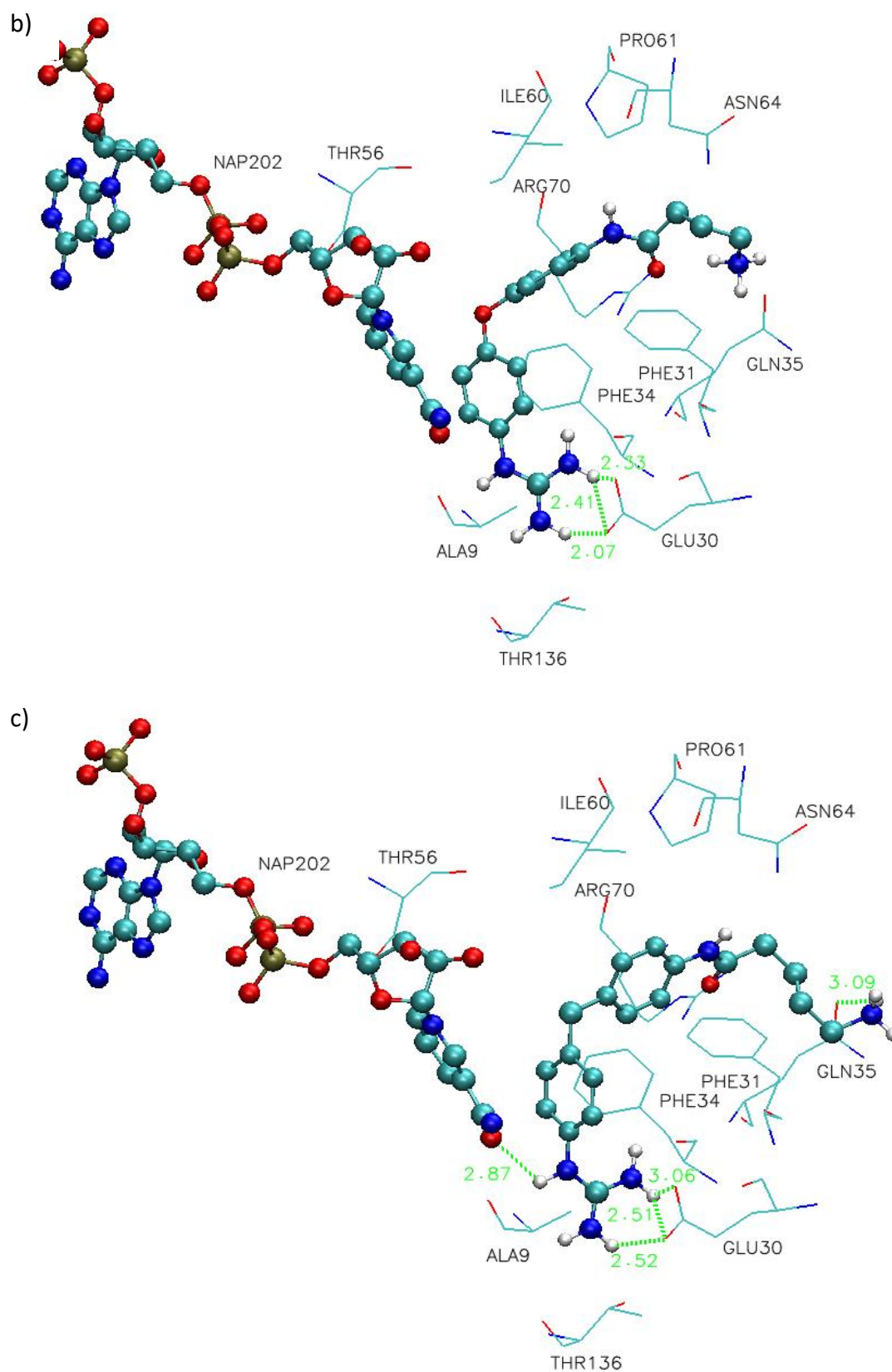
Figure 3.1.3.5. (a) Crystal structure of folic acid bound to *h*DHFR (PDB: 4M6K) and (b) superimposition of compound **33** with folic acid.

As shown in Figure 3.1.3.6, compounds **27-31** occupy the DHFR binding site in the favourable “L-shaped” conformation establishing key HB/electrostatic interactions with GLU30, HB interaction with THR136 and π - π interactions with PHE34, PHE31 and NADPH. The absence of the carboxylate functional group as seen in folic acid means that the aliphatic tail is not anchored within the hydrophobic pocket at the same position through interaction with ASN64 and ARG70. However, the terminal ammonium cation seems to play a role in situating the

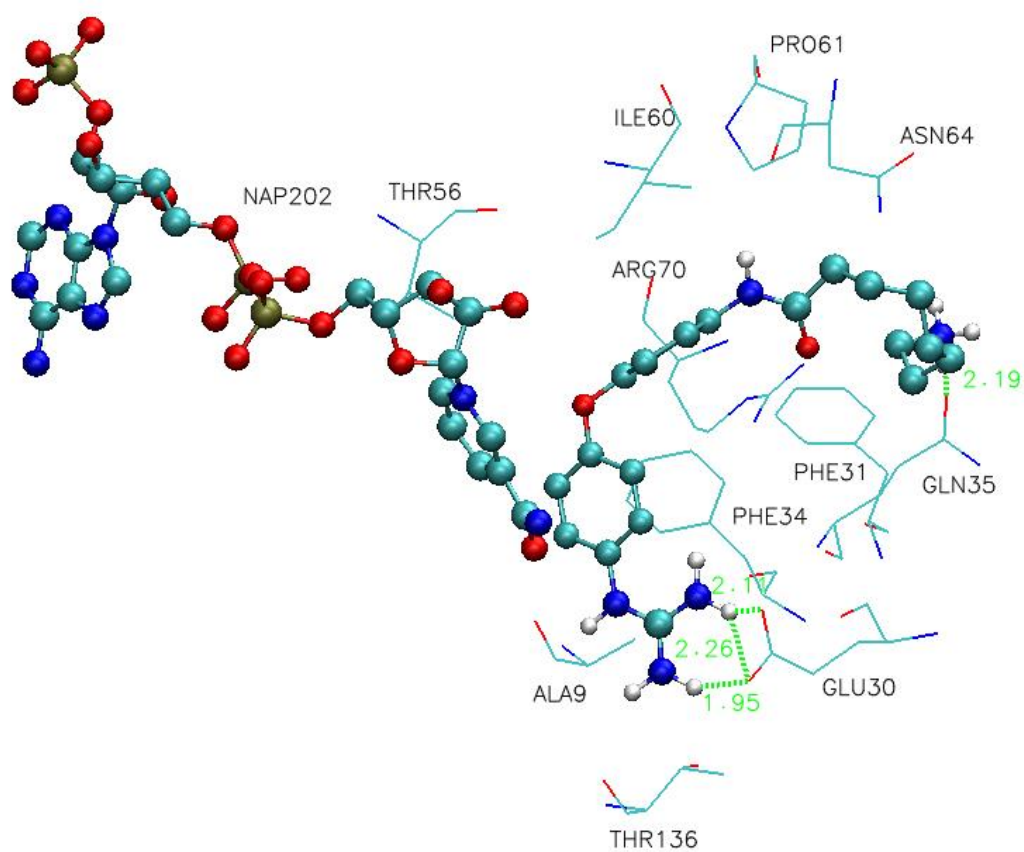
aliphatic chain of compounds **27-31** in a similar region of the hydrophobic pocket. The geometry of compounds **29** and **30** bearing an O linker allows the ammonium cation to establish π -cationic interactions with PHE31. On the other hand, the terminal ammonium cation of compounds **27** and **28** (X= CH₂ linker) forms HB interaction with GLN35 or the carbonyl of PHE31. Interestingly, when considering the effect of the carbon chain length on binding it is evident that compound **31** with a longer aliphatic chain is tolerated which differs from the outcome from *Pf*DHFR docking. This comes as a surprise as the substitution of MET55 in *Pf*DHFR with PHE31 in *h*DHFR causes a decrease in the ligand binding site space which would hinder the binding of large inhibitors. However, in *h*DHFR the aliphatic chain of compound **31** seems to be anchored within the hydrophobic region through HB interaction of the terminal ammonium cation with GLN35. This explains why compound **31** displays similar Gscore values and binding modes to the rest of the compounds in the *h*DHFR model and not the *Pf*DHFR model. Moreover, these studies explain why derivatives with shorter aliphatic chains may be eliciting their antimalarial activity through dual DNA MGB binding and *Pf*DHFR inhibition.

a)





d)



e)

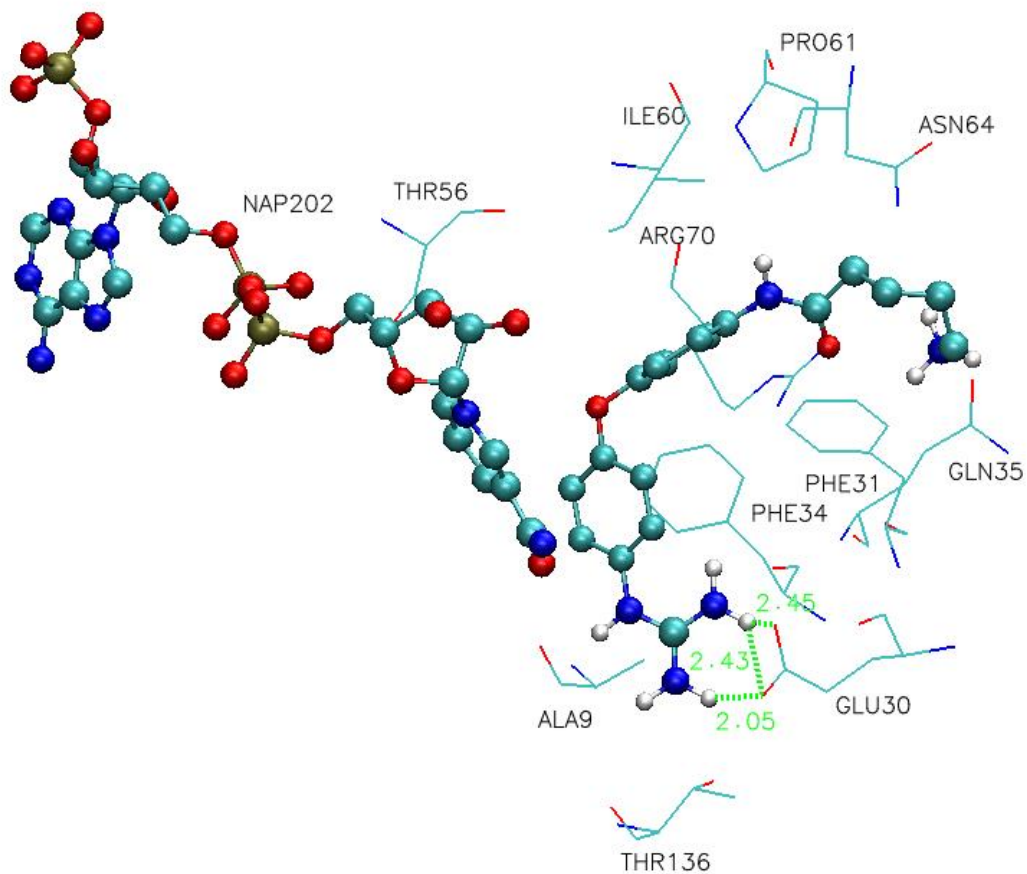
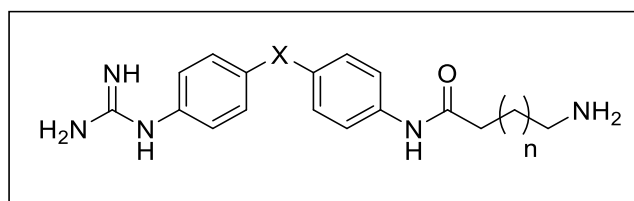


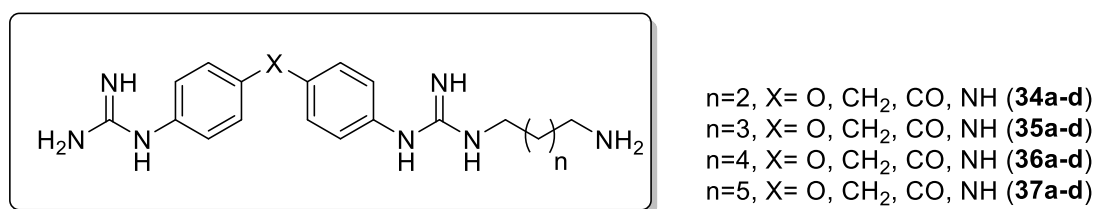
Figure 3.1.3.6. Docking poses of compounds **27** (a), **28** (b), **29** (c), **30** (d) and **31** (e) in *h*DHFR.**Table 3.1.3.2.** Gscore values (kcal mol⁻¹) obtained for compounds **27-31** bound to *h*DHFR.

Entry	Cmp	Linker (X)	Chain length (n)	Gscore (kcal mol ⁻¹)
1	27	CH ₂	1	-9.1
2	28	CH ₂	2	-9.2
3	29	O	1	-8.4
4	30	O	2	-8.3
5	31	O	5	-8.3
6	Folic acid	-	-	-9.9

Poses with appropriate binding mode and Gscore were chosen among all possible modes.

3.1.4. Molecular docking of novel Family F in the DNA minor groove and DHFR (Family F)

Given the poor to moderate DNA binding of previously made aminoalkyl derivatives (Family E) and the presence of certain miscorrelations between DNA binding affinity and antimalarial activity a new series of aminoalkyl derivatives of di-aryl bis-guanidines (Family F) was designed in order to create stronger and more potent MGBs (Figure 3.1.4.1).



Novel aminoalkyl derivatives of di-aryl bis-guanidines
Family F

Figure 3.1.4.1 General structure of newly designed MGBs, Family F.

DFT optimisation of these structures shows similar geometry to that of compounds previously studied where the aromatic rings of the di-aryl system appear twisted with respect to each other and the guanidinium cations are almost perpendicular to the adjacent phenyl rings (Figure 3.1.4.2). Similar to Family E, the aminoalkyl chain is represented in a straight chain geometry which represents the lowest energy conformation, albeit the several degrees of freedom. The addition of the aminoalkyl chains to one of the guanidine-like cations instead of attaching it directly to the aromatic ring through an amide bond may give more freedom to the alkyl chain allowing it to optimally interact with the DNA minor groove. Also, the introduction of another positive charge and a decrease in the inter-cationic distance may result in tighter binding through greater interactive contacts to the minor groove.

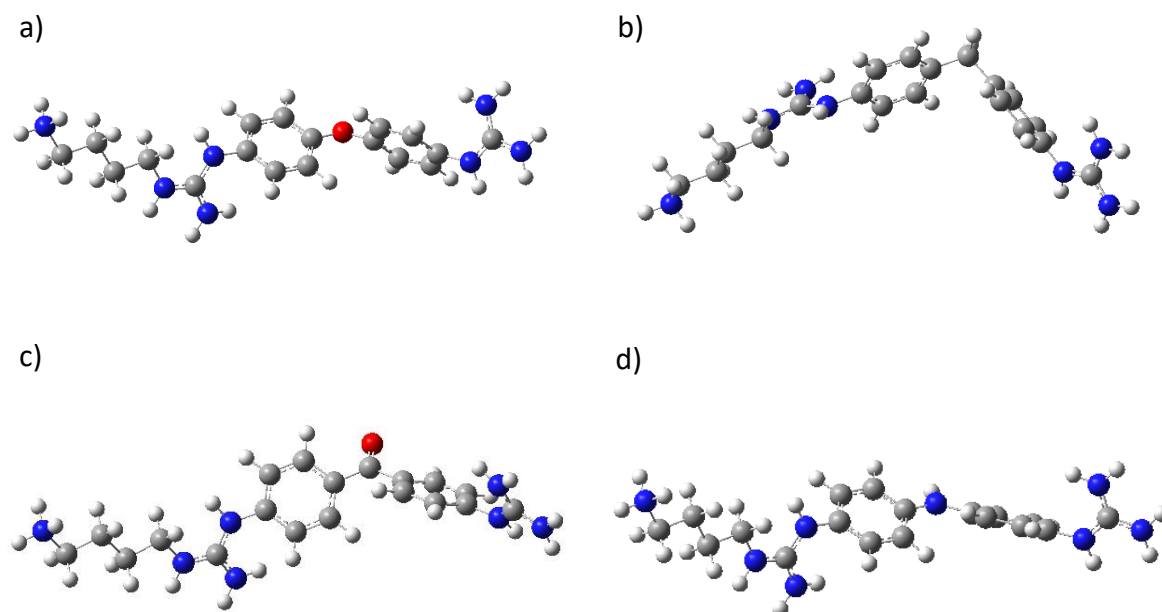


Figure 3.1.4.2. DFT optimised structures of compounds **34a** (a), **34b** (b), **34c** (c) and **34d** (d) of Family F.

Analysing the binding mode of these derivatives to the model of the DNA minor groove, it is evident that similar binding poses as previously obtained are observed. As shown in Figure 3.1.4.3, the di-aromatic system changes from a twisted arrangement to an almost co-planar geometry allowing the molecule to complement the convex curvature of the minor groove. Compounds with CO and NH linkers were shown to produce better GScore values and binding interactions than those with O and CH₂ linkers. This has been observed in previous families of MGBs produced by our group and is due to the lower energetic penalty to adopt the favourable co-planar geometry with respect to the two phenyl rings.¹³ A network of HBs is established in most of the compounds in this family where the terminal guanidinium cation forms HB with O2 of thymine and O4 of the thymine deoxyribose. Additionally, the central guanidinium cation establishes HBs with O3 of the adenine deoxyribose moiety as well as electrostatic interactions with the attached phosphate residue. These strong points of contact anchor the molecule in place allowing the aliphatic chain to wrap around the remaining available space within the groove. In addition to the further introduction of van der Waal interactions facilitated by the hydrophobic chain, the terminal ammonium cation forms HB interactions with the N2 of guanine as well as electrostatic interactions with adjacent phosphates.

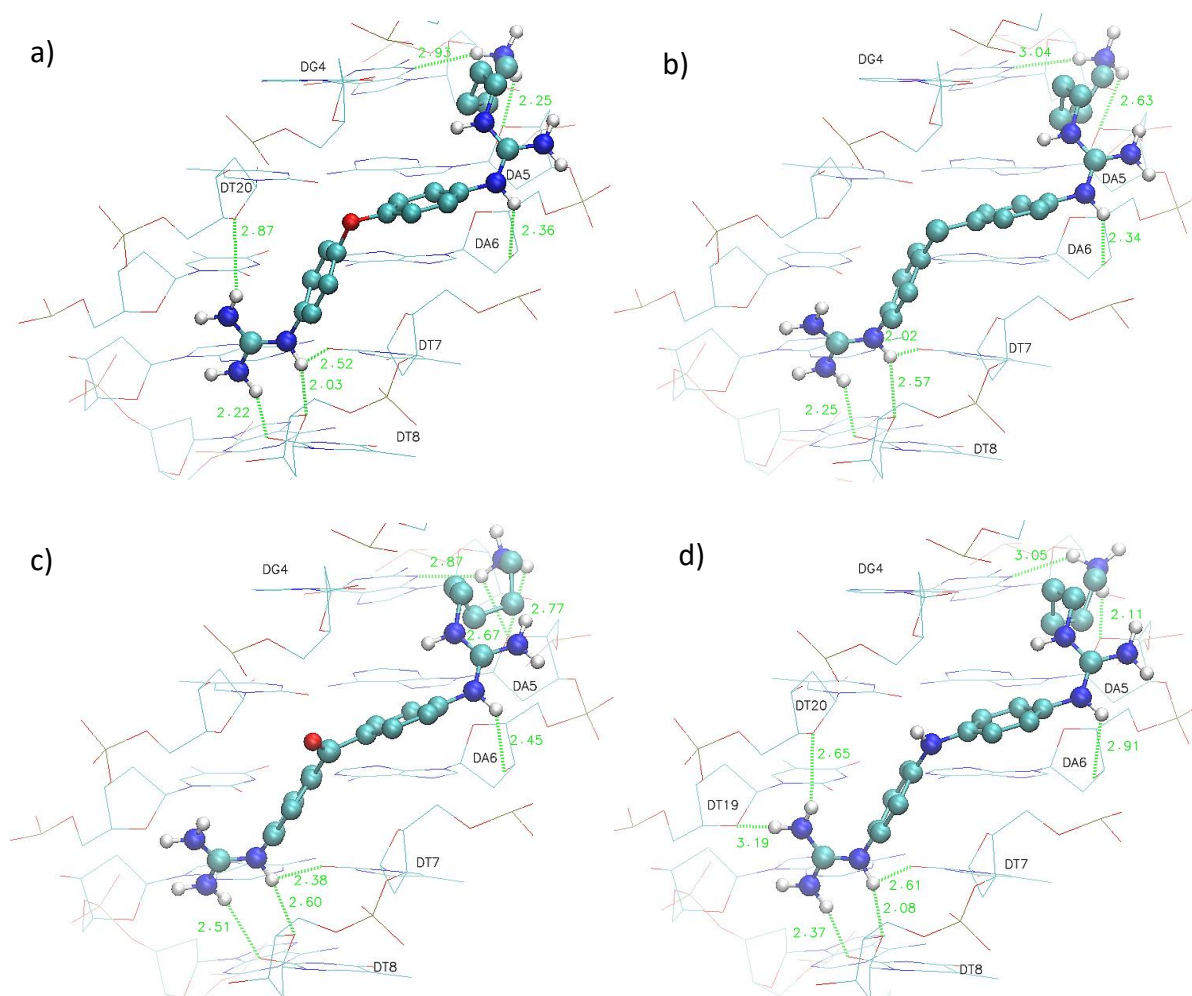
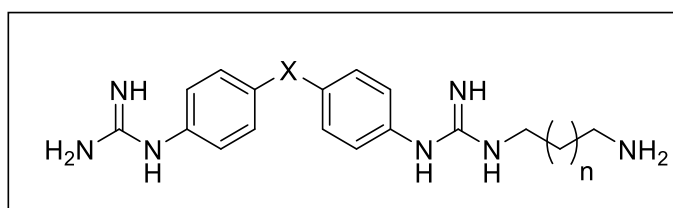


Figure 3.1.4.3. Optimal docking poses of proposed aminoalkylguanidino derivatives **34a** (a), **34b** (b) **34c** (c) and **34d** (d) in a model of the DNA minor groove.

Considering the trend in Gscore values with increasing alkyl chain length, only slight differences were observed (Table 3.1.4.1). Derivatives with the longest alkyl chains such as compounds **37a**, **37b**, **37c** and **37d**, produce higher Gscore values than shorter alkyl chain derivatives of the same linker. This suggests that we may see an increase in DNA binding affinity with increasing alkyl chain length in biophysical studies. From previous biophysical studies of compounds of Family E, derivatives with alkyl chains larger than 7 methylene groups resulted in decreased DNA binding affinity and for this reason novel derivatives with alkyl chains consisting of 4 to 7 methylene will be synthesised. Moreover, based on these positive computational results it is expected that Family F will show enhanced DNA binding affinity.

Table 3.1.4.1. Gscore values (kcal mol⁻¹) obtained for compounds **34-37** bound to the DNA minor groove.



n=2, X= O, CH₂, CO, NH (**34a-d**)

n=3, X= O, CH₂, CO, NH (**35a-d**)

n=4, X= O, CH₂, CO, NH (**36a-d**)

n=5, X= O, CH₂, CO, NH (**37a-d**)

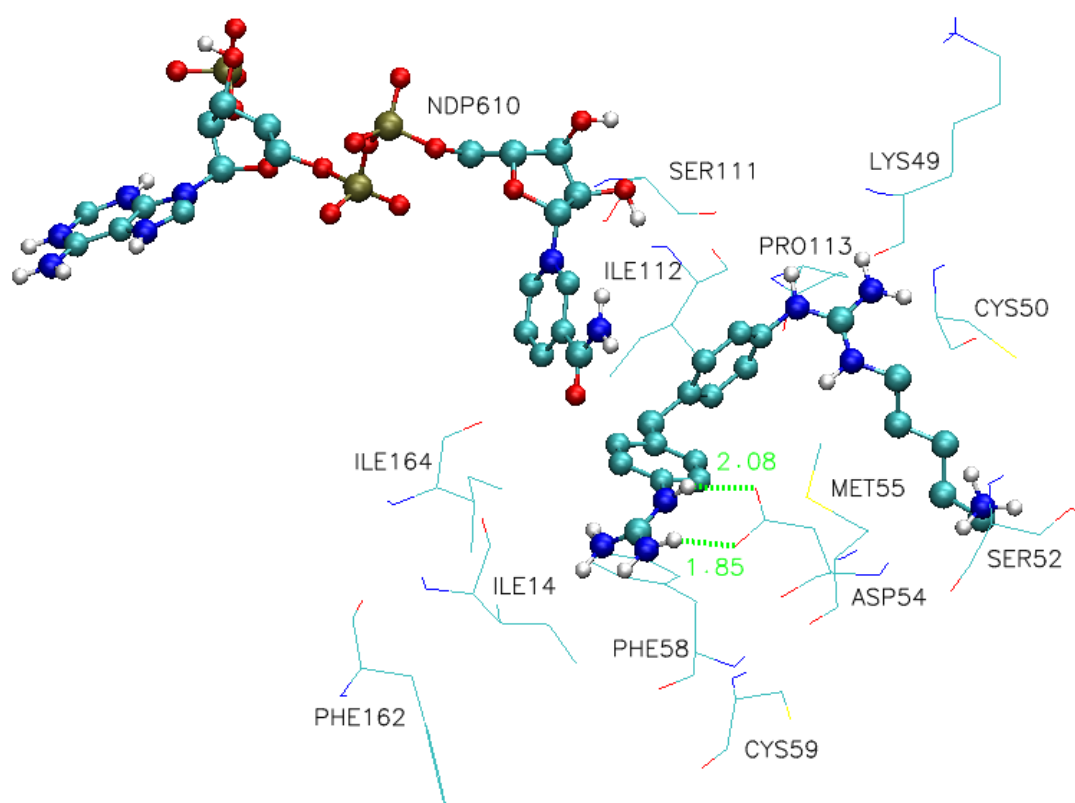
Entry	Cmp	Linker (X)	Chain length (n)	Gscore (Kcal mol ⁻¹)	Entry	Cmp	Linker (X)	Chain length (n)	Gscore (Kcal mol ⁻¹)
1	34a	O	2	-8.8	9	34b	CH ₂	2	-8.7
2	35a	O	3	-8.7	10	35b	CH ₂	3	-8.6
3	36a	O	4	-8.9	11	36b	CH ₂	4	-8.7
4	37a	O	5	-8.9	12	37b	CH ₂	5	-8.8
5	34c	CO	2	-9.1	13	34d	NH	2	-9.2
6	35c	CO	3	-9.0	14	35d	NH	3	-9.3
7	36c	CO	4	-9.1	15	36d	NH	4	-8.9
8	37c	CO	5	-9.3	16	37d	NH	5	-9.4

Poses with appropriate binding mode and Gscore were chosen among all possible modes.

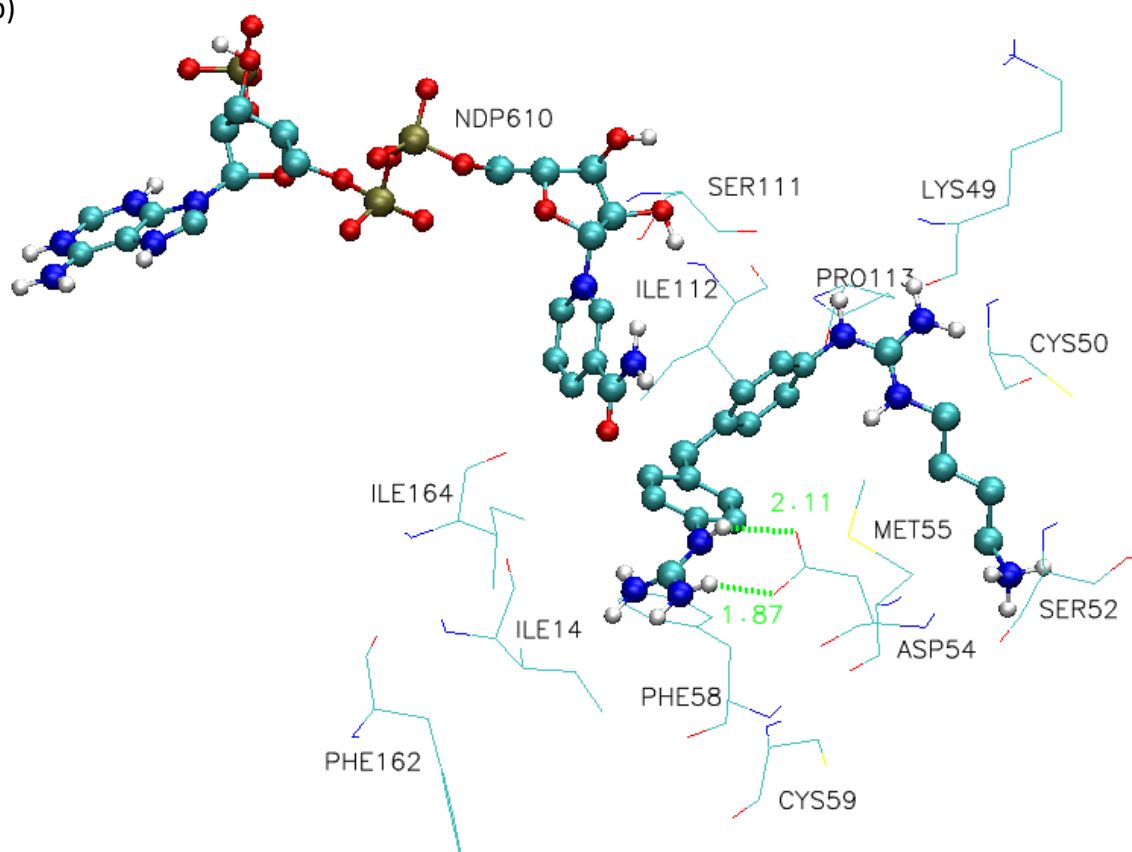
The binding of this new family of compounds to *Pf*DHFR was also investigated in order to identify any possible dual inhibition. Using the same crystal structure of WR9921018 bound to the *Pf*DHFR as a reference compounds **34a**, **34b**, **35a**, **35b** and **37a** were docked into the ligand binding site. Analysing the docking poses, similar interactions to those observed for Family E are observed (Figure 3.1.4.4). Thus, bridging HB interactions and electrostatic interactions with ASP54 are formed as well as π - π interactions with PHE58. The geometry of the linkers enables the molecule to adopt the aforementioned “L”-shaped curvature with both phenyl rings almost perpendicular to each other. The central guanidinium moiety

appears staggered with respect to the attached phenyl ring which points the aliphatic chain away from the hydrophobic pocket and towards the surface of the enzyme interacting with surrounding water molecules. As seen with compounds **34a**, **35a** and **37a** the tail of the molecule is further anchored by HB interactions established between the guanidinium NH and the backbone carbonyl of SER111 (Figure 3.1.4.4 (c), (d), and (e)). As the aliphatic chain does not seem to play a huge role in the binding of these derivatives the effect of the carbon chain length on the Gscore affinity values is negligible.

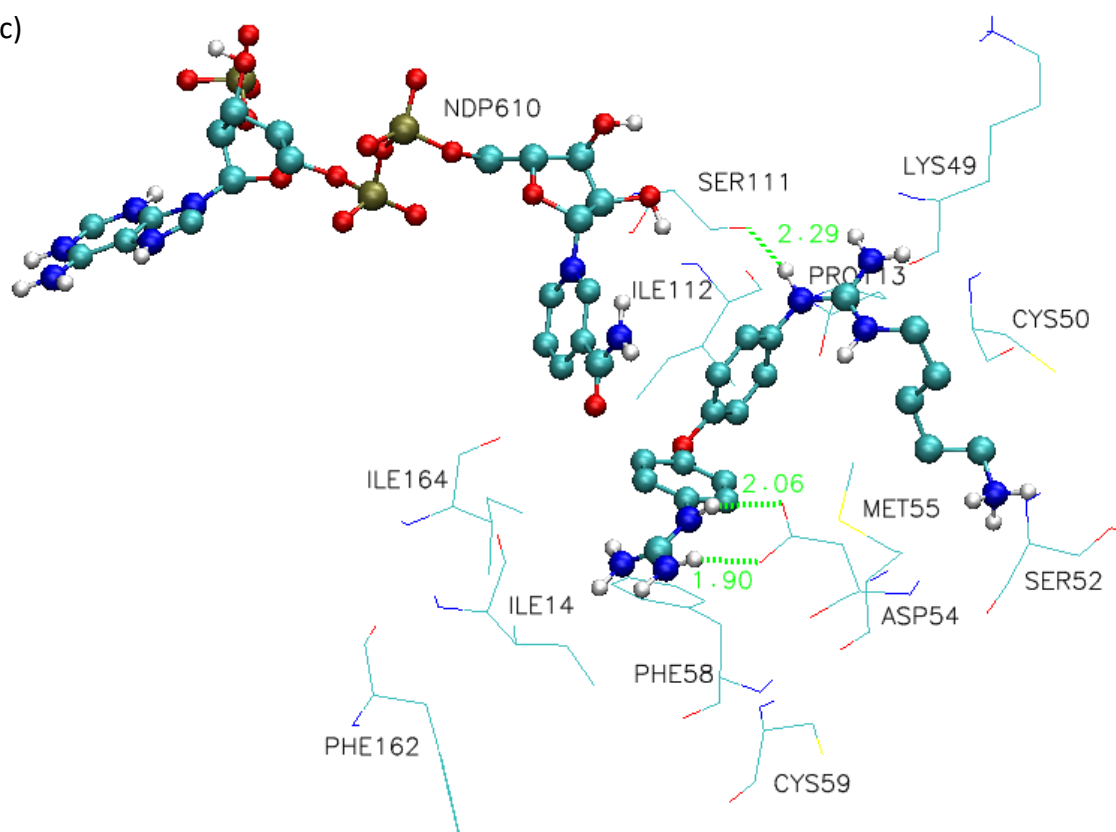
a)



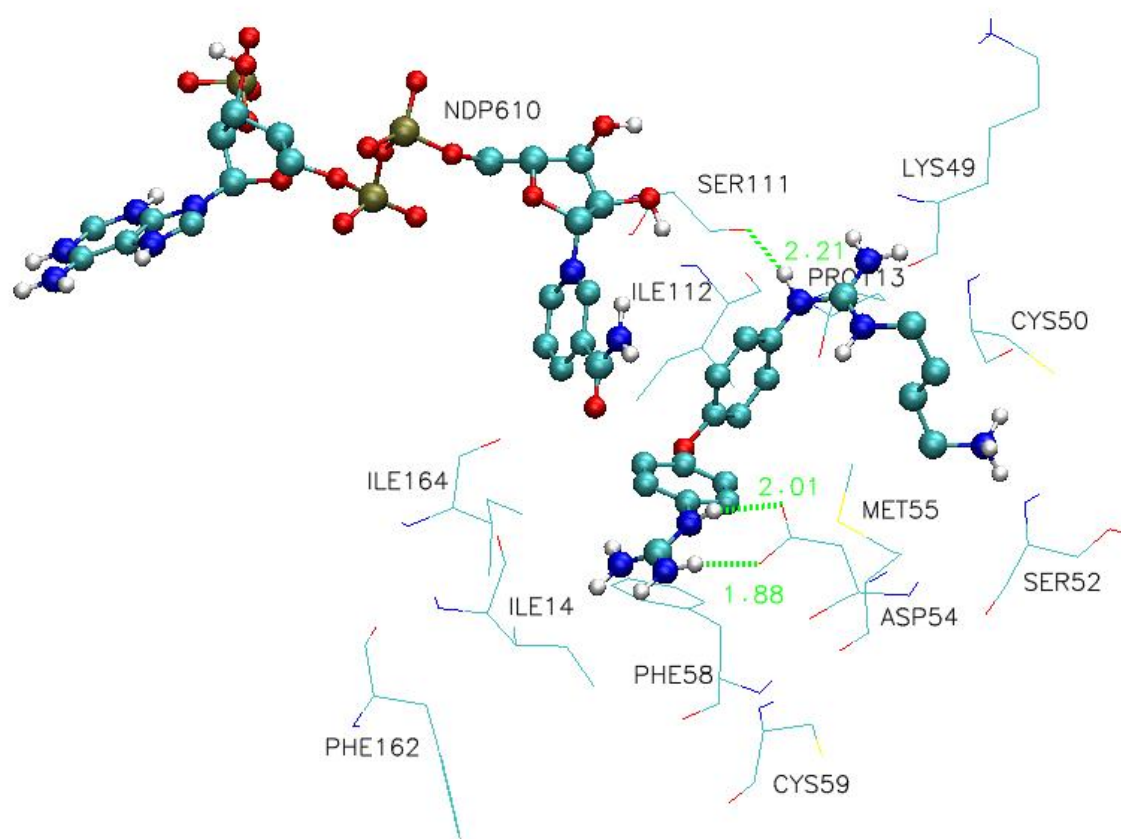
b)



c)



d)



e)

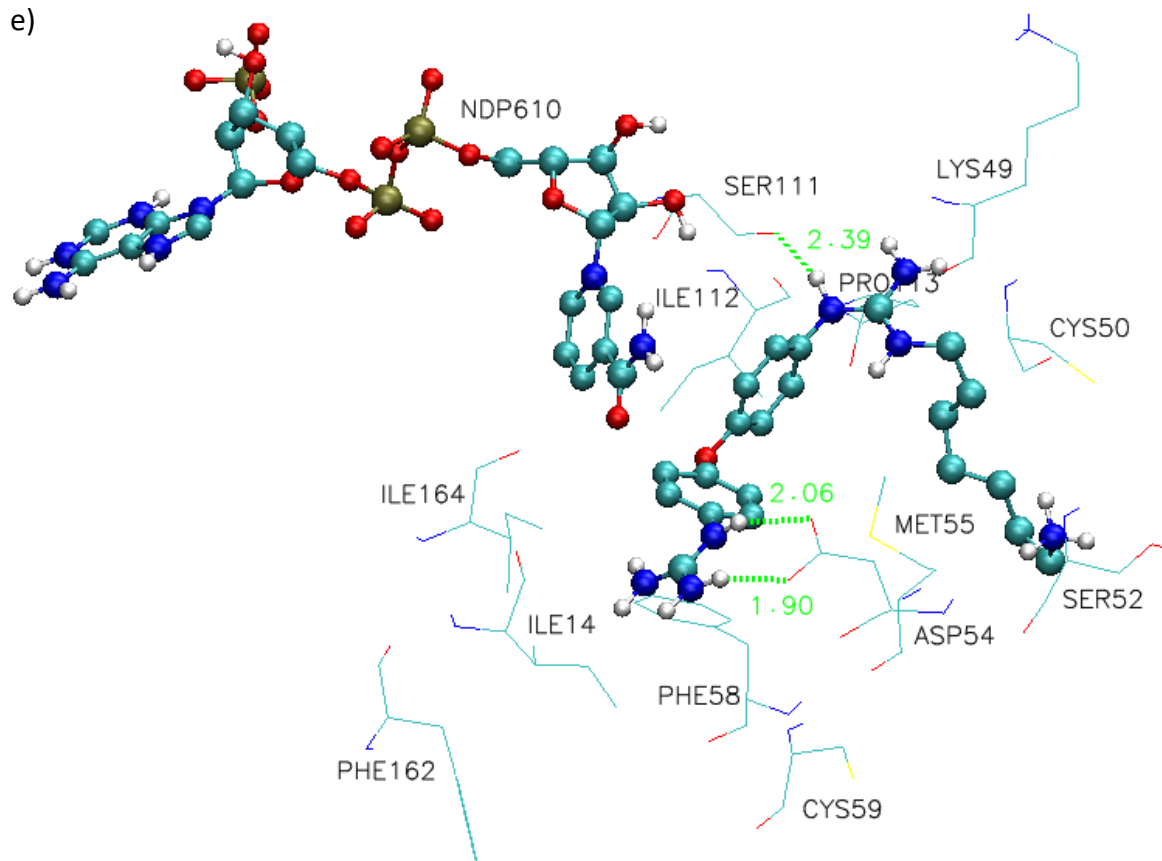
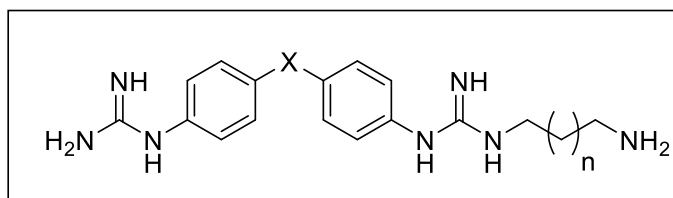


Figure 3.1.4.4. Docking poses for WR9921018 and compounds (a) **34b**, (b) **35b**, (c) **34a**, (d) **35a** and (e) **37a** in *pf*DHFR.

Table 3.1.4.1. Gscore values (Kcal mol⁻¹) obtained for compounds **34a**, **34b**, **35a**, **35b** and **37a** bound to *pf*DHFR.



Entry	Compound	Linker (X)	Chain length (n)	Gscore (Kcal mol ⁻¹)
1	34b	CH ₂	2	-6.8
2	35b	CH ₂	3	-6.8
3	34a	O	2	-6.5
4	35a	O	3	-6.5
5	37a	O	5	-6.4
6	WR9921018	-	-	-9.6

Poses with appropriate binding mode and Gscore were chosen among all possible modes.

In order to assess the possible binding of this new family of compounds towards human DHFR, compound **35b** was docked into the ligand binding site using the same crystal structure as mentioned above (PDB: 4M6K). As shown in Figure 3.1.4.5, HB/electrostatic interactions are established with GLU30 as well as HBs with the adjacent THR136. Similar to compounds **27-31**, π - π interactions with PHE34 and the nicotinamide moiety of NADPH are formed. However, the replacement of the amide with a guanidinium introduces π -cationic interactions with PHE31 which allows the aliphatic chain to form van der Waals interactions with PHE31 and GLN35. This leads to a greater binding affinity Gscore value (-8.4 kcal mol⁻¹) when compared to compound **35b** bound in *Pf*DHFR (-6.8 kcal mol⁻¹). Moreover, these studies suggest that this new family of compounds may exhibit inhibitory activity towards DHFR with an unfavourable specificity towards human DHFR. However, given the protective mechanism of human cells against DHFR inhibitors as mentioned earlier and the unnecessary need to scavenge folates from the cellular environment unlike plasmodium cells, these compounds may not exhibit

consequential side effects. Noteworthy, the weaker DHFR binding affinity scores of this novel family compared to Family E, suggests that they may exhibit less off-target effects. In order to verify these findings, DHFR enzyme-based assays will be considered in the future evaluation of this new family of compounds.

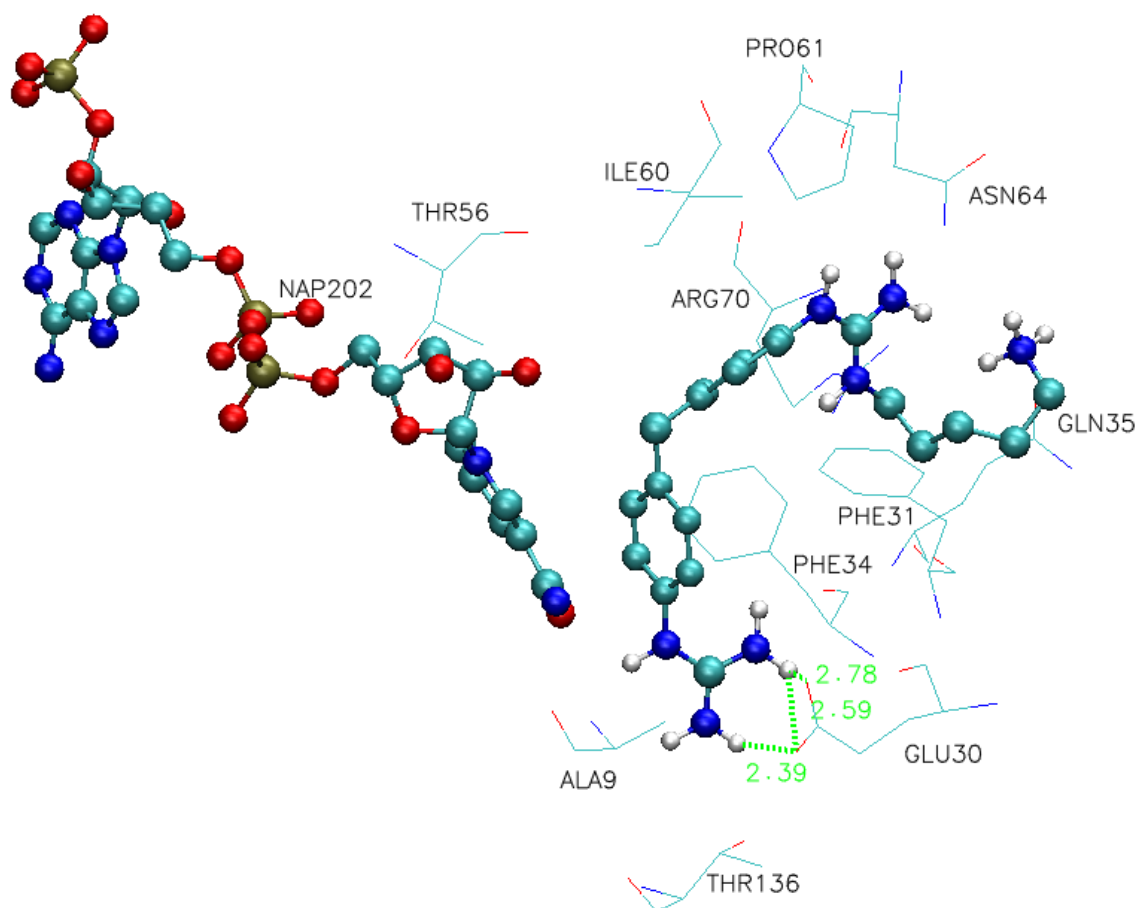


Figure 3.1.4.5. Docking pose of compound **35b** in *h*DHFR.

3.2. Synthesis

As mentioned previously, we have reported in the past the preparation of symmetric and asymmetric di-aromatic guanidinium/2-aminoimidazolinium derivatives (Families A-E), which have shown good DNA minor groove binding affinity and antiprotozoal activity. Based on these results and the computational data presented in Chapter 3.1, we have carried out the synthesis of a novel family of amino-alkyl derivatives of di-aromatic bis-guanidines (Figure 3.2.1). Using well established synthetic methods, compounds with linkers consisting of NH, O, CO and CH₂ were synthesised bearing aminoalkyl chains of four to seven methylene groups.

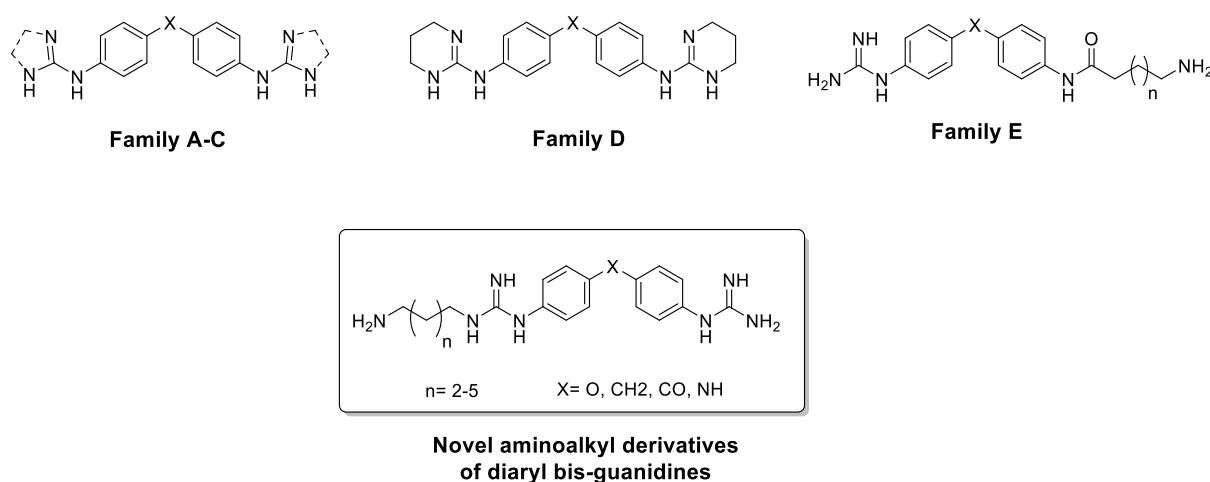


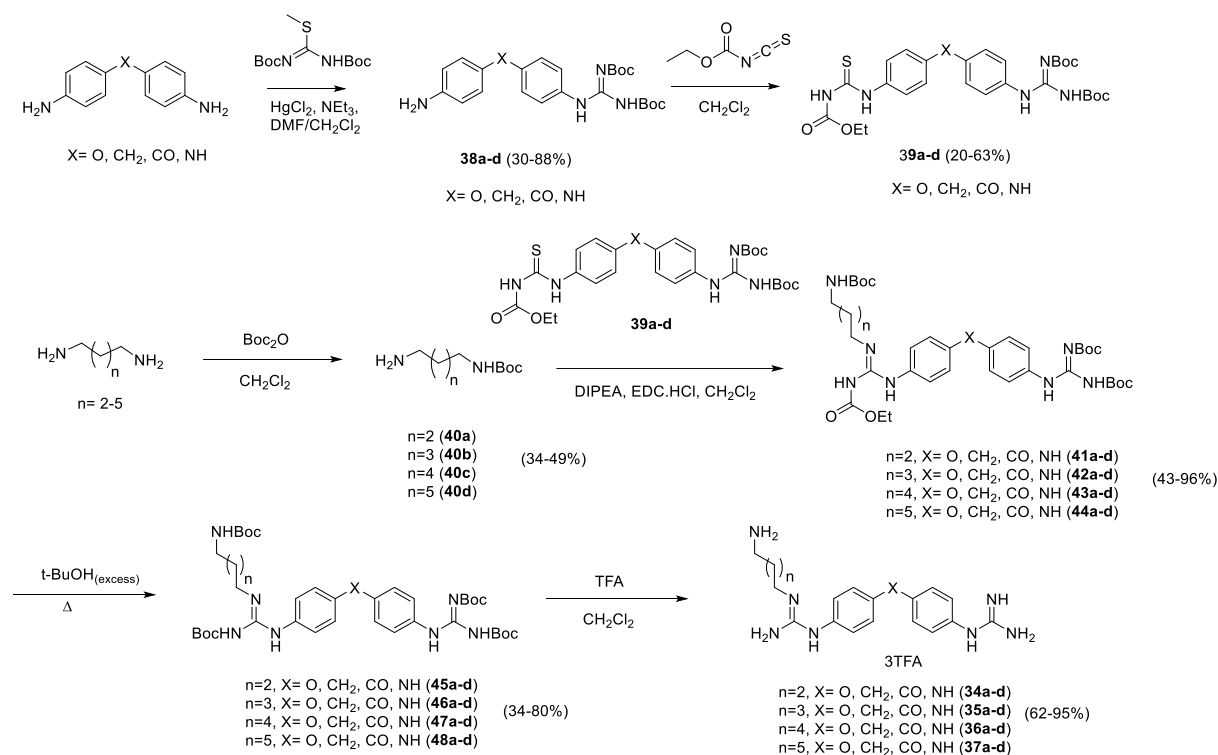
Figure 3.2.1. General structures of previously made guanidine/2-aminoimidazoline based DNA minor groove binders and novel aminoalkyl derivatives of di-aryl bis-guanidines.

3.2.1. Synthesis of novel amino-alkyl derivatives of di-aromatic bis-guanidines

The preparation of the proposed amino-alkyl derivatives of di-aromatic bis-guanidines (Scheme 3.2.1.1), involves first the synthesis of mono-Boc protected di-aromatic bis-guanidines **38a-d** which has been previously described by us.¹⁷ Using these intermediates, the corresponding *N*-ethylcarbamate thioureas (**39a-d**) were produced. In parallel, the preparation of mono-Boc (di-*tert* butyl dicarbonate) protected diamines (**40a-d**) was carried out which are used in the conversion of the *N*-ethylcarbamate thioureas (**39a-d**) into the substituted guanidines (**41a-d** to **44a-d**). Following this, the ethyl carbamates (**41a-d** to **44a-d**) were converted into the *tert*-butyl carbamates (**45a-d** to **48a-d**) via trans-esterification

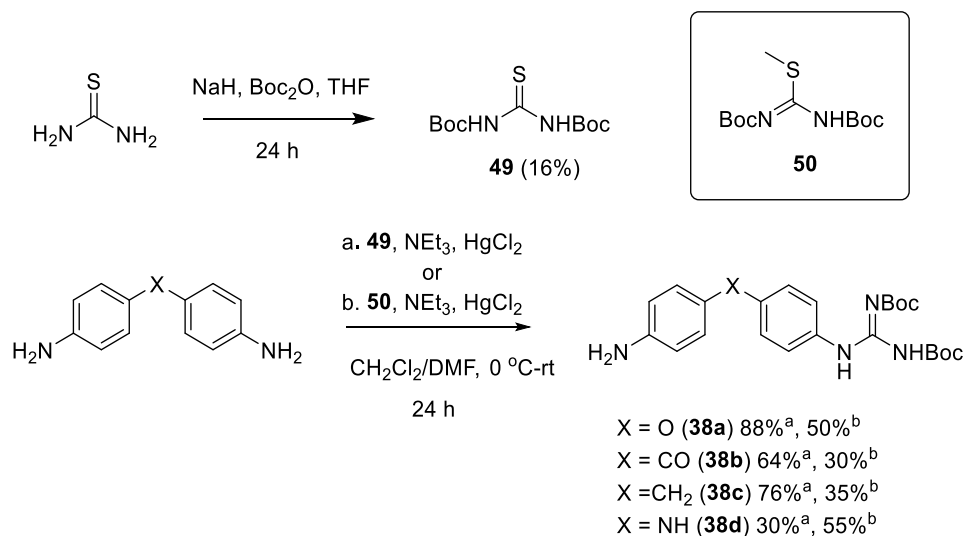
which allows for a one-step deprotection with TFA to yield the final trifluoroacetate salts (**34a-d** to **37a-d**).

Scheme 3.2.1.1. Synthetic route for the synthesis of novel aminoalkyl derivatives



3.2.2. Synthesis of mono-guanidylated di-aromatic systems

The Boc-protected di-aromatic mono-guanidines with linkers containing O (**38a**), CH₂ (**38b**), CO (**38c**) and NH (**38d**) were synthesised by reacting the corresponding commercially available diaromatic diamines with a thiourea based guanidylating agent (Scheme 3.2.2.1). Due to the low nucleophilicity of the starting aromatic amines, the first step was the activation of the corresponding commercial thiourea via Boc protection of the amino groups to yield **49** as proposed by Kim *et al.*¹⁶ Given the simplicity of the reaction and stability under basic conditions the Boc protecting group was chosen. The conditions involved the reaction of thiourea with di-*tert*-butyldicarbonate (Boc₂O) in THF in the presence of sodium hydride under N₂ at 0 °C, allowing the mixture to reach room temperature overnight.

Scheme 3.2.2.1. Synthetic route for the synthesis of Boc-protected di-aromatic mono-guanidines.

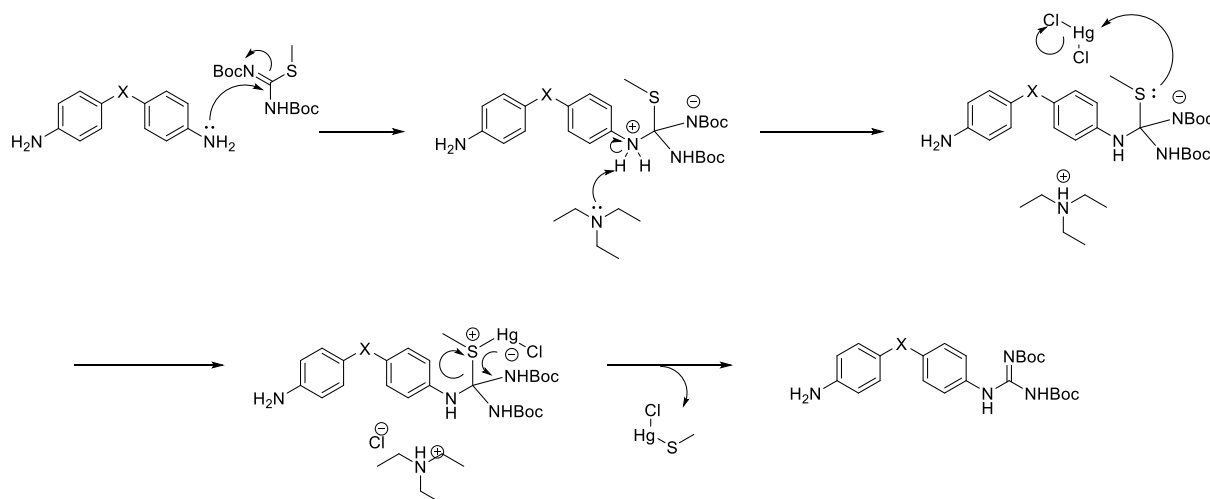
^a Yield of guanidylation using **50**; ^b Yield of guanidylation using **49**

However, due to incomplete Boc protection, solidification during the reaction and tedious workup, the resulting yield was poor (16%). Following this, guanidylation of the corresponding commercial dianilines was conducted with the use of Boc-protected thiourea **49** in a 3:1 ratio in the presence of NEt_3 and HgCl_2 in CH_2Cl_2 or DMF depending on the polarity of the diamine. Upon completion of the reaction as monitored by TLC (Thin layer chromatography), a workup involving dilution with CH_2Cl_2 followed. Next, removal of the HgS by-product by filtering the reaction mixture through a pad of Celite was performed. Following an aqueous and drying over anhydrous $\text{Na}_2\text{SO}_4/\text{MgSO}_4$ the resulting crude solid was purified by silica gel chromatography producing compounds **38a**, **38b**, **38c** and **38d**.

Given the poor yields obtained (30 – 55%), the commercially available 1,3-bis-(*tert*-butoxycarbonyl)-2-methyl-2-thiopseudourea **50** was employed as an alternative guanidylating agent. As shown in Scheme 3.2.2.2, methylation of the sulfur creates a better leaving group (Me-S-Hg adduct) upon nucleophilic attack of the starting diamine versus the S-Hg adduct formed using thiourea **49**. This alternative gave slightly better yields (30 – 88%) especially with bis-(4-aminophenyl)methanone (X= CO) which has an electron-withdrawing carbonyl linker leading to deactivation of the terminal amines. The poor yield of compound **38d** was a result of the predominant formation of the bis-guanidylated product due to the

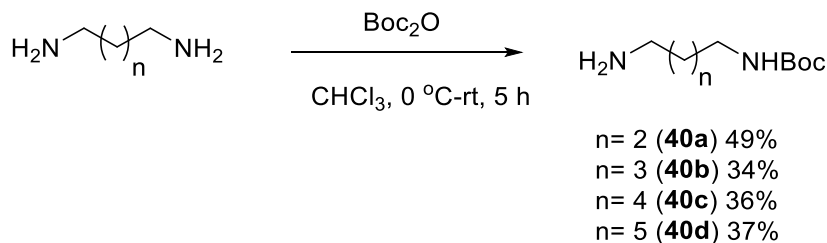
use of a more reactive thiourea and the presence of the electron-donating NH linker activating both terminal amines. Therefore, the use of the less reactive guanidylating agent **49** was chosen to obtain compound **38d** in a higher yield. These findings are similar to previous results obtained by our group.¹⁷

Scheme 3.2.2.2. Mechanism of HgCl₂ mediated guanidylation



3.2.3. Synthesis of mono-Boc-protected diamines

The next step in the synthesis of the aimed compounds was the mono-Boc protection of the desired aliphatic diamines (Scheme 3.2.3.1). The Boc protecting group was used based on its stability under basic conditions which are used in later steps of the synthesis and will allow for the deprotection of the entire compound at the final step of the synthesis. Diamines with four, five, six and seven methylene groups were used to produce the respective mono-Boc protected starting materials for the next guanidylation step. These were chosen based on previous results which showed that alkyl chain lengths less than four carbons and more than seven carbons were either detrimental or not contributory to DNA minor groove binding, as well as in our computational studies (see Chapter 3.1).¹⁵

Scheme 3.2.3.1. Mono Boc protection of alkyl diamines

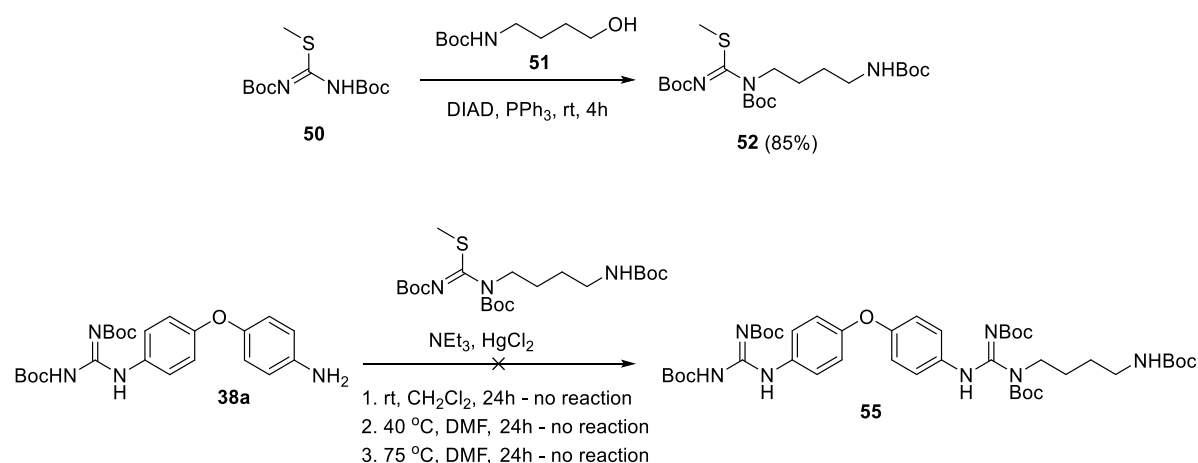
In order to obtain the mono-Boc protected diamine, a solution of di-*tert*-butyl dicarbonate (0.5 equiv) in CHCl_3 was added dropwise over 30 minutes to a solution of the corresponding diamine in CHCl_3 at 0 °C. The reaction was then left to stir for 5 hours reaching room temperature and it was monitored by TLC using ninhydrin stain to identify the formation of the desired product (dark purple stain). The work-up of the reaction involved removal of the excess diamine by washing with H_2O and brine. The resulting organic layer was then dried over anhydrous MgSO_4 and purified by column chromatography on base treated silica gel to afford the mono-Boc protected amines **40a-d** as white amorphous solids. Relatively poor yields were obtained for this process due to the formation of the di-Boc protected species *in situ* as well as product loss during column chromatography despite base treatment.

3.2.4. Synthesis of *N*-substituted guanidines

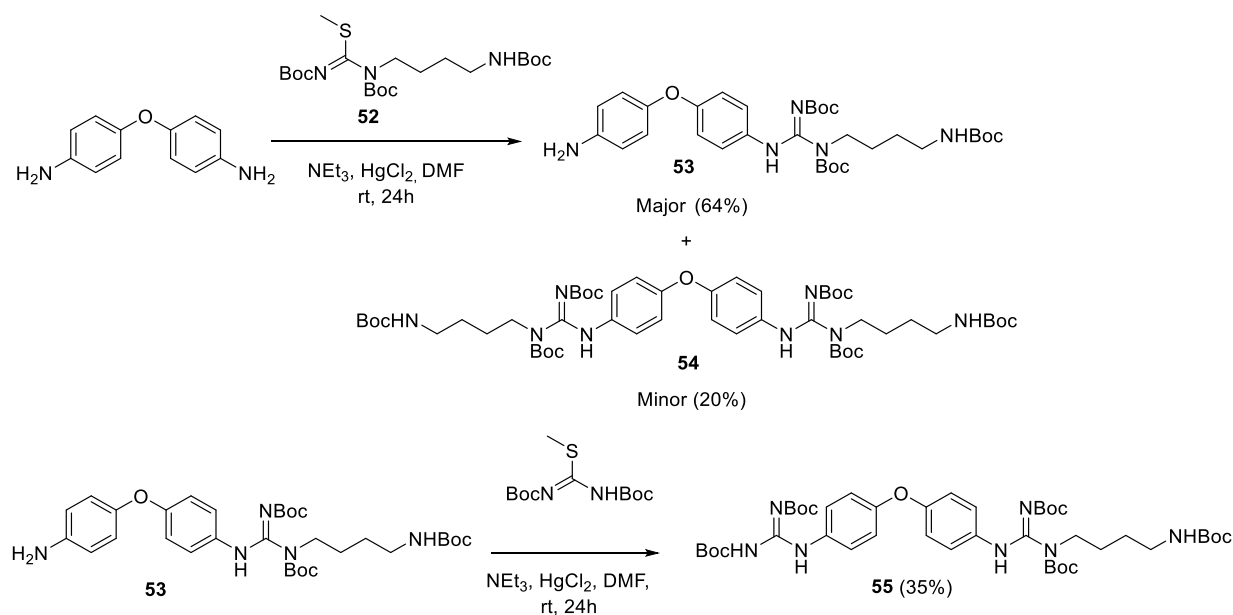
Three different approaches were explored to obtain the desired substituted guanidine intermediate from the prepared mono protected aliphatic diamines **40a-d** and Boc-protected mono-guanidylated di-aromatic systems **38a-d** through a thiourea intermediate. Similar to the approach employed by Suhs *et al.*, the Mitsunobu approach was adopted involving the reaction of the commercially available 1,3-bis-(*tert*-butoxycarbonyl)-2-methyl-2-thiopseudourea (**50**) with *N*-Boc protected aminoalcohol **51** in the presence of triphenylphosphine (PPh_3) and DIAD (diisopropylazodicarboxylate) to produce the desired thiourea (**52**) in high yield (Scheme 3.2.4.1).¹⁸ Subsequent mercury mediated guanidylation using compound **38a**, was attempted utilising conventional conditions. However, after 24

hours no evidence of product formation was observed, with both starting materials still present. The reaction was repeated in DMF at 40 °C for 24 hours followed by 75 °C for a further 24 hours after no product formation was evident.

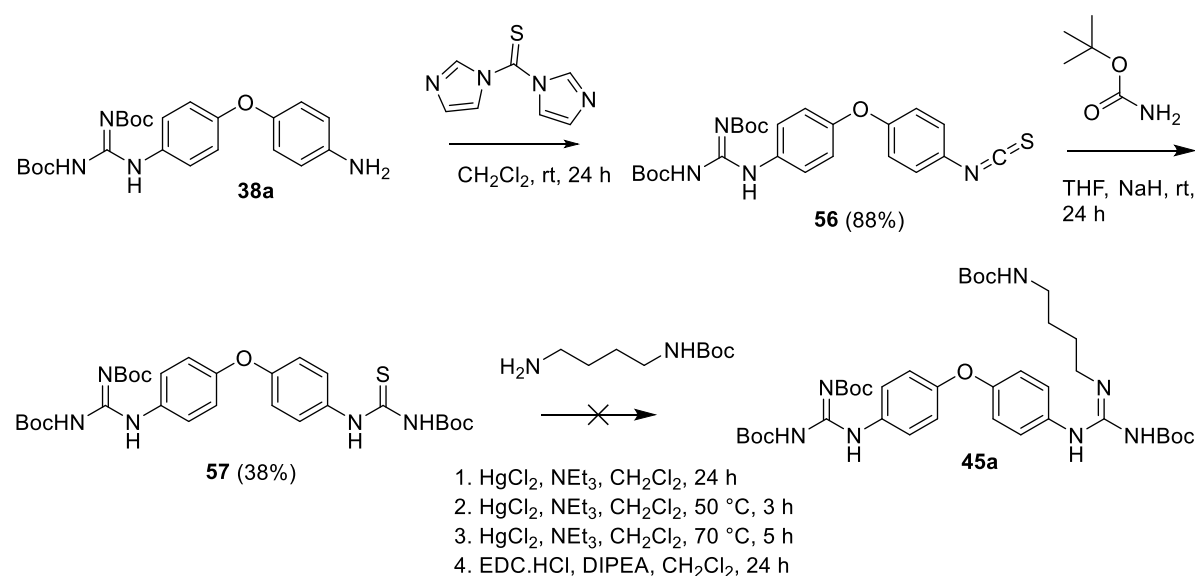
Scheme 3.2.4.1. Mitsunobu reaction and attempted guanidylation



In order to explore the lack of reactivity, both the nature of the electrophile **52** and nucleophile **38a** was investigated. Given the steric bulk surrounding the reactive carbon centre of the thiourea **52**, the need for a strong and less hindered nucleophile would be required for optimal reaction. It is important to note that the steric and electronic properties of 4,4'-oxydianiline are slightly changed upon guanidylation of the aniline amine to form compound **38a** as described in Scheme 3.2.2.1. Given this phenomenon, guanidylation using thiourea **52** was attempted using the raw starting material 4,4'-oxydianiline. This resulted in the production of compound **53** and compound **54** where the latter was produced in higher yield. Subsequent deprotection of compound **54** was conducted to explore its potential DNA binding properties and anti-parasitic activity in future works. Using the desired major product mercury mediated guanidylation with *S*-methyl thiourea **50** was conducted producing compound **55** in 35% yield.

Scheme 3.2.4.2. Guanidylation with the Mitsunobu product

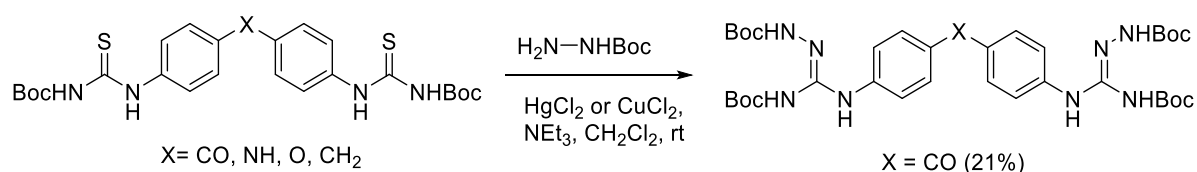
Given the low yield obtained for the production of *N,N',N''*-trisubstituted guainidine intermediate **55** an alternative approach was attempted involving the installation of the required thiourea functionality initially to the diaromatic system prior to guanidylation (Scheme 3.2.4.3). This entailed the reaction of compound **38a** with 1,1'-thiocarbonyl bis-imidazole at room temperature in CH_2Cl_2 to obtain the corresponding isothiocyanate **56** in good yield (88%). Subsequent reaction of **56** with *tert*-butyl carbamate in the presence of NaH in dry THF obtained the desired thiourea intermediate **57** as a pale yellow solid in a poor yield (38%). This may be attributed to the poor nucleophilicity of the carbamate caused by the surrounding bulky *tert*-butyl group and the delocalisation of the amine lone pair.

Scheme 3.2.4.3. Third attempted synthetic route towards substituted guanidine

In attempt to guanidylate the thiourea functionality of compound **57** under conventional conditions (use of HgCl_2 , NEt_3 and mono-Boc protected diaminopentane), it was observed that after 24 hours, both starting materials remained with no evidence of product formation as identified by TLC. In an attempt to force the reaction, the mixture was heated to 50°C for 3 hours, followed by 70°C for 5 hours; however, no evidence of product formation was evident. After recovering and analysing both starting materials another attempt was made to conduct the guanidylation using 1-ethyl-3-(3'-dimethylaminopropyl)carbodiimide HCl (EDC.HCl). This represents a greener approach for the synthesis of substituted guanidines which does not require the need for highly toxic mercury salts. Using similar conditions to those described by Flemer *et al.*, a solution of the mono-Boc protected diaminopentane and DIPEA in CH_2Cl_2 was added to a cooled solution (0°C) of thiourea **57** and allowed to stir at room temperature overnight.¹⁹ However, upon after 24 hours both starting materials still remained and no new products were identified (Scheme 3.2.4.3). This phenomenon may be a result of stereoelectronic effects around the thiocarbonyl where the bulky *tert*-butyl group hinders the nucleophilic attack of the aliphatic amine. In addition, the possible intermolecular HB between the terminal amino group of compounds **40a-d** may reduce the nucleophilicity towards the thiourea in question. This is similar to results obtained in previous studies which

showed poor reactivity of di-aryl bis-thioureas (Scheme 3.2.4.4).²⁰ Upon attempted *N*-aminoguanidylation, no reaction occurred except for the di-aryl bis-thiourea with a strong electron-withdrawing CO linker which produced the substituted guanidine in poor yield. This gives further evidence of the negative effect that the diaromatic system has on the reactivity of the thiourea.

Scheme 3.2.4.4. Previously reported guanidylation using Boc protected hydrazine



Taking into account the poor reactivity of the Boc-protected thiourea **57** other derivatives were considered that may avoid these stereoelectronic effects. Prior to carrying out the synthesis of other derivatives, DFT calculations were conducted to analyse the charge distribution around the thiocarbonyl as well as the optimised structure of the molecule. Using DFT methods (B3LYP (6-311+G (d, p), PCM solvent = CH₂Cl₂) two thioureas were considered, **57** and **39a** (Figure 3.2.4.1). One obvious feature of both structures is the intramolecular HB formed between the carbonyl of the alkyl-carbamate in the thiourea moiety and the aniline NH resulting in a planar six-membered ring arrangement. In the case of **57**, this causes two of the methyl groups of the *tert*-butyl system to reside below and above this plane, which may hinder the nucleophilic attack of the aliphatic amine. Comparing this to structure **39a**, bearing an ethyl carbamate, the steric hindrance is much less pronounced. In addition, the carbon of the thiocarbonyl group is slightly more positively charged due to the absence of the large electron cloud caused by the *tert*-butyl group.

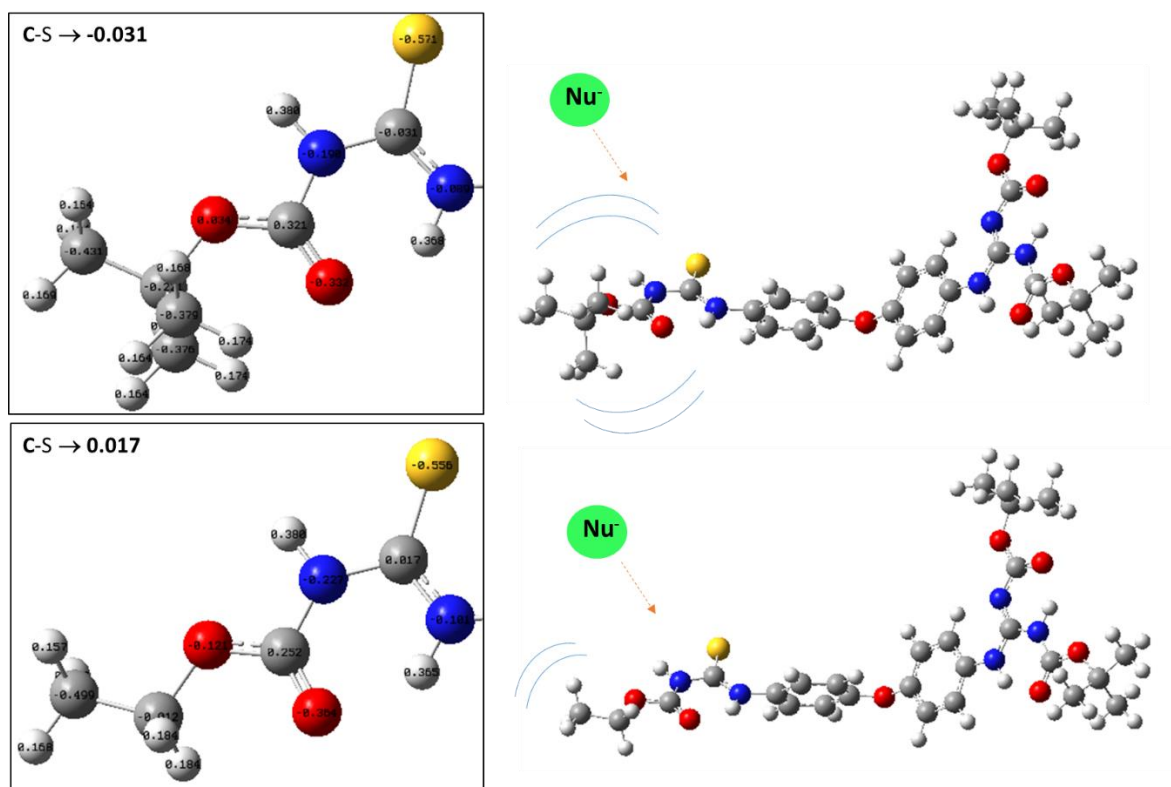
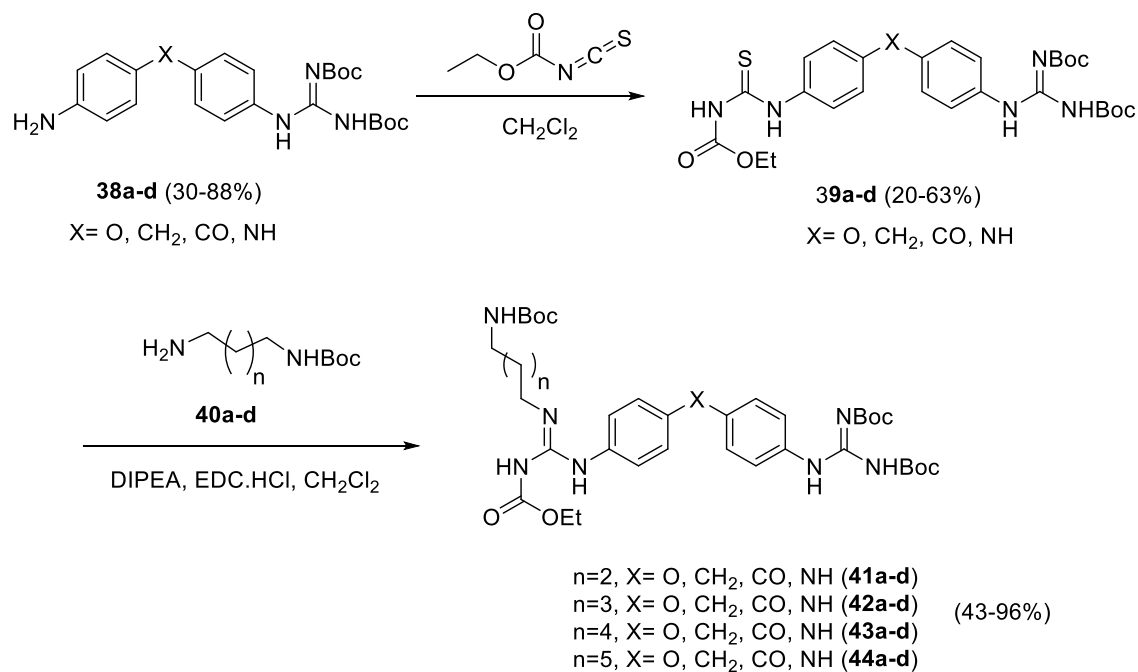
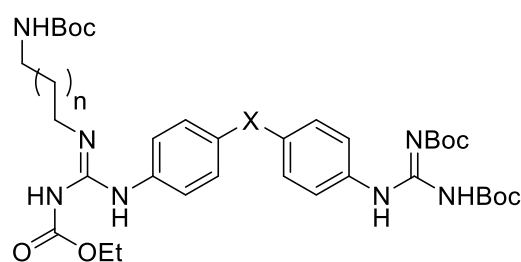


Figure 3.2.4.1. Optimised structures of compounds **57** (top) and **39a** (bottom) by DFT (B3LYP (6-311+G (d, p), PCM solvent= CH₂Cl₂)) showing the charge distribution.

Using a similar methodology employed by Manimala *et al.* to synthesise substituted guanidines, *N*-ethylcarbamate thiourea intermediates **39a-d** were prepared using the commercially available ethoxy-carbonyl isothiocyanate (ECIT) followed by guanidylation with the corresponding Boc-protected mono-guanidine diamines.²¹ Using this “greener” guanidylation approach as described earlier, compounds **39a-d** were stirred overnight in the presence of EDC.HCl and DIPEA in CH₂Cl₂ to obtain the substituted guanidines **41a-d** to **44a-d** as pale yellow amorphous solids (Scheme 3.2.4.5).

Scheme 3.2.4.5. Chosen synthetic route towards the *N*-substituted guanidines

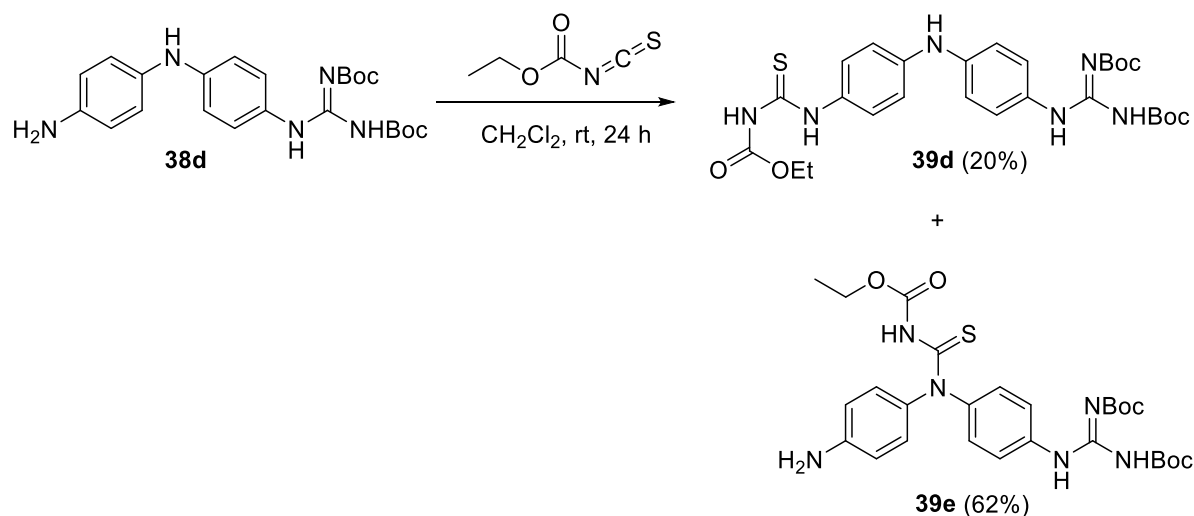
The yields for the various derivatives are summarised in Table 3.2.4.1. It is evident that with increasing carbon chain length of the aliphatic diamine there is a slight decrease in yield. This is attributed to an increase in steric bulk which affects the nucleophilicity of the protected diamine.

Table 3.2.4.1. Yields of *N,N',N''*-trisubstituted guanidines

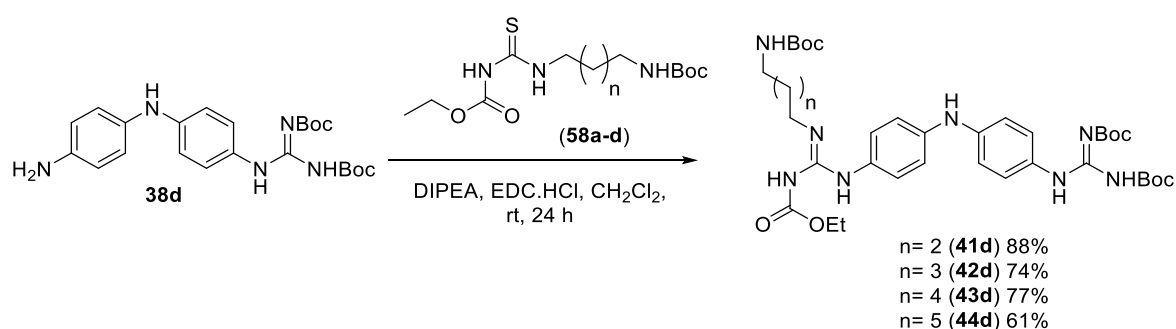
n=2, X= O, CH₂, CO, NH (**41a-d**)
 n=3, X= O, CH₂, CO, NH (**42a-d**) (43-96%)
 n=4, X= O, CH₂, CO, NH (**43a-d**)
 n=5, X= O, CH₂, CO, NH (**44a-d**)

Entry	Cmp	Linker (X)	Chain length (n)	Yield (%)	Entry	Cmp	Linker (X)	Chain length (n)	Yield (%)
1	41a	O	2	79	9	41b	CH ₂	2	96
2	42a	O	3	77	10	42b	CH ₂	3	62
3	43a	O	4	45	11	43b	CH ₂	4	65
4	44a	O	5	56	12	44b	CH ₂	5	45
5	41c	CO	2	65	13	41d	NH	2	75
6	42c	CO	3	63	14	42d	NH	3	55
7	43c	CO	4	50	15	43d	NH	4	43
8	44c	CO	5	55	16	44d	NH	5	61

Interestingly, reaction of compound **38d** with ECIT produced the unexpected product **39e** in larger amounts than to the desired compound **39d**. ¹H and ¹³C NMR analysis of this product showed that the secondary amine of compound **38d** linking the two phenyl rings had reacted with ECIT forming **39e** (Scheme 3.2.4.6). This may have been caused by the presence of the terminal ammonium sulfate salt that may not have been completely free based during the synthesis of compound **38d** from the corresponding commercially available di-aniline sulfate. Protonation of this terminal amine would thereby affect its nucleophilicity making the central amine more favourable to react.

Scheme 3.2.4.6. ECIT mediated thiourea formation

Therefore, an alternative route for the synthesis of compounds **41d**, **42d**, **43d** and **44d** was explored involving the reaction of ECIT with the mono-Boc protected diamines **40a-d** to produce the corresponding thioureas **58a-d** (Scheme 3.2.4.7). These thioureas were then used for the mercury mediated guanidylation of compound **38d**. Furthermore, this route produced the desired compounds (**41d**, **42d**, **43d** and **44d**) in higher yields.

Scheme 3.2.4.7. Alternative route towards the NH linked *N*-substituted guanidine intermediates

Evidence of the formation of protected *N,N',N''*-trisubstituted intermediates was easily identified by ^1H and ^{13}C NMR along with 2D-NMR experiments. Figure 3.2.4.2 shows the ^1H -NMR of compound **41a** where the aliphatic protons of the system are present within the expected region of 1.31 ppm to 4.14 ppm. The most upfield peak at 1.31 ppm integrating for three protons represents the methyl group of the ethyl carbamate moiety which couples ($J =$

7.1 Hz) to the 4.14 ppm quartet representing the two adjacent methylene protons. Between these shifts bears the outer methylene protons directly adjacent to the terminal Boc protected amine and guanidine nitrogen at 3.36 and 3.09 ppm. The large singlets at 1.52, 1.48 and 1.40 ppm which all integrate for nine protons each represents the methyl groups of the three Boc groups. Overlapping these signals are the four central methylene protons of the aliphatic chain which present as a multiplet from 1.47 – 1.43 ppm. Additionally, four NH proton signals are observed; three that correspond to the guanidine NH protons at 11.63, 10.53 and 10.31 ppm and one at 4.64 ppm that correlates with the proton of the terminal protected amine.

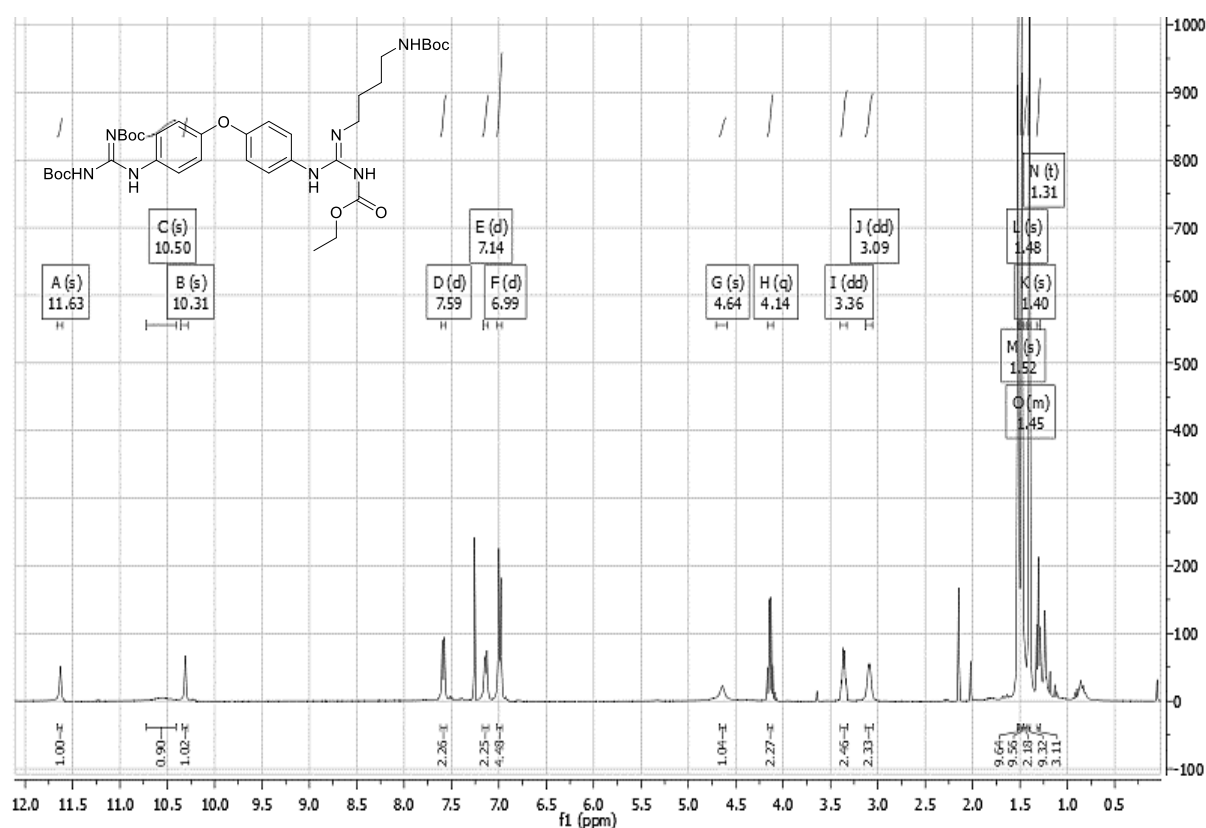


Figure 3.2.4.2. ^1H NMR of compound **41a**.

The ^{13}C NMR of compound **41a** shows the eight carbons associated with the di-aromatic system between 119 and 153 ppm (Figure 3.2.4.3). The presence of both substituted guanidine moieties is confirmed the two quaternary CN carbon signals at 164.5 and 163.4

ppm. Also, the three carbonyl signals of the Boc groups are present at 153.3, 153.5 and 156.0 which are more upfield than the ethyl carbamate carbonyl at 159.1 ppm.

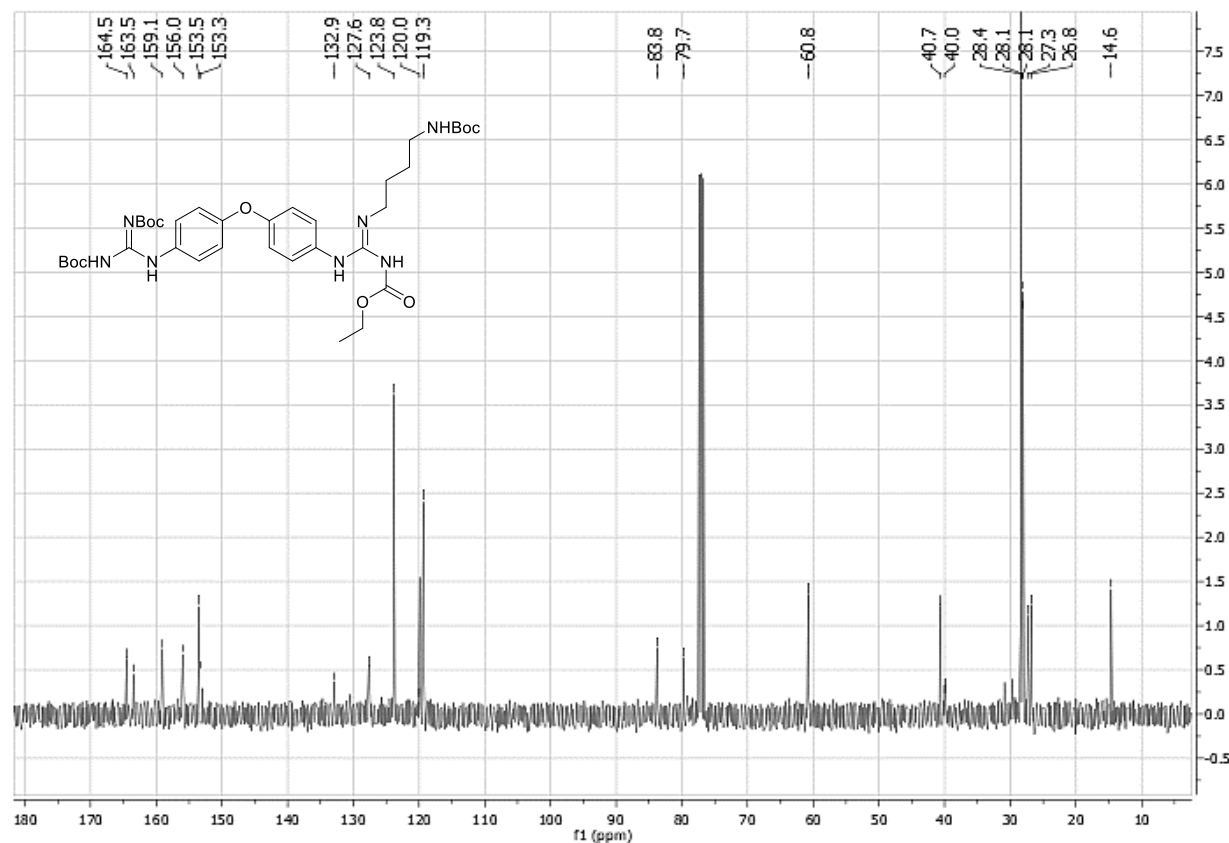


Figure 3.2.4.3. ^{13}C NMR of compound **41a**.

Given the amorphous nature of these compounds 2D NMR techniques aided in the characterisation of broad overlapping signals. ^1H - ^{13}C HSQC of compound **41a** allowed the identification of the aliphatic region of the molecule which is difficult to identify using one dimensional NMR techniques (Figure 3.2.4.4). The broad multiplet at 1.43 – 1.47 ppm representing the central methylene protons of the aliphatic chain can be discerned by correlation to two separate carbon peaks at 27.3 and 26.8 ppm. In addition, the aromatic splitting system can be identified through the correlation of the aromatic protons (7.59 – 6.92 ppm) to their respective carbons. The peaks at 119.3 and 119.9 correlate to H-7 and H-2 and the peaks at 127.7 and 123.8 represent H-6 and H-3, respectively. This indicates that compound **41a** presents an AA'BB' spin system.

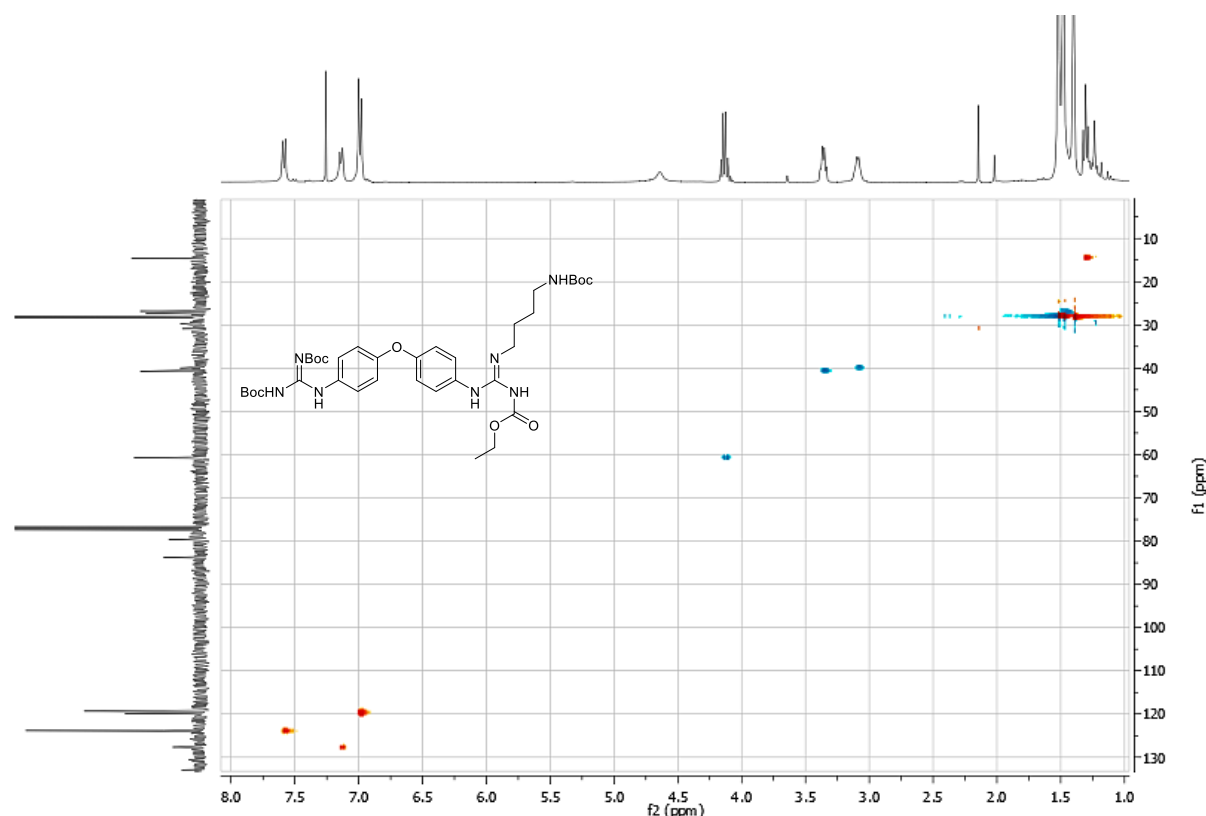


Figure 3.2.4.4. ^1H - ^{13}C HSQC of compound **41a**.

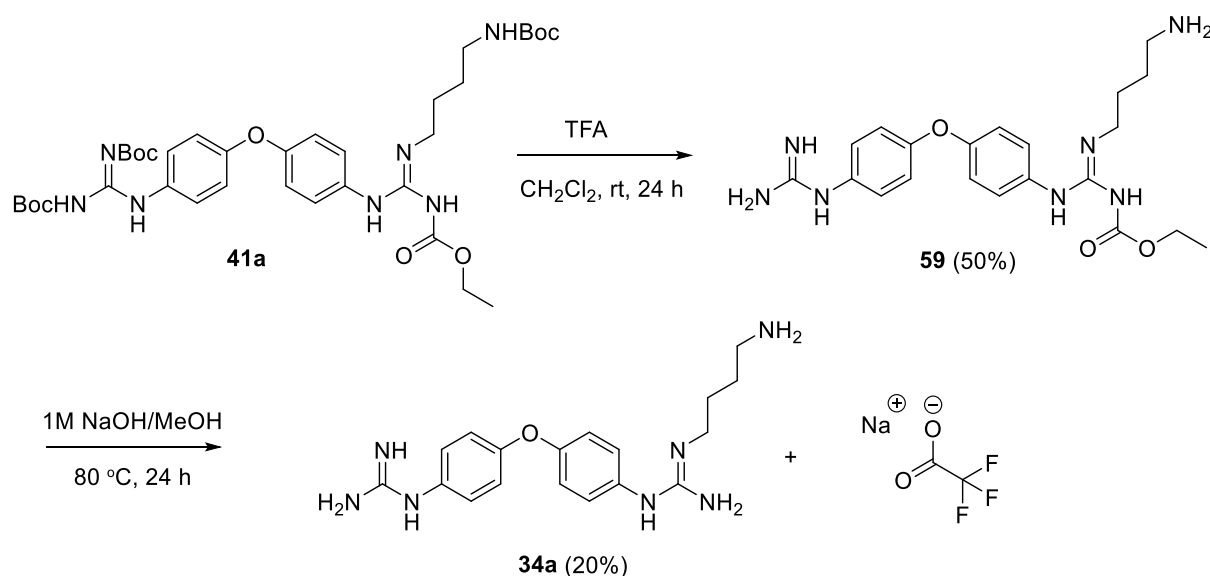
3.2.5. Trans-carbamylation of substituted guanidines

In the aforementioned study by Manimala *et al.*, various methods were attempted to selectively remove an ethyl carbamate protecting group following guanidylation with the use of ECIT. However, most of these methods failed or resulted in poor yields.²¹ As this protecting group is more base than acid sensitive, the complete deprotection of the *N*-substituted guanidine intermediates would require two steps.

Initial attempts were made using a two-step deprotection methodology which involves Boc deprotection with the use of TFA followed by ethyl carbamate removal through reflux in 1M NaOH/methanol (Scheme 3.2.5.1). The resulting compound **34a** as seen below was produced in relatively low yields which may be a result of the harsh basic conditions needed to remove the ethyl carbamate protecting group. After reverse phase column chromatography which

was needed to remove the resulting impurities, the presence of sodium trifluoroacetate salt precipitates was evident which was problematic given its similar solubility to the final product.

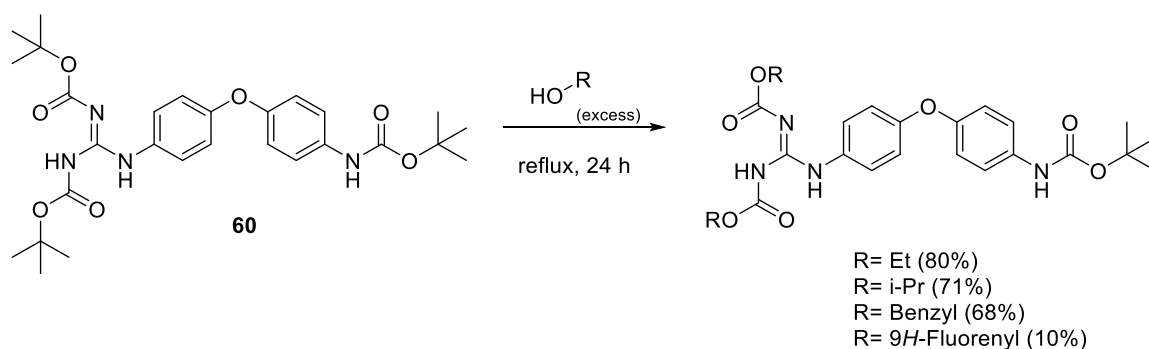
Scheme 3.2.5.1. Two step deprotection of *N*-substituted intermediates



In order to achieve a one-step global deprotection of all protecting groups, a transesterification step was employed to convert the ethyl carbamate present on the *N*-substituted guanidine to a *tert*-butyl carbamate group. This approach was employed based on a study conducted by Prabhakaran *et al.* who looked at the reaction of certain amines with *N,N',N''*-Tri-Boc guanidine under an elevated temperature to afford the corresponding amidinoureas.²² As they prove the formation of an isocyanate intermediate at high temperatures this gives rise to the possibility of conducting transesterification reactions of carbamate groups. Interestingly, this phenomenon is not solely restricted to *tert*-butyl carbamates but seems to be an interesting property of carbamate protected guanidines. In a recently published study, we explored the scope and selectivity of this reaction.²³ As shown in Scheme 3.2.5.2, it was observed that upon the reaction of compound **60** in a given alcohol, only the Boc groups of the guanidine functionality are converted to the corresponding alternative carbamate with retention of the Boc protected aniline amine. This is attributed to the greater lability of the NH of the protected guanidine due to the delocalisation of the double bond within its system versus the protected aniline NH. The use of other carbamate

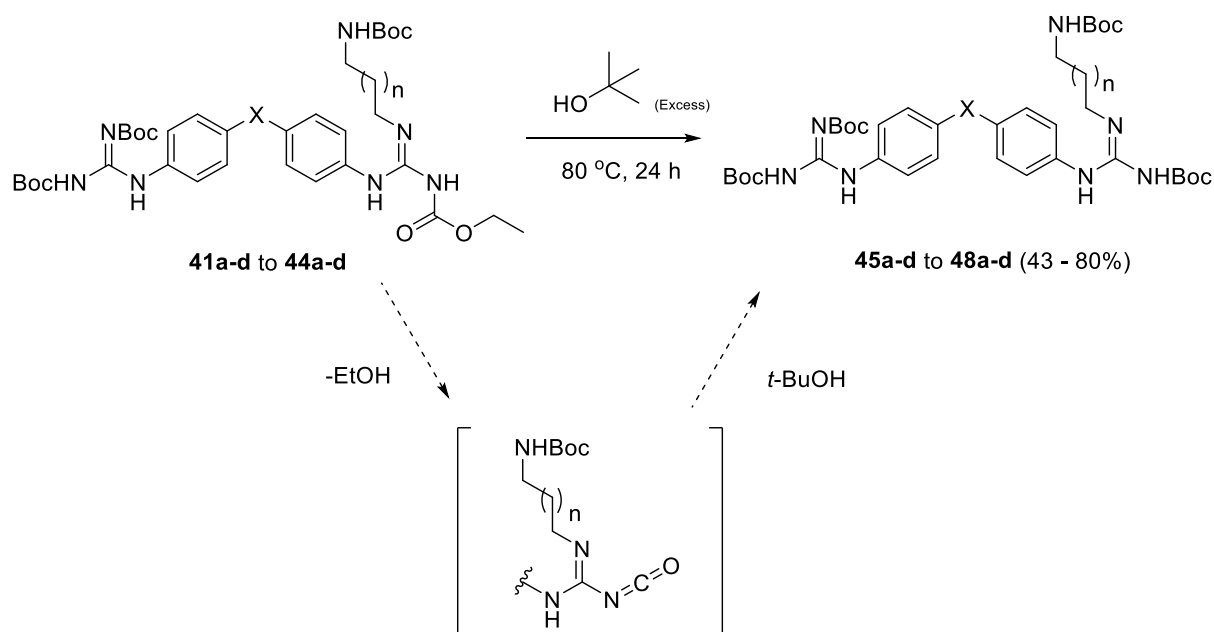
protected starting materials such as ethyl carbamate and benzyl carbamate (Cbz) protected guanidine derivatives of **60** was also explored. Interestingly, conversion of the carbamate moieties with a range of alcohols was observed in moderate to high yields. Therefore, this reaction offers a facile and selective method to exchange carbamate protecting groups attached to guanidines.

Scheme 3.2.5.2. Reported selective carbamate conversion of protected guanidines

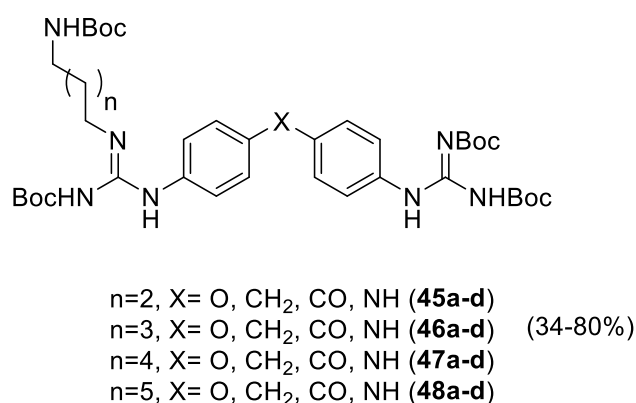


Applying this to the protected intermediates (**41a-d** to **44a-d**), *t*-butanol was used as the nucleophile in the transesterification reaction resulting in the production of the desired Boc protected intermediates (**45a-d** to **48a-d**). An excess of *t*-butanol was used given the presence of three other Boc groups within compounds **41a-d** to **44a-d** which will also undergo transesterification. However, this does not pose an issue as it inherently results in the production of the same carbamate.

Scheme 3.2.5.3. Carbamate conversion reaction for the production of the Boc protected *N*-substituted guanidines.



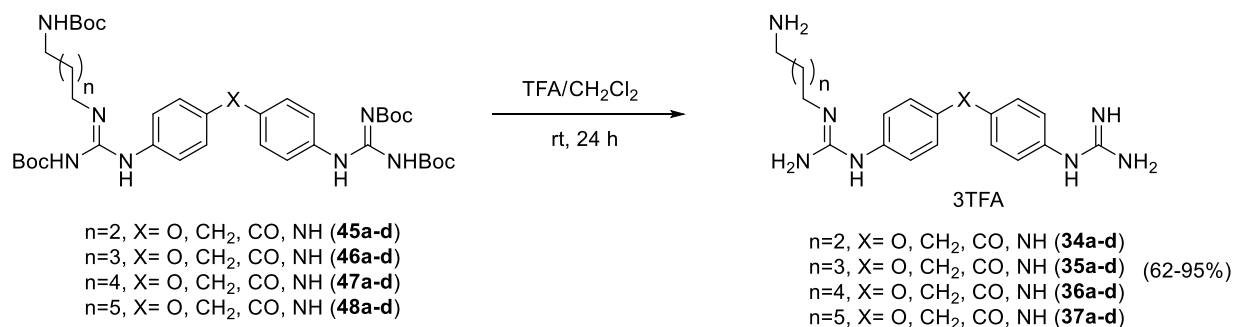
As shown in Scheme 3.2.5.3, compounds **41a-d** to **44a-d** were heated under reflux at 80 °C for 24 hours in an excess of *t*-butanol and following purification by column chromatography yielded compounds **45a-d** to **48a-d** as amorphous gums. Table 3.2.5.1 shows the yields of the carbamate conversions for each derivative. From the yields obtained, it is evident that compounds bearing larger alkyl chains result in lower yields than those bearing shorter chains. This could be attributed to the large steric bulk surrounding the carbamate group hindering the nucleophilic attack of *t*-butanol following the *in situ* formation of the isocyanate intermediate.

Table 3.2.5.1. Yields of carbamate conversion reactions

Entry	Cmp	Linker (X)	Chain length (n)	Yield (%)	Entry	Cmp	Linker (X)	Chain length (n)	Yield (%)
1	45a	O	2	77	9	45b	CH ₂	2	77
2	46a	O	3	38	10	46b	CH ₂	3	62
3	47a	O	4	49	11	47b	CH ₂	4	66
4	48a	O	5	34	12	48b	CH ₂	5	57
5	45c	CO	2	80	13	45d	NH	2	59
6	46c	CO	3	61	14	46d	NH	3	48
7	47c	CO	4	51	15	47d	NH	4	60
8	48c	CO	5	49	16	48d	NH	5	44

3.2.6. Deprotection of Boc protected substituted guanidine intermediates

tert-Butoxy carbonyl (Boc) protecting group is one of the most popular amino protecting groups due to its ease of attachment and removal as well as stability under basic conditions. The most common methods used for the removal of Boc groups are under acidic conditions with the use of formic acid, hydrochloric acid, trifluoroacetic acid, or even some Lewis acids.²⁴ The standard conditions that have proved to produce the best results and have been implemented by our group is treatment with 1:1 TFA/CH₂Cl₂.

Table 3.2.6.1. Yields obtained for the trifluoroacetate salts of the aminoalkyl derivatives of di-aromatic bis-guanidines

Entry	Cmp	Linker (X)	Chain length (n)	Yield (%)	Entry	Cmp	Linker (X)	Chain length (n)	Yield (%)
1	34a	O	2	82	9	34b	CH ₂	2	77
2	35a	O	3	81	10	35b	CH ₂	3	62
3	36a	O	4	95	11	36b	CH ₂	4	81
4	37a	O	5	77	12	37b	CH ₂	5	85
5	34c	CO	2	80	13	34d	NH	2	94
6	35c	CO	3	91	14	35d	NH	3	91
7	36c	CO	4	75	15	36d	NH	4	83
8	37c	CO	5	89	16	37d	NH	5	74

As shown in Table 3.2.6.1, the corresponding trifluoroacetate salts (**34a-d** to **37a-d**) were produced in high yields. All final compounds were fully characterised by IR, NMR, HRMS and HPLC which conclusively gave evidence of high purity. IR absorbances corresponding to the Boc carbonyls of compounds **45a-d** to **48a-d** were no longer present with the appearance of broad NH bands between 3200 and 3400 cm⁻¹. Further evidence by ¹H and ¹³C NMR in D₂O shows the disappearance of Boc proton signals at ~1.4-1.6 ppm and carbon signals at ~27-29 ppm. As for NMR spectrum of compound **37b**, shown in Figure 3.2.5.1, two signals at 3.34 ppm and 3.09 ppm represent the two terminal methylene protons attached to the guanidine and terminal ammonium cation respectively. The ten central methylene protons produce multiplets at 1.50 – 1.46 ppm and 1.34 – 1.22 ppm. Interestingly, the Boc-protected derivative of compounds **37b** (**48b**) has restricted rotation of the phenyl rings which results in an extremely short relaxation time for the aromatic signals. This gives rise to aromatic protons appearing as broad, amorphous signals such as the broad singlets at 7.19 ppm and 7.12 ppm.

Deprotection of the Boc-protected derivative reduces the steric congestion experienced by the aromatic rings causing a return to ordinary levels and sharpening of the aromatic signal as shown below.

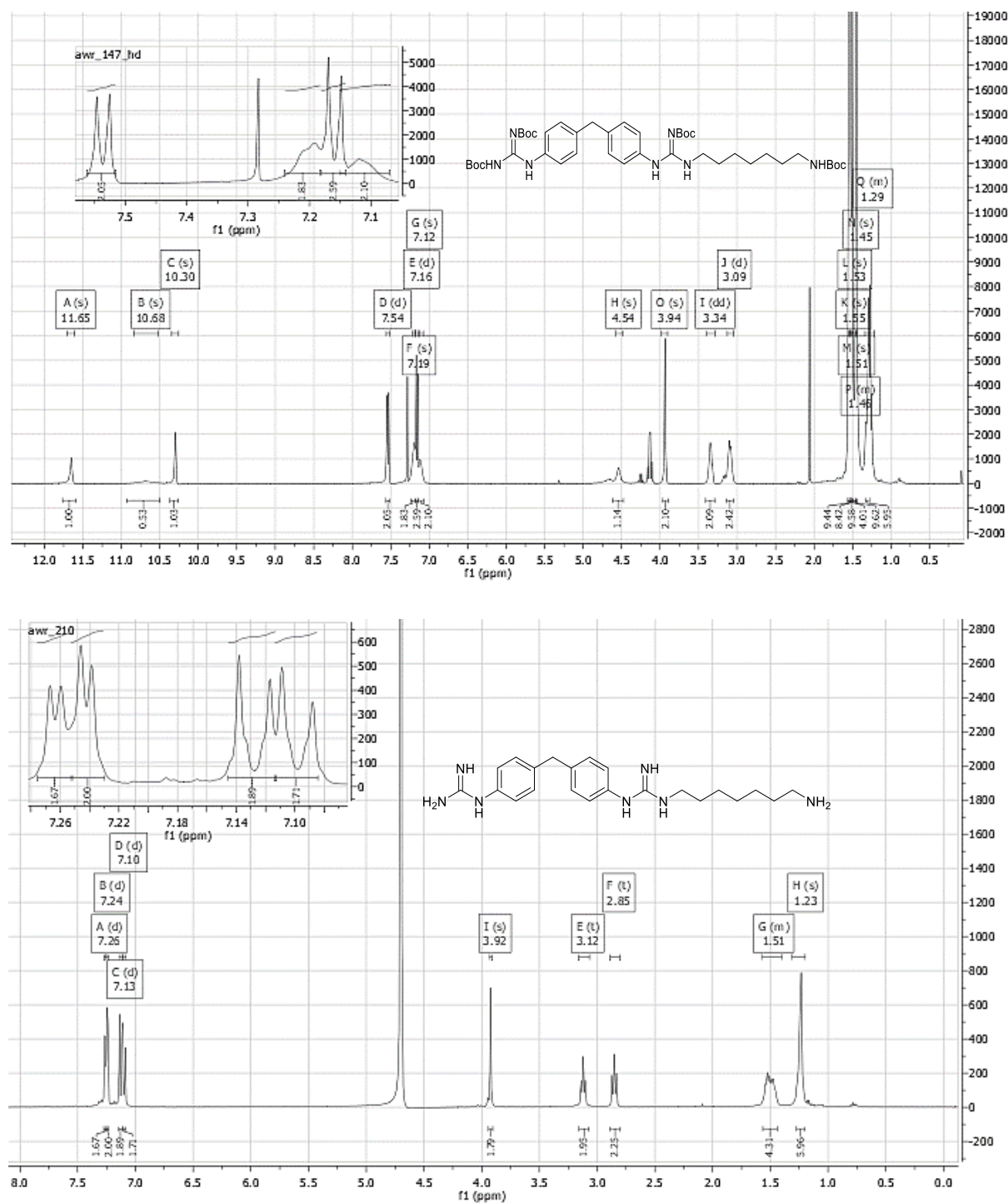


Figure 3.2.5.1. ^1H NMR of compound **48b** and **37b**.

3.3. Biophysical studies

Following the synthesis of Family F as described in the previous chapter, biophysical evaluation of these compounds was carried out to characterise their DNA binding properties. As stated in Chapter 1, the aim of this research was to design and synthesise a new series of compounds that can selectively bind to the DNA minor groove, thereby inhibiting the binding of DNA-dependent enzymes responsible for key cellular processes such as transcription and translation. Thus, confirmation of the mode of binding and determination of the DNA binding strength is essential in evaluating the potential of this new series of compounds as effective MGB anti-parasitic agents.

In order to perform this analysis, various techniques such as UV thermal denaturation, Circular Dichroism and UV-Vis titrations were carried out. These experiments were conducted with the use of salmon testes DNA (stDNA) where the best candidates were analysed for minor groove binding selectivity by testing them against a specific oligonucleotide sequence (poly (dA-dT)₂) used as a model of the DNA minor groove. Data obtained from these experiments help to identify promising candidates, which were further evaluated in biological studies as potential anti-protozoal agents.

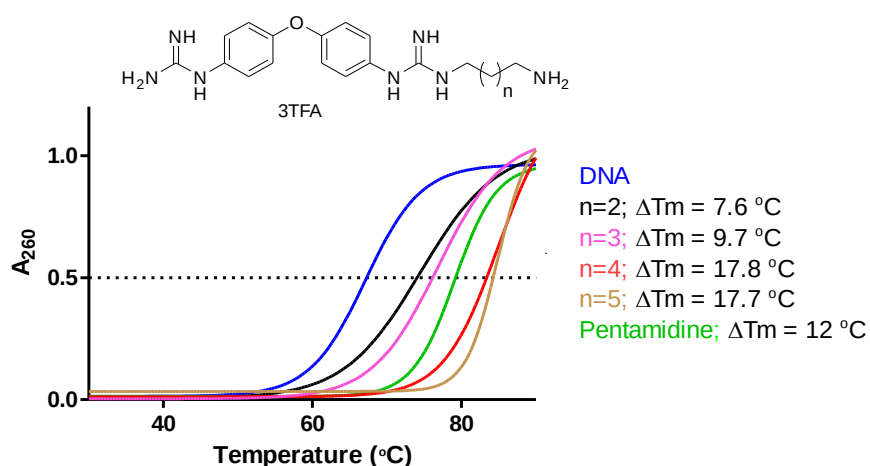
3.3.1 UV thermal denaturation of Family F

As discussed previously, the helical arrangement of DNA forces the nucleotide bases to be buried within the centre of the macromolecule while the sugar-phosphate moieties make up the external surface. Consequentially, nucleobases are shielded and can only absorb minimal quantities of light. Energy input in the form of heat can cause the HBs formed between the nucleotide bases pairs to break resulting in the splitting of the double-stranded helix into two single strands. This process is referred to as thermal denaturation (or melting). As the DNA duplex continues to “melt” with increasing temperature an increase in absorbance occurs upon irradiation of light (e.g. UV) at the wavelength maximum of DNA (260 nm). Plotting absorbance versus temperature a sigmoidal curve is obtained where the lower part of the graph indicates predominantly double-stranded DNA and the upper part indicates single-stranded DNA. Furthermore, the melting temperature (T_M) can easily be extrapolated by the

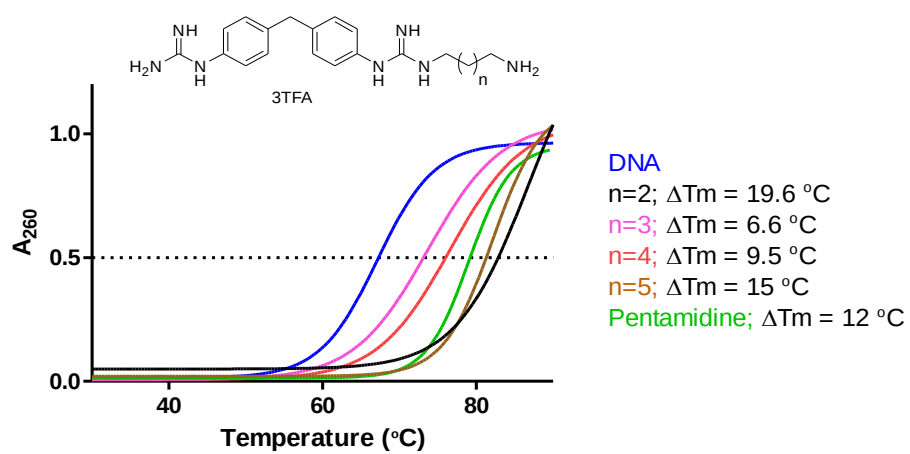
temperature at which the presence of double-stranded and single-stranded DNA is the same. The melting temperature of DNA is dependent on its sequence. The reported melting temperature of stDNA is 67 °C, while poly(dA-dT)₂ has a lower T_M of 47 °C.²⁵ The drop in temperature seen with the poly(dA-dT)₂ oligonucleotide is due to the lack of G:C bases pairs which form stronger HB interactions than the A:T base pairs. Upon binding of a ligand to DNA stabilisation of the DNA duplex may occur causing an increase in the T_M. This difference in melting temperature as referred to as the ΔT_M , is thereby used to rank various ligands in terms of DNA binding strength.

Thermal denaturation analysis of all compounds was first carried with stDNA. As stDNA is a wild-type unspecific oligonucleotide (32% G:C content), this initial screen serves as an indicator of overall DNA binding affinity. At a DNA concentration of 150 μ M (gives an absorbance unit (A.U.) of 1 at 260 nm) in 10 mM phosphate buffer (pH = 7), the T_M was measured to ensure the quality of the DNA sample. Following this, experiments were repeated by adding the specific compound to be tested (at 15 μ M) to a solution of DNA (150 μ M) to achieve a final phosphate: drug (P/D) ratio of 10 under the same conditions as stated above. DNA binding was observed for all the compounds within this family producing ΔT_M values between 7 and >20 °C. Values indicated as >20 °C are due to the absence of an upper plateau caused by extremely strong stabilisation of the DNA duplex at high temperatures. Figure 3.3.1.1 shows the thermal denaturation curves obtained for derivatives bearing the O (**34-37a**), CH₂ (**34-37b**), CO (**34-37c**) and NH (**34-37d**) linker.

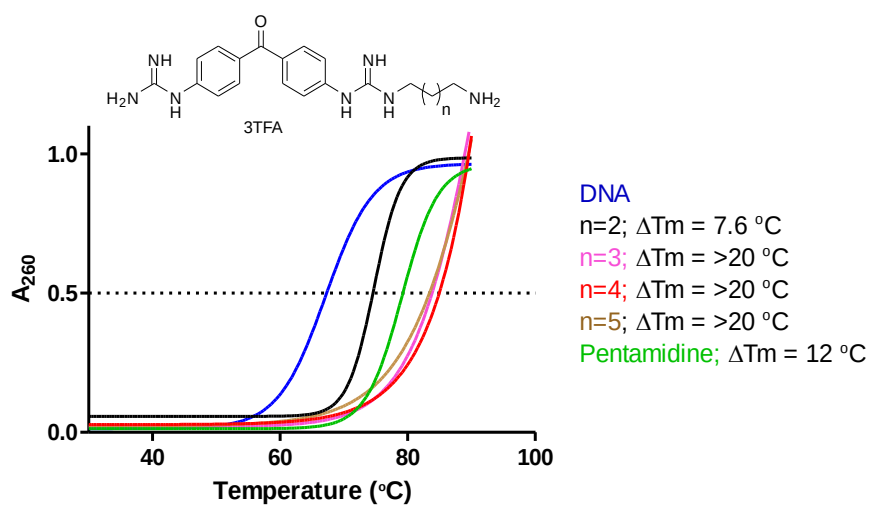
a)



b)



c)



d)

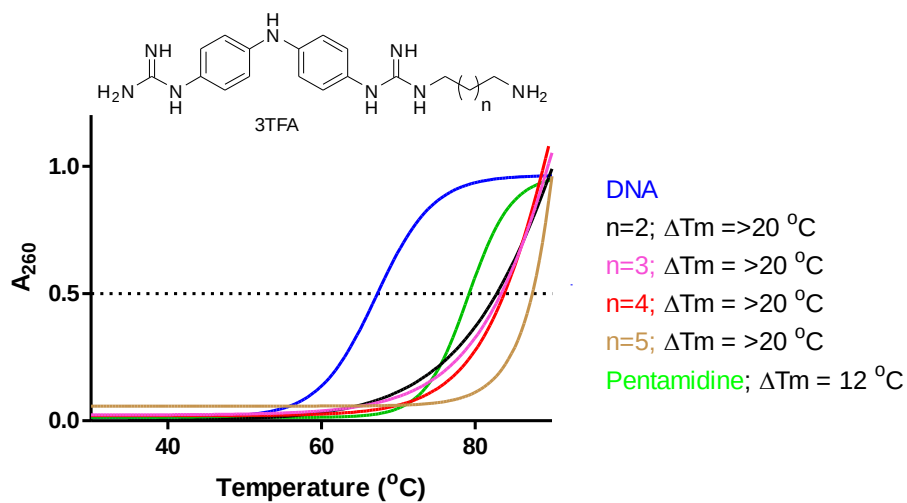


Figure 3.3.1.1. Plot showing the DNA UV thermal denaturation results of compounds **34-37a** (X=O), **34-37b** (X=CH₂), **34-37c** (X=CO) and **34-37d** (X=NH).

Comparing the extent of DNA binding to that of the known MGB pentamidine, several compounds within this new series elicit significantly stronger DNA binding affinity. Additionally, comparing the ΔT_m values obtained as shown in Table 3.3.1.1 (Entry 1-15) to those of previously reported compounds from Families A-E (Entry 17-22), it is clear that these compounds outweigh previous families in terms of DNA binding strength.

Table 3.3.1.1. Salmon testes DNA UV thermal denaturation results of novel Family F (**34a-d** to **37a-d**). Results for compounds **19**, **21**, **24**, **26** and **61** of previously synthesised families shown for comparison reasons.

3TFA

$n=2$, $X=O, CH_2, CO, NH$ (**34a-d**)
 $n=3$, $X=O, CH_2, CO, NH$ (**35a-d**)
 $n=4$, $X=O, CH_2, CO, NH$ (**36a-d**)
 $n=5$, $X=O, CH_2, CO, NH$ (**37a-d**)

$R^1, R^2 = \text{Gua}; X = NH$ (**21**)
 $R^1, R^2 = \text{Imi}; X = NH$ (**19**)
 $R^1 = \text{Gua}; R^2 = \text{Imi}; X = NH$ (**24**)
 $R^1, R^2 = \text{THP}; X = CO$ (**26**)
 $R^1 = \text{Gua}; R^2 = \text{Am}; X = NH$ (**61**)

THP =

Am =

Gua =

Imi =

Entry	Cmp	Linker (X)	Chain length (n)	$\Delta T_M (^{\circ}C)$
1	34a	O	2	7.6 ± 0.03
2	35a	O	3	9.7 ± 0.2
3	36a	O	4	17.8 ± 0.06
4	37a	O	5	17.7 ± 0.01
5	34b	CH ₂	2	19.6 ± 0.1
6	35b	CH ₂	3	6.6 ± 0.05
7	36b	CH ₂	4	9.5 ± 0.04
8	37b	CH ₂	5	15 ± 0.06
9	34c	CO	2	7.6 ± 0.01
10	35c	CO	3	>20
11	36c	CO	4	>20
12	37c	CO	5	>20

Entry	Cmp	Linker (X)	Chain length (n)	$\Delta T_M (^{\circ}C)$
13	34d	NH	2	>20
14	35d	NH	3	>20
15	36d	NH	4	>20
15	37d	NH	5	>20
17	21	NH ^a	-	8.0 ± 0.01
18	19	NH ^b	-	13.5 ± 0.03
19	24	NH ^c	-	12.0 ± 0.01
20	26	CO ^d	-	9.3 ± 0.3
21	61	NH ^e	5	5.0 ± 0.01
22	Pentamidine	-	-	12 ± 0.01

Melting temperature of stDNA in phosphate buffer (10 mM) is 67 °C. ^{a, b, c and d} taken from Ref. 12, ^e taken from Ref. 15.

Analysing the effect of the linker on DNA binding, a trend of NH>CO>O>CH₂ is observed. Comparing the ΔT_M values of CH₂-linked derivatives (Table 3.3.1.1, entry 5-8) to those of the O-linked derivatives (Table 3.3.1.1, entry 1-4), the latter produce a stronger stabilising effect

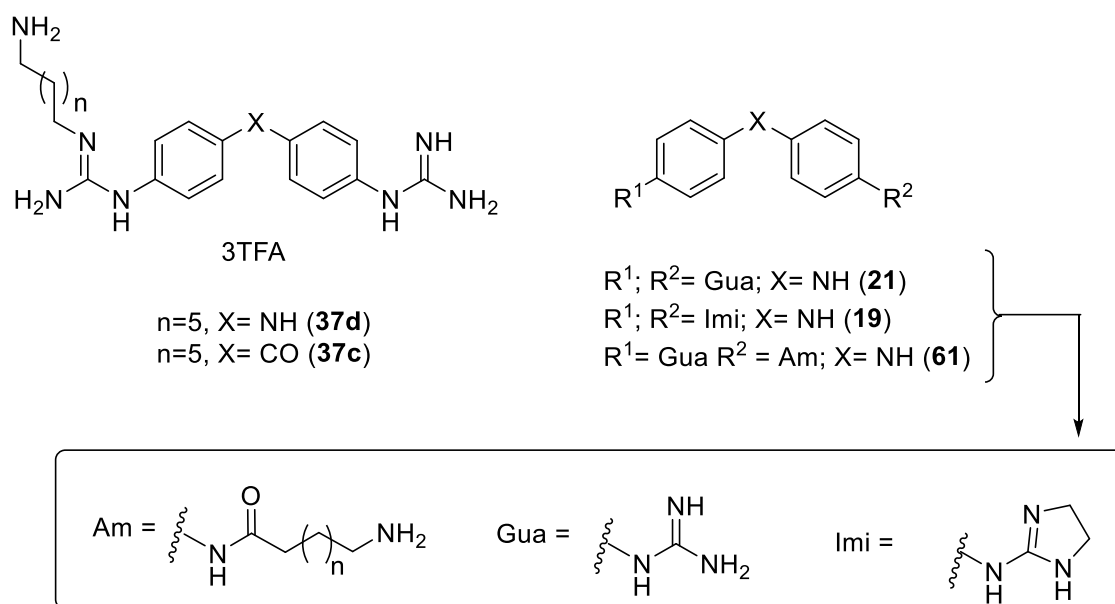
across the various alkyl chain lengths. However, the outlier, in this case, is compound **34b** which gave a ΔT_M of 19.6 °C compared to a ΔT_M of 7.6 °C produced by compound **34a**. This may be due to the smaller energy penalty paid by rotation of the molecule to adopt the co-planar mode of binding compared to compounds **35b**, **36b** and **37b** with larger alkyl chains attached. However, O-linked derivatives with a smaller dihedral angle between the phenyl rings do not have to rotate as far thus the limiting factor being alkyl chain length. Although compounds **34a** and **34c** give the same ΔT_M value (7.6 °C), an increase in carbon chain length in both cases result in drastically different values. This can be seen from the difference in ΔT_M between amino-pentyl derivatives **35a** ($\Delta T_M = 9.7$ °C) and **35c** ($\Delta T_M > 20$ °C). This comes as no surprise given the smaller dihedral angle between the phenyl rings of CO-linked derivatives versus O-linked derivatives. Additionally, the greater ΔT_M value of **34c** (7.6 °C) and **34d** (>20 °C) indicates that the NH linker is more favourable than the CO linker.

Secondly, a correlation between aminoalkyl chain length and ΔT_M is observed. As carbon chain length increases across the various derivatives, an increase in DNA binding occurs. Again, compound **34b** is an exception to this trend. Moreover, the clearest correlation is observed for the O-linked derivatives given the strong binding affinity of the NH and CO-linked derivatives producing ΔT_M values greater than 20 °C across the alkyl derivatives. Interestingly, comparing compound **61** of Family E ($\Delta T_M = 5$ °C) and **37d** ($\Delta T_M > 20$ °C), both bearing the same linker and alkyl chain length, **37d** displays significantly stronger DNA binding affinity. This indicates that condensation of the alkyl chain directly to the di-aromatic system through an amide bond is less favourable than attachment to the guanidine cation. These findings correlate with the conclusions drawn from the molecular modelling studies as shown in Chapter 3.1.

Once the experiments with stDNA were completed, the sequence selectivity of this new series of compounds was investigated by carrying out thermal denaturation experiments in poly(dA-dT)₂. As this oligonucleotide is a better representation of the DNA minor groove (due to its A:T base pair make-up), verification of minor groove binding over the major groove binding can be obtained. In order for a molecule to bind into the narrower minor groove, it should be relatively planar with a small dihedral angle between the two aromatic rings. For this reason, the most promising compounds **37c** and **37d** were chosen for these experiments. For consistency and comparison reasons, these experiments were carried out at a DNA

concentration of 5 μM and a drug concentration of 2.5 μM giving a P/D of 2. As the T_M of poly(dA-dT)₂ (47 °C) is lower than that of stDNA and given the strong DNA binding affinity expected for **37c** and **37d**, it was possible to measure very specific ΔT_M values due to the larger span in temperature after the initiation of denaturation. Table 3.3.1.2 shows the ΔT_M values of **37c**, **37d** (Entry 1-2) and the best binding candidates of previous families (Entry 3-5). From these results, it can be deduced that **37c** and **37d** display very strong minor groove binding capacity producing ΔT_M values of 30 °C which is greater than that of any previously made compounds within the Rozas group in-house library.

Table 3.3.1.2. Thermal denaturation experiment results of the Family F compounds **37c** and **37d** with poly(dA-dT)₂. Compounds **19**, **21**, and **61** from previous families were added for comparison reasons.



Entry	Cmp	Linker (X)	Chain length (n)	ΔT_M (°C)
1	37d	NH	5	30 $\pm$ 0.1
2	37c	CO	5	30 $\pm$ 0.2
3	21	NH ^a	-	17.5
4	19	NH	-	20
5	61	NH ^b	5	6.6 $\pm$ 0.01

Melting temperature of stDNA in phosphate buffer (10 mM) is 47 °C. ^a taken from Ref. 12., ^b taken from Ref. 15.

3.3.2. Circular Dichroism studies of Family F

In order to verify the mode of DNA binding of this new family of compounds, CD studies were conducted. As previously mentioned, MGBs characteristically exhibit a strong positive ICD signal due to their close proximity to chiral sugar molecules. The aminoalkyl derivatives under investigation are achiral with no inherent CD signals; however, when bound to DNA they may show an induced signal (ICD). Hence, all compounds within this family were analysed with the use of stDNA. At a fixed DNA concentration of 150 μM , aliquots of the compound were added over 10 additions achieving a final P/D ratio of 10.

Figure 3.3.2.1 shows the CD spectra of the NH-linked derivatives **34d** and **37d** with four and seven carbon chain lengths respectively. In both cases, the effects of DNA binding are clearly seen upon titration of the compound into the DNA solution. As seen from the spectrum generated for the unbound DNA (black line in Figure 3.3.2.1.), the characteristic positive signal at 275 nm and a negative signal at 240 nm are produced. Upon addition of the ligand's aliquotes, a positive ICD signal is observed at 315 nm due to the coupling of electronic transition moments between the chiral DNA and achiral ligands. The increase of this signal upon addition of the molecule is characteristic of compounds that bind to the minor groove instead of intercalating between the DNA base pairs. In conjunction with this, the ellipticity (mdeg) of the positive band at 275 nm decreased indicating the decrease in the concentration of unbound DNA. These changes are consistent with the alteration of the helical structure of B-DNA upon groove binding. Comparing both compounds studied, the increase in carbon chain length correlates with an increase in the magnitude of the effects discussed above. This gives further evidence that alkyl chains of seven methylene groups are optimal for DNA minor groove binding.

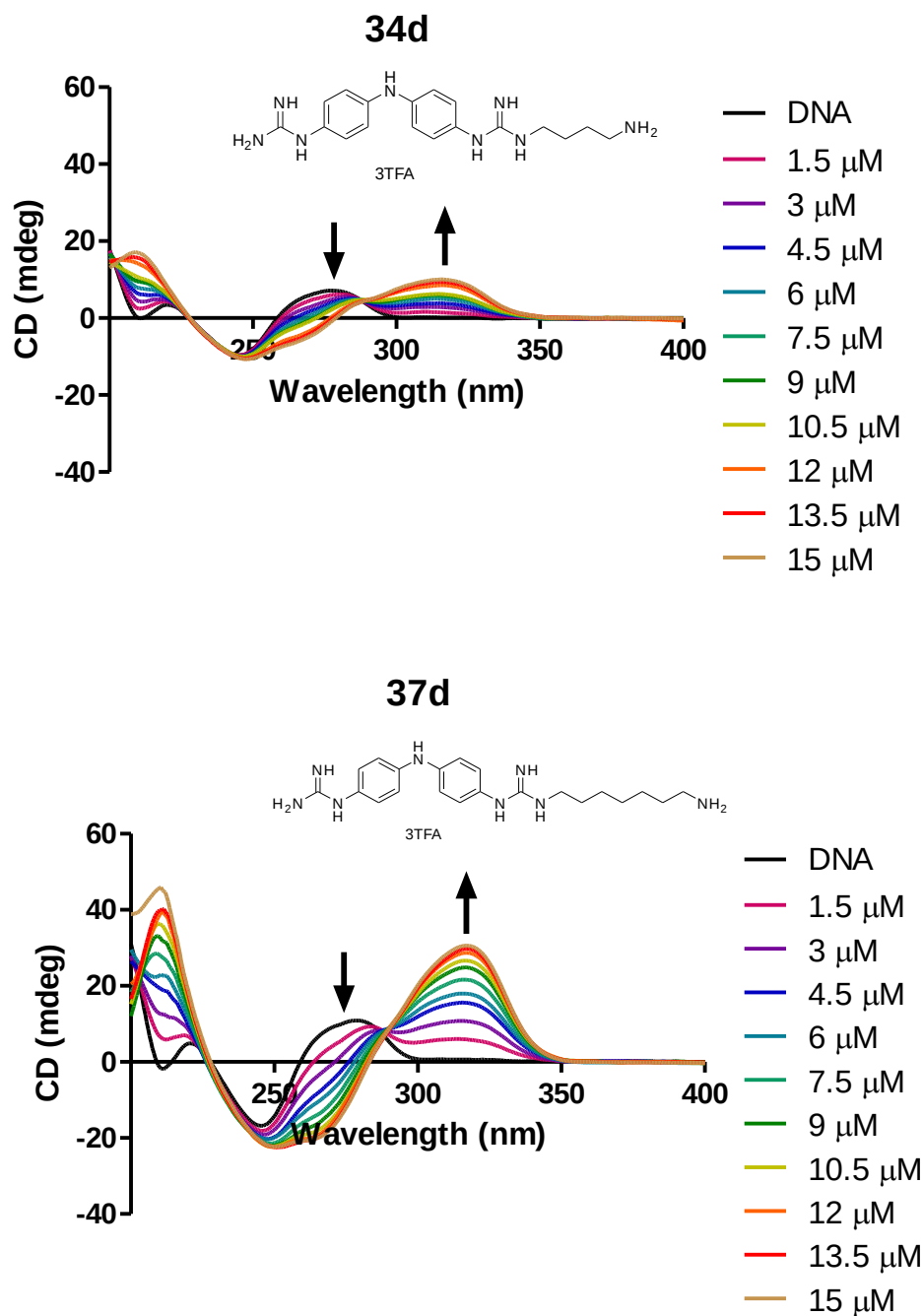


Figure 3.3.2.1 CD titrations (1.5 – 15 μM) of **34d** and **37d** with stDNA (150 μM) in 10 mM phosphate buffer.

Similarly, CO-linked derivatives **34c** ($n=2$) and **37c** ($n=5$) show a decrease in the DNA band at 275 nm and a strong positive ICD at 315 nm indicating DNA minor groove binding (Figure 3.3.2.2). However, the ellipticity of the ICD bands are not as intense, suggesting that the minor groove binding capacity is not as strong as that obtained for **37d**. Although the results

obtained from thermal denaturation were unable to rank the CO and NH linked derivatives in terms of DNA binding, these results show that the NH derivatives have greater minor groove binding potential. This is consistent with the computational results shown in Chapter 3.1 which describes the effect of linker geometry on DNA binding. Additionally, the differences in the CD spectrum of **34c** and **37c** further confirms the greater minor groove binding ability of larger alkyl chain derivatives as indicated by the higher ICD signals produced at lower concentrations.

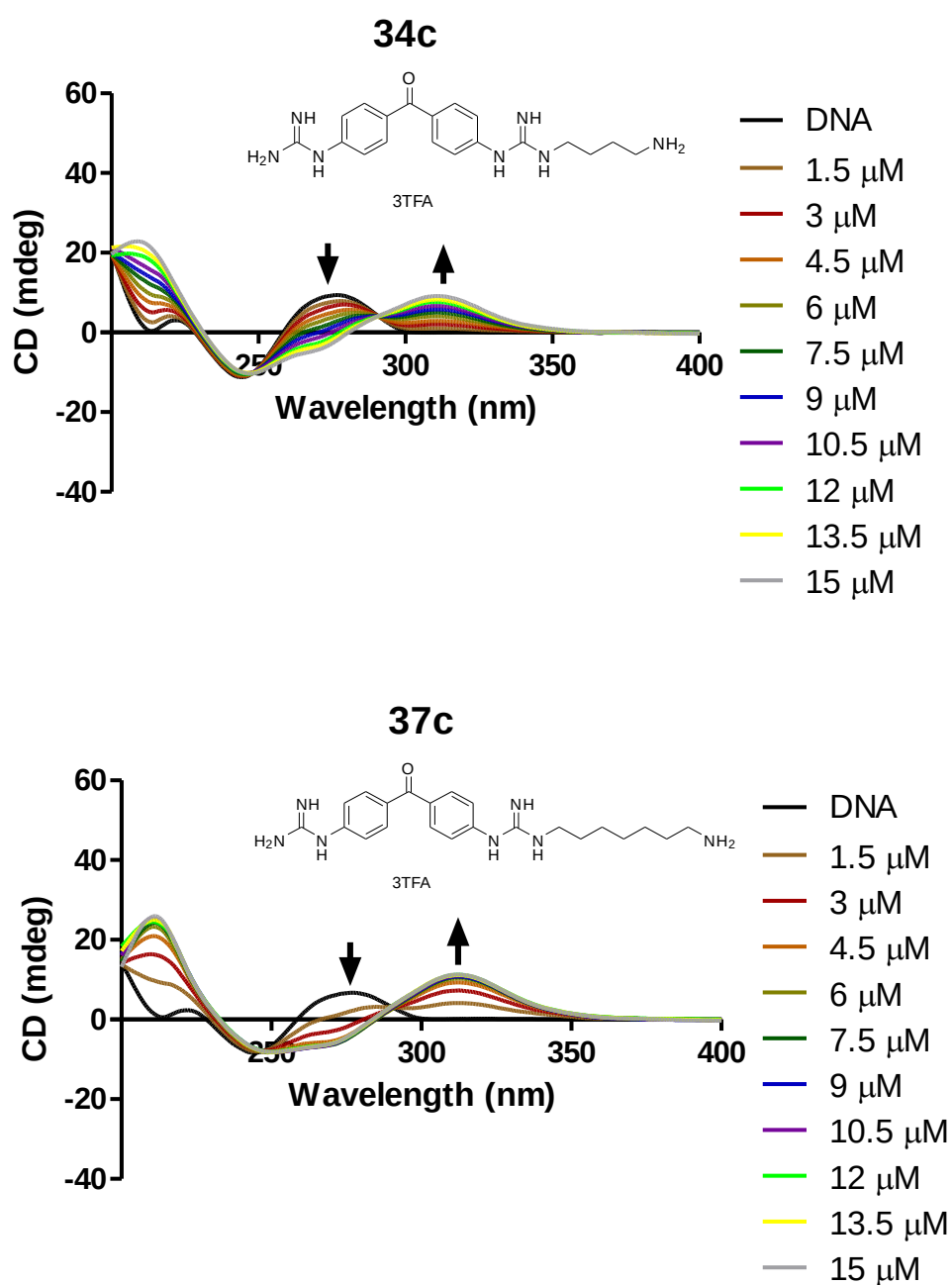
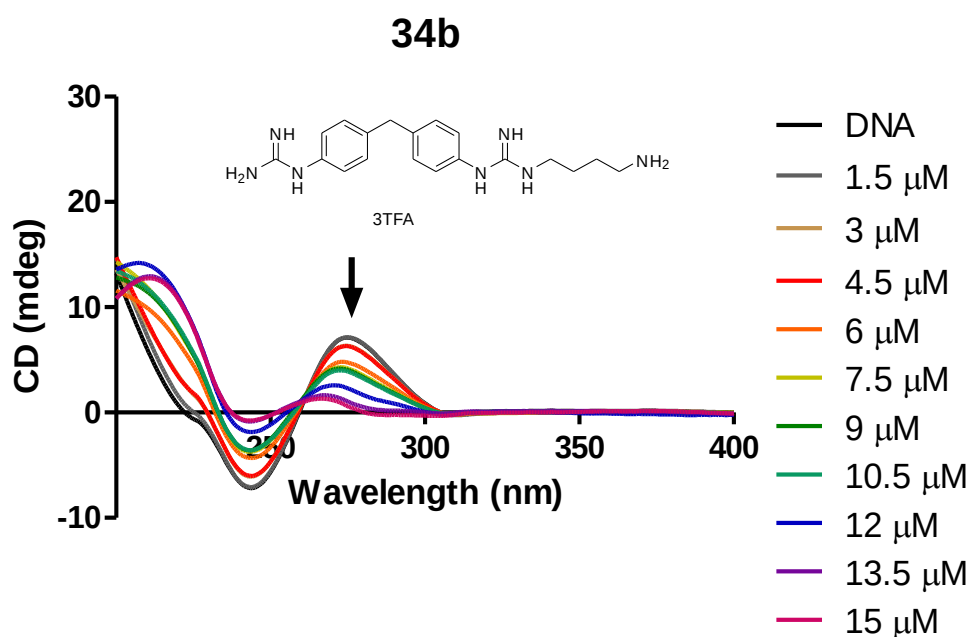


Figure 3.3.2.2 CD titrations (1.5 – 15 μM) of **34c** and **37c** with stDNA (150 μM) in 10 mM phosphate buffer.

Upon analysis of the CH_2 derivatives **34b** ($n=2$) and **37b** ($n=5$), no ICD at 315 nm was observed. Despite this, hypochromism of the DNA band at 275 nm would suggest that these compounds bind to the minor groove (Figure 3.3.2.3). The lack of a visible ICD may indicate a poor minor groove binding capacity. Additionally, a slight hypsochromic shift in the 275 nm band suggests a small change in the DNA structure. This comes as no surprise due to the nature of the linker which may result in conformational changes in the DNA structure to accommodate the molecule. As the tetrahedral geometry of the linker imposes a large dihedral angle between the phenyl rings, rotation to the favoured coplanar geometry is more energetically costly. Therefore, the diphenyl system may not be able to fully establish a dihedral angle close to zero. This effect is slightly more visible with the larger alkyl chain derivative **37b** where a clear isosbestic point at 260 nm is observed. Interestingly, these hypsochromic shifts, resulting in an A-DNA like CD spectrum (positive band at 260 nm) have been shown to occur upon binding of recognition peptides of large DNA-dependent enzymes such as RNA polymerase.²⁶



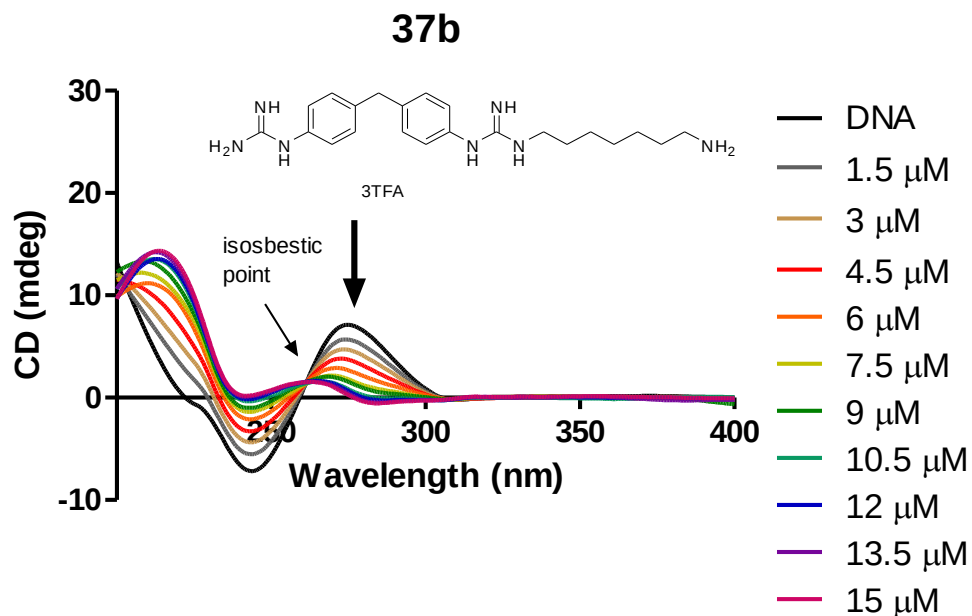


Figure 3.3.2.3 CD titrations (1.5 – 15 μM) of **34b** and **37b** with stDNA (150 μM) in 10 mM phosphate buffer.

Similar to **34b** and **37b**, O-linked derivatives **34a** ($n=2$) and **37a** ($n=5$) display no visible ICD and a more exaggerated hypsochromic shift of the 275 nm band (Figure 3.3.2.4). Again, this is consistent with the effect of tetrahedral based linkers on DNA minor groove binding. These findings correlate with our computational and thermal denaturation studies.

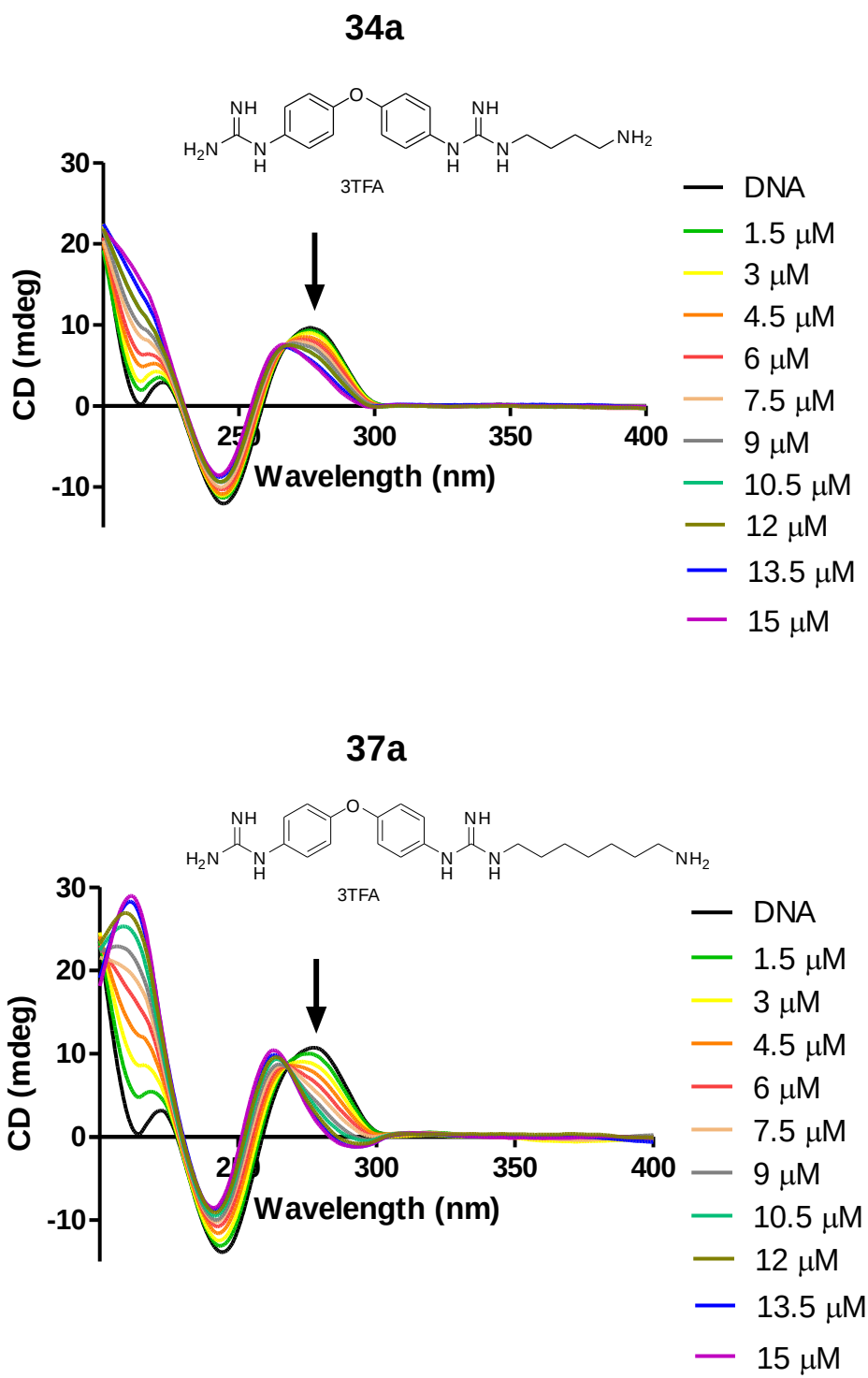
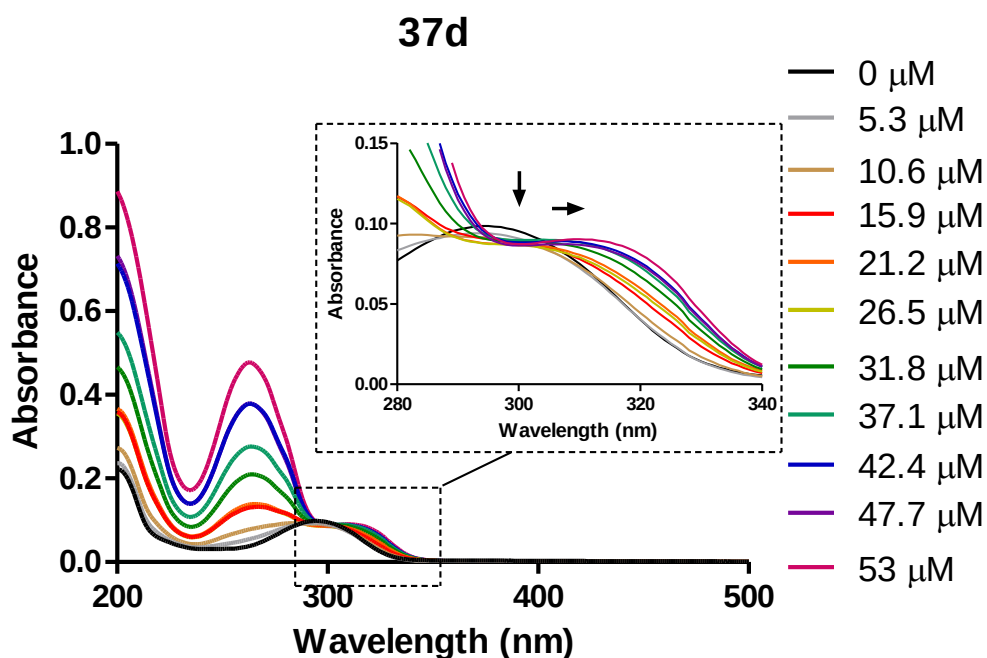


Figure 3.3.2.4 CD titrations (1.5 – 15 μM) of **34a** and **37d** with stDNA (150 μM) in 10 mM phosphate buffer.

3.3.3. UV-Vis titrations of compounds **37c** and **37d**

To get a more precise indicator of DNA binding strength, the most promising compounds **37c** and **37d** were analysed by UV-Vis titrations with the use of poly(dA·dT)₂. These experiments were carried out by adding increasing aliquots of DNA solutions to a fixed concentration (5.3 μ M) of each compound reaching a final P/D of 10. Upon addition of DNA to compound **37c**, a small hypochromic effect was observed at the λ_{max} of the ligand at 295 nm indicating the disappearance of the free molecule. That was accompanied by a bathochromic shift showing the formation of a new DNA-compound complex at 320 nm. The formation of only one isosbestic point verifies that no intermediate binding state occurs upon the interaction of the **37c** and DNA. Similarly, titration of DNA to a solution of **37d** resulted in hypochromicity of the compounds signal at 295 nm and formation of the DNA-compound signal at 320 nm. However, in comparison to the titration with **37c**, larger spectral changes were observed, reflecting the trend seen in the CD studies.



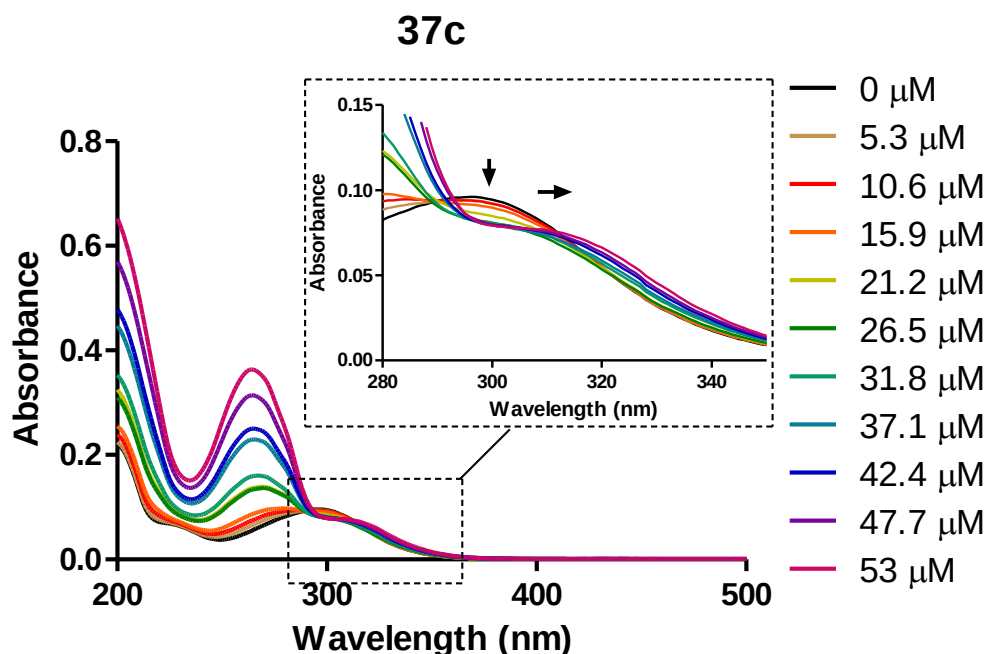
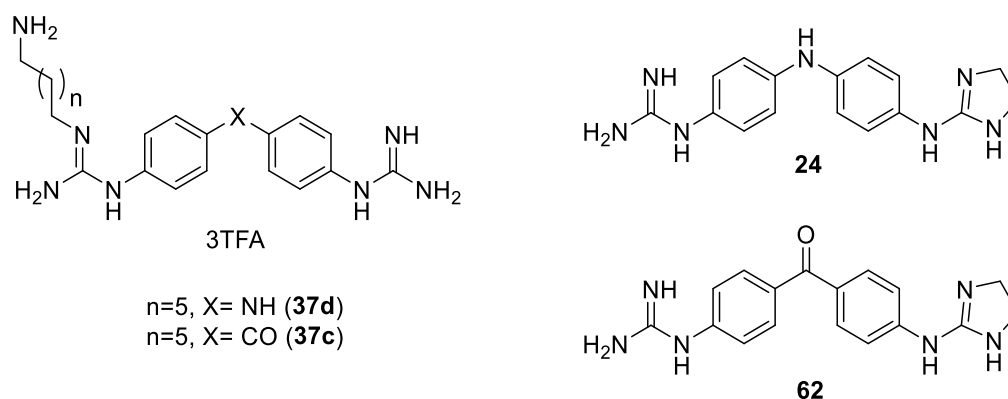


Figure 3.3.3.1. UV-Vis titrations of compound **37d** (X = NH, 5.3 μM) and compound **37c** (X = CO, 5.3 μM) with poly(dA·dT)₂ working from a P/D 0-10.

As the changes in the spectra were not readily discernible due to overlap with the intense DNA band at 260 nm, the spectrum of the molecule in the absence of DNA was subtracted from all other spectra in both cases. This allowed for the calculation of the binding constants of both compounds through Scatchard plot analysis as shown in Table 3.3.3.1. The higher binding constant of **37d** ($K_b = 5.5 \times 10^5 \text{ M}^{-1}$) versus **37c** ($K_b = 1.54 \times 10^5 \text{ M}^{-1}$) indicates a greater binding affinity towards the poly(dA·dT)₂. Comparing the binding constant of **37d** to the NH derivative of Family C (**24**, $K_b = 2.5 \times 10^5 \text{ M}^{-1}$), it is clear that incorporation of the aminoalkyl chain is massively beneficial for minor groove binding. Similarly, the binding constant of **37c** is higher than that of the CO derivative of Family C (**62**, $K_b = 0.6 \times 10^5 \text{ M}^{-1}$).

Table 3.3.3.1. Binding constants ($\times 10^5$, M^{-1}) calculated using Scatchard plot analysis from the UV titrations performed with compounds **37c** and **37d** with poly(dA·dT)₂ in phosphate buffer (10 mM).



Entry	Cmp	Linker (X)	Chain length (n)	K_b ($\times 10^5, M^{-1}$)
1	37d	NH	5	5.5 ± 0.4
2	37c	CO	5	1.5 ± 0.2
3	24	NH ^a	-	2.5 ± 0.3
4	62	CO ^a	-	0.6 ± 0.1

Melting temperature of salmon testes DNA in phosphate buffer (10 mM) is 47 °C. ^a taken from Ref. 27

3.4. Biological studies

Following the positive results obtained from biophysical experiments, *in vitro* antiparasitic activity and cytotoxicity were assessed. Six of the most promising compounds from the novel series of aminoalkyl derivatives (compounds **34a**, **34b**, **34c**, **37a**, **37c** and **37d**, Figure 3.4.1) were tested against *Trypanosoma brucei rhodesiense* (*T.b. rhodesiense*) STIB 900, *Plasmodium falciparum* (*P. falciparum*) NF54 and *Leishmania donovani* (*L. donovani*) axenic amastigotes which are the parasites responsible for human African trypanosomiasis, malaria and leishmaniasis, respectively. As stated in Chapter 1, most drugs available for protozoal diseases are outdated and exhibit unacceptable side effects such as hypotension, nephrotoxicity, cardiotoxicity and hypoglycaemia.²⁸ For this reason, the chosen compounds were tested against L6 cells which are a primary cell line derived from rat skeletal myoblasts commonly used to predict the toxicity of potential drug compounds under preclinical evaluation.²⁹ Both the antiparasitic activity and cytotoxicity was evaluated with the use of cell viability assays conducted by Dr. Marcel Kaiser at the Department of Medical Parasitology and Infection Biology in the Swiss Tropical and Public Health Institute (Basel, Switzerland). From these assays, IC₅₀ (concentration required to inhibit 50% of cellular growth) values were obtained, which allowed the ranking of the compounds in terms of antiparasitic activity. Furthermore, by comparing thermal denaturation results with the IC₅₀ values obtained we would be able to determine if DNA minor groove binding is the mechanism by which their antiparasitic activity is exerted.

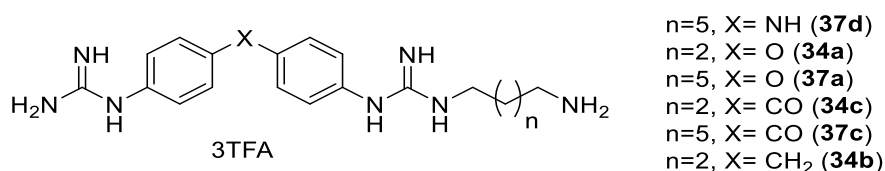
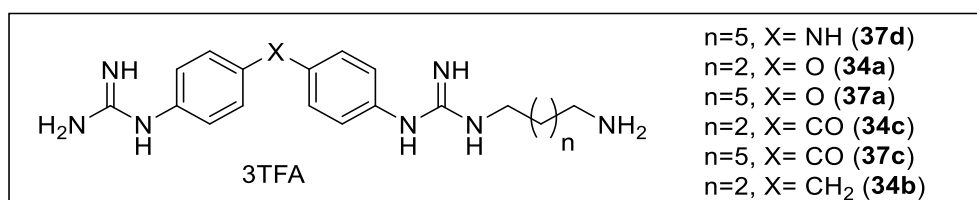


Figure 3.4.1. Aminoalkyl derivatives (**34a**, **34b**, **34c**, **37a**, **37c** and **37d**) tested for *in vitro* activity against *Trypanosoma brucei rhodesiense* STIB 900 and *Plasmodium falciparum* NF54 and *Leishmania donovani* axenic amastigotes.

3.4.1. Results obtained against the *Trypanosoma brucei rhodesiense* STIB 900 cell line

The *T. b. rhodesiense* STIB 900 cell line is known to be sensitive to all currently used antiparasitic agents and has been extensively used throughout the literature. Various MGBs including some of those developed previously by our group have been tested for anti-trypanosomal activity using this cell line.^{15, 30, 31, 32} The IC₅₀ values of compounds **34a**, **34b**, **34c**, **37a**, **37c** and **37d** against *T. b. rhodesiense* STIB 900 and L6 cells as well as their selectivity index (SI) is summarised in Table 3.4.1.1. Melarsoprol, a potent anti-trypanosomal drug with an IC₅₀ value of 0.0138 μ M, was used as a control in these experiments.

Table 3.4.1.1. *In vitro* anti-trypanosomal activity of the aminoalkyl derivatives



Entry	Cmp	<i>T. b. r</i> IC ₅₀ (μ M) ^a	<i>L6 cells</i> IC ₅₀ (μ M) ^b	Selectivity Index (SI) ^c
1	37d	0.164	39.2	239
2	34a	4.236	58.5	13.8
3	37a	2.879	12.5	4.35
4	34c	1.856	35.9	19.3
5	37c	7.497	78.6	10.4
6	34b	3.177	17.1	5.4
7	Melarsoprol	0.0138	-	-
8	Podophyllotoxin	-	0.021	-

^a*T. brucei rhodesiense* STIB 900 strain. Control: melarsoprol, IC₅₀= 0.0138 μ M.

^bL6 cells. Control: podophyllotoxin, IC₅₀= 0.021 μ M.

^cSelectivity index= (IC₅₀ L6 cells)/(IC₅₀ *T. b. rhodesiense*)

As shown in Table 3.4.1.1, all compounds exhibit very good anti-trypanosomal activity and moderate to high selectivity for *T. b. rhodesiense* versus L6 cells. The most active compound of the series was shown to be compound **37d** (X= NH, n= 5) which gave an IC₅₀ of 0.164 μ M. The lower IC₅₀ of **37d** versus **37a** (X= O, n= 5), **37c** (X= CO, n= 5) and **37b** (X= CH₂, n= 5) of the same carbon chain length indicates that the NH linker is more favourable. Additionally, **37d** exhibits the greatest selectivity towards the parasite as indicated by its selectivity index (SI = 239). Theoretically, the compounds that have the strongest affinity for the minor groove should inhibit DNA replication most effectively and, hence, elicit the strongest anti-trypanosomal activity. To investigate this, the IC₅₀ values were compared to the thermal denaturation results as shown in Table 3.4.1.2.

Table 3.4.1.2. IC₅₀ values of aminoalkyl derivatives in *T. b. rhodesiense* STIB 900 cells and stDNA binding affinity (ΔT_M)

Entry	Cmp	<i>T. b. r</i> IC ₅₀ (μ M) ^a	ΔT_M (°C) st-DNA ^b
1	37d	0.164	>20
2	34a	4.236	7.6
3	37a	2.879	17.7
4	34c	1.856	7.6
5	37c	7.497	>20
6	34b	3.177	19.6

^a*T. brucei rhodesiense* STIB 900 strain. Control: melarsoprol, IC₅₀= 0.0138 μ M

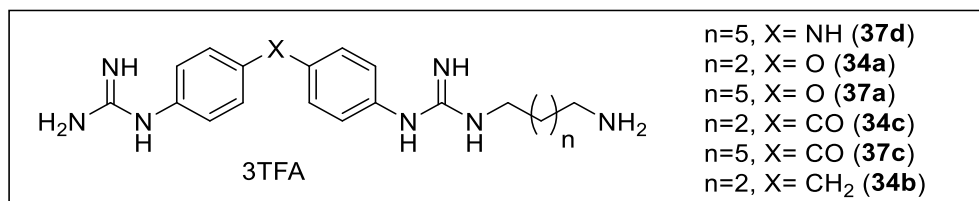
^b Using salmon-testes DNA (T_M = 67 °C); obtained in this work

A general trend was observed between the anti-trypanosomal activity (IC₅₀) and DNA binding affinity (ΔT_M). Compound **37d**, which displayed one of the strongest DNA binding affinity (stDNA ΔT_M >20 °C) also gave the lowest IC₅₀ value (0.164 μ M) of all compounds tested. Additionally, compound **34a** which displayed poor DNA binding affinity (stDNA ΔT_M = 7.6 °C)

also demonstrated poor antiparasitic activity ($IC_{50} = 4.236 \mu M$). The similar DNA binding affinities of **37a** and **34b** is also mirrored in the similar IC_{50} values obtained. Biophysical studies show that increasing carbon chain length results in greater DNA binding affinity. This also translates to better anti-trypanosomal activity when considering the differences in IC_{50} of **34a** ($X = O$, $n = 2$, $IC_{50} = 4.236 \mu M$) versus **37a** ($X = O$, $n = 5$, $IC_{50} = 2.879 \mu M$). However, this does not apply to the CO-linked derivatives where **34c** ($X = CO$, $n = 2$, $IC_{50} = 1.856 \mu M$) shows greater anti-trypanosomal activity than **37c** ($X = CO$, $n = 5$, $IC_{50} = 7.497 \mu M$) despite its lower DNA binding affinity. Previous CO-linked MGBs developed by our group have shown to exhibit good DNA binding affinity but poor anti-trypanosomal activity.^{15, 31} This may be due to off-target effects, reduced cell entry or the presence of drug metabolising enzymes. Hence, the beneficial antiparasitic effects of derivatives with larger alkyl chains cannot be ruled out.

3.4.2. Results obtained against the *Plasmodium falciparum* NF54 cell line

The *P. falciparum* NF54 cell line is susceptible to commonly used anti-malarials such as chloroquine and was used in this study to analyse the potential of these novel compounds as anti-malarial agents. Previous families A, B and E have shown to exhibit good activity against *P. falciparum* with IC_{50} values within the low micromolar range.^{15, 31} Table 3.4.2.1. summarises the results for the new compounds studied, where IC_{50} values comparable to that of chloroquine ($IC_{50} = 0.0125 \mu M$) were obtained.

Table 3.4.2.1. *In vitro* anti-plasmodial activity of the aminoalkyl derivatives

Entry	Cmp	<i>P. f</i> NF54 IC ₅₀ (μM) ^a	<i>L6 cells</i> IC ₅₀ (μM) ^b	Selectivity Index (SI) ^c
1	37d	0.062	39.2	630
2	34a	0.52	58.5	111
3	37a	0.039	12.5	314
4	34c	0.293	35.9	122
5	37c	0.366	78.6	214
6	34b	0.046	17.1	366
7	Chloroquine	0.0125	-	-
8	Podophyllotoxin	-	0.021	-

^a *P. falciparum* NF54 cell line. Control: chloroquine, IC₅₀= 0.0125 μM.

^b *L6 cells*. Control: podophyllotoxin, IC₅₀= 0.021 μM.

^c Selectivity index= (IC₅₀ *L6 cells*)/(IC₅₀ *P. falciparum*)

From the compounds tested, **37a** ($X=O$, $n=5$), **34b** ($X=CH_2$, $n=2$) and **37d** ($X=NH$, $n=5$) displayed the highest anti-plasmodial activity with IC₅₀ values of 0.039 μM, 0.046 μM and 0.062 μM respectively. Additionally, these compounds also exhibit very good selectivity for the parasite with selectivity index values above 300. Similar to the scenario observed for the CO linked derivatives **34c** and **37c** against *T. b. rhodesiense*, a slightly higher IC₅₀ was obtained for **34c** (CO, $n=2$, IC₅₀ of 0.293 μM) than **37c** (CO, $n=5$, IC₅₀ of 0.366 μM) despite the longer alkyl chain length and stronger DNA binding affinity. This may indicate that having an electron-withdrawing group as a linker is detrimental for anti-parasitic activity. Compound **37a** with a

ΔT_M of 17.7 °C gave a significantly lower IC_{50} than **34a** which has a ΔT_M of 7.6 °C. Additionally, low IC_{50} values were obtained for **34b** and **37d** which display ΔT_M values of 19.6 °C and >20 °C respectively. This suggests that strong DNA binding compounds with electron-donating linkers (i.e. NH, O, and CH₂) have greater anti-plasmodial activity. Moreover, the high selectivity indices obtained for all compounds highlights their safety and potential as effective anti-malarial agents.

Table 3.4.2.2. IC_{50} values of aminoalkyl derivatives in *P. falciparum* NF54 cells and stDNA binding affinity (ΔT_M)

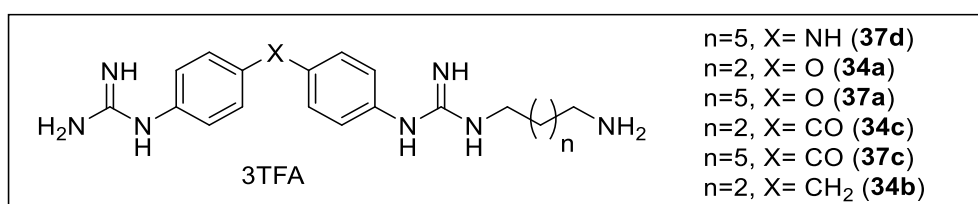
Entry	Cmp	<i>P. f</i> NF54 IC_{50} (μM) ^a	ΔT_M (°C) st-DNA ^b
1	37d	0.062	>20
2	34a	0.52	7.6
3	37a	0.039	17.7
4	34c	0.293	7.6
5	37c	0.366	>20
6	34b	0.046	19.6

^a *P. falciparum* NF54 cell line. Control: chloroquine, IC_{50} = 0.0125 μM .

^b Using salmon-testes DNA (T_M = 67 °C); obtained in this work

3.4.3. Results obtained against the *Leishmania donovani* axenic amastigotes

As stated in Chapter 1, during the leishmania life cycle, they assume different forms in different hosts which is directly associated with changes in physicochemical conditions. Prior to infecting a human host, the parasite is in the form of promastigotes within the insect vector. Upon infection of a human host, promastigotes develop into amastigotes which invade various tissues in the body. *In vitro* screening of potential anti-leishmanial agents is known to be challenging due to the difficulty in obtaining amastigotes. While promastigotes are easy to maintain, amastigotes are more complex and can only be obtained using infection assays of tissues or *in vitro* cells which allow for the differentiation of promastigotes to axenic amastigotes.³³ For these experiments, *L. donovani* axenic amastigotes were used to assess the potential anti-leishmanial activity of the chosen compounds (summarised in Table 3.4.3.1). Among the FDA approved treatments for leishmaniasis, miltefosine has shown to exhibit good potency against the parasite. For this reason, it was used as a control where it displayed a low IC₅₀ of 0.799 µM against the *L. donovani* axenic amastigote cell line.

Table 3.4.3.1. *In vitro* anti-leishmanial activity of the aminoalkyl derivatives

Entry	Cmp	<i>L. donovani</i> ax. am. IC ₅₀ (μM) ^a	<i>L6 cells</i> IC ₅₀ (μM) ^b	Selectivity Index (SI) ^c
1	37d	62.2	39.2	0.631
2	34a	93.7	58.5	0.625
3	37a	48.3	12.5	0.259
4	34c	124.2	35.9	0.289
5	37c	70.3	78.6	1.118
6	34b	70.7	17.1	0.241
7	Miltefosine	0.799	-	-
8	Podophyllotoxin	-	0.021	-

^a *L. donovani* axenic amastigotes. Control: miltefosine, IC₅₀= 0.799 μM.

^b *L6 cells*. Control: podophyllotoxin, IC₅₀= 0.021 μM.

^c Selectivity index= (IC₅₀ *L6 cells*)/(IC₅₀ *L. donovani*)

Unfortunately, all compounds tested exhibited very poor activity against the *L. donovani* axenic amastigote cell line with IC₅₀ values from 48.3 to 124.2 μM which is more than a 60 fold difference in activity compared to miltefosine. Additionally, all compounds showed very poor selectivity for the parasite with selectivity indices below 1.1. Hence, they would not be safe or effective anti-leishmanial agents. The most active compound of the series was **37a** (X= O, n= 5, IC₅₀ = 48.3 μM) which also exhibited the best activity in *P. falciparum* NF54 cells and moderate activity in the *T. b. rhodesiense* STIB 900 cell line.

Table 3.4.3.2. IC₅₀ values of aminoalkyl derivatives in *L. donovani* axenic amastigotes and stDNA binding affinity (ΔT_M)

Entry	Cmp	<i>L. don.</i> ax am IC ₅₀ (μ M) ^a	ΔT_M ($^{\circ}$ C) st-DNA ^b
1	37d	62.2	>20
2	34a	93.7	7.6
3	37a	48.3	17.7
4	34c	124.2	7.6
5	37c	70.3	>20
6	34b	70.7	19.6

^a *L. donovani* axenic amastigotes. Control: miltefosine, IC₅₀ = 0.799 μ M.

^b Using salmon-testes DNA (T_M = 67 $^{\circ}$ C); obtained in this work

Interestingly, when considering compounds **37a**, **37c** and **37d**, derivatives bearing the larger alkyl chain (n=5) displayed better activity than the derivatives of the same linker with shorter alkyl chains (n=2). The same can be seen from biophysical studies where DNA binding affinity increases upon increasing alkyl chain length. When comparing the IC₅₀ values to the DNA binding affinity, compounds **34a** (X= O, n=2) and **34c** (X= CO, n= 2) which gave ΔT_M values of 7.6 $^{\circ}$ C also gave the highest IC₅₀ values of 124.2 and 93.7 respectively (Table 3.4.3.2). Additionally, the similar DNA binding affinities of **37d** (X= NH, n= 5) and **37c** (X= CO, n= 5) are reflective in the similar IC₅₀ values obtained. However, when comparing **37a** (ΔT_M = 17.7 $^{\circ}$ C; IC₅₀ = 48.3 μ M) to **34b** (ΔT_M = 19.6 $^{\circ}$ C; IC₅₀ = 70.7 μ M) a slight miscorrelation between DNA binding affinity and anti-leishmanial activity is apparent which may be due to possible off-target effects or difference in drug transportation.

3.5. Conclusions

In this work molecular docking studies of previous MGBs produced by the Rozas group was performed using a reference crystal structure of the DNA minor groove. These experiments gave a better understanding of the importance of certain structural features on DNA binding. Additionally, molecular docking studies of Family E in *pf*DHFR showed that short-chained derivatives exhibit similar binding interactions as known *pf*DHFR inhibitors. This indicates that Family E may also be targeting *pf*DHFR explaining the miscorrelations observed between DNA binding affinity and anti-parasitic activity in previous works. Docking studies of newly designed MGBs (Family F) in the DNA minor groove showed that these derivatives exhibit better DNA binding properties than previous families. A subset of these compounds were also docked into the *pf*DHFR active site which resulted in less favourable binding poses and lower DHFR binding affinity scores than Family E.

Following these promising results, the synthesis of Family F was conducted. The synthetic route first involved the production of mono-guanidylated di-aromatic intermediates (**38a-d**) followed by ethyl carbamate protected thiourea formation (**39a-d**). These thioureas were then converted to the desired protected *N*-aminoalkyl guanidines (**41a-d** to **44a-d**) through guanidylation using various mono Boc-protected alkyl diamines (**40a-d**). Using methods recently reported by our group, the ethyl carbamate functional groups of **41a-d** to **44a-d** were converted to *tert*-butyl carbamates (**45a-d** to **48a-d**) in moderate to high yield. This allowed for global deprotection using TFA affording the final trifluoroacetate salts (**34a-d** to **37a-d**) in good yield and purity.

Biophysical evaluation of Family F showed that they exhibit very impressive DNA binding properties. DNA UV thermal denaturation experiments using stDNA and poly(dA-dT)₂ oligonucleotide showed that these novel compounds have stronger DNA binding affinity than previous families and known MGB, pentamidine. The high affinity for poly(dA-dT)₂ as well as the results obtained from CD experiments indicate that these compounds have a strong DNA minor groove binding capacity. This was confirmed by the superior DNA binding constants of the most promising compounds, **37c** and **37d**, versus that of previous MGBs developed by our group.

Lastly, biological analysis of the six most promising compounds (**34a**, **34b**, **34c**, **37a**, **37c** and **37d**) against *T. b. rhodesiense* STIB 900 cells, *P. falciparum* NF54 cells, *L. donovani* axenic amastigotes and L6 skeletal cells was performed by the Department of Medical Parasitology and Infection Biology at the Swiss Tropical and Public Health Institute. Low micromolar to nanomolar IC₅₀ values were obtained against *T. b. rhodesiense* and *P. falciparum* with high selectivity for the parasites over healthy L6 cells. Unfortunately, poor activity and selectivity was observed in *L. donovani*. However, a general trend between DNA binding affinity and anti-parasitic activity was observed indicating that DNA binding is the main mechanism at play. Compound **37d** which has one of the strongest DNA binding affinities within Family F ($\Delta T_M = >20\text{ }^\circ\text{C}$) produced the greatest activity against *T. b. rhodesiense* (IC₅₀ = 0.164 μM). Additionally, **37d** as well as strongly binding derivatives **37a** and **34b** exhibit potent anti-parasitic activity against *P. falciparum* displaying IC₅₀ values of 0.06 μM , 0.039 μM and 0.046 μM , respectively. Given the promising results obtained against *T. b. rhodesiense* and *P. falciparum*, these novel MGBs could be good candidates for further *in vivo* analysis to assess their potential as effective treatments for HAT and malaria.

3.6. References

1. <https://www.rcsb.org/structure/1D64>
2. <https://www.rcsb.org/structure/1Z8V>
3. <https://www.rcsb.org/structure/4OCD>
4. Becke, A. D., *J. Chem. Phys.*, **1993**, *98*, 1372–1377.
5. Zhao, Y.; Truhlar, D. G., *Theor. Chem. Acc.*, **2008**, *120*, 215–241.
6. Hehre, W. J.; Ditchfield, R. and Pople, J. A., *J Chem. Phys.*, **1972**, *56*, 2257–2262.
7. Marenich, A. V.; Cramer, C. J.; Truhlar, D. G., *J. Phys. Chem. B.*, **2009**, *113*, 6378–6396.
8. Gaussian 16, Revision C.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian, Inc., Wallingford CT, **2016**.
9. Trott, O and Olson, A. J., *Journal of Computational Chemistry*, **2010**, *31*, 455–461
10. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC
11. Humphrey, W.; Dalke, A. and Schulten, K., "VMD - Visual Molecular Dynamics", *J. Molec. Graphics*, **1996**, *14*, 33–38.
12. O'Sullivan, P.; Rozas, I. *ChemMedChem*, **2014**, *9*, 2065–2073.
13. Acosta-Reyes, F. J.; Dardonville, C.; De Koning, H. P.; Natto, M.; Subirana, J. A.; Campos, J. L. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **2014**, *70*, 1614–1621.

14. Nagle, P. S.; Rodriguez, F.; Kahvedžić, A.; Quinn, S. J.; Rozas, I., *J. Med. Chem.*, **2009**, 52, 7113–7121.
15. McKeever, C.; Kaiser, M.; Rozas, I., *J. Med. Chem.*, **2013**, 56, 700–711.
16. Kim, K. S.; Qian, L. *Tetrahedron Lett.*, **1993**, 34, 7677.
17. Nagle, P. S.; Rodriguez, F.; Kahvedžić, A.; Quinn, S. J.; Rozas, I., *J. Med. Chem.*, **2009**, 52, 7113–7121.
18. Suhs, T.; Kçnig, B., *Chem. Eur. J.*, **2006**, 12, 8150 – 8157.
19. Flemer, S.; Madalengoitia, J. S., *Synthesis*, **2007**, 12, 1848-1860.
20. Patrick O’Sullivan, PhD Thesis, University of Dublin, **2016**
21. Manimala, J. C.; Anslyn, E. V., *Tetrahedron Lett.*, **2002**, 109, 2455-2504
22. Prabhakaran, P.; Sanjayan, G. J., *Tetrahedron Lett.*, **2007**, 48, 1725–1727.
23. Rahman, A.; Stipaničev, N.; Keogh, A.; Twamley, B.; Rozas, I., *Tetrahedron Lett.*, **2021**, 153160.
24. Sartori, G.; Maggi, R., *Chem. Rev.*, **2010**, 113, PR1
25. Yakovchuk, P.; Protozanova, E. and Frank-Kamenetskii, M. D., *Nucleic. Acids. Res.*, **2006**, 34, 564.
26. Ivanov, V. I.; Minchenkova, L. E.; Schyolkina, A. K.; Poletayev, A. I. *Biopolymers*, **1973**, 12, 89–110.
27. Nagle, P. S.; Quinn, S. J.; Kelly, J.; O'Donovana, D; Khanb, A. R.; Fernando Rodriguez, F.; Nguyenc, B.; Wilson, W. D. and Rozas, I., *Org. Biomol. Chem.*, **2013**, 8, 1–25.
28. O'Brien, J. G.; Dong, B. J.; Coleman, R. L.; Gee, L.; and Balano, K. B., *Clin. Infect. Dis.*, **1997**, 24, 854.
29. Richler C. and Yaffe, D., *Dev. Biol.*, **1970**, 23, 1.
30. Movassaghi M. and Schmidt, M. A., *Org. Lett.*, **2005**, 7, 2453
31. Lemster, T.; Pindur, U.; Lenglet, G.; Depauw, S.; Dassi, C.; David-Cordonnier, M., *Eur. J. Med. Chem.*, **2009**, 44, 3235.
32. Dardonville, C.; Brun, R., *J. Med. Chem.*, **2004**, 47, 2296–2307.
33. Dias-Lopes, G.; Zabala-Peñafiel, A.; de Albuquerque-Melo, B. C.; Souza-Silva, F.; Menaguali do Canto, L.; Cysne-Finkelstein, L.; Alves, C. R., *Acta Trop.*, **2021**, 220, 105956

Chapter 4
***Guanidine based derivatives as potential
anti-tubercular agents***

4.1. Preliminary investigations

As stated in Chapter 1, small molecules containing piperazine (PZ) and piperidine (PD) moieties have a long track record of behaving as potent anti-infective agents. More recently, a study conducted by Forge *et al*, reported a series of PD and PZ containing bis-amidines with good anti-mycobacterial activity and bear high structural similarity to a subset of containing within our in-house family of compounds (Figure 4.1.1).¹ In this particular study, PZ derivatives (**13-18**) showed better *in vitro* activity ($MIC_{50} = 2.162 - 0.052 \mu M$) against the drug-sensitive *M. tuberculosis* strain H37Rv than isoniazid ($MIC_{50} = 0.465 \mu M$), a first-class anti-tubercular agent. The low MIC_{50} values obtained probed investigation into the intracellular activity of these compounds by assessing their ability to selectively target mycobacteria within human macrophages. Results from this experiment showed 88% H37Rv inhibition at $2 \mu M$ with the most active compound (**17**) versus 52% inhibition with isoniazid. Additionally, compounds **13-18** were shown to be effective against drug-resistant clinical isolates of *M. tuberculosis* highlighting the potential of this drug class as novel anti-tubercular agents.

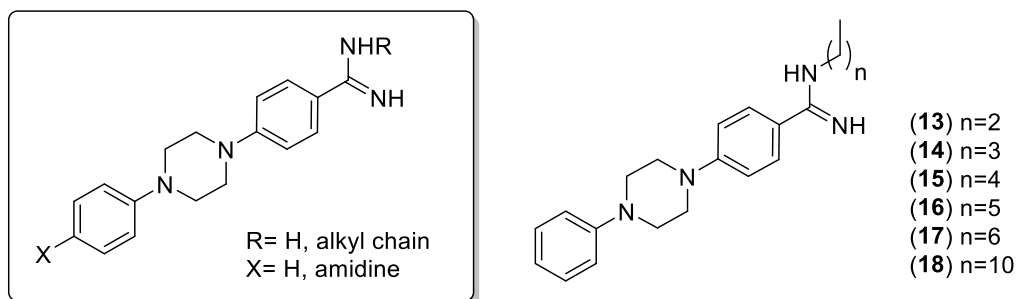


Figure 4.1.1. Structures of di-aryl piperazine based compounds produced by Forge *et al*. with anti-mycobacterial activity

This study along with a regular appearance of PD and PZ moieties within the pharmacophore of current and emerging anti-infective agents encouraged investigation of some guanidine/2-aminoimidazoline derivatives from our in-house library (Families A-C) containing these functionalities (Figure 4.1.2). In collaboration with the group of Prof. Stephen Gordon at the School of Veterinary Medicine in University College Dublin, a series of these compounds were

screened against *Mycobacterium smegmatis* and *Mycobacterium bovis* BCG as surrogate strains of *M. tuberculosis*. Additionally, with the help of Dr. Apoorva Bhatt and Dr. Sarah Batt from the University of Birmingham, whole-cell phenotypic assays and enzyme-based biochemical assays were conducted to investigate the mechanism of action of these compounds.

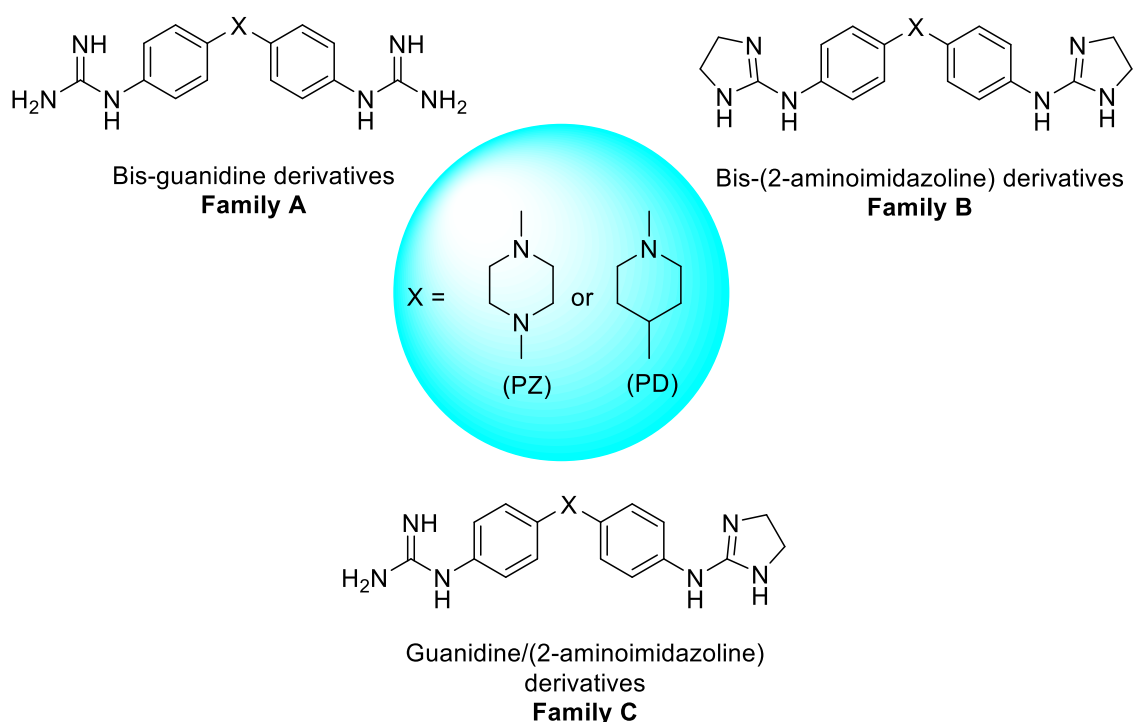


Figure 4.1.2. Structures of piperazine/piperidine based derivatives of Families A, B and C under investigation.

4.1.1. Resazurin based microtiter plate assay (REMA) analysis of previously made compounds against mycobacterial strains

Resazurin based microtiter plate assay (REMA) is a well-known methodology used to assess drug susceptibility. Due to its simplicity, speed, sensitivity and reproducibility, REMAs have been employed to determine the anti-mycobacterial activity of current and emerging anti-tubercular agents *in vitro*. This method uses the redox colorimetric dye resazurin which is blue in its oxidised state. When subjected to the reducing environment of the viable cells, resazurin is reduced to the pink and highly fluorescent resorufin (Figure 4.1.1.2).² Hence, the amount

of fluorescence is proportional to the number of viable cells. The REMA experiment consists of a 96-well plate with a serial dilution of drug compounds in the presence of the mycobacterial cells. The cell viability at each concentration is determined by comparing the fluorescence of test wells to that of the controls. From this, the minimum inhibitory concentration (MIC) can be determined which is the drug concentration at 10% cell viability.

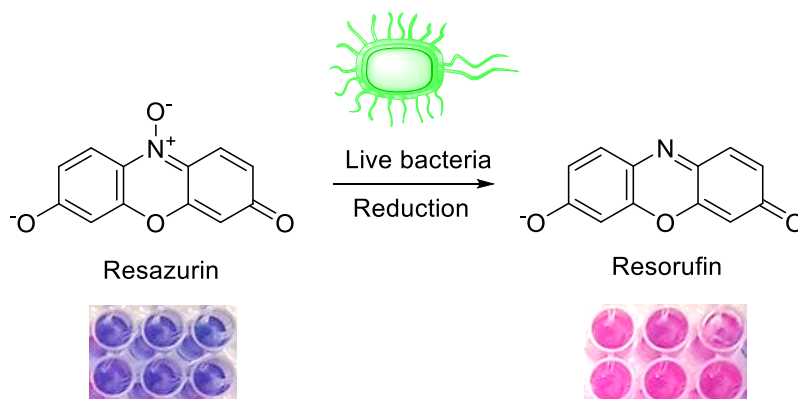


Figure 4.1.1.1. The reaction involved in REMA.

Using these experiments a high throughput screen was conducted with the previously synthesised derivatives of interest, where the resulting cell viability curves were used to calculate the MIC for each compound (Figure 4.1.1.2). Due to the potential threat of performing these assays with *M. tuberculosis* (H37Rv strain), safer and non-pathogenic surrogate strains were used to perform these experiments. These include *M. bovis* BCG (Tokyo and Danish strain) and *M. smegmatis*. *M. bovis* BCG is an attenuated strain of *mycobacterium bovis* (*M. bovis*) and is the active agent used in the prophylactic vaccine against tuberculosis. As *M. bovis* BCG is among the *Mycobacterium tuberculosis* complex (MTBC) and shares up to 99% genetic similarity with *M. tuberculosis*, use of this strain is a good predictor form anti-tubercular activity.³ Similarly, the fast-growing *M. smegmatis* bears high genetic similarity to *M. tuberculosis* and can be used in standard Biosafety Level 2 containment facilities.

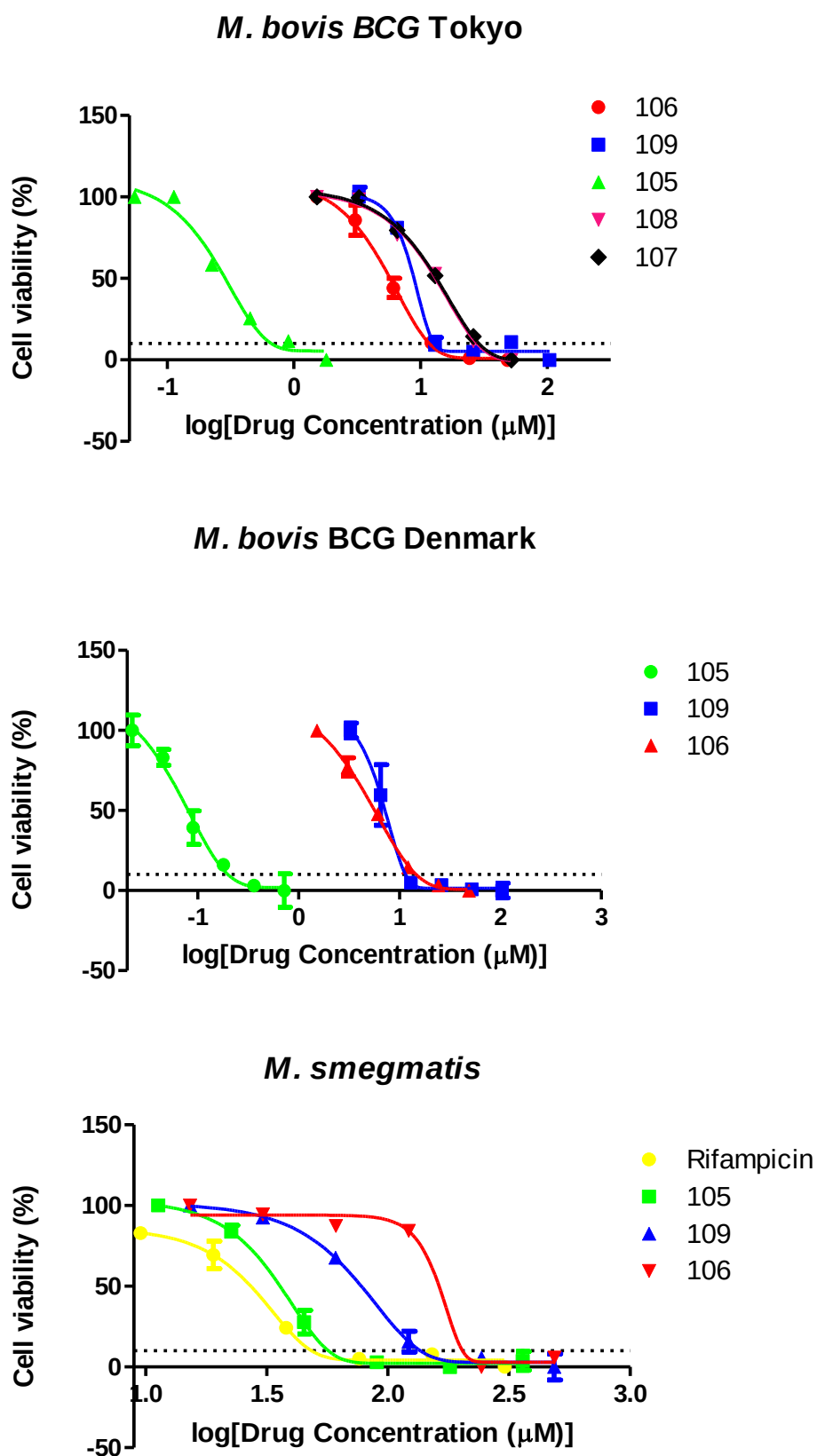
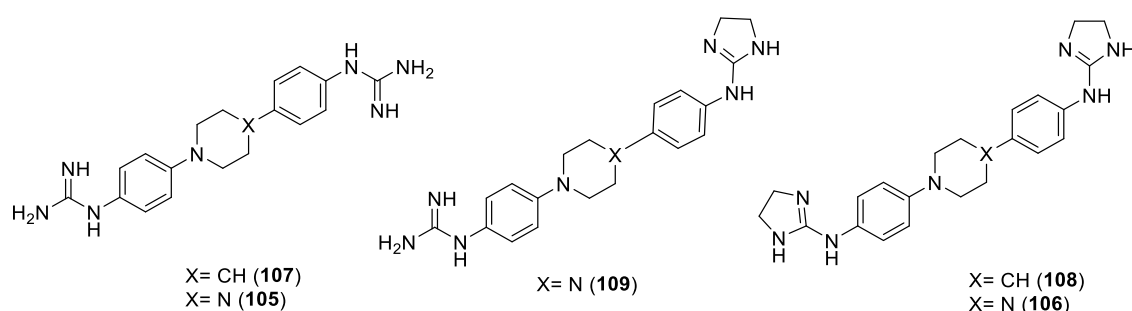


Figure 4.1.1.2. Cell viability curves generated from REMA analysis of our in-house MGBs against *M. bovis* BCG Tokyo, *M. bovis* BCG Denmark and *M. smegmatis* (the dotted line represents 10% cell viability mark which corresponds to the MIC).

Table 4.1.1.1 shows the MIC values obtained for all compounds against all the three mycobacterial strains. Analysing the results obtained from the initial screen against the *M. bovis* BCG Tokyo strain, the piperazine-linked derivatives (**105**, **106** and **109**) displayed better anti-mycobacterial activity than the piperidine linked derivatives (**107** and **108**) as indicated by the lower MICs. Given the lower MICs obtained for **105**, **106** and **109**, they were also screened against *M. bovis* BCG Denmark. Similar anti-mycobacterial activity was observed for these compounds with **105** proving to be the most active bearing a MIC of 0.38 μM . Comparing their activity to that observed against *M. smegmatis*, an increase in MIC was evident which can be explained by the increase in metabolic rate of the fast-growing bacteria. This can also be seen in the difference in the MIC of rifampicin in *M. bovis* BCG Tokyo (0.45 μM) versus *M. smegmatis* (50.2 μM).

Table 4.1.1.1. Minimum inhibitory concentration (MIC) of previously made in-house family of compounds against mycobacterial strains with their corresponding DNA binding affinity against the poly(dA)-poly(dT) oligonucleotide.



Entry	Cmp	<i>M. bovis</i> BCG Tokyo MIC (μM)	<i>M. bovis</i> BCG Den. MIC (μM)	<i>M. smeg.</i> MIC (μM)	ΔT_M ($^{\circ}\text{C}$) Poly(dA)- Poly(dT) ^a
1	108	>20	-	-	35
2	107	>20	-	-	30
3	106	12.5 $\pm$ 0.185	15.33 $\pm$ 0.97	205.9 $\pm$ 0.4	38.8
4	105	1 $\pm$ 0.5	0.38 $\pm$ 0.27	60.7 $\pm$ 9.7	27.8
5	109	12.65 $\pm$ 2.3	13.3 $\pm$ 6.2	142.6 $\pm$ 20.8	15
6	RIF	0.45 $\pm$ 0.15 ^b	-	50.2 $\pm$ 7.8	-

Cells were used at an OD_{600} of 0.1 ($\sim 3 \times 10^7$ CFU/ mL) in a 96-well plate and treated with a serial dilution of drug compound dissolved in H_2O or DMSO. Once treated, cells were incubated for 3 days for *M. smeg* or 5 days for BCG strains at 37°C then treated with resazurin and left to incubate for 24 or 48 h respectively. The resulting fluorescence was read using a plate reader from which percentage viability was calculated. MIC was defined by the drug concentration at 10% cell viability and was calculated using the "Gompertz equation for MIC determination" through GraphPad; ^a ΔT_M values obtained from Ref. 4. ^b MIC of rifampicin against *M. bovis* BCG Tokyo obtained from Ref. 5.

The comparable activity of **105** to rifampicin highlights the potential of this drug class as anti-tubercular agents. From these results, a structure-activity relationship (SAR) was developed which shows the key structural features that are beneficial for anti-mycobacterial activity (Figure 4.1.1.3). As stated earlier, compounds with piperazine linkers show greater activity and

of these derivatives the less bulky guanidine functionality is more tolerated at both terminals of the molecule. It is also important to note that the counter-ion of the salt does not play an essential role in the activity of the compounds as shown in the appendix (Table 1).

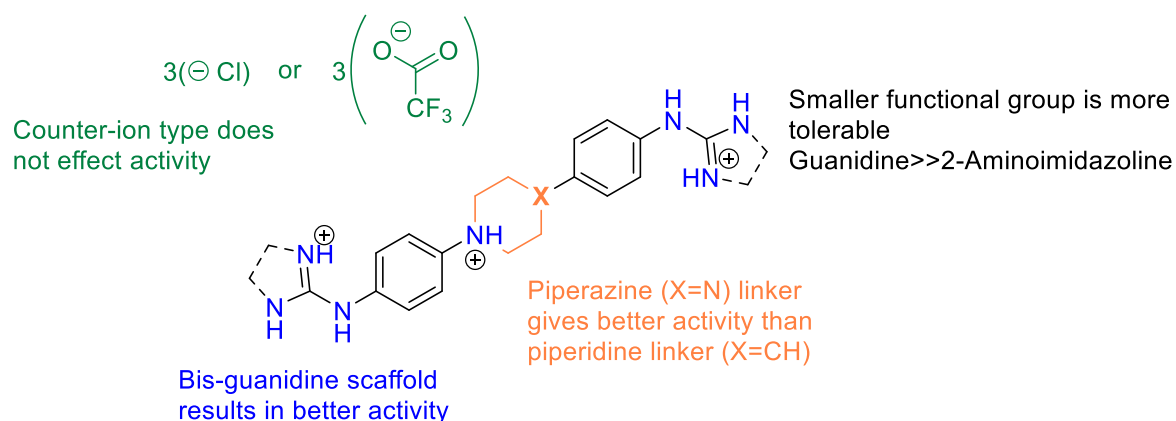


Figure 4.1.1.3. Structure-activity relationship (SAR) of previously made in-house family of compounds.

As the compounds investigated above were previously designed as DNA MGBs, the relationship between DNA binding and anti-mycobacterial activity was analysed (Table 4.1.1.1). Surprisingly, compounds that exhibit large ΔT_M (strong DNA binding affinity) values such as **108** (35 °C) and **106** (38 °C) display higher MIC values than **109** and **105** which exhibit weaker DNA binding affinity. This miscorrelation between DNA binding and anti-mycobacterial activity is consistent among all strains used which led to the hypothesis that an alternative mechanism of action may be at play. Based on structural similarities between these compounds and known inhibitors of mycobacterial cell wall biosynthesis that act through mycolic acid or DprE1 inhibition, it was hypothesised that these compounds may be acting in the same manner. To confirm this hypothesis whole-cell phenotypic assays and enzyme-based assays were conducted. Additionally, further computer-aided investigation was performed which is described in Chapter 4.2.

4.1.2. Analysis of mycolic acid biosynthesis through TLC FAMES and MAMES assays

Whole-cell phenotypic screening is the visual observations of the effects that chemical entities have on whole cells *in vitro*. Although a shift from whole-cell phenotypic screening to target-based approaches for drug discovery has occurred within recent years, there have been certain problems faced by compounds derived from target-based approaches in translating their activity into cell-based assays.⁶ For this reason, analysis of drug compounds that have been repurposed for new indications has been subject to whole-cell phenotypic assays to identify their mechanism of action. This is particularly true in anti-tubercular drug discovery. Structural similarity between the above PZ/PD-containing compounds and known anti-tubercular agents known to target mycobacterial mycolic acid synthesis encouraged the investigation into this biosynthetic pathway (Figure 4.1.2.1).

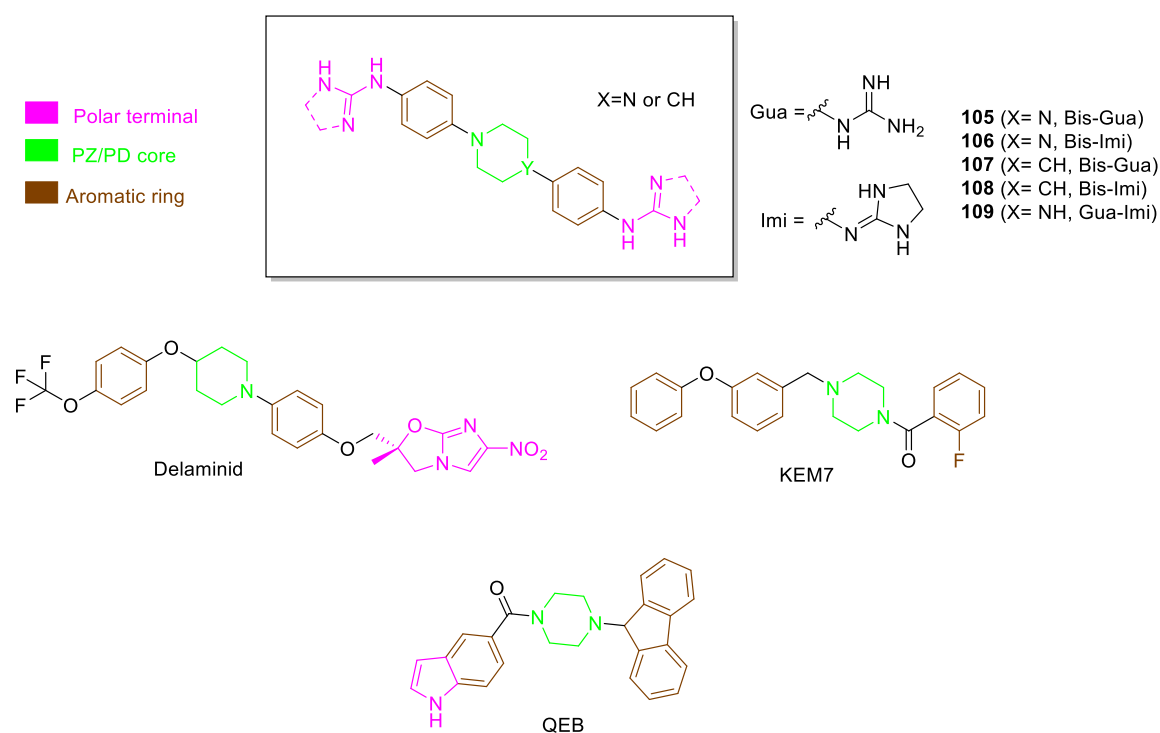


Figure 4.1.2.1. Structures of some compounds from our in-house library (**105**, **106**, **107**, **108** and **109**) and structurally similar mycolic acid inhibitors, delamanid, KEM7 and QEB.

Mycolic acids are the primary constituent of the mycobacterial cell wall and contribute to outer membrane permeability and integrity as well as virulence. In mycobacteria, the

synthesis of mycolic acids is governed by the fatty acid synthase pathway I (FAS I) and II (FAS II). As shown in Figure 4.1.2.2, the saturated α -alkyl chain (C22 – C26 fatty acids) and the long meromycoloyl chain (C40 – C60 fatty acids) are synthesized by the FAS-I and FAS-II complexes, respectively. The C16-AcpM formed in the FAS I pathway is channeled to the FAS II pathway, where it undergoes a round of keto-reduction, dehydration and enoyl- reduction which is catalysed by: MabA (β -ketoacyl-AcpM reductase), HadAB/BC (β - hydroxyacyl-AcpM hydratase) and InhA (enoyl-AcpM reductase). The resulting saturated long-chain meromycolate can either undergo further elongation by KasA/B or undergo modifications such as cis-/trans-cyclopropanation, and the addition of methoxy and keto groups to yield the desired mycolic acids. The intact mycolic acid is then shuttled to the periplasm as a trehalose ester by the membrane transporter, MmpL3.^{7, 8}

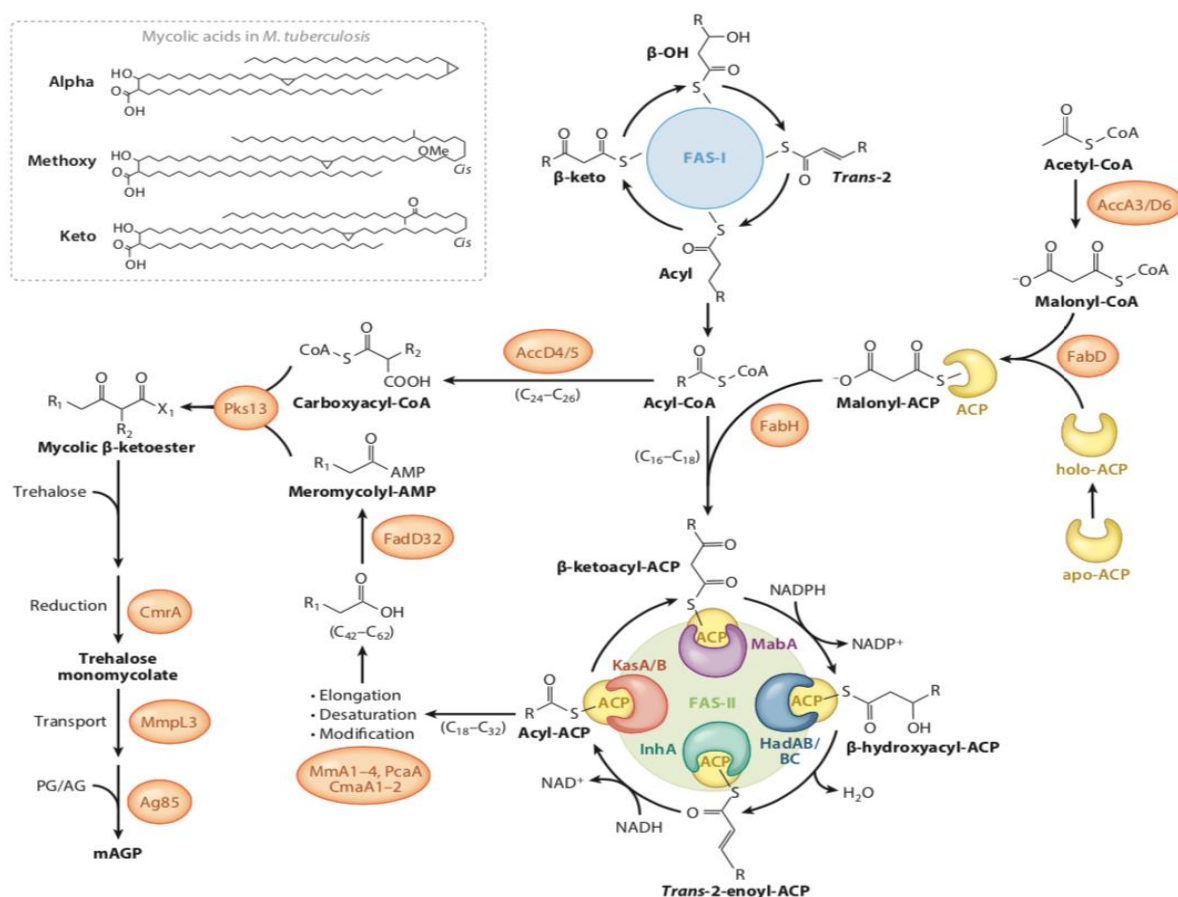


Figure 4.1.2.2. Mycolic acid biosynthesis pathways in *M. tuberculosis*.⁹

To investigate the effect of the most active compounds **105** and **109** on the mycolic acid biosynthetic pathway, *M. smegmatis* prelabelled with ¹⁴C-acetate was treated with each compound. After lipid extraction and esterification to afford the corresponding methyl esters,

the samples were analysed for mycolic acid methyl esters (MAMEs) and their precursors, short-chain fatty acid methyl esters (FAMES) via thin-layer chromatography (TLC) (Figure 4.1.2.3). As shown in the vehicle sample (DMSO), the characteristic fatty acid and mycolic acid distribution are normal with the lower molecular weight FAMES at the top of the TLC plate and the three spots below this representing the alpha (α), alpha prime (α') and epoxy mycolates which are known to be found in *M. smegmatis*. Incubation with **109** at 50 μ M resulted in an increase in the presence of FAMES which indicates an accumulation of the shorter chain FAS I end-products. This is accompanied by truncation of the α -mycolates into two smaller fragments and a decrease in the quantity of α' - and epoxy-mycolates which is suggestive of the inhibition of the FAS II complex. Additionally, the presence of another spot at the midway point of the TLC plate indicates the accumulation of unsaturated end products of the FAS II pathway. These changes are indicative of classical mycolic acid inhibition and show similar results to studies conducted with known anti-tubercular agent isoniazid which inhibits mycolic acid biosynthesis through InhA inhibition.¹⁰ These changes are also apparent after incubation with **105**, however, to a lesser extent. This comes as a surprise given the higher MIC against *M. smegmatis*. However, given the high DNA binding affinity of **105**, InhA inhibition may be its secondary mode of action.

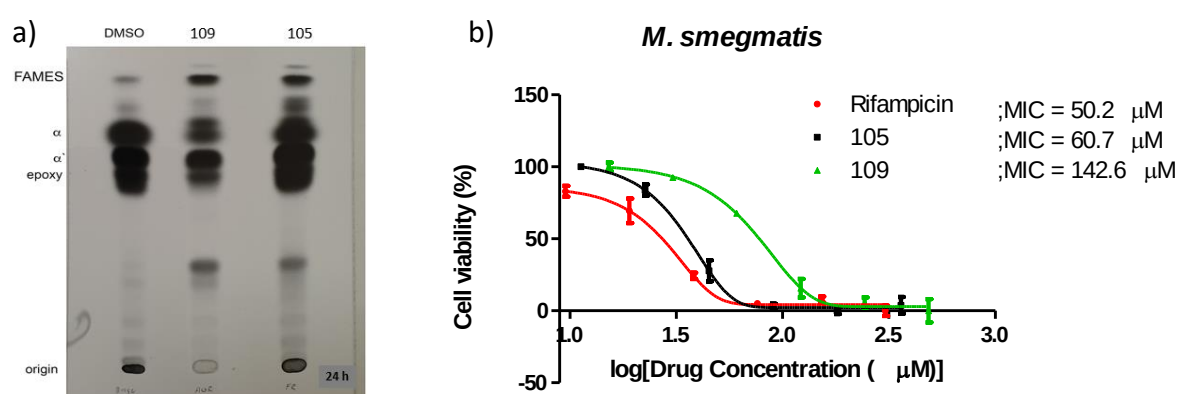


Figure 4.1.2.3. (a) FAMES and MAMEs TLC plate of ^{14} C-acetate labeled lipid extracts. Assay conducted using *M. smegmatis* at 50 μ M of **105** and **109**. (b) REMA analysis of **105**, **109** and rifampicin against *M. smegmatis* showing their respective MICs.

4.1.3. Biochemical evaluation of piperidine/piperazine derivatives against DprE1

Decaprenylphosphoryl-b-D-ribose-20-epimerase (DprE) is a heterodimeric enzyme comprising the proteins DprE1 and DprE2. As shown in Figure 4.1.3.1, DprE1 oxidises decaprenylphosphoryl-D-ribose (DPR) to decaprenylphosphoryl-2-ketoribose (DPX) which is then reduced by DprE2 to form decaprenylphosphoryl arabinose (DPA), a precursor of the mycobacterial cell wall component arabinogalactan. DprE1 has become a highly vulnerable and chemically validated target for TB drug discovery.¹¹ In 2009, benzothiazinones were identified as covalent inhibitors of DprE1 showing low nanomolar inhibitory activity against *M. tuberculosis* in *in vitro* studies. One of the most active benzothiazinones, **BTZ043**, was shown to exhibit an MIC of 2.3 nM and an IC₅₀ of 4.25 nM against DprE1. Structural modification of **BTZ043** led to the production of piperazine-containing derivatives such as **PBTZ169** which showed better anti-tubercular activity (MIC = 0.66 nM) and is under clinical evaluation.² The 8-nitro group of the benzothiazinone scaffold has been shown to be crucial for the mechanism of action, which involves the formation of a semimercaptal bond with Cys387 in the active site of DprE1. However, in recent studies conducted by Makarov *et al.* a series of 8-pyrrole-benzothiazinones, non-nitro-benzothiazinones that retain significant DprE1 inhibition and antimycobacterial activity highlighting the possibility of noncovalent inhibition of DprE1.¹²

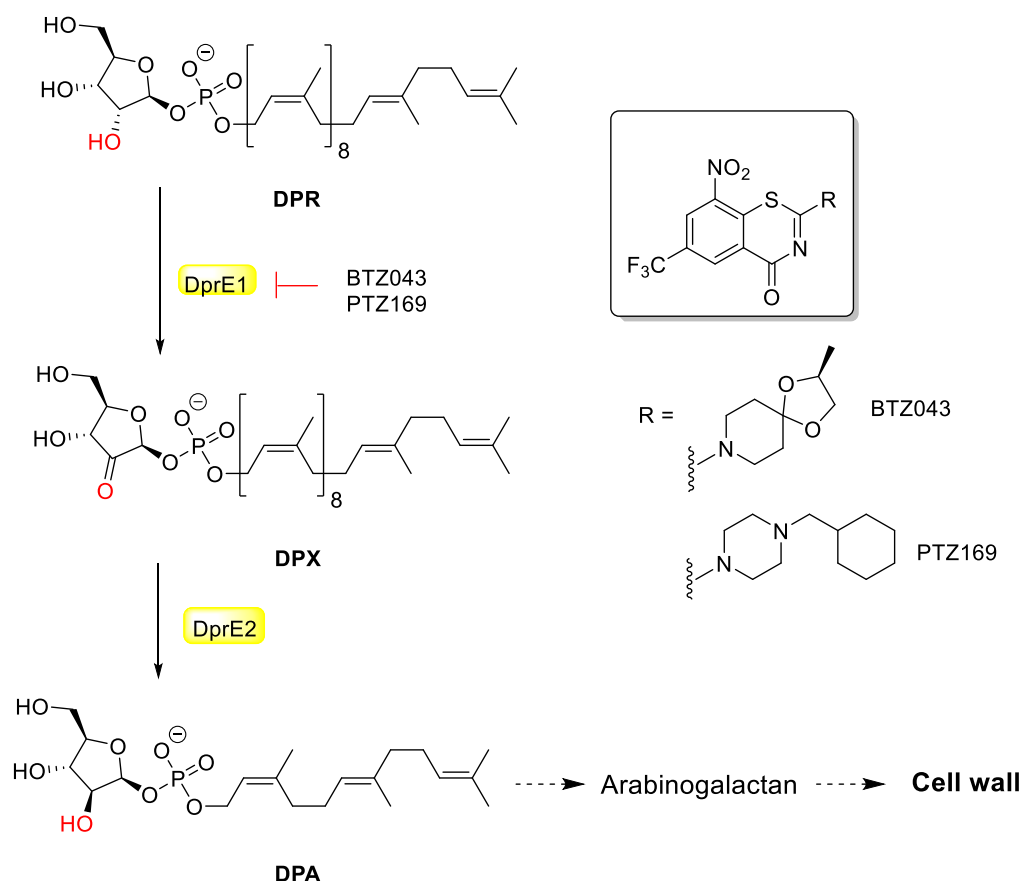


Figure 4.1.3.1. Role of DprE1/E2 in mycobacterial cell wall biosynthesis with structures of known DprE1 inhibitors **BTZ043** and **PTZ169**.

Analysis of the activity of the PZ/PD derivatives under investigation against DprE1 showed that none of the compounds exhibited significant inhibition with IC_{50} values of greater than 1000 μ M (Figure 4.1.3.2 (a)). A fluorescence-based enzyme assay for Mtb-DprE1 was employed that involves oxidation of farnesylphosphoryl-β-D-ribose (FPR) to farnesylphosphoryl-β-D-2'-keto-erythro-penta-furanose (FPX) as resazurin is reduced to the highly fluorescent resorufin. As shown in Figure 4.1.3.2 (b), no significant reduction of DprE1 was observed after 6 minutes of incubation with **105**. Only at concentrations greater than 1000 μ M was any inhibition observed, with significant inhibition occurring at 2000 μ M. These results in addition to the large IC_{50} difference between these compounds and the known DprE1 inhibitor **BTZ043** (0.00425 μ M), indicates that DprE1 inhibition is not their mechanism of action.

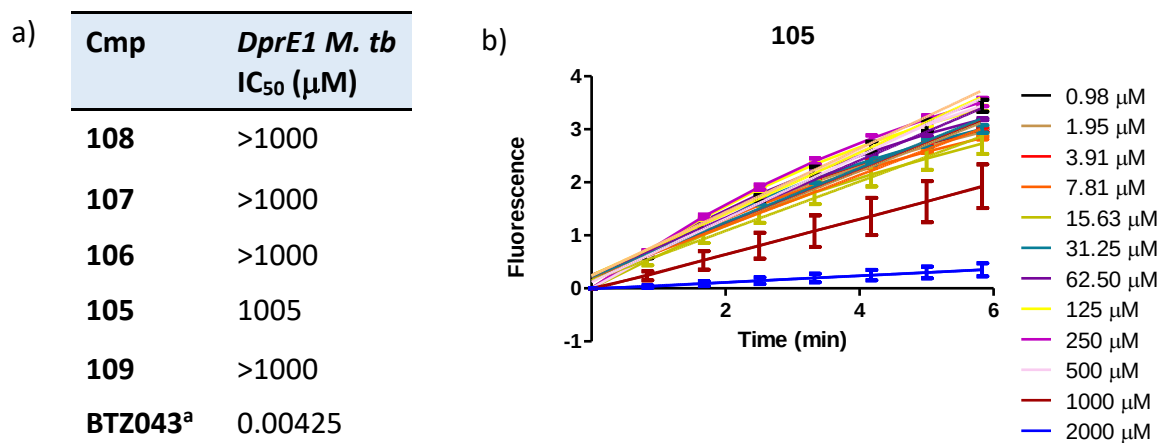


Figure 4.1.3.2. (a) IC₅₀ values measured for **108**, **107**, **106**, **105** and **109** and **BTZ043** against wild-type *M. tuberculosis* DprE1. ^a IC₅₀ of BTZ043 obtained from Ref. 2. (b) Progress curve analysis of inhibition of Mtb-DprE1 by **105**.

4.1.4. Time-kill assay

To determine whether the compounds under investigation are killing the bacteria (bactericidal) or inhibiting their growth (bacteriostatic), a time-kill experiment was conducted. This involved incubation of the respective compound at twice the MIC over 7 days where optical density (OD₆₀₀) values were obtained at specific time points as a measure of the amount of cells present. For this experiment, *M. smegmatis* was used with a starting optical density of 0.1 (CFU/mL 3x10⁷ cells). As the sensitivity profile of *M. smegmatis* towards rifampicin has been shown to be identical to that of multi-drug resistant *M. tuberculosis*, this experiment along with the previously conducted REMA analysis serves to potentially identify effective leads for the treatment of MDR-TB. Bactericidal activity is characterised by a significant (≥ 1 fold) decrease in the bacterial concentration from the initial bacterial inoculum. On the other hand, bacteriostatic activity is characterised by the inhibition of bacterial growth which can be identified by a retention in bacterial concentration from the initial inoculum or slightly decreased growth with respect to the control. Figure 4.1.4.1 shows the time-kill kinetics upon incubation with **105**, **106**, **109** and rifampicin in terms of OD₆₀₀. Analysing the effect of **106** at twice its MIC, a weak bacteriostatic effect is observed with inhibition of growth only occurring after 2 days of incubation. This is in line with the poor MIC value obtained from REMA analysis. The asymmetric derivative **109** shows an almost identical

profile to that of rifampicin. Due to a negligible change in the growth of the bacteria from the initial inoculum over the 7 days, both of these compounds can be characterised as being strongly bacteriostatic. Even though rifampicin is known to be bactericidal against wild-type *M. tuberculosis* strains, the natural resistance that has been reported to occur in *M. smegmatis* is replicative in the change of activity observed in this experiment. **105** was shown to display a sharp decrease in OD₆₀₀ which was observed after just 2 days of incubation. Additionally, a 1 fold decrease in OD₆₀₀ was observed after 7 days of incubation indicating a bactericidal mechanism.

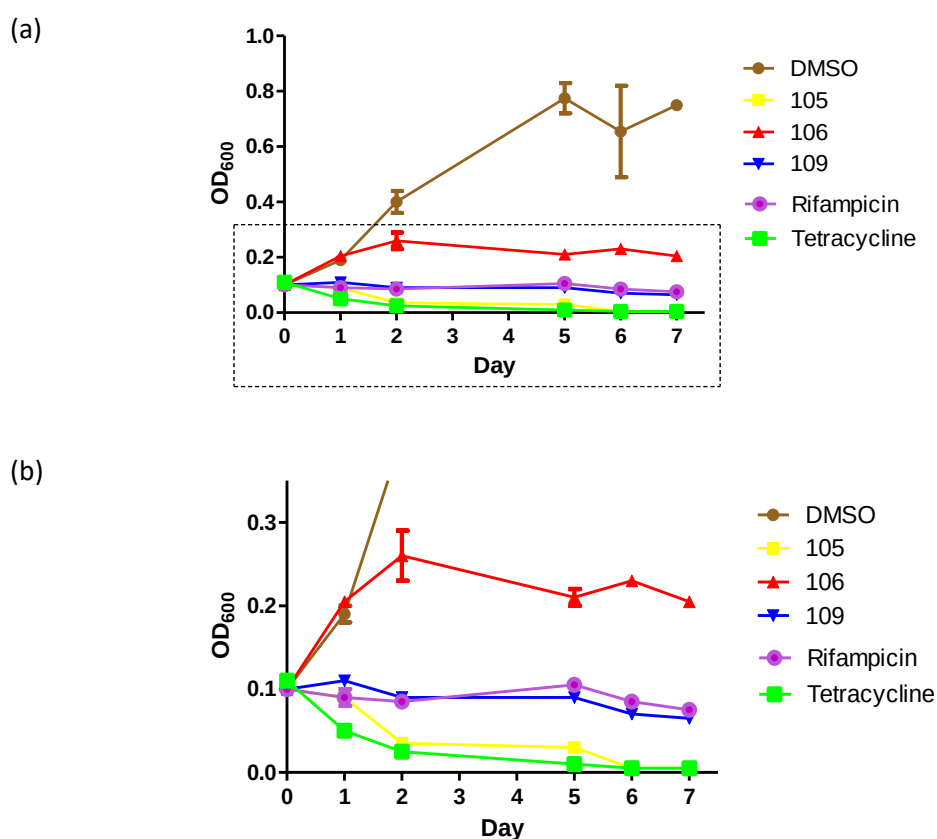
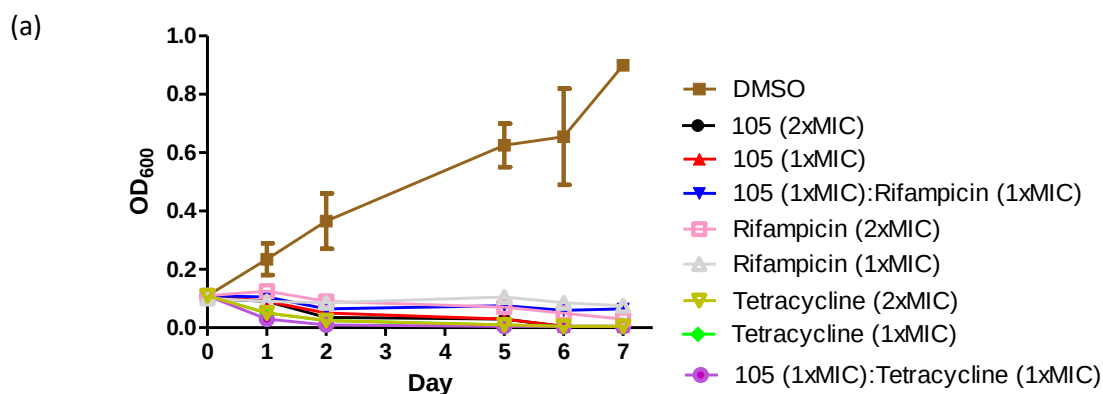


Figure 4.1.4.1. (a) Time-kill kinetics assay of **105**, **106**, **109**, rifampicin and tetracycline (tested in triplicates at 2× MIC) against *M. smegmatis* over a 7 day incubation period at 37 °C. DMSO (solvent for the compounds) served as a negative control; (b) shows an expansion of (a) in the region of interest indicated by the dotted box.

As stated earlier, treatment of tuberculosis requires a long dosage regime consisting of a cocktail of anti-mycobacterial agents. To assess the potential of **105** to be used in combination, incubation with rifampicin and tetracycline was considered (Figure 4.1.4.2). Given the different modes of action of all three compounds a synergistic effect was expected. However, upon analysis of the **105** and rifampicin combination at 1xMIC no synergistic effect was observed. Moreover, an antagonist effect with respect to the activity of **105** at 1x and 2x MIC was obtained. Although this came as a surprise, recent literature has shown that rifampicin induces mycomembrane remodeling in *M. smegmatis* which reduces the ability of certain antimicrobial cationic peptides to disrupt or penetrate the membrane.¹³ This explains the weakened activity of **105** in the presence of rifampicin. On the other hand, incubation with tetracycline at 1xMIC was shown to induce a synergistic effect, producing greater killing kinetics than **105** and tetracycline alone. It is also important to note that a combination of **105** and tetracycline at 1xMIC each gives greater anti-mycobacterial activity than 2xMIC of **105**, rifampicin and tetracycline alone. These positive results suggest that this family of compounds deserve to be investigated further and used as a scaffold for the design of novel anti-tubercular agents.



(b)

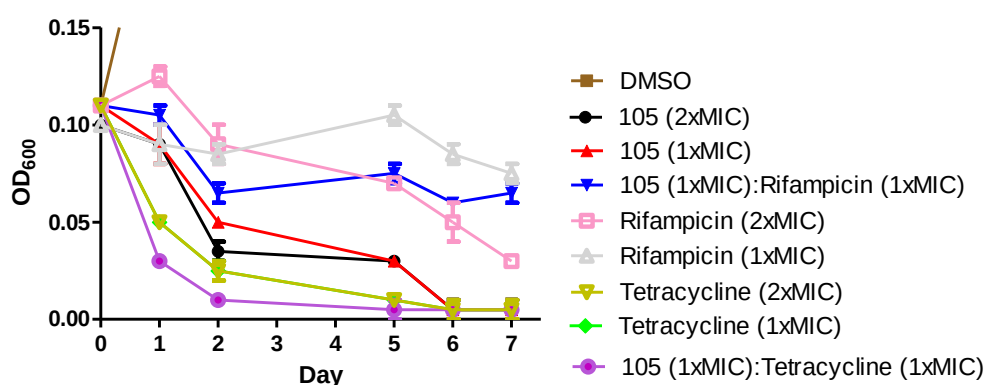
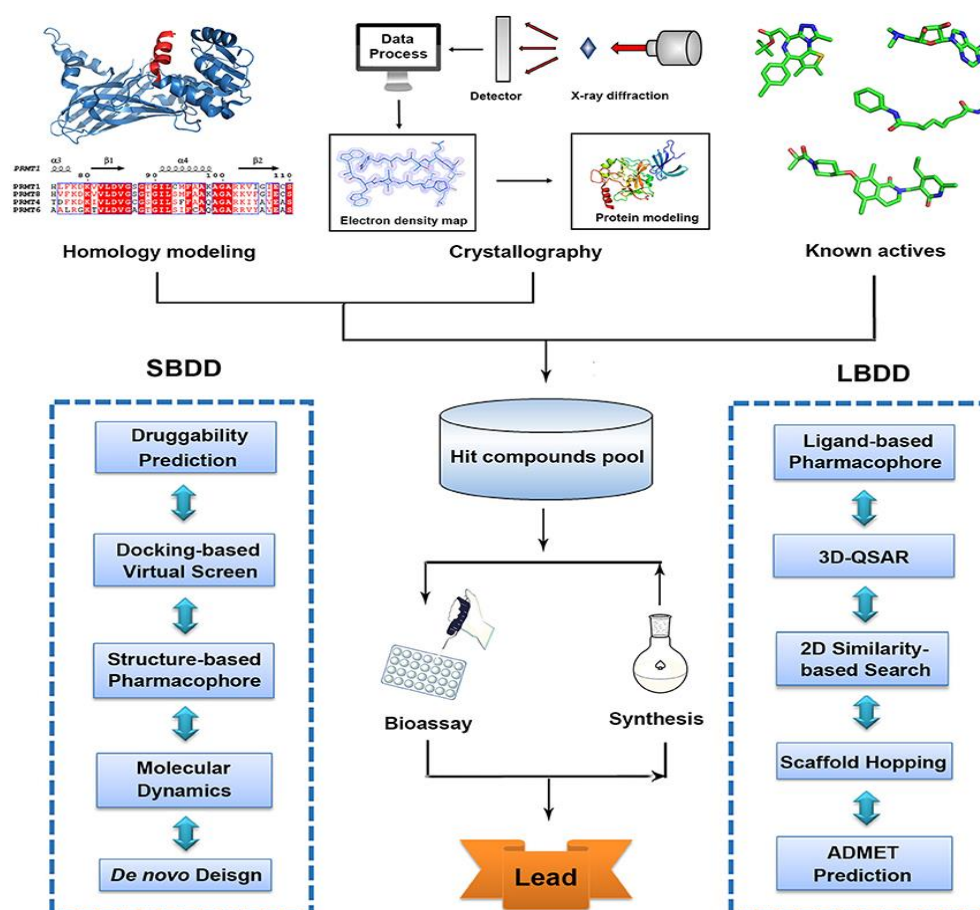


Figure 4.1.4.2. (a) Time-kill kinetics assay of **105** alone and in combination with rifampicin and tetracycline (single treatments tested at 1x and 2x MIC and combination treatment tested at 1x MIC of each drug) against *M. smegmatis* over a 7 day incubation period at 37 °C. DMSO (solvent for the compounds) served as a negative control; (b) shows an expansion of (a) in the region of interest as indicated by the dotted box.

4.2. Computer-aided investigations

Based on the results presented in the previous section, the mechanism of action of the PZ/PD derivatives under investigation is suggested to be the inhibition of mycolic acid synthesis possibly due to InhA blocking. To explore this further, computer-aided methods were employed consisting of ligand-based pharmacophore screening and structure-based molecular docking. As shown in Figure 4.2.1, a traditional drug discovery workflow consists of either a ligand-based approach (LBDD) or structure-based approach (SBDD).¹⁴ Using both of these methodologies, the most active compounds of our in-house library of compounds were screened against a database of FDA approved anti-tubercular agents with known targets and then docked into reference crystal structures of potential targets. Outcomes from these studies serve to identify the structural components responsible for their therapeutic activity and to recognise an optimal scaffold for lead optimisation.



4.2.1. Ligand-based pharmacophore screen

Within the last decade, the relationship between the structure of a molecule and its activity has been well documented. The specific structural components of molecules responsible for their biological or pharmacological activity is known as its pharmacophore. Hence, one of the purposes of ligand-based pharmacophore screening (LBPS) is to find the pharmacophore of compounds with unknown targets and map it against drug compounds with known targets to identify their potential mechanism of action (Figure 4.2.1.1). Given the need for novel anti-tubercular agents, we have used LBPS to identify bioactive molecules with different basic scaffolds but similar essential functional groups. This concept is also referred to as scaffold hopping.¹⁵

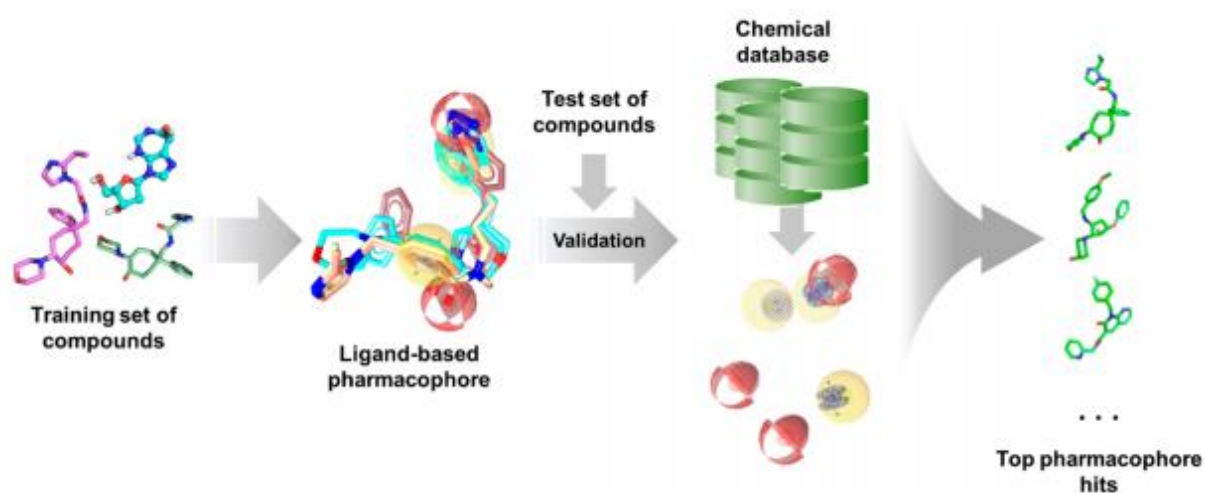


Figure 4.2.1.1. Typical ligand-based pharmacophore generation and screening workflow.¹⁵

To conduct a ligand-based screen for target prediction, chemical databases containing FDA approved anti-tubercular agents with known targets were extracted from the ChEMBL and ZINC databases.^{16, 17} This afforded a dataset of 32 compounds with diverse mechanisms of action. From our most active compounds, **105**, **106** and **109**, several 3D pharmacophores hypotheses were developed using the Schrödinger Maestro molecular modelling software which does so by using a common pharmacophore perception algorithm based on conformational alignment of pharmacophore features (HB acceptor, HB donor, hydrophobic, negative ionic, positive ionic, aromatic ring).¹⁸ The best fit 3D pharmacophore hypothesis

(**HYP1**) was chosen based on the highest “phase hypo score” and “survival score” (Figure 4.2.1.2). The phase hypo score is a scoring function that rank orders the hypotheses based on ligand alignment whereas the survival score also takes into consideration volume overlap, number of matches and relative energy of the reference ligand.

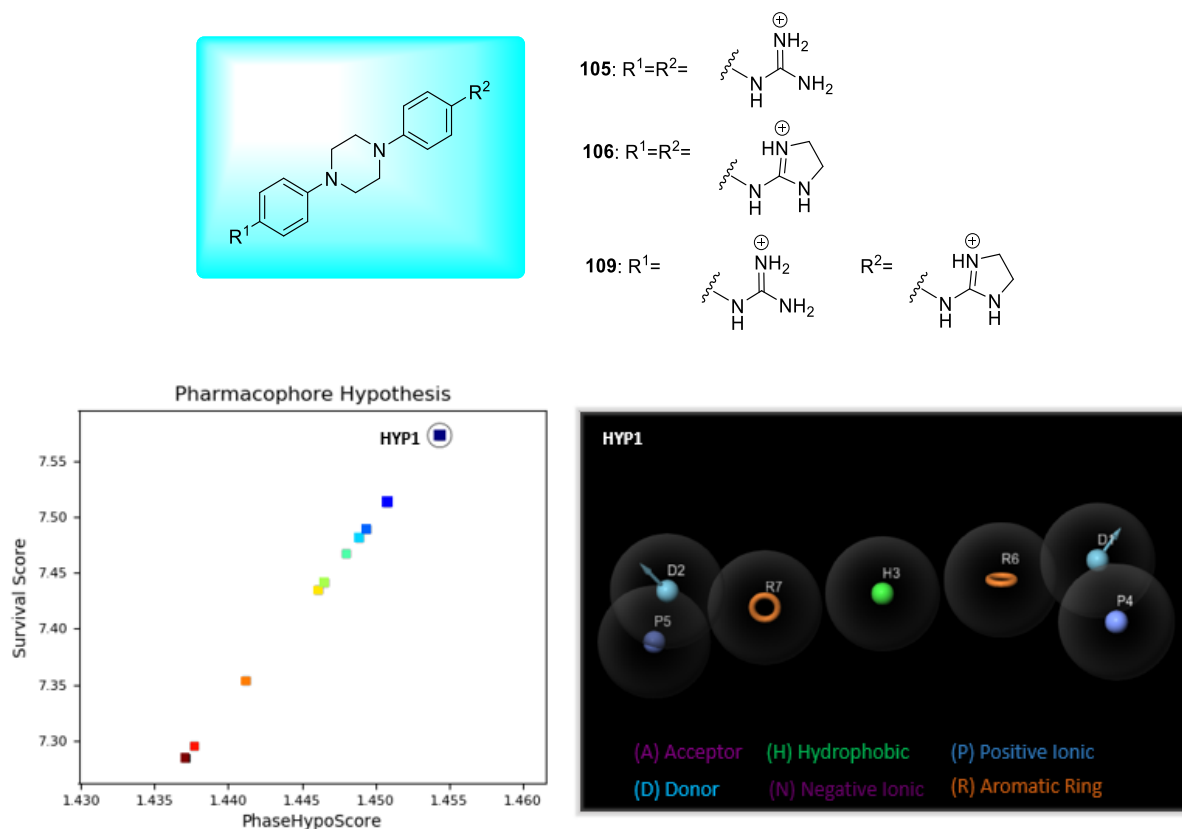


Figure 4.2.1.2. 3D pharmacophore hypothesis.

The chosen pharmacophore (**HYP1**) was screened against the dataset of FDA approved anti-tubercular agents to identify any positive hits. From the 32 compounds within the chemical database, 10 hits were identified which are summarised in Table 4.2.1.1. Each hit is ranked in terms of its “phase screen score” which is a scoring function that assesses both the quantity and quality of ligand feature matching. Interestingly, the top two highest scoring drug compounds, delamanid and ethambutol, are known to elicit their anti-tubercular activity through the inhibition of cell wall biosynthesis. Both delamanid and ethionamide are known to target the FAS II pathway through inhibition of the Enoyl ACP Reductase (InhA) which corroborates the conclusions drawn from our phenotypic whole-cell assays shown in Section 4.1.

Table 4.2.1.1. Positive hits obtained from ligand-based pharmacophore screen against 32 FDA approved anti-tubercular agents

Entry	Drug	Phase Screen Score	MOA	Target
1	Ethionamide	1.369	Inhibition of cell wall synthesis	InhA
2	Delamanid	1.307	Inhibition of cell wall synthesis	InhA
3	Rifamycin B	1.283	Inhibition of protein synthesis	RNA polymerase
4	Bedaquiline	1.282	Interference with Energy metabolism	ATP synthase
5	Clofazimine	1.263	ROS generation	2 NADH-quinone oxidoreductase
6	Rifampicin	1.262	Inhibition of protein synthesis	RNA polymerase
7	Rifabutin	1.26	Inhibition of protein synthesis	RNA polymerase
8	Rifapentine	1.239	Inhibition of protein synthesis	RNA polymerase
9	PASA	1.078	Inhibition of folate metabolism	DHFR

Some interesting findings can be drawn from the structural analysis of ethionamide and delamanid against the constructed 3D pharmacophore. As shown in Figure 4.2.1.3 below, significant similarity was obtained for ethionamide where the aryl-thioamide overlaps with the cationic terminal of the pharmacophore. Interestingly, only the centre of the pharmacophore constructed from the di-aryl piperazine motif was shown to overlap with the di-aryl core of delamanid. However, it is important to note that the active drug produced upon activation of the delamanid prodrug bears a cyclic isouronium which is structurally and chemically similar to the guanidine found in our compounds.¹⁹ These findings indicate that the 4-phenyl-guanidino and di-aryl-piperazine core found in our in-house compounds play an important role in their anti-mycobacterial activity. In addition, it is worth noting that delamanid bears a polar isouronium head group and a lipophilic aromatic tail which suggests that an amphiphilic motif is more beneficial for InhA binding.

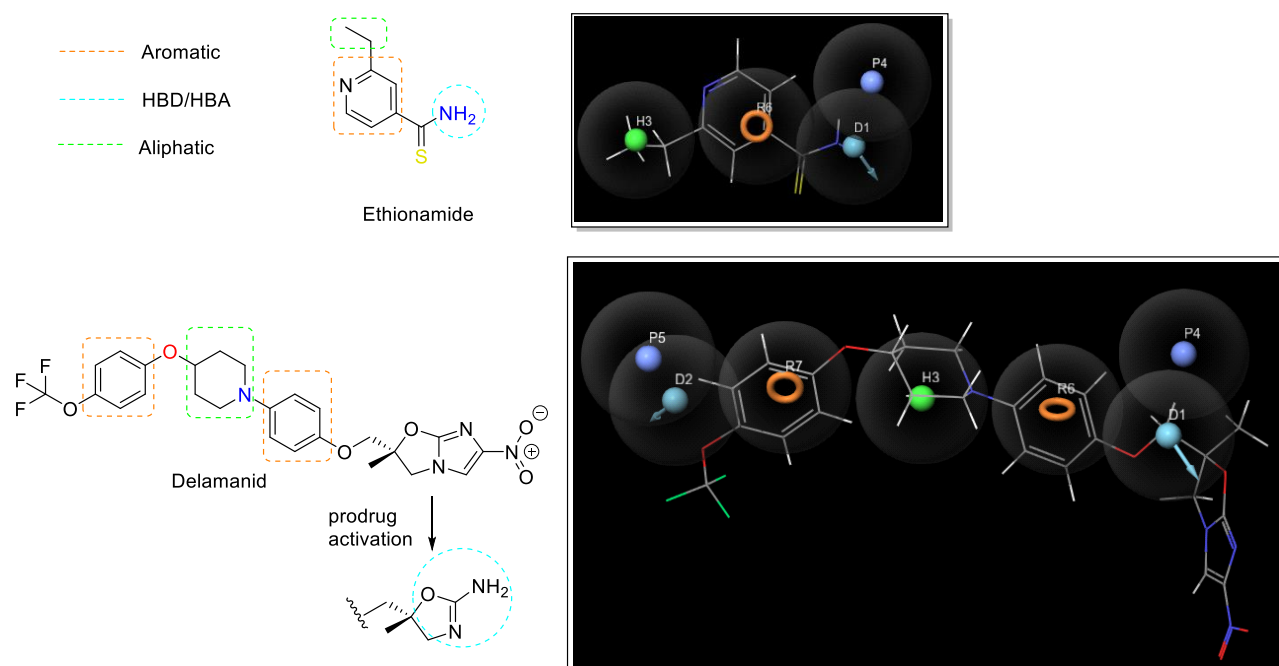


Figure 4.2.1.3. Overlap of the constructed 3D-pharmacophore with hit compounds ethionamide and delamanid.

In summary, the findings from the LBPS indicate that our in-house compounds share key functionalities with compounds that inhibit cell wall synthesis through InhA inhibition. However, to identify the key binding interactions at play further analysis was conducted through SBDD molecular docking studies.

4.2.2. Structure-based molecular docking in InhA

As stated previously, InhA catalyses the NADH-dependent reduction of enoyl-ACP at the later stages of the FAS II pathway leading to the production of mycolic acids which are essential for the biosynthesis of the mycobacterial cell wall. The reaction is thought to proceed via hydride transfer from the C4 of NADH to the β -carbon of the enoyl-ACP, followed by enolate formation and stereospecific protonation to yield the desired saturated acyl-ACP (Figure 4.2.2.1). The amino acid residue, TYR158, has been suggested to play a key role in the stabilisation of the transition state via hydrogen bonding to the carbonyl of the substrate.²⁰

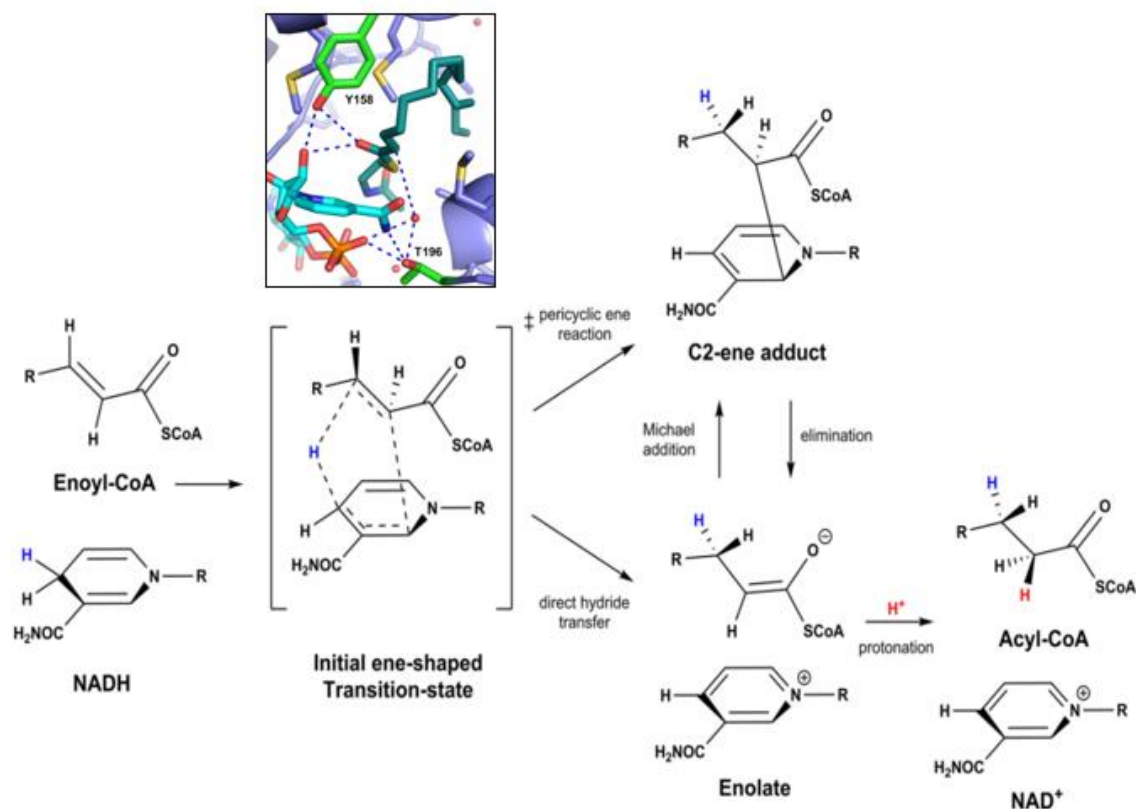


Figure 4.2.2.1. Reaction pathway for the NADH mediated reduction of acyl-substrate analogue (Enoyl CoA) by InhA.²⁰

Various strategies have been attempted to target InhA and inhibit this essential cellular process. The early success of isoniazid and ethionamide as inhibitors of InhA relied on NAD adduct formation which competes kinetically with the cofactor NADH. Other attempted methods include the design of small molecule mimics of the isoniazid-NAD complex. However, the large molecular weight of these mimics limits the opportunity for oral administration. More recently, substrate competitive inhibitors have emerged as effective InhA inhibitors with good *in vitro* potency; some examples include triclosan, methyl-thiazoles (MTH) and piperazine derivatives (GENZ-10850) (Figure 4.2.2.2).^{7, 21, 22}

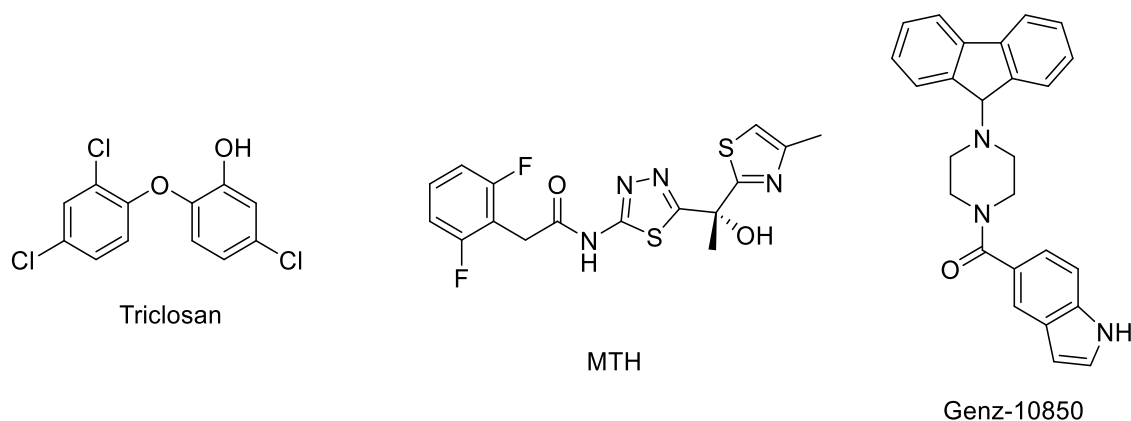


Figure 4.2.2.2. Substrate competitive InhA inhibitors triclosan, MTH and Genz-10850.

Taking the reference crystal structure of MTH bound to *M. tuberculosis* InhA (Figure 4.2.2.3, PDB: 4BQR)²³, compounds **105**, **109** and **106** were docked into the substrate-binding site to identify the potential key interactions responsible for enzyme inhibition. Firstly, taking a look at the original crystal structure and the key interactions MTH forms in the active site, the methyl-thiazole group engages with the nicotinamide and ribose groups of the NAD cofactor which is common amongst substrate competitive inhibitors. The N atom of the thiazole ring acts as a HB acceptor to the hydroxyl group of the NADH ribose. The close proximity of the thiazole to the nicotinamide of the NADH results in the formation of π - π stacking interactions. The methyl group of the methyl thiazole system is shown to force the TYR158 to maintain an outward facing conformation by preventing the residue from adopting the active inward facing conformation. The tetrahedral carbon linking the thiazole and 1,3,4-thiadiazole rings gives the molecule the flexibility to orientate away from the rest of the NADH, thus, avoiding steric clashes. Additionally, a HB donor-acceptor pair is formed between one N of the 1,3,4-thiadiazole ring and the backbone of MET98, which anchors the molecule in place allowing the difluoro-phenyl ring to access a hydrophobic pocket.

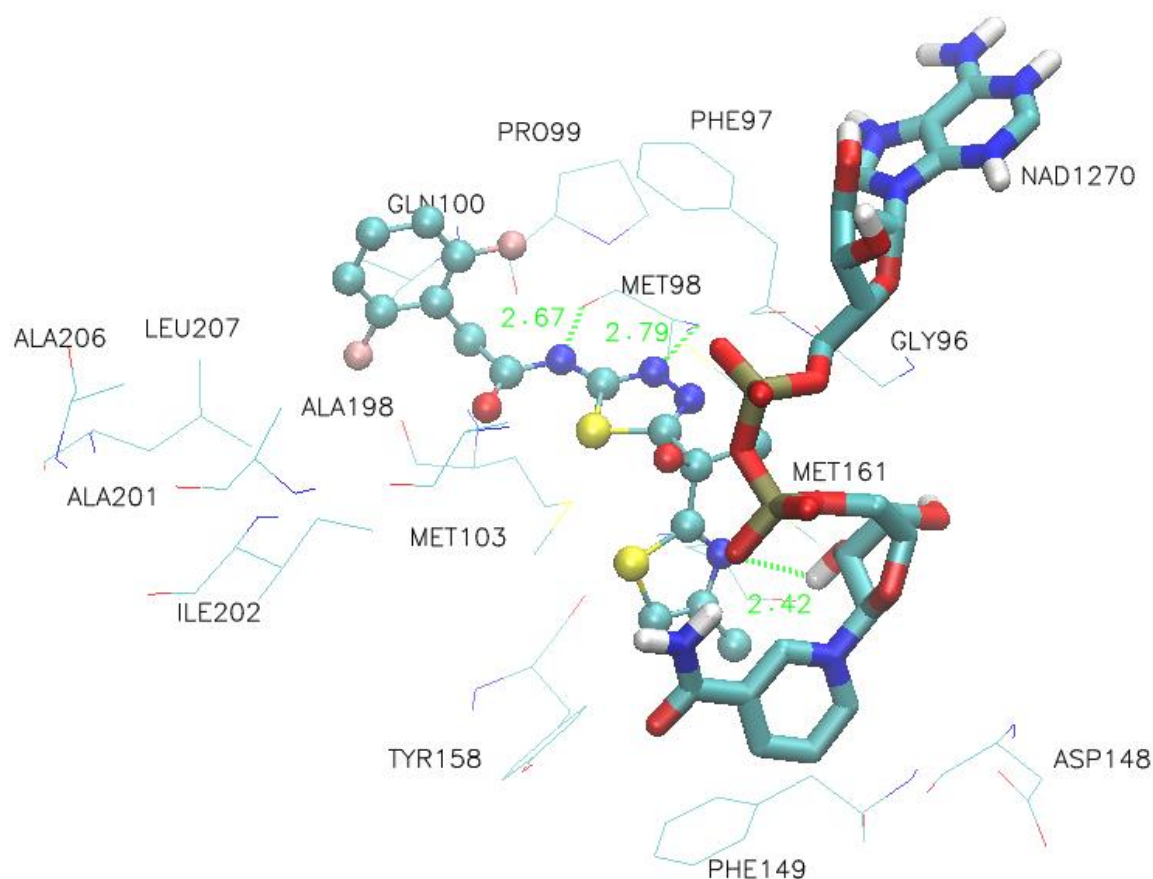


Figure 4.2.2.3. Crystal structure of MTH (CPK model) bound to *M. tuberculosis* InhA (line model) showing NADH co-factor (licorice style) (PDB: 4BQR).

Removal of MTH and docking of **105** in a flexible-rigid approach revealed that similar binding interactions are formed with the active site residues. In the inner pocket of the binding site, the guanidinium NH forms HB interactions with the hydroxyl group of the NADH ribose. Additionally, π -cation interactions are formed between the guanidinium and the nicotinamide ring of NADH as well as the phenyl ring of PHE97. This dual T-shape and stacked π -cation interaction has not been reported among InhA inhibitors and may be unique to our class of compounds. The conformation of the piperazine ring orientates the rest of the molecule away from the NADH cofactor preventing steric clash which is similar to the effect of the tetrahedral carbon of MTH. Consequently, the other terminal guanidinium orients towards the hydrophobic pocket; although this is not ideal, an anchoring HB is formed between the guanidinium NH and the backbone carbonyl of ALA198 within this pocket. The proximity of

the adjacent NHs of the guanidinium to the opening of the binding pocket creates the possibility of HB interactions with water molecules.

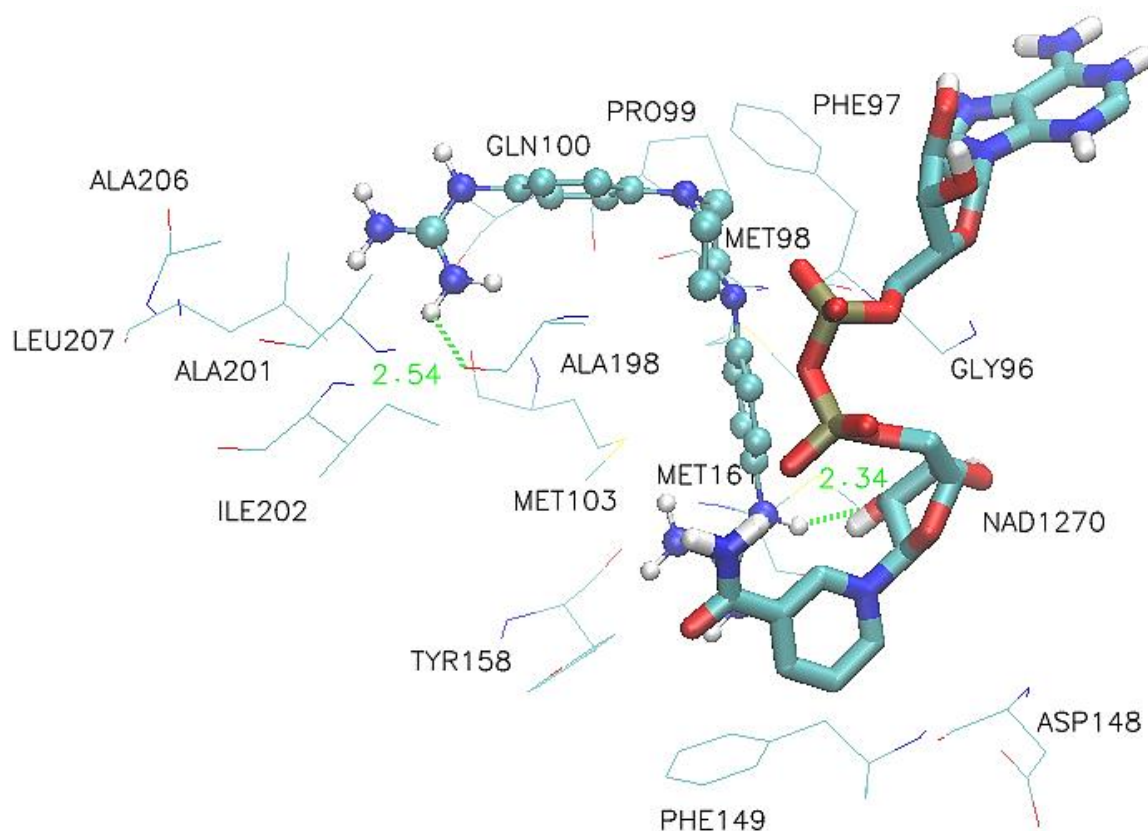


Figure 4.2.2.4. Docking pose of **105** bound to the active site of InhA.

Comparing the binding pose of **105** to that of **106**, slight changes were observed in the interactions formed (Figure 4.2.2.5). Firstly, replacement of the guanidiniums for 2-aminoimidazoline moieties creates a staggering of the π -cation interaction between the nicotinamide moiety and the terminal 2-aminoimidazoline within the inner pocket to avoid steric clashes with PHE149. As a consequence of this, the T-shaped π -cation interaction with PHE149 is lost. Additionally, a weaker HB is formed (2.67 Å) between the NH connecting the imidazoline ring and the hydroxyl group of the NADH ribose. However, the positioning of the methylene carbons of the imidazoline ring allows for van der Waals interactions with MET161, TYR158 and PHE149. Similar to **105**, the piperazine ring of **106** forces the other terminal cation to orientate towards the hydrophobic pocket. The presence of the 2-aminoimidazoline moiety as opposed to the guanidine moiety in this position allows for van der Waals

interactions with ALA201, ALA206, ILE202 and LEU207 of the hydrophobic pocket. The lower *in vitro* activity of **106** over **105**, may indicate the importance of the intermolecular interactions formed between the guanidine and the residues of the inner pocket of the binding site. This is also observed in the ligand based pharmacophore screen, where a higher structural pharmacophore similarity with current InhA inhibitors was observed for **105**.

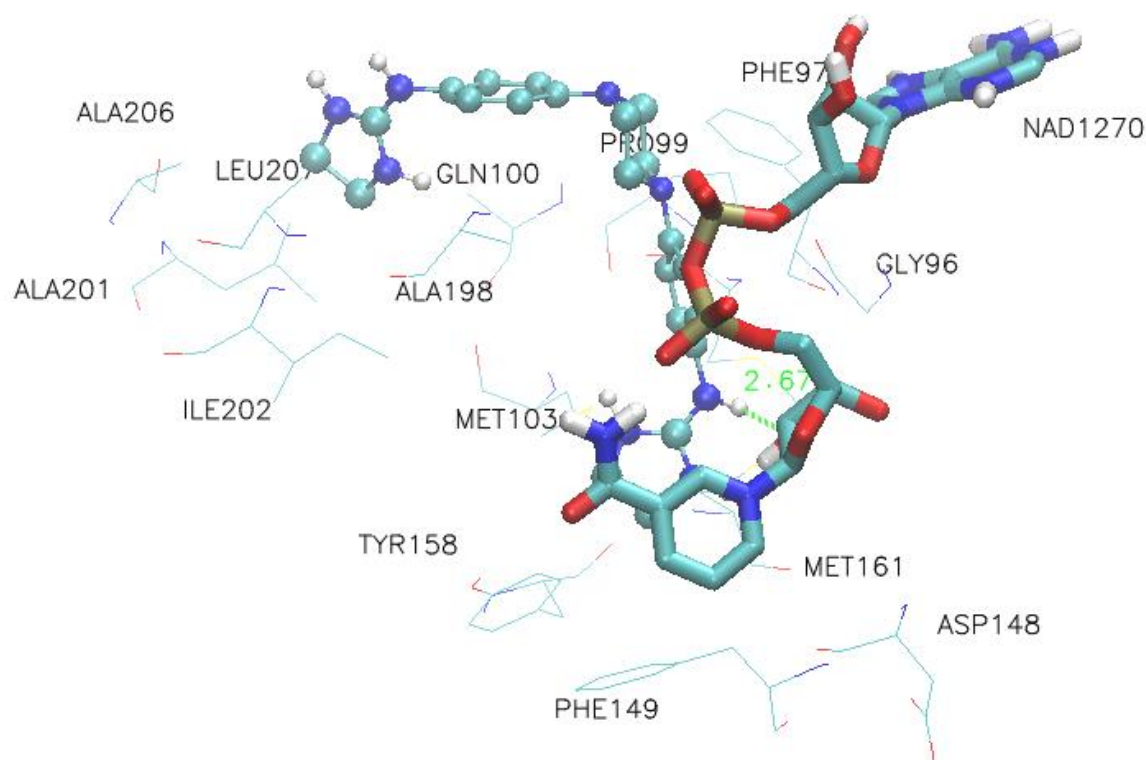


Figure 4.2.2.5. Docking pose of **106** bound to the active site of InhA.

From both binding poses, it can be deduced that a more amphiphilic motif would be more beneficial for binding where a small polar functional group occupies the inner pocket while the more lipophilic groups interact with the hydrophobic pocket. This is exactly what we see with **109**, where the terminal guanidinium forms a HB and stacked π -cation interaction with NADH and a T-shaped π -cation interaction with PHE149. The 2-aminoimidazolinium moiety then occupies the hydrophobic pocket forming additional van der Waals interactions. The favourability of this motif is also reflected in the higher Gscore value of **109** ($-7.5 \text{ kcal mol}^{-1}$) versus **105** ($-7.0 \text{ kcal mol}^{-1}$) and **106** ($-6.8 \text{ kcal mol}^{-1}$). Moreover, this correlates with the greater mycolic acid inhibition of **109** observed in our FAMES and MAMES TLC analysis.

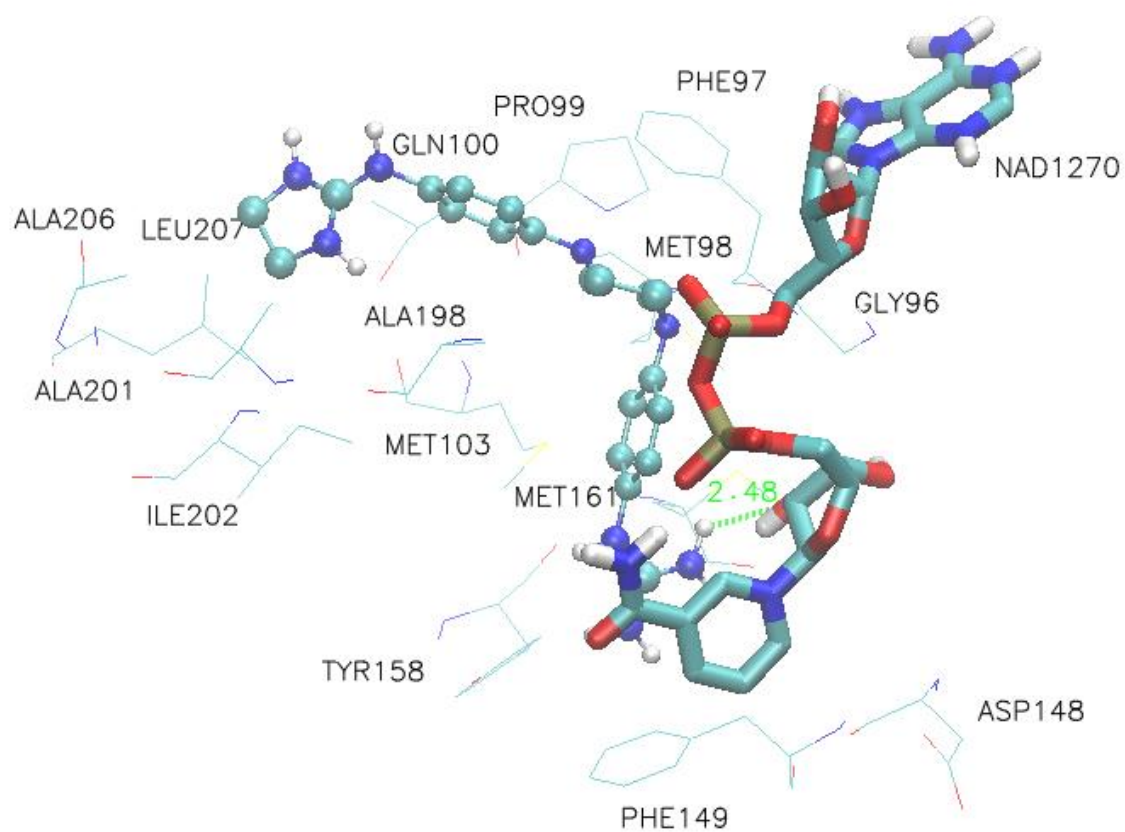


Figure 4.2.2.6. Docking pose of **109** bound to the active site of InhA.

4.3. Whole Genome Sequencing

Whole Genome Sequencing (WGS) technologies are currently used as a tool to guide components of TB control including diagnosis, treatment, surveillance and source investigation. More recently, WGS has been employed in preclinical drug development settings where information on the resistance mechanism(s) and mechanism(s) of actions of novel anti-tubercular agents can be attained. The beauty of this technology is its ability to identify key mutations, whether they be Single Nucleotide Polymorphisms (SNPs) or Insertions/deletions (INDELs), that can inform us how a particular drug is acting based on the response of the microbe towards the selective pressure. By target confirmation and resistance mechanism elucidation through variant analysis, we have the ability to design more potent anti-infective agents. Various attempts within the literature to purposely generate mycobacterial mutants have involved single-shot approaches on solid media and continual passaging in liquid media.^{24,25,26,27} Using *M. smegmatis* and our most active compound, **105**, both approaches were attempted in collaboration with Prof. Stephen Gordon from the School of Veterinary Medicine at University College Dublin. Drug susceptibility testing with the use of REMA analysis was conducted to monitor the development of resistance through changes in the MIC. DNA extraction of the resulting mutants followed by DNA sequencing allowed for the characterisation of the causative mutations. Following these findings, gene expression analysis and cross-resistance evaluation gave further insight into mechanisms induced by **105** overexposure.

4.3.1 Resistance generation

4.3.1.1. Liquid media continual passaging approach

Most efforts to generate drug-resistant mutants in liquid media have involved the incubation of a drug compound with a wild-type mycobacterial strain through a continual passaging approach. This approach involves the extraction of surviving cells from the culture media and re-exposure to the compound under investigation at higher concentrations (usually multiples of its MIC). Repetition of this procedure has been shown to result in the selection of drug-

resistant mutations in *M. smegmatis*.²⁴ Using this approach, we attempted to raise mutants in *M. bovis* BCG Denmark through continual passaging in 7H9 medium (5 mL culture volume) against **105**. However, upon re-growth of the surviving cells after the first passage at 1xMIC, subsequent exposure at 2xMIC resulted in complete cell death. For this reason, we conducted a continual exposure experiment at a constant concentration of 1xMIC through 3 passages where each incubation period lasted 7 days. Unfortunately, we observed that each passage produced the same MIC once analysed by REMA (Figure 4.3.1.1.1). Additionally, DNA extraction and sequencing of isolates from each passage showed only background SNPs that were also found in the wild-type and no causative mutations as a result of **105** exposure (Table 4.3.1.1.1).

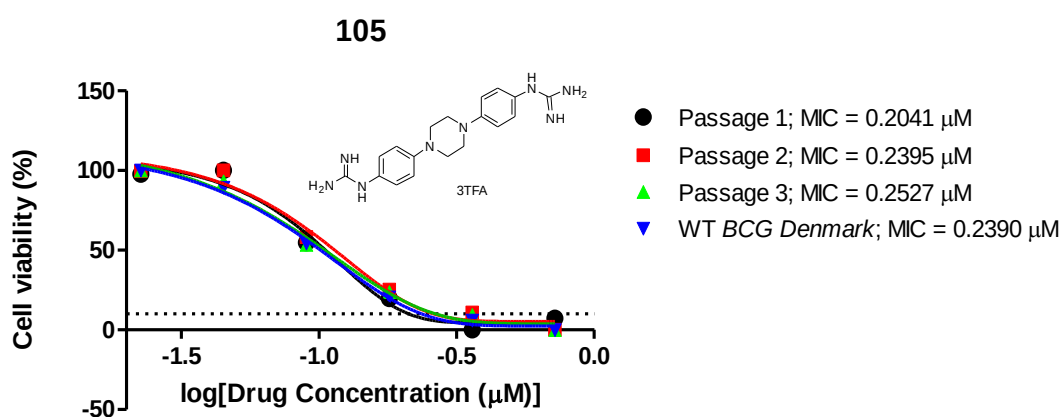


Figure 4.3.1.1.1. REMA drug susceptibility curves of *M. bovis* BCG Denmark isolates from continual passaging of **105** in 7H9 liquid media.

Table 4.3.1.1.1. Mutations observed from the DNA sequencing of *M. bovis* BCG Denmark isolates from passages 1, 2 and 3.

Position	SNP	Gene	Protein	Mutation	Type
593769	C>T	FCU26_RS02705	SDR oxidoreductase	323Q>323*	Stop gained
2500583	T>C	FCU26_RS12145	sulfotransferase	260H>260R	Missense
3586995	C>G	FCU26_RS17645	carbonic anhydrase	24S>24*	Stop gained
3745609	G>T	FCU26_RS18410	PPE family protein	434S>434Y	Missense

Base calling of DNA extracts was conducted using the Illumina platform; Variant calling was conducted using the SNiPgenie pipeline; Variants displayed are present in wild-type *M. bovis* BCG Denmark and are deemed background mutations.

The lack of selection of resistance mutations may be a result of the slow-growing nature of *M. bovis* BCG Denmark.²⁸ Due to these properties, the chances of mutations occurring due to the selective pressure is low and may require a large number of passages. As this experiment is done at relatively large culture volumes, the amount of drug compound needed to meet the adequate concentration would be very high if multiple passages were required. For this reason, the faster growing *M. smegmatis* was employed using a novel small scale high throughput approach developed by us that involves continual incubation of **105** with *M. smegmatis* in LB broth using a 96-well plate. This method allows for a greater number of test replicates, lower culture volume (200 μ L/well), higher throughput passaging and requires less drug compound. Taking advantage of the higher capacity, each passage consisted of a serial dilution of **105** with a constant starting concentration of *M. smegmatis* (O.D = 0.1, CFU/mL $\sim$ 3×10^7 cells). By coupling this experiment with REMA analysis, the cells at the MIC point were inoculated and re-grown for the next exposure. With an increase in drug concentration range after each passage, the development of drug resistance was observed by a gradual increase in MIC (Figure 4.3.1.1.2).

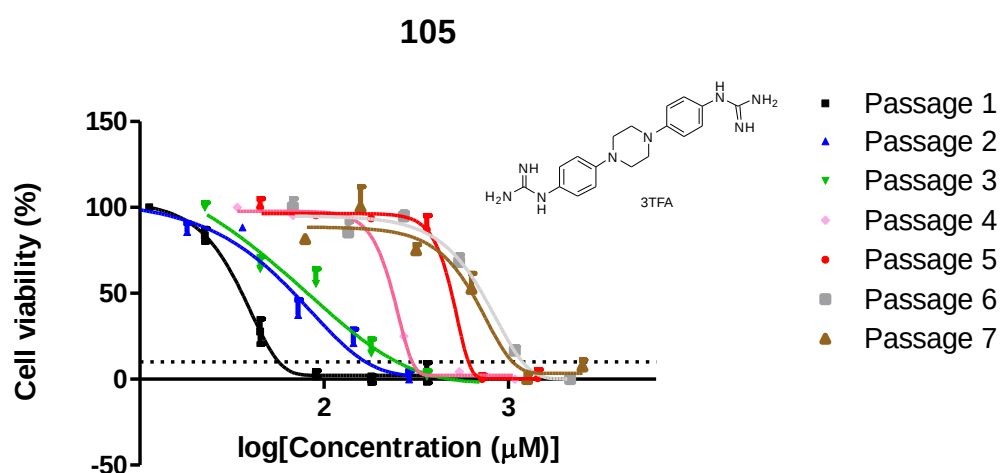


Figure 4.3.1.1.2. REMA drug susceptibility curves of *M. smegmatis* isolates from continual passaging of **105** in small-scale high throughput experiments (the dotted line represents the 10% cell viability mark which equates to the MIC).

After passage 1, no changes in **105** drug susceptibility was observed with respect to the initial exposure to wild-type *M. smegmatis* (MIC = 60.7 μ M). This came as no surprise as just one drug exposure is unlikely to generate drug resistance. However, passage 2 resulted in an

increase in the MIC to 179 μM . The 2.95-fold increase in MIC with respect to the initial exposure suggests that the initiation of resistant mechanisms may be occurring. This is confirmed by the 4.8-fold increase in MIC observed for passage 3 (MIC = 294 μM). Surprisingly, exposure of cells from passage 3 to a higher drug concentration range resulted in minimal changes to the MIC (passage 4 MIC = 301 μM). This suggests that the concentration of **105** present may be too high for the specific drug resistance mechanism to prevent microbial cell death. Following this stall in MIC, a further increase in drug concentration range in passage 5 resulted in an increase in MIC to 624 μM which may be a result of the initiation of an additional resistance mechanism. A similar magnitude of increase was observed for passage 6 where a MIC of 1281 μM was obtained. The slight drop in MIC shown for passage 7 may indicate another plateau in drug susceptibility due to the extremely high concentration of **105**.

Table 4.3.1.1.2. Minimal inhibitory concentration (MIC) of *M. smegmatis* isolates from continual passaging of **105**

Passage	105 Concentration range (μM)	MIC (μM)	ΔMIC (Calculated MIC – Initial MIC)	Fold increase in MIC
1	11 – 360	58	-2.7	0.95
2	11 – 360	179.2	118.5	2.95
3	22 – 720	294	233.3	4.8
4	34 – 1079	300.8	240.1	4.95
5	45 – 1439	623.8	563.1	10.3
6	68 – 2159	1281	1220.3	21.1
7	78 – 2519	1077	1016.3	17.7

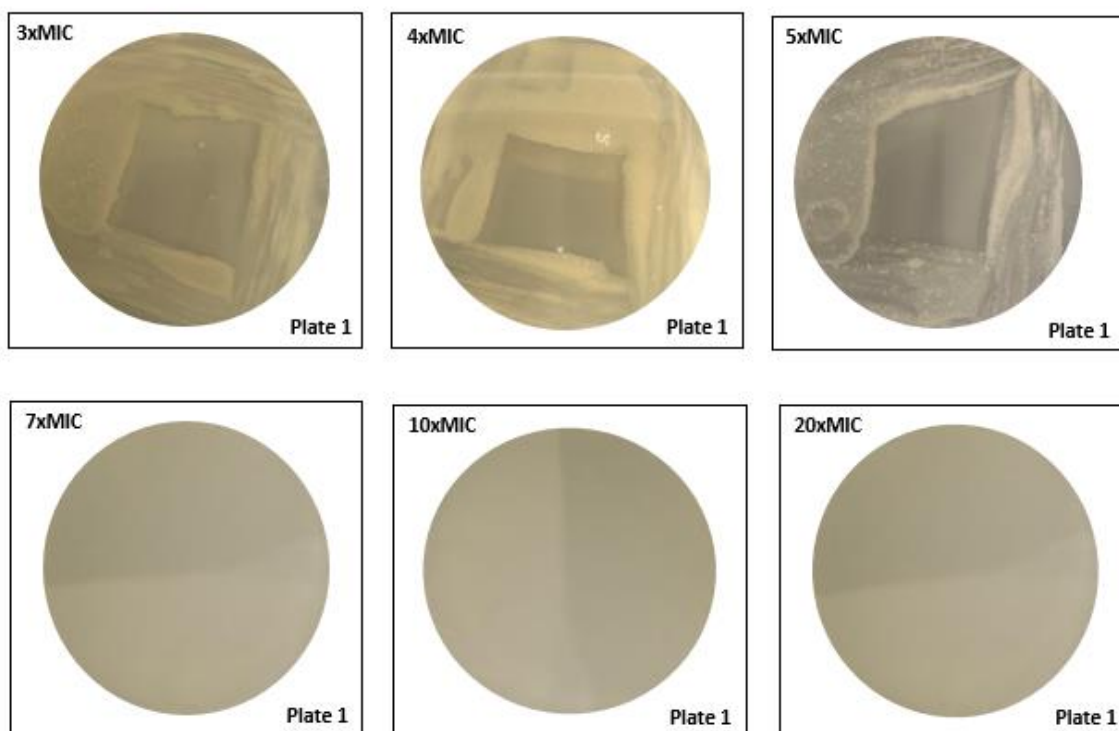
MIC obtained from REMA analysis where it was defined by the drug concentration at 10% cell viability; Initial MIC of **105** against wild-type *M. smegmatis* is 60.7 μM ; Fold increase in MIC is defined by the ratio of the calculated MIC for a specific passage versus the initial MIC.

Overall, a fold increase in MIC as high as 21.1 was obtained after several passages which highlights the capability of this novel method in generating drug resistance. Given these positive results, an investigation into the mutations responsible for these changes in drug susceptibility was conducted with the use of whole-genome sequencing.

4.3.1.2. Solid media single-shot approach

As a comparative approach, we explored the generation of drug resistance on solid media. The use of drug treated agar has been employed in various studies to identify drug resistance mutations by sequencing mutant colonies.^{25, 26} These approaches involve single or double drug exposures at very high concentrations in order to extract any surviving mutant bacteria. Using a slightly modified protocol as described by *Garcia et al.*, *M. smegmatis*, that was adjusted to an optical density (OD₆₀₀) of 0.1 at a volume of 100 µl (~3x10⁶ CFU), was spread on LB agar plates containing various concentrations of **105**.²⁴ Concentrations of 3x, 4x, 5x, 7x, 10x and 20x MIC were attempted where colonies appearing on these plates were scored as resistant mutants only if they grew again on drug-treated LB agar.

a)



b)

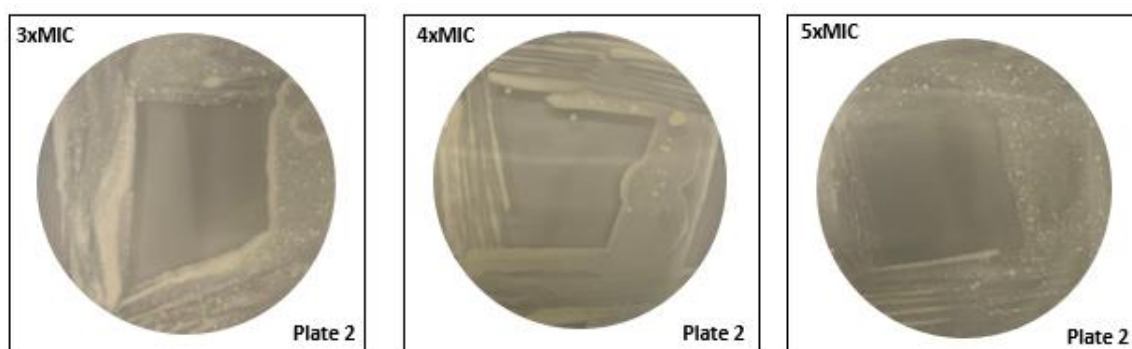


Figure 4.3.1.2.1. (a) Growth of *M. smegmatis* on **105** treated LB agar plates at 3x, 4x, 5x, 7x, 10x and 20x MIC after 3 days of incubation at 37 °C (b) Bacterial growth of colonies isolated from plate 1 of 3x, 4x and 5x MIC that were restreaked onto new **105** treated agar plates.

As shown in Figure 4.3.1.2.1 (a), a dose-dependent decrease in bacterial density can be seen from 3xMIC to 5xMIC of **105**. However, incubation of *M. smegmatis* with **105** at 7xMIC, 10xMIC and 20xMIC resulted in no growth on the agar plate. For this reason, only visible colonies from 3x, 4x and 5x MIC treated plates were picked and re-suspended in 100 μ L of LB

broth followed by re-streaking on new **105** treated plates. The bacterial growth observed in these plates after 2 days of incubation suggested the presence of mutant colonies that could be resistant to **105** (Figure 4.3.1.2.1 (b)). To assess the drug susceptibility of these isolates, REMA analysis was conducted where the MIC obtained was compared to the MIC of **105** against wild-type *M. smegmatis*.

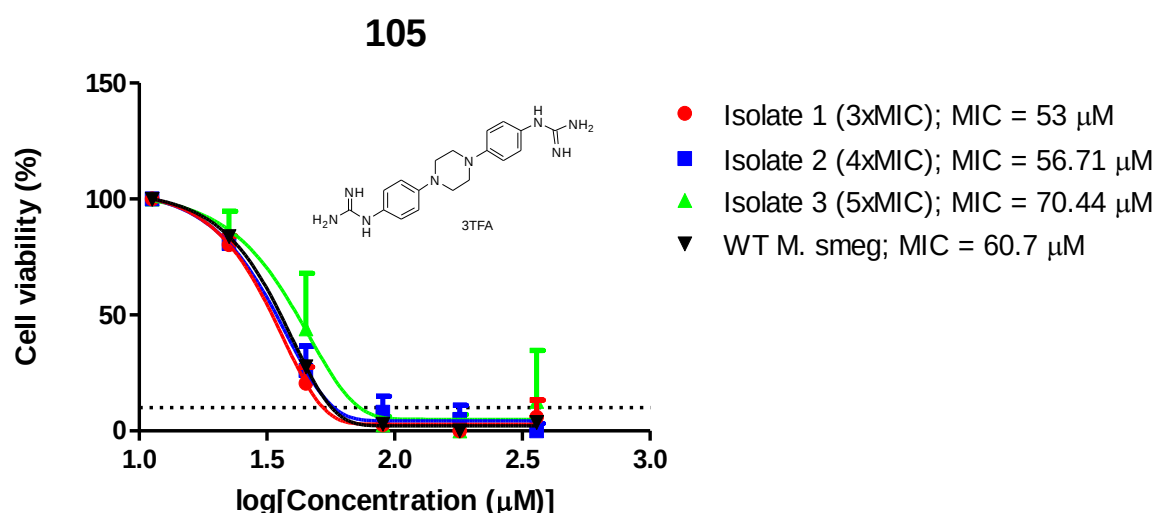


Figure 4.3.1.2.2. REMA drug susceptibility curves of *M. smegmatis* isolates of colonies picked from 3x, 4x and 5x MIC **105** treated LB agar plates (the dotted line represents the 10% cell viability mark which equates to the MIC).

As shown in the figure above, we observed minimal changes in the MIC of the three passages compared to the MIC of **105** against wild-type *M. smegmatis*. The highest change in MIC was observed for colonies isolated from 5xMIC treated plates; however, as this falls within the error span in MIC ($\pm 9.7 \mu\text{M}$) against wild-type *M. smegmatis*, the development of resistant strains could not be verified. Noteworthy, these results are not consistent with the growth of colonies in concentrations as high as 5xMIC (Figure 4.3.1.2.2). However, the MIC of certain antimicrobial agents has been shown to differ in solid and liquid media due to differences in drug diffusion, surface area and oxygen availability.²⁹ Therefore, as the MIC of **105** has been determined by REMA which uses liquid media, it could be considered that the growth seen on the 3x, 4x and 5x MIC treated plates may be a result of incubation at concentrations lower than the true “solid-media MIC”. To analyse whether the development of drug-resistant mutants has occurred or not, the DNA of each isolate was extracted and sequenced to identify any causative mutations.

4.3.2. Characterisation of resistant mutants through DNA sequencing

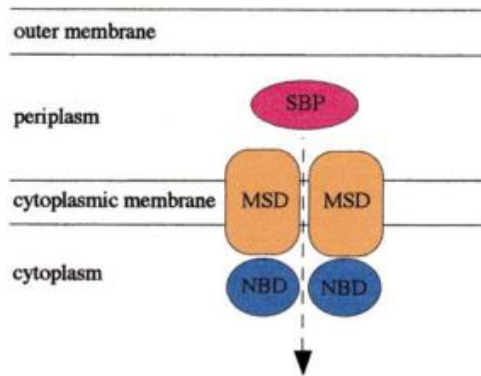
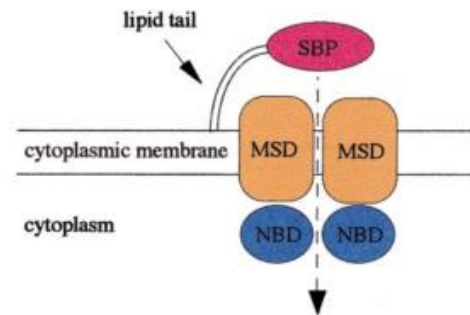
To identify genetic variations in the isolates obtained from the liquid media and solid media experiments shown above, we conducted whole-genome sequencing with the use of the Illumina next-generation sequencer. The SNiPgenie pipeline developed by our collaborators at the School of Veterinary Medicine (UCD) was used to process the reads obtained after DNA sequencing.³⁰ Alignment of these nucleotide reads with the reference genome sequence of *M. smegmatis* mc²155 (NCBI reference sequence = NC_018289.1) allowed for the identification of single nucleotide polymorphisms (SNPs) and insertions or deletions (INDELs). To accommodate for the presence of mutations that may have occurred over time from the ancestor strain, any variants that we identified that were common across all of the strains (wild-type and mutants) were deemed 'background mutations' while only those mutations identified in the bacteria passaged in **105** were potential resistance mutations. From our liquid media experiments shown in section 4.3.1.1 DNA extracts from passages 1 to 7 were sequenced and compared to wild-type *M. smegmatis*. Although several background mutations were identified (See Appendix, Table 2), three interesting INDELs in the ABC transporter genes and one SNP in the TetR family transcriptional regulator gene were identified (Table 4.3.2.1).

Table 4.3.2.1. Mutations identified in each passage from liquid media experiments

Passage	Position	SNP/INDEL	Gene	Protein	Mutation	Type
1-4	728258	GCCCCC*>GCCCCCCC	MSMEI_RS03175	ABC transporter substrate-binding protein	Insertion	Frameshift
	726428	CGGGGG*>CGGGG	MSMEI_RS03170	ABC transporter permease	deletion	Frameshift
	723342	GCCCCC*>GCCCCCCC	MSMEI_RS03155	Dipeptide ABC transporter ATP-binding protein	Insertion	Frameshift
5-7	727562	G>GC	MSMEI_RS03175	ABC transporter substrate-binding protein	Insertion	Frameshift
	6288289	A>G	MSMEI_RS19970	TetR family transcriptional regulator	14T>14A	mis

Base calling of DNA extracts was conducted using the Illumina platform; Variant calling was conducted using the SNIpGenie pipeline; Variants displayed are present in wild-type *M. smegmatis* and are deemed background mutations

From passages 1-4, the only variants present are those within the ATP-binding cassette (ABC) transporter genes, namely, the substrate-binding protein (SBP), permease (or transmembrane domain (TMD)) and ATP binding protein (or nucleotide-binding domain (NBD)). Two insertions in the SBP and the ATP binding protein and one deletion in the permease were identified, all resulting in frameshift mutations. Mycobacteria utilise ATP transporters to import ('importers') and expel ('exporters') substances such as amino acids, peptides, sugars, antibiotics and ions. As the SBP is only found in bacterial ABC importers, it can be concluded that the INDELs present are within an ABC importer rather than an exporter. Therefore, the presence of this variant indicates that this ABC importer is responsible for the uptake of **105** into the mycobacterial cell. Noteworthy, this specific transporter is a dipeptide ABC importer which may explain the selectivity for **105** given its size and the presence of guanidine groups which is also found in the amino acid arginine. Moreover, mutations within this system prevent the uptake of **105** into the cell through reduced binding to the SBP, reduced activity of the TMD or reduced activation of the ATP binding protein. Due to *M. smegmatis* being a phylogenetically gram-positive bacteria, the SBP is attached to the cytoplasmic membrane through a lipid anchor as shown in Figure 4.3.2.1.

IMPORTERS**Prototype in Gram-negative bacteria****Prototype in Gram-positive bacteria****Figure 4.3.2.1.** Topological organisation of the bacterial ABC importers.³¹

Mechanistically, the SBP binds to its respective substrate to form the SBP-substrate complex which interacts with the TMDs leading to an outward-facing conformation and release of the substrate into the TMDs binding cavity (Figure 4.3.2.2). ATP hydrolysis at the consensus site of one nucleotide-binding domain (NBD) occurs, leading to another ATP hydrolysis at the second consensus site of the second NBD. Conformational changes in TMDs lead to an inward-facing conformation that allows translocation of the substrate into the cytosol allowing the importer to return to its basal state by nucleotide exchange.

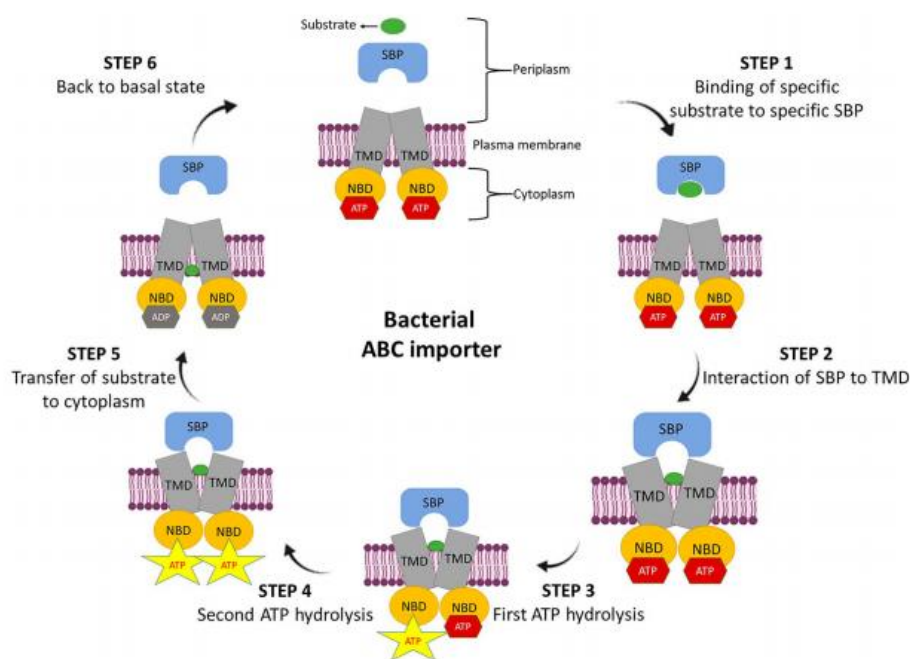


Figure 4.3.2.2. Schematic representation of the ABC import system.³²

The gradual increase in MIC displayed from passage 1 to 4 may be due to a gradual increase in expression of mutants of the ABC transporter making the bacteria less susceptible to **105**. As stated previously, a stall in MIC was observed from passages 3 to 4. This may be due to the inability of the ABC transport mutants to survive the increasing presence of **105** at such high concentrations. Interestingly, from passage 5 onwards, we identified a new mutation in the TetR family transcriptional regulator gene which coincided with a large increase in MIC. TetR family transcriptional regulators (TFRs) are widely associated with antibiotic resistance and the regulation of genes encoding small-molecule exporters through a series of signal transduction cascades. TetR is known to regulate the expression of an efflux pump encoded by the gene *tetA* which is responsible for the efflux of common antibiotics such as tetracycline, fluoroquinolones and acridine.^{33,34,35} The TetR protein consists of an N-terminal DNA-binding domain and a larger C-terminal ligand-binding domain which alters the regulator's ability to bind DNA. As shown in Figure 4.3.2.3, in the absence of tetracycline, a pair of TetR dimers bind to overlapping operator DNA sequences in the intergenic region between the *tetR* and *tetA* genes. In the presence of tetracycline, it binds directly to TetR, trapping it in a conformation that is incompatible with DNA binding. This allows transcription of both *tetR* and *tetA* leading to increased efflux of tetracycline out of the cell.

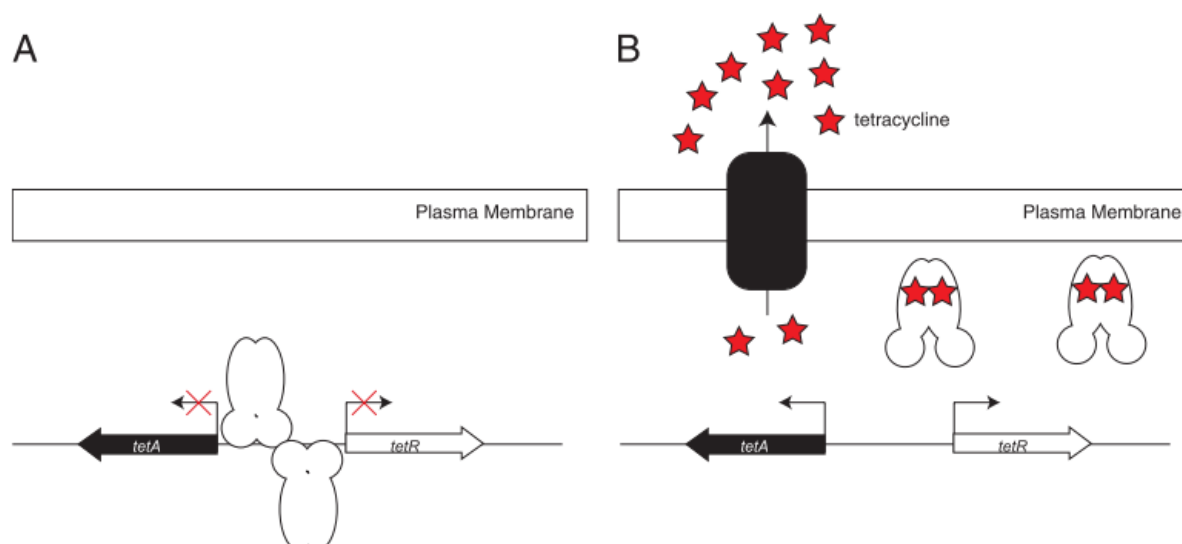


Figure 4.3.2.3. Schematic representation of the TetR family transcriptional regulator system. (a) In the absence of tetracycline, TetR protein is bound to the repeating operator sequence between the *tetR* and *tetA* gene (b) When tetracycline enters the cell it binds to TetR causing its release from the operator sequence and expression of *tetR* and *tetA*.³⁴

Intriguingly, the mutation identified in the *tetR* gene was a SNP from adenine to guanine which caused a missense mutation resulting in the conversion of a threonine at position 14 (T14) into an alanine. As the T14 is located at the N-terminal DNA binding domain as shown in Figure 4.3.2.4, it can be suggested that this polymorphism results in the reduced affinity of TetR towards the operator sequence leading to an increase in *tetA* expression. Noteworthy, the change from threonine to alanine results in the loss of a HB donor/acceptor which may be important for DNA binding explaining the reduced affinity. Additionally, taking the reference crystal structure of structurally similar acridine bound to the C-terminal ligand-binding domain (PDB= 2V57)³⁶, we docked **105**. This showed that one guanidine forms electrostatic interactions with ASP125 and HB interactions with ASN71. Two π - π interactions are also formed with TYR106 and TRP152 which holds the molecule with the binding site allowing the other terminal guanidine to interact with water molecules on the aqueous surface.

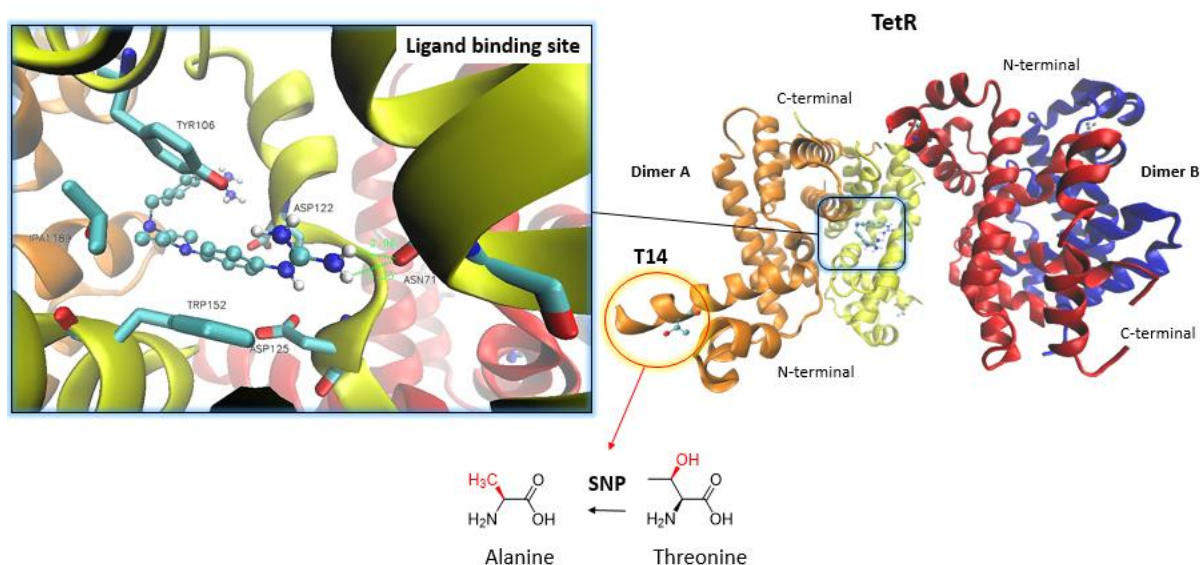


Figure 4.3.2.4. Optimal docking pose of **105** in the ligand-binding site of TetR showing the location of the SNP obtained from continual passaging (reference crystal structure PDB: 2V57).

The initiation of this efflux system explains the increase in MIC from the fourth passage (300.8 μ M) to the fifth passage (623.8 μ M) followed by an increase of similar magnitude shown in the sixth passage (1281 μ M). However, a drop in MIC was observed when exposed to higher concentrations in the seventh passage (1077 μ M) which may be due to the inability of this new combined *ABC-tetR* mutant to survive such high concentrations of **105**.

Thus far, this continual passaging experiment has yielded information on the transporter responsible for the uptake of **105** as well as the specific efflux system responsible for resistance towards the drug compound. Although these two aims have been met, information on the mechanism of action of **105** was not gained as indicated by the absence of mutations in the proposed target, *InhA*, or any other genes essential for bacterial survival. It is well known that mutations in the coding region of *inhA* are rare and resistance towards *InhA* covalent inhibitors arises from mutations in the gene encoding the prodrug activating protein KatG.^{37, 38} Consequently, it was hypothesised that mutations within *inhA* may be detrimental to cell viability. Given the fact that this continual passaging approach to generate mutants involves the selection of the most “fit” bacteria from the population of cells, a bottleneck scenario may be occurring resulting in the cells bearing any *inhA* mutations being lost. To investigate this, the DNA of *M. smegmatis* isolates from the solid media experiments was

sequenced. As these isolates were obtained without continual passaging and from colonies picked off the surface of the drug-treated plates, the cells present may give more information on the effects of the selective pressure.

Table 4.3.2.2 shows the results obtained from the whole-genome sequencing of the *M. smegmatis* isolates generated from single-shot solid media experiments described in section 4.3.1.2. Similar to isolates generated from the continual passaging experiment, we identified various background SNPs and INDELs which are not a result of **105** exposure (see Appendix, Table 2). No causative mutations were identified from the 3xMIC and 4xMIC isolates which is consistent with the minimal changes to the MIC compared to the wild-type *M. smegmatis* (Figure 4.3.1.2.2).

Table 4.3.2.2. Mutations identified from single-shot solid media experiments

Isolate	Position	SNP/INDEL	Gene	Protein	Mutation	Type
3xMIC	-	-	-	-	-	-
4xMIC	-	-	-	-	-	-
5xMIC	6288675	TCCCCC>TCCCCC	MSMEI_RS30470	TetR family transcriptional regulator	Insertion	Frameshift

Base calling of DNA extracts was conducted using the Illumina platform; Variant calling was conducted using the SNiPgenie pipeline; Variants displayed are present in wild-type *M. smegmatis* and are deemed background mutations

Interestingly, the 5xMIC isolates presented an INDEL in *tetR* gene which was previously identified from our continual passaging experiments. The presence of this mutation may indicate the initiation of **105** resistance. This would suggest that more exposures at higher concentrations would be needed to facilitate the formation of more causative variants. However, exposure to concentrations of **105** higher than 5xMIC was shown to result in complete cell death and no colony growth. This confirms that gradual continual passaging is more optimal for drug-resistant mutant generation. Moreover, the inability to identify mutations in the proposed target using both methods led to investigations into gene expression rather than gene mutation.

4.3.3. InhA expression analysis

It is important to note that gene-mediated resistance not only occurs through SNPs and INDELs but also through the overexpression of the target protein. As no mutations were identified in *inhA* to conclusively verify it as the drug target, an investigation into the possibility of *inhA* overexpression was conducted. Firstly, to assess whether *inhA* overexpression has an effect on the activity of **105**, we performed drug susceptibility analysis against an InhA overexpressing mycobacterial strain. To generate an *inhA* overexpressor, we cloned *inhA* independently downstream of the strong *hsp60* promoter in a multicopy plasmid pMV261 which was then transformed into *M. bovis* BCG. To confirm the generation of the *M. bovis* BCG *inhA* overexpressor, its susceptibility against isoniazid, a known InhA inhibitor, was assessed through an MTT assay. As shown in Figure 4.3.3.1, a significant decrease in cell viability as indicated by the decreased presence of the tetrazolium dye in its reduced form, was observed between 0.1-1 $\mu\text{g mL}^{-1}$ of isoniazid in the wild-type strain (pMV261). However, no anti-mycobacterial activity was observed in the transformed strain (pMV261-InhA) indicating isoniazid resistance through InhA overexpression.

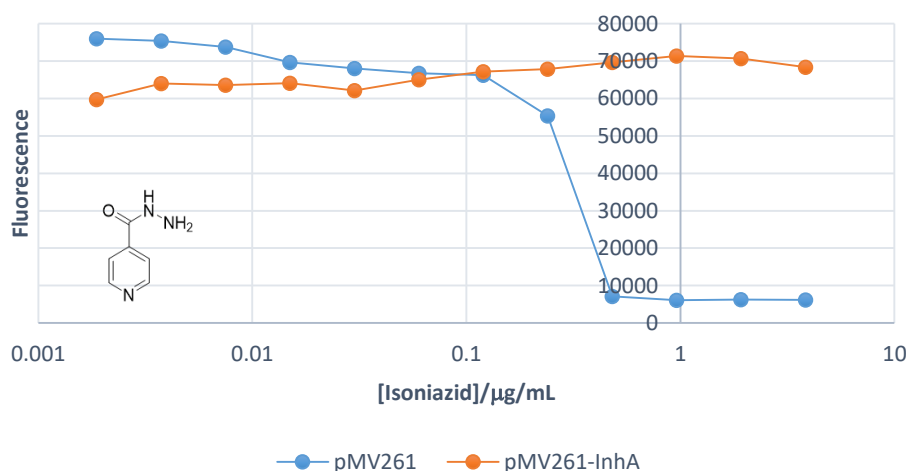


Figure 4.3.3.1. Drug susceptibility curves of wild-type (pMV261) and *inhA* overexpressing (pMV261-InhA) *M. bovis* BCG against isoniazid.

Given the fact that one of the proposed targets for **105** is InhA, a similar profile was expected. However, an opposite effect was observed upon exposing the compound to both strains (Figure 4.3.3.2). Surprisingly, the transformed strain was more susceptible to **105** than the

wild-type strain causing a greater decrease in cell viability between 0.03-1 $\mu\text{g mL}^{-1}$. While overexpression of a drug's molecular target increases drug resistance, offering a tool for target identification, studies have shown that this phenomenon applies to some drugs but not to all. Although the reasoning behind this is still quite unclear, it has been theorised that growth may be inhibited by the toxic activity of a drug-bound target, rather than a lack of normal target activity therefore target overexpression enhances drug sensitivity instead of conferring resistance. This has been shown for two DNA gyrase inhibitors ciprofloxacin and coumermycin A1 where ciprofloxacin causes increased susceptibility and coumermycin A1 causes decreased susceptibility in *M. tuberculosis* DNA gyrase overexpressors.³⁹

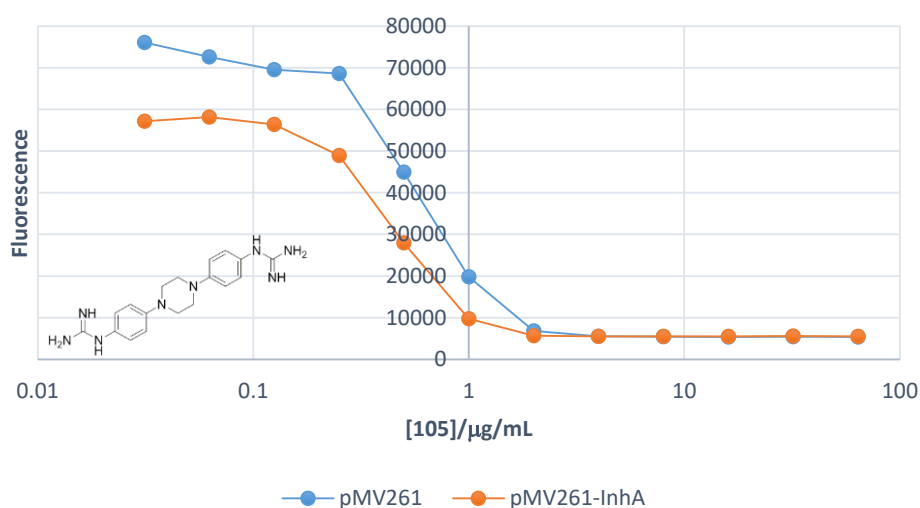


Figure 4.3.3.2. Drug susceptibility curves of wild-type (pMV261) and *InhA* overexpressing (pMV261-InhA) *M. bovis* BCG against **105**.

These results suggest that the first defense mechanism of the bacteria, when exposed to **105**, may have been to overexpress *inhA*. However, as this negatively affects cell viability in the presence of **105** as shown above, other mechanisms such as the ABC importer inactivation and TetR suppression were activated to reduce the presence of **105** in the cell. To investigate this, we replaced the *tetR* mutant gene of a passage 7 isolate with the wild-type *tetR* gene with the use of a constructed plasmid vector (pMV361-TetR). This was performed to potentially reverse the suspected interplay between *inhA* overexpression and reduced **105** intracellular presence. Through qPCR analysis, as shown in Figure 4.3.3.3, the level of *inhA*

gene expression of the passage 7 isolate (P7a-c) was slightly lower than that of wild-type *M. smegmatis* (WT1-3). Additionally, the replacement of the mutant *tetR* gene with wildtype *tetR* (C1-3) was shown to cause a drastic increase in *inhA* expression. This confirms that *inhA* overexpression occurs initially which causes an increased susceptibility to **105** resulting in the initiation of non-target-based mechanisms of resistance.

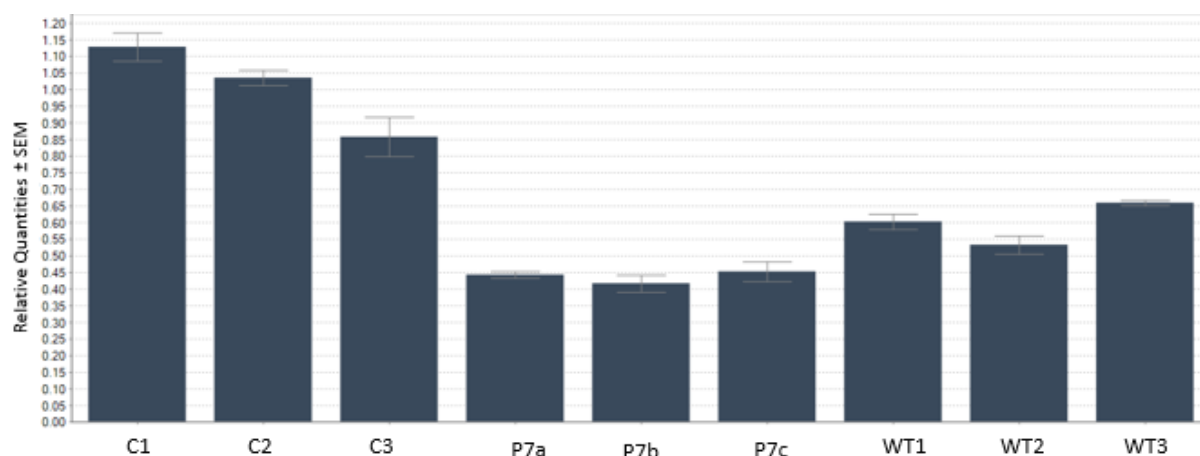


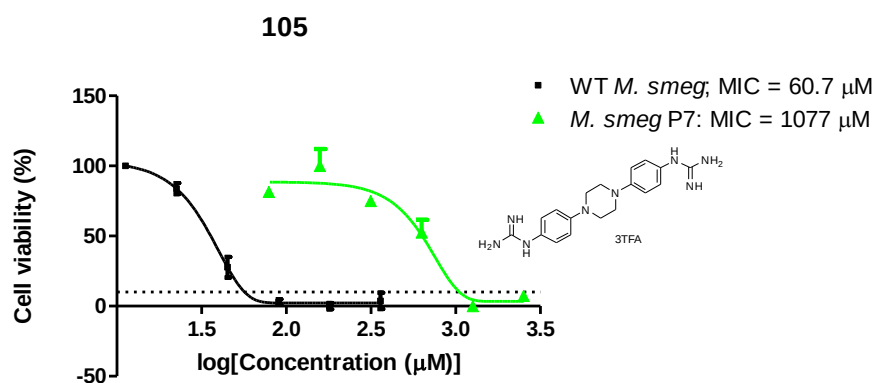
Figure 4.3.3.3. Quantitative PCR (qPCR) analysis showing the level of *inhA* expression of the *M. smegmatis* passage 7 isolates complemented with wild-type *tetR* (C1-3), *M. smegmatis* passage 7 isolates (P7a-c) and wild-type *M. smegmatis* (WT1-3).

4.3.4. Evaluation of cross-resistance

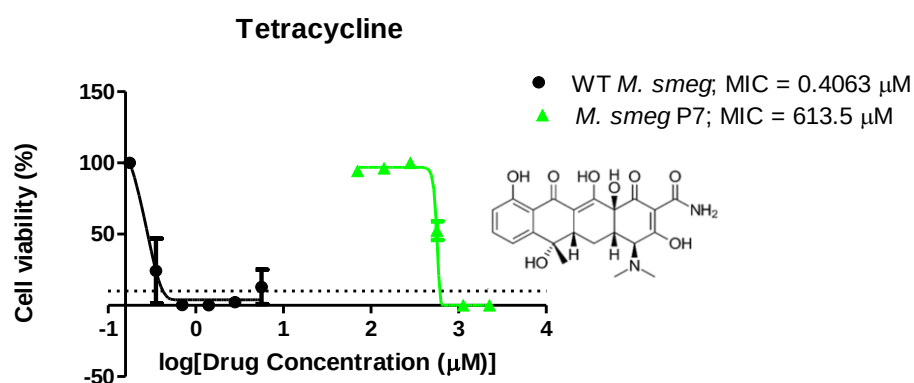
One of the advantages of being able to deliberately generate drug resistance at a preclinical stage is the ability to predict certain drug-resistant mutations before the drug is marketed. This allows for the preparation of second-generation agents that evade these resistant mechanisms as well as rapid allocation of effective alternative treatments. To demonstrate an example of this, we tested known anti-tubercular agents, rifampicin and tetracycline, against a **105** resistant strain of *M. smegmatis* (passage 7 isolate). Figure 4.3.4.1 (a) shows that **105** is 17.7 times less active against the resistant strain versus the wild-type strain. Analysis of this **105** resistant strain against tetracycline resulted in a 1510-fold increase in MIC compared to the wild-type strain. This huge decrease in drug susceptibility can be explained by the mutation found in the TetR protein which is a known target for tetracycline. The consequence of this mutation caused by overexposure to **105** results in the increased efflux

of tetracycline due to the reduced ability of TetR to suppress the expression of the TetA efflux pump. As a relevant side note, this shows how the misuse of a certain antibiotic can affect the use of other agents through antibiotic resistance. Examination of rifampicin against the **105** resistant strain showed only minimal changes to the MIC indicating that rifampicin would be a more suitable alternative treatment as opposed to tetracycline.

a)



b)



c)

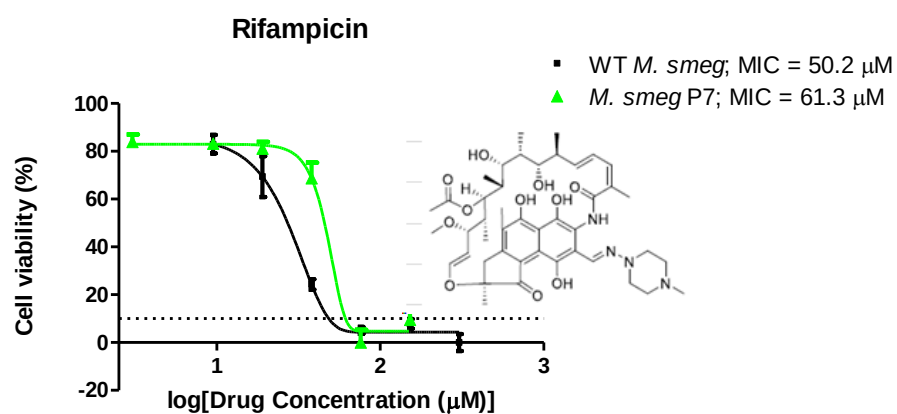


Figure 4.3.4.1. REMA drug susceptibility curves of (a) **105**, (b) rifampicin and (c) tetracycline against wild-type *M. smegmatis* and **105** resistant *M. smegmatis* (*M. smeg* P7) generated from passage 7 of the continual passing experiment (the dotted line represents the 10% cell viability mark which equates to the MIC).

4.4. Design and Molecular Modelling

Based on the molecular docking and FAMES and MAMES TLC analysis shown previously, it was concluded that a more amphiphilic motif would be optimal for InhA binding. Inspection of the key interactions and binding mode of previous compounds under investigation showed that the guanidinium moiety acts as the “war-head” occupying the inner binding pocket. The piperazine linker in its chair conformation allows the other terminal of the molecule to orient towards the hydrophobic pocket resulting in an “L-shaped” binding pose. This binding mode is similar to that found in crystal structures of known InhA inhibitors bound to the active site of the enzyme. Examples of these structurally diverse InhA inhibitors include **63**, **64**, **65**, QEB and triclosan (Figure 4.4.1).^{40, 41} A distinct feature found in almost all of these inhibitors is the presence of a central amide functional group which forms HBs with TYR158 and NAD. As TYR158 has been suggested to be key in the stabilisation of the transition state of the natural substrate, the observed HB interaction plays a key role in the inhibition of the enzyme.

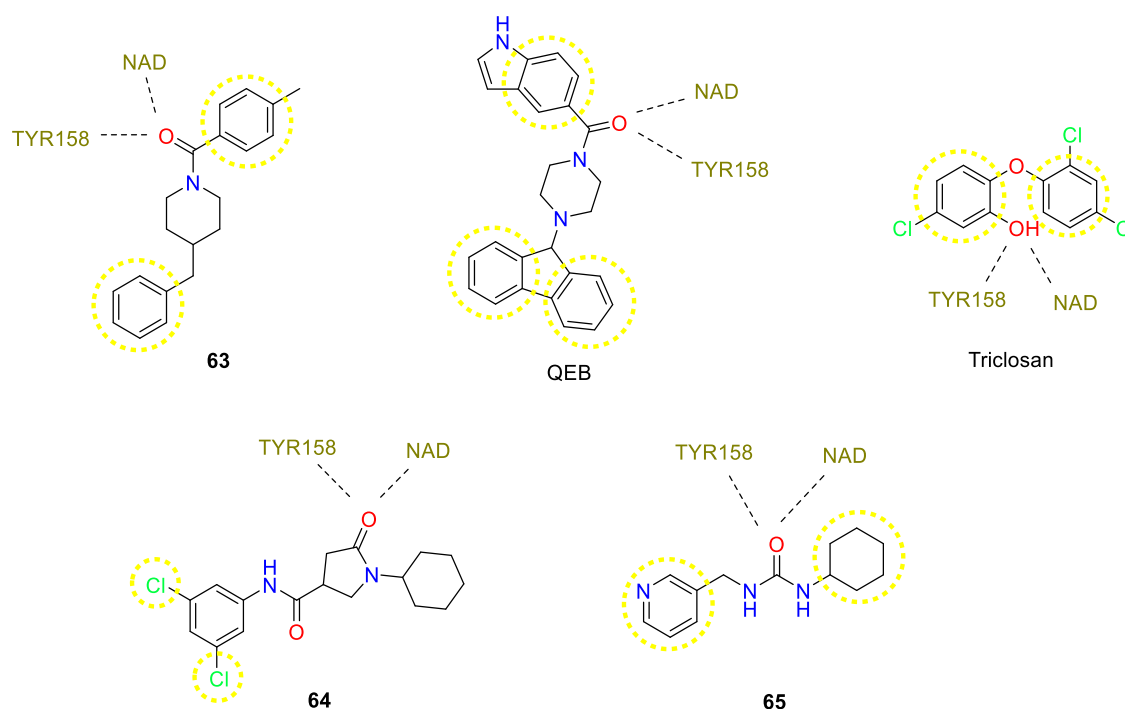


Figure 4.4.1. 2D representations of the binding orientation of InhA inhibitors in the InhA active site. Parts of the molecules participating in InhA binding are highlighted: HBs are represented as black dotted lines and hydrophobic interactions as yellow dotted circles.

As shown in the docking poses of **105**, **106** and **109**, no interactions with TYR158 was observed. Therefore, to probe this region of the binding pocket, a novel series of compounds were designed, namely, Family G and Family H (Figure 4.4.2). Family G consists of derivatives containing an amido-piperazine linker where both bis-cationic and amphiphilic motifs are considered. On the other hand, Family H retains the same diaryl piperazine scaffold as previous compounds with an amphiphilic motif. Hence, both the ability to further probe the hydrophobic pocket and the inner binding are under investigation. Additionally, in previous *in vivo* experiments to analyse the anti-trypanosomal potency of DNA minor groove binders **105** and **106**, severe cytotoxicity was observed.⁴⁷ Therefore, by incorporating the amide bond and/or removal of one cation may serve to diminish the DNA binding ability of these novel family of compounds limiting any possible cytotoxicity. To analyse the effect of these changes molecular docking analysis was conducted both in InhA and DNA.

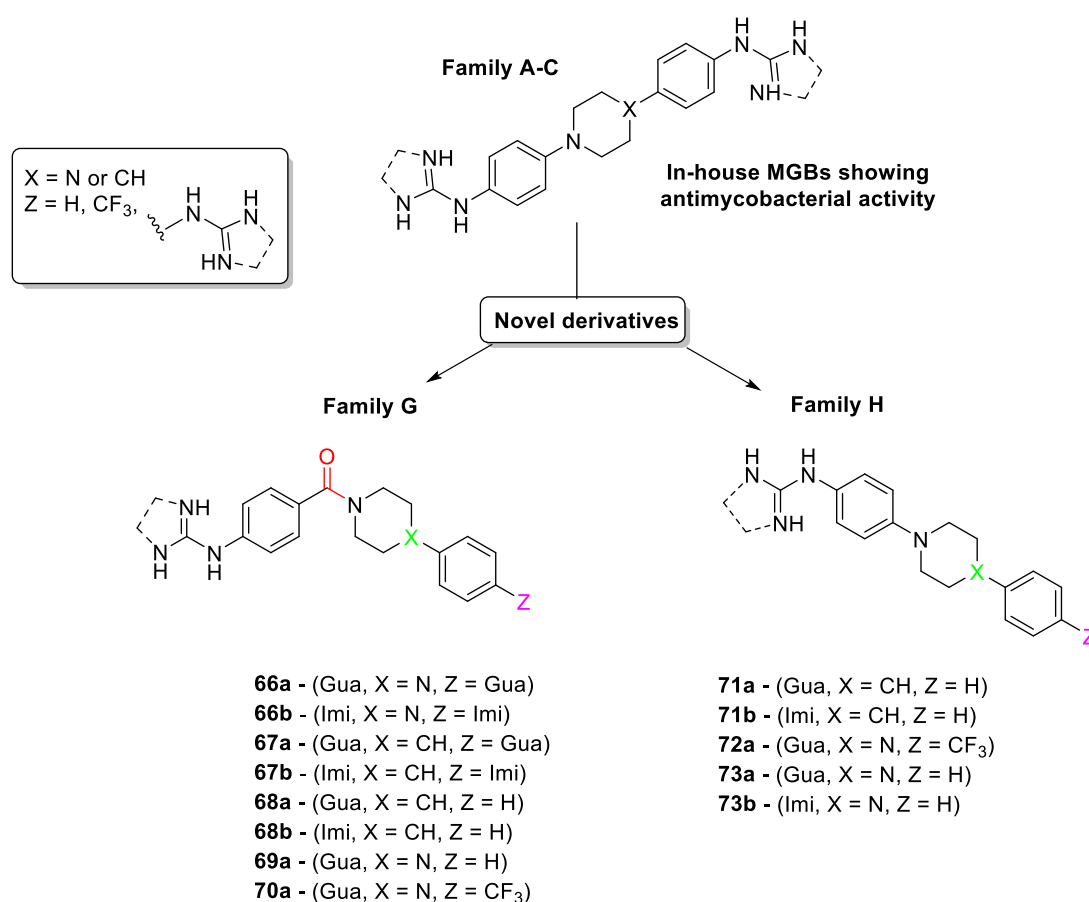
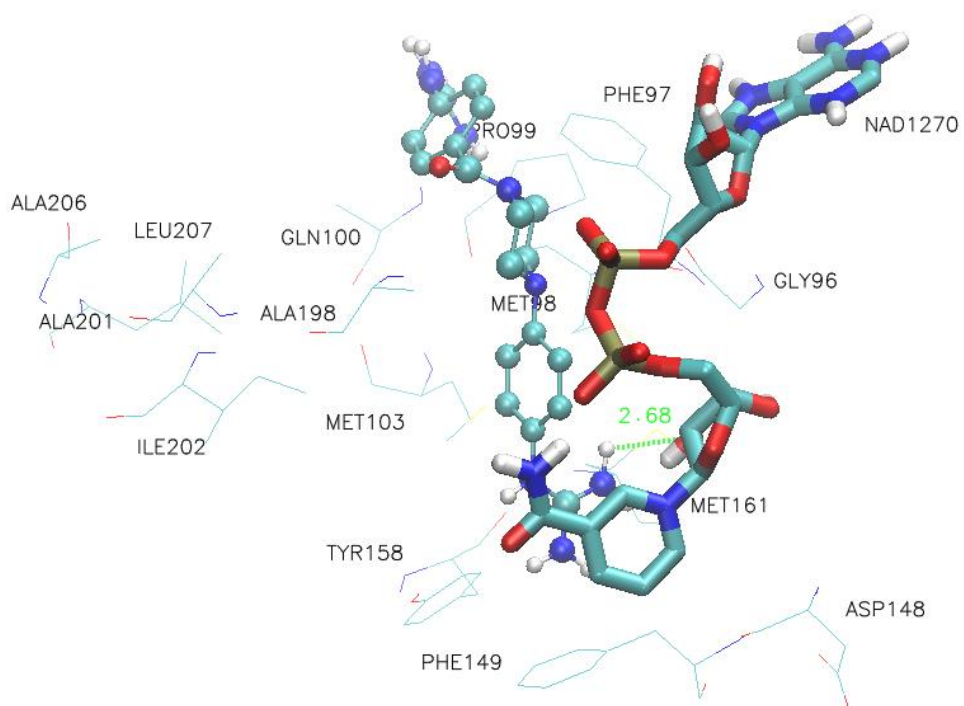


Figure 4.4.2. Scheme of the structural modifications made to previous in-house compounds to produce novel families G and H.

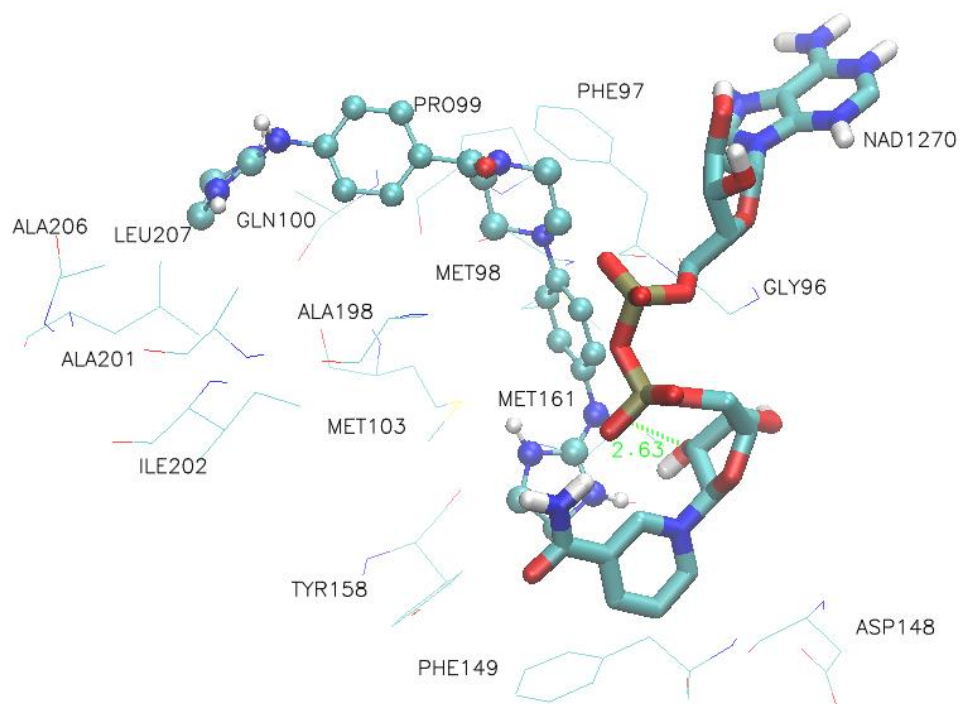
4.4.1. Molecular docking studies of novel Families G and H in InhA

Using the reference crystal structure of MTH bound to the InhA active site as shown in Section 4.2 (PDB: 4BQR), the binding modes of Family G compounds were first analysed. Considering the bis-guanidine/2-aminoimidazoline derivatives **66a** and **66b**, similar modes of binding were obtained (Figure 4.4.1.1). It was previously hypothesised that the guanidinium or 2-aminoimidazolinium moiety closest to the amide carbonyl would occupy the inner binding pocket where the carbonyl could interact with the NAD cofactor or TYR158. However, upon docking these compounds, the favoured orientation was one where the amide carbonyl points towards the outer part of the active site allowing the guanidinium-like cation furthest away to occupy the inner binding pocket. In the case of compound **66a**, the absence of a more lipophilic functionality at the other terminal of the molecule causes the guanidinium moiety to point away from the hydrophobic pocket and towards the aqueous surface. This is reflected in the slightly better Gscore obtained for **66b** ($-7.4 \text{ kcal mol}^{-1}$) versus **66a** ($-6.9 \text{ kcal mol}^{-1}$). When comparing this scenario to binding modes of **105** and **106** shown in the previous chapter, we see that the HB formed by the guanidinium of **105** with ALA198 is lost when incorporating the amide bond as in **66a**; this is due to a lack of flexibility and increased length of the molecule. When analysing the effect of having a piperidine over a piperazine linker no significant differences were observed in the mode of binding seen with **66b** and **67b**.

a)



b)



c)

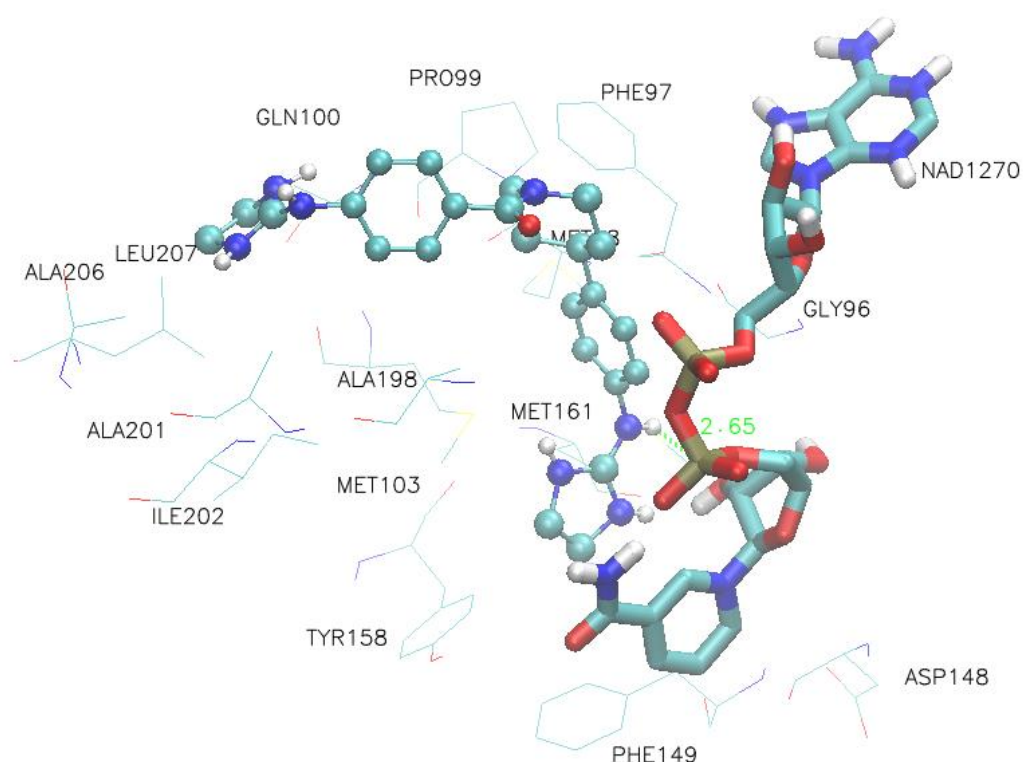
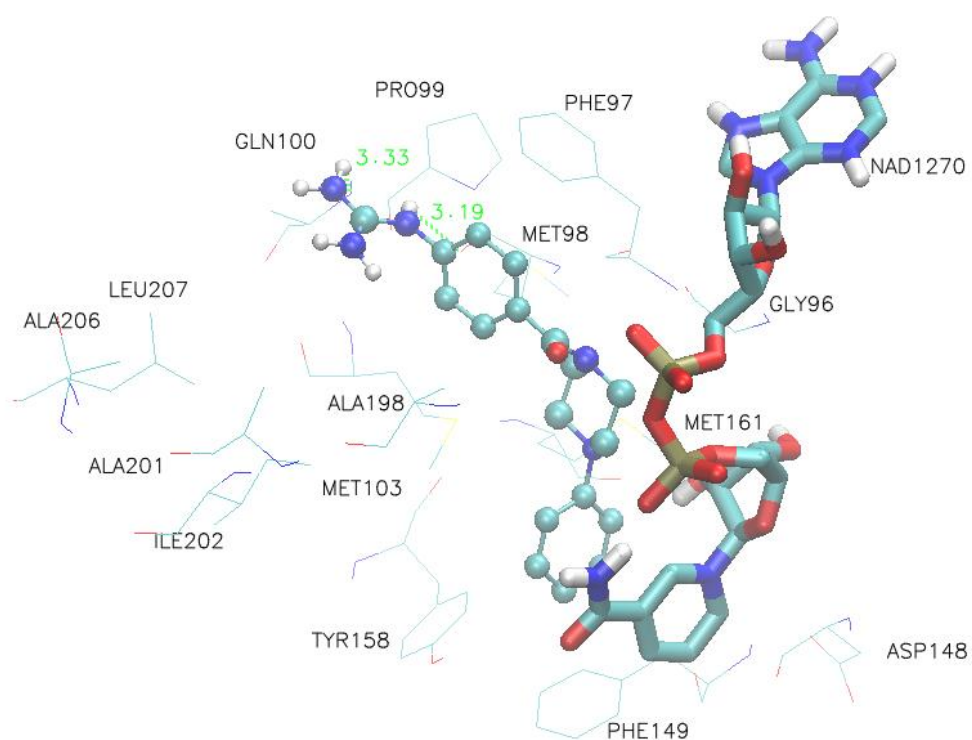


Figure 4.4.1.1. Docking poses of (a) **66a**, (b) **66b** and (c) **67b** within the active site of InhA.

These results as well as conclusions drawn from previous experiments indicate that a more amphiphilic motif would be optimal for binding. Surprisingly, upon docking compounds **68a** and **69a** the guanidinium appears to point away from the inner binding pocket forming very weak HB interactions with GLN100 and MET98 towards the surface (Figure 4.4.1.2). The terminal phenyl ring is shown to occupy the inner binding pocket forming a π - π interaction with the nicotinamide moiety of NAD. These results indicate that a less rigid functional group may be required to allow the interaction of the cationic “warhead” as well as limiting steric clashes with NAD and residues of the inner binding pocket.

a)



b)

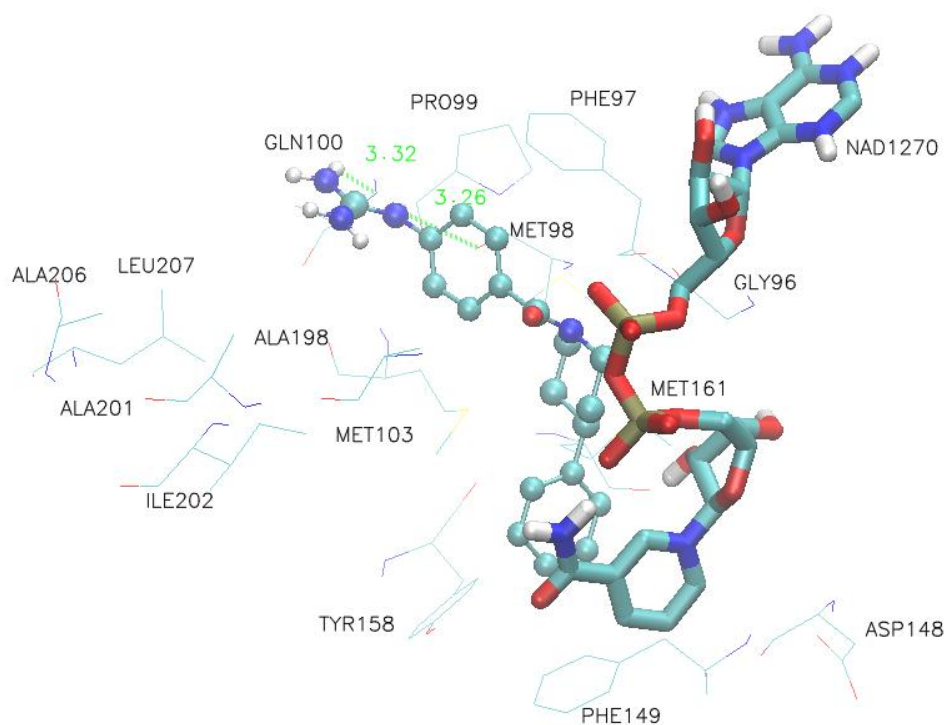
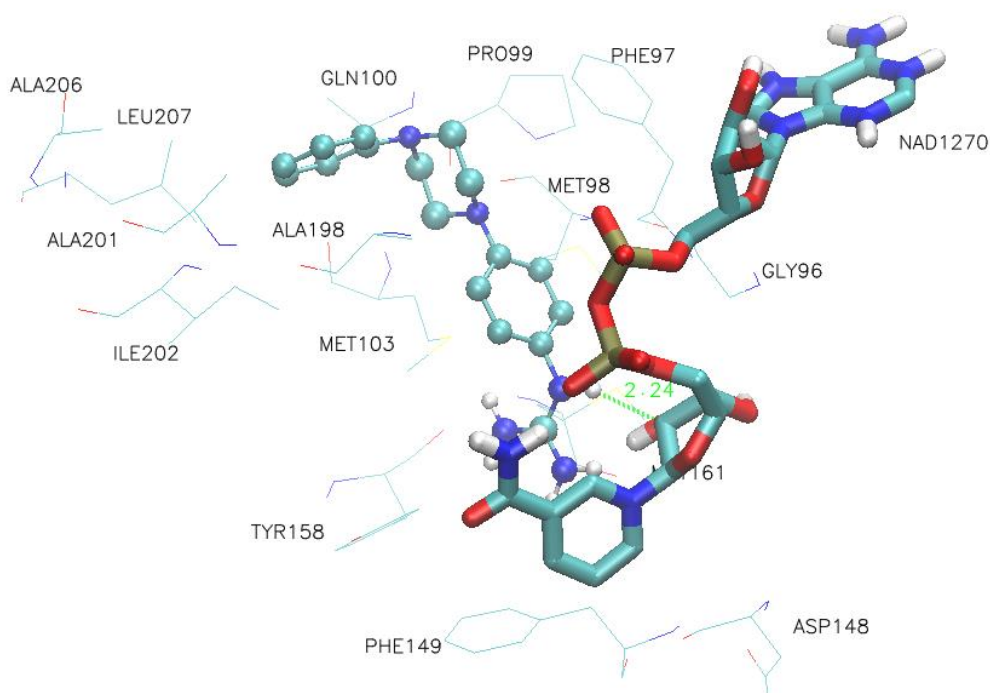


Figure 4.4.1.2. Docking poses of (a) **69a** and (b) **68a** within the active site of InhA.

Docking of Family H compounds **73a** and **72a** resulted in a favourable binding geometry where the guanidinium interacts within the inner binding pocket and the lipophilic portion of the molecule occupies the hydrophobic pocket (Figure 4.4.1.3). The absence of the amide bond and retention of the original piperazine linker confirms that the flexibility of the linker plays an important role in the binding mode of the ligand within the active site. Interestingly, shortening of the molecule by replacement of the phenyl guanidine with phenyl or p-CF₃-phenyl results in a stronger HB interaction with NAD and closer contact to the residues of the hydrophobic pocket. The better Gscore (**72a** = -8.0 kcal mol⁻¹ and **73a** = -7.7 kcal mol⁻¹) and stronger intermolecular interactions of **72a** over **73a** suggests that functionalisation of the phenyl ring with a lipophilic functional group would be beneficial for binding.

a)



b)

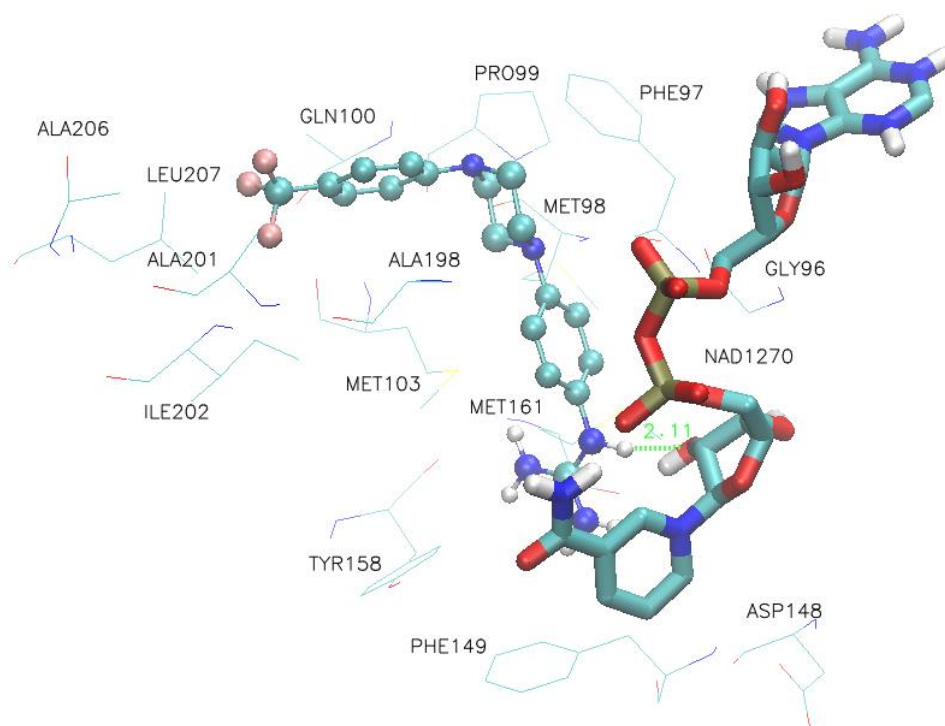


Figure 4.4.1.3. Docking poses of (a) **73a** and (b) **72a** within the active site of InhA.

In summary, Family G compounds exhibit similar binding modes to those seen for previous compounds. Despite the pre-organised “L-shaped” conformation created by the incorporation of the amide bond, the resulting rigidity and increased length causes steric hindrance with residues of the inner pocket and NAD resulting in unfavourable binding modes. Additionally, no significant interactions between the amide carbonyl and active site residues were observed. Analysis of binding modes obtained for Family H compounds highlights the importance of the flexibility of the linker created by the piperazine/piperidine ring. As shown in Table 4.4.1.1, the stark difference in Gscore between the amphiphilic derivatives and the dicationic derivatives indicates that an amphiphilic motif is more beneficial for InhA inhibition.

Table 4.4.1.1. G-score values (kcal mol⁻¹) obtained from the best poses of the novel compounds bound to InhA.

Entry	Family	Cmp	Gscore (kcal mol ⁻¹)
1	G	67a	-7.0
2	G	67b	-6.9
3	G	70a	-7.1
4	G	69a	-7.1
5	G	48b	-6.2
6	G	66a	-6.9
7	G	66b	-7.4
8	G	68a	-6.6
9	G	68b	-6.8
10	H	71a	-7.4
11	H	73a	-7.7
12	H	72a	-8.0
13	H	71b	-7.0
14	H	73b	-7.1

4.4.2. Molecular docking studies of novel Families G and H in DNA

As stated earlier, the previously made compounds under investigation were designed as DNA MGBs for anti-parasitic activity. *In vivo* studies in acute and late-stage models of HAT were conducted with **105** and **106** as these compounds elicited the best DNA binding affinity and *in vitro* anti-parasitic activity. From this study, both compounds were shown to be toxic after the first treatment at 20 mg kg⁻¹. To avoid such toxicity, Family G and H were designed to limit the DNA binding capability by modifying the shape and charge of the molecule. As shown in Figure 4.4.2.1, molecular docking of compound **66a** in a model of the DNA minor groove shows a similar binding mode to compound **105**. However, incorporation of the amide bond linker seems to force one side of the molecule further out from the minor groove and prevents the slight twist in the molecule required to form HB interactions with nucleotides

from both strands. It is key to note that the di-cationic nature of the molecule may still create the possibility for moderate binding due to the electrostatic interactions formed.

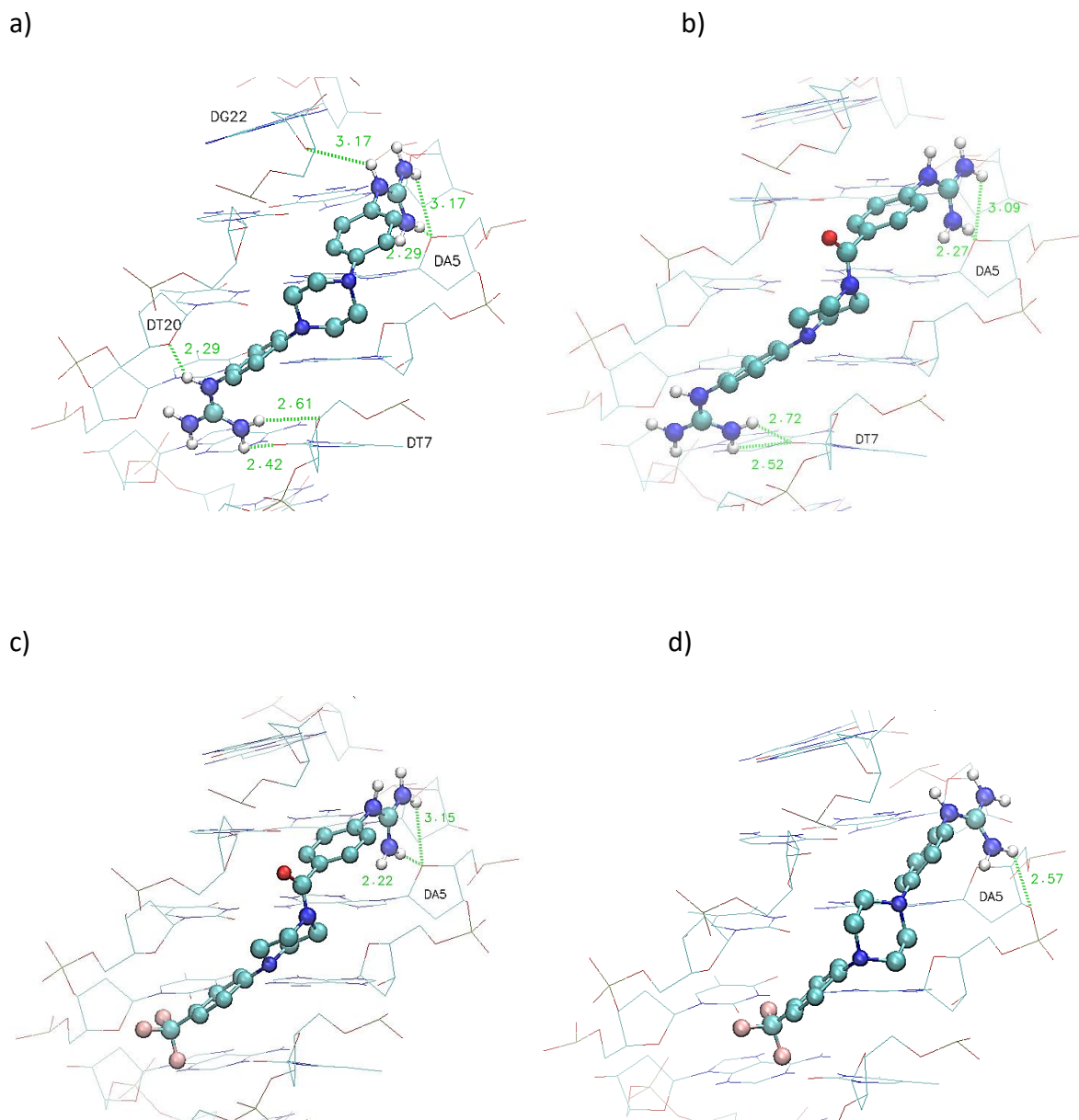


Figure 4.4.2.1. Docking poses of (a) **105**, (b) **66a**, (c) **70a** and (d) **72a** in the minor groove of DNA (reference crystal structure PDB:1D64).

Replacement of one of the guanidines with a CF_3 as seen with compound **70a**, reduces the electrostatic and HB interactions that can be formed within the minor groove and has translated to a worse Gscore score than that of **66a** as seen in Table 4.4.2.1. Analysis of the

docking pose obtained for **72a** shows that only one HB is formed with the deoxyribose of adenine as well as electrostatic interactions with the backbone phosphate. Although the concave geometry of the diarylpiperazine linker is optimal for DNA binding the loss of the second terminal guanidine and replacement with a CF₃ significantly reduces the ability of the molecule to bind tightly into the groove. This is reflective of the less negative Gscore value obtained compared to that of **105**. In summary, compounds of Family G and H exhibit weak DNA binding interactions and may therefore avoid the cytotoxicity observed in previously made compounds.

Table 4.4.2.1. G-score values (kcal mol⁻¹) obtained from the best poses of compounds **105**, **66a**, **70a** and **72a** bound to the minor groove of DNA.

Entry	Family	Cmp	Gscore (kcal mol ⁻¹)
1	A	105	-8.5
2	G	66a	-8.1
3	G	70a	-7.8
4	H	72a	-6.6

4.5. Synthesis

A series of di-aryl guanidinium/2-aminomidazolinium based compounds bearing piperazine and piperidine linkers have been previously produced by our group (Figure 4.5.1 – Family A-C).^{42, 47} These compounds were initially designed as DNA MGBs to be used as potential anti-protozoal agents. As stated earlier, based on the work of Forge *et al.* who reported a series of structurally similar bis-amidine based compounds with good anti-mycobacterial potency, we were encouraged to screen our in-house family of compounds for anti-mycobacterial activity.¹ Preliminary screening against surrogate strains of *M. tuberculosis* as described in Section 4.1 has shown that our previously prepared compounds elicit good anti-mycobacterial activity with MICs in the low μM range. Further optimisation of these compounds was conducted following an investigation into a potential mechanism of action after a miscorrelation between DNA binding affinity and anti-mycobacterial activity was observed. This led to the design of Families G and H, which have been docked into the potential target Enoyl ACP reductase (InhA) as shown in Section 4.4. This section will describe the synthesis of previously reported Family A-C as well as new families G and H.

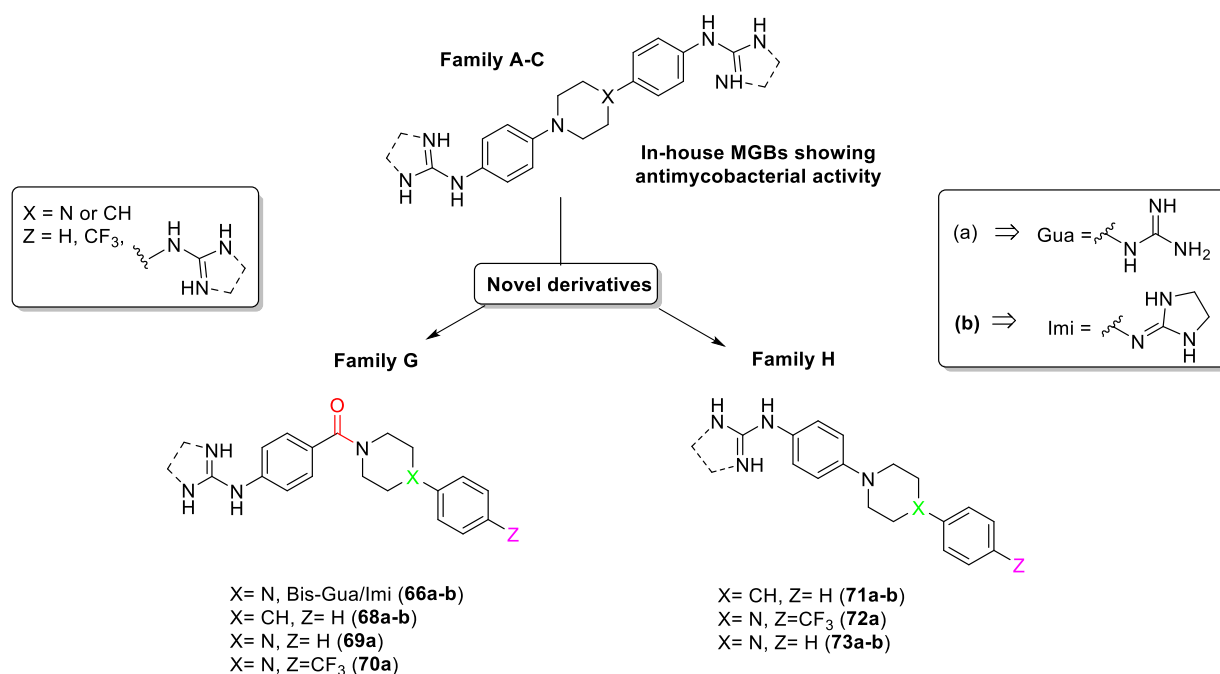


Figure 4.5.1. Scheme showing original compound family prepared by the Rozas group (Family A-C) and newly designed Families G and H.

Synthesis of the piperazine/piperidine based bis-guanidine (**105**) and bis-(2-aminoimidazoline) (**106** and **108**) first involves a nucleophilic aromatic substitution (SNAr) with the commercially available 1-(4-nitrophenyl)piperazine/piperidine to afford the corresponding bis-nitrophenyl intermediates (**74** and **75**). Reduction of the nitro groups via catalytic hydrogenation followed by guanidylation with the use of commercially available 1,3-bis-(*tert*-butoxycarbonyl)-2-methyl-2-thiopseudourea or *N,N'*-bis-(*tert*-butoxycarbonyl)imidazoline-2-thione introduces the Boc protected guanidine or 2-aminoimidazoline moieties (**78-80**). Deprotection with TFA followed by treatment with Amberlyte exchange resin afforded the final hydrochloride salts (**105**, **106** and **108**).

$X = N \text{ or } CH$

74 ($X = N$) (83%)
 75 ($X = CH$) (90%)

76 ($X = N$) (79%)
 77 ($X = CH$) (50%)

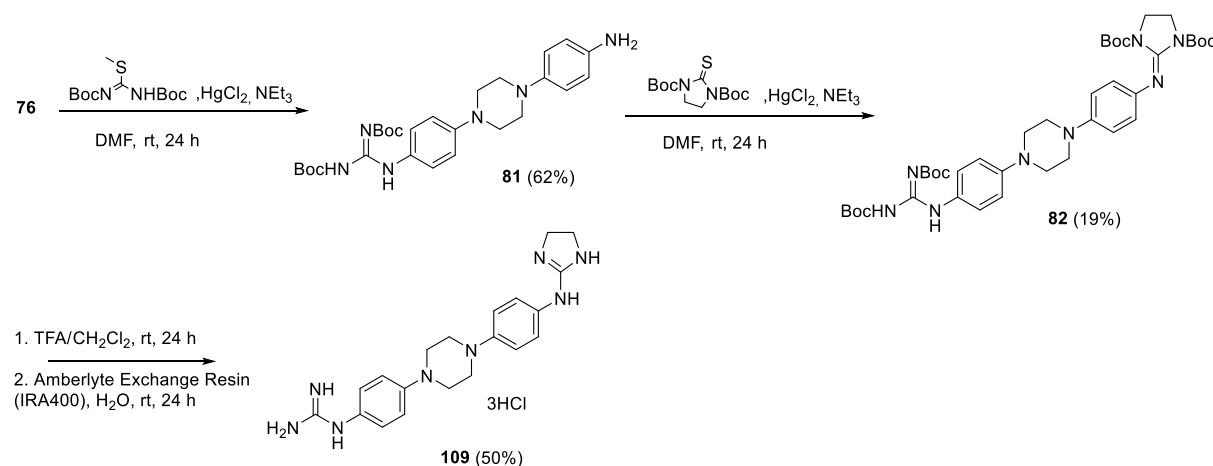
78 ($X = N$, Bis-Boc-Gua) (72%)
 79 ($X = N$, Bis-Boc-Imi) (28%)
 80 ($X = CH$, Bis-Boc-Imi) (28%)

105 ($X = N$, Bis-Gua) (70%)
 106 ($X = N$, Bis-Imi) (65%)
 108 ($X = CH$, Bis-Imi) (28%)

$3HCl$

To synthesise the asymmetric guanidine/2-aminoimidazoline derivatives, guanidylation of one of the terminal amines was performed by varying the equivalents of the reactants. To 1 equivalent of **76**, 0.3 equivalents of 1,3-bis-(*tert*-butoxycarbonyl)-2-methyl-2-thiopseudourea, 0.36 equivalents of HgCl₂ and 1 equivalent of NEt₃ was added. This followed attachment of the 2-aminimidazoline moiety by reaction of the free terminal amine with *N,N'*-bis-(*tert*-butoxycarbonyl)imidazoline-2-thione (Scheme 4.5.1.2). Deprotection was achieved by treatment with TFA followed by Amberlyte exchange resin to afford the final hydrochloride salt (**109**).

Scheme 4.5.1.2. Synthetic route for the production of asymmetric guanidine/2-aminoimidazoline derivatives of Family C.

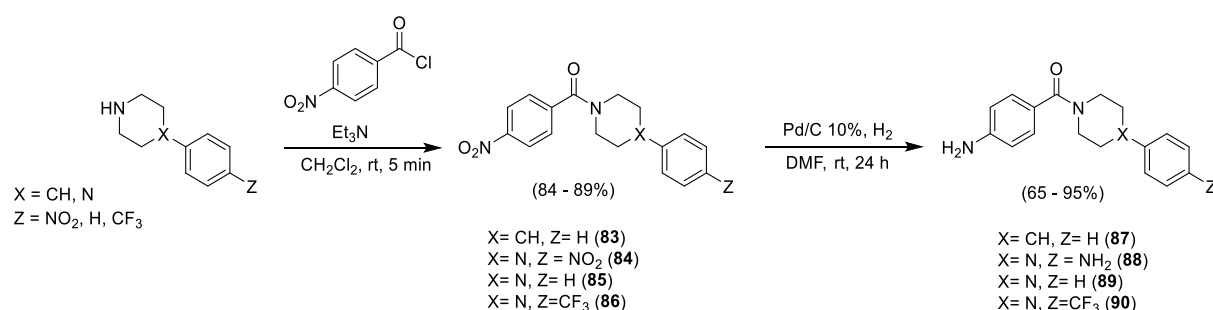


4.5.2. Synthesis of novel families of guanidine based anti-mycobacterial agents

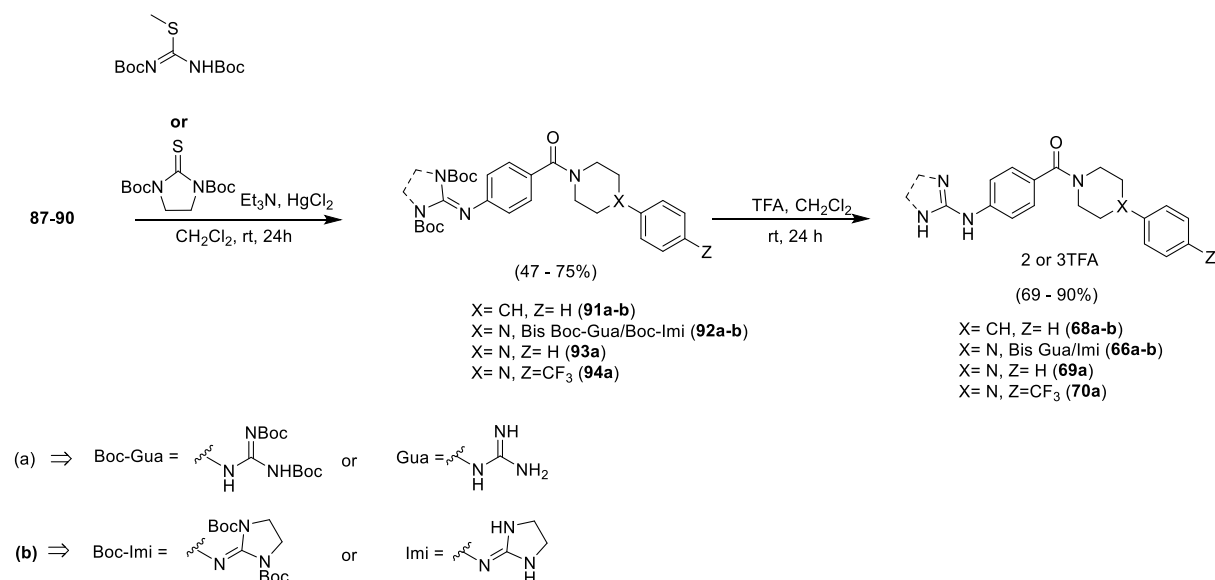
To incorporate the amide linker required for the production of Family G, various methods were considered. Most commonly used methods for the synthesis of amide bonds include the use of acid chlorides through the Schotten–Baumann reaction or the use of coupling reagents such as EDC, DCC, or HOBt. Comparing the yields and reaction times of acid chloride versus coupling reagent-aided amide bond formation with structurally related starting materials within the literature, acid chlorides afforded higher yields and shorter reaction times.^{43,44,45,46} For this reason, amide linked intermediates were synthesised from the commercially available

4-nitrobenzoyl chloride. As shown in Scheme 4.5.2.1, reaction conditions were modified from those reported in the literature by the addition of triethylamine to compensate for the production of HCl in-situ. Reaction of the para-phenyl piperazine/piperidine starting materials afforded the corresponding amides (**83-86**) in good yield (84-89%) where completion of the reaction occurred within 5 minutes. After quick and easy purification by H₂O:chloroform partitioning subsequent reaction involved palladium-catalysed hydrogenation to obtain the desired amines (**87-90**).

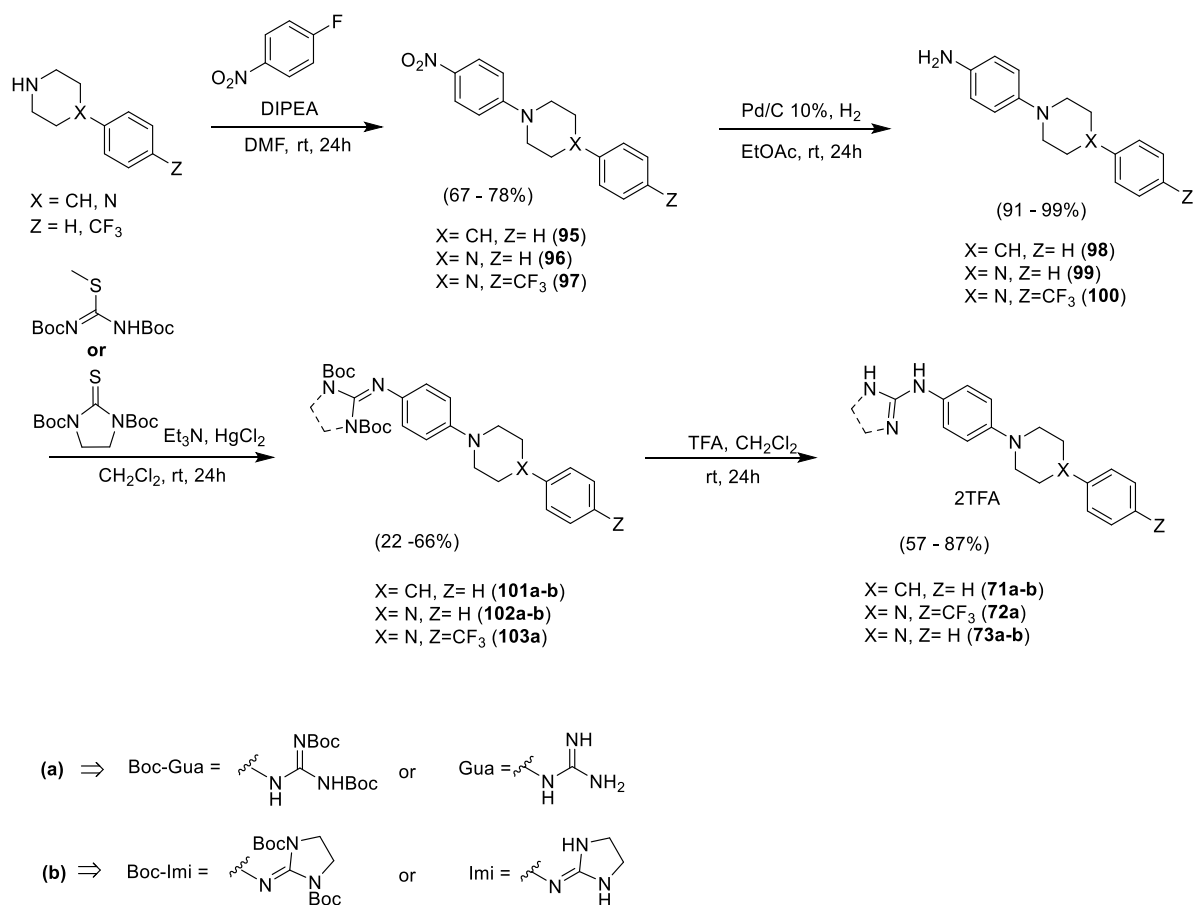
Scheme 4.5.2.1. Amide coupling and hydrogenation reaction to obtain amido-piperazine/piperidine intermediates



Introduction of Boc-protected guanidine and 2-aminoimidazoline moieties was conducted using HgCl₂ mediated guanidylation as illustrated in Scheme 4.5.2.2. 2-Aminoimidazoline derivatives (**91b** and **92b**) were obtained in lower yields due to the increase in steric bulk surrounding the electrophilic thiocarbonyl as well as the instability on silica-based column chromatography during purification. For this reason, the equivalents of NEt₃ and HgCl₂ were increased slightly to accommodate for the decreased electrophilicity and the purification was conducted using alumina-based column chromatography for subsequent reactions. Synthesis of bis-cationic derivatives (**66a** and **66b**) was achieved by adding double the equivalents of starting materials for the incorporation of the guanidine/2-aminoimidazoline moieties. Boc deprotection of intermediates (**91a-b**, **92a-b**, **93a** and **94a**) was conducted using equivolume amounts of TFA and CH₂Cl₂. Purification via H₂O : CH₂Cl₂ partitioning followed by Et₂O trituration resulted in the desired trifluoroacetate salts (**66a-b**, **68a-b**, **69a** and **70a**) in good yield (69-69%) and above 95% purity as identified by HPLC.

Scheme 4.5.2.2. Guanidylation and Boc deprotection reaction to obtain desired TFA salts

For Family H, a similar synthetic route to previously reported Family A and B was adopted (Scheme 4.5.2.3). The first step involves a nucleophilic aromatic substitution with the use of commercially available phenyl piperazine/piperidine derivatives with 1-fluoro-4-nitrobenzene to afford compounds **95-97**. Unlike the $\text{S}_{\text{N}}\text{Ar}$ methodology reported by Dardonville *et al.* and used for the synthesis of previously prepared bis-cationic derivatives (Scheme 4.5.1.1.), compounds **95-97** were synthesised under much milder conditions.^{47,48} Modifying the procedure reported in the literature by adding DIPEA to limit the formation of toxic and potentially dangerous HF, **95-97** were obtained in high yield. Subsequent reduction, guanidylation/2-aminoimidazolation and TFA mediated Boc deprotection produced final salts (**71a-b**, **72a** and **73a-b**) in good yield.

Scheme 4.5.2.3. Synthetic route for the production of Family H compounds

Interestingly, ^1H -NMR analysis of compound **72a** compared to compound **70a** from Family G shows a difference in the nature of the piperazine methylene proton signals. The restricted rotation caused by the implementation of the amide bond causes a broadening and shortening of the signals. For example, the piperazine methylene singlets of **70a** at 3.92, 3.65 and 3.33 ppm are significantly broader and less intense than those of **72a** with two sharp multiplets between 3.45 and 3.37 ppm (Figure 4.5.2.1). This is similar to the phenomenon observed in Chapter 3.2 when comparing the aromatic signals of the final TFA salts with the Boc-protected aminoalkyl intermediates. However, in this scenario, the amorphous nature is retained in the final product due to the nature of the linker.

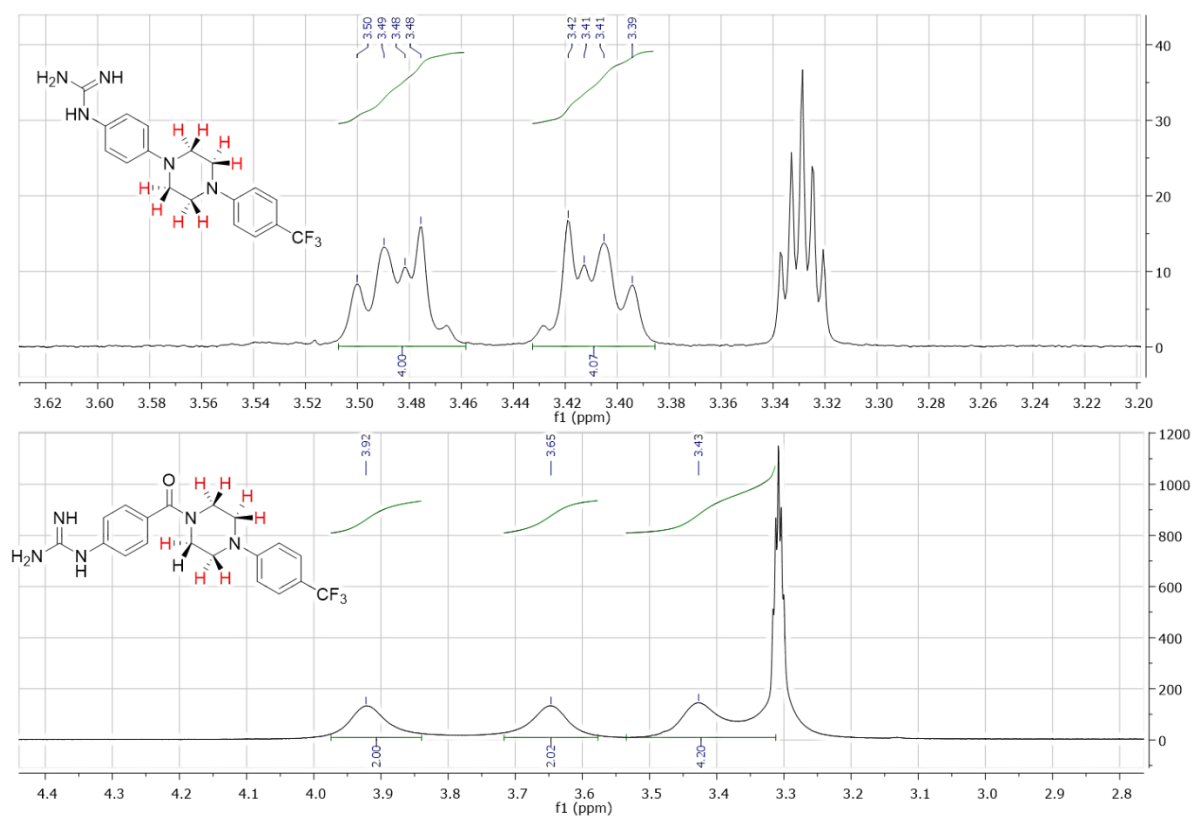


Figure 4.5.2.1. ^1H -NMR of compounds **72a** (top) and **70a** (bottom), showing proton signals of the piperazine linker.

4.6. Biological evaluation

Based on investigations into the activity and mechanism of action of previously made in-house compounds that show good anti-mycobacterial activity, computational studies were performed to optimise these compounds using a structure-based approach. This led to the design of Family G and H as potential InhA inhibitors. To assess the drug susceptibility of these compounds against *M. smegmatis*, REMA analysis was conducted to obtain their respective MICs which can be used to rank the various derivatives in turns of activity. Given the toxicity observed with **105** and **106** in previous *in vivo* studies, the DNA binding affinity of the newly designed compounds was assessed with the aim of creating novel derivatives with no DNA binding affinity and increased InhA inhibition. Moreover, FAMEs and MAMEs TLC analysis was used to verify the mechanism of action by monitoring the effect of drug treatment on mycolic acid biosynthesis. Further comparison of the most active compound from this series against previous compounds and known anti-tubercular agents was conducted using time-kill assays. Lastly, the cytotoxicity of previous and novel series of compounds will be assessed against healthy mammalian cells to determine their safety.

4.6.1. REMA analysis of Family G and H against *Mycobacterium smegmatis*

Analysis of the anti-mycobacterial activity of Family G showed that these amido-piperazine/piperidine derivatives bear similar activity to that seen in previous compounds (Figure 4.6.1.1). Considering compounds **66a** and **66b**, which are the closest derivatives to **105** and **106**, incorporation of the amide bond into the linker of the bis-imidazoline motif is beneficial for activity whereas incorporation into the bis-guanidine motif results in a loss of activity. This is evident from the differences in MIC as shown in Table 4.6.1.1, where an increase in MIC from 60.7 μ M to 139.6 μ M was obtained in the case of **105** versus **66a** and a decrease in MIC from 205.9 μ M to 121 μ M in the case of **106** versus **66b**. Interestingly, the greater activity of **66b** over **66a** is consistent with the results obtained from molecular docking studies. These studies showed that the increased length gained by addition of the amide bond results in the loss of an HB interaction between the terminal guanidinium of **66a** and

backbone residues of the hydrophobic pocket. In the case of **66b**, van der Waals interactions can be established which holds the molecule within the binding site.

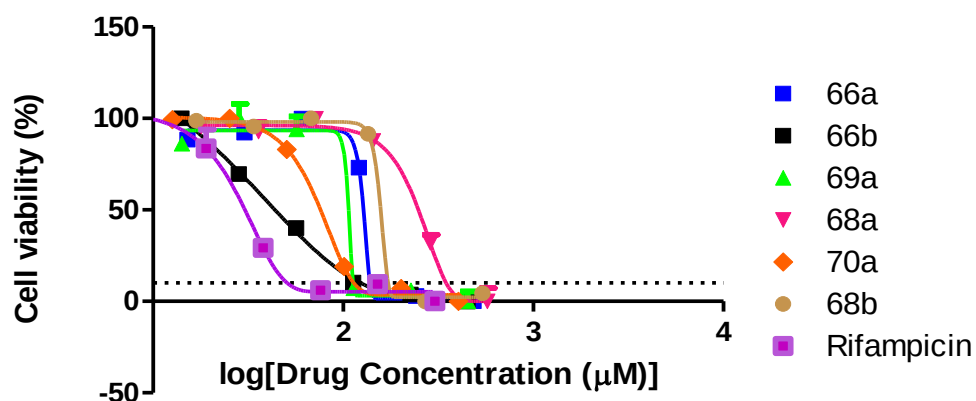
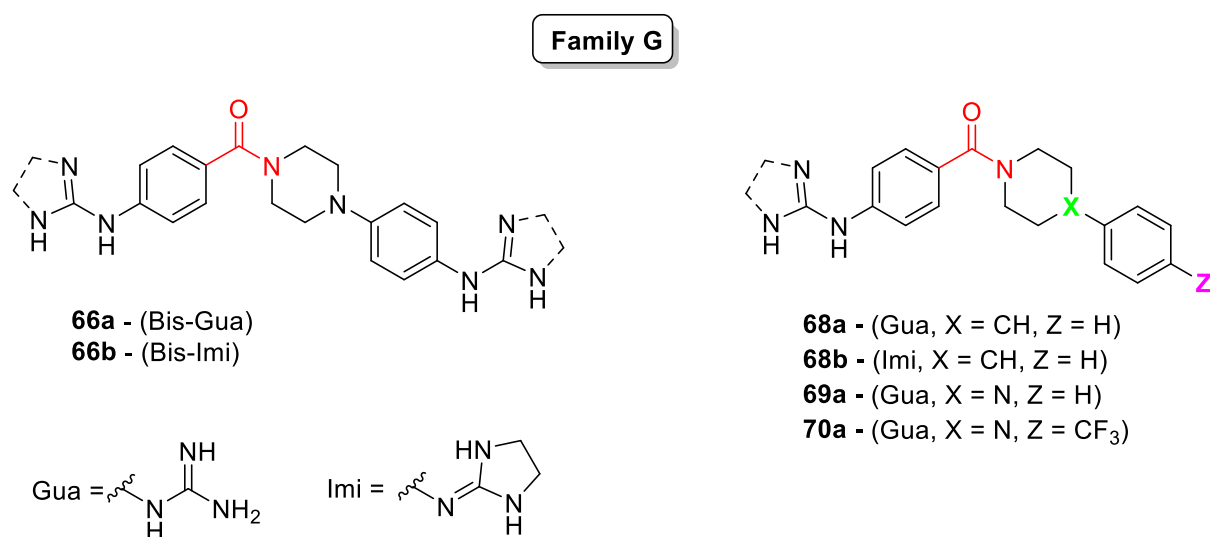


Figure 4.6.1.1. Cell viability curves generated from REMA analysis of compounds **68a**, **68b**, **66a**, **66b**, **69a**, **70a** and rifampicin (the dotted line represents 10% cell viability mark which corresponds to the MIC).

Comparing these di-cationic derivatives to the piperazine linked amphiphilic derivatives **66a** and **66b**, a slight increase in activity was observed with a MIC of 112.8 and 113.9 μM respectively. The lower activity compared to **105** may be explained by the docking output which shows that the cationic “warhead” points away from the inner binding pocket and is actually occupied by the phenyl or p- CF_3 -phenyl portion of the molecule which is less favourable. This indicates that the rigidity created by the amide bond may cause steric hindrance preventing the guanidinium from occupying its desired pocket. The piperidine derivatives **68a** and **68b** displayed the worse activity which is consistent with the results obtained in preliminary screens of in-house compounds.

Table 4.6.1.1. Minimum inhibitory concentrations (MIC) for Family G compounds obtained by REMA analysis using *M. smegmatis*

Entry	Cmp	<i>M. smegmatis</i> MIC (μM)
1	66a	139.6 ±20
2	66b	121 ±20
3	68a	365.9 ±40
4	69a	112.8 ±5
5	70a	113.9 ±7
6	68b	172.9 ±20
7	RIF	50.2 ±7.8

Cells were used at an OD₆₀₀ of 0.1 (~3 × 10⁷ CFU/ mL) in a 96-well plate and treated with a serial dilution of drug compound dissolved in H₂O or DMSO. Once treated, cells were incubated for 3 days at 37 °C then treated with resazurin and left to incubate for 24 h. The resulting fluorescence was read using a plate reader from which percentage viability was calculated. MIC was defined by the drug concentration at 10% cell viability and was calculated using the "Gompertz equation for MIC determination" through GraphPad.

Removal of the amide bond and retention of the amphiphilic motif was shown to result in no significant increments in activity considering compounds **71a**, **71b**, **73a**, **73b** and **72a** of Family H. However, a drastic increase in activity was observed for compound **72a** which shows better activity than **105** and almost identical activity to rifampicin. This indicates the importance of the functionalisation of the aromatic ring at the lipophilic portion of the molecule. Consistent with the observations from our docking studies, functionalisation of the phenyl ring allows for closer contact van der Waals interactions within the hydrophobic pocket which results in tighter binding.

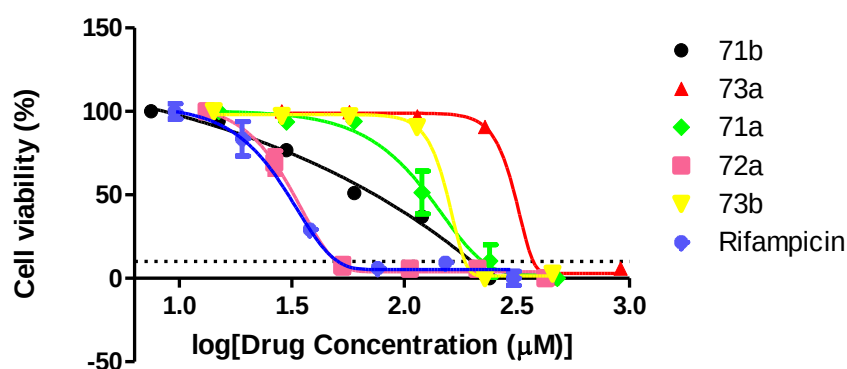
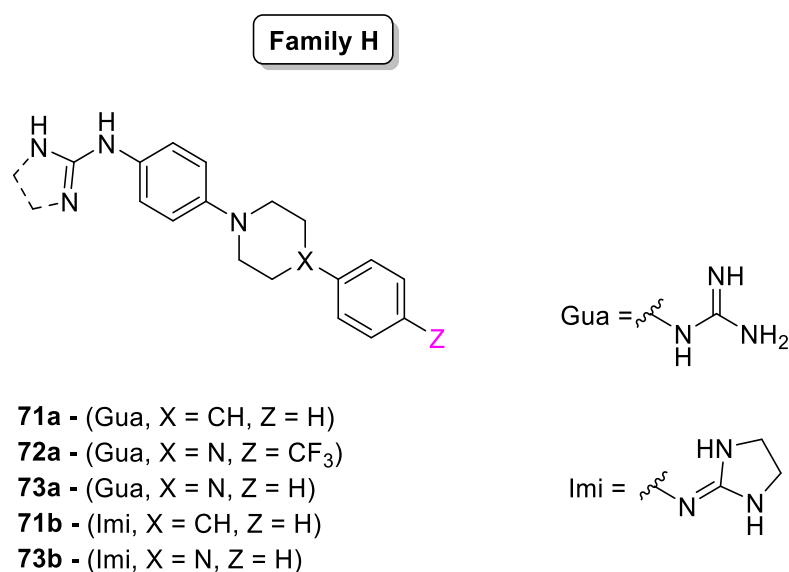


Figure 4.6.1.2. Cell viability curves generated from REMA analysis of compounds **71a**, **71b**, **73a**, **73b**, **72a** and rifampicin (the dotted line represents 10% cell viability mark which corresponds to the MIC).

Table 4.6.1.2. Minimum inhibitory concentrations (MIC) for Family H compounds obtained by REMA analysis using *M. smegmatis*

Entry	Cmp	<i>M. smegmatis</i> MIC (μM)
1	71a	128.5 ±11
2	73a	231.2 ±23
3	72a	49.3 ±10
4	71b	376.4 ±15
5	73b	188.9 ±10
6	RIF	50.2 ±7.8

Cells were used at an OD₆₀₀ of 0.1 (~3 × 10⁷ CFU/ mL) in a 96-well plate and treated with a serial dilution of drug compound dissolved in H₂O or DMSO. Once treated, cells were incubated for 3 days at 37 °C then treated with resazurin and left to incubate for 24 h. The resulting fluorescence was read using a plate reader from which percentage viability was calculated. MIC was defined by the drug concentration at 10% cell viability and was calculated using the "Gompertz equation for MIC determination" through GraphPad.

From these results, a structure-activity relationship (SAR) can be drawn combining the results obtained from preliminary *in vitro* studies and molecular docking studies. As shown in Figure 4.6.1.3, a mono-cationic motif is more favourable where the guanidine results in better activity over the 2-aminoimidazoline. Additionally, the ranking order of linker type in terms of activity follows the order of the piperazine, piperidine, amido-piperazine and amido-piperidine linker. Lastly, the functionalisation of the aromatic ring with a lipophilic functional group is favourable and leads to an increase in activity.

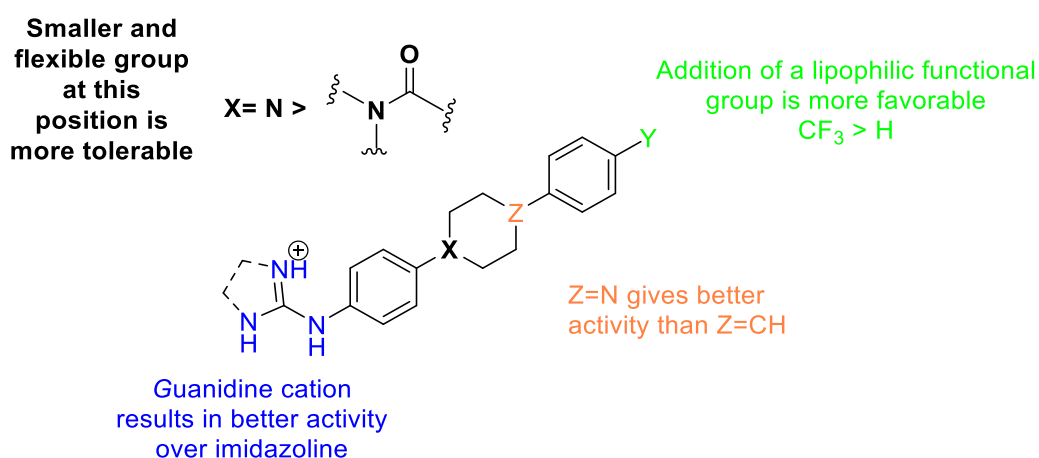


Figure 4.6.1.3. Structure-activity relationship (SAR) generated from REMA studies of Family G and H.

4.6.2. DNA binding of Family G and H

As stated in Section 4.1 **105** and **106** were shown to be toxic after the first treatment at 20 mg/kg in previous *in vivo* studies. To avoid such toxicity, Family G and H were designed to limit their DNA binding capability by modifying the shape and charge of the molecule while increasing InhA binding capacity. To assess the DNA binding properties of this new series of compounds UV thermal denaturation studies were conducted with the use of salmon testes DNA. Analysis of the di-cationic derivatives of Family G showed that **66a** and **66b** bind to DNA with compound **66b** displaying a ΔT_m greater than known MGB agent pentamidine. The ΔT_m values smaller than those of **105** ($\Delta T_m > 20^\circ C$) and **106** ($\Delta T_m > 20^\circ C$) can be explained by the reduced binding interactions caused by the incorporation of the amide bond as shown in the previous section. From this outcome, it can be concluded that the di-cationic motif may not

be optimal for selective InhA binding and may still result in off-target DNA binding toxicity. The lack of DNA binding of the amphiphilic derivatives **68b**, **69b** and **70a** is consistent with the observations from the DNA docking studies and reinforces the importance of an amphiphilic motif.

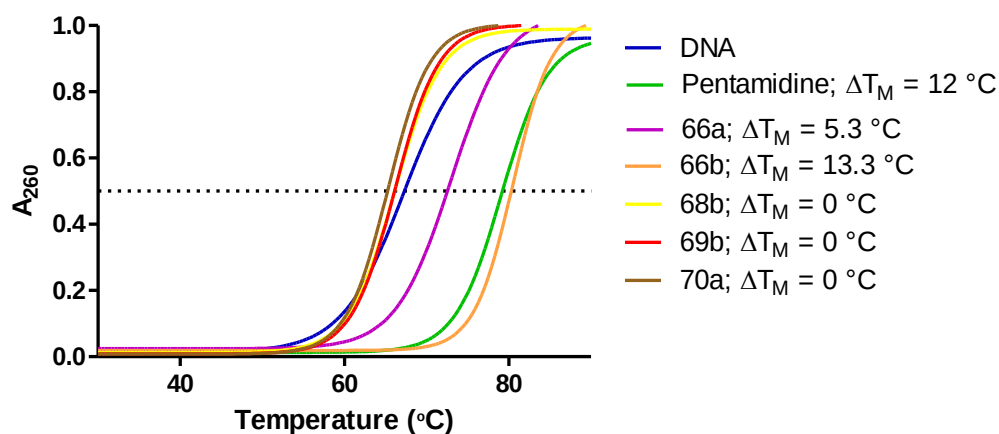


Figure 4.6.2.1. Plot showing the UV thermal denaturation results of compounds **66a**, **66b**, **68b**, **69b**, **70b** and pentamidine using salmon testes DNA in phosphate buffer (10 mM).

Similarly, compounds of Family H, **70a** and **72a**, show no DNA binding affinity as indicated by the ΔT_M values of 0 °C (Figure 4.6.2.2). This suggests that removal of one terminal guanidine alone is enough to significantly reduce the ability of the molecule to bind tightly into the DNA minor groove. The lack of DNA binding and superior anti-mycobacterial activity of **72a** suggests that this compound may serve as an effective anti-tubercular agent and deserves to be investigated further.

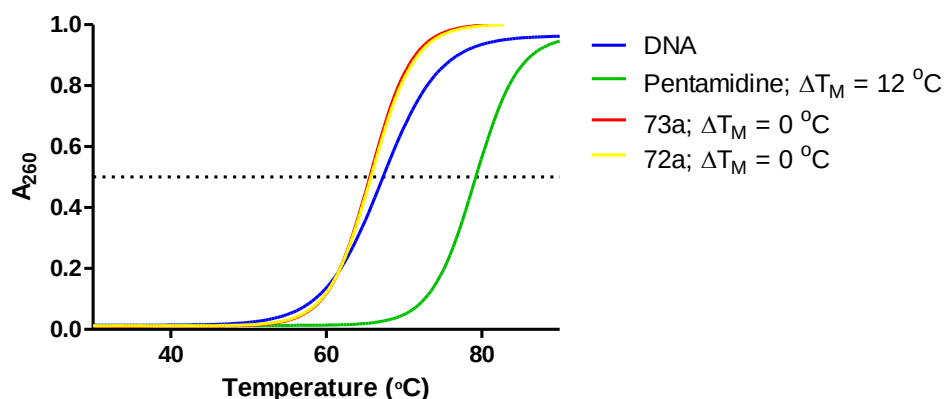


Figure 4.6.2.2. Plot showing the UV thermal denaturation results of compounds **73a**, **72a**, and pentamidine using salmon testes DNA in phosphate buffer (10 mM).

4.6.3. Time-kill assay

Time kill kinetics of the most active compound, **72a**, was analysed in *M. smegmatis*. Using the same methodology utilised to determine the bactericidal or bacteriostatic nature of our previously made compounds, a concentration of 2xMIC (100 μM) was used with an incubation period of 7 days at 37 $^{\circ}\text{C}$. Figure 4.6.3.1 shows the killing profile of **72a** with an overlay of the profiles obtained for **105**, **106**, **109**, rifampicin and tetracycline. **72a** was shown to display a sharp decrease in OD_{600} which was observed after just 24 hours of incubation. A more rapid, potent bactericidal activity than **106**, **109** and rifampicin was observed with complete eradication of the *M. smegmatis* inoculum after 7 days. Similar to **105** and tetracycline, a 1 fold decrease in OD_{600} from the initial inoculum was observed after 7 days of incubation indicating a strong bactericidal mechanism.

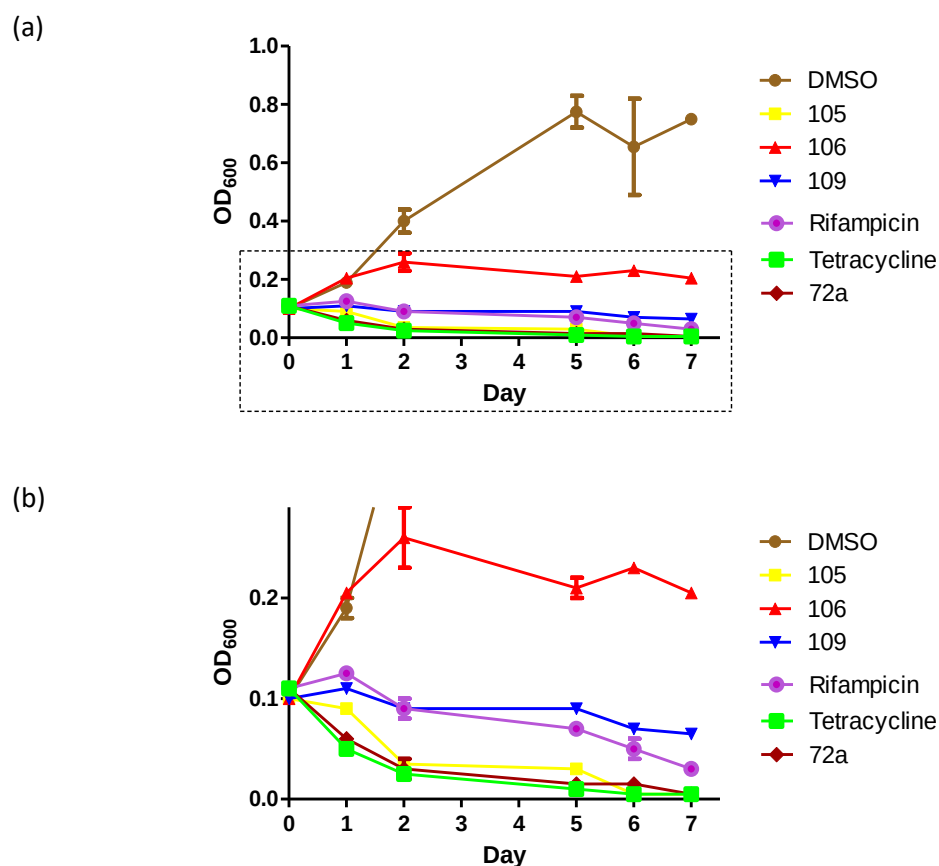


Figure 4.6.3.1. (a) Time-kill kinetics assay of **105**, **106**, **109**, **72a** rifampicin and tetracycline (tested in triplicates at 2× MIC) against *M. smegmatis* over a 7 day incubation period at 37 °C. DMSO (solvent for the compounds) served as a negative control; (b) shows an expansion of (a) in the region of interest as indicated by the dotted box.

As shown in Figure 4.6.3.2, a concentration-dependent killing was observed by comparing the profile of 1xMIC and 2xMIC treatments of **72a**. To probe for greater anti-mycobacterial potency, combination treatments of **72a** with rifampicin and tetracycline were investigated at 1xMIC of each compound. The combination of **72a** and rifampicin displayed slightly better activity than 1xMIC of rifampicin alone but a significant antagonist effect with respect to the activity of 1x and 2xMIC of **72a** was observed. This is similar to the effect seen with the combination of **105** and rifampicin shown in Section 4.1 and is a result of the modification of the *M. smegmatis* mycomembrane upon exposure to rifampicin.¹³ However, the combination of **72a** and tetracycline produced a synergistic effect resulting in a more rapid and potent bactericidal effect than tetracycline and **72a** at 1x and 2x MIC alone. An 80% reduction in

mycobacterial density was observed after just 24 hours of incubation reaching complete eradication after the 7 day incubation period.

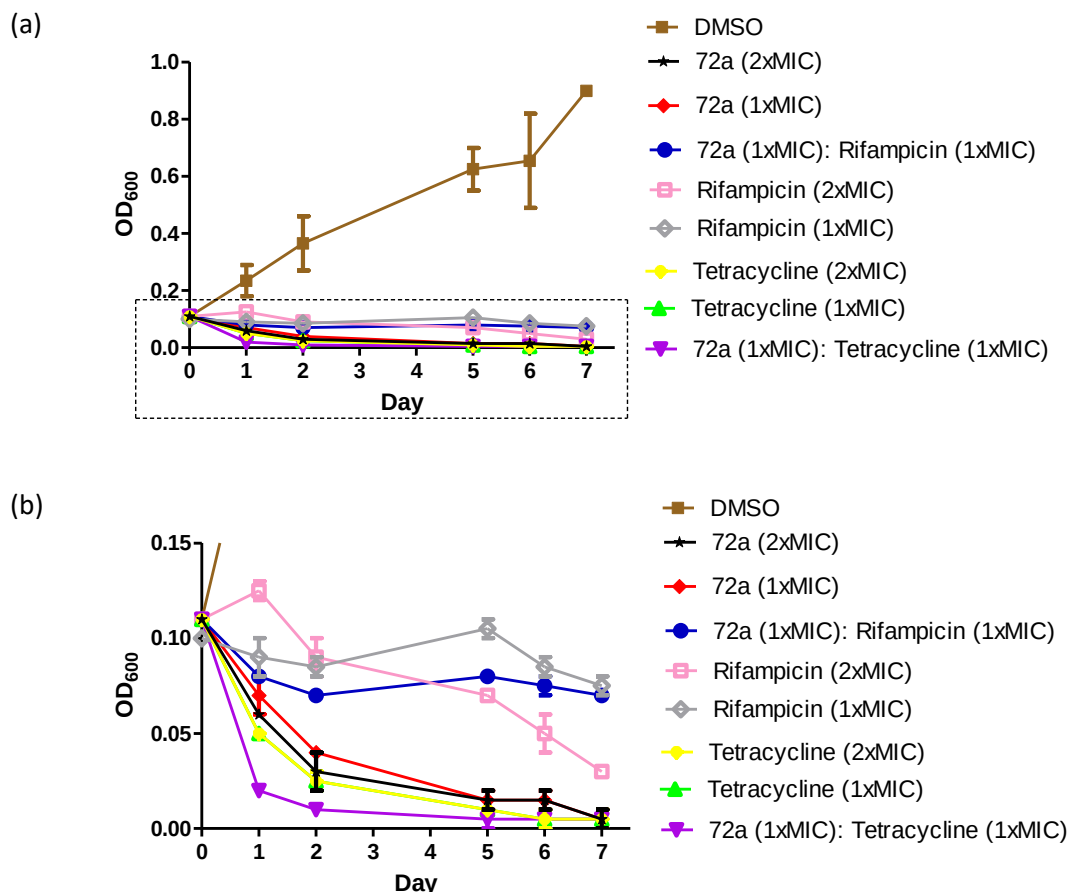


Figure 4.6.3.2. (a) Time-kill kinetics assay of **72a** alone and in combination rifampicin and tetracycline (single treatment at 1x and 2xMIC and combination treatment tested at 1xMIC of each drug) against *M. smegmatis* over a 7 day incubation period at 37 °C. DMSO (solvent for the compounds) served as a negative control; (b) shows an expansion of (a) in the region of interest as indicated by the dotted box.

4.6.4. FAMES and MAMES TLC assays of *M. smegmatis* with compound 72a

Given the positive results obtained for compound **72a**, FAMES and MAMES TLC analysis was performed to confirm its inhibitory effect on mycolic acid synthesis, more specifically, through inhibition of the proposed target InhA. As a structure-based approach was employed in the design of **72a**, it was expected that a more significant effect on mycolic acid synthesis would

be observed. Fortunately, after the treatment of *M. smegmatis* with **72a** at 50 μ M for 8 hours, a huge decrease in alpha (α), alpha prime (α') and epoxy mycolates (MAMEs) was observed (Figure 4.6.4.1 (b)). In addition, after 24 hours, we observed a complete eradication of the mycolates. Given the stark difference between the treatments of **105** and **109** versus **72a** using the same concentrations, it is clear that **72a** exhibits a significantly greater inhibitory effect on mycolic acid biosynthesis. Interestingly, as **72a** is very efficient at killing *M. smegmatis*, after 8 and 24 hours we also see the reduction of the fatty acids (FAMES) which are the FAS I end products.

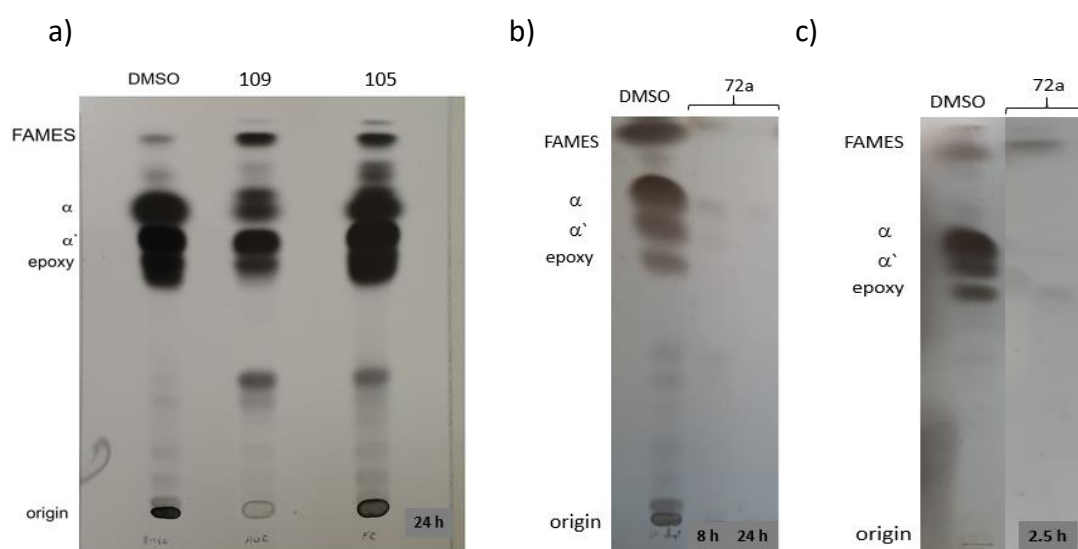


Figure 4.6.4.1. FAMES and MAMEs TLC plate of C-14 acetate labelled *M. smegmatis* lipid extracts; (a) 24 hour treatment with **105** and **109** at 50 μ M; (b) 8 and 24 hour treatment with **72a** at 50 μ M; (c) 2.5 hour treatment with **72a** at 50 μ M.

As stated previously, InhA inhibition is characterised by membrane lysis (or “bleeding”), accumulation of FAMES and reduction in MAMEs. Although **72a** showed membrane lysis and MAMEs reduction, an accumulation of FAMES was not observed. For this reason, a shorter incubation period of 2.5 hours was used which showed that the proportion of FAMES is higher than that of MAMEs, thus, completing the phenotype to verify InhA inhibition (Figure 4.6.4.1 (c)).

4.6.5. Cytotoxicity analysis

To assess the safety of previous in-house compounds (**105**, **106** and **109**) and a subset of novel compounds from Family G (**68a**, **66a** and **69a**) and H (**71a**, **73a** and **72a**), *in vitro* cytotoxicity analysis was conducted with the use of human leukemia monocytic (THP-1) cells. As stated in Chapter 1, the organism that causes TB, *M. tuberculosis* is an intracellular pathogen, and its first target is alveolar macrophages. Hence, as THP-1 cells can be differentiated *in vitro* to macrophages, potential toxicity against alveolar macrophages can be assessed.⁴⁹ As the goal of TB chemotherapy is to selectively target the intracellular microbe without harming the macrophage or any surrounding human cells, low toxicity towards the THP-1 cells is favourable. The cytotoxicity toward these cells were assessed via MTT based assays and was carried by the group of Prof. Lydia Tabernero from the University of Manchester (UK).

As shown in Figure 4.6.5.1, previously produced compounds **106**, **105** and **109** exhibit low toxicity towards THP-1 cells. At a concentration as high as 100 μ M cell survival rates of around 50% was obtained for each compound. Their activity towards *M. bovis* BCG Denmark (MIC = 15.33 μ M (**106**), 0.38 μ M (**105**), 13.3 μ M (**109**)) and lack of toxicity against THP-1 cells suggest that they may be safe for the treatment of drug-sensitive TB. However, the activity of these compounds versus *M. smegmatis* was shown to be significantly lower (MIC = 205.9 μ M (**106**), 60.7 μ M (**105**), 142.6 μ M (**109**)). Hence, given the similar drug sensitivity of drug-resistant *M. tuberculosis* and *M. smegmatis*, these compounds may not be safe for the treatment of MDR-TB due to the high concentrations needed to elicit their anti-mycobacterial activity.

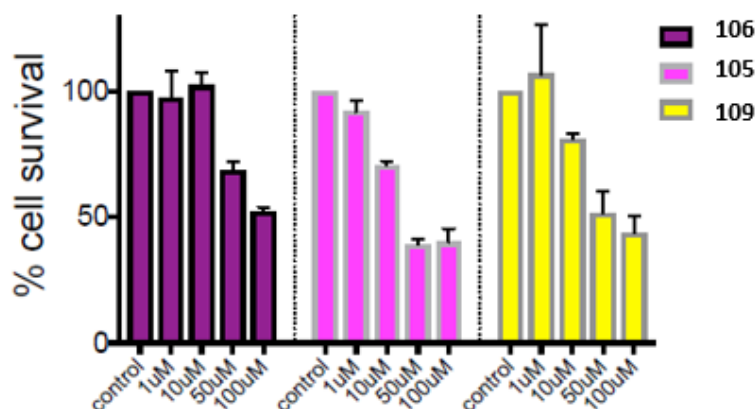


Figure 4.6.5.1. Cell survival graphs of THP-1 cells treated with **106**, **105** and **109** from 1 – 100 μ M for 72 hours.

Novel compounds **71a**, **73a** and **68a** were shown to exhibit lower toxicity than compounds **105**, **106** and **109** with >50% cell survival rates at 100 μ M (Figure 4.6.5.2). This may be attributed to the lack of DNA binding affinity as a result of their amphiphilic motif preventing off-target toxicity. Despite the lower toxicity, the anti-mycobacterial activity of **68a**, **71a** and **73a** is lower than that of **105** which is the most active compound of the previous in-house Family. Moreover, the scaffold of these novel derivatives serve as an optimum starting point for further optimisation.

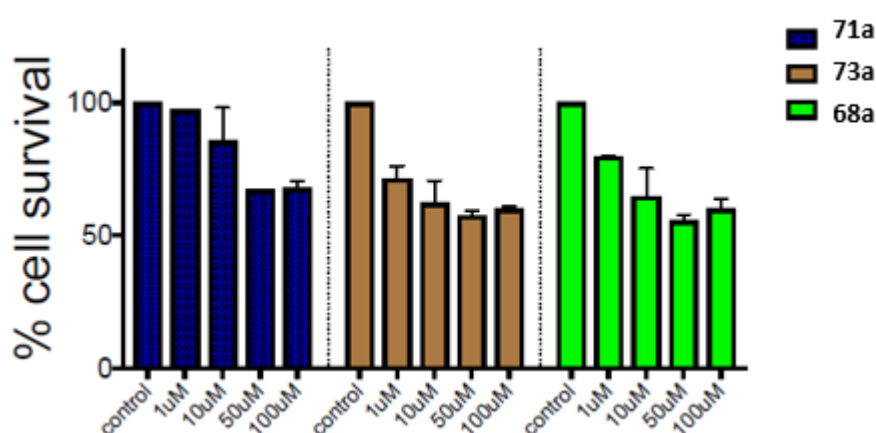


Figure 4.6.5.2. Cell survival graphs of THP-1 cells treated with **71a**, **73a** and **68a** from 1 – 100 μ M for 72 hours.

Surprisingly, compounds **66a**, **69a** and **72a** were shown to exhibit greater toxicity than other compounds tested. From the three compounds, **66a** displayed the greatest toxicity with > 50% cell death at a concentration as low as 10 μ M. For compound **69a**, no cell death was observed at 10 μ M, however, a significant drop below 50% was observed at 50 μ M. **72a** was shown to exhibit a more gradual decline in cell survival with around 50% cell death at 10 μ M. This came as a surprise as **72a** was shown to exhibit no DNA binding affinity ($\Delta T_M = 0$ $^{\circ}$ C) and significantly greater inhibition of bacterial mycolic acid synthesis than previous derivatives. However, a recent study has shown that the C-F bond of aryl- CF_3 containing drugs can be easily cleaved by mammalian drug-metabolising enzymes causing the release of fluoride which is toxic to the cell.⁵⁰ This may explain the toxicity of **72a** given the lack of toxicity observed for **73a** which does not bear the 4-trifluoromethyl group.

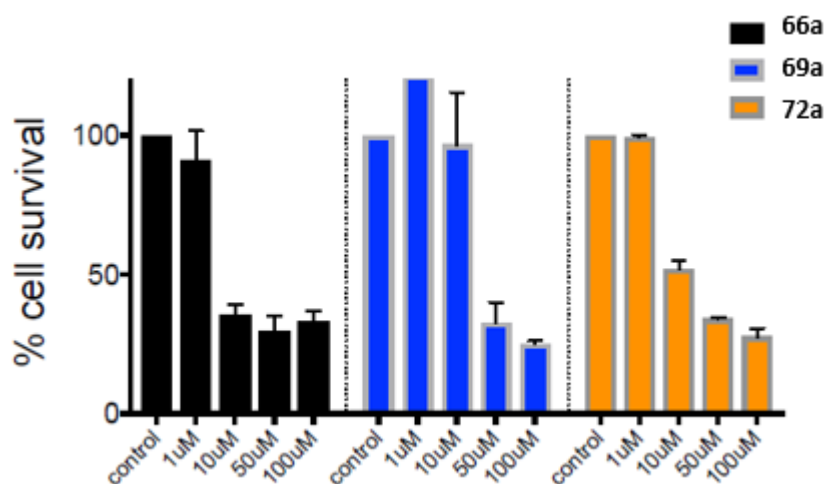


Figure 4.6.5.3. Cell survival graphs of THP-1 cells treated with **66a**, **69a** and **72a** from 1 – 100 μ M for 72 hours.

As previous experiments showed that **72a** was the most potent anti-mycobacterial compound, further investigation into potential toxicity was conducted before considering the compound for further development. Firstly, the IC_{50} of **72a** against *M. bovis* BCG Denmark and *M. smegmatis* was calculated and compared to the IC_{50} against THP-1 cells and another human cell line, HEK293 (Human Embryonic Kidney cells). Figure 4.6.5.4, shows the cell viability curves obtained from the treatment of **72a** in the HEK293 and mycobacterial cell lines. From the THP-1 MTT assays as shown above, an IC_{50} of 10 μ M was extrapolated. Comparing this to the IC_{50} against *M. bovis* BCG and *M. smegmatis*, selectivity indices (SI = IC_{50} human cell/ IC_{50} bacterial cell line) of 1.6 and 0.33 was obtained, respectively. When comparing the IC_{50} values to that obtained for HEK293 cells (IC_{50} = 27.74 μ M), selectivity indices of 4.4 and 0.92 were obtained for *M. bovis* BCG and *M. smegmatis* respectively. These results demonstrate that **72a** is more selective towards *M. bovis* BCG Denmark than *M. smegmatis*. Hence, **72a** may be more suited for the safe and effective treatment of drug-sensitive TB. Noteworthy, treatment of **72a** at lower concentrations with tetracycline has a synergetic effect as shown in the Time-kill assays above. Hence, use of **72a** in combination treatments may provide an alternative method for the treatment of drug-sensitive and MDR-TB while avoiding toxic side effects.

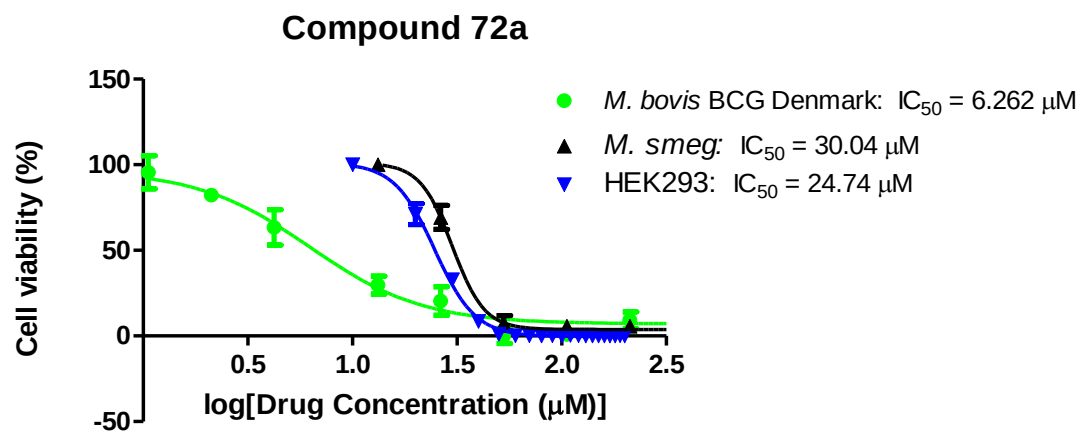


Figure 4.6.5.4. Cell viability curves generated from REMA analysis of **72a** with *M. bovis* BCG Denmark, *M. smegmatis* and HEK293 cells.

4.7. Conclusions

Preliminary drug susceptibility screening of our in-house compounds against surrogate strains of *M. tuberculosis* showed that **105** was the most active compound with comparable activity to the known anti-tubercular agent, rifampicin. Time-kill assays showed that **106** and **109** exhibit a bacteriostatic mechanism whereas **105** displays a bactericidal mechanism. These promising results encouraged the investigation into the molecular target of these compounds to allow for structure-based optimisation. Biochemical experiments by means of FAMES and MAMES TLC analysis showed that these compounds inhibit mycolic acid synthesis through the inhibition of the enzyme InhA. Interestingly, **109** gave a more significant decrease in mycolic acid synthesis despite the greater anti-mycobacterial activity of **105**. This was attributed to the higher DNA binding affinity of **105** which may be acting in conjunction with InhA inhibition.

Using Whole Genome Sequencing, mutations in the ABC transporter genes and the TetR family transcriptional regulator gene were identified in **105** resistant strains of *M. smegmatis* which gave information on how **105** is being imported into the cell and what specific transport system is responsible for its efflux. Unexpectedly, no InhA mutations were identified in the **105** resistant strains which may be due to the possible detriment to cell viability. For this reason, investigations into the possibility of InhA overexpression were conducted through drug susceptibility screening against InhA overexpressors. Surprisingly, **105** was shown to be more active in *M. bovis* BCG InhA overexpressors. Through RT-qPCR analysis of **105** resistant *M. smegmatis* strains, it was discovered that InhA overexpression occurs in the initial exposure to **105**, but due to the detrimental effects on cell viability InhA expression decreases below normal levels and other resistant mechanisms such as reduced uptake and increase efflux initiate.

Following these experiments, ligand-based pharmacophore screening (LBPS) and structure-based molecular docking experiments were performed to identify and understand the key interactions with the proposed drug target, InhA. Through LBPS, the importance of the guanidinium “warhead” was made apparent due to positive hits with InhA inhibitors containing structurally similar polar head groups. By conducting molecular docking experiments, we were able to identify the mode of binding of the previous compounds under investigation. As the InhA active site was found to mainly consist of a polar pocket and an

adjacent hydrophobic pocket, it was concluded that a more amphiphilic ligand would be optimal for binding. This explained the greater mycolic acid inhibition observed for **109** in preliminary biochemical studies. Taking this into account Families G and H was designed using a structure-based approach and docking studies showed that Family H has the best theoretical binding properties with compound **72a** displaying the best binding pose and strongest binding affinity score. In addition, the poor DNA binding poses obtained for these novel families suggests that these compounds may yield a more selective and targeted treatment approach.

Given these promising results, Families G and H were synthesised. The synthesis of the Family G was achieved through a series of steps including amide coupling, catalytic hydrogenation, guanidylation and deprotection to afford the desired TFA salts (**66a**, **66b**, **68a**, **68b**, **69a** and **70a**). Using similar synthetic routes as described for previously made in-house compounds, Family H (**71a**, **71b**, **72a**, **73a** and **73b**) was produced using cheap and commercially available starting materials. All final compounds were produced in high yield and purity.

REMA analysis of novel Families G and H showed comparable activity to previous in-house compounds. Interestingly, compound **72a** from Family H displayed the best activity showing greater anti-mycobacterial activity than rifampicin. Treatment of *M. smegmatis* with compound **72a** was shown to result in a drastic decrease in mycolic acid biosynthesis producing a much greater inhibitory effect than previous compounds. In addition, no DNA binding was observed for **72a** indicating the production of a more selective InhA inhibitor. Time-kill assays showed that **72a** exhibits a bactericidal mechanism and shows a better killing profile than rifampicin and a similar profile to that of **105** and tetracycline. A synergistic effect was also observed upon treatment of **72a** with tetracycline resulting in faster killing kinetics than **72a** or tetracycline alone. However, biological analysis against healthy THP-1 cells showed that **72a** exhibits toxicity at moderate to high concentrations. This was theorised to be due to the release of fluoride upon hydrolysis of the 4-CF₃ group by drug metabolising enzymes which has been previously reported in the literature. However, given some of the very positive results obtained, further development and optimisation of **72a** may yield an effective treatment for tuberculosis.

4.8. References

1. Forge, D.; Cappoen, D.; Laurent, J.; Stanicki, D.; Mayence, A.; Huang, T. L.; Verschaeve, L.; Huygen, K.; Eynde, J. J. Vanden. *Eur. J. Med. Chem.*, **2012**, *49*, 95–101.
2. Richter, A.; Rudolph, I.; Möllmann, U.; Voigt, K.; Chung, C. wa; Singh, O. M. P.; Rees, M.; Mendoza-Losana, A.; Bates, R.; Ballell, L.; Batt, S.; Veerapen, N.; Fütterer, K.; Besra, G.; Imming, P.; Argyrou, A. *Sci. Rep.*, **2018**, *8*, 1–12.
3. Brosch, R.; Gordon, S. V.; Pym, A.; Eiglmeier, K.; Garnier, T.; Cole, S. T., *Int. J. Med. Microbiol.*, **2000**, *290*, 143–152.
4. 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.
5. Shishido, Y.; Mitarai, S.; Otomo, K.; Seki, M.; Sato, A.; Yano, I.; Koyama, A.; Hattori, T., *Int. J. Tuberc. Lung Dis.*, **2007**, *11*, 1334–1338.
6. Mendes, V.; Blundell, T. L., *Drug Discov. Today.*, **2017**, *22*, 546–554.
7. North, E.; Jackson, M.; Lee, R., *Curr. Pharm. Des.*, **2014**, *20*, 4357–4378.
8. Abrahams, K. A.; Besra, G. S., *Parasitology*, **2018**, *145*, 116–133.
9. Jankute, M.; Cox, J. A. G.; Harrison, J.; Besra, G. S., *Annu. Rev. Microbiol.*, **2015**, *69*, 405–423.
10. Morbidoni, H. R.; Weisbrod, T. R., *J. Bacteriol.*, **2000**, *182*, 4059–4067.
11. Bhat, Z. S.; Rather, M. A.; Maqbool, M.; Lah, H. U.; Yousuf, S. K.; Ahmad, Z., *Biomed. Pharmacother.*, **2017**, *95*, 1520–1534.
12. Makarov, V.; Neres, J.; Hartkoorn, R. C.; Ryabova, O. B.; Kazakova, E.; Šarkan, M.; Huszár, S.; Piton, J.; Kolly, G. S.; Vocat, A.; Conroy, T. M.; Mikušová, K.; Cole, S. T., *Antimicrob. Agents Chemother.*, **2015**, *59*, 4446–4452.
13. Man, D. K.-W.; Kanno, T.; Manzo, G.; Robertson, B. D.; Lam, J. K. W.; Mason, A. J., *mSphere.*, **2018**, *3*, 1–19.
14. Lu, W.; Zhang, R.; Jiang, H.; Zhang, H.; Luo, C., *Front. Chem.*, **2018**, *6*, 1–23.
15. Macalino, S. J. Y.; Billones, J. B.; Organo, V. G.; Carrillo, M. C. O., *Molecules*, **2020**, *25*, 1–32.
16. <https://www.ebi.ac.uk/chembl>
17. <https://zinc.docking.org/substances/subsets/fda/>

18. Zhang, M.; Prior, A. M.; Maddox, M. M.; Shen, W. J.; Hevener, K. E.; Bruhn, D. F.; Lee, R. B.; Singh, A. P.; Reinicke, J.; Simmons, C. J.; Hurdle, J. G.; Lee, R. E.; Sun, D., *ACS Omega*, **2018**, 3, 18343–18360.
19. Shimokawa, Y.; Sasahara, K.; Koyama, N.; Kitano, K.; Shibata, M.; Yoda, N.; Umehara, K., *Drug Metab. Dispos.*, **2015**, 43, 1277–1283.
20. Vögeli, B.; Rosenthal, R. G.; Stoffel, G. M. M.; Wagner, T.; Kiefer, P.; Cortina, N. S.; Shima, S.; Erb, T. J., *J. Biol. Chem.*, **2018**, 293, 17200–17207.
21. Shirude, P. S.; Madhavapeddi, P.; Naik, M.; Murugan, K.; Shinde, V.; Nandishaiah, R.; Bhat, J.; Kumar, A.; Hameed, S.; Holdgate, G.; Davies, G.; McMiken, H.; Hegde, N.; Ambady, A.; Venkatraman, J.; Panda, M.; Bandodkar, B.; Sambandamurthy, V. K.; Read, J. A., *J. Med. Chem.*, **2013**, 56, 8533–8542.
22. Rotta, M.; Pissinate, K.; Villela, A. D.; Back, D. F.; Timmers, L. F. S. M.; Bachega, J. F. R.; De Souza, O. N.; Santos, D. S.; Basso, L. A.; Machado, P., *Eur. J. Med. Chem.*, **2015**, 90, 436–447.
23. <https://www.rcsb.org/structure/4BQR>
24. Garcia, P. K.; Annamalai, T.; Wang, W.; Bell, R. S.; Le, D.; Pancorbo, P. M.; Sikandar, S.; Seddek, A.; Yu, X.; Sun, D.; Uhlemann, A. C.; Tiwari, P. B.; Leng, F.; Tse-Dinhi, Y. C., *PLoS One.*, **2019**, 14, 1–18.
25. Poce, G.; Consalvi, S.; Venditti, G.; Alfonso, S.; Desideri, N.; Fernandez-Menendez, R.; Bates, R. H.; Ballell, L.; Barros Aguirre, D.; Rullas, J.; De Logu, A.; Gardner, M.; Ioerger, T. R.; Rubin, E. J.; Biava, M., *ACS Med. Chem. Lett.*, **2019**, 10, 1423–1429.
26. Gurcha, S. S.; Cox, J. A. G.; Mendoza, A.; Jime, E.; Casanueva, R.; Loman, N. J.; Moynihan, P. J.; Izquierdo, C.; Kremer, L.; Bantscheff, M.; Cammack, N. C.; Drewes, G.; Ballell, L.; Barros, D.; Besra, G. S.; Bates, R. H., *Nature Comm.*, **2016**, 7:12581.
27. Foo, C. S.; Lupien, A.; Kienle, M.; Vocat, A.; Benjak, A.; Sommer, R.; Lamprecht, D. A.; Steyn, A. J. C.; Pethe, K.; Piton, J.; Altmann, K.; Cole, T., *A. S. Micr.*, **2018**, 9, 1–13.
28. Altaf, M.; Miller, C. H.; Bellows, D. S.; O'Toole, R., *Tuberculosis*, **2010**, 90, 333–337.
29. Brown, K. J.; Tannock, G. W.; Elliott, R. B.; Lines, D. R.; Tannock, G. W.; Elliott, R. B.; Lines, D. R., *Microbiol. Immunol.*, **1980**, 24, 603–615.
30. <https://github.com/dmnfarrell/snipgenie>
31. Braibant, M.; Gilot, P.; Content, J., *FEMS Microbiol. Rev.*, **2000**, 24, 449–467.
32. Soni, D. K.; Dubey, S. K.; Bhatnagar, R., *Emerg. Microbes Infect.*, **2020**, 9, 207–220.
33. Ramos, J. L.; Marti, M.; Molina-henares, A. J.; Tera, W.; Brennan, R.; Tobes, R., *Microbiol Mol. Biol. Rev.*, **2005**, 69, 326–356.

34. Cuthbertson, L.; Nodwell, R., **2013**, 77, 440–475.
35. Bellinzoni, M.; Buroni, S.; Schaeffer, F.; Riccardi, G.; De Rossi, E.; Alzari, P. M., *J. Bacteriol.*, **2009**, 191, 7531–7537.
36. <https://www.rcsb.org/structure/2V57>
37. Chawla, K.; Martinez, E.; Kumar, A.; Shenoy, V. P.; Sintchenko, V., *J. Glob. Antimicrob. Resist.*, **2018**, 15, 103–104.
38. M. Christopher, A. M. L. S., *Physiol. Behav.*, **2016**, 176, 100–106.
39. Palmer, A. C.; Kishony, R., *Nat. Commun.*, **2014**, 5, 4296
40. Doğan, Ş. D.; Gündüz, M. G.; Doğan, H.; Krishna, V. S.; Lherbet, C.; Sriram, *Eur. J. Med. Chem.* **2020**, 199.
41. He, X.; Alian, A.; Ortiz de Montellano, P. R., *Bioorganic Med. Chem.* **2007**, 15, 6649–6658.
42. Nagle, P. S.; Rodriguez, F.; Kahvedžić, A.; Quinn, S. J.; Rozas, I., *J. Med. Chem.*, **2009**, 52, 7113–7121
43. Espacenet - Bibliographic Data. **2021**, 2012059932.
44. Ding, X.; Stasi, L. P.; Ho, M. H.; Zhao, B.; Wang, H.; Long, K.; Xu, Q.; Sang, Y.; Sun, C.; Hu, H.; Yu, H.; Wan, Z.; Wang, L.; Edge, C.; Liu, Q.; Li, Y.; Dong, K.; Guan, X.; Tattersall, F. D.; Reith, A. D.; Ren, F. *Bioorganic Med. Chem. Lett.*, **2018**, 28 (9), 1615–1620.
45. Heuvel, V.; Den, Johannes, M. A.; Jong, D.; Dossier, G. Espacenet Bibliographic Data : US2018223077 (A1) — 2018-08-09. **2021**, 2018223077, 1–2.
46. Nguyen, T.; Sakasegawa, Y.; Doh-Ura, K.; Go, M. L., *Eur. J. Med. Chem.*, **2011**, 46, 2917–2929.
47. Dardonville, C. and Brun, R. *J. Med. Chem.*, **2004**, 47, 2296–2307
48. Kumari, S.; Mishra, C. B.; Tiwari, M., *Bioorganic Med. Chem. Lett.*, **2015**, 25, 1092–1099.
49. Pick, N.; Cameron, S.; Arad, D.; Av-Gay, Y. *Biol. Proced. Online*, **2004**, 6, 220–225.
50. Pan, Y., *ACS Med. Chem. Lett.*, **2019**, 10, 1016–1019.

Chapter 5

Conclusions and Future Work

5.1. Conclusions: DNA Minor Groove Binders as potential anti-protozoal agents

The objective of this work was to firstly re-evaluate the mechanism of action of a series of guanidine-based MGBs (Family E) previously produced by our group based on miscorrelations between DNA binding affinity and anti-parasitic activity. Additionally, we proposed to synthesise a novel family of MGBs (Family F) which combine key structural features of previous promising compounds. The computational, biophysical and biological studies of this newly designed family of compounds are also detailed in this thesis and have allowed us to make some interesting conclusions.

5.1.1. Computational studies

Molecular docking studies of previous families A (bis-guanidines), B (bis-2-aminoimidazolines), C (asymmetric guanidines/2-aminomidazolines) and D (bis-2-amino-1,4,5,6-tetrahydropyrimidines) gave a greater understanding of why the diphenyl bis-cationic motif is beneficial for DNA minor groove binding as well as the importance of linker geometry. Previous biophysical studies showed that increasing the lipophilicity of the MGBs resulted in an increase in DNA binding and this was attributed to the additional van der Waals interactions formed within the minor groove. Consequently, Family B was shown to display superior DNA binding strength. Unexpectedly, Family D, which is more lipophilic showed poor DNA binding properties. Hence, it was previously concluded that the size and properties of the cationic moieties of these guanidine-like di-aromatic MGBs have an optimal limit which is that of the 2-aminoimidazolinium cation. However, the development of the aminoalkyl derivatives of di-aryl mono-guanidines (Family E) showed that added lipophilicity was tolerated if incorporated as an aliphatic chain. Docking of these compounds within the minor groove showed that the aliphatic tail wraps around the minor groove occupying more space than previous derivatives. In addition, the terminal amine forms HB interactions with the nucleotide bases of the groove as well as electrostatic interactions with the backbone phosphates.

Although these derivatives showed good anti-malarial activity, certain miscorrelations between DNA binding affinity and anti-parasitic activity were observed. For this reason, molecular docking studies were conducted with *pf*DHFR as a potential alternative target. These studies show that short-chained derivatives which were shown to display poor DNA binding exhibit similar binding interactions in the *pf*DHFR active site as WR9921918, a known *pf*DHFR inhibitor.

Taking the desirable structural components of Families A and E gave rise to the design of aminoalkyl derivatives of di-aryl bis-guanidines (Family F). Docking studies in DNA showed that these derivatives form stronger and significantly more HB and electrostatic interactions within the minor groove. This is due to the incorporation of an additional positive charge and reduction of the intercationic distance compared to Family E. In addition, derivatives with longer aliphatic chains displayed greater DNA binding affinity scores which can be explained by the greater number of van der Waals interactions formed. To investigate any potential *pf*DHFR inhibition, a set of compounds were docked into the *pf*DHFR active site. These results showed that similar interactions are formed to those observed for Family E. However, Family F displayed lower DHFR binding affinity scores which may be a result of the bulky *N*-aminoalkyl guanidine moiety which appears to protrude out from the substrate binding pocket. Moreover, these studies showed that this novel family of MGBs has the potential to exhibit very good DNA binding and anti-parasitic properties.

5.1.2. Synthesis

The production of the novel family of aminoalkyl bis-guanidine derivatives (Family F) was initiated by the synthesis of mono-guanidylated di-aromatic intermediates (**38a-d**) with various linkers through conventional HgCl_2 mediated guanidylation in moderate yields (30 - 88%). Higher yields were obtained with the use of the cheap and commercially available 1,3-bis-(*tert*-butoxycarbonyl)-2-methyl-2-thiopseudourea (**50**) due to the better leaving group capacity of the Me-S-Hg adduct formed. Incorporation of the *N*-aminoalkyl substituted guanidine was conducted through a specific thiourea intermediate which was produced by the Mitsunobu reaction (**52**) or isocyanate mediated coupling (**56** or **39a-d**). Reaction of the mono-guanidylated di-aromatic intermediate **38a** with the Mitsunobu product **52** resulted in no desired product formation. However, reaction of **52** with the commercially available 4,4'-

oxydianiline under conventional guanidylation procedure yielded the required protected *N*-aminoalkyl substituted guanidine **53** in moderate yield. Attachment of the Boc-protected guanidine through the free terminal amino group yielded desired Boc-protected intermediate **55** in relatively low yield. Given the low yield obtained, another method was attempted involving the reaction of *tert*-butyl carbamate with an isocyanate intermediate (i.e. **56**) to yield the desired thiourea **57** suitable for guanidylation. However, attempts to yield the desired *N*-aminoalkyl substituted guanidine **45a** from the thiourea **57** and mono Boc-protected alkyl diamine **40a** was futile due to the stereoelectronic properties of the thiourea as indicated by the DFT studies performed. Alternatively, ethyl carbamate derivatives (**39a-d**) of the thioureas mentioned above were synthesised in good yield which gave rise to successful guanidylation upon reaction with the mono Boc-protected alkyl diamines **40a-d**. To facilitate global deprotection, the ethyl carbamate functional group of **41a-d** to **44a-d** was converted to a *tert*-butyl carbamate (**45a-d** to **48a-d**) through the selective carbamate conversion methodology recently reported by our group. Deprotection of the Boc protected *N*-aminoalkyl substituted guanidine intermediates **45a-d** to **48a-d** with TFA yielded the desired trifluoroacetate salts **34a-d** to **37a-d** in good yield (62 - 95%) and purity.

5.1.3. Biophysical experiments

Family F displayed strong DNA binding affinity against stDNA and poly(dA-dT)₂ oligonucleotide in thermal denaturation experiments with some derivatives achieving ΔT_M values >20 °C. Compounds bearing longer alkyl chains and linkers composed of NH or CO show greater binding strength which is in agreement with our computational studies. CD experiments show that these compounds bind to DNA through the minor groove. CO and NH linked derivatives displayed strong ICDs indicating a greater minor groove binding capability. The tetrahedral-based linkers (O and CH₂) resulted in a slight hypsochromic shift in the DNA band (275 nm) indicating conformational changes in the DNA minor groove structure to accommodate the molecules. UV-Vis titration studies of most promising compounds **37c** and **37d** with poly(dA-dT)₂ show spectral changes consistent with strong binding. In addition, the calculated DNA binding constants were greater than that of the previous MGBs developed by our group. These positive results implied that these compounds may exhibit potent anti-parasitic effects.

5.1.4. Biological experiments

The six chosen compounds (**34a**, **34b**, **34c**, **37a**, **37c** and **37d**) from the novel family of aminoalkyl derivatives were analysed for *in vitro* anti-parasitic activity and cytotoxicity. These compounds bearing different linkers and alkyl chain length were tested in three different parasitic cell lines including; *T. b. rhodesiense* STIB 900 cells, *P. falciparum* NF54 cells, *L. donovani* axenic amastigotes. The safety of these compounds were also assessed with the use of L6 skeletal cells. From the cell viability assays, IC₅₀ values were calculated which allowed the ranking of compounds in terms of anti-parasitic activity. By comparing these values to those obtained against L6 cells, the selectivity indices could be calculated. In addition, IC₅₀ values were compared to ΔT_M values obtained from thermal denaturation experiments.

In *T. b. rhodesiense* STIB 900 cells, IC₅₀ values from 0.164 to 7.497 μ M and selectivity indices from 4.35 to 239 were obtained where compound **37d** exhibited the best activity and selectivity. This is consistent with its superior DNA binding affinity obtained from thermal denaturation experiments. A general correlation between DNA binding affinity and anti-trypansomal activity was observed with the exception of the CO-linked derivatives (**12c** and **15c**) which did not follow the trend. This may be due to a possible alternative mechanism of action or reduced importation of the compound into the cell. Interestingly, this is consistent with studies of previous CO-linked MGBs produced within our group.^{1, 2} The aminoalkyl derivatives proved to be more potent and selective towards *P. falciparum* NF54 with IC₅₀ values in the low micromolar – nanomolar range. Compounds **37a**, **34b** and **37d** displayed excellent *in vitro* activity exhibiting IC₅₀ values of 0.039 μ M, 0.046 μ M and 0.062 μ M respectively. These compounds also showed very good selectivity for *P. falciparum* with selectivity indices of 314, 366 and 630 respectively. A similar correlation between DNA binding affinity and anti-parasitic activity was observed against the *P. falciparum* NF54 cells, again, with the exception of CO-linked derivatives **34c** and **37c**. This indicates that strong DNA binding compounds with electron-donating linkers (NH, O, and CH₂) is more favorable for anti-parasitic activity. Unfortunately, all compounds displayed very poor activity against *L. donovani* axenic amastigotes with IC₅₀ values ranging from 48.3 to 124.2 μ M. In addition, these compounds were shown to be poorly selective for the parasite with selectivity indices of less than 1.1. Despite the poor results obtained, some correlation between DNA binding

affinity and anti-leishmanial activity was observed. However, given the promising results obtained against *T. b. rhodesiense* STIB 900 and *P. falciparum* NF54 cells, the aminoalkyl derivatives, especially compounds **37a**, **37d** and **34b**, could be good candidates for further analysis *in vivo* to assess their ability to treat HAT and malaria.

5.2. Conclusions: Guanidine based derivatives as potential anti-tubercular agents

The first objective of this work was to investigate the potential anti-mycobacterial activity of some of our in-house family of compounds based on structural similarities to anti-mycobacterial agents within the literature. Using various computational and biological techniques a drug target was identified which allowed for the design and synthesis of a novel series of guanidine-based derivatives with very good anti-mycobacterial activity.

5.2.1. Computational studies

Computer-aided methods such as ligand-based pharmacophore screening and structure-based molecular docking experiments were employed to identify and understand the key interactions with the proposed drug target. Construction of a 3D pharmacophore and screening against a database of FDA approved anti-tubercular agents identified positive hits against InhA inhibitors. The similarity between the polar head groups of the known anti-tubercular agents and the aryl-guanidinium moiety of our in-house compounds indicates that a polar “warhead” is important for InhA binding. This also explains the higher activity of **105** (bis-guanidinium derivative) over **106** (bis-2-aminoimidazoline derivative). Further insight through molecular docking studies showed that a more amphiphilic motif is beneficial for binding where a polar functional group occupies the inner pocket while the more lipophilic groups interact with the hydrophobic pocket. This is exactly what we see with **109**, where the terminal guanidine forms an HB and a stacked π -cation interaction with NADH as well as a T-shaped π -cation interaction with PHE149. The 2-aminoimidazoline moiety then occupies the hydrophobic pocket forming additional van der Waals interactions. These results correlate with the greater mycolic acid inhibition of **109** observed in preliminary biological studies.

The Identification of InhA as a target led to the design of Families G and H using a structure-based approach. Family H retains the same di-aryl piperazine scaffold as previous compounds with a more amphiphilic motif which has been shown to result in greater InhA inhibition from previous biological studies. Incorporation of an amido-piperazine linker similar to those found in known InhA inhibitors yielded Family G. The resulting geometry of these molecules was anticipated to result in greater binding due to binding pre-organisation as well as potentially introducing new HB interactions within the enzyme's binding pocket. However, despite the pre-organised "L-shaped" conformation, docking experiments showed that the resulting rigidity and increased length causes steric clash with residues of the inner pocket and NAD resulting in unfavourable binding modes. Additionally, no significant interactions between the amide carbonyl and active site residues were observed. Docking of Family H compounds showed a more favourable binding geometry where the guanidinium interacts within the inner binding pocket and the lipophilic portion of the molecule occupies the hydrophobic pocket. From this novel family, compound **72a** proved to be the most promising compound by displaying the best binding pose and strongest binding affinity score. With the aim of creating derivatives with improved InhA binding and reduced DNA binding, molecular docking studies were conducted with a reference crystal structure of the DNA minor groove (PDB: 1D64). The results from these studies show that bis-guanidinium derivative, **66a**, exhibits similar DNA binding interactions and Gscore value to **105** while the amphiphilic derivatives **70a** and **72a** show reduced DNA binding potential. Furthermore, these *in silico* experiments aided in the design of novel derivatives that exhibit promising theoretical anti-mycobacterial properties.

5.2.2. Synthesis

Previously reported symmetric and asymmetric compounds (**105**, **106**, **108** and **109**) were synthesised using methods reported by our group, yielding the required HCl salts in moderate to high yields (28-70%). Noteworthy, the TFA salts prior to the Amberlyte exchange were purified to assess the biological effect of the counter-ion. Negligible differences in the anti-mycobacterial activity between the TFA and HCl salts suggests that the counter-ion does not play a major role. For this reason, TFA salts of newly designed derivatives were utilised for

biological analysis which resulted in higher overall yield due to the removal of the additional ion-exchange step. The synthesis of the amido-piperazine/piperidine derivatives (Family G) was achieved through conventional Schotten Baumann amide coupling, catalytic hydrogenation, guanidylation and deprotection to afford the desired TFA salts (**66a**, **66b**, **68a**, **68b**, **69a** and **70a**) in good yield (69-90%). Using similar synthetic routes as described for previously made in-house compounds, Family H was produced starting from cheap commercially available para-phenyl piperazine/piperidine derivatives. Desired TFA salts (**71a**, **71b**, **72a**, **73a** and **73b**) were produced in moderate to high yield (57-87%) and purity as determined by HPLC.

5.2.3. Biological studies

Initial drug susceptibility screening against *M. bovis* BCG Tokyo showed that the piperazine-linked derivatives (**105**, **106** and **109**) displayed better anti-mycobacterial activity than the piperidine-linked derivatives (**107** and **108**) as indicated by the lower MICs obtained. Further analysis of the most active compounds **105**, **106** and **109** against *M. bovis* BCG Denmark and *M. smegmatis* showed that **105** was the most active compound with a MIC of 0.38 μ M and 60.7 μ M, respectively. Noteworthy, the activity of **105** against *M. smegmatis* is comparable to that of known anti-tubercular agent rifampicin (MIC = 50.2 μ M). Given these promising results investigations into the mechanism of action of these compounds was performed.

Miscorrelations between the previously conducted thermal denaturation results and anti-mycobacterial activity suggested that an additional target may be involved. Structural similarities to anti-tubercular agents targeting cell wall biosynthesis through mycolic acid synthesis and DprE1 inhibition led to the use of whole-cell and enzyme-based assays to explore these targets. Treatment of *M. smegmatis* with **105** and **109** was shown to result in cell lysis through membrane “bleeding”. In addition, FAMES and MAMES TLC analysis showed that both compounds caused an accumulation of FAS I end products and a reduction in mycolic acids which is consistent with InhA inhibition. These experiments showed that treatment with **109** results in a more significant decrease in mycolic acid synthesis despite the lower MIC of **105**. This was attributed to the higher DNA binding affinity of **105** which may be acting in conjunction with InhA inhibition. Moreover, the scaffold for **109** may be more optimal for targeting InhA alone. As no significant inhibition of DprE1 was identified through

enzyme-based assays, it was concluded that these compounds are targeting cell wall biosynthesis mainly through InhA inhibition.

Time-kill assays showed that **106** and **109** exhibit a bacteriostatic mechanism whereas **105** displays a bactericidal mechanism. Combination treatment of **105** with rifampicin resulted in an antagonistic effect where the activity was lower than that of **105** alone. This was suggested to be due to the membrane remodelling that has been shown to occur in *M. smegmatis* upon treatment with rifampicin, thus, affecting the importation of **105** into the cell. However, treatment with tetracycline exhibited a synergistic effect where an increase in killing kinetics compared to **105** and tetracycline alone was observed. These promising results indicate that this family of compounds deserved to be investigated further and used as a scaffold for the design of novel anti-tubercular agents.

Using whole-genome sequencing, drug-resistant mutations were identified that informed us on how **105** crosses the bacterial cell as well as how it is shuttled out of the cell in resistant mutants. This involved the use of the fast-growing *M. smegmatis* and a liquid media continual passaging approach. Using this method, mutations in the ABC transporter genes and the TetR family transcriptional regulator gene were identified. This gave information on how **105** is being imported into the cell and what specific transport system is responsible for its efflux. Due to the absence of mutants in the proposed target, DNA extracts from a solid media single-shot approach were sequenced to compensate for any possible bottleneck effects through continual passaging. Failure to identify any InhA mutations with these approaches led to the investigation into the possibility of InhA overexpression. Surprisingly, drug susceptibility assays showed that **105** was more active in *M. bovis* BCG InhA overexpressors. Through closer inspection, it was concluded that the toxic activity arose from the drug-bound target, rather than a lack of normal target activity leading to increased susceptibility. Thus, InhA overexpression negatively impacts cell viability in the presence of **105**. Through gene expression analysis it was discovered that InhA overexpression occurs initially in the presence of **105**. However, due to the detrimental effects of this on cell viability, InhA expression decreases below normal levels and other resistant mechanisms such as reduced uptake and increase efflux occurs. Cross-resistance evaluation showed that **105** resistant mutants were also resistant to tetracycline but minimally affected the activity of rifampicin. This highlights

the possibility of being able to tailor alternative treatments or design new agents that evade drug resistance before the drug is marketed.

REMA analysis of novel Families G and H showed comparable activity to previous in-house compounds which were originally designed as anti-parasitic DNA MGBs. Based on these results an experimental SAR was drawn which showed that a mono-cationic amphiphilic motif was favoured with compounds bearing a piperazine linker exhibiting better activity. From the compounds tested, **72a** displayed the best activity with a MIC (MIC = 49.3 μ M) greater than that of **105** and known first-line anti-TB agent rifampicin. With the aim of creating new derivatives that do not bind DNA, amphiphilic derivatives **68b**, **69b**, **70b**, **72a** and **73a** showed no DNA binding affinity. Additionally, the most active compound, **72a**, was shown to exhibit a drastic increase in mycolic acid inhibition compared to previous compounds. This indicates a superior affinity towards the proposed target, InhA. These findings support the conclusions drawn from the computational and preliminary biochemical studies.

Time-kill assays showed that **72a** exhibits a bactericidal mechanism and shows a better killing profile than rifampicin and a similar profile to that of **105** and tetracycline. Combination treatment of **72a** with rifampicin showed an antagonistic effect as a result of reduced anti-mycobacterial activity compared to **72a** alone. This was attributed to findings within the literature that shows that rifampicin exposure causes alteration to the mycomembrane of *M. smegmatis* which affects the transport of antimicrobials into the cytosol. However, a synergistic effect was observed upon treatment of **72a** with tetracycline resulting in faster killing kinetics than **72a** or tetracycline alone. To assess the safety of these compounds, MTT assays were conducted with THP-1 cells which showed that both previous compounds and some derivatives from the newly designed family of compounds display minimal toxicity. Surprisingly, toxicity was observed with high concentrations of **72a** which may be attributed to the release of fluoride upon hydrolysis of the 4-CF₃ group by drug metabolising enzymes. When comparing the IC₅₀ values of **72a** against the bacterial strains versus healthy human cells (THP-1 and HEK293), **72a** was shown to be selective towards *M. bovis* BCG *Denmark* but poorly selective for *M. smegmatis*. Given the phenotypic similarities, **72a** may display a greater safety profile for the treatment of drug-sensitive TB over MDR-TB. Moreover, further development of **72a** may yield an effective treatment against tuberculosis.

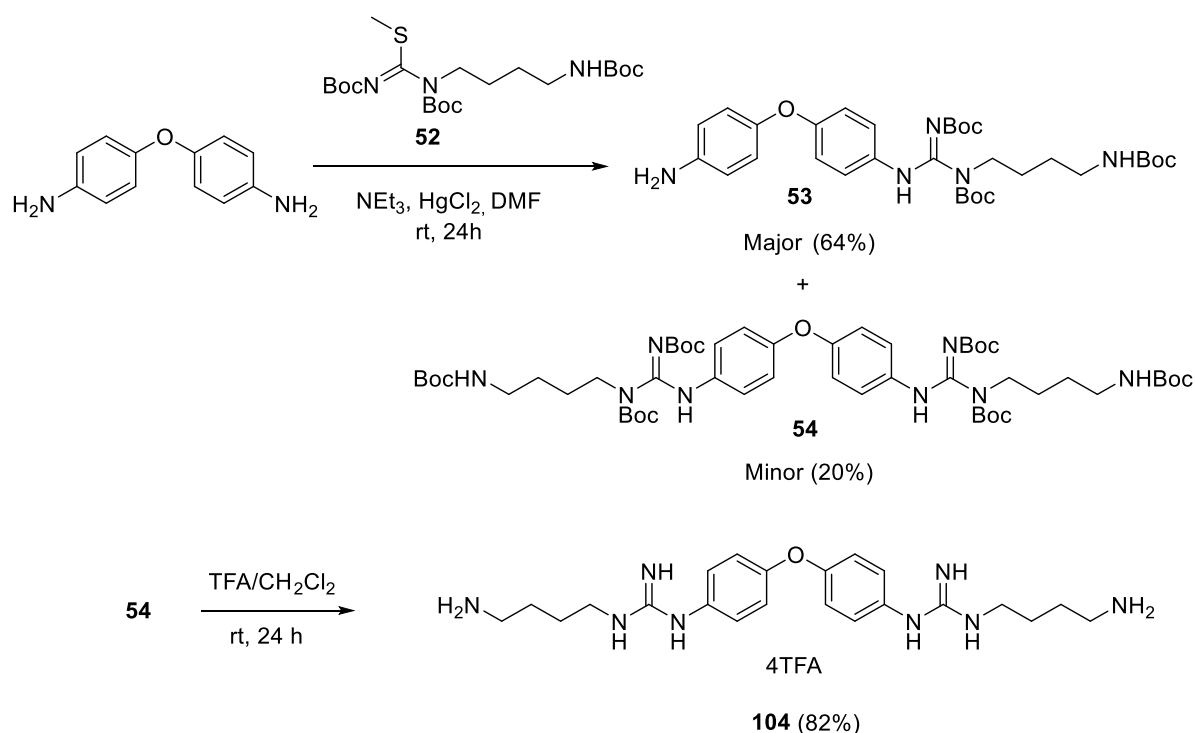
5.3. Future work: DNA Minor Groove Binders as potential anti-protozoal agents

The results obtained and presented in this work were very interesting and promising. This has led to a number of ideas for possible future work.

5.3.1. Synthesis

In an attempt to install the *N*-aminoalkyl guanidine required to produce the proposed novel MGBs, compound **54** was obtained as a minor side-product upon guanidylation of 4,4'-oxydianiline with compound **52** (Scheme 5.3.1.1).

Scheme 5.3.1.1.



Deprotection of **54** using TFA yielded the bis-*N*-aminoalkyl guanidine product **104** which was subject to UV thermal denaturation studies with stDNA. A ΔT_M value of $>20^\circ\text{C}$ was obtained which is greater than the corresponding mono *N*-aminoalkyl guanidine derivative (**34a**; $\text{X} = \text{O}$,

$n = 2$). Therefore, using a similar route as shown above it would be interesting to synthesise other bis-*N*-aminoalkyl guanidines with varying linkers and alkyl chain length to access new MGBs with potentially increased DNA binding strength and anti-parasitic properties (Figure 5.3.1.1).

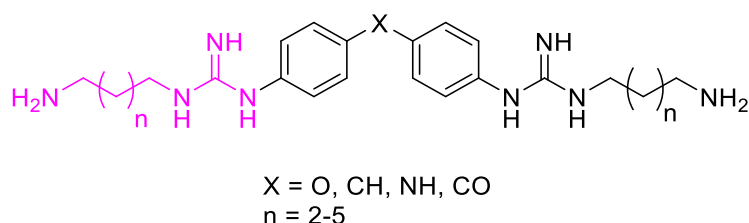


Figure 5.3.1.1. Bis-*N*-aminoalkyl guanidines as potential MGBs

Secondly, the substitution of a guanidine with a 2-aminoimidazoline moiety has resulted in the production of stronger MGBs in previous studies conducted in our group. For this reason, it would be interesting to substitute the terminal guanidine of Family F with a 2-aminoimidzoline to further improve their DNA binding affinity. The synthesis of these derivatives would involve the same synthetic route as reported for the production of Family F with the use of Boc protected thione instead of the commercially available 1,3-bis-(*tert*-butoxycarbonyl)-2-methyl-2-thiopseudourea (**50**).

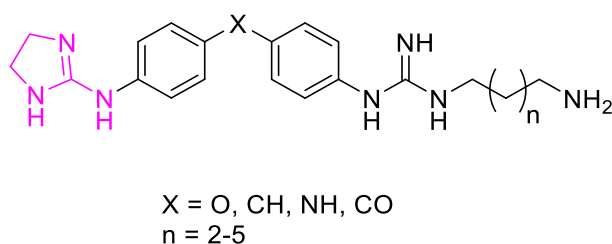


Figure 5.3.1.2. Asymmetric 2-aminoimidazoline/*N*-aminoalkyl guanidines as potential MGBs

5.3.2. Biophysical studies

Firstly, biophysical analysis with novel Family F could be carried out using a GC-rich sequence such as poly(dG-dC)₂ DNA. This would give a better indication of their selectivity for the minor groove over the major groove. Additionally, more sophisticated techniques such as Isothermal Calorimetry (ITC) and Surface Plasmon Resonance (SPR) could be carried out to determine more precise DNA binding affinities and kinetics.

5.3.3. Biochemical and biological studies

The computational studies of previously reported Family E showed that they could also be targeting *pfDHFR*. To confirm these findings enzyme-based assays need to be conducted. Due to the limited availability of *pfDHFR* containing drug screening kits, these experiments are commonly conducted with the use of expression systems such as *E. coli* to produce *pfDHFR*.³ Following on from this, standard colorimetric assays can be performed to obtain IC₅₀ values.

Additionally, it would be interesting to explore the *in vivo* anti-malarial and anti-trypanosomal activity of compounds **37a**, **37d** and **34b** which gave great results in *in vitro* experiments. Common *in vivo* malaria models for drug screening involves the use of female National Medical Research Institute (NMRI) mice that have been intravenously infected with *P. berghei*. In terms of HAT, two acute stage models STIB 795 and STIB 900 are commonly used which are mice infected with *T. b. brucei* and *T. b. rhodesiense* respectively.⁴ In addition, effectiveness against late-stage HAT can be screened for using the GVR 35 model which is created by the infection of the mouse CNS with *T. b. brucei*.⁵

5.4. Future work: Guanidine based derivatives as potential anti-tubercular agents

This work has involved the use of various interesting methods and technologies and has consequently yielded very promising results. Some future work can be done to potentially develop new and improved anti-tubercular drugs.

5.4.1. Synthesis

To further optimise the most active compound **72a** various structural modifications could be made (Figure 5.4.1.1). As stated previously, the investigation into our in-house family compounds was influenced by the work of Forge *et al.* who produced structurally similar diaryl piperazines with proven anti-mycobacterial activity.⁶ It was observed that the *N*-alkyl derivatives (**13-18**) produced better *in vitro* activity than Isoniazid against drug-sensitive *M. tuberculosis* strain H37Rv and very good activity against drug-resistant clinical isolates of *M. tuberculosis*. Hence, it would be interesting to see the effect of introducing alkyl chains of various lengths to the guanidine of compound **72a**. Secondly, molecular docking studies of **72a** in InhA has shown that the hydrophobic pocket could be further probed by additional functionalisation of the aromatic ring with hydrophobic functional groups. In addition, the replacement of the 4-CF₃ group with another lipophilic group such as a methyl may improve the safety profile of **72a** by avoiding the suspected fluoride mediated toxicity as a result of CF₃ hydrolysis.

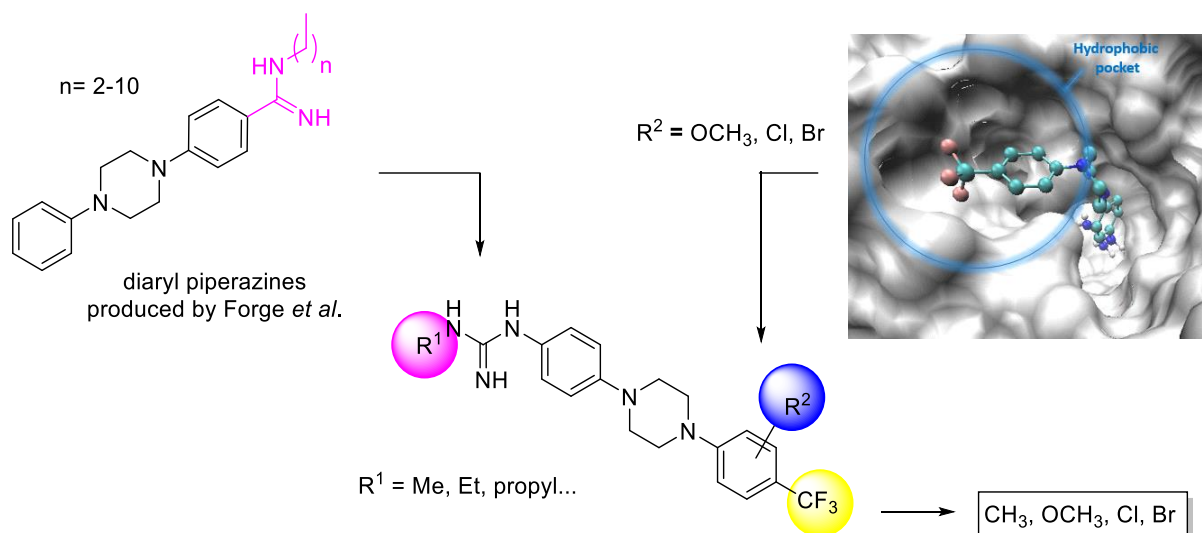


Figure 5.4.1.1. Potential modifications for the optimisation of compound **72a**

5.4.2. Biochemical and Biological studies

Similar to the WGS experiments conducted with **105** it would be interesting to perform these experiments with compound **72a** to determine if similar drug-resistant mechanisms are activated. In addition, the use of InhA enzyme-based assays would provide a more direct approach to quantify InhA inhibition.

For further biological assessment of compound **72a**, *in vitro* analysis against drug-sensitive and drug-resistant *M. tuberculosis* strain H37Rv should be conducted. In addition, as *M. tuberculosis* is an intracellular microbe, performing drug susceptibility assays against H37Rv infected macrophages would determine the ability of compound **72a** to cross the macrophage membrane and target the bacteria.

5.5. References

1. Dardonville, C.; Brun, R., *J. Med. Chem.*, **2004**, 47 (9), 2296–2307.
2. McKeever, C.; Kaiser, M.; Rozas, I., *J. Med. Chem.*, **2013**, 56 (3), 700–711.
3. Bunyarataphan, S.; Leartsakulpanich, U.; Taweethai, S.; Tarnchompoo, B.; Kamchonwongpaisan, S.; Yuthavong, Y. *Antimicrob. Agents Chemother.*, **2006**, 50, 3631–3637.
4. Koch, O.; Jäger, T.; Flohé, L., *Trypanosomatid Diseases: Molecular Routes to Drug Discovery*, **2013**, 13-22
5. Rodríguez, F.; Rozas, I.; Kaiser, M.; Brun, R.; Nguyen, B.; Wilson, W. D.; García, R. N.; Dardonville, C., *J. Med. Chem.*, **2008**, 51 (4), 909–923.
6. Forge, D.; Cappoen, D.; Laurent, J.; Stanicki, D.; Mayence, A.; Huang, T. L.; Verschaeve, L.; Huygen, K.; Eynde, J. J. Vanden. *Eur. J. Med. Chem.*, **2012**, 49, 95–101.

Chapter 6

Experimental

6.1. Synthetic Chemistry

6.1.1. Materials and Methods

All commercial chemicals were obtained from Sigma-Aldrich or Fisher and were used without further purification. Dry solvents were obtained from PureSolvTM solvent purification system (Innovative Technology, Inc.) or otherwise purchased from Sigma-Aldrich or Fisher. Solvents for synthesis purposes were used at HPLC grade. Chromatographic columns were run using Silica gel 60 (230-400 mesh ASTM). Analytical TLC was performed using Merck Kieselgel 60 F₂₅₄ silica gel plates or aluminium oxide plates. Visualisation was by UV light (254 nm) or by staining with ninhydrin. NMR spectra were recorded in a Bruker DPX-400 Avance spectrometer, operating at 400.13 MHz and 600.1 MHz for ¹H-NMR and 100.6 MHz and 150.9 MHz for ¹³C-NMR. Shifts are referenced to the internal solvent signals. NMR data were processed using the MestreNova software. The following abbreviations (or combinations thereof) were used in NMR characterisation: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br. s = broad singlet, H = proton, C = carbon, Ar = aromatic proton or carbon, qC = quaternary carbon. Compounds are assigned arbitrary numbers for assignment of the NMR spectrum and these numbers do not correspond to the naming of the compound.

HRMS spectra were measured on a Micromass LCT electrospray TOF instrument with a WATERS 2690 autosampler with methanol 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.

6.1.2. General Procedures

Method A: *General procedure for the mono-Boc protection of the alkyl diamines*

To a suspension of diamine (1 equiv.) in dry CH₂Cl₂ at 0 °C was added a solution of di-*tert*butyl-dicarbonate (0.5 equiv.) in CH₂Cl₂ (10 mL) dropwise over 30 min. After stirring for 5 hours

reaching room temperature, the mixture was filtered under suction and washed through with CH_2Cl_2 . This solution was then washed with H_2O (2x 20 mL), brine (20 mL), dried over MgSO_4 anh. and purified by NEt_3 treated (5%) silica gel column chromatography eluting with a $\text{CH}_2\text{Cl}_2/\text{MeOH}/\text{NEt}_3$ mixture.

Method B: *General method for the preparation of mono-functionalised guanidines*

To a solution of dianiline (1 equiv.) in CH_2Cl_2 , 1,3-bis-(*tert*-butoxycarbonyl)-2-methyl-2-thiopseudourea (0.3 equiv.) and NEt_3 (1.03 equiv.) at 0 °C was added HgCl_2 (0.36 equiv.). This cloudy mixture was stirred at 0 °C for 30 min followed by room temperature overnight. The reaction mixture was then diluted with CH_2Cl_2 , filtered through a path of Celite and rinsed through with CH_2Cl_2 . The organic phase was then washed with brine, dried over MgSO_4 anh. and concentrated under reduced pressure. The crude residue was then purified by silica gel column chromatography eluting with a hexane/EtOAc mixture.

Method C: *General method for the preparation of ethyl-carbamate protected thiourea intermediates*

To a solution of amine (1 equiv.) in dry CH_2Cl_2 (8 mL) was added ethoxycarbonyl isothiocyanate (ECIT) (1 equiv.) at 0 °C and allowed to stir overnight reaching room temperature. Upon completion, the reaction mixture was reduced and purified by silica gel column chromatography (hexane/EtOAc) followed by crystallisation from Et_2O .

Method D: *General method for the preparation of substituted guanidines*

To a solution of thiourea intermediate (1 equiv.) in dry CH_2Cl_2 (5 mL) was added a solution of DIPEA (1.5 equiv.) and mono-Boc protected alkyl diamine (1.5 equiv.) in CH_2Cl_2 (2 mL) at 0 °C followed by the addition of EDC.HCl (2 equiv.). The reaction was allowed to stir overnight reaching room temperature. Upon completion, the mixture was reduced and purified by silica gel column chromatography a hexane/EtOAc mixture.

Method E: *General method for the carbamate conversion of ethyl carbamate protected guanidines*

Ethyl carbamate protected substituted guanidines were dissolved in excess *tert*-butanol and heated under reflux (80 °C) for 24 h. Upon completion, the reaction mixture was concentrated and purified by silica gel column chromatography eluting with a hexane/EtOAc mixture.

Method F: *General procedure for the synthesis of 1,4-diphenyl piperazine or piperidine intermediates*

To a solution of the appropriate phenyl piperazine/piperidine derivative (1 equiv.) in DMF, DIPEA (1 equiv.) and 1-Fluoro-4-nitrobenzene (1 equiv.) was added. The reaction mixture was stirred at room temperature for 24 hours and quenched with H₂O. The product was extracted with CH₂Cl₂, concentrated and triturated with EtOH to yield the desired diphenyl intermediate.

Method G: *General procedure for the amide coupling of acyl chloride and secondary amines*

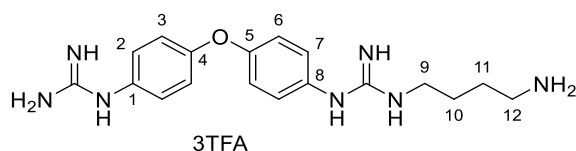
To a solution of an appropriate phenyl piperazine/piperidine (1 equiv.) and NEt₃ (1.2 equiv.) in CH₂Cl₂ at 0 °C, 4-nitrobenzoyl chloride (1.2 equiv.) was slowly added, and the reaction was left stirring for 5 minutes. Upon completion, the reaction mixture was condensed under reduced pressure and the product purified using silica gel column chromatography eluting with a hexane/EtOAc mixture.

Method H: *General procedure for nitroreduction using Palladium on carbon (Pd/C)*

To a solution of the appropriate nitro derivative in a specified solvent, was added 10% Pd/C. The reaction mixture was stirred at room temperature under a balloon of H₂ for 24 h. The reaction mixture was filtered off and concentrated to yield the desired product.

Method I: General procedure for Boc deprotection

A solution of Boc protected intermediate was dissolved in a 50:50 mixture of CH₂Cl₂ and TFA at 0 °C and allowed to stir overnight at room temperature. Upon completion, the reaction mixture was reduced under pressure, dissolved in H₂O and washed with CH₂Cl₂ (2x 5 mL). The aqueous layer was then concentrated to yield the final TFA salt.

6.1.3. Synthesis and Characterisation**TFA salt of 1-(4-aminobutyl)-3-(4-(4-guanidinophenoxy)phenyl)guanidine (34a)**

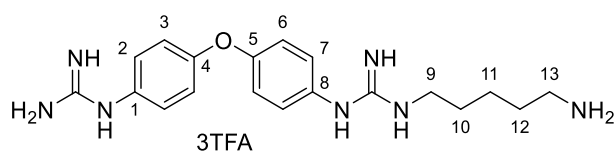
Following method I, from **45a** (7.8 mg, 0.0103 mmol), **34a** was obtained as a pale yellow gum (5.9 mg, 82% yield); HPLC = 95% pure (R_T = 17.11 min)

δ_H (400 MHz, D₂O): 7.20 (d, J = 8.9 Hz, 2H, H-2 and H-2'), 7.18 (d, J = 8.9 Hz, 2H, H-3 and H-3'), 7.02 – 6.97 (m, 4H, H-6/H-6' and H-7/H-7'), 3.18 (t, J = 6.4 Hz, 2H, H-12), 2.89 (t, J = 6.6 Hz, 2H, H-9), 1.62 – 1.49 (m, 4H, H-10 and H-11).

δ_C (101 MHz, D₂O): 163.1 (s, CO), 162.8 (s, CO), 156.6 (s, CN), 156.0 (s, CN), 155.9 (s, C-1), 155.5 (s, C-8), 129.7 (s, C-4), 129.6 (s, C-5), 128.2 (s, C-2), 128.2 (s, C-3), 120.1 (s, C-7), 120.0 (s, C-6), 40.7 (s, C-12), 39.0 (s, C-9), 25.0 (s, C-11), 24.0 (s, C-10).

HRMS-ESI (+): m/z calculated for C₁₈H₂₅N₇O: 356.2121 found: 356.2183 [M+H]⁺

ν_{max} (ATR)/cm⁻¹: 3072 (CH₂), 1640 (CN), 1501, 1438, 795, 704.

TFA salt of 1-(5-aminopentyl)-3-(4-(4-guanidinophenoxy)phenyl)guanidine (35a)

Following method I, from **46a** (7.4 mg, 0.0100 mmol), **35a** was obtained as a pale yellow gum (5.6 mg, 81% yield); HPLC = 98% pure (R_T = 18.28 min)

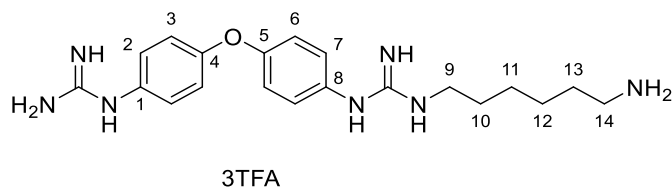
δ_H (400 MHz, D_2O): 7.20 (d, J = 8.9 Hz, 2H, H-2 and H-2'), 7.17 (d, J = 8.9 Hz, 2H, H-3 and H-3'), 7.04 – 6.99 (m, 4H, H-6/H-6' and H-7/H-7'), 3.14 (t, J = 7.0 Hz, 2H, H-13), 2.86 (t, J = 7.6 Hz, 2H, H-9), 1.60 – 1.47 (m, 4H, H-10 and H-12), 1.29 – 1.18 (m, 2H, H-11).

δ_C (101 MHz, D_2O): 163.1 (s, CO), 162.8 (s, CO), 156.5 (s, CN), 156.0 (s, CN), 155.9 (s, C-1), 155.4 (s, C-8), 129.8 (s, C-4), 129.6 (s, C-5), 128.2 (s, C-2 and C-3), 120.1 (s, C-7), 120.0 (s, C-6), 41.1 (s, C-13), 39.2 (s, C-9), 27.4 (s, C-12), 26.3 (s, C-10), 22.8 (s, C-11).

HRMS-ESI (+): m/z calculated for $C_{19}H_{27}N_7O$: 370.2277 found: 370.2357 $[M+H]^+$

$\nu_{max}(ATR)/cm^{-1}$: 3016 (CH_2), 1650 (CN), 1123, 1175, 833, 796.

TFA salt of 1-(6-aminohexyl)-3-(4-(4-guanidinophenoxy)phenyl)guanidine (**36a**)



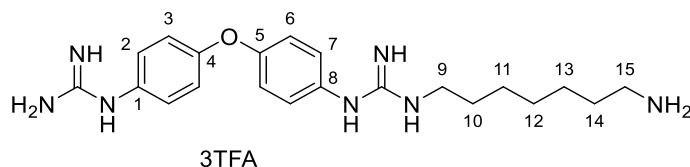
Following method I, from **47a** (14 mg, 0.0175 mmol), **36a** was obtained as a pale yellow gum (12 mg, 95% yield); HPLC = 95% pure (R_T = 19.01 min)

δ_H (400 MHz, D_2O): 7.18 (d, J = 8.9 Hz, 2H, H-2 and H-2'), 7.15 (d, J = 8.8 Hz, 2H, H-3 and H-3'), 7.00 – 6.97 (m, 4H, H-6/H-6' and H-7/H-7'), 3.11 (t, J = 6.8 Hz, 2H, H-14), 2.83 (t, J = 7.5 Hz, 2H, H-9), 1.49 – 1.37 (m, 4H, H-10 and H-13), 1.24 (br. s, H-12 and H-11).

δ_C (101 MHz, D_2O): 162.9 (s, CO), 162.6 (s, CO), 155.8 (s, CN), 155.7 (s, C-1 and C-8), 155.3 (s, CN), 129.8 (s, C-4), 129.5 (s, C-5), 128.1 (s, C-2 and C-3), 120.0 (s, C-7), 119.9 (s, C-6), 41.1 (s, C-14), 39.2 (s, C-9), 27.6 (s, C-13), 26.5 (s, C-10), 25.2 (s, C-12), 25.1 (s, C-11).

HRMS-ESI (-): m/z calculated for $C_{20}H_{29}N_7O$: 384.2434: found 384.2466 $[M+H]^+$

$\nu_{max}(ATR)/cm^{-1}$: 3163 (NH), 2159, 2027, 1662 (CN), 1500, 881, 837.

TFA salt of 1-(7-aminoheptyl)-3-(4-(4-guanidinophenoxy)phenyl)guanidine (37a)

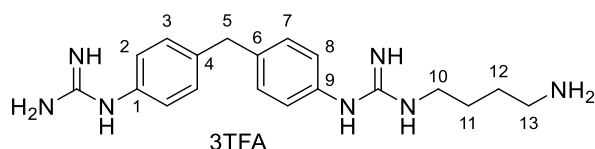
Following method I, from **48a** (29 mg, 0.036 mmol), **37a** was obtained as a pale yellow gum (20 mg, 77% yield); HPLC = 99.2% pure (R_T = 20.41 min)

δ_H (400 MHz, D_2O): 7.28 (d, J = 8.6 Hz, 2H, H-2 and H-2'), 7.25 (d, J = 8.6 Hz, 2H, H-3 and H-3'), 7.09 – 6.99 (m, 4H, H-6/H-6' and H-7/H-7'), 3.20 (t, J = 6.6 Hz, 2H, H-15), 2.92 (t, J = 7.5 Hz, 2H, H-9), 1.57 – 1.41 (m, 4H, H-10 and H-14), 1.30 (br. s, 6H, H-11, H-12, and H-13).

δ_C (101 MHz, D_2O): 156.6 (s, CN), 156.1 (s, CN), 155.8 (s, C-1), 155.4 (s, C-8), 129.9 (s, C-4), 129.6 (s, C-5), 128.2 (s, C-2 and C-3), 120.1 (s, C-7), 120.0 (s, C-6), 41.3 (s, C-15), 39.4 (s, C-9), 27.8 (s, C-11 and C-14), 26.6 (s, C-10), 25.6 (s, C-13), 25.4 (s, C-12).

HRMS-ESI (-): m/z calculated for $C_{21}H_{31}N_7O$: 397.2590: found 397.2466 [$M+H$] $^-$

$\nu_{max}(\text{ATR})/\text{cm}^{-1}$: 2943 (CH_2), 1660 (CN), 1500, 836.

TFA salt of 1-(4-aminobutyl)-3-(4-(4-guanidinobenzyl)phenyl)guanidine (34b)

Following method I, from **45b** (23 mg, 0.031 mmol), **34b** was obtained as a pale yellow gum (17 mg, 77% yield); HPLC = 99.8% pure (R_T = 18.15 min)

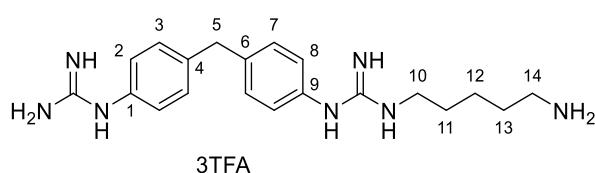
δ_H (400 MHz, D_2O): 7.28 – 7.21 (m, 4H, H-2/H-2' and H-3/H-3'), 7.15 – 7.08 (m, 4H, H-7/H-7' and H-8/H-8'), 3.92 (br. s, 2H, H-5), 3.18 (t, J = 6.3 Hz, 2H, H-13), 2.94 – 2.86 (m, 2H, H-10), 1.61 – 1.52 (m, 4H, H-11, H-12).

δ_c (101 MHz, D₂O): 162.9 (s, CN), 162.6 (s, CN), 141.3 (s, C-9), 141.1 (s, C-1), 132.2 (s, C-6), 132.1 (s, C-4), 130.1 (s, C-2), 130.1 (s, C-3), 126.3 (s, C-8), 126.2 (s, C-7), 40.6 (s, C-5), 40.0 (s, C-13), 38.9 (s, C-10), 24.9 (s, C-12), 23.9 (s, C-11).

HRMS-ESI (+): m/z calculated for C₁₉H₂₇N₇: 354.2328 found: 354.3104 [M+H]⁺

$\nu_{\max}$ (ATR)/cm⁻¹: 3102 (CH₂), 1663 (CN), 1598, 837.

TFA salt of 1-(5-aminopentyl)-3-(4-(4-guanidinobenzyl)phenyl)guanidine (35b)



Following method I, from **46b** (101 mg, 0.131 mmol), **35b** was obtained as a pale yellow gum (58 mg, 62 % yield); HPLC = 98.7% pure (R_T = 18.91 min)

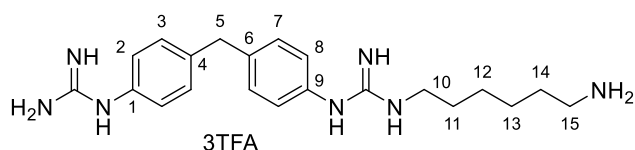
δ_H (400 MHz, D₂O): 7.27 – 7.20 (m, 4H, H-2/H-2' and H-3/H-3'), 7.11 – 7.00 (m, 4H, H-7/H-7' and H-8/H-8'), 3.91 (br. s, 2H, H-5), 3.13 (t, J = 7.0 Hz, 2H, H-14), 2.85 (t, J = 7.5 Hz, 2H, H-10), 1.59 – 1.44 (m, 4H, H-11 and H-13), 1.28 – 1.18 (m, 2H, H-12).

δ_c (101 MHz, D₂O): 156.4 (s, CN), 155.2 (s, CN), 141.3 (s, C-1), 141.2 (s, C-9), 132.4 (s, C-4), 132.2 (s, C-6), 130.2 (s, C-2), 130.2 (s, C-3), 126.4 (s, C-8), 126.3 (s, C-7), 41.0 (s, C-5), 40.1 (s, C-14), 39.2 (s, C-10), 27.4 (s, C-11), 26.3 (s, C-13), 22.8 (s, C-12).

HRMS-ESI (+): m/z calculated for C₂₀H₂₉N₇: 368.2484 found: 368.2558 [M+H]⁺

$\nu_{\max}$ (ATR)/cm⁻¹: 3089 (NH), 2935 (CH₂), 1680 (CO), 1512, 1488, 799, 726.

TFA salt of 1-(6-aminoheptyl)-3-(4-(4-guanidinobenzyl)phenyl)guanidine (36b)



Following method I, from **47b** (29 mg, 0.038 mmol), **36b** was obtained as a pale yellow gum (17 mg, 62% yield); HPLC = 97% pure (R_T = 19.43 min)

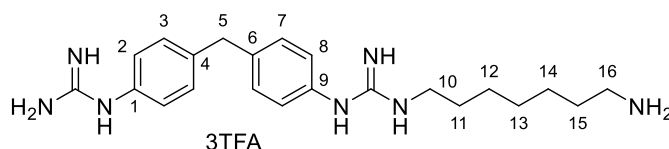
δ_H (400 MHz, D_2O): 7.26 - 7.20 (m, 4H, H-2/H-2' and H-3/H-3'), 7.13 - 7.00 (m, 4H, H-7/H-7' and H-8/H-8'), 3.92 (s, 2H, H-5), 3.12 (t, J = 7.0 Hz, 2H, H-15), 2.85 (t, J = 7.5 Hz, 2H, H-10), 1.50 - 1.45 (m, 4H, H-11 and H-14), 1.29 - 1.23 (m, 4H, H-12 and H-13).

δ_C (101 MHz, D_2O): 156.3 (s, CN), 155.2 (s, CN), 141.3 (s, C-9), 132.3 (s, C-4), 132.1 (s, C-6), 130.1 (s, C-2 and C-3), 126.3 (s, C-7 and C-8), 41.2 (s, C-5), 40.0 (s, C-15), 39.3 (s, C-10), 27.7 (s, C-14), 27.7 (s, C-11), 26.5 (s, C-12), 25.5 (s, C-13).

HRMS-ESI (+): m/z calculated for $C_{21}H_{31}N_7$: 382.2641 found: 382.2789 $[M+H]^+$

$\nu_{max}(ATR)/cm^{-1}$: 3110 (NH), 2938 (CH_2), 1665 (CN), 1512, 1469, 836, 798.

TFA salt of 1-(7-aminoheptyl)-3-(4-(4-guanidinobenzyl)phenyl)guanidine (**37b**)



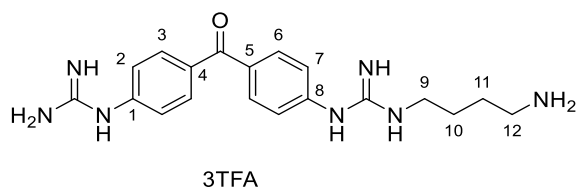
Following method I, from **48b** (25 mg, 0.032 mmol), **37b** was obtained as a pale yellow gum (20 mg, 85% yield); HPLC = 96% pure (R_T = 20.64 min)

δ_H (400 MHz, D_2O): 7.26 (d, J = 8.0 Hz, 2H, H-2 and H-2'), 7.24 (d, J = 8.0 Hz, 2H, H-3 and H-3'), 7.13 (d, J = 8.4 Hz, 2H, H-7/H-7'), 7.10 (d, J = 8.4 Hz, 2H, H-8 and H-8'), 3.92 (s, 2H, H-5), 3.12 (t, J = 6.9 Hz, 2H, H-16), 2.85 (t, J = 7.5 Hz, 2H, H-10), 1.48 - 1.32 (m, 4H, H-11 and H-15), 1.23 (br. s, 6H, H-12, H-13 and H-14).

δ_C (101 MHz, D_2O): 163.1 (s, CO), 162.7 (s, CO), 156.4 (s, CN), 155.3 (s, CN), 141.3 (s, C-1), 141.1 (s, C-9), 132.4 (s, C-4), 132.2 (s, C-6), 130.1 (s, C-2 and C-3), 126.3 (s, C-7 and C-8), 41.3 (s, C-16), 40.1 (s, C-5), 39.4 (s, C-10), 27.8 (s, C-13), 27.7 (s, C-15), 26.6 (s, C-11), 25.6 (s, C-14), 25.4 (s, C-12).

HRMS-ESI (+): m/z calculated for $C_{22}H_{33}N_7$: 396.2797 found: 396.2870 $[M+H]^+$

$\nu_{max}(ATR)/cm^{-1}$: 3088 (NH), 2939 (CH_2), 2864 (CH_2), 1664 (CN), 1603, 836, 798, 720.

TFA salt of 1-(4-aminobutyl)-3-(4-(4-guanidinobenzoyl)phenyl)guanidine (34c)

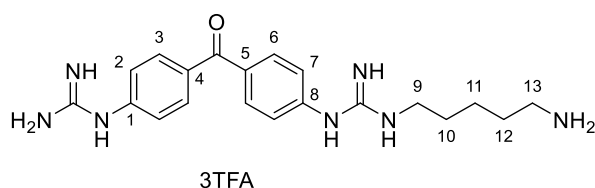
Following method I, from **45c** (14 mg, 0.0176 mmol), **34c** was obtained as a pale yellow gum (10 mg, 80% yield); HPLC = 97% pure (R_T = 16.37 min)

δ_H (400 MHz, D_2O): 7.81 – 7.67 (m, 4H, H-2/H-2' and H-3/H-3'), 7.41 – 7.37 (m, 2H, H-6/H-6' and H-7/H-7'), 3.31 (t, J = 6.3 Hz, 2H, H-12), 2.96 (t, J = 6.9 Hz, 2H, H-9), 1.72 – 1.62 (m, 4H, H-10 and H-11)

δ_C (101 MHz, D_2O): 198.3 (s, CO), 163.0 (s, CO), 162.8 (s, CO), 155.8 (s, CN), 154.7 (s, CN), 139.8 (s, C-1), 139.3 (s, C-8), 134.9 (s, C-4 and C-5), 132.1 (s, C-6), 132.0 (s, C-7), 124.1 (s, C-2 and C-3), 41.0 (s, C-12), 39.0 (s, C-9), 24.9 (s, C-11), 24.0 (s, C-10).

HRMS-ESI (+): m/z calculated for $C_{19}H_{25}N_7O$: 368.2121 found: 368.2195 $[M+H]^+$

$\nu_{max}(\text{ATR})/\text{cm}^{-1}$: 3181 (NH), 1668 (CN), 1596, 1176, 1066, 931, 838, 798, 721.

TFA salt of 1-(5-aminopentyl)-3-(4-(4-guanidinobenzoyl)phenyl)guanidine (35c)

Following method I, from **46c** (29 mg, 0.0377 mmol), **35c** was obtained as a pale yellow gum (25 mg, 91% yield); HPLC = 97% pure (R_T = 18.48 min)

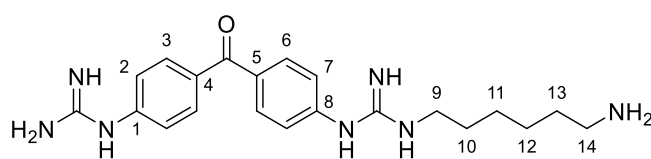
δ_H (400 MHz, D_2O): 7.82 – 7.69 (m, 4H, H-2/H-2' and H-3/H-3'), 7.41 – 7.37 (m, 2H, H-6/H-6' and H-7/H-7'), 3.27 (t, J = 7.1 Hz, 1H, H-13), 2.96 – 2.92 (m, 2H, H-9), 1.69 – 1.58 (m, 2H, H-10 and H-12), 1.44 – 1.33 (m, 2H, H-11).

δ_c (101 MHz, D₂O): 198.3 (s, CO), 155.8 (s, CN), 154.7 (s, CN), 139.9 (s, C-1), 139.3 (s, C-8), 134.9 (s, C-4, C5), 132.1 (s), 132.0 (s), 124.1 (s), 123.5 (s), 41.4 (s, C-13), 39.2 (s, C-9), 27.3 (s, C-12), 26.3 (s, C-10), 22.9 (s, C-11).

HRMS-ESI (+): m/z calculated for C₂₀H₂₇N₇O: 382.2277 found: 382.2359 [M+H]⁺

$\nu_{\max}(\text{ATR})/\text{cm}^{-1}$: 3165 (NH), 1667 (CN), 1595, 837, 798, 768, 721.

TFA salt of 1-(6-aminohexyl)-3-(4-(4-guanidinobenzoyl)phenyl)guanidine (36c)



3TFA

Following method I, from **47c** (28 mg, 0.0352 mmol), **36c** was obtained as a pale yellow gum (20 mg, 75% yield); HPLC = 99.6% pure (R_T = 18.21 min)

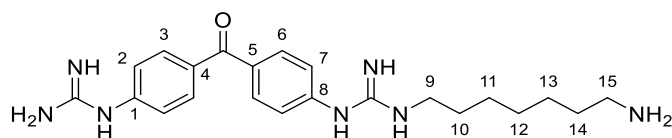
δ_H (400 MHz, D₂O): 7.71 – 7.60 (m, 4H, H-3/H-3' and H-6/H-6'), 7.30 (d, J = 8.5 Hz, 2H, H-7 and H-7'), 7.25 (d, J = 8.3 Hz, 2H, H-2 and H-2'), 3.18 (t, J = 7.0 Hz, 2H, H-14), 2.87 (t, J = 7.5 Hz, 2H, H-9), 1.53 – 1.42 (m, 4H, H-13 and H-10), 1.38 – 1.17 (m, 4H, H-11 and H-12).

δ_c (101 MHz, D₂O): 197.90 (s, CO), 162.8 (s, CO), 162.5 (s, CO), 155.7 (s, CN), 139.2 (s, C-1 and C-8), 134.6 (s, C-4 and C-5), 132.0 (s, C-2), 131.9 (s, C-3), 123.8 (s, C-6, C-7), 41.5 (s, C-14), 39.3 (s, C-9), 27.5 (s, C-13), 26.5 (s, C-10), 25.3 (s, C-12), 25.1 (s, C-11).

HRMS-ESI (+): m/z calculated for C₂₁H₂₉N₇O: 396.2434: found: 396.2512 [M+H]⁺

$\nu_{\max}(\text{ATR})/\text{cm}^{-1}$: 3095 (NH), 1664 (CN), 1594, 1473, 836, 798, 767, 720.

TFA salt of 1-(7-aminoheptyl)-3-(4-(4-guanidinobenzoyl)phenyl)guanidine (37c)



3TFA

Following method I, from **48c** (86 mg, 0.106 mmol), **37c** was obtained as a pale yellow gum (75 mg, 89% yield); HPLC = 98% pure (R_T = 19.01 min)

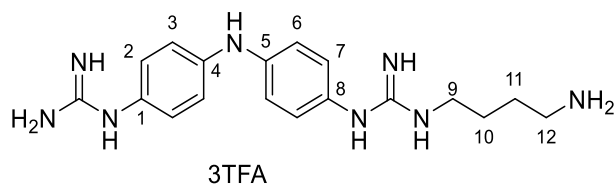
δ_H (400 MHz, D_2O): 7.76 – 7.70 (m, 4H, H-2/H-2' and H-3/H-3'), 7.35 (d, J = 8.5 Hz, 2H, H-6/H-6'), 7.30 (d, J = 8.5 Hz, 2H, H-7/H-7'), 3.20 (t, J = 6.9 Hz, 2H, H-15), 2.87 (t, J = 7.4 Hz, 2H, H-9), 1.54 (br. s, 4H, H-10 and H-14), 1.29 (br. s, 6H, H-11, H-12 and H-13).

δ_C (101 MHz, D_2O): 198.3 (s, CO), 163.4 (s, CO), 163.1 (s, CO), 155.8 (s, CN), 154.6 (s, CN), 139.3 (s, C-1), 134.8 (s, C-8), 132.0 (s, C-2), 131.9 (s, C-3), 124.0 (s, C-6), 123.4 (s, C-7), 41.5 (s, C-15), 39.3 (s, C-9), 27.5 (s, C-10), 26.5 (s, C-14), 25.3 (s, C-11 and C-13), 25.2 (s, C-12).

HRMS-ESI (+): m/z calculated for $C_{22}H_{31}N_7O$: 410.2590 found: 410.2583 $[M+H]^+$

$\nu_{max}(ATR)/cm^{-1}$: 3095 (NH), 2932 (CH_2), 1664 (CN), 1594, 835, 798.

TFA salt of 1-(4-aminobutyl)-3-((4-((4-guanidinophenyl)amino)phenyl)guanidine (**34d**)



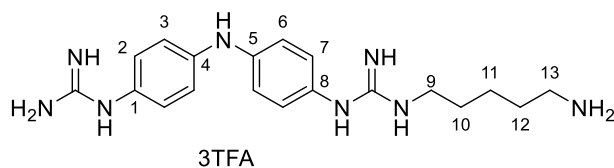
Following method I, from **45d** (14 mg, 0.0183 mmol), **34d** was obtained as a pale yellow gum (12 mg, 94% yield); HPLC = 96% pure (R_T = 15.32 min)

δ_H (400 MHz, D_2O): 7.15 (s, 8H, H-2/H-2', H-3/H-3', H-6/H-6' and H-7/H-7'), 3.25 (br. s, 2H, H-9), 2.96 (br. s, 2H, H-12), 1.64 (br. s, 4H, H-10 and H-11).

δ_C (151 MHz, D_2O): 156.7 (s, CN), 154.9 (s, CN), 142.9 (s, C-1), 142.8 (s, C-8), 127.0 (s, C-2 and C-3), 117.8 (s, C-6 and C-7), 40.6 (s, C-12), 38.7 (s, C-9), 25.2 (s, C-11), 24.2 (s, C-10).

HRMS-ESI (+): m/z calculated for $C_{18}H_{26}N_8$: 355.2280 found: 355.2357 $[M+H]^+$

$\nu_{max}(ATR)/cm^{-1}$: 3181 (NH), 2610, 2115, 2037, 1667 (CN), 1509, 856, 800.

TFA salt of 1-(5-aminopentyl)-3-(4-((4-guanidinophenyl)amino)phenyl)guanidine (35d)

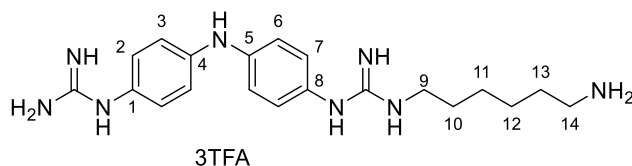
Following method I, from **46d** (17.8 mg, 0.0232 mmol), **35d** was obtained as a pale yellow gum (15 mg, 91% yield); HPLC = 99% pure (R_T = 16.80 min)

δ_H (400 MHz, D_2O): 7.20 – 7.10 (m, 8H, H-2/H-2', H-3/H-3', H-6/H-6' and H-7/H-7'), 3.20 (d, J = 6.8 Hz, 2H, H-13), 2.92 (d, J = 7.4 Hz, 2H, H-9), 1.69 – 1.51 (m, 4H, H-10 and H-12), 1.42 – 1.30 (m, 2H, H-11).

δ_C (151 MHz, D_2O): 156.7 (s, CN), 155.6 (s, CN), 142.9 (s, C-1), 142.8 (s, C-8), 127.9 (s, C-2), 127.8 (s, C-3), 126.9 (s, C-4), 126.8 (s, C-5), 118.8 (s, C-7), 118.7 (s, C-6), 41.0 (s, C-13), 39.2 (s, C-9), 27.4 (s, C-12), 26.4 (s, C-10), 22.8 (s, C-11).

HRMS-ESI (+): m/z calculated for $C_{19}H_{28}N_8$: 369.2437 found: 369.2513 [$M+H$] $^+$

ν_{max} (ATR)/ cm^{-1} : 3175 (NH), 2585, 2057, 2030, 1666 (CN), 1509, 841, 799.

TFA salt of 1-(6-aminoheptyl)-3-(4-((4-guanidinophenyl)amino)phenyl)guanidine (36d)

Following method I, from **47d** (22 mg, 0.0282 mmol), **36d** was obtained as a pale yellow gum (17 mg, 83% yield); HPLC = 98% pure (R_T = 18.05 min)

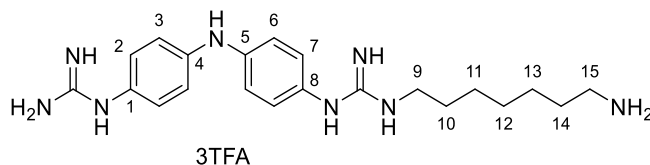
δ_H (400 MHz, D_2O): 7.21 – 7.13 (m, 8H, H-2/H-2', H-3/H-3', H-6/H-6' and H-7/H-7'), 3.20 (t, J = 6.9 Hz, 2H, H-14), 2.92 (t, J = 7.3 Hz, 2H, H-9), 1.64 – 1.51 (m, 4H, H-10 and H-13), 1.39 – 1.29 (m, 4H, H-11 and H-12).

δ_C (101 MHz, D_2O): 155.6 (s, CN), 143.0 (s, C-1), 142.8 (s, C-8), 127.9 (s, C-2), 127.8 (s, C-3), 126.9 (s, C-4), 126.8 (s, C-5), 118.8 (s, C-7), 118.7 (s, C-6), 41.2 (s, C-14), 39.3 (s, C-9), 27.7 (s, C-13), 26.5 (s, C-10), 25.3 (s, C-12), 25.2 (s, C-11).

HRMS-ESI (+): m/z calculated for $C_{20}H_{30}N_8$: 383.2593 found: 383.2610 $[M+H]^+$

$\nu_{\max}(\text{ATR})/\text{cm}^{-1}$: 3166 (NH), 2570, 2155, 2028, 1666 (CN), 1510, 835, 799.

TFA salt of 1-(7-aminoheptyl)-3-(4-((4-guanidinophenyl)amino)phenyl)guanidine (37d)



Following method I, from **48d** (14.6 mg, 0.0183 mmol), **37d** was obtained as a pale yellow gum (10 mg, 74% yield); HPLC = 99.5% pure (R_T = 19.80 min)

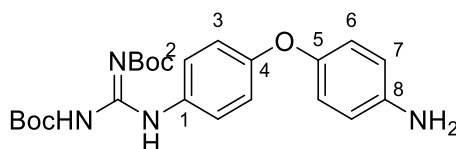
δ_H (400 MHz, D_2O): 7.21 – 7.08 (m, 8H, H-2/H-2', H-3/H-3', H-6/H-6' and H-7/H-7'), 3.18 (br. s, 2H, H-15), 2.91 (br. s, 2H, H-9), 1.56 (br. s, 4H, H-10 and H-14), 1.30 (br. s, 6H, H-11, H-12 and H-13).

δ_C (101 MHz, D_2O): 156.7 (s, CN), 155.5 (s, CN), 142.9 (s, C-1), 142.4 (s, C-8), 128.0 (s, C-2), 127.8 (s, C-3), 126.9 (s, C-4), 126.8 (s, C-5), 118.8 (s, C-7), 118.7 (s, C-6), 41.2 (s, C-15), 39.4 (s, C-9), 27.8 (s, C-14), 27.8 (s, C-10), 26.6 (s, C-13), 25.6 (s, C-11), 25.4 (s, C-12).

HRMS-ESI (+): m/z calculated for $C_{21}H_{32}N_8$: 397.2750: found: 397.2823 $[M+H]^+$

$\nu_{\max}(\text{ATR})/\text{cm}^{-1}$: 3162 (NH), 2556, 2160, 2028, 1663 (CN), 1509, 837, 798.

***tert*-Butyl-*N*-[(1*E*)-{[4-(4-aminophenoxy)phenyl]amino}{[(*tert*-butoxy)carbonyl]imino)}-2-methyl-2-thiopseudourea (38a)¹**



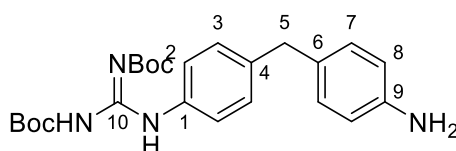
Following method B, from 4,4'-oxydianiline (1000 mg, 5 mmol), 1,3-bis-(*tert*-butoxycarbonyl)-2-methyl-2-thiopseudourea (436 mg, 1.5 mmol), NEt_3 (0.7 mL, 5.15 mmol) and $HgCl_2$ (448 mg,

1.8 mmol), **38a** was obtained as a white solid (595 mg, 88% yield); R_f 0.4 (1:1 Hex: EtOAc); mp decomposes over 255 °C (lit. >260 °C)¹

δ_H (400 MHz, $CDCl_3$): 11.64 (s, 1H, NH), 10.22 (s, 1H, NH), 7.49 (d, J = 8.8 Hz, 2H, H-7/H-7'), 6.91 – 6.88 (m, 2H, H-6/H-6'), 6.87 – 6.85 (m, 2H, H-3/H-3'), 6.69 – 6.65 (m, 2H, H-2/H-2'), 3.59 (s, 2H, NH_2), 1.54 (s, 9H, $C(CH_3)_3$), 1.50 (s, 9H, $C(CH_3)_3$).

HRMS-ESI: m/z calculated for $C_{23}H_{30}N_4O_5$: 443.2216; found: 443.2292 $[M+H]^+$

tert-Butyl-N-[(1E)-({4-[(4-aminophenyl)methyl]phenyl}amino){[(*tert*-butoxy)carbonyl]imino})methyl]carbamate (38b**)¹**

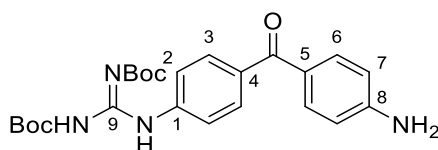


Following method B, from 4,4'-diaminodiphenylmethane (1784 mg, 9 mmol), 1,3 bis-(*tert*-butoxycarbonyl)-2-methyl-2-thiopseudourea (876 mg, 3 mmol), NEt_3 (1.3 mL, 9.3 mmol) and $HgCl_2$ (896 mg, 3.3 mmol), **38b** was obtained as pale yellow solid (1000 mg, 76% yield); R_f 0.5 (1:1 Hex: EtOAc); mp 104-107 °C (lit. 106 - 108 °C)¹

δ_H (400 MHz, $CDCl_3$): 11.62 (s, 1H, NH), 10.21 (s, 1H, NH), 7.50 – 7.46 (m, 2H, H-8/H-8'), 7.11 (d, J = 8.3 Hz, 2H, H-7/H-7'), 6.95 (m, 2H, H-2/H-2'), 6.64 – 6.59 (m, 2H, H-3/H-3'), 3.83 (s, 2H, H-5), 1.52 (s, 9H, $(C(CH_3)_3)$), 1.50 (s, 9H, $(C(CH_3)_3)$).

HRMS-ESI (+): m/z calculated for $C_{24}H_{32}N_4O_4$: 441.2424; found: 441.2456 $[M+H]^+$

tert-Butyl-N-[(1E)-{4-(4-aminobenzoyl)phenyl}amino){[(*tert*-butoxy)carbonyl]imino})methyl]carbamate (38c**)¹**



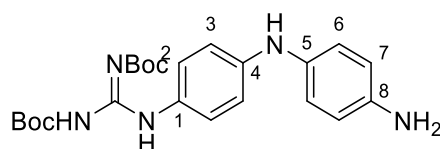
Following method B, from 4,4'-diaminodiphenylmethane (1910 mg, 9 mmol), 1,3-bis-(*tert*-butoxycarbonyl)-2-methyl-2-thiopseudourea (876 mg, 3 mmol), NEt_3 (1.3 mL, 9.3 mmol) and

HgCl₂ (896 mg, 3.3 mmol), **38c** was obtained as a white powder (600 mg, 64% yield); *R_f* 0.5 (1:1 Hex: EtOAc); mp decomposes over 200 °C (lit. >205 °C)¹

δ_{H} (400 MHz, CDCl₃): 11.66 (s, 1H, NH), 10.54 (s, 1H, NH), 7.71 (br. s, 4H, H-6/H-6' and H-7/H-7'), 7.66 (d, *J* = 8.6 Hz, 2H, H-3/H-3'), 6.63 (d, *J* = 8.6 Hz, 2H, H-2/H-2'), 1.52 (s, 18H, 2x C(CH₃)₃).

HRMS-ESI (+): *m/z* calculated for C₂₄H₃₀N₄O₅: 455.2216; found: 455.2239 [M+H]⁺

***tert*-Butyl-*N*-[(1*E*)-({4-[(4-aminophenyl)amino]phenyl}amino){[(*tert*-butoxy)carbonyl]imino})methyl]carbamate (**38d**)¹**



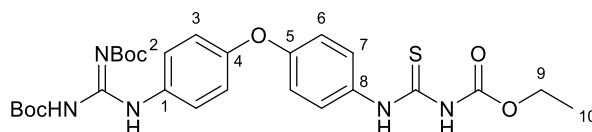
Following method B, from 4,4'-diaminodiphenylamine sulfate (2676 mg, 9 mmol), 1,3-bis-(*tert*-butoxycarbonyl)-2-methyl-2-thiopseudourea (876 mg, 3 mmol), NEt₃ (1.3 mL, 9.3 mmol) and HgCl₂ (896 mg, 3.3 mmol), **38d** was obtained as a purple solid. (1186 mg, 30% yield); *R_f* 0.4 (1:1 Hex: EtOAc); mp decomposes over 225 °C (lit. >230 °C)¹

δ_{H} (400 MHz, CDCl₃): 11.63 (s, 1H, NH), 10.12 (s, 1H, NH), 7.37 (d, *J* = 8.7 Hz, 2H, H-7/H-7'), 6.94 (d, *J* = 8.5 Hz, 2H, H-2/H-2'), 6.81 (d, *J* = 8.7 Hz, 2H, H-6/H-6'), 6.66 (d, *J* = 8.5 Hz, 2H, H-3/H-3'), 5.36 (s, 1H, NH), 3.53 (s, 2H, NH₂), 1.53 (s, 9H, C(CH₃)₃), 1.49 (s, 9H, C(CH₃)₃).

HRMS-ESI (+): *m/z* calculated for C₂₃H₃₁N₅O₄: 442.2376; found: 442.2432 [M+H]⁺

ν_{max} (ATR)/cm⁻¹: 3395 (NH₂), 3300 (NH), 3100 (NH), 1720 (CO), 1634, 1601 (CN).

Ethyl-*N*-{[4-(4-{[(1*E*)-{[(*tert*-butoxy)carbonyl]amino){[(*tert*-butoxy)carbonyl]imino})methyl]amino}phenoxy)phenyl]carbamothioyl}carbamate (39a**)**



Following method C, from **38a** (350 mg, 0.79 mmol) and ECIT (0.09 mL, 0.79 mmol), **39a** was obtained as a pale yellow powder (300 mg, 62%); *R_f* 0.6 (6: 4 Hex: EtOAc); mp 170-175 °C

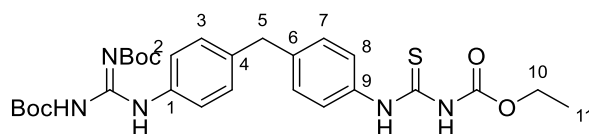
δ_{H} (400 MHz, CDCl_3): 11.66 (s, 1H, NH), 11.38 (s, 1H, NH), 10.33 (s, 1H, NH), 8.13 (s, 1H, NH), 7.61 (d, $J = 8.9$ Hz, 2H, H-3/H-3'), 7.56 (d, $J = 8.9$ Hz, 2H, H-6/H-6'), 7.06 – 6.99 (m, H-2/H-2' and H-7/H-7'), 4.31 (q, $J = 7.1$ Hz, 2H, H-9), 1.56 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.52 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.36 (d, $J = 7.1$ Hz, 3H, H-10).

δ_{C} (151 MHz, CDCl_3): 178.0 (s, CS), 163.5 (s, CN), 156.1 (s, C-5), 153.5 (s, CO), 153.3 (s, CO), 153.3 (s, C-4), 152.7 (s, CO), 132.6 (s, C-1), 132.5 (s, C-8), 126.0 (s, C-6), 123.8 (s, C-3), 119.7 (s, C-2), 118.4 (s, C-7), 83.7 (s, $\text{C}(\text{CH}_3)_3$), 79.6 (s, $\text{C}(\text{CH}_3)_3$), 63.0 (s, C-9), 28.2 (s, $\text{C}(\text{CH}_3)_3$), 28.1 (s, $\text{C}(\text{CH}_3)_3$), 14.2 (s, C-10).

HRMS-ESI (+): m/z calculated for $\text{C}_{27}\text{H}_{35}\text{N}_5\text{O}_7\text{S}$: 574.2257; found: 574.2327 $[\text{M}+\text{H}]^+$

$\nu_{\text{max}}(\text{ATR})/\text{cm}^{-1}$: 1500, 1224, 1155 (CS), 746.

Ethyl-*N*-{[4-[(4-[(1*E*)-[[(*tert*-butoxy)carbonyl]amino}{[(*tert*-butoxy)carbonyl]imino)}methyl]amino}phenyl)methyl]phenyl}carbamothioyl)carbamate (39b**)**



Following method C, from **38b** (200 mg, 0.45 mmol) and ECIT (0.05 mL, 0.45 mmol), **39b** was obtained as a pale yellow powder (131 mg, 51% yield); R_f 0.7 (6: 4 Hex: EtOAc); mp 130-135 °C

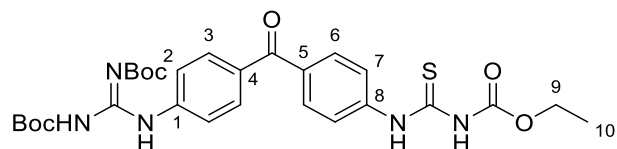
δ_{H} (400 MHz, CDCl_3): 11.66 (s, 1H, NH), 11.41 (s, 1H, NH), 10.30 (s, 1H, NH), 8.31 (s, 1H, NH), 7.55 – 7.39 (m, H-7/H-7' and H-8/H-8'), 7.18 (d, $J = 8.2$ Hz, 2H, H-2/H-2'), 7.13 (d, $J = 8.2$ Hz, 2H, H-3/H-3'), 4.27 (q, $J = 7.0$ Hz, 2H, H-10), 3.94 (s, 2H, H-5), 1.53 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.49 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.33 (t, $J = 7.0$ Hz, 3H, H-11).

δ_{C} (151 MHz, CDCl_3): 177.8 (s, CS), 153.4 (s, CO), 152.8 (s, CO), 139.7 (s, C-1), 137.0 (s, C-9), 135.6 (s, C-4), 134.9 (s, C-6), 129.3 (s, C-7), 129.2 (s, C-8), 124.3 (s, C-2), 122.3 (s, C-3), 83.6 (s, $\text{C}(\text{CH}_3)_3$), 75.6 (s, $\text{C}(\text{CH}_3)_3$), 65.2 (s, C-10), 28.1 (s, $\text{C}(\text{CH}_3)_3$), 28.0 (s, $\text{C}(\text{CH}_3)_3$), 14.1 (s, C-11).

HRMS-ESI (+): m/z calculated for $\text{C}_{28}\text{H}_{37}\text{N}_5\text{O}_6\text{S}$: 572.2465; found: 572.2498 $[\text{M}+\text{H}]^+$

$\nu_{\text{max}}(\text{ATR})/\text{cm}^{-1}$: 1556, 1304, 1155 (CS), 706.

Ethyl-*N*-{[4-(4-{[(1*E*)-{[(*tert*-butoxy)carbonyl]amino}{[(*tert*-butoxy)carbonyl]imino})methyl]amino}benzoyl)phenyl]carbamothioyl}carbamate (39c**)**



Following method C, from **38c** (500 mg, 1.1 mmol) and ECIT (0.13 mL, 1.1 mmol), **39c** was obtained as a pale yellow powder (327 mg, 63%); R_f 0.7 (6: 4 Hex: EtOAc); mp 130-138 °C

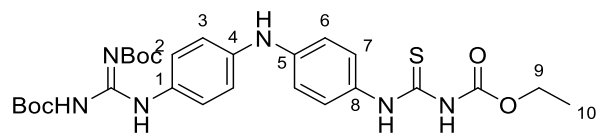
δ_H (400 MHz, $CDCl_3$): 11.74 (s, 1H, NH), 11.63 (s, 1H, NH), 10.61 (s, 1H, NH), 8.11 (s, 1H, NH), 7.88 – 7.83 (m, 4H, H-7/H-7' and H-6/H-6'), 7.83 – 7.77 (m, 4H, H-3/H-3' and H-2/H-2'), 4.31 (q, J = 7.1 Hz, 2H, H-9), 1.55 (s, 9H, $C(CH_3)_3$), 1.52 (s, 9H, $C(CH_3)_3$), 1.37 (t, J = 7.1 Hz, 3H, H-10).

δ_C (151 MHz, $CDCl_3$): 194.4 (CO), 177.8 (CS), 169.9 (CN), 153.2 (CO), 152.9 (CO), 152.8 (CO), 141.0 (C-1), 140.9 (s, C-8), 135.5 (s, C-5), 133.1 (s, C-4), 130.8 (s, C-7), 127.0 (s, C-3), 122.8 (C-6), 121.1 (s, C-2), 63.2 (C-9), 28.1 ($C(CH_3)_3$), 28.0 ($C(CH_3)_3$), 14.1 (C-10).

HRMS-ESI (+): m/z calculated for $C_{28}H_{35}N_5O_7S$: 586.2257; found: 586.2333 $[M+H]^+$

ν_{max} (ATR)/ cm^{-1} : 1723 (CO), 1498, 1227, 1143 (CS), 742.

***tert*-Butyl-*N*-[(1*Z*)-{[(*tert*-butoxy)carbonyl]imino}{[4-[4-{[(ethoxycarbonyl)amino]methanethioyl}amino)phenyl]amino}phenyl]amino]methyl]carbamate (**39d**)**



Following method C, from **38d** (200 mg, 0.45 mmol) and ECIT (0.05 mL, 0.45 mmol), **39d** was obtained as a pale yellow powder (114 mg, 20%); R_f 0.5 (6: 4 Hex: EtOAc)

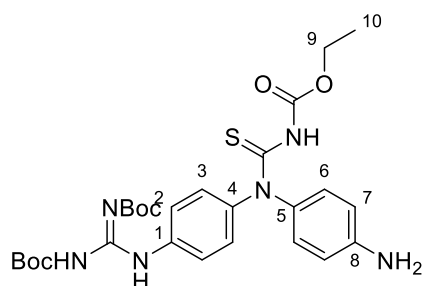
δ_H (400 MHz, $CDCl_3$): 11.64 (s, 1H, NH), 11.27 (s, 1H, NH), 10.17 (s, 1H, NH), 8.24 (s, 1H, NH), 7.39 (d, J = 8.5 Hz, 4H, H-3/H-3' and H-6/H-6'), 6.97 (d, J = 8.5 Hz, 4H, H-2/H-2' and H-7/H-7'), 6.03 (s, 1H, NH), 4.25 (q, J = 7.1 Hz, 2H, H-9), 1.52 (s, 9H, $C(CH_3)_3$), 1.46 (s, 9H, $C(CH_3)_3$), 1.32 (t, J = 7.1 Hz, 3H, H-10).

δ_c (151 MHz, $CDCl_3$): δ 177.7 (s, CS), 163.5 (s, CN), 153.7 (s, CO), 153.2 (s, CO), 152.8 (s, CO), 142.4 (s, C-1), 139.6 (s, C-8), 130.2 (s, C-4), 129.8 (s, C-5), 125.5 (s, C-3), 123.9 (s, C-6), 118.9 (s, C-2), 116.6 (s, C-7), 83.5 (s, $C(CH_3)_3$), 79.5 (s, $C(CH_3)_3$), 62.8 (s, C-9), 28.1 (s, $C(CH_3)_3$), 28.0 (s, $C(CH_3)_3$), 14.1 (s, C-10).

HRMS-ESI (+): m/z calculated for $C_{27}H_{36}N_6O_6S$: 573.2417: found: 573.2495 $[M+H]^+$

$\nu_{max}(ATR)/cm^{-1}$: 3201 (NH), 1716 (CO), 1637, 1536, 1391, 1147 (CS), 1096, 809.

Ethyl-*N*-({4-[(4-{[(1*E*)-{[(*tert*-butoxy)carbonyl]amino}{[(*tert*-butoxy)carbonyl]imino)}methyl]amino}phenyl)amino]phenyl}carbamoithioyl)carbamate (39e**)**



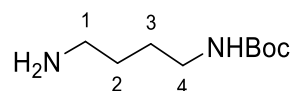
Following method C with the intention of synthesising compound **39d**, **39e** was obtained, as a pale yellow powder (374 mg, 62%); R_f 0.5 (6: 4 Hex: EtOAc); mp 145-151 °C

δ_H (400 MHz, $CDCl_3$): 11.59 (s, 1H, NH), 11.55 (s, 1H, NH), 10.45 (s, 1H, NH), 8.22 (s, 1H, NH), 7.73 (s, 1H, NH), 7.69 (d, J = 8.8 Hz, 2H, H-2/H-2'), 7.64 (d, J = 8.8 Hz, 2H, H-3/H-3'), 7.26 (d, J = 9.0 Hz, 2H, H-6/H-6'), 7.22 (d, J = 9.0 Hz, 2H, H-7 and H-7'), 4.25 (q, J = 7.1 Hz, 2H, H-9), 1.50 (s, 9H, $C(CH_3)_3$), 1.47 (s, 9H, $C(CH_3)_3$), 1.31 (t, J = 7.1 Hz, 3H, H-10).

δ_c (151 MHz, $CDCl_3$): 177.3 (s, CS), 163.4 (s, CN), 153.3 (s, CO), 152.8 (s, CO), 149.8 (s, CO), 142.7 (s, C-1), 140.3 (s, C-8), 136.5 (s, C-4), 136.4 (s, C-5), 127.6 (s, C-2), 124.8 (s, C-6), 124.6 (s, C-7), 122.9 (s, C-3), 84.0 (s, $C(CH_3)_3$), 79.8 (s, $C(CH_3)_3$), 63.1 (s, C-9), 28.1 (s, $C(CH_3)_3$), 14.1 (s, C-10).

HRMS-ESI (+): m/z calculated for $C_{27}H_{36}N_6O_6S$: 573.2417: found: 573.2495 $[M+H]^+$

$\nu_{max}(ATR)/cm^{-1}$: 3130 (NH), 1629 (CO), 1371, 1340, 1145 (CS), 831, 746.

tert-Butyl N-(4-aminobutyl)carbamate (40a)

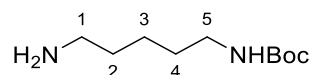
Following method A, from 1,7-diaminobutane (3000 mg, 34 mmol) and di-*tert*-butyl-dicarbonate (3700 mg, 17 mmol), **40a** was obtained as a white amorphous gum. (1580 mg, 49% yield); R_f 0.3 (7:2.5:0.5 CH₂Cl₂: MeOH: NEt₃)

δ_H (400 MHz, DMSO-*d*₆): 6.74 (s, 1H, NH), 2.85 – 2.77 (m, 2H, H-1), 2.53 – 2.47 (m, 2H, H-4), 1.31 (br. s, 9H, (C(CH₃)₃), 1.30 (br. s, 4H, H-2 and H-3).

δ_C (151 MHz, DMSO- *d*₆): 155.9 (s, CO), 77.7 (s, C(CH₃)₃), 41.3 (s, C-1), 40.5 (s, C-4), 30.0 (s, H-2), 28.7 (s, C(CH₃)₃), 27.3 (s, H-3).

HRMS-ESI (+): m/z calculated for C₉H₂₀N₂O₂: 189.1525: found: 189.1603 [M+H]⁺

$\nu_{\max}$ (ATR)/cm⁻¹: 3368 (NH₂), 2943 (CH₂), 1680 (CO), 1388, 1364, 740.

tert-Butyl N-(5-aminopentyl)carbamate (40b)

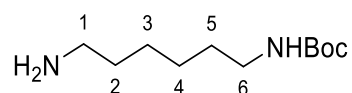
Following method A, from 1,5-diaminopentane (3000 mg, 29 mmol) and di-*tert*-butyl-dicarbonate (3200 mg, 15 mmol), **40b** was obtained as a white amorphous gum. (1020 mg, 34% yield); R_f 0.3 (7:2.5:0.5 CH₂Cl₂: MeOH: NEt₃)

δ_H (400 MHz, DMSO- *d*₆): 6.72 (s, 1H, NH), 2.85 – 2.75 (m, 2H, H-1), 2.52 – 2.48 (m, J = 2.9 Hz, 2H, H-5), 1.33 (s, 9H, (C(CH₃)₃), 1.32 – 1.26 (m, 4H, H-2 and H-4), 1.23 – 1.15 (m, 2H, H-3).

δ_C (151 MHz, DMSO- *d*₆): 155.9 (s, CO), 77.6 (s, C(CH₃)₃), 41.5 (s, C-1), 40.5 (s, C-5), 32.5 (s, C-2), 29.7 (s, C-4), 28.7 (s, C(CH₃)₃), 24.0 (s, C-3).

HRMS-ESI (+): m/z calculated for C₁₀H₂₂N₂O₂: 203.1681: found: 203.1761 [M+H]⁺

$\nu_{\max}$ (ATR)/cm⁻¹: 3368 (NH₂), 2937 (CH₂), 1682 (CO), 1389, 1364, 760.

***tert*-Butyl *N*-(6-aminohexyl)carbamate (**40c**)**

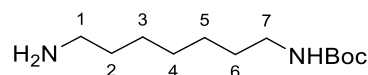
Following method A, from 1,6-diaminohexane (3000 mg, 26 mmol) and di-*tert*-butyl-dicarbonate (2817 mg, 13 mmol), **40c** was obtained as a white amorphous gum. (1005 mg, 36% yield); R_f 0.3 (7:2.5:0.5 CH₂Cl₂: MeOH: NEt₃)

δ_H (400 MHz, DMSO- *d*₆): 6.71 (s, 1H, NH), 2.83 – 2.76 (m, 2H, H-1), 2.51 – 2.47 (m, 2H, H-5), 1.33 (s, 9H, (C(CH₃)₃), 1.29 – 1.27 (m, 4H, H-2 and H-5), 1.25 – 1.14 (m, 4H, H-3 and H-4).

δ_C (151 MHz, DMSO- *d*₆): 155.9 (s, CO), 77.6 (s, C(CH₃)₃), 41.7 (s, C-1), 40.5 (s, C-6), 33.2 (s, C-2), 29.9 (s, C-5), 28.6 (s, C(CH₃)₃), 26.6 (s, C-3), 26.5 (s, C-4).

HRMS-ESI (+): m/z calculated for C₁₁H₂₄N₂O₂: 217.1838: found: 217.1919 [M+H]⁺

$\nu_{\max}$ (ATR)/cm⁻¹: 3368 (NH₂), 2943 (CH₂), 1679 (CO), 1387, 1364, 729.

***tert*-Butyl *N*-(7-aminoheptyl)carbamate (**40d**)**

Following method A, from 1,7-diaminoheptane (3000 mg, 23 mmol) and di-*tert*-butyl-dicarbonate (2518 mg, 11.5 mmol), **40d** was obtained as a white amorphous gum. (980 mg, 37% yield); R_f 0.4 (7:2.5:0.5 CH₂Cl₂: MeOH: NEt₃)

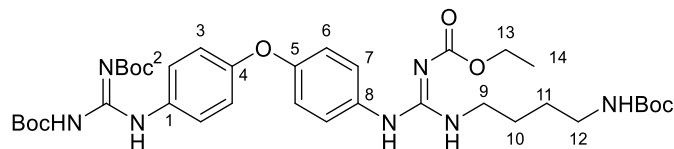
δ_H (400 MHz, DMSO- *d*₆): 6.71 (s, 1H, NH), 2.84 – 2.73 (m, H-1), 2.51 – 2.48 (m, 2H, H-7), 1.33 (s, 9H, (C(CH₃)₃), 1.32 – 1.25 (m, 4H, H-2 and H-6), 1.20 (br. s, 6H, H-3, H-4 and H-5).

δ_C (151 MHz, DMSO- *d*₆): 155.9 (s, CO), 77.6 (s, C(CH₃)₃), 41.6 (s, C-1), 40.5 (s, C-7), 33.0 (s, C-2), 29.8 (s, C-6), 29.0 (s, C-3), 28.7 (s, C(CH₃)₃), 26.7 (s, C-5), 26.7 (s, C-4).

HRMS-ESI (+): m/z calculated for C₁₂H₂₆N₂O₂: 231.1994: found 231.2108 [M+H]⁺

$\nu_{\max}$ (ATR)/cm⁻¹: 3368 (NH₂), 2931 (CH₂), 1679 (CO), 1389, 1364, 780.

***tert*-Butyl-*N*-[(1*E*)-{[4-(4-{[(1*Z*)-[4-{[(*tert*-butoxy)carbonyl]amino}butyl]imino] [(ethoxycarbonyl)amino]methyl]amino}phenoxy)phenyl]amino}{[(*tert*-butoxy)carbonyl]imino)}methyl]carbamate (**41a**)**



Following method D, from **39a** (100 mg, 0.1740 mmol), **40a** (50 mg, 0.261 mmol), EDC.HCl (66 mg, 0.348 mmol) and DIPEA (0.05 mL, 0.261 mmol), **41a** was obtained as a yellow amorphous gum (100 mg, 79% yield); R_f 0.3 (7: 3 Hex: EtOAc)

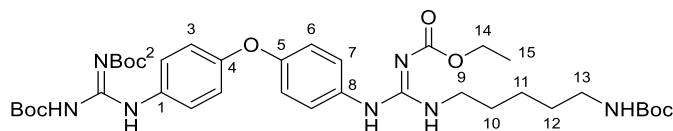
δ_H (400 MHz, $CDCl_3$): 11.63 (s, 1H), 10.50 (s, 1H), 10.31 (s, 1H), 7.59 (d, $J = 8.8$ Hz, 2H, H-3/H-3'), 7.14 (d, $J = 8.4$ Hz, 2H, H-6/H-6'), 6.99 – 6.92 (m, 4H, H-2/H-2' and H-7/H-7'), 4.64 (br. s, 1H, NH), 4.14 (q, $J = 7.1$ Hz, 2H, H-13), 3.36 – 3.25 (m, 2H, H-12), 3.09 – 2.98 (m, 2H, H-9), 1.52 (s, 9H, $(C(CH_3)_3)$), 1.48 (s, 9H, $(C(CH_3)_3)$), 1.47 – 1.43 (m, 4H, H-10 and H-11), 1.40 (s, 9H, $(C(CH_3)_3)$), 1.31 (t, $J = 7.1$ Hz, 3H, H-14).

δ_C (151 MHz, $CDCl_3$): 164.5 (s, CN), 163.4 (s, CN), 159.0 (s, CO), 156.6 (s, C-4), 155.9 (s, CO), 153.5 (s, CO), 153.2 (s, CO), 152.9 (s, C-5), 132.9 (s, C-1), 130.4 (s, C-8), 127.6 (s, C-6), 123.8 (s, C-3), 119.9 (s, C-2), 119.3 (s, C-7), 83.7 (s, $\underline{C}(CH_3)_3$), 79.6 (s, $\underline{C}(CH_3)_3$), 60.7 (s, C-13), 40.7 (s, C-12), 39.9 (s, C-9), 28.3 (s, $C(\underline{CH}_3)_3$), 28.1 (s, $C(\underline{CH}_3)_3$), 28.0 (s, $C(\underline{CH}_3)_3$), 27.3 (s, C-10), 26.8 (s, C-11), 14.6 (s, C-14).

HRMS-ESI (+): m/z calculated for $C_{36}H_{53}N_7O_9$: 728.3905; found: 728.3980 $[M+H]^+$

$\nu_{max}(ATR)/cm^{-1}$: 3211 (NH), 2987 (CH_2), 1626, 1533, 1111, 839, 806.

***tert*-Butyl-*N*-[(1*E*)-{[4-(4-{[(1*Z*)-[(5-{[(*tert*-butoxy)carbonyl]amino}pentyl)imino] [(ethoxycarbonyl)amino]methyl]amino}phenoxy)phenyl]amino}{[(*tert*-butoxy)carbonyl]imino}]methyl]carbamate (**42a**)**



Following method D, from **39a** (116 mg, 0.2022 mmol), **40b** (62 mg, 0.303 mmol), EDC.HCl (78 mg, 0.404 mmol) and DIPEA (0.05 mL, 0.303 mmol), **42a** was obtained as a yellow amorphous gum (115 mg, 77% yield); R_f 0.4 (7: 3 Hex: EtOAc)

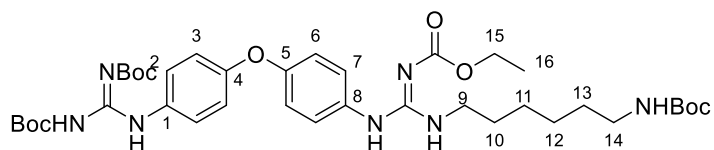
δ_H (400 MHz, $CDCl_3$): 11.66 (s, 1H, NH), 10.60 (s, 1H, NH), 10.33 (s, 1H, NH), 7.60 (d, $J = 8.8$ Hz, 2H, H-3/H-3'), 7.15 (d, $J = 7.4$ Hz, 2H, H-6/H-6'), 7.01-6.95 (m, 4H, H-2/H-2' and H-7/H-7'), 4.58 (br. s, 1H, NH), 4.16 (q, $J = 7.1$ Hz, 2H, H-14), 3.36 – 3.28 (m, 2H, H-13), 3.10 – 2.99 (m, 2H, H-9), 1.54 (s, 9H, $C(CH_3)_3$), 1.50 (s, 9H, $C(CH_3)_3$), 1.49 – 1.45 (m, , 4H, H-10 and H-12), 1.43 (s, 9H, $C(CH_3)_3$), 1.33 (t, $J = 7.1$ Hz, 3H, H-15), 1.31 – 1.29 (m, 2H, H-11)

δ_C (151 MHz, $CDCl_3$): 164.5 (s, CN), 163.4 (s, CN), 159.0 (s, CO), 156.7 (s, C-4), 156.0 (s, CO), 153.5 (s, CO), 153.3 (s, CO), 153.0 (s, C-5), 132.9 (s, C-1), 130.5 (s, C-8), 127.6 (s, C-6), 123.8 (s, C-3), 119.9 (s, C-2), 119.4 (s, C-7), 83.8 (s, $C(CH_3)_3$), 79.7 (s, $C(CH_3)_3$), 79.0 (s, $C(CH_3)_3$), 60.8 (s, C-11), 41.0 (s, C-13), 40.2 (s, C-9), 29.6 (s, C-10), 29.2 (s, C-12), 28.4 (s, $C(CH_3)_3$), 28.1 (s, $C(CH_3)_3$), 28.0 (s, $C(CH_3)_3$), 23.8 (s, C-11), 14.6 (s, C-15).

HRMS-ESI (+): m/z calculated for $C_{37}H_{55}N_7O_9$: 742.4061; found: 742.4149 $[M+H]^+$

$\nu_{max}(ATR)/cm^{-1}$: 3265 (NH), 2927 (CH_2), 1627 (CO), 1597 (CO), 1499 (CN), 1395, 801, 755.

***tert*-Butyl-*N*-[(1*E*)-{[4-(4-{[(1*Z*)-[(6-{[(*tert*-butoxy)carbonyl]amino}hexyl)imino] [(ethoxycarbonyl)amino]methyl]amino}phenoxy)phenyl]amino}{[(*tert*-butoxy)carbonyl]imino}]methyl]carbamate (**43a**)**



Following method D, from **39a** (200 mg, 0.35 mmol), **40c** (114 mg, 0.524 mmol), EDC.HCl (134 mg, 0.698 mmol) and DIPEA (0.09 mL, 0.524 mmol), **43a** was obtained as a yellow amorphous gum (120 mg, 45% yield); R_f 0.4 (7: 3 Hex: EtOAc)

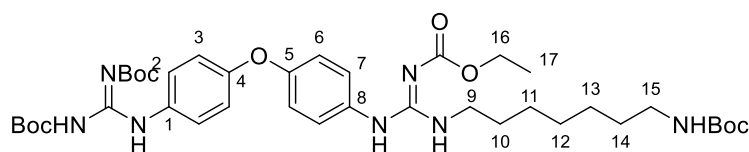
δ_H (400 MHz, $CDCl_3$): 11.65 (s, 1H, NH), 10.60 (s, 1H, NH), 10.33 (s, 1H, NH), 7.61 (d, J = 8.6 Hz, 2H, H-3/H-3'), 7.17 (d, J = 8.8 Hz, 2H, H-6/H-6'), 7.01 - 6.95 (m, 4H, H-2/H-2' and H-7/H-7'), 4.64 (s, 1H, NH), 4.17 (q, J = 7.1 Hz, 2H, H-15), 3.35 (br. s, 2H, H-14), 3.07 (br. s, 2H, H-9), 1.55 (s, 9H, $C(CH_3)_3$), 1.50 (s, 9H, $C(CH_3)_3$), 1.46 – 1.44 (m, 4H, H-10 and H-13), 1.43 (s, 9H, $C(CH_3)_3$), 1.33 (t, J = 7.1 Hz, 3H, H-16), 1.31 – 1.28 (m, 4H, H-11 and H-12).

δ_C (151 MHz, $CDCl_3$): 163.4 (s, CN), 159.0 (s, CO), 156.7 (s, C-4), 156.0 (s, CO), 153.5 (s, CO), 153.3 (s, CO), 153.1 (s, C-5), 132.8 (s, C-1), 130.6 (s, C-8), 127.7 (s, C-6), 123.8 (s, C-3), 119.8 (s, C-2), 119.4 (s, C-7), 83.8 (s, $C(CH_3)_3$), 79.7 (s, $C(CH_3)_3$), 79.0 (s, $C(CH_3)_3$), 60.8 (s, C-15), 40.9 (s, C-14), 40.2 (s, C-9), 29.9 (s, C-10), 29.6 (s, C-13), 29.4 (s, C-11), 28.4 (s, $C(CH_3)_3$), 28.1 (s, $C(CH_3)_3$), 28.0 (s, $C(CH_3)_3$), 26.2 (s, C-12), 14.6 (s, C-16).

HRMS-ESI (+): m/z calculated for $C_{38}H_{57}N_7O_9$: 756.4218; found: 756.4282 $[M+H]^+$

$\nu_{max}(ATR)/cm^{-1}$: 3044 (NH), 2900 (CH_2), 2850 (CH_2), 1597 (CO), 1394, 1112, 801, 752.

tert-Butyl-N-[(1E)-{[4-(4-{[(1Z)-[(7-{[(tert-butoxy)carbonyl]amino}heptyl)imino] [(ethoxycarbonyl)amino]methyl]amino}phenoxy)phenyl]amino}{[(tert-butoxy)carbonyl]imino}]methyl]carbamate (44a**)**



Following method D, from **39a** (200 mg, 0.35 mmol), **40d** (120 mg, 0.524 mmol), EDC.HCl (134 mg, 0.698 mmol) and DIPEA (0.09 mL, 0.524 mmol), **44a** was obtained as a yellow amorphous gum was obtained (151 mg, 56% yield); R_f 0.3 (7: 3 Hex: EtOAc)

δ_H (400 MHz, $CDCl_3$): 11.62 (s, 1H, NH), 10.55 (s, 1H, NH), 10.28 (s, 1H, NH), 7.55 (d, J = 8.8 Hz, 2H, H-3/H-3'), 7.10 (d, J = 7.6 Hz, 2H, H-6/H-6'), 6.97 (d, J = 8.8 Hz, 2H, H-2/H-2'), 6.94 (d, J = 7.6 Hz, 2H, H-7/H-7'), 4.64 (s, 1H, NH), 4.10 (q, J = 7.1 Hz, 2H, H-16), 3.30 (br. s, 2H, H-9), 3.02

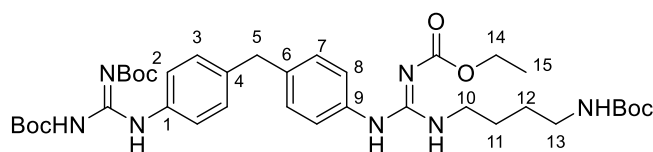
(br. s, 2H, H-15), 1.49 (s, 9H, C(CH₃)₃), 1.45 (s, 9H, C(CH₃)₃), 1.44 – 1.39 (m, 10H, H-10, H-11, H-12, H-13 and H-14), 1.38 (s, 9H, C(CH₃)₃), 1.27 (t, *J* = 7.1 Hz, 3H, H-16).

δ_c (151 MHz, CDCl₃): 164.4 (s, CN), 163.4 (s, CN), 158.9 (s, CO), 156.5 (s, C-4), 155.9 (s, CO), 153.5 (s, CO), 153.2 (s, CO), 153.0 (s, C-5), 132.7 (s, C-1), 130.7 (s, C-8), 127.5 (s, C-6), 123.8 (s, C-3), 119.8 (s, C-2), 119.3 (s, C-7), 83.7 (s, C(CH₃)₃), 79.6 (s, C(CH₃)₃), 78.8 (s, C(CH₃)₃), 60.7 (s, C-16), 41.1 (s, C-15), 40.4 (s, C-9), 29.9 (s, C-10), 29.3 (s, C-14), 28.8 (s, C-11), 28.3 (s, C(CH₃)₃), 28.1 (s, C(CH₃)₃), 28.0 (s, C(CH₃)₃), 26.6 (s, C-13), 26.5 (s, C-12), 14.6 (s, C-7).

HRMS-ESI (+): *m/z* calculated for C₃₉H₅₉N₇O₉: 770.4374; found: 770.4465 [M+H]⁺

vmax(ATR)/cm⁻¹: 2998 (NH), 2950 (CH₂), 1720 (CO), 1626 (CO), 1499 (CN), 1434, 802, 750.

***tert*-Butyl-*N*-{4-[(*E*)-{[(4-[(1*E*)-{[(*tert*-butoxy)carbonyl]amino}{[(*tert*-butoxy)carbonyl]imino})methyl]amino}phenyl)methyl]phenyl]amino}[(ethoxycarbonyl)amino]methylidene]amino]butyl}carbamate (**41b**)**



Following method D, from **39b** (130 mg, 0.226 mmol), **40a** (64 mg, 0.339 mmol), EDC.HCl (87 mg, 0.453 mmol) and DIPEA (0.06 mL, 0.339 mmol), **41b** was obtained as a yellow amorphous gum (160 mg, 96% yield); *R_f* 0.4 (7: 3 Hex: EtOAc)

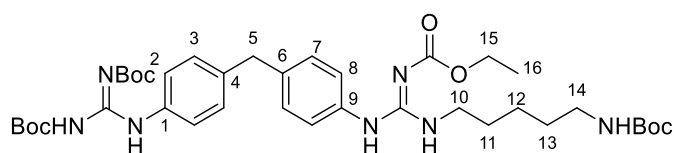
δ_H (400 MHz, CDCl₃): 11.62 (s, 1H, NH), 10.60 (s, 1H, NH), 10.27 (s, 1H, NH), 7.51 (d, *J* = 8.5 Hz, 2H, H-8/H-8'), 7.19 (d, *J* = 8.1 Hz, 2H, H-2/H-2'), 7.13 (d, *J* = 8.5 Hz, 2H, H-7/H-7'), 7.08 (d, *J* = 8.1 Hz, 2H, H-3/H-3'), 4.64 (s, 1H, NH), 4.14 (q, *J* = 7.1 Hz, 2H, H-14), 3.91 (br. s, 2H, H-5), 3.34 (br. s, 2H, H-13), 3.08 (br. s, 2H, H-10), 1.51 (s, 9H, C(CH₃)₃), 1.47 (s, 9H, C(CH₃)₃), 1.45 – 1.43 (m, 4H, H-11 and H-12), 1.40 (s, 9H, C(CH₃)₃), 1.30 (t, *J* = 7.1 Hz, 3H, H-15).

δ_c (151 MHz, CDCl₃): 164.5 (s, CN), 163.5 (s, CN), 158.7 (s, CO), 155.9 (s, CO), 153.5 (s, CO), 153.2 (s, CO), 140.5 (s, C-1), 136.9 (s, C-6), 135.0 (C-9), 130.3 (s, C-2), 129.2 (s, C-7), 125.8 (s, C-3), 122.4 (s, C-8), 83.6 (s, C(CH₃)₃), 79.5 (s, C(CH₃)₃), 60.7 (C-14), 40.9 (s, C-5), 40.6 (s, C-13), 39.9 (s, C-10), 28.3 (s, C(CH₃)₃), 28.1 (s, C(CH₃)₃), 28.0 (s, C(CH₃)₃), 27.2 (s, C-11), 26.8 (s, C-12), 14.6 (s, C-15).

HRMS-ESI (-): m/z calculated for C₃₇H₅₅N₇O₈: 724.4112; found: 724.4101 [M-H]⁻

v_{max}(ATR)/cm⁻¹: 3005 (NH), 2986 (CH₂), 2925 (CH₂), 1626 (CO), 1597 (CN), 1391, 802, 749.

tert-Butyl-N-{5-[(E)-[[(4-[(1E)-[[(tert-butoxy)carbonyl]amino}{[(tert-butoxy)carbonyl]imino}]methyl]amino]phenyl)methyl]phenyl]amino}[(ethoxycarbonyl)amino]methylidene]amino]pentyl}carbamate (42b**)**



Following method D, from **39b** (206 mg, 0.359 mmol), **40b** (109 mg, 0.538 mmol), EDC.HCl (138 mg, 0.718 mmol) and DIPEA (0.09 mL, 0.538 mmol), **42b** was obtained as a yellow amorphous gum (166 mg, 62%); R_f 0.4 (7: 3 Hex: EtOAc)

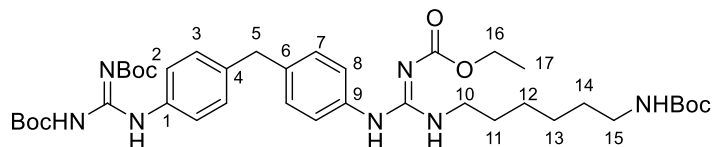
δ_H (400 MHz, CDCl₃): 11.65 (s, 1H, NH), 10.63 (s, 1H, NH), 10.29 (s, 1H, NH), 7.53 (d, *J* = 8.4 Hz, 2H, H-8/H-8'), 7.22 (d, *J* = 8.1 Hz, 2H, H-2/H-2'), 7.16 (d, *J* = 8.4 Hz, 2H, H-7/H-7'), 7.11 (d, *J* = 8.1 Hz, 2H, H-3/H-3'), 4.57 (s, 1H, NH), 4.16 (q, *J* = 7.2 Hz, 2H, H-14), 3.93 (br. s, 2H, H-5), 3.35 (br. s, 2H, H-14), 3.07 (br. s, 2H, H-10), 1.54 (s, 9H, C(CH₃)₃), 1.50 (s, 9H, C(CH₃)₃), 1.49 – 1.44 (m, 4H, H-11 and H-13), 1.43 (s, 9H, C(CH₃)₃), 1.33 (t, *J* = 7.1 Hz, 3H, CH₃), 1.30 (s, 2H, H-12).

δ_C (151 MHz, CDCl₃): 164.5 (s, CN), 163.5 (s, CN), 158.7 (s, CO), 155.9 (s, CO), 153.5 (s, CO), 153.3 (s, CO), 140.5 (s, C-1), 136.9 (s, C-6), 135.0 (s, C-9), 134.4 (s, C-4), 130.3 (s, C-2), 129.3 (s, C-7), 125.8 (s, C-3), 122.4 (s, C-8), 83.7 (s, C(CH₃)₃), 79.6 (s, C(CH₃)₃), 60.8 (s, CH₂), 41.0 (s, C-14), 40.9 (s, C-5), 40.3 (s, C-10), 29.6 (s, C-11), 29.2 (s, C-13), 28.4 (s, C(CH₃)₃), 28.1 (s, C(CH₃)₃), 28.0 (s, C(CH₃)₃), 23.8 (s, C-12), 14.6 (s, C-16).

HRMS-ESI (-): m/z calculated for C₃₈H₅₇N₇O₈: 738.4269; found: 738.4205 [M-H]⁻

v_{max}(ATR)/cm⁻¹: 3230 (NH), 2968 (CH₂), 1598, 1366, 1300, 1232, 1148, 748.

***tert*-Butyl-*N*-{6-[(*E*)-[[(4-{[(1*E*)-[[(*tert*-butoxy)carbonyl]amino}{[(*tert*-butoxy)carbonyl]imino})methyl]amino}phenyl)methyl]phenyl]amino)[(ethoxycarbonyl)amino]methylidene]amino]hexyl}carbamate (**43b**)**



Following method D, from **39b** (203 mg, 0.355 mmol), **40c** (122 mg, 0.533 mmol), EDC.HCl (136 mg, 0.7106 mmol) and DIPEA (0.09 mL, 0.533 mmol), **43b** was obtained as a yellow amorphous gum (174 mg, 65%); R_f 0.4 (7: 3 Hex: EtOAc)

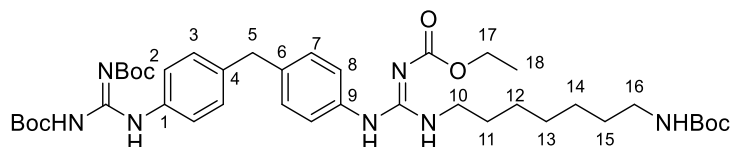
δ_H (400 MHz, $CDCl_3$): 11.64 (s, 1H, NH), 10.62 (s, 1H, NH), 10.28 (s, 1H, NH), 7.51 (d, J = 8.4 Hz, 2H, H-8/H-8'), 7.20 (d, J = 8.0 Hz, 2H, H-2/H-2'), 7.14 (d, J = 8.5 Hz, 2H, H-7/H-7'), 7.10 (d, J = 7.3 Hz, 2H, H-3/H-3'), 4.83 (s, 1H, NH), 4.64 (s, 1H, NH), 4.15 (q, J = 6.9 Hz, 2H, H-16), 3.92 (br. s, 2H, H-5), 3.34 (br. s, 2H, H-25), 3.02 (br. s, 2H, H-10), 1.53 (s, 9H, $C(CH_3)_3$), 1.52 – 1.49 (m, 4H, H-11 and H-14), 1.49 (s, 9H, $C(CH_3)_3$), 1.42 (s, 9H, $C(CH_3)_3$), 1.32 (t, J = 6.9 Hz, 3H, H-17), 1.29 – 1.27 (m, 4H, H-12 and H-13).

δ_C (151 MHz, $CDCl_3$): 164.5 (s, CN), 163.5 (s, CN), 158.8 (s, CO), 155.9 (s, CO), 153.5 (s, CO), 153.3 (s, CO), 140.1 (s, C-1), 136.9 (s, C-6), 135.0 (s, C-9), 134.0 (s, C-4), 130.3 (s, C-2), 129.3 (s, C-7), 125.8 (s, C-3), 122.4 (s, C-8), 83.7 (s, $C(CH_3)_3$), 79.6 (s, $C(CH_3)_3$), 60.7 (s, C-16), 40.9 (s, C-5), 40.8 (s, C-15), 40.2 (s, C-10), 29.9 (s, C-11), 29.3 (s, C-14), 28.4 (s, C-12), 28.1 (s, $C(CH_3)_3$), 28.0 (s, $C(CH_3)_3$), 26.1 (s, $C(CH_3)_3$), 26.1 (s, C-13), 14.6 (s, C-17).

HRMS-ESI (-): m/z calculated for $C_{39}H_{59}N_7O_8$: 752.4425; found: 752.4331 $[M-H]^-$

$\nu_{max}(ATR)/cm^{-1}$: 3245 (NH), 2925 (CH_2), 1625, 1595, 1394, 801, 750.

***tert*-Butyl-*N*-{5-[(*E*)-[[(4-{[(1*E*)-[[(*tert*-butoxy)carbonyl]amino}{[(*tert*-butoxy)carbonyl]imino})methyl]amino}phenyl)methyl]phenyl]amino)[(ethoxycarbonyl)amino]methylidene]amino]pentyl}carbamate (**44b**)**



Following method D, **39b** (130 mg, 0.2266 mmol), **40d** (98 mg, 0.339 mmol), EDC.HCl (87 mg, 0.453 mmol) and DIPEA (0.06 mL, 0.339 mmol), **44b** was obtained as a yellow amorphous gum (75 mg, 45% yield); R_f 0.4 (7: 3 Hex: EtOAc)

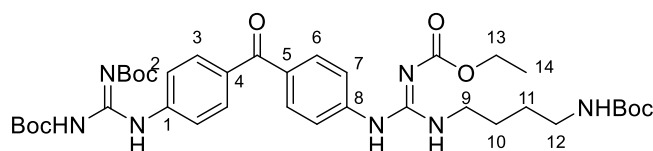
δ_H (600 MHz, $CDCl_3$): 11.64 (s, 1H, NH), 10.65 (s, 1H, NH), 10.29 (s, 1H, NH), 7.53 (d, $J = 8.4$ Hz, 2H, H-8/H-8'), 7.21 (d, $J = 7.5$ Hz, 2H, H-2/H-2'), 7.15 (d, $J = 8.4$ Hz, 2H, H-7/H-7'), 7.14 – 7.08 (m, 2H, H-3/H-3'), 4.69 (s, 1H, NH), 4.55 (s, 1H, NH), 4.16 (q, $J = 7.3$ Hz, 2H, H-17), 3.93 (br. s, 2H, H-5), 3.35 (br. s, 2H, H-16), 3.08 (br. s, 2H, H-10), 1.54 (s, 9H, $C(CH_3)_3$), 1.50 (s, 9H, $C(CH_3)_3$), 1.49 – 1.44 (m, 6H, H-11, H-12 and H-15), 1.44 (s, 9H, $C(CH_3)_3$), 1.33 (t, $J = 7.3$ Hz, 3H, H-18), 1.27-1.22 (m, 4H, H-13 and H-14).

δ_C (151 MHz, $CDCl_3$): 164.5 (s, CN), 163.5 (s, CN), 158.7 (s, CO), 156.0 (s, CO), 153.5 (s, CO), 153.3 (s, CO), 140.1 (s, C-1), 136.9 (s, C-6), 135.0 (s, C-9), 134.0 (s, C-4), 130.3 (s, C-2), 129.3 (s, C-7), 125.8 (s, C-3), 122.4 (s, C-8), 83.7 (s, $C(CH_3)_3$), 79.6 (s, $C(CH_3)_3$), 79.0 (s, $C(CH_3)_3$), 60.8 (s, C-17), 41.1 (s, C-16), 40.9 (s, C-5), 40.5 (s, C-10), 29.9 (s, C-11), 29.4 (s, C-15), 28.8 (s, C-12), 28.4 (s, $C(CH_3)_3$), 28.1 (s, $C(CH_3)_3$), 28.0 (s, $C(CH_3)_3$), 26.6 (s, C-14), 26.6 (s, C-13), 14.6 (s, C-18).

HRMS-ESI (-): m/z calculated for $C_{40}H_{61}N_7O_8$: 766.4582; found: 766.4506 $[M-H]^-$

ν_{max} (ATR)/ cm^{-1} : 3010 (NH), 2949 (CH_2), 1625, 1595, 1393, 801, 751.

tert-Butyl-N-[(1E)-{[4-(4-{[(1Z)-[(4-{[(tert-butoxy)carbonyl]amino}butyl)imino] [(ethoxycarbonyl)amino]methyl]amino}benzoyl)phenyl]amino}{[(tert-butoxy)carbonyl]imino})methyl]carbamate (41c)



Following method D, **39c** (100 mg, 0.1698 mmol), **40a** (45 mg, 0.254 mmol), EDC.HCl (65 mg, 0.339 mmol) and DIPEA (0.04 mL, 0.254 mmol), **41c** was obtained as a yellow amorphous gum (82 mg, 65%); R_f 0.2 (6: 4 Hex: EtOAc)

δ_H (400 MHz, $CDCl_3$): 11.65 (s, 1H, NH), 10.63 (s, 1H, NH), 7.94 – 7.69 (m, 8H, H-2/H-2', H-3/H-3', H-6/H-6' and H-7/H-7'), 7.26 (s, 2H, NH), 4.72 (s, 1H, NH), 4.18 (q, $J = 7.0$ Hz, 2H, H-13),

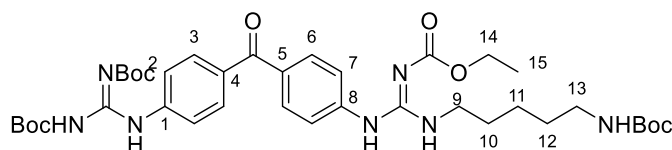
3.43 (br. s, 2H, H-12), 3.16 (br. s, 2H, H-9), 1.68 – 1.57 (m, 4H, H-10 and H-11), 1.56 (s, 9H, C(CH₃)₃), 1.54 (s, 9H, C(CH₃)₃), 1.43 (s, 9H, C(CH₃)₃), 1.32 (t, *J* = 7.0 Hz, 3H, H-14).

δ_c (151 MHz, CDCl₃): 194.5 (s, CO), 163.2 (s, CN), 156.1 (s, CO), 153.3 (s, CO), 140.8 (s, (C-1 and C-8), 133.3 (s, C-4 and C-5), 131.9 (s, C-7), 131.2 (s, C-3), 123.2 (s, C-6), 121.0 (s, C-2), 84.1 (s, C(CH₃)₃), 80.1 (s, C(CH₃)₃), 79.2 (s, C(CH₃)₃), 60.3 (s, C-13), 40.9 (s, C-12), 39.8 (s, C-9), 28.3 (s, C(CH₃)₃), 28.0 (s, 2xC(CH₃)₃), 27.4 (s, C-10), 26.5 (s, C-11), 14.4 (s, C-14).

HRMS-ESI (+): *m/z* calculated for C₃₇H₅₃N₇O₉: 740.3905: found: 740.3911 [M+H]⁺

$\nu_{\max}$ (ATR)/cm⁻¹: 3106 (NH), 2943 (CH₂), 1631, 1589, 1367, 1336, 1301, 767.

***tert*-Butyl-*N*-[[(1*E*)-{[4-(4-{[(1*Z*)-[(5-{[(*tert*-butoxy)carbonyl]amino}pentyl)imino] [(ethoxycarbonyl)amino]methyl]amino}benzoyl)phenyl]amino}]{[(*tert*-butoxy)carbonyl]imino}]methyl]carbamate (**42c**)**



Following method D, **39c** (100 mg, 0.1698 mmol), **40b** (52 mg, 0.254 mmol), EDC.HCl (65 mg, 0.339 mmol) and DIPEA (0.04 mL, 0.254 mmol), **42c** was obtained as a yellow amorphous gum (81 mg, 63%); *R_f* 0.3 (6: 4 Hex: EtOAc)

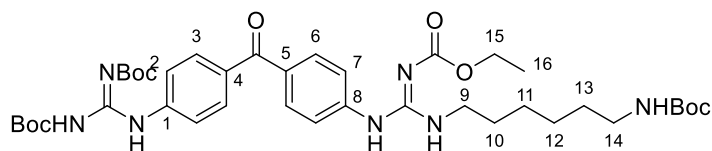
δ_H (600 MHz, CDCl₃): 11.60 (s, 1H, NH), 10.57 (s, 1H, NH), 7.82 – 7.69 (m, 8H, H-2/H-2', H-3/H-3', H-6/H-6' and H-7/H-7'), 7.17 (s, 2H, NH), 4.57 (s, 1H, NH), 4.12 (q, *J* = 7.0 Hz, 2H, H-15), 3.35 (br. s, 3H, H-13), 3.07 (br. s, 2H, H-9), 1.67 – 1.53 (m, 4H, H-10 and H-12), 1.50 (s, 9H, C(CH₃)₃), 1.48 (s, 8H, C(CH₃)₃), 1.38 (s, 9H, C(CH₃)₃), 1.37 – 1.30 (m, 2H, H-11), 1.26 (t, *J* = 7.0 Hz, 3H, H-15).

δ_c (151 MHz, CDCl₃): 194.4 (s, CO), 163.1 (s, CN), 155.9 (s, CO), 153.2 (s, CO), 140.7 (s, (C-1 and C-8), 133.3 (s, C-4 and C-5), 131.8 (s, C-7), 131.1 (s, C-3), 123.2 (s, C-6), 121.0 (s, C-2), 84.0 (s, C(CH₃)₃), 80.0 (s, C(CH₃)₃), 79.0 (s, C(CH₃)₃), 60.3 (s, C-14), 41.2 (s, C-13), 40.1 (s, C-9), 29.6 (s, C-10), 28.8 (s, C-12), 28.3 (s, C(CH₃)₃), 28.0 (s, C(CH₃)₃), 23.9 (s, C-11), 14.4 (s, C-15).

HRMS-ESI (+): *m/z* calculated for C₃₈H₅₅N₇O₉: 754.4061: found: 754.4085 [M+H]⁺

$\nu_{\max}(\text{ATR})/\text{cm}^{-1}$: 3299 (NH), 2937 (CH₂), 1631, 1589, 1366, 1302, 730.

***tert*-Butyl-*N*-[(1*E*)-{4-[4-[[[(1*Z*)-[(6-[[*tert*-butoxy)carbonyl]amino]hexyl]imino] [(ethoxycarbonyl)amino]methyl]amino]benzoyl]phenyl]amino}{[[*tert*-butoxy)carbonyl] imino]}methyl]carbamate (**43c**)**



Following method D, from **39c** (100 mg, 0.169 mmol), **40c** (52 mg, 0.254 mmol), EDC.HCl (65 mg, 0.339 mmol) and DIPEA (0.04 mL, 0.2548 mmol), **43c** was obtained as a yellow amorphous gum (64 mg, 50%); R_f 0.3 (6: 4 Hex: EtOAc)

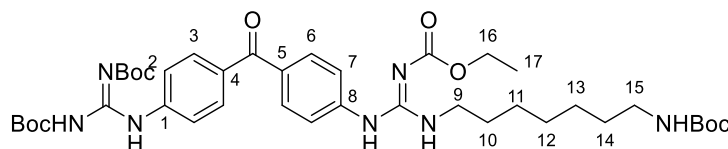
δ_H (400 MHz, CDCl₃): 11.65 (s, 1H, NH), 11.02 (s, 1H, NH), 10.63 (s, 1H, NH), 7.83 - 7.80 (m, 8H, H-2/H-2', H-3/H-3', H-6/H-6' and H-7/H-7'), 7.22 (s, 1H, NH), 4.58 (s, 1H, NH), 4.18 (q, J = 7.1 Hz, 2H, H-15), 3.40 (br. s, 2H, H-14), 3.14 – 3.07 (m, 2H, H-9), 1.55 (s, 9H, C(CH₃)₃), 1.53 (s, 9H, C(CH₃)₃), 1.51 – 1.46 (m, 4H, H-10 and H-13), 1.43 (s, 9H, C(CH₃)₃), 1.37 (br. s, 4H, H-11 and H-12), 1.32 (t, J = 7.1 Hz, 2H, H-16).

δ_C (151 MHz, CDCl₃): 194.5 (s, CO), 163.2 (s, C-N), 156.0 (s, CO), 153.3 (s, CO), 140.7 (s, (C-1 and C-8), 133.3 (s, C-4 and C-5), 131.9 (s, C-7), 131.2 (s, C-3), 123.2 (s, C-6), 121.0 (s, C-2), 84.2 (s, C(CH₃)₃), 80.1 (s, C(CH₃)₃), 60.3 (s, C-16), 41.1 (s, C-14), 40.3 (s, C-9), 29.9 (s, C-10), 29.1 (s, C-13), 28.4 (s, C(CH₃)₃), 28.1 (s, C(CH₃)₃), 28.0 (s, C(CH₃)₃), 26.4 (s, C-11), 26.2 (s, C-12), 14.4 (s, C-16).

HRMS-ESI (+): m/z calculated for C₃₉H₅₇N₇O₉: 767.4218; found: 802.3923 [M+Na]⁺

$\nu_{\max}(\text{ATR})/\text{cm}^{-1}$: 3299 (NH), 2943 (CH₂), 1630, 1589, 1367, 1337, 1301, 749.

***tert*-Butyl-*N*-[(1*E*)-{4-(4-{[(1*Z*)-[(7-{[(*tert*-butoxy)carbonyl]amino}heptyl)imino] [(ethoxycarbonyl)amino]methyl]amino}benzoyl)phenyl]amino){[(*tert*-butoxy)carbonyl]imino})methyl]carbamate (**44c**)**



Following method D, from **39c** (100 mg, 0.1698 mmol), **40d** (59 mg, 0.254 mmol), EDC.HCl (65 mg, 0.339 mmol) and DIPEA (0.04 mL, 0.254 mmol), **44c** was obtained as a yellow amorphous gum (70 mg, 55%); R_f 0.3 (6: 4 Hex: EtOAc)

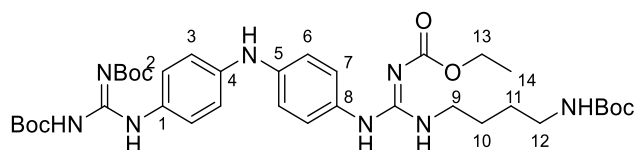
δ_H (400 MHz, $CDCl_3$): 11.65 (s, 1H, NH), 10.63 (s, 1H, NH), 7.85 – 7.77 (m, 8H, H-2/H-2', H-3/H-3', H-6/H-6' and H-7/H-7'), 7.28 (s, 1H, NH), 7.21 (s, 2H, NH), 4.56 (s, 1H, NH), 4.18 (q, J = 7.1 Hz, 2H, H-16), 3.43 – 3.36 (m, 2H, H-15), 3.14 – 3.06 (m, 2H, H-9), 1.63 – 1.57 (m, 4H, H-10 and H-14), 1.56 (s, 9H, $C(CH_3)_3$), 1.54 (s, 9H, $C(CH_3)_3$), 1.53 – 1.50 (m, 4H, H-11 and H-13), 1.44 (s, 9H, $C(CH_3)_3$), 1.34 (br. s, 2H, H-12), 1.31 (t, J = 7.1 Hz, 3H, H-17).

δ_C (151 MHz, $CDCl_3$): 194.5 (s, CO), 163.1 (s, CN), 156.0 (s, CO), 153.3 (s, CO), 140.8 (s, (C-1 and C-8), 133.3 (s, C-4 and C-5), 131.9 (s, C-7), 131.2 (s, C-3), 123.1 (s, C-6), 121.0 (s, C-2), 41.3 (s, C-15), 40.5 (s, C-9), 29.9 (s, C-10), 29.1 (s, C-14), 28.8 (s, $C(CH_3)_3$), 28.4 (s, $C(CH_3)_3$), 28.1 (s, $C(CH_3)_3$), 26.7 (s, C-11), 26.6 (s, C-13), 21.0 (s, C-12), 14.4 (s, C-17)

HRMS-ESI (+): m/z calculated for $C_{40}H_{59}N_7O_9$: 782.4374; found: 782.4422 $[M+H]^+$

$\nu_{max}(ATR)/cm^{-1}$: 3301 (NH), 2943 (CH_2), 1629, 1588, 1367, 1301, 767.

***tert*-Butyl *N*-[(1*E*)-{4-[(1*E*)-[(4-{[(*tert*-butoxy)carbonyl]amino}butyl)amino] [(ethoxycarbonyl)imino]methyl]amino}phenyl]amino]phenyl]amino){[(*tert*-butoxy)carbonyl]imino})methyl]carbamate (**41d**)**



Following method D, from **39d** (87 mg, 0.152 mmol), **40a** (43 mg, 0.228 mmol), EDC.HCl (58 mg, 0.304 mmol) and DIPEA (0.04 mL, 0.228 mmol), **41d** was obtained as a purple gum (82 mg, 75% yield); R_f 0.5 (6: 4 Hex: EtOAc)

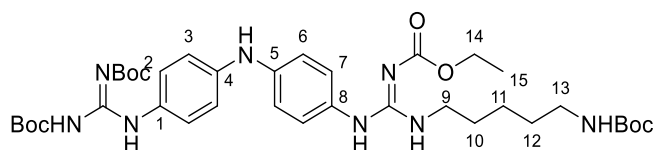
δ_H (400 MHz, $CDCl_3$): 11.67 (s, 1H, NH), 10.47 (s, 1H, NH), 10.27 (s, 1H, NH), 7.53 (d, $J = 8.8$ Hz, 2H, H-2/H-2'), 7.11 – 7.00 (m, 8H, H-3/H-3', H-6/H-6' and H-7/H-7'), 5.85 (s, 1H, NH), 4.64 (s, 1H, NH), 4.19 (q, $J = 7.1$ Hz, 2H, H-13), 3.39 (br. s, 2H, H-12), 3.13 (br. s, 2H, H-9), 1.56 (s, 9H, $C(CH_3)_3$), 1.51 (s, 9H, $C(CH_3)_3$), 1.47 – 1.44 (m, 4H, H-10 and H-11), 1.44 (s, 9H, $C(CH_3)_3$), 1.34 (t, $J = 7.1$ Hz, 3H, H-14).

δ_C (151 MHz, $CDCl_3$): 164.5 (s, CN), 163.6 (s, CN), 159.4 (s, CO), 156.0 (s, CO), 153.6 (s, CO), 153.3 (s, CO), 143.2 (s, C-1), 138.9 (s, C-8), 131.1 (s, C-4 and C-5), 127.7 (s, C-7), 123.7 (s, C-2), 119.7 (s, C-6), 117.4 (s, C-3), 83.6 (s, $C(CH_3)_3$), 79.5 (s, $C(CH_3)_3$), 79.1 (s, $C(CH_3)_3$), 60.3 (s, C-13), 40.6 (s, C-12), 40.0 (s, C-9), 28.4 (s, $C(CH_3)_3$), 28.2 (s, $C(CH_3)_3$), 28.1 (s, $C(CH_3)_3$), 27.3 (s, C-11), 26.9 (s, C-10), 14.7 (s, C-14).

HRMS-ESI (+): m/z calculated for $C_{36}H_{54}N_8O_8$: 727.4065 found: 727.4138 $[M+H]^+$

$\nu_{max}(ATR)/cm^{-1}$: 3328 (NH), 2950 (CH_2), 2918 (CH_2), 2160, 2032, 1688 (CN), 1592, 799.

tert-Butyl-N-[(1E)-({4-[(1E)-[(5-[(tert-butoxy)carbonyl]amino}pentyl)amino] [(ethoxycarbonyl)imino]methyl]amino}phenyl)amino]phenyl)amino){[(tert-butoxy)carbonyl]imino})methyl]carbamate (42d**)**



Following method D, from **39d** (70 mg, 0.122 mmol), **40b** (37 mg, 0.183 mmol), EDC.HCl (47 mg, 0.244 mmol) and DIPEA (0.03 mL, 0.183 mmol), **42d** was obtained as a purple gum (50 mg, 55% yield) R_f 0.4 (6: 4 Hex: EtOAc)

δ_H (400 MHz, $CDCl_3$): 11.62 (s, 1H, NH), 10.41 (s, 1H, NH), 10.20 (s, 1H, NH), 7.45 (d, $J = 8.9$ Hz, 2H, H-2/H-2'), 7.01 – 6.99 (m, 4H, H-6/H-6' and H-7/H-7'), 6.97 (d, $J = 8.9$ Hz, 2H, H-3/H-3'), 5.99 (s, 1H, NH), 4.49 (s, 1H, NH), 4.12 (q, $J = 7.1$ Hz, 2H, H-14), 3.30 (br. s, 2H, H-13), 3.04 (br.

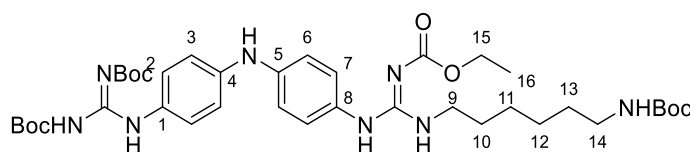
s, 2H, H-9), 1.51 (s, 9H, C(CH₃)₃), 1.45 (s, 9H, C(CH₃)₃), 1.40 (s, 9H, C(CH₃)₃), 1.29 (t, *J* = 7.1 Hz, 2H, H-15).

δ_c (151 MHz, CDCl₃): 164.5 (s, CN), 163.5 (s, CN), 159.4 (s, CO), 155.9 (s, CO), 153.6 (s, CO), 153.3 (s, CO), 143.2 (s, C-1), 138.9 (s, C-8), 130.7 (s, C-4 and C-5), 127.6 (s, C-7), 123.8 (s, C-2), 119.4 (s, C-6), 117.3 (s, C-3), 83.6 (s, C(CH₃)₃), 79.5 (s, C(CH₃)₃), 60.3 (s, C-14), 40.9 (s, C-13), 40.1 (s, C-9), 29.6 (s, C-12), 29.1 (s, C-10), 28.3 (s, C(CH₃)₃), 28.1 (s, C(CH₃)₃), 28.0 (s, C(CH₃)₃), 23.8 (s, C-11), 14.6 (s, C-15).

HRMS-ESI (+): *m/z* calculated for C₃₇H₅₆N₈O₈: 741.4221 found: 741.4294 [M+H]⁺

$\nu_{\max}$ (ATR)/cm⁻¹: 3300 (NH), 2991 (CH₂), 2900 (CH₂), 2111, 2032, 1691 (CO), 1592, 799.

***tert*-Butyl -*N*-[[(1*E*)-{4-[[[(1*E*)-[(6-[(*tert*-butoxy)carbonyl]amino]hexyl)amino] [(ethoxycarbonyl)imino]methyl]amino]phenyl]amino]phenyl]amino){[(*tert*-butoxy) carbonyl]imino)}methyl]carbamate (**43d**)**



Following method D, from **39d** (98 mg, 0.172 mmol), **40c** (56 mg, 0.258 mmol), EDC.HCl (66 mg, 0.344 mmol) and DIPEA (0.07 mL, 0.258 mmol), **43d** was obtained as a purple gum (56 mg, 43% yield); *R_f* 0.3 (6: 4 Hex: EtOAc)

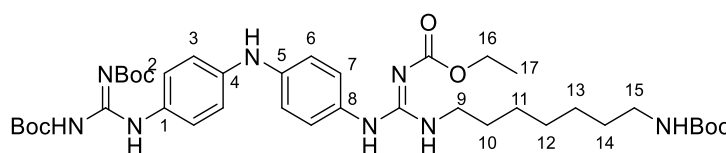
δ_H (400 MHz, CDCl₃): 11.67 (s, 1H, NH), 10.42 (s, 1H, NH), 10.26 (s, 1H, NH), 7.52 (d, *J* = 8.6 Hz, 2H, H-2/H-2'), 7.16 – 6.93 (m, 6H, H-3/H-3', H-6/H-6' and H-7/H-7'), 5.95 (s, 1H, NH), 4.56 (s, 1H, NH), 4.21 (q, *J* = 7.8 Hz, 1H, H-15), 3.40 – 3.32 (m, 2H, H-14), 3.15 – 3.04 (m, 2H, H-9), 1.56 (s, 9H, C(CH₃)₃), 1.51 (s, 9H, C(CH₃)₃), 1.51 – 1.47 (m, 8H, H-10, H-11, H-12 and H-13), 1.45 (s, 18H, 2xC(CH₃)₃), 1.40 – 1.31 (m, 2H, H-16).

δ_c (151 MHz, CDCl₃): 164.9 (s, CN), 163.6 (s, CN), 153.6 (s, CO), 153.5 (s, CO), 153.3 (s, CO), 143.0 (s, C-1), 138.9 (s, C-8), 131.1 (s, C-4 and C-5), 127.8 (s, C-2), 123.6 (s, C-7), 119.6 (s, C-6), 117.5 (s, C-3), 83.6 (s, C(CH₃)₃), 79.5 (s, C(CH₃)₃), 79.0 (s, C(CH₃)₃), 60.7 (s, C-16), 40.9 (s, C-14), 40.4 (s, C-9), 29.9 (s, C-10), 29.4 (s, C-13), 28.4 (s, C(CH₃)₃), 28.2 (s, C(CH₃)₃), 28.1 (s, C(CH₃)₃), 26.3 (s, C-11), 21.0 (s, C-12), 14.7 (s, C-16).

HRMS-ESI (+): m/z calculated for $C_{38}H_{58}N_8O_8$: 755.4378 found: 755.4469 $[M+H]^+$

$\nu_{\max}(\text{ATR})/\text{cm}^{-1}$: 3321 (NH), 2995 (CH_2), 2912 (CH_2), 2098, 2021, 1700 (CO), 800.

***tert*-Butyl -N-[(1E)-{4-[(4-[(1E)-[(7-[(*tert*-butoxy)carbonyl]amino}heptyl]amino] [(ethoxycarbonyl)imino]methyl]amino}phenyl)amino]phenyl)amino){[(*tert*-butoxy) carbonyl]imino)}methyl]carbamate (**44d**)**



Following method D, from **39d** (110 mg, 0.192 mmol), **40d** (66 mg, 0.288 mmol), EDC.HCl (73 mg, 0.384 mmol) and DIPEA (0.05 mL, 0.288 mmol), **44d** was obtained as a purple gum (90 mg, 61% yield); R_f 0.3 (6: 4 Hex: EtOAc)

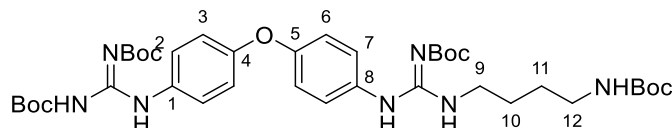
δ_H (400 MHz, CDCl_3): 11.67 (s, 1H, NH), 10.47 (s, 1H, NH), 10.28 (s, 1H, NH), 7.54 (d, $J = 8.7$ Hz, 2H, H-2/H-2'), 7.10 – 7.07 (m, H-6/H-6' and H-7/H-7'), 7.05 (d, $J = 8.7$ Hz, 2H, H-3/H-3'), 5.87 (s, 1H, NH), 4.54 (s, 1H, NH), 4.14 (q, $J = 7.1$ Hz, 2H, H-16), 3.37 (br. s, 2H, H-15), 3.09 (br. s, 2H, H-9), 1.60 (br. s, 4H, H-10 and H-14), 1.56 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.52 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.51 – 1.47 (m, 4H, H-11 and H-13), 1.45 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.35 (t, $J = 7.1$ Hz, 4H, H-17), 1.30 – 1.27 (m, 2H, H-12).

δ_C (151 MHz, CDCl_3): 164.5 (s, CN), 163.6 (s, CN), 159.4 (s, CO), 155.9 (s, CO), 153.6 (s, CO), 153.3 (s, CO), 143.2 (s, C-1), 138.7 (s, C-8), 131.0 (s, C-4 and C-5), 127.7 (s, C-7), 123.6 (s, C-2), 119.6 (s, C-6), 117.4 (s, C-3), 83.6 (s, $\text{C}(\text{CH}_3)_3$), 79.5 (s, $\text{C}(\text{CH}_3)_3$), 60.3 (s, C-16), 41.1 (s, C-15), 40.4 (s, C-9), 29.9 (s, C-14), 29.3 (s, C-10), 28.8 (s, C-13), 28.4 (s, C-11 and $\text{C}(\text{CH}_3)_3$), 28.2 (s, $\text{C}(\text{CH}_3)_3$), 28.1 (s, $\text{C}(\text{CH}_3)_3$), 26.6 (s, C-12), 14.6 (s, C-17).

HRMS-ESI (+): m/z calculated for $C_{39}H_{60}N_8O_8$: 769.4534 found: 769.4608 $[M+H]^+$

$\nu_{\max}(\text{ATR})/\text{cm}^{-1}$: 3213 (NH), 2894 (CH_2), 2799 (CH_2), 1699 (CO), 1586, 1088, 801.

***tert*-Butyl-*N*-[(1*E*)-{[4-(4-{[(1*Z*)-{[(*tert*-butoxy)carbonyl]amino}[(4-{[(*tert*-butoxy)carbonyl]amino}butyl)imino]methyl]amino}phenoxy)phenyl]amino}{[(*tert*-butoxy)carbonyl]imino}) methyl]carbamate (45a)**



Following method E, from **41a** (100 mg, 0.137 mmol) and *t*-BuOH (5 mL), **45a** was obtained as pale brown amorphous gum. (80 mg, 77% yield); R_f 0.4 (7:3 Hex: EtOAc)

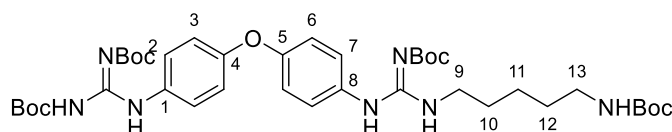
δ_H (400 MHz, $CDCl_3$): 11.62 (s, 1H, NH), 10.29 (s, 1H, NH), 7.56 (d, $J = 8.5$ Hz, 2H, H-3/H-3'), 7.08 (br. s, 2H, H-6/H-6'), 6.97 – 6.89 (m, 4H, H-2/H-2' and H-7/H-7'), 4.71 (s, 1H, NH), 3.28 (br. s, 2H, H-12), 3.09 (br. s, 2H, H-9), 1.50 (s, 9H, $C(CH_3)_3$), 1.48 (s, 18H, $2 \times C(CH_3)_3$), 1.47 – 1.43 (m, 4H, H-10 and H-11), 1.40 (s, 9H, $C(CH_3)_3$).

δ_c (151 MHz, $CDCl_3$): 163.5 (s, CN), 158.9 (s, CO), 156.6 (s, C-4), 156.0 (s, CO), 153.5 (s, CO), 153.3 (s, CO), 153.0 (s, C-5), 132.8 (s, C-1), 130.7 (s, C-8), 127.6 (s, C-6), 123.8 (s, C-3), 119.6 (s, C-2), 119.2 (s, C-7), 82.6 (s, $\underline{C(CH_3)_3}$), 79.6 (s, $\underline{C(CH_3)_3}$), 79.0 (s, $\underline{C(CH_3)_3}$), 40.8 (s, C-12), 33.6 (s, C-9), 28.3 (s, $\underline{C(CH_3)_3}$), 28.2 (s, $\underline{C(CH_3)_3}$), 28.1 (s, $\underline{C(CH_3)_3}$), 28.0 (s, $\underline{C(CH_3)_3}$), 27.1 (s, C-10), 26.6 (s, C-11).

HRMS-ESI (+): m/z calculated for $C_{38}H_{57}N_7O_9$: 756.9140; found: 756.9182 $[M+H]^+$

$\nu_{max}(ATR)/cm^{-1}$: 3200 (NH), 2899 (CH_2), 1639 (CN), 1597, 1114, 803.

***tert*-Butyl-*N*-[(1*Z*)-{[4-(4-{[(1*Z*)-{[(*tert*-butoxy)carbonyl]amino}[(5-{[(*tert*-butoxy)carbonyl]amino}pentyl)imino]methyl]amino}phenoxy)phenyl]amino}{[(*tert*-butoxy)carbonyl]imino}) methyl]carbamate (46a)**



Following method E, from **42a** (60 mg, 0.080 mmol) and *t*-BuOH (5 mL), **46a** was obtained as pale brown amorphous gum. (24 mg, 38% yield); R_f 0.4 (7: 3 Hex: EtOAc)

δ_c (151 MHz, $CDCl_3$): 162.9 (s, CN), 158.9 (s, CO), 156.6 (s, C-4), 155.6 (s, CO), 154.3 (s, CO), 153.6 (s, CO), 152.8 (s, C-5), 133.5 (s, C-1), 129.9 (s, C-8), 126.8 (s, C-6), 122.6 (s, C-3), 119.8 (s, C-2), 119.3 (s, C-7), 83.0 (s, $C(CH_3)_3$), 78.9 (s, $C(CH_3)_3$), 78.8 (s, $C(CH_3)_3$), 42.3 (s, C-13), 41.6 (s, C-9), 28.6 (s, C-10), 28.2 (s, C-12), 28.1 (s, $C(CH_3)_3$), 27.9 (s, $C(CH_3)_3$), 27.8 (s, $C(CH_3)_3$), 22.8 (s, C-11).

$$\nu_{\max}(\text{ATR})/\text{cm}^{-1}: 3208 (\text{NH}), 2987 (\text{CH}_2), 1623, 1585, 1499, 1355, 1151, 832.$$
CC(C)(C)C(=O)Nc1ccc(cc1)C(=O)Nc2ccc(cc2)Oc3ccc(cc3)C(=O)Nc4ccc(cc4)C(=O)Nc5ccc(cc5)C(=O)Nc6ccc(cc6)C(=O)Nc7ccc(cc7)C(=O)Nc8ccc(cc8)C(=O)Nc9ccc(cc9)C(=O)Nc10ccc(cc10)C(=O)Nc11ccc(cc11)C(=O)Nc12ccc(cc12)C(=O)Nc13ccc(cc13)C(=O)Nc14ccc(cc14)C(=O)Nc15ccc(cc15)C(=O)Nc16ccc(cc16)C(=O)Nc17ccc(cc17)C(=O)Nc18ccc(cc18)C(=O)Nc19ccc(cc19)C(=O)Nc20ccc(cc20)C(=O)Nc21ccc(cc21)C(=O)Nc22ccc(cc22)C(=O)Nc23ccc(cc23)C(=O)Nc24ccc(cc24)C(=O)Nc25ccc(cc25)C(=O)Nc26ccc(cc26)C(=O)Nc27ccc(cc27)C(=O)Nc28ccc(cc28)C(=O)Nc29ccc(cc29)C(=O)Nc30ccc(cc30)C(=O)Nc31ccc(cc31)C(=O)Nc32ccc(cc32)C(=O)Nc33ccc(cc33)C(=O)Nc34ccc(cc34)C(=O)Nc35ccc(cc35)C(=O)Nc36ccc(cc36)C(=O)Nc37ccc(cc37)C(=O)Nc38ccc(cc38)C(=O)Nc39ccc(cc39)C(=O)Nc40ccc(cc40)C(=O)Nc41ccc(cc41)C(=O)Nc42ccc(cc42)C(=O)Nc43ccc(cc43)C(=O)Nc44ccc(cc44)C(=O)Nc45ccc(cc45)C(=O)Nc46ccc(cc46)C(=O)Nc47ccc(cc47)C(=O)Nc48ccc(cc48)C(=O)Nc49ccc(cc49)C(=O)Nc50ccc(cc50)C(=O)Nc51ccc(cc51)C(=O)Nc52ccc(cc52)C(=O)Nc53ccc(cc53)C(=O)Nc54ccc(cc54)C(=O)Nc55ccc(cc55)C(=O)Nc56ccc(cc56)C(=O)Nc57ccc(cc57)C(=O)Nc58ccc(cc58)C(=O)Nc59ccc(cc59)C(=O)Nc60ccc(cc60)C(=O)Nc61ccc(cc61)C(=O)Nc62ccc(cc62)C(=O)Nc63ccc(cc63)C(=O)Nc64ccc(cc64)C(=O)Nc65ccc(cc65)C(=O)Nc66ccc(cc66)C(=O)Nc67ccc(cc67)C(=O)Nc68ccc(cc68)C(=O)Nc69ccc(cc69)C(=O)Nc70ccc(cc70)C(=O)Nc71ccc(cc71)C(=O)Nc72ccc(cc72)C(=O)Nc73ccc(cc73)C(=O)Nc74ccc(cc74)C(=O)Nc75ccc(cc75)C(=O)Nc76ccc(cc76)C(=O)Nc77ccc(cc77)C(=O)Nc78ccc(cc78)C(=O)Nc79ccc(cc79)C(=O)Nc80ccc(cc80)C(=O)Nc81ccc(cc81)C(=O)Nc82ccc(cc82)C(=O)Nc83ccc(cc83)C(=O)Nc84ccc(cc84)C(=O)Nc85ccc(cc85)C(=O)Nc86ccc(cc86)C(=O)Nc87ccc(cc87)C(=O)Nc88ccc(cc88)C(=O)Nc89ccc(cc89)C(=O)Nc90ccc(cc90)C(=O)Nc91ccc(cc91)C(=O)Nc92ccc(cc92)C(=O)Nc93ccc(cc93)C(=O)Nc94ccc(cc94)C(=O)Nc95ccc(cc95)C(=O)Nc96ccc(cc96)C(=O)Nc97ccc(cc97)C(=O)Nc98ccc(cc98)C(=O)Nc99ccc(cc99)C(=O)Nc100ccc(cc100)C(=O)Nc101ccc(cc101)C(=O)Nc102ccc(cc102)C(=O)Nc103ccc(cc103)C(=O)Nc104ccc(cc104)C(=O)Nc105ccc(cc105)C(=O)Nc106ccc(cc106)C(=O)Nc107ccc(cc107)C(=O)Nc108ccc(cc108)C(=O)Nc109ccc(cc109)C(=O)Nc110ccc(cc110)C(=O)Nc111ccc(cc111)C(=O)Nc112ccc(cc112)C(=O)Nc113ccc(cc113)C(=O)Nc114ccc(cc114)C(=O)Nc115ccc(cc115)C(=O)Nc116ccc(cc116)C(=O)Nc117ccc(cc117)C(=O)Nc118ccc(cc118)C(=O)Nc119ccc(cc119)C(=O)Nc120ccc(cc120)C(=O)Nc121ccc(cc121)C(=O)Nc122ccc(cc122)C(=O)Nc123ccc(cc123)C(=O)Nc124ccc(cc124)C(=O)Nc125ccc(cc125)C(=O)Nc126ccc(cc126)C(=O)Nc127ccc(cc127)C(=O)Nc128ccc(cc128)C(=O)Nc129ccc(cc129)C(=O)Nc130ccc(cc130)C(=O)Nc131ccc(cc131)C(=O)Nc132ccc(cc132)C(=O)Nc133ccc(cc133)C(=O)Nc134ccc(cc134)C(=O)Nc135ccc(cc135)C(=O)Nc136ccc(cc136)C(=O)Nc137ccc(cc137)C(=O)Nc138ccc(cc138)C(=O)Nc139ccc(cc139)C(=O)Nc140ccc(cc140)C(=O)Nc141ccc(cc141)C(=O)Nc142ccc(cc142)C(=O)Nc143ccc(cc143)C(=O)Nc144ccc(cc144)C(=O)Nc145ccc(cc145)C(=O)Nc146ccc(cc146)C(=O)Nc147ccc(cc147)C(=O)Nc148ccc(cc148)C(=O)Nc149ccc(cc149)C(=O)Nc150ccc(cc150)C(=O)Nc151ccc(cc151)C(=O)Nc152ccc(cc152)C(=O)Nc153ccc(cc153)C(=O)Nc154ccc(cc154)C(=O)Nc155ccc(cc155)C(=O)Nc156ccc(cc156)C(=O)Nc157ccc(cc157)C(=O)Nc158ccc(cc158)C(=O)Nc159ccc(cc159)C(=O)Nc160ccc(cc160)C(=O)Nc161ccc(cc161)C(=O)Nc162ccc(cc162)C(=O)Nc163ccc(cc163)C(=O)Nc164ccc(cc164)C(=O)Nc165ccc(cc165)C(=O)Nc166ccc(cc166)C(=O)Nc167ccc(cc167)C(=O)Nc168ccc(cc168)C(=O)Nc169ccc(cc169)C(=O)Nc170ccc(cc170)C(=O)Nc171ccc(cc171)C(=O)Nc172ccc(cc172)C(=O)Nc173ccc(cc173)C(=O)Nc174ccc(cc174)C(=O)Nc175ccc(cc175)C(=O)Nc176ccc(cc176)C(=O)Nc177ccc(cc177)C(=O)Nc178ccc(cc178)C(=O)Nc179ccc(cc179)C(=O)Nc180ccc(cc180)C(=O)Nc181ccc(cc181)C(=O)Nc182ccc(cc182)C(=O)Nc183ccc(cc183)C(=O)Nc184ccc(cc184)C(=O)Nc185ccc(cc185)C(=O)Nc186ccc(cc186)C(=O)Nc187ccc(cc187)C(=O)Nc188ccc(cc188)C(=O)Nc189ccc(cc189)C(=O)Nc190ccc(cc190)C(=O)Nc191ccc(cc191)C(=O)Nc192ccc(cc192)C(=O)Nc193ccc(cc193)C(=O)Nc194ccc(cc194)C(=O)Nc195ccc(cc195)C(=O)Nc196ccc(cc196)C(=O)Nc197ccc(cc197)C(=O)Nc198ccc(cc198)C(=O)Nc199ccc(cc199)C(=O)Nc200ccc(cc200)C(=O)Nc201ccc(cc201)C(=O)Nc202ccc(cc202)C(=O)Nc203ccc(cc203)C(=O)Nc204ccc(cc204)C(=O)Nc205ccc(cc205)C(=O)Nc206ccc(cc206)C(=O)Nc207ccc(cc207)C(=O)Nc208ccc(cc208)C(=O)Nc209ccc(cc209)C(=O)Nc210ccc(cc210)C(=O)Nc211ccc(cc211)C(=O)Nc212ccc(cc212)C(=O)Nc213ccc(cc213)C(=O)Nc214ccc(cc214)C(=O)Nc215ccc(cc215)C(=O)Nc216ccc(cc216)C(=O)Nc217ccc(cc217)C(=O)Nc218ccc(cc218)C(=O)Nc219ccc(cc219)C(=O)Nc220ccc(cc220)C(=O)Nc221ccc(cc221)C(=O)Nc222ccc(cc222)C(=O)Nc223ccc(cc223)C(=O)Nc224ccc(cc224)C(=O)Nc225ccc(cc225)C(=O)Nc226ccc(cc226)C(=O)Nc227ccc(cc227)C(=O)Nc228ccc(cc228)C(=O)Nc229ccc(cc229)C(=O)Nc230ccc(cc230)C(=O)Nc231ccc(cc231)C(=O)Nc232ccc(cc232)C(=O)Nc233ccc(cc233)C(=O)Nc234ccc(cc234)C(=O)Nc235ccc(cc235)C(=O)Nc236ccc(cc236)C(=O)Nc237ccc(cc237)C(=O)Nc238ccc(cc238)C(=O)Nc239ccc(cc239)C(=O)Nc240ccc(cc240)C(=O)Nc241ccc(cc241)C(=O)Nc242ccc(cc242)C(=O)Nc243ccc(cc243)C(=O)Nc244ccc(cc244)C(=O)Nc245ccc(cc245)C(=O)Nc246ccc(cc246)C(=O)Nc247ccc(cc247)C(=O)Nc248ccc(cc248)C(=O)Nc249ccc(cc249)C(=O)Nc250ccc(cc250)C(=O)Nc251ccc(cc251)C(=O)Nc252ccc(cc252)C(=O)Nc253ccc(cc253)C(=O)Nc254ccc(cc254)C(=O)Nc255ccc(cc255)C(=O)Nc256ccc(cc256)C(=O)Nc257ccc(cc257)C(=O)Nc258ccc(cc258)C(=O)Nc259ccc(cc259)C(=O)Nc260ccc(cc260)C(=O)Nc261ccc(cc261)C(=O)Nc262ccc(cc262)C(=O)Nc263ccc(cc263)C(=O)Nc264ccc(cc264)C(=O)Nc265ccc(cc265)C(=O)Nc266ccc(cc266)C(=O)Nc267ccc(cc267)C(=O)Nc268ccc(cc268)C(=O)Nc269ccc(cc269)C(=O)N

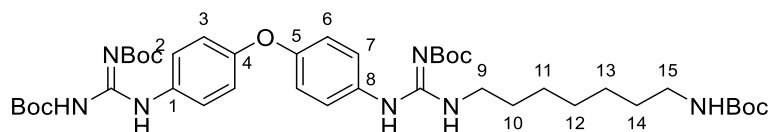
δ_{H} (400 MHz, CDCl_3): 11.62 (s, 1H, NH), 10.30 (s, 1H, NH), 7.57 (d, $J = 7.8$ Hz, 2H, H-3/H-3'), 7.12 (br. s, 2H, H-6/H-6'), 6.98 (br. s, 4H, H-2/H-2' and H-7/H-7'), 4.53 (s, 1H, NH), 3.27 (br. s, 2H, H-14), 3.05 (br. s, 2H, H-9), 1.52 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.49 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.48 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.46 – 1.44 (m, 4H, H-10 and H-13), 1.41 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.27 (s, 4H, H-11 and H-12).

δ_c (101 MHz, $CDCl_3$): 163.5 (s, CN), 158.9 (s, CO), 156.6 (s, C-4), 156.0 (s, CO), 153.5 (s, CO), 153.3 (s, CO), 153.0 (s, C-5), 132.8 (s, C-1), 130.7 (s, C-8), 127.6 (s, C-6), 123.8 (s, C-3), 119.8 (s, C-2), 119.4 (s, C-7), 83.8 (s, $\underline{C}(CH_3)_3$), 79.7 (s, $\underline{C}(CH_3)_3$), 79.0 (s, $\underline{C}(CH_3)_3$), 40.8 (s, C-14), 40.3 (s, C-9), 29.9 (s, C-10), 29.7 (s, C-13), 29.3 (s, C-11), 28.4 (s, $C(\underline{C}H_3)_3$), 28.3 (s, $C(\underline{C}H_3)_3$), 28.1 (s, $C(\underline{C}H_3)_3$), 28.0 (s, $C(\underline{C}H_3)_3$), 26.2 (s, C-12).

HRMS-ESI (+): m/z calculated for C₄₀H₆₁N₇O₉: 784.4531; found: 784.4580 [M+H]⁺

$\nu_{\max}$ (ATR)/cm⁻¹: 3201 (NH), 2919 (CH₂), 1595, 1365, 1111, 802.

***tert*-Butyl-N-[(1E)-{[4-(4-{[(1Z)-{[(*tert*-butoxy)carbonyl]amino}{7-{[(*tert*-butoxy)carbonyl]amino}heptyl]imino)methyl]amino}phenoxy)phenyl]amino}{[(*tert*-butoxy)carbonyl]imino})methyl]carbamate (**48a**)**



Following method E, from **44a** (100 mg, 0.1299 mmol) and *t*-BuOH (5 mL), **48a** was obtained as pale brown amorphous gum. (35 mg, 34% yield); R_f 0.4 (7: 3 Hex: EtOAc).

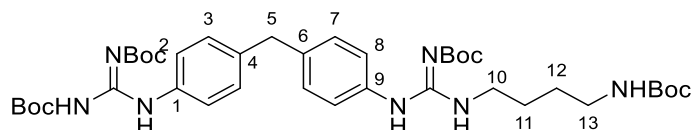
δ_{H} (400 MHz, CDCl₃): 11.62 (s, 1H, NH), 10.31 (s, 1H, NH), 7.58 (d, *J* = 8.2 Hz, 2H, H-3/H-3'), 7.11 (br. s, 2H, H-6/H-6'), 6.99 (br. s, 2H, H-2/H-2' and H-7/H-7'), 4.57 (s, 1H, NH), 3.67 – 3.53 (m, 2H, H-15), 3.06 (br. s, 2H, H-9), 1.52 (s, 9H, C(CH₃)₃), 1.50 (s, 9H, C(CH₃)₃), 1.48 (s, 9H, C(CH₃)₃), 1.41 (s, 9H, C(CH₃)₃), 1.30 – 1.26 (m, 10H, H-10, H-11, H-12, H-13 and H-14).

δ_{C} (151 MHz, CDCl₃): 163.4 (s, CN), 158.8 (s, CO), 156.3 (s, C-4), 154.9 (s, CO), 153.3 (s, CO), 152.9 (s, CO), 153.0 (s, C-5), 132.5 (s, C-1), 129.9 (s, C-8), 127.2 (s, C-6), 123.7 (s, C-3), 119.7 (s, C-2), 119.2 (s, C-7), 83.5 (s, C(CH₃)₃), 78.9 (s, C(CH₃)₃), 78.5 (s, C(CH₃)₃), 42.0 (s, C-15), 41.4 (s, C-9), 29.0 (s, C-10), 28.9 (s, C-14), 28.8 (s, C-13), 28.1 (s, C(CH₃)₃), 27.9 (s, C(CH₃)₃), 27.7 (s, C(CH₃)₃), 26.6 (s, C-11), 25.9 (s, C-12).

HRMS-ESI (+): m/z calculated for C₄₁H₆₃N₇O₉: 798.4687; found: 798.4695 [M+H]⁺

$\nu_{\max}$ (ATR)/cm⁻¹: 3233 (NH), 2989 (CH₂), 1691 (CO), 1641 (CN), 1490, 823.

***tert*-Butyl -N-[(1E)-{4-[4-[(1E)-{[(*tert*-butoxy)carbonyl]amino}{4-[(*tert*-butoxy)carbonyl]amino}butyl]imino)methyl]amino}phenyl)methyl]phenyl]amino){[(*tert*-butoxy)carbonyl]imino})methyl]carbamate (**45b**)**



Following method E, from **41b** (268 mg, 0.370 mmol) and *t*-BuOH (5 mL), **45b** was obtained as pale brown amorphous gum (210 mg, 77% yield); R_f 0.5 (7: 3 Hex: EtOAc)

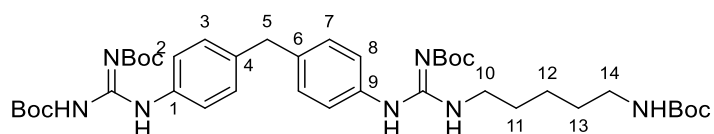
δ_H (400 MHz, $CDCl_3$): δ 11.61 (s, 1H, NH), 10.65 (s, 1H, NH), 10.26 (s, 1H, NH), 7.50 (d, J = 8.5 Hz, H-2/H-2'), 7.14 (br. s, 2H, H-8/H-8'), 7.12 (d, J = 8.5 Hz, 2H, H-3/H-3'), 7.08 (br. s, 2H, H-7/H-7'), 4.69 (s, 1H, NH), 3.90 (br. s, 2H, H-5), 3.32 (br. s, 2H, H-13), 3.09 (br. s, 2H, H-10), 1.51 (s, 9H, $C(CH_3)_3$), 1.50 – 1.46 (m, 4H, H-11 and H-12), 1.49 (s, 9H, $C(CH_3)_3$), 1.47 (s, 9H, $C(CH_3)_3$), 1.40 (s, 9H, $C(CH_3)_3$).

δ_C (101 MHz, $CDCl_3$): δ 164.2 (s, CN), 163.5 (s, CN), 158.7 (s, CO), 156.0 (s, CO), 153.5 (s, CO), 153.3 (s, CO), 140.0 (s, C-1), 136.9 (s, C-4), 135.0 (s, C-6), 134.2 (s, C-9), 130.2 (s, C-8), 129.3 (s, C-3), 125.7 (s, C-7), 122.4 (s, C-2), 83.7 (s, $C(CH_3)_3$), 79.6 (s, $C(CH_3)_3$), 79.0 (s, $C(CH_3)_3$), 78.4 (s, $C(CH_3)_3$), 40.9 (s, C-5), 40.5 (s, C-13), 39.9 (s, C-10), 28.4 (s, $C(CH_3)_3$), 28.1 (s, $C(CH_3)_3$), 28.1 (s, $C(CH_3)_3$), 27.2 (s, C-12), 26.8 (s, C-11).

HRMS-ESI (+): m/z calculated for $C_{39}H_{59}N_7O_8$: 754.4425 found: 754.4511[M+H]⁺

ν_{max} (ATR)/ cm^{-1} : 3297 (NH), 3157 (NH), 2929 (CH_2), 1667 (CN), 1602, 800, 776.

***tert*-Butyl -N-[(1E)-{4-[4-[(1E)-{[(*tert*-butoxy)carbonyl]amino}pentyl]amino){[(*tert*-butoxy)carbonyl]imino})methyl]amino}phenyl)methyl]phenyl]amino){[(*tert*-butoxy)carbonyl]imino})methyl]carbamate (**46b**)**



Following method E, from **42b** (174 mg, 0.324 mmol) and *t*-BuOH (5 mL), **46b** was obtained as pale brown amorphous gum. (154 mg, 62% yield); R_f 0.4 (7: 3 Hex: EtOAc)

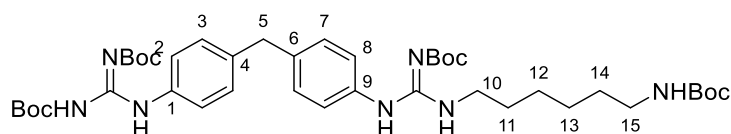
δ_{H} (600 MHz, CDCl_3): 11.65 (s, 1H, NH), 10.67 (s, 1H, NH), 10.30 (s, 1H, NH), 7.54 (d, $J = 8.1$ Hz, 2H, H-2/H-2'), 7.20 (s, 2H, H-8/H-8'), 7.16 (d, $J = 8.1$ Hz, 2H, H-3/H-3'), 7.11 (s, 2H, H-7/H-7'), 4.56 (s, 1H, NH), 3.94 (s, 2H, H-5), 3.33 (br. s, 2H, H-14), 3.09 (br. s, 2H, H-10), 1.55 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.53 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.51 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.47 (s, 4H, H-11 and H-13), 1.45 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.35 – 1.28 (m, 2H, H-12).

δ_{C} (151 MHz, CDCl_3): δ 164.4 (s, CN), 163.5 (s, CN), 158.6 (s, CO), 155.9 (s, CO), 153.5 (s, CO), 153.3 (s, CO), 140.0 (s, C-1), 136.9 (s, C-9), 135.0 (s, C-4), 134.1 (s, C-6), 130.2 (s, C-2), 129.3 (s, C-7), 125.7 (s, C-3), 122.4 (s, C-8), 83.7 (s, $\text{C}(\text{CH}_3)_3$), 79.6 (s, $\text{C}(\text{CH}_3)_3$), 79.0 (s, $\text{C}(\text{CH}_3)_3$), 40.8 (s, C-5), 40.3 (s, C-14), 29.7 (s, C-10), 29.6 (s, C-13), 29.6 (s, C-11), 28.4 (s, $\text{C}(\text{CH}_3)_3$), 28.2 (s, $\text{C}(\text{CH}_3)_3$), 28.0 (s, $\text{C}(\text{CH}_3)_3$), 23.8 (s, C-12).

HRMS-ESI (+): m/z calculated for $\text{C}_{40}\text{H}_{61}\text{N}_7\text{O}_8$: 768.4582 found: 768.4664 $[\text{M}+\text{H}]^+$

$\nu_{\text{max}}(\text{ATR})/\text{cm}^{-1}$: 3301 (NH), 2921 (CH_2), 2852 (CH_2), 1631, 804.

***tert*-Butyl -*N*-[(1*E*)-{4-[(4-[(1*E*)-{[(*tert*-butoxy)carbonyl]amino}[(6-[(*tert*-butoxy)carbonyl]amino}hexyl]imino)methyl]amino}phenyl)methyl]phenyl]amino){[(*tert*-butoxy)-carbonyl]imino)methyl]carbamate (**47b**)**



Following method E, from **43b** (129 mg, 0.172 mmol) and *t*-BuOH (5 mL), **47b** was obtained as pale brown amorphous gum. (89 mg, 66% yield); R_f 0.4 (7: 3 Hex: EtOAc)

δ_{H} (400 MHz, CDCl_3): 11.61 (s, 1H, NH), 10.25 (s, 1H, NH), 7.49 (d, $J = 8.2$ Hz, 2H, H-2/H-2'), 7.15 (d, $J = 8.1$ Hz, 2H, H-8/H-8'), 7.12 (d, $J = 8.2$ Hz, 2H, H-3/H-3'), 7.06 (br. s, 2H, H-7/H-7'), 4.51 (s, 2H, H-5), 3.28 (s, 2H, H-15), 3.05 (br. s, 2H, H-10), 1.57 – 1.52 (m, 4H, H-11 and H-14), 1.50 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.48 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.47 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.41 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.36 – 1.25 (m, 4H, H-12 and H-13).

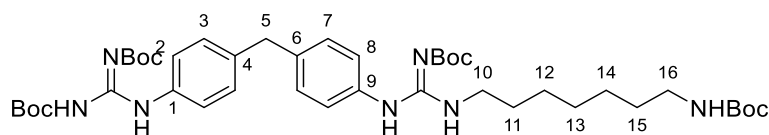
δ_{C} (101 MHz, CDCl_3): 164.3 (s, CN), 163.5 (s, CN), 158.6 (s, CO), 155.9 (s, CO), 153.5 (s, CO), 153.3 (s, CO), 139.7 (s, C-1), 136.9 (s, C-9), 135.0 (s, C-4), 134.3 (s, C-6), 130.2 (s, C-2), 129.3 (s, C-7), 125.6 (s, C-3), 122.4 (s, C-8), 83.7 (s, $\text{C}(\text{CH}_3)_3$), 79.6 (s, $\text{C}(\text{CH}_3)_3$), 79.0 (s, $\text{C}(\text{CH}_3)_3$), 78.4

(s, $\underline{\text{C}}(\text{CH}_3)_3$), 41.1 (s, C-5), 40.8 (s, C-15), 40.5 (s, C-10), 29.9 (s, C-11), 29.7 (s, C-14), 29.4 (s, C-12), 28.4 (s, $\text{C}(\underline{\text{CH}_3})_3$), 28.3 (s, $\text{C}(\underline{\text{CH}_3})_3$), 28.1 (s, $\text{C}(\underline{\text{CH}_3})_3$), 28.0 (s, $\text{C}(\underline{\text{CH}_3})_3$), 26.6 (s, C-13).

HRMS-ESI (+): m/z calculated for $\text{C}_{41}\text{H}_{63}\text{N}_7\text{O}_8$: 782.4738 found: 782.4911 $[\text{M}+\text{H}]^+$

$\nu_{\text{max}}(\text{ATR})/\text{cm}^{-1}$: 3320 (NH), 2972 (CH_2), 2933 (CH_2), 2855 (CH_2), 1631, 803, 768.

***tert*-Butyl -N-[(1E)-{4-[(4-{[(1Z)-{[(*tert*-butoxy)carbonyl]amino}{7-{[(*tert*-butoxy)carbonyl]amino}heptyl]imino]methyl]amino}phenyl)methyl]phenyl]amino){[(*tert*-butoxy)carbonyl]imino})methyl]carbamate (**48b**)**



Following method E, from **44b** (110 mg, 0.143 mmol) and *t*-BuOH (5 mL), **48b** was obtained as pale brown amorphous gum. (65 mg, 57% yield); R_f 0.4 (7: 3 Hex: EtOAc)

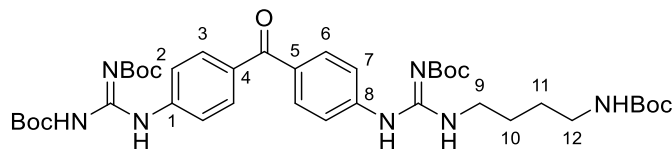
δ_{H} (400 MHz, CDCl_3): 11.65 (s, 1H, NH), 10.68 (s, 1H, NH), 10.30 (s, 1H, NH), 7.54 (d, $J = 8.4$ Hz, 2H, H-2), 7.19 (br. s, 2H, H-8 and H-8'), 7.16 (d, $J = 8.4$ Hz, 2H, H-3 and H-3'), 7.12 (br. s, 2H, H-7 and H-7'), 4.54 (s, 1H, NH), 3.94 (s, 2H, H-5), 3.34 (br. s, 2H, H-16), 3.09 (br. s, 2H, H-10), 1.55 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.53 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.51 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.50 – 1.46 (m, 4H, H-11 and H-15), 1.45 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.34 – 1.22 (m, 6H, H-12, H-13, H-14).

δ_{C} (101 MHz, CDCl_3): 164.3 (s, CN), 163.5 (s, CN), 158.6 (s), 155.9 (s), 153.5 (s, CO), 153.3 (s, CO), 139.8 (s, C-1), 136.9 (s, C-4), 135.0 (s, C-6), 134.3 (s, C-9), 130.2 (s, C-8), 129.3 (s, C-3), 125.6 (s, C-7), 122.4 (s, C-2), 83.7 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 79.6 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 78.9 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 78.4 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 41.1 (s, C-16), 40.8 (s, C-5), 40.5 (s, C-10), 29.9 (s, C-11), 29.3 (s, C-15), 28.9 (s, C-14), 28.8 (s, $\text{C}(\underline{\text{CH}_3})_3$), 28.4 (s, $\text{C}(\underline{\text{CH}_3})_3$), 28.1 (s, $\text{C}(\underline{\text{CH}_3})_3$), 28.0 (s, $\text{C}(\underline{\text{CH}_3})_3$), 26.6 (s, C-12), 26.6 (s, C-13).

HRMS-ESI (+): m/z calculated for $\text{C}_{42}\text{H}_{65}\text{N}_7\text{O}_8$: 796.4895 found: 796.4755 $[\text{M}+\text{H}]^+$

$\nu_{\text{max}}(\text{ATR})/\text{cm}^{-1}$: 3325 (NH), 2977 (CH_2), 2931 (CH_2), 2861 (CH_2), 1680 (CN), 1600, 803, 768.

***tert*-Butyl -N-{4-[(E)-{[(*tert*-butoxy)carbonyl]amino}{[4-(4-{[(1E)-{[(*tert*-butoxy)carbonyl] amino}{[(*tert*-butoxy)carbonyl] imino})methyl]amino}benzoyl) phenyl]amino} methylidene)amino]butyl}carbamate (**45c**)**



Following method E, from **41c** (22.6 mg, 0.031 mmol) and *t*-BuOH (5 mL), **45c** was obtained as pale brown amorphous gum. (19 mg, 80% yield); R_f 0.5 (7: 3 Hex: EtOAc)

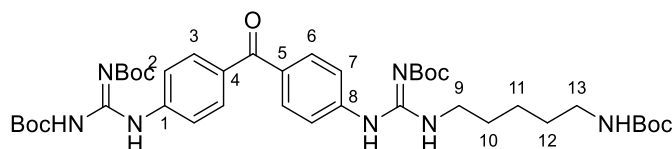
δ_H (400 MHz, $CDCl_3$): 11.62 (s, 1H, NH), 10.58 (s, 1H, NH), 7.85 – 7.71 (m, 8H, H-2/H-2', H-6/H-6' and H-7/H-7'), 7.04 (br. s, 2H, H-3/H-3'), 4.72 (s, 1H, NH), 3.34 (br. s, 2H, H-12), 3.16 (br. s, 2H, H-9), 1.64 (br. s, 4H, H-10 and H-11), 1.51 (s, 18H, 2x $C(CH_3)_3$), 1.45 (s, 9H, $C(CH_3)_3$), 1.41 (s, 9H, $C(CH_3)_3$).

δ_C (101 MHz, $CDCl_3$): 194.5 (s, CO), 165.2 (s, CN), 163.2 (s, CN), 156.0 (s, CO), 153.6 (s, CO), 141.0 (s, C-1), 133.5 (C-8), 131.9 (s, C-4), 131.1 (s, C-5), 129.7 (s, C-2), 123.6 (s, C-7), 121.1 (s, C-3), 120.9 (s, C-6), 82.7 (s, $C(CH_3)_3$), 79.3 (s, $C(CH_3)_3$), 33.7 (s, C-12), 29.6 (s, C-9), 28.3 (s, $C(CH_3)_3$), 28.3 (s, $C(CH_3)_3$), 28.0 (s, $C(CH_3)_3$), 28.0 (s, $C(CH_3)_3$), 24.8 (s, C-10), 24.3 (s, C-11).

HRMS-MALDI (+): m/z calculated for $C_{39}H_{57}N_7O_9$: 768.4218; found 768.4296

ν_{max} (ATR)/ cm^{-1} : 3321 (NH), 3001 (CH_2), 1672 (CN), 1559, 815, 788.

***tert*-Butyl N-[(1E)-{[4-(4-{[(1E)-[(5-{[(*tert*-butoxy)carbonyl]amino}pentyl)amino]{[(*tert*-butoxy)carbonyl]imino})methyl]amino}benzoyl)phenyl]amino}{[(*tert*-butoxy)carbonyl] imino})methyl]carbamate (**46c**)**



Following method E, from **42c** (87 mg, 0.115 mmol) and *t*-BuOH (5 mL), **46c** was obtained as pale brown amorphous gum (55 mg, 61% yield); R_f 0.5 (7: 3 Hex: EtOAc)

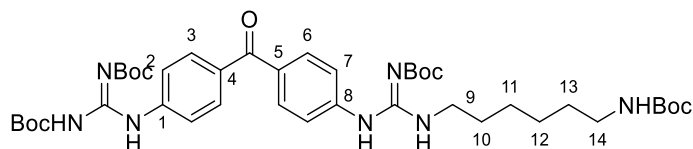
δ_{H} (400 MHz, CDCl_3): 11.66 (s, 1H, NH), 10.65 (s, 1H, NH), 8.12 (s, 1H, NH), 7.85 (d, $J = 8.3$ Hz, 2H, H-2 and H-2'), 7.82 (br. s, 4H, H-6/H-6' and H-7/H-7'), 7.21 (br. s, 2H, H-3 and H-3'), 4.62 (s, 1H, NH), 3.38 – 2.99 (m, 4H, H-9 and H-13), 1.57 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.55 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.53 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.52 – 1.47 (m, 4H, H-10 and H-12), 1.45 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.42 – 1.34 (m, 2H, H-11).

δ_{C} (151 MHz, CDCl_3): 194.4 (s, CO), 165.2 (s, CN), 163.2 (s, CN), 156.0 (s, CO), 153.3 (s, CO), 141.0 (s, C-1), 133.7 (s, C-8), 133.0 (s, C-4), 131.8 (s, C-5), 131.2 (s, C-2), 129.7 (s, C-7), 123.6 (s, C-3), 121.1 (s, C-6), 84.2 (s, $\text{C}(\text{CH}_3)_3$), 80.1 (s, $\text{C}(\text{CH}_3)_3$), 79.1 (s, $\text{C}(\text{CH}_3)_3$), 40.1 (s, C-13), 31.9 (s, C-9), 29.7 (s, C-12), 28.7 (s, C-11), 28.4 (s, $\text{C}(\text{CH}_3)_3$), 28.1 (s, $\text{C}(\text{CH}_3)_3$), 28.0 (s, $\text{C}(\text{CH}_3)_3$), 23.8 (s, C-11).

HRMS-ESI (+): m/z calculated for $\text{C}_{40}\text{H}_{59}\text{N}_7\text{O}_9$: 782.4374 found: 782.4449 $[\text{M}+\text{H}]^+$

$\nu_{\text{max}}(\text{ATR})/\text{cm}^{-1}$: 3310 (NH), 2997 (CH_2), 1621, 1578, 810, 790.

***tert*-Butyl -*N*-[(1*E*)-{[4-(4-{[(1*E*)-{[(*tert*-butoxy)carbonyl]amino}[(6-{[(*tert*-butoxy)carbonyl]amino}hexyl]imino)methyl]amino}benzoyl)phenyl]amino}{[(*tert*-butoxy)carbonyl]imino}]methyl]carbamate (**47c**)**



Following method E, from **43c** (396 mg, 0.517 mmol) and *t*-BuOH (5 mL), **47c** was obtained as pale brown amorphous gum. (210 mg, 51% yield); R_f 0.4 (7: 3 Hex: EtOAc)

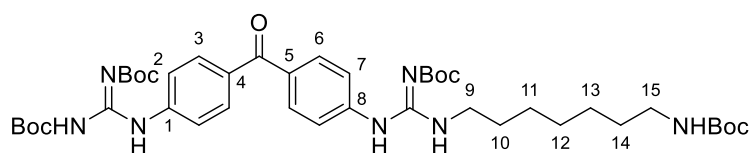
δ_{H} (400 MHz, CDCl_3): 11.63 (s, 1H, NH), 10.59 (s, 1H, NH), 7.84 – 7.73 (m, 6H, H-2/H-2', H-6/H-6' and H-7/H-7'), 6.99 (br. s, 2H, H-3/H-3'), 4.52 (s, 1H, NH), 3.35 (br. s, 2H, H-14), 3.10 (br. s, H-9), 1.75 – 1.54 (m, 4H, H-10 and H-13), 1.54 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.51 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.45 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.42 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.41 – 1.28 (m, 4H, H-11 and H-12).

δ_{C} (101 MHz, CDCl_3): 194.9 (s, CO), 163.3 (s, CN), 156.0 (s, CO), 153.3 (s, CO), 140.6 (s, C-1), 133.6 (s, C-8), 132.1 (s, C-4), 131.1 (s, C-5), 122.9 (s, C-2 and C-3), 121.0 (s, C-6 and C-7), 84.1 (s, $\text{C}(\text{CH}_3)_3$), 80.0 (s, $\text{C}(\text{CH}_3)_3$), 79.0 (s, $2 \times \text{C}(\text{CH}_3)_3$), 41.0 (s, C-14), 40.4 (s, C-9), 30.0 (s, C-10), 29.1 (s, C-13), 28.4 (s, $\text{C}(\text{CH}_3)_3$), 28.1 (s, $\text{C}(\text{CH}_3)_3$), 28.0 (s, $2 \times \text{C}(\text{CH}_3)_3$), 26.6 (s, C-11), 26.4 (s, C-12).

HRMS-ESI (+): m/z calculated for C₄₁H₆₁N₇O₉: 796.4531 found: 796.4619 [M+H]⁺

v_{max}(ATR)/cm⁻¹: 3322 (NH), 2995 (CH₂), 1641 (CN), 1600, 806.

tert-Butyl -N-{7-[(E)-{[(tert-butoxy)carbonyl]amino}{[4-(4-{[(1E)-{[(tert-butoxy)carbonyl]amino}{[(tert-butoxy)carbonyl]imino})methyl]amino}benzoyl]phenyl]amino})methylidene)amino]heptyl}carbamate (48c)



Following method E, from **44c** (120 mg, 0.154 mmol) and *t*-BuOH (5 mL), **48c** was obtained as pale brown amorphous gum. (61 mg, 49% yield); *R_f* 0.4 (7: 3 Hex: EtOAc)

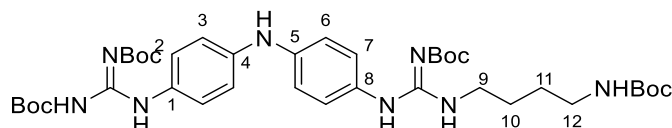
δ_H (400 MHz, CDCl₃): 11.66 (s, 1H, NH), 10.62 (s, 1H, NH), 7.86 – 7.76 (m, 6H, H-2/H-2', H-6/H-6 and H-7/H-7'), 7.04 (br. s, 2H, H-3/H-3'), 4.54 (s, 1H, NH), 3.37 (t, *J* = 6.6 Hz, 2H, H-15), 3.12 (d, *J* = 5.9 Hz, 2H, H-9), 1.68 – 1.58 (m, 4H, H-10 and H-14), 1.56 (s, 9H, C(CH₃)₃), 1.55 (s, 9H, C(CH₃)₃), 1.52 – 1.44 (m, 4H, H-11 and H-13), 1.49 (s, 9H, C(CH₃)₃), 1.46 (s, 9H, C(CH₃)₃), 1.41 – 1.33 (m, 4H, H-12).

δ_C (101 MHz, CDCl₃): 194.9 (s, CO), 163.3 (s, CN), 156.0 (s, CO), 153.3 (s, CO), 140.6 (s, C-1), 133.5 (s, C-8), 132.1 (s, C-4), 131.1 (s, C-5), 122.9 (s, C-2 and C-3), 121.0 (s, C-6 and C-7), 84.1 (s, C(CH₃)₃), 80.0 (s, C(CH₃)₃), 79.0 (s, 2x C(CH₃)₃), 41.1 (s, C-15), 40.5 (s, C-9), 30.0 (s, C-14), 29.7 (s, C-10), 29.1 (s, C-13), 28.9 (s, C(CH₃)₃), 28.4 (s, C(CH₃)₃), 28.2 (s, C(CH₃)₃), 28.1 (s, C(CH₃)₃), 26.9 (s, C-11), 26.6 (s, C-12).

HRMS-ESI (+): m/z calculated for C₄₂H₆₃N₇O₉: 810.4687: found: 810.5122 [M+H]⁺

v_{max}(ATR)/cm⁻¹: 3301 (NH), 2921 (CH₂), 1631, 1591, 804.

***tert*-Butyl-*N*-[(1*E*)-{(4-[(4-[(1*E*)-[(4-[(*tert*-butoxy)carbonyl]amino)butyl]amino){[(*tert*-butoxy)carbonyl]imino)}methyl]amino}phenyl)amino]phenyl)amino){[(*tert*-butoxy)carbonyl]imino)}methyl]carbamate (**45d**)**



Following method E, from **41d** (82 mg, 0.112 mmol) and *t*-BuOH (5 mL), **45d** was obtained as purple gum (50 mg, 59% yield); R_f 0.4 (7: 3 Hex: EtOAc)

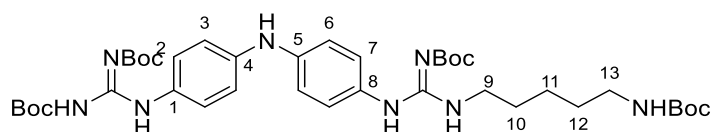
δ_H (400 MHz, $CDCl_3$): 11.67 (s, 1H, NH), 10.50 (s, 1H, NH), 10.27 (s, 1H, NH), 7.53 (d, $J = 8.2$ Hz, 2H, H-2 and H-2'), 7.12 – 7.05 (m, 4H, H-6/H-6' and H-7/H-7'), 7.03 (d, $J = 8.2$ Hz, 2H, H-3 and H-3'), 5.82 (s, 1H, NH), 4.71 (s, 1H, NH), 3.37 – 3.23 (m, 2H, H-12), 3.14 (br. s, 2H, H-9), 1.56 (s, 9H, $C(CH_3)_3$), 1.55 (s, 9H, $C(CH_3)_3$), 1.52 (s, 9H, $C(CH_3)_3$), 1.45 (s, 9H, $C(CH_3)_3$), 1.47 – 1.44 (m, 4H, H-10 and H-11).

δ_C (101 MHz, $CDCl_3$): 164.2 (s, CN), 163.6 (s, CN), 159.7 (s, CO), 156.5 (s, CO), 154.1 (s, CO), 153.9 (s, CO), 139.4 (s, C-1), 135.0 (s, C-8), 134.1 (s, C-4), 133.1 (s, C-5), 128.1 (s, C-7), 124.2 (s, C-2), 119.9 (s, C-6), 118.0 (s, C-3), 84.2 (s, $C(CH_3)_3$), 80.0 (s, $C(CH_3)_3$), 79.5 (s, $C(CH_3)_3$), 78.8 (s, $C(CH_3)_3$), 41.1 (s, C-12), 40.5 (s, C-9), 29.0 (s, $C(CH_3)_3$), 28.9 (s, $C(CH_3)_3$), 28.7 (s, $C(CH_3)_3$), 28.6 (s, $C(CH_3)_3$), 27.7 (s, C-11), 27.4 (s, C-10).

HRMS-ESI (+): m/z calculated for $C_{38}H_{58}N_8O_8$: 754.4378 found: 755.2811 $[M+H]^+$

$\nu_{max}(ATR)/cm^{-1}$: 3295 (NH), 2992 (CH_2), 2915 (CH_2), 1641 (CN), 1055, 1031, 811, 709.

***tert*-Butyl-*N*-[(1*E*)-{(4-[(4-[(1*E*)-[(5-[(*tert*-butoxy)carbonyl]amino)pentyl]amino){[(*tert*-butoxy)carbonyl]imino)}methyl]amino}phenyl)amino]phenyl)amino){[(*tert*-butoxy)carbonyl]imino)}methyl]carbamate (**46d**)**



Following method E, from **42d** (313 mg, 0.422 mmol) and *t*-BuOH (5 mL), **46d** was obtained as purple gum (156 mg, 48% yield); R_f 0.4 (7: 3 Hex: EtOAc)

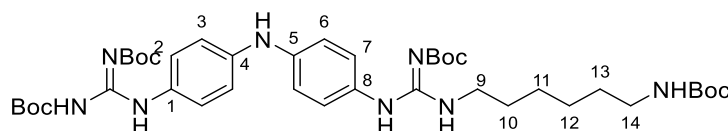
δ_{H} (400 MHz, CDCl_3): 11.64 (s, 1H, NH), 10.49 (s, 1H, NH), 10.27 (s, 1H, NH), 7.76 (s, 1H, NH), 7.53 (d, $J = 8.3$ Hz, 2H, H-2 and H-2'), 7.18 – 6.93 (m, 6H, H-3/H-3', H-6/H-6' and H-7/H-7'), 5.86 (s, 1H, NH), 4.55 (s, 1H, NH), 3.36 – 3.24 (m, 2H, H-13), 3.17 – 3.05 (m, 2H, H-9), 1.56 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.55 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.52 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.52 – 1.44 (m, 6H, H-10, H-11 and H-12), 1.45 (s, 9H, $\text{C}(\text{CH}_3)_3$).

δ_{C} (101 MHz, CDCl_3): 164.4 (s, CN), 163.6 (s, CN), 159.4 (s, CO), 155.9 (s, CO), 153.6 (s, CO), 153.3 (s, CO), 139.0 (s, C-1 and C-8), 130.9 (s, C-4 and C-5), 127.7 (s, C-7), 123.7 (s, C-2), 119.5 (s, C-6), 117.4 (s, C-3), 83.6 (s, $\text{C}(\text{CH}_3)_3$), 79.5 (s, $\text{C}(\text{CH}_3)_3$), 79.0 (s, $\text{C}(\text{CH}_3)_3$), 78.36 (s, $\text{C}(\text{CH}_3)_3$), 40.9 (s, C-13), 39.6 (s, C-9), 29.7 (s, C-12), 29.3 (s, C-10), 29.2 (s, C-11), 28.5 (s, $\text{C}(\text{CH}_3)_3$), 28.4 (s, $\text{C}(\text{CH}_3)_3$), 28.2 (s, $\text{C}(\text{CH}_3)_3$), 28.1 (s, $\text{C}(\text{CH}_3)_3$).

HRMS-ESI (+): m/z calculated for $\text{C}_{39}\text{H}_{60}\text{N}_8\text{O}_8$: 769.4534 found: 769.4600 $[\text{M}+\text{H}]^+$

$\nu_{\text{max}}(\text{ATR})/\text{cm}^{-1}$: 3332 (NH), 2983 (CH_2), 2925 (CH_2), 1600, 1509, 802, 776.

***tert*-butyl -*N*-[(1*E*)-{4-[(4-[(1*E*)-[(6-[(*tert*-butoxy)carbonyl]amino]hexyl)amino]{{{*tert*-butoxy)carbonyl}imino}}methyl]amino}phenyl]amino]phenyl]amino){[(*tert*-butoxy)carbonyl]imino}}methyl]carbamate (**47d**)**



Following method E, from **43d** (243 mg, 0.328 mmol) and *t*-BuOH (5 mL), **47d** was obtained as purple gum (154 mg, 60% yield); R_f 0.3 (7: 3 Hex: EtOAc)

δ_{H} (400 MHz, CDCl_3): 11.65 (s, 1H, NH), 10.44 (s, 1H, NH), 10.25 (s, 1H, NH), 7.54 – 7.35 (m, 2H, H-2/H-2'), 7.10 – 6.98 (m, 6H, H-3/H-3', H-6/H-6' and H-7/H-7'), 5.83 (s, 1H, NH), 4.52 (s, 1H, NH), 3.34 (br. s, 2H, H-14), 3.06 – 2.98 (m, 2H, H-9), 1.57 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.54 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.50 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.49 – 1.44 (m, 4H, H-10 and H-13), 1.43 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.30 – 1.25 (m, 4H, H-11 and H-12).

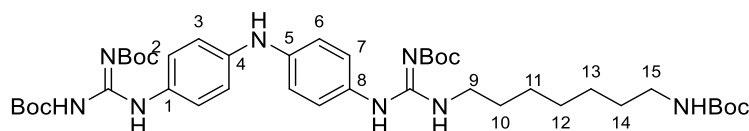
δ_{C} (151 MHz, CDCl_3): 164.9 (s, CN), 163.6 (s, CN), 153.6 (s, CO), 153.5 (s, CO), 153.3 (s, CO), 143.0 (s, C-1), 138.9 (s, C-8), 131.1 (s, C-4 and C-5), 127.7 (s, C-2), 123.5 (s, C-7), 119.5 (s, C-6), 117.6 (s, C-3), 83.6 (s, $\text{C}(\text{CH}_3)_3$), 79.5 (s, $\text{C}(\text{CH}_3)_3$), 79.0 (s, 2x $\text{C}(\text{CH}_3)_3$), 40.7 (s, C-14), 40.3 (s, C-

9), 29.9 (s, C-10), 29.4 (s, C-13), 28.4 (s, C(CH₃)₃), 28.2 (s, C(CH₃)₃), 28.1 (s, C(CH₃)₃), 28.0 (s, C(CH₃)₃), 26.3 (s, C-11), 21.0 (s, C-12).

HRMS-ESI (+): m/z calculated for C₄₀H₆₂N₈O₈: 783.4691 found: 783.4556 [M+H]⁺

v_{max}(ATR)/cm⁻¹: 3332 (NH), 2933 (CH₂), 2925 (CH₂), 2863 (CH₂), 1600, 1509, 802.

***tert*-Butyl -N-[(1E)-{4-[(4-{[(1E)-[7-[(*tert* butoxy)carbonyl]amino}heptyl)amino]({[(*tert*-butoxy)carbonyl]imino)}methyl)amino]phenyl)amino]phenyl)amino]({[(*tert*-butoxy)carbonyl]imino)}methyl]carbamate (**48d**)**



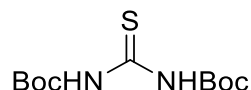
Following method E, from **44d** (230 mg, 0.299 mmol) and *t*-BuOH (5 mL), **48d** was obtained as purple gum (105 mg, 44% yield); R_f 0.3 (7: 3 Hex: EtOAc)

δ_H (400 MHz, CDCl₃): 11.68 (s, 1H, NH), 10.51 (s, 1H, NH), 10.28 (s, 1H, NH), 7.54 (d, *J* = 8.7 Hz, 2H, H-2 and H-2'), 7.09 (d, *J* = 7.7 Hz, 4H, H-6/H-6' and H-7/H-7'), 7.04 (d, *J* = 8.7 Hz, 2H, H-3 and H-3'), 5.87 (s, 1H, NH), 4.54 (s, 1H, NH), 3.36 – 3.23 (m, 2H, H-15), 3.10 – 2.99 (m, 2H, H-9), 1.56 (s, 18H, 2(C(CH₃)₃)), 1.53 (s, 9H, C(CH₃)₃), 1.51 – 1.43 (m, 4H, H-10 and H-14), 1.46 (s, 9H, C(CH₃)₃), 1.33 – 1.26 (m, 6H, C(CH₃)₃).

δ_C (151 MHz, CDCl₃): 164.5 (s, CN), 163.6 (s, CN), 159.3 (s, CO), 155.9 (s, CO), 153.5 (s, CO), 153.3 (s, CO), 142.4 (s, C-1), 139.2 (s, C-8), 130.9 (s, C-4 and C-5), 127.7 (s, C-7), 123.6 (s, C-2), 119.5 (s, C-6), 117.4 (s, C-3), 83.6 (s, C(CH₃)₃), 79.4 (s, C(CH₃)₃), 79.0 (s, C(CH₃)₃), 78.2 (s, C(CH₃)₃), 41.0 (s, C-15), 40.5 (s, C-9), 29.9 (s, C-14), 29.7 (s, C-10), 29.3 (s, C-13), 28.8 (s, C-11), 28.5 (s, C(CH₃)₃), 28.4 (s, C(CH₃)₃), 28.2 (s, C(CH₃)₃), 28.1 (s, C(CH₃)₃), 26.6 (s, C-12).

HRMS-ESI (+): m/z calculated for C₄₁H₆₄N₈O₈: 797.4847 found: 797.4917 [M+H]⁺

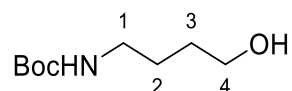
v_{max}(ATR)/cm⁻¹: 3300 (NH), 2977 (CH₂), 2961 (CH₂), 1600, 1566, 1031, 1011, 812, 771.

***N,N'*-bis-(*tert*-butoxycarbonyl)thiourea (49)²**

A solution of thiourea (1000 mg, 13 mmol) in dry THF (200 mL) was prepared and cooled to 0 °C under N₂. To this, NaH (1600 mg, 57 mmol, 60% in mineral oil) was added. After 5 min, the reaction was left to stir at room temperature for a further 30 min. The mixture was then cooled to 0 °C and di-*tert*-butyldicarbonate (1000 mg, 46 mmol) was added. After 30 min, the reaction was left to stir overnight. The reaction was then quenched by the careful addition of saturated NaHCO₃ (aq) (15 mL) and H₂O (60 mL). The aqueous layer was extracted using CH₂Cl₂ and the combined organic phases were washed with brine, dried over MgSO₄ anhydrous and concentrated under reduced pressure. The concentrated crude product was then recrystallised from 8:2 hexane: CH₂Cl₂ to give a fine white powder (592 mg, 16%); *R_f* 0.3 (4:1 Hex: EtOAc); mp 120-125 °C (lit. 127-129 °C)²

δ_{H} (400 MHz, CDCl₃): 1.52 (s, 18H, 2x C(CH₃)₃).

HRMS-ESI (+) *m/z* calculated for C₁₁H₂₀N₂O₄S: 276.1144; found: 299.1034 [M+Na]⁺

***tert*-Butyl (4-hydroxybutyl) carbamate (51)**

To a stirring solution of 4-aminobutan-1-ol (1.03 mL, 11.2 mmol) in anhydrous CH₂Cl₂ was added anhydrous triethylamine (1.87 mL, 13.5 mmol). To this stirring solution was added Boc₂O (2446 mg, 11.21 mmol) in anhydrous CH₂Cl₂ via addition funnel. The solution was then stirred for 24 h. Upon completion the reaction mixture was washed three times with 10 mL of H₂O. The organic phase was dried and concentrated by rotary evaporation at 90 °C for 10 min affording **51** as a pale yellow oil (1800 mg, 85% yield); *R_f* 0.7 (7:2.5:0.5 CHCl₃: MeOH: NEt₃)

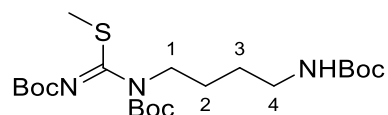
δ_{H} (400 MHz, *d*₆-DMSO): 6.67 (t, *J* = 5.3 Hz, 1H, NH), 4.49 (t, *J* = 5.1 Hz, 1H, OH), 3.41 – 3.27 (m, 2H, H-4), 2.91 – 2.78 (m, 2H, H-1), 1.35 – 1.32 (m, 4H, H-2 and H-3), 1.31 (s, 9H, C(CH₃)₃).

δ_{C} (101 MHz, *d*₆-DMSO): δ 156.0 (s, CO), 77.7 (s, C(CH₃)₃), 60.8 (s, C-4), 40.0 (s, C-1), 30.0 (s, C-3), 28.5 (s, C(CH₃)₃), 26.4 (s, C-2).

HRMS-ESI (+): m/z calculated for $C_9H_{19}NO_3$: 190.1365 found: 190.1443 $[M+H]^+$

$\nu_{\max}(\text{ATR})/\text{cm}^{-1}$: 3301 (NH), 2950 (CH_2), 1696 (CO), 1409, 1332, 726.

***tert*-Butyl -*N*-(4-{{{*tert*-butoxy}carbonyl}amino}butyl)-*N*-[(1*Z*)-{{{*tert*-butoxy}carbonyl}imino}{methylsulfanyl)methyl]carbamate (52)**



1,3-bis-(*tert*-butoxycarbonyl)-2-methyl-2-thiopseudourea (200 mg, 0.6887 mmol), **51** (130 mg, 0.6887 mmol) and PPh_3 (271 mg, 1.033 mmol) were dissolved in THF (10 mL) and DIAD (209 mg, 0.2 mL, 1.033 mmol) was added. The reaction mixture was stirred at room temperature for 4 h. The solvent was evaporated and the crude product was purified by silica gel column chromatography (8:2 EtOAc: Hex) to give **52** as a colourless oil. (132 mg, 42 % yield) R_f 0.6 (1:1 Hex: EtOAc)

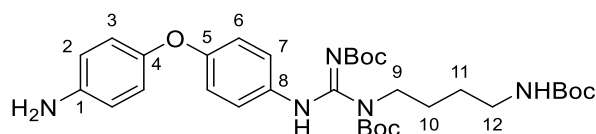
δ_H (400 MHz, CDCl_3): 4.69 (s, 1H, NH), 3.51 – 3.41 (m, 2H, H-4), 3.07 – 2.99 (m, 2H, H-1), 2.32 (s, 3H, S- CH_3), 1.66 – 1.54 (m, 2H, H-3), 1.44 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.41 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.40 – 1.39 (m, 2H, H-2), 1.37 (s, 9H, $\text{C}(\text{CH}_3)_3$).

δ_C (101 MHz, CDCl_3): 162.7 (s, CN), 157.7 (s, CO), 155.9 (s, CO), 151.7 (s, CO), 82.2 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 81.7 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 78.8 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 48.3 (s, C-4), 40.0 (s, C-1), 28.3 (s, $\text{C}(\underline{\text{C}}\text{H}_3)_3$), 28.0 (s, $\text{C}(\underline{\text{C}}\text{H}_3)_3$), 27.9 (s, $\text{C}(\underline{\text{C}}\text{H}_3)_3$), 27.0 (s, C-3), 26.1 (s, C-2), 15.4 (s, S- $\underline{\text{C}}\text{H}_3$).

HRMS-ESI (+): m/z calculated for $C_{21}H_{39}N_3O_6S$: 462.2560 found: 462.2633 $[M+H]^+$

$\nu_{\max}(\text{ATR})/\text{cm}^{-1}$: 2977 (CH_2), 2160, 1711 (CO), 1616 (CN), 1514, 970, 853.

***tert*-Butyl -*N*-(4-aminobutyl)-*N*-[(1*Z*)-{[4-(4-aminophenoxy)phenyl]amino}{{{*tert*-butoxy}carbonyl}imino})methyl]carbamate (53)**



Following method B, from 4,4-oxydianiline (50 mg, 0.25 mmol), **52** (173 mg, 0.375 mmol) and HgCl_2 (101 mg, 0.375 mmol), NEt_3 (0.1 mL, 0.75 mmol), **53** was obtained as a cloudy white oil (83 mg, 64% yield); R_f 0.5 (1:1 Hex: EtOAc)

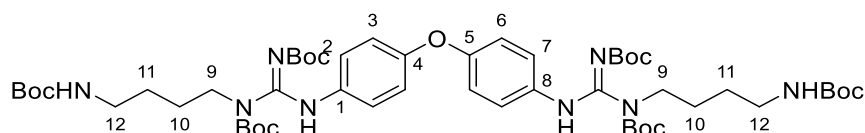
δ_H (400 MHz, CDCl_3): 11.07 (s, 1H, NH), 7.00 (br. s, 2H, H-7/H-7'), 6.86 (d, $J = 8.3$ Hz, 2H, H-2/H-2'), 6.80 (d, $J = 8.3$ Hz, 2H, H-3/H-3'), 6.65 (d, $J = 8.7$ Hz, 2H, H-6/H-6'), 4.79 (s, 1H, NH), 3.82 – 3.43 (m, 2H, H-12), 3.13 (br. s, 2H, H-9), 1.72 (br. s, 4H, H-10 and H-11), 1.46 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.42 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.24 (s, 9H, $\text{C}(\text{CH}_3)_3$).

δ_C (101 MHz, CDCl_3): δ 156.0 (s, CO), 152.2 (s, CO), 148.4 (s, C-8), 142.8 (s, C-1), 123.5 (s, C-4), 121.9 (s, C-5), 120.9 (s, C-7), 120.6 (s, C-2), 117.8 (s, C-6), 116.2 (s, C-3), 82.4 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 81.8 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 78.9 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 47.3 (s, C-12), 39.8 (s, C-9), 28.4 (s, $\text{C}(\underline{\text{C}}\text{H}_3)_3$), 28.1 (s, $\text{C}(\underline{\text{C}}\text{H}_3)_3$), 27.7 (s, $\text{C}(\underline{\text{C}}\text{H}_3)_3$), 27.1 (s, C-11), 25.8 (s, C-10).

HRMS-ESI (+): m/z calculated for $\text{C}_{27}\text{H}_{39}\text{N}_5\text{O}_5$: 514.2951 found: 514.3025 $[\text{M}+\text{H}]^+$

$\nu_{\text{max}}(\text{ATR})/\text{cm}^{-1}$: 3232 (NH), 2985 (CH_2), 1700 (CO), 1616 (CN), 1551, 825, 735.

***tert*-Butyl -N-(4-{[(1E)-{[4-(4-{[(1Z)-[(4-aminobutyl)][(*tert*-butoxy)carbonyl] amino} {[(*tert*butoxy)carbonyl]imino})methyl]amino}phenoxy)phenyl]amino}{[(*tert*-butoxy)carbonyl]imino})methyl}[(*tert*-butoxy)carbonyl]amino)butyl)carbamate (**54**)**



Following method B with the intention of synthesising **53**, **54** was obtained as a minor product (46 mg, 20% yield); R_f 0.8 (1:1 Hex: EtOAc)

δ_H (400 MHz, CDCl_3): 11.09 (s, 1H, NH), 7.12 – 6.82 (m, 8H, H-2/H-2', H-3/H-3', H-6/H-6' and H-7/H-7'), 4.75 (s, 2H, NH), 3.83 – 3.44 (m, 4H, H-12), 3.14 (s, 4H, H-9), 1.72 (s, 4H, H-11), 1.47 (s, 18H, $\text{C}(\text{CH}_3)_3$), 1.42 (s, 18H, $\text{C}(\text{CH}_3)_3$), 1.26 (s, 18H, $\text{C}(\text{CH}_3)_3$).

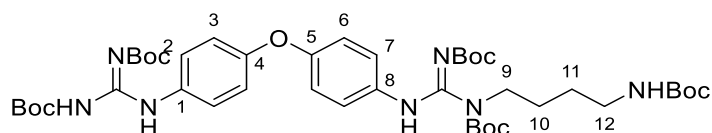
δ_C (101 MHz, CDCl_3): 155.9 (s, CO), 148.4 (s, C-8), 142.8 (s, C-1), 135.5 (s, C-4 and C-5), 123.5 (s, C-2 and C-3), 119.2 (s, C-6 and C-7), 81.6 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 79.4 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 78.8 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 78.6

(s, $\underline{\text{C}}(\text{CH}_3)_3$), 39.9 (s, C-12), 39.1 (s, C-9), 28.3 (s, 4x $\underline{\text{C}}(\text{CH}_3)_3$), 28.0 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 27.7 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 27.1 (s, C-11), 25.8 (s, C-10).

HRMS-ESI (+): m/z calculated for $\text{C}_{47}\text{H}_{74}\text{N}_8\text{O}_{11}$: 927.5477 found: 927.5547 $[\text{M}+\text{H}]^+$

$\nu_{\text{max}}(\text{ATR})/\text{cm}^{-1}$: 3211 (NH), 2974 (CH_2), 1710 (CO), 1623 (CN), 1498, 801, 755.

***tert*-Butyl -N-[(1E)-{[4-(4-{[(1Z)-[(4-aminobutyl)[(*tert*-butoxy)carbonyl]amino}{[(*tert*-butoxy)carbonyl]imino)}methyl]amino}phenoxy)phenyl]amino}{[(*tert*-butoxy)carbonyl]imino)}methyl]carbamate (**55**)**



Following method B, from **53** (66 mg, 0.128 mmol), **50** (56 mg, 0.192 mmol), HgCl_2 (52 mg, 0.192 mmol) and NEt_3 (0.05 mL, 0.32 mmol), **55** was obtained as a pale yellow gum (34 mg, 35% yield); R_f 0.7 (1:1 Hex: EtOAc)

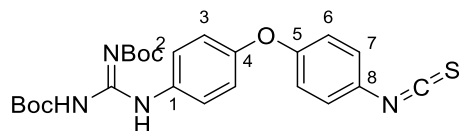
δ_{H} (400 MHz, CDCl_3): 11.65 (s, 1H, NH), 11.11 (s, 1H, NH), 10.29 (s, 1H, NH), 7.56 (d, $J = 7.7$ Hz, 2H, H-2/H-2'), 7.04 – 6.85 (m, 6H, H-3/H-3', H-6/H-6' and H-7/H-7'), 4.81 (s, 1H, NH), 3.81 – 3.62 (m, 2H, H-12), 3.18 (s, 2H, H-9), 1.78 (s, 2H, H-10), 1.64 (s, 2H, H-11), 1.56 (s, 9H, $\underline{\text{C}}(\text{CH}_3)_3$), 1.52 (s, 9H, $\underline{\text{C}}(\text{CH}_3)_3$), 1.52 (s, 9H, $\underline{\text{C}}(\text{CH}_3)_3$), 1.50 (s, 9H, $\underline{\text{C}}(\text{CH}_3)_3$), 1.45 (s, 9H, $\underline{\text{C}}(\text{CH}_3)_3$).

δ_{C} (101 MHz, CDCl_3): 163.4 (s, CN), 156.0 (s, CO), 154.6 (s, CO), 153.7 (s, CO), 153.3 (s, CO), 152.5 (s, CO), 143.2 (s, C-8), 142.4 (s, C-1), 133.0 (s, C-4), 132.3 (s, C-5), 124.0 (s, C-2 and C-3), 119.4 (s, C-7), 119.2 (s, C-6), 83.2 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 82.5 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 81.9 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 79.8 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 78.9 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 47.3 (s, C-12), 39.8 (s, C-9), 28.4 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 28.1 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 28.1 (s, 2x $\underline{\text{C}}(\text{CH}_3)_3$), 28.0 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 27.1 (s, C-11), 25.8 (s, C-10).

HRMS-ESI (+): m/z calculated for $\text{C}_{38}\text{H}_{57}\text{N}_7\text{O}_9$: 756.4218 found: 756.4510 $[\text{M}+\text{H}]^+$

$\nu_{\text{max}}(\text{ATR})/\text{cm}^{-1}$: 3211 (NH), 2895 (CH_2), 1720 (CO), 1600 (CN), 1499, 1210, 876, 721.

***tert*-Butyl-*N*-[(1*E*)-{[(*tert*-butoxy)carbonyl]imino}{[4-(4-isothiocyanatophenoxy)phenyl]amino}]methyl]carbamate (56)**



To a solution of **38a** (500 mg, 1.13 mmol) in dry CH₂Cl₂ (10 mL) was added 1,1'-Thiocarbonylbisimidazole (289 mg, 1.6 mmol) at 0 °C. The reaction mixture was then left to stir at room temperature for 5 hours. Upon completion, the reaction mixture was reduced and purified by silica gel column chromatography to produce a white solid (482 mg, 88%); *R_f* 0.8 (1:1 Hex: EtOAc); mp 130-135 °C

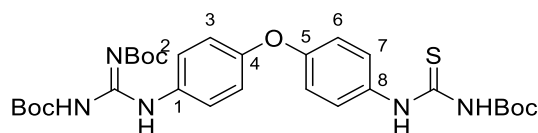
δ_{H} (400 MHz, CDCl₃): 11.64 (s, 1H, NH), 10.32 (s, 1H, NH), 7.66 – 7.54 (m, 2H, H-7/H-7'), 7.20 – 7.15 (m, 2H, H-2/H-2'), 7.00 – 6.96 (m, 2H, H-6/H-6'), 6.96 – 6.92 (m, 2H, H-3/H-3'), 1.54 (s, 9H, C(CH₃)), 1.50 (s, 9H, C(CH₃)).

δ_{C} (151 MHz, CDCl₃): 163.6 (s, CS), 156.7 (s, CO), 153.6 (s, CO), 152.8 (s, CN), 150.0 (s, C-1), 146.8 (s, C-8), 133.0 (s, C-4), 132.4 (s, C-5), 127.9 (s, C-2), 123.8 (s, C-7), 119.9 (s, C-3), 119.0 (s, C-6), 79.7 (s, C(CH₃)₃), 79.6 (s, C(CH₃)₃), 28.1 (s, C(CH₃)₃), 28.0 (s, C(CH₃)₃).

HRMS-ESI (+): *m/z* calculated for C₂₄H₂₈N₄O₅S: 485.1780; found: 485.1844 [M+H]⁺

ν_{max} (ATR)/cm⁻¹: 2100 (N=C=S), 1718 (CO), 1661, 1492 (CN), 1303, 1228, 1158, 1115.

***tert*-Butyl-*N*-[(1*E*)-[4-{4-[[[(*tert*-butoxy)carbonyl]amino]methanethioyl]amino}phenoxy]phenyl]amino]{[(*tert*-butoxy)carbonyl]imino}]methyl]carbamate (57)**



To a cooled solution of *tert*-butyl carbamate (1 equiv.) in dry THF was added NaH (60% immersion oil, 4.5 equiv) and allowed to stir for 20 mins. **38a** was then added and allowed to stir at room temperature for 5 h. The reaction was quenched by the addition of H₂O (5 mL) and stirred for 15 min. The mixture was then reduced, partitioned between H₂O and CH₂Cl₂ and dried under anhydrous MgSO₄. The crude product was purified by silica gel column

chromatography (8:2 Hex: EtOAc) to yield a pale yellow solid (48mg, 38% yield) R_f 0.35; mp decomposes over 200 °C

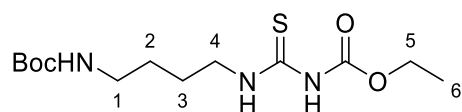
δ_H (400 MHz, $CDCl_3$): 11.64 (s, 1H, NH), 10.27 (s, 1H, NH), 9.83 (s, 1H, NH), 7.53 (d, J = 8.9 Hz, 2H, H-7/H-7'), 7.46 (d, J = 8.9 Hz, 2H, H-6/H-6'), 7.19 (s, 1H, NH), 6.95 – 6.87 (m, 4H, H-2/H-2', and H-3/H-3'), 1.52 (s, 18H, 2(C(CH₃)₃), 1.49 (s, 9H, C(CH₃)₃).

δ_C (151 MHz, $CDCl_3$): 178.0 (s, CS), 163.5 (s, CN), 154.3 (s, CO), 153.5 (s, CO), 153.2 (s, CO), 150.7 (s, CO), 132.6 (s, C-1 and C-8), 131.9 (s, C-4 and C-5), 123.8 (s, C-7), 121.7 (s, C-6), 119.3 (s, C-2), 118.9 (s, C-3), 83.7 (s, C(CH₃)₃), 83.4 (s, C(CH₃)₃), 79.6 (s, C(CH₃)₃), 28.1 (s, 2(C(CH₃)₃), 28.0 (s, C(CH₃)₃).

HRMS-ESI (+): m/z calculated for C₂₉H₃₉N₅O₇S: 601.2570; found: 624.2471 [M+Na]⁺

ν_{max} (ATR)/cm⁻¹: 1636, 1500, 1408, 1368, 1149 (CS), 1219, 806.

Ethyl *N*-[(4-[(*tert*-butoxy)carbonyl]amino}butyl)carbamothioyl]carbamate (**58a**)



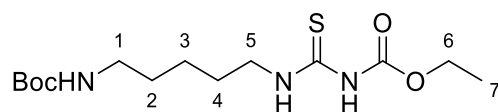
Following method C, **58a** was obtained as a yellow solid (230 mg, 77% yield); R_f 0.7 (1:1 Hex: EtOAc); mp 85-90 °C

δ_H (400 MHz, $CDCl_3$): 9.72 (s, 1H, NH), 8.07 (s, 1H, NH), 4.61 (s, 1H, NH), 4.24 (q, J = 7.1 Hz, 2H, H-5), 3.69 – 3.51 (m, 2H, H-4), 3.18 (br. s, 2H, H-1), 1.76 – 1.67 (m, 2H, H-2), 1.62 – 1.54 (m, 2H, H-3), 1.46 (s, 9H, C(CH₃)₃), 1.33 (t, J = 7.1 Hz, 3H, H-6).

δ_C (151 MHz, $CDCl_3$): 179.0 (s, CS), 155.9 (s, CO), 152.7 (s, CO), 79.2 (s, C(CH₃)₃), 62.7 (s, C-5), 45.1 (s, C-4), 40.1 (s, C-1), 28.4 (s, C-2), 27.4 (s, C-3), 25.6 (s, C(CH₃)₃), 14.1 (s, C-6).

HRMS-ESI (+): m/z calculated for C₁₃H₂₅N₃O₄S: 319.1566 found: 342.1459 [M+Na]⁺

ν_{max} (ATR)/cm⁻¹: 1714 (CO), 1679 (CO), 1546, 1141 (CS), 1042, 1012, 772, 738.

Ethyl *N*-[(5-[[*tert*-butoxy)carbonyl]amino}pentyl)carbamothioyl]carbamate (58b)

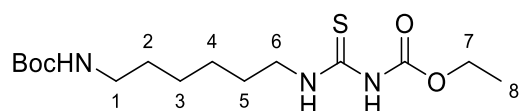
Following method C, **58b** was obtained as a yellow solid (150 mg, 91% yield); R_f 0.7 (1:1 Hex: EtOAc); mp 70-75 °C

δ_H (400 MHz, $CDCl_3$): 9.70 (s, 1H, NH), 8.12 (s, 1H, NH), 4.57 (s, 1H, NH), 4.23 (q, $J = 7.1$ Hz, 2H, H-6), 3.66 – 3.52 (m, 2H, H-1), 3.14 – 3.00 (m, 2H, H-5), 1.75 – 1.65 (m, 2H, H-4), 1.54 – 1.49 (m, 2H, H-2), 1.45 (s, 9H, $C(CH_3)_3$), 1.44 – 1.36 (m, 2H, H-3), 1.32 (t, $J = 7.1$ Hz, 3H, H-7).

δ_C (151 MHz, $CDCl_3$): 179.1 (s, CS), 155.8 (s, CO), 152.7 (s, CO), 79.1 (s, $C(CH_3)_3$), 62.7 (s, C-6), 45.4 (s, C-1), 40.3 (s, C-5), 29.7 (s, C-2), 28.4 (s, $C(CH_3)_3$), 27.9 (s, C-4), 24.0 (s, C-3), 14.1 (s, C-7).

HRMS-ESI (+): m/z calculated for $C_{14}H_{27}N_3O_4S$: 334.1722 found: 334.1609 $[M+H]^+$

ν_{max} (ATR)/ cm^{-1} : 3308 (NH), 3239 (NH), 2933 (CH_2), 1709 (CO), 1682, 1148 (CS), 1123, 609.

Ethyl *N*-[(6-[[*tert*-butoxy)carbonyl]amino}hexyl)carbamothioyl]carbamate (58c)

Following method C, **58c** was obtained as a yellow solid (105 mg, 79% yield); R_f 0.7 (1:1 Hex: EtOAc)

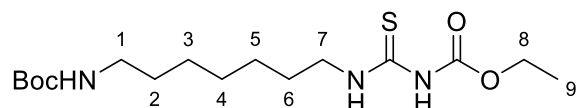
δ_H (400 MHz, $CDCl_3$): 9.69 (s, 1H, NH), 8.10 (s, 1H, NH), 4.23 (q, $J = 7.1$ Hz, 2H, H-7), 3.70 – 3.59 (m, 2H, H-1), 3.12 – 2.98 (m, 2H, H-6), 1.72 – 1.63 (m, 2H, H-2), 1.51 – 1.47 (m, 2H, H-5), 1.45 (s, 9H, $C(CH_3)_3$), 1.41 – 1.36 (m, 4H, H-3 and H-4), 1.32 (t, $J = 7.1$ Hz, 8H, H-8).

δ_C (151 MHz, $CDCl_3$): 179.0 (s, CS), 155.9 (s, CO), 152.7 (s, CO), 79.0 (s, $C(CH_3)_3$), 62.6 (s, C-7), 45.5 (s, C-1), 40.4 (s, C-6), 29.9 (s, C-5), 28.4 (s, $C(CH_3)_3$), 28.1 (s, C-2), 26.5 (s, C-3), 26.3 (s, C-4), 14.1 (s, C-8).

HRMS-ESI (+): m/z calculated for $C_{15}H_{29}N_3O_4S$: 348.1879 found: 348.1922 $[M+H]^+$

$\nu_{\max}(\text{ATR})/\text{cm}^{-1}$: 3311 (NH), 3241 (NH), 2988 (CH₂), 1714 (CO), 1148 (CS), 611.

Ethyl *N*-[(7-[[*tert*-butoxy)carbonyl]amino}heptyl]carbamoithioyl]carbamate (58d)



Following method C, **38d** was obtained as a yellow oil (313 mg, 95% yield); R_f 0.8 (1:1 Hex: EtOAc); mp 70-75 °C

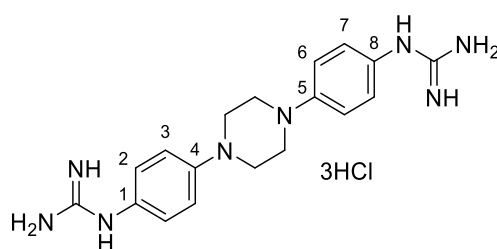
δ_H (400 MHz, CDCl₃): 9.69 (s, 1H, NH), 8.05 (s, 1H, NH), 4.53 (s, 1H, NH), 4.24 (q, J = 7.1 Hz, 2H, H-8), 3.65 – 3.51 (m, 2H, H-7), 3.11 (t, J = 6.5 Hz, 2H, H-1), 1.71 – 1.62 (m, 2H, H-2), 1.53 – 1.46 (m, 2H, H-6), 1.46 (s, 9H), 1.41 – 1.34 (m, 6H, H-3, H-4, H-5), 1.32 (t, J = 7.1 Hz, 3H, H-9).

δ_C (151 MHz, CDCl₃): 178.9 (s, CS), 152.7 (s, CO), 62.6 (s, C-8), 45.6 (s, C-7), 40.5 (s, C-1), 29.9 (s, C-2), 28.8 (s, C-4), 28.4 (s, C(CH₃)₃), 28.1 (s, C-6), 26.7 (s, C-3), 26.6 (s, C-5), 14.1 (s, C-9).

HRMS-ESI (+): m/z calculated for C₁₆H₃₁N₃O₄S: 361.2035 found: 384.1927 [M+Na]⁺

$\nu_{\max}(\text{ATR})/\text{cm}^{-1}$: 2930 (CH₂), 2857 (CH₂), 1708 (CO), 1513, (CS), 733.

***N*-{4-[4-(4-Carbamimidamidophenyl)piperazin-1-yl]phenyl}guanidine (105)²**

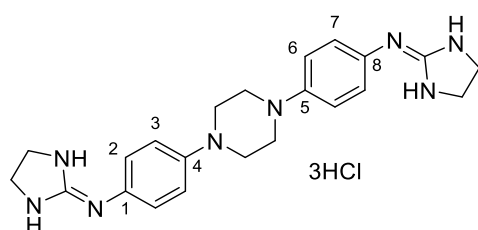


Compound **78** (47 mg, 0.0619 mmol) was stirred in a 50:50 mixture of CH₂Cl₂: TFA (2 mL) for 24 h. Upon completion, TFA was removed and the crude product was dissolved in H₂O. IRA400 Amberlyte resin was added and the mixture was stirred for 24 h. The resin was then filtered off and the aqueous layer was washed with CH₂Cl₂ (3 x 10 mL) and reduced to yield the corresponding HCl salt as a pale brown powder (20 mg, 70% yield); mp 120–128 °C (no lit. mp reported)²; HPLC = 95.5% pure (R_T = 21.88 min)

δ_{H} (400 MHz, D_2O): 7.19–7.15 (m, 8H, H-2/H-2', H-3/H-3', H-6/H-6' and H-7/H-7'), 3.36 (s, 8H, $\text{CH}_2(\text{PZ})$).

HRMS-ESI (+) m/z calculated for $\text{C}_{18}\text{H}_{24}\text{N}_8$: 353.2124; found: 353.2186 $[\text{M}+\text{H}]^+$

***N*-[4-(4-{4-[(Imidazolidin-2-ylidene)amino]phenyl}piperazin-1-yl)phenyl]imidazolidin-2-imine (106)²**

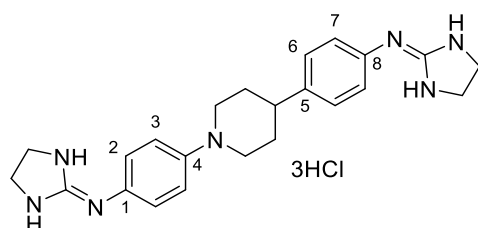


Compound **79** (60 mg, 0.075 mmol) was stirred in a 50:50 mixture of CH_2Cl_2 : TFA (2 mL) for 24 h. Upon completion, TFA was removed and the crude product was dissolved in H_2O . IRA400 Amberlyte resin was added and the mixture was stirred for 24 h. The resin was then filtered off and the aqueous solution was washed with CH_2Cl_2 (3 x 10 mL) and reduced to yield the corresponding HCl salt as a pale brown powder (25 mg, 65% yield); mp 110–115 °C (no lit. mp reported)²; HPLC = 99.8% pure (R_{T} = 22.3 min)

δ_{H} (400 MHz, D_2O): 7.30 – 7.00 (br. s, 8H, H-2/H-2', H-3/H-3', H-6/H-6' and H-7/H-7'), 3.63 (s, 8H, $2 \times \text{CH}_2(\text{Imi})$), 3.59 (s, 8H, $\text{CH}_2(\text{PZ})$).

HRMS-ESI (+) m/z calculated for $\text{C}_{22}\text{H}_{28}\text{N}_8$: 405.2437; found: 405.2491 $[\text{M}+\text{H}]^+$

***N*-[4-(1-{4-[(Imidazolidin-2-ylidene)amino]phenyl}piperidin-4-yl)phenyl]imidazolidin-2-imine (108)²**

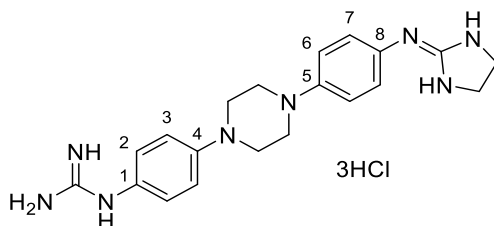


Compound **80** (91 mg, 0.113 mmol) was stirred in a 50:50 mixture of CH₂Cl₂: TFA (2 mL) for 24 h. Upon completion, TFA was removed and the crude product was dissolved in H₂O. IRA400 Amberlyte resin was added and the mixture was stirred for 24 h. The resin was then filtered off and the aqueous solution was washed with CH₂Cl₂ (3 x 10 mL) and reduced to yield the corresponding HCl salt as a pale brown powder (28 mg, 48% yield); mp 100–105 °C (no lit. mp reported)²; HPLC = 99% pure (R_T = 22 min)

δ_{H} (400 MHz, *d*-DMSO): 10.40 (s, 1H, NH), 10.12 (s, 1H, NH), 8.28 (s, 1H, NH), 8.12 (s, 1H, NH), 7.34 (d, *J* = 8.5 Hz, 2H, H-2 and H-2'), 7.19 (d, *J* = 8.2 Hz, 2H, H-7 and H-7'), 7.12 – 7.02 (m, 4H, H-3/H-3' and H-6/H-6'), 3.83 – 3.72 (m, 4H, CH₂ (PD)), 3.62 (s, 4H, CH₂ (Imi)), 3.25 (s, 4H, CH₂ (Imi)), 3.08 – 2.97 (m, 1H, CH (PD)), 2.90 – 2.61 (m, 4H, CH₂ (PD)).

HRMS-ESI (+) *m/z* calculated for C₂₃H₂₉N₇: 404.2484: found: 404.2559 [M+H]⁺

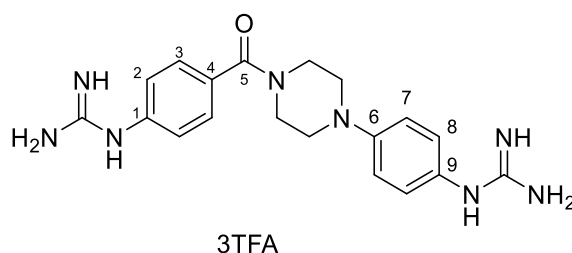
***N*-[4-(4-{4-[(Imidazolidin-2-ylidene)amino]phenyl}piperazin-1-yl)phenyl]guanidine (109)¹**



Compound **82** (48 mg, 0.061 mmol) was stirred in a 50:50 mixture of CH₂Cl₂: TFA (2 mL) for 24 h. Upon completion, TFA was removed and the crude product was dissolved in H₂O. IRA400 Amberlyte resin was added and the mixture was stirred for 24 h. The resin was then filtered off and the aqueous solution was washed with CH₂Cl₂ (3 x 10 mL) and reduced to yield the corresponding HCl salt as a pale brown powder (15 mg, 50% yield); mp 170–175 °C (lit. 168–170 °C)¹; HPLC = 95% pure (R_T = 21.19 min);

δ_{H} (400 MHz, D₂O): 7.37 (d, *J* = 8.6 Hz, 2H, H-2 and H-2'), 7.33 (d, *J* = 9.0 Hz, 2H, H-7 and H-7'), 7.27 – 7.21 (m, 4H, H-3/H-3' and H-6/H-6'), 3.67 (br. s, 8H, CH₂ (PZ)), 3.62 (br. s, 4H, CH₂ (Imi)).

HRMS-ESI (+) *m/z* calculated for C₂₀H₂₆N₈: 379.2280: found: 379.2356 [M+H]⁺

TFA salt of 1-(4-(4-(4-guanidinophenyl)piperazine-1-carbonyl)phenyl)guanidine (66a)

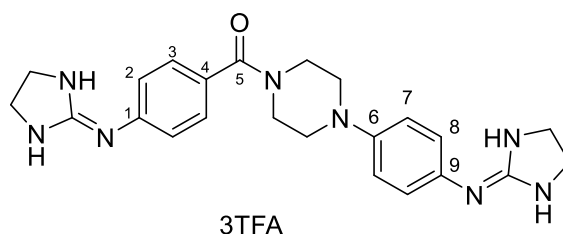
Following method I, from **92a** (278 mg, 0.356 mmol), **66a** was obtained as light brown crystalline solid. (220 mg, 86% yield); mp 51 - 54 °C; HPLC = 95% pure (R_T = 20.05 min)

δ_H (400 MHz, MeOD): 7.56 (d, J = 8.5 Hz, 2H, H-2 and H-2'), 7.39 (d, J = 8.5 Hz, 2H, H-3 and H-3'), 7.18 (d, J = 9.0 Hz, 2H, H-8 and H-8'), 7.08 (d, J = 9.0 Hz, 2H, H-7 and H-7'), 3.92 (br. s, 2H, CH₂), 3.65 (br. s, 2H, CH₂), 3.32-3.22 (m, 4H, 2xCH₂).

δ_C (101 MHz, MeOD): 170.5 (s, CO), 157.4 (s, CN), 157.0 (s, CN), 151.0 (s, C-1), 137.1 (s, C-4), 134.2 (s, C-8), 128.9 (s, C-2), 127.2 (s, C-3), 126.6 (s, C-7), 125.2 (s, C-8), 29.6 (s, CH₂), 23.2 (s, CH₂)

HMRS-ESI (+): m/z calculated for C₁₉H₂₄N₈O: 381.2073 Found: 381.2150 [M+H]⁺

ν_{max} (ATR)/ cm⁻¹: 3131 (NH), 2849 (CH₂), 1666 (CO), 1603, 834, 799.

TFA salt of (4-(imidazolidin-2-ylideneamino)phenyl)(4-(4-(imidazolidin-2-ylideneamino)phenyl)piperazin-1-yl)methanone (66b)

Following method I, from **92b** (119 mg, 0.143 mmol), **66b** was obtained as a dark brown glassy solid. (22 mg, 20% yield); mp 137 - 141 °C; HPLC = 97% pure (R_T = 21.45 min)

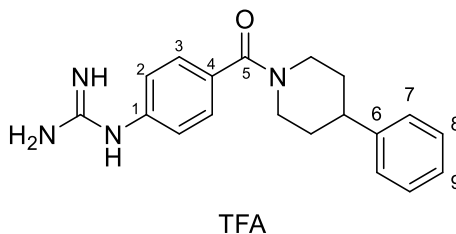
δ_{H} (400 MHz, MeOD): 7.56 (d, $J = 8.5$ Hz, 2H, H-2 and H-2'), 7.39 (d, $J = 8.5$ Hz, 2H, H-3 and H-3'), 7.18 (d, $J = 9.0$ Hz, 2H, H-8 and H-8'), 7.07 (d, $J = 9.0$ Hz, 2H, H-7 and H-7'), 3.92 (br. s, 2H, CH₂), 3.81 (s, 4H, 2xCH₂), 3.74 (s, 4H, 2xCH₂), 3.64 (br. s, 2H, CH₂), 3.33 - 3.17 (m, 4H, 2xCH₂).

δ_{C} (101 MHz, MeOD): 170.4 (s, CO), 159.7 (s, CN), 159.0 (s, CN), 150.8 (s, C-1), 137.7 (s, C-4), 134.0 (s, C-8), 129.0 (s, C-2), 127.6 (s, C-3), 125.9 (s, C-7), 123.8 (s, C-6), 47.9 (s, CH₂), 43.1 (s, CH₂), 43.1 (s, CH₂).

HMRS-ESI (+): m/z calculated for C₂₃H₂₈N₈O: 433.2386 Found: 433.2459 [M+H]⁺

ν_{max} (ATR)/ cm⁻¹: 3182 (NH), 2916 (CH₂), 1674 (CO), 1603, 832, 800.

TFA salt of *N*-[4-(4-phenylpiperidine-1-carbonyl)phenyl]guanidine (68a)



Following method I, from **91a** (113 mg, 0.216 mmol), **68a** was obtained as a white powder. (83 mg, 88% yield); mp 215-219 °C; HPLC = 99.6% pure ($R_{\text{T}} = 29.43$ min)

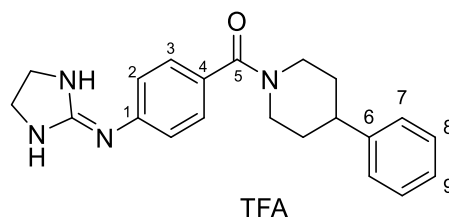
δ_{H} (400 MHz, MeOD): 7.55 (d, $J = 8.5$ Hz, 2H, H-2 and H-2'), 7.42 (d, $J = 8.5$ Hz, 2H, H-3 and H-3'), 7.33 - 7.29 (m, 4H, H-7/H-7' and H-8/H-8'), 7.31 - 7.19 (m, 1H, H-9), 4.77 (br. s, 2H, CH₂), 3.88 (br. s, 2H, CH₂), 3.27 - 2.83 (m, 4H, 2xCH₂), 1.88 (br. s, 1H, CH).

δ_{C} (151 MHz, MeOD): 168.5 (s, CO), 157.1 (s, CN), 145.7 (s, C-1), 136.5 (s, C-4), 135.1 (s, C-6), 128.5 (s, C-2), 128.4 (s, C-7), 126.7 (s, C-8), 126.2 (s, C-9), 124.8 (s, C-3), 42.5 (s, CH₂), 33.3 (s, CH).

HMRS-ESI (+): m/z calculated for C₁₉H₂₂N₄O: 323.1794 Found: 323.1866 [M+H]⁺

ν_{max} (ATR)/ cm⁻¹: 3009 (NH), 2871 (CH₂) 1698 (CO), 1625(NH), 1581 (CN), 1571, 761.

TFA salt of 4-(imidazolidin-2-ylideneamino)phenyl(4-phenylpiperidin-1-yl)methanone (68b)



Following method I, from **91b** (93 mg, 0.168 mmol), **68b** was obtained as a dark green solid (71 mg, 90% yield); mp 219-223 °C; HPLC = 98.7% pure (R_T = 29.87 min)

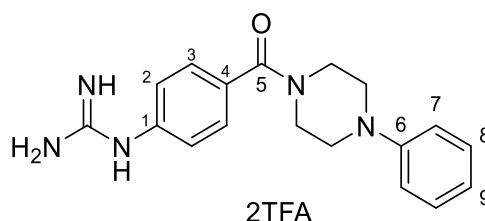
δ_H (400 MHz, MeOD): 7.57 (d, J = 8.5 Hz, 2H, H-2 and H-2'), 7.40 (d, J = 8.5 Hz, 2H, H-3 and H-3'), 7.33 - 7.23 - 7.19 (m, 4H, H-7/H-7' and H-8/H-8'), 7.21 (m, 1H, H-9), 4.83 - 4.79 (m, 2H, CH₂) 3.89 - 3.87 (m, 2H, CH₂), 3.82 (s, 4H, 2xCH₂), 3.29 (s, 4H, 2xCH₂), 2.98 (m, 1H, CH).

δ_C (151 MHz, MeOD): 170.0 (s, CO), 158.6 (s, CN), 145.1 (s, C-1), 136.9 (s, C-4), 134.4 (s, C-6), 128.3 (s, C-2), 128.1 (s, C-7), 126.4 (s, C-8), 126.1 (s, C-9), 123.6 (s, C-3), 65.5 (s, CH₂), 42.7 (s, CH₂), 42.3 (s, CH₂), 36.4 (s, CH₂), 33.1 (s, CH).

HMRS-ESI (+): m/z calculated for C₂₁H₂₄N₄O: 349.1950 Found: 349.2023 [M + H]⁺

ν_{max} (ATR)/ cm⁻¹: 2948 (CH₂), 1684 (CO), 1601, 830, 797.

TFA salt of 1-(4-(4-phenylpiperazine-1-carbonyl)phenyl)guanidine (69a)



Following method I, from **93a** (330 mg, 0.630 mmol), **69a** was obtained as a light pink powder. (274 mg, 79% yield); mp >230 °C; HPLC = 99.7% pure (R_T = 28.00 min)

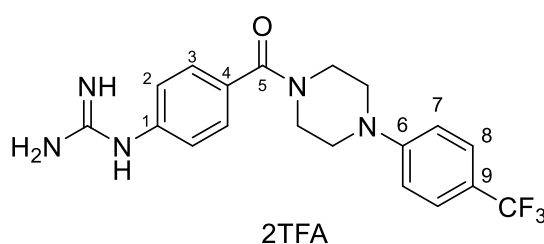
δ_H (400 MHz, MeOD): 7.59 - 7.50 (m, 2H, H-2 and H-2'), 7.42 - 7.40 (m, 2H, H-3 and H-3'), 7.29 - 7.27 (m, 2H, H-7 and H-7'), 7.06 - 7.03 (m, 2H, H-8 and H-8'), 6.94 - 6.91 (m, 1H, H-9), 3.95 (s, 2H, CH₂), 3.67 (s, 2H, CH₂), 3.35 - 3.14 (m, 4H, 2xCH₂).

δ_c (101 MHz, MeOD): 168.5 (s, CO), 155.0 (s, CN), 149.2 (s, C-1), 135.1 (s, C-4), 132.4 (s, C-2), 127.2 (s, C-3), 127.0 (s, C-7), 123.3 (s, C-8), 119.0 (s, C-9), 48.3 (s, CH₂), 40.4 (s, CH₂).

HMRS-ESI (+): m/z calculated for C₁₈H₂₁N₅O: 324.1746 Found: 324.1820 [M+H]⁺

$\nu_{\max}$ (ATR)/ cm⁻¹: 3100 (NH), 1678 (CO), 1624, 799.

TFA salt of 1-(4-(4-(4-(trifluoromethyl)phenyl)piperazine-1-carbonyl)phenyl)guanidine (70a)



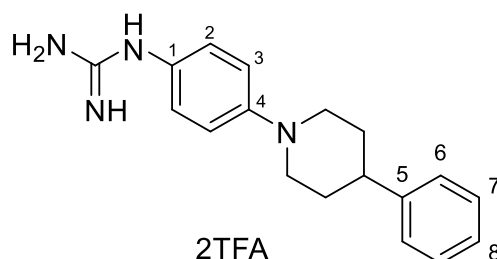
Following method I, from **94a** (52 mg, 0.088 mmol), **70a** was obtained as a white powder. (37 mg, 69% yield); mp >230 °C; HPLC = 99.8% pure (R_T = 31.07 min)

δ_H (400 MHz, MeOD): 7.56 (d, *J* = 8.5 Hz, 2H, H-2 and H-2'), 7.51 (d, *J* = 8.7 Hz, 2H, H-8 and H-8'), 7.39 (d, *J* = 8.7 Hz, 2H, H-7 and H-7'), 7.08 (d, *J* = 8.5 Hz, 2H, H-3 and H-3'), 3.92 (br. s, 2H, CH₂), 3.65 (br. s, 2H, CH₂), 3.45 - 3.28 (m, 4H, 2xCH₂).

δ_c (101 MHz, MeOD): 162.1 (s, CO), 152.7 (s, CN), 148.5 (s, C-1), 145.1 (s, C-9), 128.6 (s, C-4), 125.7 (s, C-6), 120.3 (s, C-2 and C-3), 117.9 (s, CF₃), 116.7 (s, C-8), 106.8 (s, C-7)

HMRS-ESI (+): m/z calculated for C₁₉H₂₀F₃N₅O: 392.1620 Found: 392.1693 [M+H]⁺

$\nu_{\max}$ (ATR)/ cm⁻¹: 3111 (NH), 1698 (CO), 1615, 822, 800.

TFA salt of *N*-[4-(4-phenylpiperidin-1-yl)phenyl]guanidine (71a)

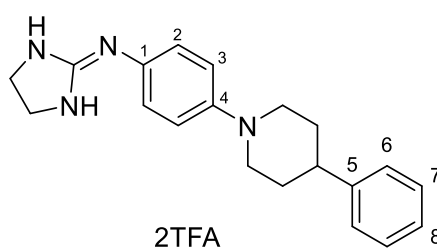
Following method I, from **101a** (189 mg, 0.382 mmol), **71a** was obtained as a light pink powder. (173 mg, 87% yield); mp 172 – 176 °C; HPLC = 99.9% pure (R_T = 30.13 min)

δ_H (400 MHz, MeOD): 7.44 (d, J = 8.9 Hz, 2H, H-2 and H-2'), 7.35 - 7.26 (m, 6H, H-3/H-3', H-7/H-7' and H-8'/H-8), 7.21 (m, 1H, H-9), 3.39 - 3.30 (m, 4H, 2xCH₂), 2.87 (q, J = 7.7 Hz, 1H, CH), 2.10 - 2.02 (m, 4H, 2xCH₂).

δ_C (101 MHz, MeOD): 151.1 (s, CN), 144.6 (s, C-1), 128.2 (s, C-2), 126.5 (s, C-3), 126.3 (s, C-5), 126.2 (s, C-6 and C-7), 119.7 (s, C-8), 53.1 (s, CH₂), 40.5 (s, CH₂), 31.8 (s, CH₂).

HRMS-ESI (+): m/z calculated for C₁₈H₂₂N₄: 295.1844: Found: 295.1917 [M+H]⁺

$\nu_{\max}$ (ATR)/ cm⁻¹: 3136 (NH), 3085 (NH), 1666, 797, 767, 720.

TFA salt of *N*-[4-(4-phenylpiperidin-1-yl)phenyl]imidazolidin-2-imine (71b)

Following method I, from **101b** (38 mg, 0.074 mmol), **71b** was obtained as a dark-brown solid. (34 mg, 84% yield); mp 145-149 °C; HPLC = 96% pure (R_T = 30.75 min)

δ_H (400 MHz, MeOD): 7.69 (d, J = 8.9 Hz, 2H, H-2 and H-2'), 7.42 (d, J = 8.9 Hz, 2H, H-3 and H-3'), 7.36 – 7.27 (m, 4H, H-6/H-6' and H-7/H-7'), 7.23 - 7.19 (m, 1H, H-8), 3.82 – 3.76 (m, 6H,

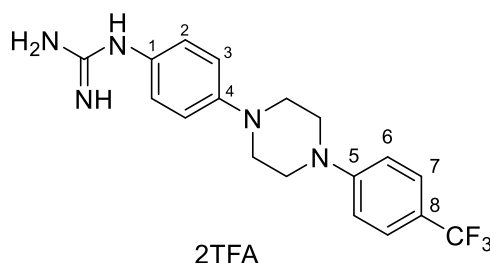
4xCH₂ (Imi), CH₂), 3.60-3.57 (m, 2H, CH₂), 3.01 - 2.95 (m, 1H, CH), 2.30 – 2.18 (m, 2H, CH₂), 2.15 – 2.08 (m, 2H, CH₂).

δ_c (151 MHz, MeOD): 158.6 (s, CN), 146, (s, C-1), 144 (s, C-4), 142.7 (s, C-5), 128.3 (s, C-2), 126.4 (s, C-7), 126.3 (s, C-3), 125.1 (s, C-6), 55.1 (s, CH₂), 42.7 (s, CH₂), 39.6 (s, CH₂), 31.1 (s, CH).

HRMS-ESI (+): m/z calculated for C₂₀H₂₄N₄: 321.2001: Found: 321.2074 [M+H]⁺

$\nu_{\max}$ (ATR)/ cm⁻¹: 3203 (NH), 2943 (CH₂), 1673 (CO), 1607, 1093, 818, 799.

TFA salt of *N*-(4-{4-[4-(trifluoromethyl)phenyl]piperazin-1-yl}phenyl)guanidine (72a)



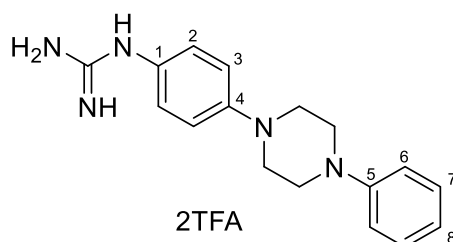
Following method I, from **103a** (100 mg, 0.177 mmol), **72a** was obtained as a white crystalline solid. (73 mg, 70% yield); mp 150 - 154 °C; HPLC = 99% pure (R_T = 31.07 min)

δ_H (400 MHz, MeOD): 7.49 (d, *J* = 8.6 Hz, 2H, H-2 and H-2'), 7.17 (d, *J* = 9.1 Hz, 2H, H-7 and H-7'), 7.12 - 7.05 (m, 4H, H-3/H-3' and H-6/H-6'), 3.50 - 3.41 (m, 4H, 2xCH₂), 3.41 - 3.33 (m, 4H, 2xCH₂).

δ_c (101 MHz, MeOD): 151.2 (s, CN), 143.8 (s, C-1), 140.6 (s, C-8), 129.3 (s, C-4), 128.9 (s, C-2), 126.0 (s, C-5), 120.0 (s, CF₃), 118.9 (s, C-7), 116.3 (s, C-3), 116.2 (s, C-6), 51.3 (s, CH₂), 49.4 (s, CH₂).

HRMS-ESI (+): m/z calculated for C₁₈H₂₀F₃N₅: 364.1671: Found: 364.1747 [M+H]⁺

$\nu_{\max}$ (ATR)/ cm⁻¹: 3135 (NH), 1668, 1616, 830.

TFA salt of *N*-[4-(4-phenylpiperazin-1-yl)phenyl]guanidine (73a**)**

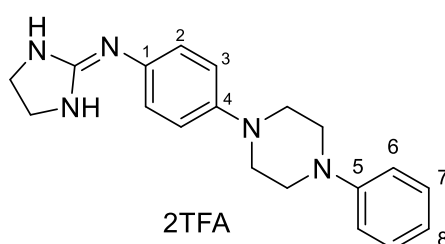
Following method I, from **102a** (129 mg, 0.259 mmol), **73a** was obtained as a pink powder. (118 mg, 87% yield); mp 140-143 °C; HPLC = 98% pure (R_T = 28.87 min)

δ_H (400 MHz, MeOD): 7.31 - 7.21 (m, 2H, H-2 and H-2'), 7.19 - 7.14 (m, 2H, H-6 and H-6'), 7.12 - 7.08 (m, 4H, H-3/H-3' and H-7/H-7'), 6.94 (t, J = 7.4, H-8), 3.41 (m, 4H, 2xCH₂), 3.37 (m, 4H, 2xCH₂).

δ_C (151 MHz, MeOD): 150.9 (s, CN), 150.7 (s, C-1), 128.8 (s, C-2), 126.8 (s, C-6), 125.8 (s, C-4), 116.8 (s, C-9), 116.6 (s, C-3), 49.7 (s, CH₂), 48.7 (s, CH₂), 29.2 (s, CH₂), 22.8 (s, CH₂).

HRMS-ESI (+): m/z calculated for C₁₇H₂₁N₅: 296.1797: Found: 296.1870 [M+H]⁺

ν_{max} (ATR)/ cm⁻¹: 2936 (CH₂), 1629 (NO), 1498, 1395, 829, 807.

TFA salt of *N*-[4-(4-phenylpiperazin-1-yl)phenyl]imidazolidin-2-imine (73b**)**

Following method I, from **102b** (143 mg, 0.275 mmol), **73b** was obtained as a light-green powder. (89 mg, 58% yield); mp 125-130 °C; HPLC = 96% pure (R_T = 28.22 min)

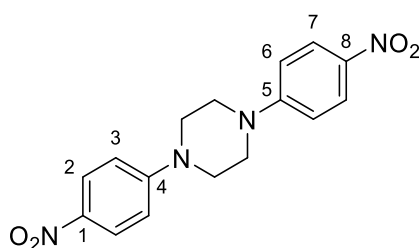
δ_H (400 MHz, MeOD): 7.38 - 7.35 (m, 2H, H-2 and H-2'), 7.23 - 7.21 (m, 4H, H-6/H-6' and H-7/H-7'), 7.17 - 7.13 (m, 2H, H-3 and H-3'), 7.09 - 7.04 (m, 1H, H-8), 3.77 (s, 4H, 2xCH₂), 3.47 (s, 8H, 4xCH₂).

δ_c (151 MHz, MeOD): 159.3 (s, CN), 150.1 (s, C-1), 149.0 (s, C-4), 129.1 (s, C-2), 127.3 (s, C-5), 125.6 (s, C-6), 122.4 (s, C-8), 117.4 (s, C-7), 115.0 (s, C-3), 50.7 (s, CH₂), 48.5 (s, CH₂), 42.7 (s, CH₂), 29.2 (s, CH₂), 22.8 (s, CH₂).

HRMS-ESI (+): m/z calculated for C₁₉H₂₃N₅: 322.1953: Found: 322.2026 [M+H]⁺

$\nu_{\max}$ (ATR)/ cm⁻¹: 3182 (NH), 2962 (CH₂), 2911 (CH₂), 1674 (CO), 1600, 830.

1,4-Bis(4-nitrophenyl) piperazine (74)²

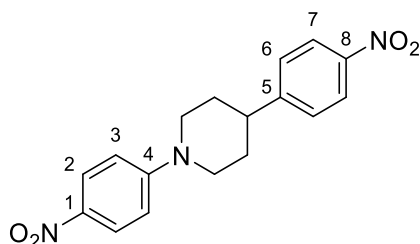


A solution of 1-(4-nitrophenyl) piperazine (1000 mg, 4.826 mmol) and 1-fluoro-4-nitrobenzene (229 mg, 1.61 mmol) in DMSO (20 mL) was heated at 100 °C for 58 h. The reaction was then cooled and poured into ice water (50 mL). The tangy yellow precipitate was collected and rinsed with water. Recrystallisation with Tol/EtOH afforded a dark red fine solid (437 mg, 83%); *R_f* 0.4 (1:1 Hex: EtOAc); mp 260 -265 °C (lit. 265 - 266 °C)²

δ_H (400 MHz, DMSO-*d*₆): 8.08 – 7.9 (m, 4H, Ar), 7.00 – 6.8 (m, 4H, Ar), 3.6 (m, 4H, 2xCH₂), 2.49 (m, 4H, 2xCH₂).

δ_c (101 MHz, DMSO-*d*₆): 154.5 (C-1 and C-8), 137.3 (C-4 and C-5), 125.2 (Ar), 112.4 (Ar), 45.5 (CH₂).

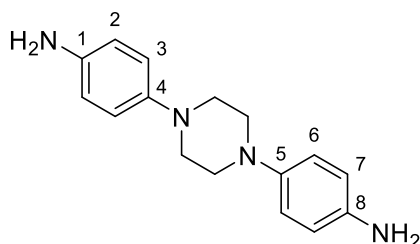
HRMS-ESI (+) m/z calculated for C₁₆H₁₆N₄O₄: 329.1172; found: 329.1258 [M+H]⁺

1,4-Bis(4-nitrophenyl) piperidine (75)²

A solution of 4-(4-nitrophenyl)piperidine (1000 mg, 4.84 mmol) and 1-fluoro-4-nitrobenzene (229 mg, 1.61 mmol) in DMSO (20 mL) was heated at 100 °C for 58 h. The reaction was then cooled and poured into ice water (50 mL). The dark yellow precipitate was collected and rinsed with water. Recrystallisation with Tol/EtOH afforded an orange solid (473 mg, 90%) R_f 0.4 (1:1 Hex: EtOAc); mp 170 -175 °C (lit. 170 - 172 °C)²

δ_H (400 MHz, DMSO-*d*₆): 8.18 (d, J = 8.0 Hz, 2H, H-2 and H-2'), 8.10 (d, J = 9.0 Hz, 2H, H-7 and H-7'), 7.58 (d, J = 8.0 Hz, 2H, H-3 and H-3'), 7.10 (d, J = 9.0 Hz, 2H, H-6 and H-6'), 4.12 – 4.32 (m, 2H, CH₂(PD)), 3.00 – 3.20 (m, 3H, CH₂(PD) and CH(PD)), 1.85 – 1.95 (m, 2H, CH₂(PD)), 1.60 – 1.78 (m, 2H, CH₂(PD)).

HRMS-ESI (+) m/z calculated for C₁₇H₁₇N₃O₄: 328.1219; found: 328.1272 [M+H]⁺

1,4-Bis(4-aminophenyl) piperazine (76)²

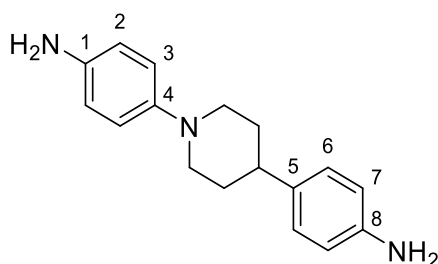
Compound **74** (300 mg, 0.914 mmol) was dissolved in HCl saturated methanolic solution (1.25 M, 35 mL) and stirred at room temperature under H₂ gas in the presence of 10% Pd/C (70 mg) for 14 h. Upon completion, the catalyst was filtered and the solvent was removed to obtain the solid crude HCl salt which was recrystallized with ethanol to yield a brown crystalline solid. (194 mg, 79%) R_f 0.5 (4:1 MeOH: Hex); mp decomposes over 350 °C (lit. >350 °C)²

δ_{H} (400 MHz, D_2O): 7.32 (br. s, 8H, Ar), 3.57 (s, 8H, $4\times\text{CH}_2$ (PZ)).

δ_{C} (101MHz, D_2O): 123.8 (Ar), 119.8 (Ar), 50.3 (CH_2).

HRMS-ESI (+) m/z calculated for $\text{C}_{16}\text{H}_{20}\text{N}_4$: 269.1688; found: 269.1757 $[\text{M}+\text{H}]^+$

4-[4-(4-Aminophenyl)piperidin-1-yl]aniline (**77**)²

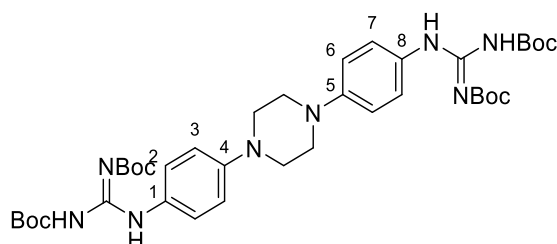


26 (300 mg, 0.916 mmol) was dissolved in HCl saturated methanolic solution (1.25 M, 35 mL) and stirred at room temperature under H_2 gas in the presence of 10% Pd/C (70 mg) for 14 h. Upon completion, the catalyst was filtered and the solvent was removed to obtain the solid crude HCl salt which was recrystallized with ethanol to yield a brown crystalline solid. (123 mg, 50%) R_f 0.6 (4:1 MeOH: Hex); mp 155-160 °C (lit. 156 - 158 °C)²

δ_{H} (400 MHz, $\text{DMSO}-d_6$): 6.91 (d, $J = 8.0$ Hz, 2H, H-2 and H-2'), 6.82 (d, $J = 8.8$ Hz, 2H, H-7 and H-7'), 6.37 - 6.42 (m, 4H, H-3/H-3' and H-6/H-6'), 4.60 (br. s, 2H, NH_2), 4.75 (br. s, 2H, NH_2), 3.67 – 3.79 (m, 2H, CH_2 (PD)), 2.77 – 2.85 (m, 2H, CH_2 (PD)), 2.15 – 2.22 (m, 1H, CH (PD)), 1.65 – 1.82 (m, 2H, CH_2 (PD)).

HRMS-ESI (+) m/z calculated for $\text{C}_{17}\text{H}_{21}\text{N}_3$: 268.1725; found: 268.1777 $[\text{M}+\text{H}]^+$

1,4-Bis[4-(N^2,N^3 -bis-(*tert*-butyloxycarbonyl)guanidino)- phenyl]piperazine (**78**)²



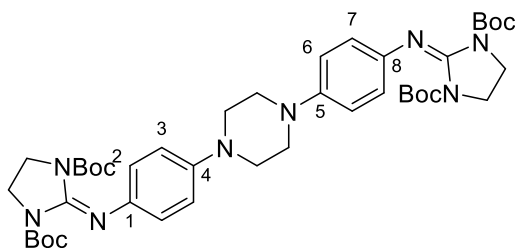
To a solution of **27** (100 mg, 0.373 mmol), 1,3-bis-(*tert*-butoxycarbonyl)-2-methyl-2-thiopseudourea (270 mg, 0.933 mmol) and triethylamine (2.1 mL) in DMF (10 mL) at 0°C was added HgCl₂ (253 mg, 0.933 mmol). This cloudy mixture was stirred at 0°C for 30 min followed by room temperature overnight. The reaction mixture was then diluted with CH₂Cl₂, filtered through a path of Celite and rinsed through with CH₂Cl₂. The organic phase was then washed with brine, dried over anhydrous MgSO₄ and concentrated under reduced pressure. The dark crude residue was then purified by alumina gel column chromatography (1:1 Hex: EtOAc) to yield a pale yellow solid. (200 mg, 72% yield) *R*_f 0.6; mp decomposes over 295 °C (lit. >300 °C)²

δ_{H} (400 MHz, CDCl₃): 11.64 (s, 2H, NH), 10.17 (s, 2H, NH), 7.49 (d, *J* = 9.0 Hz, 4H, H-2/H-2' and H-7/H-7'), 6.94 – 6.89 (m, 4H, H-3/H-3' and H-6/H-6'), 3.28 (s, 8H, 2 x CH₂ (PZ)), 1.51-1.49 (m, 36H, 4x C(CH₃)₃).

δ_{C} (151 MHz, CDCl₃): 164.0 (CN), 153.5 (CO), 152.5 (CO), 148.5 (Ar), 129.4 (Ar), 123.4 (Ar), 116.8 (Ar), 83.4 (C(CH₃)₃), 79.3 (C(CH₃)₃), 49.7 (CH₂), 28.2 (CH₃).

HRMS-ESI (+) *m/z* calculated for C₃₈H₅₆N₈O₈: 753.4221; found 753.4277 [M+H]⁺

Di-*tert*-butyl-2-(4-[4-(4-[1,3-bis-(*tert*-butyloxycarbonyl)-tetrahydro-1H-2-imidazolyliden]aminophenyl)piperazino]-phenylimino)-1,3-imidazolidinedicarboxylate (79**)²**



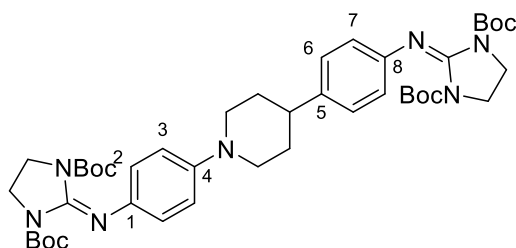
To a solution of **76** (50 mg, 0.186 mmol), *N,N'*-bis-(*tert*-butoxycarbonyl)-imidazoline-2-thione (169 mg, 0.558 mmol) and triethylamine (0.25 mL) in DMF (8 mL) was added HgCl₂ (126 mg, 0.465 mmol) at 0 °C. This cloudy mixture was stirred at 0°C for 30 min followed by room temperature overnight. The reaction mixture was then diluted with CH₂Cl₂, filtered through a path of Celite and rinsed through with CH₂Cl₂. The organic phase was then washed with brine, dried over anhydrous MgSO₄ and concentrated under reduced pressure. The dark crude

residue was then purified by alumina gel column chromatography (1:1 Hex: EtOAc) to yield a yellow solid (42 mg, 28% yield) R_f 0.5; mp decomposes over 250 °C (no lit. mp reported)²

δ_H (400 MHz, $CDCl_3$): 7.05 - 6.92 (m, 8H, Ar), 3.81 (br. s, 8H, CH_2 (PZ)), 3.22 (s, 8H, CH_2 (Imi)), 1.58 – 1.49 (m, 36H, 4x $C(CH_3)_3$).

HRMS-ESI (+) m/z calculated for $C_{42}H_{60}N_8O_8$: 805.4534: found: 805.4581 $[M+H]^+$

1,3-Di-*tert*-butyl-2-[(4-{1-[4-({1,3-bis[(*tert*-butoxy)carbonyl]imidazolidin-2-ylidene}amino)phenyl]piperidin-4-yl}phenyl)imino]imidazolidine-1,3-dicarboxylate (80)²

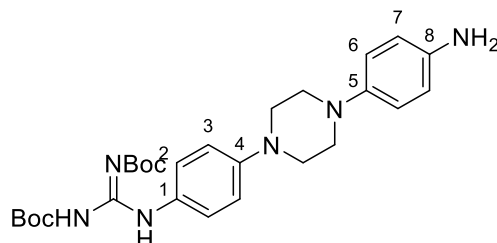


To a solution of **77** (50 mg, 0.186 mmol), *N,N'*-bis-(*tert*-butoxycarbonyl)-imidazoline-2-thione (169 mg, 0.558 mmol) and triethylamine (0.25 mL) in DMF (8 mL) was added $HgCl_2$ (126 mg, 0.465 mmol) at 0 °C. This cloudy mixture was stirred at 0 °C for 30 min followed by room temperature overnight. The reaction mixture was then diluted with CH_2Cl_2 , filtered through a path of Celite and rinsed through with CH_2Cl_2 . The organic phase was then washed with brine, dried over anhydrous $MgSO_4$ and concentrated under reduced pressure. The dark crude residue was then purified by alumina gel column chromatography (1:1 Hex: EtOAc) to yield a yellow solid (42 mg, 28% yield) R_f 0.7 (1:1 Hex: EtOAc); mp 190-195 °C (lit. 191 - 193 °C)²

δ_H (400 MHz, $CDCl_3$): 7.09 (d, J = 7.9 Hz, 2H, H-2 and H-2'), 7.00 – 6.85 (m, 6H, H-3/H-3', H-6/H-6' and H-7/H-7'), 3.81 (s, 8H, CH_2 (Imi)), 3.73 – 3.60 (m, 2H, CH_2 (PZ)), 2.71 (s, 2H, CH_2 (PZ)), 1.83–1.98 (m, 4H, 2x CH_2 (PZ)), 1.33 (s, 18H, 2x $C(CH_3)_3$), 1.30 (s, 18H, 2x $C(CH_3)_3$)

HRMS-ESI (+) m/z calculated for $C_{43}H_{61}N_7O_8$: 804.4582: found: 804.4650 $[M+H]^+$

***tert*-Butyl-*N*-[(1*E*)-({4-[4-(4-aminophenyl)piperazin-1-yl]phenyl}amino){[(*tert*-butoxy)carbonyl]imino)}methyl]carbamate (**81**)¹**

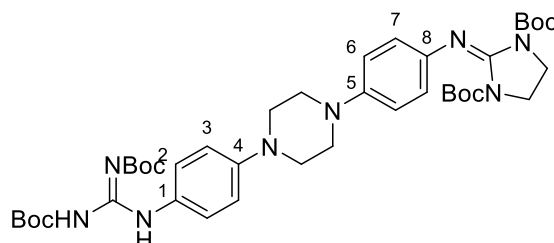


Following method B, from **27** (50 mg, 0.186 mmol), 1,3-bis-(*tert*-butoxycarbonyl)-2-methyl-2-thiopseudourea (16 mg, 0.056 mmol), triethylamine (0.05 mL) and HgCl₂ (18 mg, 0.067 mmol) a light brown solid was obtained. (100 mg, 62% yield) R_f 0.3 (1:1 Hex: EtOAc); mp decomposes over 200 °C (lit. >220 °C)¹

δ_{H} (400 MHz, CDCl₃): 11.64 (s, 1H, NH), 10.16 (s, 1H, NH), 7.46 (d, *J* = 9.0 Hz, 2H, H-2 and H-2'), 7.04 (d, *J* = 8.4 Hz, 2H, H-7 and H-7'), 6.93 (d, *J* = 9.0 Hz, 2H, H-3 and H-3'), 6.66 (d, *J* = 8.4 Hz, 2H, H-6 and H-6'), 3.73 (br. s, 8H, CH₂ (PD)) , 1.50 (s, 18H, 2xC(CH₃)₃)

HRMS-ESI (+) *m/z* calculated for C₂₇H₃₈N₆O₄: 511.2955: found: 511.2985 [M+H]⁺

1,3-Di-*tert*-butyl-2-({4-[4-(4-{[(1*E*)-{[(*tert*-butoxy)carbonyl]amino){[(*tert*-butoxy)carbonyl]imino)}methyl]amino}phenyl)piperazin-1-yl]phenyl]imino)imidazolidine-1,3-dicarboxylate (82**)¹**



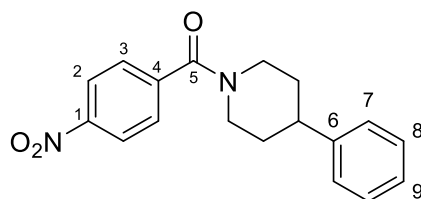
To a solution of **81** (50 mg, 0.097 mmol) in CH₂Cl₂, *N,N'*-bis-(*tert*-butoxycarbonyl)-imidazoline-2-thione (37 mg, 0.122 mmol) and triethylamine (0.1 mL) in DMF (8 mL) was added HgCl₂ (33 mg, 0.122 mmol) at 0 °C. This cloudy mixture was stirred at 0°C for 30 min followed by room

temperature overnight. The reaction mixture was then diluted with CH_2Cl_2 , filtered through a path of Celite and rinsed through with CH_2Cl_2 . The organic phase was then washed with brine, dried over anhydrous MgSO_4 and concentrated under reduced pressure. The crude residue was then purified by alumina gel column chromatography (1:1 Hex: EtOAc) to yield a pale brown solid (15 mg, 19% yield) R_f 0.7 (1:1 Hex: EtOAc); mp 85-88 °C (lit. 88-90 °C)¹

δ_{H} (400 MHz, CDCl_3): 11.64 (s, 1H, NH) 10.16 (s, 1H, NH), 7.46 (d, J = 8.3 Hz, 2H, H-2 and H-2'), 7.04 - 6.93 (m, 6H, H-3/H-3', H-6/H-6' and H-7/H-7'), 3.80 (s, 8H, $\text{CH}_2(\text{PZ})$), 3.73 (m, 8H, $\text{CH}_2(\text{Imi})$), 1.50 (s, 18H, 2 x $\text{C}(\text{CH}_3)_3$), 1.35 (s, 18H, 2 x $\text{C}(\text{CH}_3)_3$).

HRMS-ESI (+) m/z calculated for $\text{C}_{40}\text{H}_{58}\text{N}_8\text{O}_8$: 779.4378; found 779.4389 $[\text{M}+\text{H}]^+$

1-(4-Nitrobenzoyl)-4-phenylpiperidine (**83**)



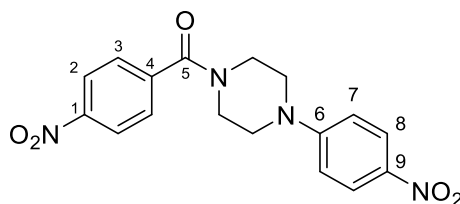
Following method G, 4-nitrobenzoyl chloride (150 mg, 0.808 mmol), 4-phenylpiperidine (156 mg, 0.970 mmol) and NEt_3 (0.14 mL, 0.970 mmol) in CH_2Cl_2 (6 mL) afforded **83** as a white crystalline solid. (211 mg, 84% yield); R_f 0.7 (1:1 EtOAc: Hex); mp 159-161 °C

δ_{H} (400 MHz, CDCl_3): 8.32 (d, J = 8.7 Hz, H-2 and H-2'), 7.64 (d, J = 8.7 Hz, H-3 and H-3'), 7.36 (d, J = 8.0 Hz, 2H, H-7 and H-7'), 7.27-7.24 (m, 3H, H-8/H-8' and H-9), 4.92 (br. s, 2H, CH_2), 3.75 (br. s, 2H, CH_2), 3.22 (br. s, 2H, CH_2), 2.95 (br. s, 2H, CH_2), 2.86-2.82 (m, 1H, CH).

δ_{C} (150 MHz, CDCl_3): 168.0 (s, CO), 148.3 (s, C-1), 144.6 (s, C-4), 142.4 (s, C-11), 128.7 (s, C-2), 127.9 (s, C-3), 126.6 (s, C-7), 126.7 (s, C-8), 123.9 (s, C-9), 48.3 (s, CH_2), 42.9 (s, CH_2), 42.6 (s, CH_2), 33.9 (s, CH_2), 32.7 (s, CH).

HMRS-ESI (+): m/z calculated for $\text{C}_{18}\text{H}_{18}\text{N}_2\text{O}_3$: 311.1317 Found: 311.1391 $[\text{M}+\text{H}]^+$

ν_{max} (ATR)/ cm^{-1} : 2941 (CH_2), 2865 (CH_2), 1631 (CO), 1601 (NO_2), 1510, 1346 (NO_2), 845.

1-(4-Nitrobenzoyl)-4-(4-nitrophenyl)piperazine (84**)³**

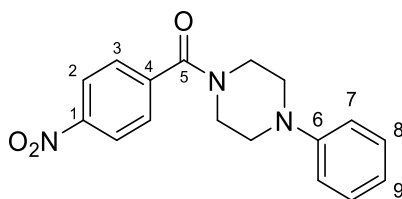
Following method G, from 4-nitrobenzoyl chloride (672 mg, 3.621 mmol), 1-(4-nitrophenyl)piperazine (900 mg, 4.343 mmol) and NEt₃ (0.6 mL, 4.329 mmol) in CH₂Cl₂ (2 mL), **84** was obtained as a dark-orange solid. (1140 mg, 88% yield); R_f 0.4 (6:4 EtOAc: Hex); mp 135-140 °C (lit. 145 °C)³

δ_{H} (400 MHz, CDCl₃): 8.34 (d, J = 8.6 Hz, 2H, H-2 and H-2'), 8.17 (d, J = 9.3 Hz, 2H, H-8 and H-8'), 7.66 (d, J = 8.6 Hz, 2H, H-3 and H-3'), 6.88 (d, J = 9.3 Hz, 2H, H-7 and H-7'), 4.00 (s, 2H, CH₂), 3.59 (s, 4H, 2xCH₂), 3.44 (s, 2H, CH₂).

δ_{C} (151 MHz, CDCl₃): 168.2 (s, CO), 154.3 (s, C-1), 148.6 (s, C-9), 141.1 (s, C-4), 139.5 (s, C-6), 128.2 (s, C-2), 125.9 (s, C-8), 124.0 (s, C-3), 113.4 (s, C-7), 47.2 (s, CH₂), 41.7 (s, CH₂), 30.9 (s, CH₂).

HMRS-ESI (+): m/z calculated for C₁₇H₁₆N₄O₅: 356.1121: Found: 379.1009 [M+ Na]⁺

ν_{max} (ATR)/ cm⁻¹: 2921 (CH₂), 2849 (CH₂), 1598 (NO₂), 1314 (NO₂), 1116, 1004, 862.

1-(4-Nitrobenzoyl)-4-phenylpiperazine (85**)³**

Following method G, 4-nitrobenzoyl chloride (286 mg, 1.541 mmol), 1-phenylpiperazine (300 mg, 1.849 mmol) and NEt₃ (0.26 mL, 1.849 mmol) in CH₂Cl₂ (8 mL) afforded **85** as a yellow powder. (422 mg, 88% yield); R_f 0.8 (1:1 EtOAc: Hex); mp 128-130 °C (lit. 122 °C)³

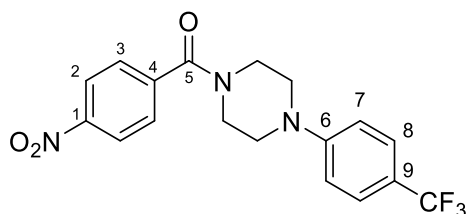
δ_{H} (400 MHz, CDCl_3): 8.33 (d, $J = 8.7$ Hz, 2H, H-2 and H-2'), 7.64 (d, $J = 8.7$ Hz, 2H, H-3 and H-3'), 7.32 (m, 3H, H-8/H-8' and H-9), 6.97 (d, $J = 6.7$ Hz, 2H, H-7 and H-7'), 4.00 (br. s, 2H, CH_2), 3.57 (br. s, 2H, CH_2), 3.32 (br. s, 2H, CH_2), 3.16 (br. s, 2H, CH_2).

δ_{C} (151 MHz, CDCl_3): 168.1 (s, CO), 150.6 (s, C-1), 148.5 (s, C-4), 141.7 (s, C-6), 129.7 (s, C-2), 128.1 (s, C-3), 124.0 (s, C-7), 121.0 (s, C-8), 116.9 (s, C-9), 50.1 (s, CH_2), 49.6 (s, CH_2), 47.4 (s, CH_2), 42.2 (s, CH_2).

HMRS-ESI (+): m/z calculated for $\text{C}_{17}\text{H}_{17}\text{N}_3\text{O}_3$: 312.1270 Found: 312.1368 $[\text{M}+\text{H}]^+$

ν_{max} (ATR)/ cm^{-1} : 2829 (CH_2), 1625 (CO), 1598 (NO_2), 1347 (NO_2), 845.

1-(4-Nitrobenzoyl)-4-[4-(trifluoromethyl)phenyl]piperazine (**86**)



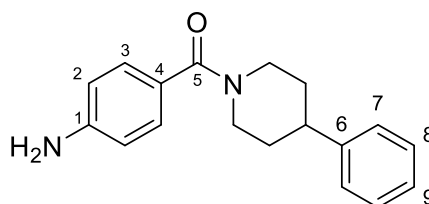
Following method G, 4-nitrobenzoyl chloride (39 mg, 0.201 mmol), 1-[4-(trifluoromethyl)phenyl]-piperazine (58 mg, 0.252 mmol) and NEt_3 (0.03 mL, 0.252 mmol) in CH_2Cl_2 (6 mL) afforded **86** as a yellow crystalline solid. (71 mg, 89% yield); R_f 0.5 (1:1 EtOAc: Hex); mp 165-167 °C.

δ_{H} (400 MHz, CDCl_3): 8.34 (d, $J = 8.8$ Hz, H-2 and H-2'), 7.64 (d, $J = 8.8$ Hz, 2H, H-3 and H-3'), 7.56-7.54 (m, 2H, H-8 and H-8'), 6.98-6.96 (m, 2H, H-7 and H-7'), 3.99 (br. s, 2H, CH_2), 3.58 (br. s, 2H, CH_2), 3.41 (br. s, 2H, CH_2), 3.26 (br. s, 2H, CH_2).

δ_{C} (151 MHz, CDCl_3): 168.1 (s, CO), 152.7 (s, C-1), 148.5 (s, C-11), 147.2 (s, C-4), 141.4 (s, C-8), 128.2 (s, C-2), 126.6 (s, C-8), 124.0 (s, C-3), 115.4 (s, C-7), 41.8 (s, CH_2).

HMRS-ESI (-): m/z calculated for $\text{C}_{18}\text{H}_{16}\text{F}_3\text{N}_3\text{O}_3$: 378.1144 Found: 378.1153 $[\text{M}-\text{H}]^-$

ν_{max} (ATR)/ cm^{-1} : 2926 (CH_2), 2859 (CH_2), 1636 (CO), 1613 (NO_2), 1348 (NO_2), 1226 (CF_3), 847.

4-(4-Phenylpiperidine-1-carbonyl)aniline (87**)**

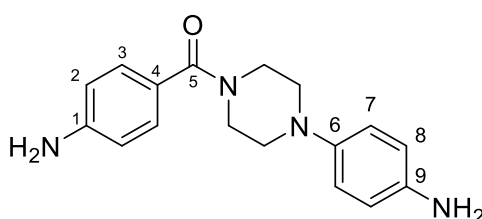
Following method H, from **83** (193 mg, 0.622 mmol) and 10% Pd/C (75 mg, 0.249 mmol) in EtOAc (10 mL), **87** was obtained as a light-brown crystalline solid. (165 mg, 95% yield); R_f 0.1 (1:1 EtOAc: Hex); mp 145-147 °C

δ_H (400 MHz, $CDCl_3$): 7.35 - 7.33 (m, 4H, H-2/H-2' and H-3/H-3'), 7.25-7.23 (m, 3H, H-8/H-8' and H-9), 6.69 (d, J = 8.4 Hz, 2H, H-7 and H-7'), 4.48 (br. s, 2H, NH_2), 3.46 (s, 4H, $2 \times CH_2$), 3.01 (s, 4H, $2 \times CH_2$), 2.83 - 2.78 (m, 1H, CH).

δ_C (151 MHz, $CDCl_3$): 170.9 (s, CO), 147.8 (s, C-1), 145.3 (s, C-4), 129.1 (s, C-2), 128.5 (s, C-3), 126.7 (s, C-8), 126.2 (s, C-7), 114.3 (s, C-9), 42.9 (s, CH_2), 33.4 (s, CH_2).

HMRS-ESI (+): m/z calculated for $C_{18}H_{20}N_2O$: 281.1576 Found: 281.1648 $[M+H]^+$

ν_{max} (ATR)/ cm^{-1} : 3403 (NH_2), 2926 (CH_2), 2854 (CH_2), 1590, 995, 841.

4-[4-(4-Aminobenzoyl)piperazin-1-yl]aniline (88**)³**

Following method H, from **84** (400 mg, 1.123 mmol) and 10% Pd/C (130 mg, 0.449 mmol) in DMF (4 mL), **88** was obtained as a purple crystalline solid. (301 mg, 90% yield); R_f 0.1 (7:3 EtOAc: Hex); mp 122-127 °C (lit. 126-127 °C)³

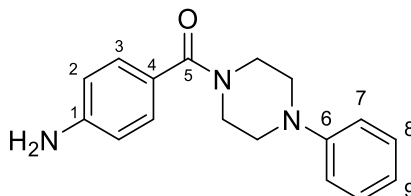
δ_H (400 MHz, $CDCl_3$): 7.25 (d, J = 8.5 Hz, 2H, H-2 and H-2'), 6.87 (d, J = 8.8 Hz, 2H, H-7 and H-7'), 7.69 - 7.63 (m, 4H, H-3/H-3' and H-6/H-6'), 3.79 (s, 4H, $2 \times CH_2$), 3.05 - 3.02 (m, 4H, $2 \times CH_2$).

δ_C (151 MHz, $CDCl_3$): 172.1 (s, CO), 150.6 (s, C-1), 143.9 (s, C-9), 141.5 (s, C-4), 139.5 (s, C-6), 128.9 (s, C-2), 122.5 (s, C-8), 119.0 (s, C-3), 113.5 (s, C-7), 51.4 (s, CH_2), 22.2 (s, CH_2).

HMRS-ESI (+): m/z calculated for C₁₇H₂₀N₄O: 297.1637 Found: 297.1710 [M+H]⁺

v_{max} (ATR)/ cm⁻¹: 3449 (NH₂), 3218 (NH₂), 2803 (CH₂), 1604 (CO), 1587, 811.

4-(4-Phenylpiperazine-1-carbonyl)aniline (**89**)³



Following method H, from **85** (373 mg, 1.198 mmol) and 10% Pd/C (140 mg, 0.479 mmol) in EtOAc (15 mL), **89** was obtained as a white crystalline solid. (319 mg, 95% yield); *R_f* 0.1 (1:1 EtOAc: Hex); mp 145-150 °C (lit. 149 °C)³

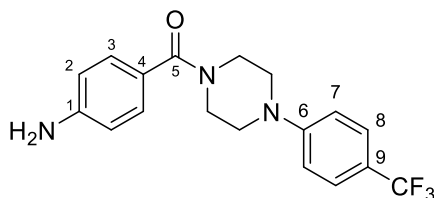
δ_H (400 MHz, CDCl₃): 7.32 - 7.30 (m, 4H, H-2/H-2' and H-3/H-3'), 6.96-6.92 (m, 3H, H-8/H-8' and H-9), 6.69 (d, *J* = 8.4 Hz, 2H, H-7 and H-7'), 3.92 (s, 2H, NH₂), 3.81 (s, 4H, 2xCH₂), 3.21 (s, 4H, 2xCH₂).

δ_C (151 MHz, CDCl₃): 170.8 (s, CO), 151.0 (s, C-1), 148.2 (s, C-4), 141.7 (s, C-6), 129.4 (s, C-2), 129.2 (s, C-3), 124.9 (s, C-8), 120.5 (s, C-7), 116.6 (s, C-9), 49.7 (s, CH₂).

HMRS-ESI (+): m/z calculated for C₁₇H₁₉N₃O: 282.1528 Found: 282.1596 [M+H]⁺

v_{max} (ATR)/ cm⁻¹: 3218 (NH₂), 2885 (CH₂), 2818 (CH₂), 1619 (CO), 1598, 761.

4-{4-[4-(Trifluoromethyl)phenyl]piperazine-1-carbonyl}aniline (**90**)



Following method H, from **86** (78 mg, 0.206 mmol) and 10% Pd/C (30 mg, 0.082 mmol) in EtOAc (4 mL), **90** was obtained as a purple crystalline solid. (47mg, 65% yield); *R_f* 0.1 (1:1 Hex: EtOAc); mp 155-159 °C.

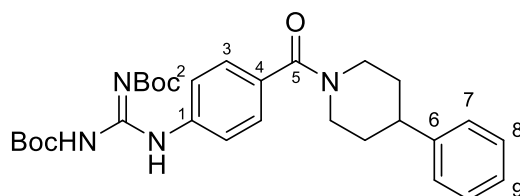
δ_{H} (400 MHz, CDCl_3): 7.52 (d, $J = 8.6$ Hz, 2H, H-2 and H-2'), 7.32 (d, $J = 8.3$ Hz, 2H, H-8 and H-8'), 6.95 (d, $J = 8.6$ Hz, 2H, H-3 and H-3'), 6.69 (d, $J = 8.3$ Hz, 2H, H-7 and H-7'), 3.95 (s, 2H, NH_2), 3.82 (br. s, 4H, CH_2), 3.31 (br. s, 2H, CH_2), 3.26 (s, 2H, CH_2)

δ_{C} (151 MHz, CDCl_3): 160.9 (s, CO), 153.0 (s, C-1), 148.4 (s, C-4), 129.4 (s, C-2), 126.5 (s, C-3), 124.5 (s, C-8), 115.0 (s, C-7), 48.4 (s, CH_2)

HMRS-ESI (+): m/z calculated for $\text{C}_{18}\text{H}_{18}\text{F}_3\text{N}_3\text{O}$: 350.1402 Found: 350.1417 $[\text{M}+\text{H}]^+$

ν_{max} (ATR)/ cm^{-1} : 2929 (CH_2), 1636 (CO), 1617, 1292, 1219 (CF_3), 817.

***tert*-Butyl *N*-[(*E*)-{[(*tert*-butoxy)carbonyl]amino}{[4-(4-phenylpiperidine-1-carbonyl)phenyl]amino}}methylidene] carbamate (91a)**



Following method B, from **87** (115 mg, 0.410 mmol), 1,3-bis-(*tert*-butoxycarbonyl)-2-methyl-2-thiopseudourea (125 mg, 0.431 mmol, 1.05 equiv.), HgCl_2 (123 mg, 0.451 mmol, 1.1 equiv.) and NEt_3 (0.1 mL, 0.820 mmol, 2 equiv.) in CH_2Cl_2 (3 mL), **91a** was obtained as a white powder. (160 mg, 75% yield); R_f 0.6 (1:1 EtOAc: Hex); mp 163-165 °C

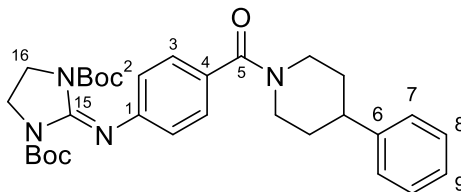
δ_{H} (400 MHz, CDCl_3): 11.66 (s, 1H, NH), 10.49 (s, 1H, NH), 7.71 (d, $J = 8.5$ Hz, 2H, H-2 and H-2'), 7.46 (d, $J = 8.5$ Hz, 2H, H-3 and H-3'), 7.36 - 7.33 (m, 2H, H-7 and H-7'), 7.26 - 7.22 (m, 3H, H-8/H-8' and H-9), 3.11 (s, 2H, CH_2), 2.91 (s, 2H, CH_2), 2.82-2.79 (m, 4H, $2\times\text{CH}_2$), 1.55 (br. s, 18H, $2\times\text{C}(\text{CH}_3)_3$).

δ_{C} (151 MHz, CDCl_3): 170.0 (s, CO), 153.4 (s, CO), 145.2 (s, C-1), 138.1 (s, C-4), 132.1 (s, C-6), 128.6 (s, C-2), 128.0 (s, C-3), 126.7 (s, C-7), 126.5 (s, C-8), 121.6 (s, C-9), 42.8 (s, CH_2), 28.1 (s, CH_2), 28.0 (s, $\text{C}(\text{CH}_3)_3$).

HMRS-ESI (+): m/z calculated for $\text{C}_{29}\text{H}_{38}\text{N}_4\text{O}_5$: 523.2842 Found: 523.2908 $[\text{M}+\text{H}]^+$

ν_{max} (ATR)/ cm^{-1} : 2986 (CH_2), 2935 (CH_2), 1631 (CO), 1602, 843.

1,3-Di-*tert*-butyl 2-[[4-(4-phenylpiperidine-1-carbonyl)phenyl]imino]imidazolidine-1,3-dicarboxylate (91b)



Following method B, from **87** (120 mg, 0.464 mmol), *N,N'*-bis-(*tert*-butoxycarbonyl)-imidazoline-2-thione (147 mg, 0.4869 mmol, 1.05 equiv.), HgCl₂ (138 mg, 0.510 mmol, 1.1 equiv.) and NEt₃ (0.13 mL, 0.927 mmol, 2 equiv.) in CH₂Cl₂ (10 mL), **91b** was obtained as a white crystalline solid. (119 mg, 47% yield); *R_f* 0.3 (1:1 EtOAc: Hex); mp 176-181 °C

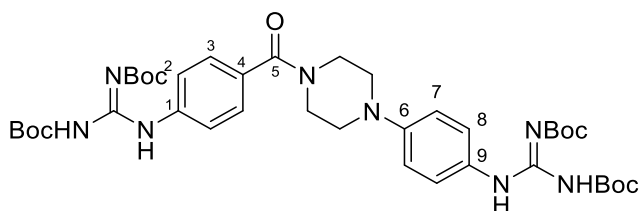
δ_H (400 MHz, CDCl₃): 7.37 - 7.34 (m, 4H, H-2/H-2' and H-3/H-3'), 7.26 - 7.23 (m, 3H, H-8/H-8' and H-9), 7.03 (d, *J* = 8.3 Hz, 2H, H-7 and H-7'), 3.87 (s, 4H, CH₂), 3.25 (s, 2H, CH₂), 2.99 (s, 2H, CH₂), 2.79 - 2.76 (m, 4H, CH₂), 1.88 (s, 2H, CH), 1.36 (s, 18H, 2x C(CH₃)₃).

δ_C (151 MHz, CDCl₃): 170.6 (s, CO), 145.3 (s, C-1), 128.5 (s, C-2), 128.1 (s, C-3), 126.7 (s, C-8), 126.5 (s, C-7), 121.1 (s, C-9), 83.0 (s, C(CH₃)₃), 43.2 (s, CH₂), 42.9 (s, CH₂), 28.3 (s, C(CH₃)₃), 27.9 (s, C(CH₃)₃).

HMRS-ESI (+): *m/z* calculated for C₃₁H₄₀N₄O₅: 549.2999 Found: 549.3071 [M+H]⁺

ν_{max} (ATR)/ cm⁻¹: 3376 (CH₂), 2930 (CH₂), 1627 (CO), 1368, 842.

***tert*--Butyl-*N*-[1-({4-[4-(4-{[(*Z*)-{[(*tert*-butoxy)carbonyl](methyl)amino}]{[(*tert*-butoxy)carbonyl]imino})methyl]amino}phenyl)piperazine-1-carbonyl]phenyl)amino)ethenyl]-*N*-methylcarbamate (92a)**



Following method B, from **88** (130 mg, 0.439 mmol), 1,3-bis-(*tert*-butoxycarbonyl)-2-methyl-2-thiopseudourea (280 mg, 0.965 mmol), HgCl₂ (262 mg, 0.965 mmol) and NEt₃ (0.31 mL, 2.193 mmol) in DMF (3 mL), **92a** was obtained as a white crystalline solid. (288 mg, 84% yield); R_f 0.3 (1:1 EtOAc: Hex); mp >230 °C

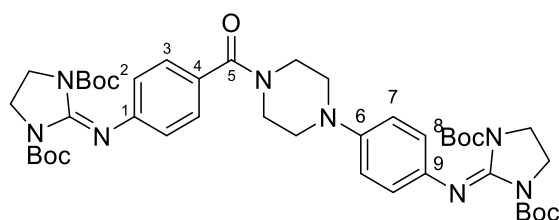
δ_H (400 MHz, CDCl₃): 11.67 (s, 2H, NH), 10.51 (s, 1H, NH), 10.22 (s, 1H, NH), 7.72 (d, *J* = 8.6 Hz, 2H, H-2 and H-2'), 7.51 (d, *J* = 9.0 Hz, H-8 and H-8'), 7.46 (d, *J* = 8.6 Hz, 2H, H-3 and H-3'), 6.92 (d, *J* = 9.0 Hz, 2H, H-7 and H-7'), 3.91 (s, 2H, CH₂), 3.68 (s, 2H, CH₂), 3.16 (s, 4H, CH₂), 1.56 - 1.52 (m, 36H, 4xC(CH₃)₃).

δ_C (151 MHz, CDCl₃): 170.0 (s, CO), 163.6 (s, CN), 163.3 (s, CN), 153.5 (s, CO), 153.4 (s, CO), 153.3 (s, CO), 148.2 (s, CO), 138.4 (s, C-1), 137.8 (s, C-9), 131.4 (s, C-4), 129.9 (s, C-8), 128.2 (s, C-2), 123.5 (s, C-8), 121.7 (s, C-3), 117.2 (s, C-7), 84.0 (s, C(CH₃)₃), 83.5 (s, C(CH₃)₃), 79.4 (s, C(CH₃)₃), 59.9 (s, CH₂), 28.0 (s, C(CH₃)₃).

HMRS-ESI (+): m/z calculated for C₃₉H₅₆N₈O₉: 781.4170 Found: 781.4243 [M+H]⁺

$\nu_{\max}$ (ATR)/ cm⁻¹: 2982 (CH₂), 2936 (CH₂), 1627 (CO), 1408, 806, 761.

1,3-Di-*tert*-butyl-2-[(4-{4-[4-({1,3-bis[(*tert*-butoxy)carbonyl]imidazolidin-2-ylidene} amino) benzoyl]piperazin-1-yl}phenyl)imino]imidazolidine-1,3-dicarboxylate (92b)



Following method B, from **88** (150 mg, 0.506 mmol), *N,N'*-bis-(*tert*-butoxycarbonyl)-imidazoline-2-thione (460 mg, 1.518 mmol), HgCl₂ (412 mg, 1.518 mmol) and NEt₃ (0.35 mL, 2.531 mmol) in DMF (20 mL), **92b** was obtained as a light-brown glassy solid. (303 mg, 78% yield); R_f 0.4 (7:3 EtOAc: Hex); mp 95-100 °C

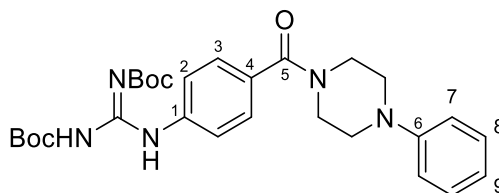
δ_H (400 MHz, CDCl₃): 7.36 (d, *J* = 8.4 Hz, 2H, H-2 and H-2'), 7.02 (d, *J* = 8.4 Hz, 2H, H-3 and H-3'), 6.97 (d, *J* = 8.9 Hz, 2H, H-8 and H-8'), 6.85 (d, *J* = 8.9 Hz, 2H, H-7 and H-7'), 3.84 (br. s, 8H, CH₂ (imi)), 3.06 (s, 4H, CH₂), 2.59 (s, 4H, CH₂), 1.35 (br. s, 36H, 4xC(CH₃)₃).

δ_c (151 MHz, $CDCl_3$): 170.6 (s, CO), 150.3 (s, C-1), 150.0 (s, C-9), 129.0 (s, C-4), 128.3 (s, C-2), 122.3 (s, C-7), 121.1 (s, C-3), 117.8 (s, C-8), 83.0 (s, $C(\underline{CH}_3)_3$), 82.7 (s, $C(\underline{CH}_3)_3$), 43.2 (s, CH_2), 39.5 (s, CH_2), 27.9 (s, $C(CH_3)_3$).

HMRS-ESI (+): m/z calculated for $C_{43}H_{60}N_8O_9$: 833.4483 Found: 833.4555 $[M+H]^+$

ν_{max} (ATR)/ cm^{-1} : 2982 (CH_2), 1700 (CO), 1601 (CO), 1367, 846, 821.

***tert*-Butyl-*N*-[(*E*)-{[(*tert*-butoxy)carbonyl]amino}{[4-(4-phenylpiperazine-1-carbonyl)phenyl]amino}]methylidene]carbamate (**93a**)**



Following method B, from **89** (300 mg, 1.066 mmol), 1,3-bis-(*tert*-butoxycarbonyl)-2-methyl-2-thiopseudourea (325 mg, 1.112 mmol, 1.05 equiv.), $HgCl_2$ (318 mg, 1.173 mmol, 1.1 equiv.) and NEt_3 (0.3 mL, 2.132 mmol, 2 equiv.) in CH_2Cl_2 (10 mL), **93a** was obtained as a white crystalline solid. (360 mg, 65% yield); R_f 0.6 (1:1 EtOAc: Hex); mp 171-174 °C

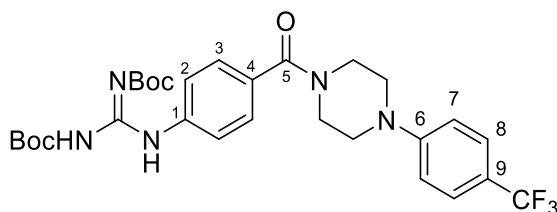
δ_H (400 MHz, $CDCl_3$): 11.67 (s, 1H, NH), 10.51 (s, 1H, NH), 7.72 (d, J = 8.6 Hz, 2H, H-2 and H-2'), 7.46 (d, J = 8.6 Hz, 2H, H-3 and H-3'), 7.34 - 7.30 (m, 2H, H-7 and H-7'), 6.96 - 6.92 (m, 3H, H-8/H-8' and H-9), 3.93 (s, 2H, CH_2), 3.69 (s, 2H, CH_2), 3.21 (s, 4H, $2 \times CH_2$), 1.65 (br. s, 18H, $2 \times C(CH_3)_3$).

δ_c (151 MHz, $CDCl_3$): 170.0 (s, CO), 163.3 (s, CN), 153.4 (s, CO), 138.4 (s, C-1), 131.3 (s, C-4), 129.2 (s, C-2), 128.2 (s, C-3), 121.7 (s, C-7), 120.6 (s, C-8), 116.7 (s, C-9), 84.0 (s, $C(\underline{CH}_3)_3$), 80.0 (s, $C(\underline{CH}_3)_3$), 42.9 (s, CH_2), 33.4 (s, CH_2), 28.1 (s, $C(\underline{CH}_3)_3$).

HMRS-ESI (+): m/z calculated for $C_{28}H_{37}N_5O_5$: 524.2795 Found: 524.2698 $[M+H]^+$

ν_{max} (ATR)/ cm^{-1} : 3116 (NH), 2982 (CH_2), 2813 (CH_2), 1627 (CO), 1600, 755, 724.

***tert*-Butyl-*N*-[(*E*)-{[(*tert*-butoxy)carbonyl]amino}[(4-{4-[4-(trifluoromethyl)phenyl]piperazine-1-carbonyl}phenyl)amino]methylidene]carbamate (**94a**)**



Following method B, from **90** (48 mg, 0.137 mmol), 1,3-bis-(*tert*-butoxycarbonyl)-2-methyl-2-thiopseudourea (42 mg, 0.144 mmol, 1.05 equiv.), HgCl₂ (41 mg, 0.151 mmol, 1.1 equiv.) and NEt₃ (0.04 mL, 0.2748 mmol, 2 equiv.) in CH₂Cl₂ (5 mL), **94a** was obtained as a white crystalline solid. (52 mg, 64% yield); R_f 0.7 (1:1 EtOAc: Hex); mp 176-179 °C

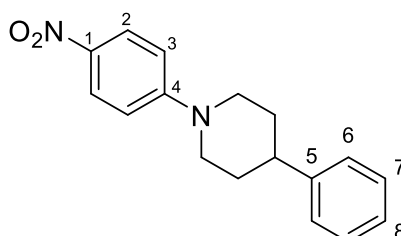
δ_{H} (400 MHz, CDCl₃): 11.67 (s, 1H, NH), 10.52 (s, 1H, NH), 7.73 (d, *J* = 8.3 Hz, 2H, H-2 and H-2'), 7.53 (d, *J* = 8.6 Hz, 2H, H-8 and H-8'), 7.45 (d, *J* = 8.3 Hz, 2H, H-3 and H-3'), 6.96 (d, *J* = 8.6 Hz, 2H, H-7 and H-7'), 3.88 (s, 2H, CH₂), 3.75 (s, 2H, CH₂), 3.31 (s, 4H, CH₂), 1.55 (br. s, 18H, 2xC(CH₃)₃).

δ_{C} (151 MHz, CDCl₃): 170.1 (s, CO), 163.3 (s, CN), 153.5 (s, C-1), 138.5 (s, C-9), 131.1 (s, C-4), 128.2 (s, C-2), 126.5 (s, C-8), 121.7 (s, C-3 and C-7), 115.2 (s, C-9), 84.1 (s, CH₂), 80.0 (s, CH₂), 48.4 (br, s, CH₂), 28.9 (s, 2xC(CH₃)₃).

HMRS-ESI (+): m/z calculated for C₂₉H₃₆F₃N₅O₅: 592.2669 Found: 592.2741 [M+H]⁺

ν_{max} (ATR)/ cm⁻¹: 2988 (CH₂), 1617 (CO), 1407, 824.

1-(4-Nitrophenyl)-4-phenylpiperidine (95**)³**



Following method F, 1-fluoro-4-nitrobenzene (173 mg, 1.240 mmol), 4-phenylpiperidine (200 mg, 1.240 mmol), DIPEA (0.22 mL, 1.240 mmol) in DMF (4 mL) afforded **95** as a yellow powder. (219 mg, 62% yield); R_f 0.5 (8:2 He: EtOAc); mp 175-177 °C. (lit. 174 °C)³

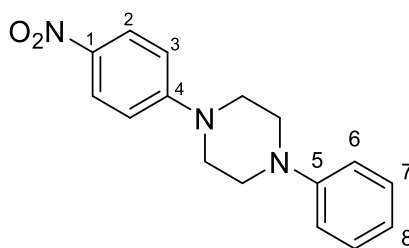
δ_{H} (400 MHz, CDCl_3): 8.16 (d, $J = 9.5$ Hz, 2H, H-2 and H-2'), 7.37 – 7.34 (m, 2H, H-3 and H-3'), 7.29 – 7.24 (m, 3H, H-7/H-7' and H-8), 6.90 – 6.78 (m, 2H, H-6 and H-6'), 4.15 – 4.12 (m, 2H, CH_2), 3.13 – 3.10 (m, 2H, CH_2), 2.85 – 2.82 (m, 1H, H-6), 2.02 (br. s, Hz, 2H, CH_2), 1.87 – 1.84 (m, 2H, CH_2).

δ_{C} (151 MHz, CDCl_3): 154.7 (s, C-1), 145.0 (s, C-4), 137.9 (s, C-5), 128.6 (s, C-2), 126.7 (s, C-3), 126.6 (s, C-6), 126.1 (s, C-7), 112.6 (s, C-8), 48.2 (s, CH_2), 42.4 (s, CH_2), 32.6 (s, CH_2).

HRMS-ESI (+): m/z calculated for $\text{C}_{17}\text{H}_{18}\text{N}_2\text{O}_2$: 283.1368: Found: 283.1440 $[\text{M}+\text{H}]^+$

ν_{max} (ATR)/ cm^{-1} : 2946 (CH_2), 2925 (CH_2), 2849 (CH_2), 1583 (NO_2), 1491, 828.

1-(4-Nitrophenyl)-4-phenylpiperazine (**96**)³



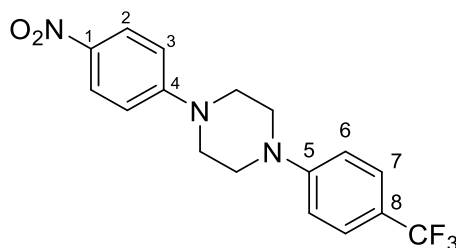
Following method F, 1-fluoro-4-nitrobenzene (348 mg, 2.465 mmol), 1-phenylpiperazine (400 mg, 2.465 mmol), DIPEA (0.43 mL, 2.465 mmol) in DMF (4 mL) afforded **96** as a dark-brown crystalline solid. (545 mg, 78% yield); R_f 0.6 (7:3 Hex: EtOAc); mp 186-188 °C (lit. 184 °C)³

δ_{H} (400 MHz, CDCl_3): 8.15 (d, $J = 9.4$ Hz, 2H, H-2 and H-2'), 7.30 (t, $J = 8.4$ Hz, 2H, H-7 and H-7'), 6.96 (d, $J = 8.4$ Hz, 2H, H-6 and H-6'), 6.92 (t, $J = 7.4$ Hz, 1H, H-8), 6.87 (d, $J = 9.4$ Hz, 2H, H-3 and H-3'), 3.61 – 3.55 (m, 4H, CH_2), 3.38 – 3.33 (m, 4H, CH_2).

δ_{C} (151 MHz, CDCl_3): 154.7 (s, C-1), 150.7 (s, C-4), 138.7 (s, C-6), 129.3 (s, C-2), 126.0 (s, C-3), 120.4 (s, C-6), 116.3 (s, C-7), 112.7 (s, C-8), 48.9 (s, CH_2), 47.0 (s, CH_2).

HRMS-ESI (+): m/z calculated for $\text{C}_{16}\text{H}_{17}\text{N}_3\text{O}_2$: 284.1321: Found: 284.1394 $[\text{M}+\text{H}]^+$

ν_{max} (ATR)/ cm^{-1} : 3213 (CH_2), 2982 (CH_2), 2829 (CH_2), 1598 (NO_2), 1516, 941, 861.

1-(4-Nitrophenyl)-4-[4-(trifluoromethyl)phenyl]piperazine (97)³

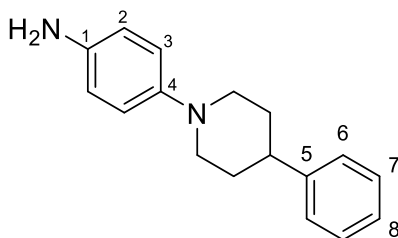
Following method F, 1-fluoro-4-nitrobenzene (306 mg, 2.172 mmol), 1-[4-(trifluoromethyl)phenyl]piperazine (500 mg, 2.172 mmol), DIPEA (0.38 mL, 2.172 mmol) in DMF (5 mL) afforded **97** as an orange crystalline solid. (560 mg, 73% yield); R_f 0.6 (7:3 Hex: EtOAc); mp 176-181 °C (lit. 181-183 °C)³

δ_H (400 MHz, $CDCl_3$): 8.19 (d, J = 8.9 Hz, 2H, H-2 and H-2'), 7.55 (d, J = 8.1 Hz, 2H, H-6 and H-6'), 6.97 (d, J = 8.1 Hz, 2H, H-7 and H-7'), 6.88 (d, J = 8.9 Hz, 2H, H-3 and H-3'), 3.64 (br. s, 4H, CH_2), 3.51 (br. s, 4H, CH_2).

δ_C (150.6 MHz, $CDCl_3$): 154.4 (s, C-1), 152.5 (s, C-4), 138.8 (s, C-8), 126.6 (s, C-6), 126.0 (s, C-2), 114.5 (s, C-7), 112.6 (s, C-3), 47.4 (s, CH_2), 46.6 (s, CH_2).

HRMS-ESI (+): m/z calculated for $C_{17}H_{16}F_3N_3O_2$: 352.1195: Found: 352.1262 $[M+H]^+$

ν_{max} (ATR)/ cm^{-1} : 3085 (CH_2), 2854 (CH_2), 1592 (NO_2), 1323, 940, 824.

4-(4-Phenylpiperidin-1-yl)-alanine (98)³

Following method H, from **95** (210 mg, 0.744 mmol) and 10% Pd/C (37 mg, 0.298 mmol) in EtOAc (15 mL), **98** was obtained as a purple crystalline solid. (212 mg, 99% yield); R_f 0.3 (1:1 Hex: EtOAc); mp 139-140 °C (lit. 138-140 °C)³

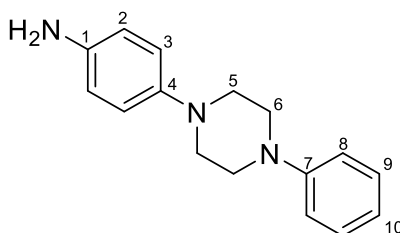
δ_{H} (400 MHz, CDCl_3): 7.32 (d, $J = 7.8$ Hz, 2H, H-2 and H-2'), 7.26 (d, $J = 7.8$ Hz, 2H, H-3 and H-3'), 7.21-7.19 (m, 1H, H-8), 6.87 (d, $J = 8.7$ Hz, 2H, H-6 and H-6'), 6.66 (d, $J = 8.7$ Hz, 2H, H-7 and H-7'), 3.58 (br. s, 2H, NH_2), 3.41 (br. s, 2H, CH_2), 2.71 (br. s, 2H, CH_2), 2.59-2.56 (m, 1H, H-CH), 1.97-1.94 (m, 4H, CH_2).

δ_{C} (151 MHz, CDCl_3): 146.2 (s, C-1), 145.1 (s, C-4), 140.0 (s, C-5), 128.4 (s, C-2), 126.8 (s, C-3), 126.2 (s, C-6), 119.2 (s, C-7), 116.1 (s, C-8), 52.4 (s, CH_2), 42.4 (s, CH_2), 33.6 (s, C-CH).

HRMS-ESI (+): m/z calculated for $\text{C}_{17}\text{H}_{20}\text{N}_2$: 253.1626: found: 253.1704 $[\text{M}+\text{H}]^+$

ν_{max} (ATR)/ cm^{-1} : 2839 (CH_2), 1439, 829, 752, 699, 652.

4-(4-Phenylpiperazin-1-yl)aniline (**99**)³



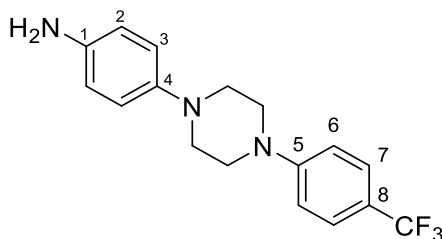
Following method F, from **96** (510 mg, 1.800 mmol) and 10% Pd/C (87 mg, 0.720 mmol) in EtOAc (25 mL), **99** was obtained as a dark-brown powder. (454 mg, 99% yield); R_f 0.3 (1:1 Hex: EtOAc); mp 146-150 °C (lit. 128-130 °C)³

δ_{H} (400 MHz, CDCl_3): 7.27 (d, $J = 8.8$ Hz, 2H, H-2 and H-2'), 6.97 (d, $J = 7.8$ Hz, 2H, H-6 and H-6'), 6.89 – 6.85 (m, 3H, H-7/H-7' and H-8), 6.66 (d, $J = 8.8$ Hz, 2H, H-3 and H-3'), 3.35 – 3.32 (m, 4H, 2x CH_2), 3.23-3.21 (m, 4H, 2x CH_2).

δ_{C} (151 MHz, CDCl_3): 151.2 (s, C-1), 144.0 (s, C-4), 140.6 (s, C-5), 129.1 (s, C-2), 120.0 (s, C-3), 119.0 (s, C-6), 116.3 (s, C-7), 116.2 (s, C-8), 51.3 (s, CH_2), 49.4 (s, CH_2).

HRMS-ESI (+): m/z calculated for $\text{C}_{16}\text{H}_{19}\text{N}_3$: 254.1579: found: 254.1652 $[\text{M}+\text{H}]^+$

ν_{max} (ATR)/ cm^{-1} : 2885 (CH_2), 2834 (CH_2), 1481, 827.

4-{4-[4-(Trifluoromethyl)phenyl]piperazin-1-yl}aniline (100**)³**

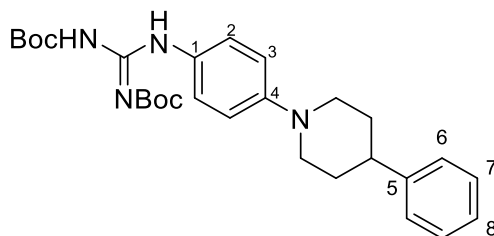
Following method F, from **97** (400 mg, 1.139 mmol) and 10% Pd/C (37 mg, 0.456 mmol) in EtOAc (25 mL), **100** was obtained as a dark-green powder. (332 mg, 99% yield); R_f 0.2 (1:1 Hex: EtOAc); mp 201-205 °C (lit. 203-205 °C)³

δ_H (400 MHz, $CDCl_3$): 7.53 (d, J = 8.7 Hz, 2H, H-2 and H-2'), 7.00 (d, J = 8.7 Hz, 2H, H-3 and H-3'), 6.90 (d, J = 8.5 Hz, 2H, H-6 and H-6'), 6.71 (d, J = 8.5 Hz, 2H, H-7 and H-7'), 3.40 – 3.37 (m, 4H, 2xCH₂), 3.18 – 3.15 (m, 4H, 2xCH₂).

δ_C (151 MHz, $CDCl_3$): 153.2 (s, C-1), 143.9 (s, C-4), 140.8 (s, C-5), 126.4 (s, C-2), 119.4 (s, C-3), 119.0 (s, C-6), 116.2 (s, C-7), 51.0 (s, CH₂), 48.3 (s, CH₂).

HRMS-ESI (+): m/z calculated for $C_{17}H_{18}F_3N_3$: 322.1453: found: 322.1461 [M+H]⁺

ν_{max} (ATR)/ cm^{-1} : 2822 (CH₂), 1514, 1323, 820.

***tert*-Butyl-*N*-[(*E*)-{[(*tert*-butoxy)carbonyl]imino}{[4-(4-phenylpiperidin-1-yl)phenyl]amino}) methyl]carbamate (**101a**)**

Following method B, from **98** (177 mg, 0.701 mmol), 1,3-bis-(*tert*-butoxycarbonyl)-2-methyl-2-thiopseudourea (214 mg, 0.736 mmol, 1.05 equiv.), $HgCl_2$ (209 mg, 0.772 mmol, 1.1 equiv.) and NEt_3 (0.2 mL, 1.403 mmol, 2 equiv.) in CH_2Cl_2 (3 mL), **101a** was obtained as a white solid. (189 mg, 54% yield); R_f 0.6 (7:3 Hex: EtOAc); mp 211-215 °C

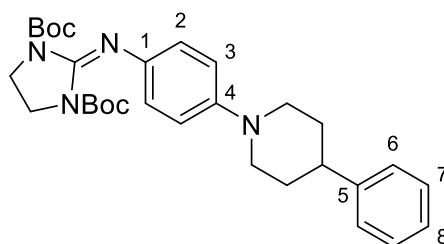
δ_{H} (400 MHz, CDCl_3): 11.71 (s, 1H, NH), 10.20 (s, 1H, NH), 7.49 (d, $J = 8.9$ Hz, 2H, H-2 and H-2'), 7.34 (d, $J = 8.1$ Hz, 2H, H-6 and H-6'), 7.29 – 7.20 (m, 3H, H-7/H-7' and H-8), 6.95 (d, $J = 8.9$ Hz, 2H, H-3 and H-3'), 3.76 (br. s, 2H, CH_2), 2.81 (br. s, 2H, CH_2), 2.67 – 2.64 (m, 1H, CH), 1.95 – 1.92 (m, 4H, $2\times\text{CH}_2$), 1.54 (br. s, 18H, $2\times\text{C}(\text{CH}_3)_3$).

δ_{C} (151 MHz, CDCl_3): 163.7 (s, CN), 153.3 (s, CO), 150.2 (s, C-1), 148.5 (s, C-4), 129.3 (s, C-2), 123.4 (s, C-3), 120.0 (s, C-6), 116.9 (s, C-7), 116.6 (s, C-8), 83.5 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 78.8 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 49.8 (s, CH_2), 29.3 (s, $\text{C}(\underline{\text{C}}\text{H}_3)_3$).

HRMS-ESI (+): m/z calculated for $\text{C}_{28}\text{H}_{38}\text{N}_4\text{O}_4$: 495.2893: Found: 495.2904 $[\text{M}+\text{H}]^+$

ν_{max} (ATR)/ cm^{-1} : 2970 (CH_2), 1631 (CO), 1411, 1372, 808.

1,3-Di-*tert*-butyl-2-[[4-(4-phenylpiperidin-1-yl)phenyl]imino]imidazolidine-1,3-dicarboxylate (101b**)**



Following method F, from **98** (95 mg, 0.375 mmol), thione (113 mg, 0.375 mmol, 1 equiv.), HgCl_2 (102 mg, 0.375 mmol, 1 equiv.) and NEt_3 (0.05 mL, 0.375 mmol, 1 equiv.) in CH_2Cl_2 (3 mL), **101b** was obtained as a light-yellow powder. (42 mg, 422% yield); R_f 0.9 (1:1 EtOAc: Hex); mp 172-176 °C.

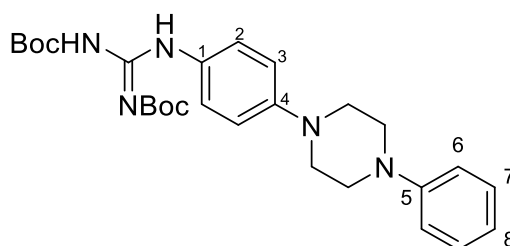
δ_{H} (400 MHz, CDCl_3): 7.34 (d, $J = 8.1$ Hz, 2H, H-6 and H-6'), 7.25 – 7.17 (m, 3H, H-7/H-7' and H-8), 6.98 (d, $J = 8.9$ Hz, 2H, H-2 and H-2'), 6.92 (d, $J = 8.9$ Hz, 2H, H-3 and H-3'), 3.69 (br. s, 2H, CH_2), 3.84 (s, 4H, CH_2), 2.78-7.73 (m, 2H, CH_2), 2.69-2.63 (m, 1H, CH), 1.98-1.92 (m, 4H, CH_2), 1.36 (s, 18H, $\text{C}(\text{CH}_3)_3$).

δ_{C} (151 MHz, CDCl_3): 150.4 (s, C-1), 147.9 (s, C-4), 146.2 (s, C-5), 128.4 (s, C-2), 126.8 (s, C-3), 126.1 (s, C-6), 122.2 (s, C-7), 117.8 (s, C-8), 82.5 (s, $\underline{\text{C}}(\text{CH}_3)_3$), 51.8 (s, CH_2), 43.0 (s, CH_2), 42.5 (s, CH_2), 33.3 (s, CH_2), 27.9 (s, $\text{C}(\underline{\text{C}}\text{H}_3)_3$).

HRMS-ESI (+): m/z calculated for $\text{C}_{30}\text{H}_{40}\text{N}_4\text{O}_4$: 521.3050: Found: 521.3125 $[\text{M}+\text{H}]^+$

$\nu_{\max}$ (ATR)/ cm^{-1} : 2979 (CH_2), 2817 (CH_2), 1689 (CO), 1366, 835.

***tert*-Butyl-*N*-[(*E*)-{[(*tert*-butoxy)carbonyl]imino}{[4-(4-phenylpiperazin-1-yl)phenyl]amino}]methyl]carbamate (**102a**)**



Following method F, from **99** (200 mg, 0.789 mmol), 1,3-bis-(*tert*-butoxycarbonyl)-2-methyl-2-thiopseudourea (241 mg, 0.829 mmol, 1.05 equiv.), HgCl_2 (236 mg, 0.868 mmol, 1.1 equiv.) and NEt_3 (0.22 mL, 1.579 mmol, 2 equiv.) in CH_2Cl_2 (6 mL), **102a** was obtained as an orange powder. (148 mg, 38% yield); R_f 0.5 (8:2 Hex: EtOAc); mp >230 °C

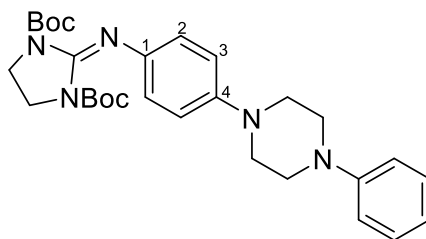
δ_{H} (400 MHz, CDCl_3): 11.62 (s, 1H, NH), 10.17 (s, 1H, NH), 7.48 (d, 2H, $J = 9.0$ Hz, H-2 and H-2'), 7.28 (t, 2H, $J = 8.7$ Hz, H-7 and H-7'), 6.97 (d, $J = 8.7$ Hz, 2H, H-6 and H-6'), 6.93 (d, $J = 9.0$ Hz, 2H, H-3 and H-3'), 6.89 – 6.86 (m, 1H, H-8), 3.34 – 3.31 (m, 4H, $2 \times \text{CH}_2$), 3.30 – 3.27 (m, 4H, $2 \times \text{CH}_2$), 1.52 (s, 9H, $\text{C}(\text{CH}_3)_3$), 1.48 (s, 9H, $\text{C}(\text{CH}_3)_3$).

δ_{C} (150.6 MHz, CDCl_3): 163.7 (s, CN), 153.5 (s, CO), 153.3 (s, CO), 151.2 (s, C-1), 148.5 (s, C-4), 129.4 (s, C-7), 129.2 (s, C-2), 123.4 (s, C-3), 120.0 (s, C-6), 116.8 (s, C-7), 116.3 (s, C-8), 83.5 (s, $\text{C}(\text{CH}_3)_3$), 79.4 (s, $\text{C}(\text{CH}_3)_3$), 49.8 (s, CH_2), 49.4 (s, CH_2), 28.2 (s, $\text{C}(\text{CH}_3)_3$), 28.12 (s, $\text{C}(\text{CH}_3)_3$).

HRMS-ESI (+): m/z calculated for $\text{C}_{27}\text{H}_{37}\text{N}_5\text{O}_4$: 496.6240: Found: 496.2918 $[\text{M}+\text{H}]^+$

$\nu_{\max}$ (ATR)/ cm^{-1} : 2972 (CH_2), 2829 (CH_2), 1631 (CO), 1600 (CN), 1407, 807.

1,3-Di-*tert*-butyl-2-{[4-(4-phenylpiperazin-1-yl)phenyl]imino}imidazolidine-1,3-dicarboxylate (102b)



Following method F, from **99** (200 mg, 0.789 mmol), thione (239 mg, 0.789 mmol, 1 equiv.), HgCl₂ (236 mg, 0.868 mmol, 1.1 equiv.) and NEt₃ (0.22 mL, 1.579 mmol, 2 equiv.) in CH₂Cl₂ (6 mL), **102b** was obtained as a light-pink powder. (42 mg, 40% yield); R_f 0.5 (1:1 Hex: EtOAc); mp 184-186 °C

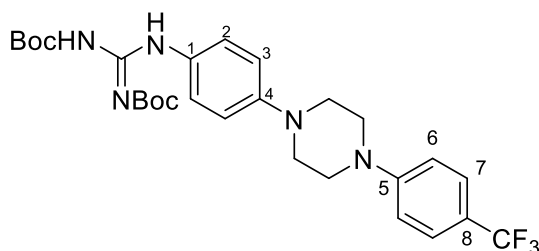
δ_{H} (400 MHz, CDCl₃): 7.31 (d, J = 8.0 Hz, 2H, H-2 and H-2'), 7.00 – 6.96 (m, 4H, H-H-3/H-3' and H-6/H-6'), 6.94 – 6.90 (m, 3H, H-7/H-7' and H-8), 3.84 (s, 4H, CH₂), 3.36 (br. s, 4H, CH₂), 3.26 (br. s, 4H, CH₂), 1.36 (s, 18H, 2xC(CH₃)₃).

δ_{C} (151 MHz, CDCl₃): 151.3 (s, C-1), 150.4 (s, C-4), 129.1 (s, C-5), 122.3 (s, C-6), 120.0 (s, C-7), 117.5 (s, C-8), 82.6 (s, C(CH₃)₃), 50.6 (s, CH₂), 49.4 (s, CH₂), 43.1 (s, CH₂), 30.9 (s, CH₂), 27.9 (s, C(CH₃)₃).

HRMS-ESI (+): m/z calculated for C₂₉H₃₉N₅O₄: 522.3002: Found: 522.3079 [M+H]⁺

ν_{max} (ATR)/ cm⁻¹: 2977 (CH₂), 2834 (CH₂), 1664 (CO), 1376, 841.

***tert*-Butyl-*N*-[(*E*)-{[(*tert*-butoxy)carbonyl]imino}][4-{4-[4-(trifluoromethyl)phenyl]piperazin-1-yl}phenyl]amino]methyl]carbamate (103a)**



Following method F, from **100** (150 mg, 0.467 mmol), 1,3-bis-(*tert*-butoxycarbonyl)-2-methyl-2-thiopseudourea (142 mg, 0.490 mmol, 1.05 equiv.), HgCl₂ (139 mg, 0.514 mmol, 1.1 equiv.) and NEt₃ (0.13 mL, 0.934 mmol, 2 equiv.) in CH₂Cl₂ (4 mL), **103a** was obtained as a white crystalline solid. (174 mg, 66% yield); R_f 0.4 (8:2 Hex: EtOAc). mp >230 °C

δ_H (400 MHz, CDCl₃): 11.67 (s, 1H, NH), 10.22 (s, 1H, NH), 7.54 (d, J = 9.0 Hz, 2H, H-2 and H-2'), 7.52 (d, J = 9.0 Hz, 2H, H-3 and H-3'), 7.00 (d, J = 8.8 Hz, 2H, H-6 and H-6'), 6.96 (d, J = 8.8 Hz, 2H, H-7 and H-7'), 3.45 (br. s, 4H, 2xCH₂), 3.37-3.28 (m, 4H, 2xCH₂), 1.59 (s, 9H, C(CH₃)₃), 1.56 (s, 9H, C(CH₃)₃).

δ_C (151 MHz, CDCl₃): 164.2 (s, CN), 155.1 (s, CO), 153.5 (s, C-1), 126.4 (s, C-2), 123.5 (s, C-3), 119.3 (s, C-6) 114.8 (s, C-7), 48.1 (s, 2xCH₂), 28.2 (s, 2xCH₂), 28.1 (s, C(CH₃)₃).

HRMS-ESI (+): m/z calculated for C₂₈H₃₆F₃N₅O₄: 564.2719: Found: 564.2792 [M+H]⁺

$\nu_{\max}$ (ATR)/ cm⁻¹: 2986 (CH₂), 2842 (CH₂), 1612 (CO), 1409, 1391, 821, 772.

6.1.4. HPLC purity assessment of final compounds

HPLC purity analysis was conducted using a Varian ProStar system equipped with a Varian Prostar 335 diode array detector and a manual injector (20 μ L). For purity assessment, UV detection at 245 nm was performed and peak purity was confirmed using a purity channel. The stationary phase consisted of an ACE 5 C18-AR column (150 $\times$ 4.6 mm), and the mobile phase used the following gradient system, eluting at 1 mL/min: aqueous ammonium formate buffer (30mM, 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%.

6.2. Computational methods

6.2.1. DFT calculations

Density Functional Theory (DFT) was used to optimise the structures of all the compounds under investigation. This was performed using the Gaussian16 package⁴ at the B3LYP⁵ or M06-2X⁶ computational level with the 6-311++G** basis set⁷ and a SMD⁸ or PCM⁹ solvation model. Frequency calculations have been performed at the same computational level to confirm that the resulting optimized structures are energetic minima.

6.2.2. Molecular Docking

Receptor preparation: The pdb file for specific macromolecule was downloaded from the RSCB PDB databank¹⁰. The file was then imported into the AutoDock Vina software¹¹. Polar hydrogen atoms were added and the coordinates of the binding site were determined using the AutoGrid function and the ligand was removed. The Gasteiger charges were added and the macromolecule was saved as a pdbqt file.

Ligand preparation: The optimised structure was imported into GaussView where it was converted to pdb format. The pdb structure was imported to Autodock Vina where its torsions were set. The ligand structure was then converted to pdbqt format which was used for the docking experiment.

Docking: After receptor and ligand preparation, the configuration file was prepared in txt format containing the correct coordinates. The docking experiment was then run through the command terminal using the Autodock Vina software. Upon completion of the calculation, output files were generated containing binding poses which are ranked based on their theoretical binding affinity in the form of Gscores (Kcal mol⁻¹).

Visualisation: Visualisation of the best binding poses was conducted using PyMoL¹² and VMD¹³.

6.2.3. Ligand based pharmacophore screening

The ligand based pharmacophore screen was performed using the Schrodinger Maestro Molecular Modelling software.¹⁴

3D pharmacophore construction: The 3D optimised structures of **105**, **106** and **109** were imported into the Maestro workspace in PDB format. With the use of the LigPrep tool, hydrogen atoms were added, physiological pH was set (pH = 7) and the specified chiralities were retained. The ligand set was then converted to maestro format. For the pharmacophore hypothesis development, the match percentage was set to 15% with a maximum of 7 features where each feature has a tolerance sphere of 2.0 Å. The scoring function was set to the default phase hypo score, where hypothesis with the highest score was deemed the best fit.

Pharmacophore screen: The 3D structure of each FDA approved anti-tubercular agent from the ChEMBL and ZINC databases were prepared with the use of the LigPrep tool. These ligands were then gathered into a phase database prior to screening. Using the phase ligand screening tool, the best fit pharmacophore hypothesis was screened against the constructed phase database where positive hits were allocated a specific phase screen score.

6.3. Biophysical experiments

6.3.1. UV-thermal denaturation

Thermal melting experiments were conducted with a Varian Cary 300 Bio spectrophotometer equipped with a 6×6 multicell temperature-controlled block. Temperatures were monitored with a thermistor inserted into a 1 cm quartz cuvette containing the same volume of water as in the sample cells. Absorbance changes at 260 nm were monitored from a range of 30 °C to 90 °C with a heating rate of 1 °C min⁻¹ and a data collection rate of five points per °C. The salmon sperm DNA and poly (dA-dT)₂ were purchased from Sigma Aldrich (extinction coefficient ϵ_{260} =6600 M⁻¹ cm⁻¹ for both oligomers). A quartz cell with a 1 cm path length was filled with a 1 mL solution of DNA polymer or DNA-compound complex. For thermal denaturation using salmon sperm DNA, the DNA polymer (150 µM base) and the compound

solution (15 μM) were prepared in a phosphate buffer ([10 mM $\text{K}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$], adjusted to pH 7) so that a compound to DNA base ratio of 0.1 was obtained. For thermal denaturation using poly (dA-dT)₂ (extinction coefficient $\epsilon_{260}=6600 \text{ cm}^{-1}\text{M}^{-1}$ base), the DNA polymer (50 μM base) and the compound solution (5 μM) were prepared in a phosphate buffer (10 mM) so that a compound to DNA base ratio of 0.1 was obtained. The thermal melting temperatures of the duplex or duplex-compound complex obtained from the first derivative of the melting curves are reported.

6.3.2. Circular Dichroism

CD spectra were collected with a JASCO J-800 spectrometer at different ratios of compound to DNA polymer (stDNA) at 25 °C in phosphate buffer at pH 7.0. Titrations were carried out by adding the compound (1.5 μM to 15 μM over 10 additions) in a DNA polymer solution (stDNA: 150 μM) in a 1 cm quartz cuvette and scanned over a desired wavelength range. Data points for stDNA were taken at an interval of 0.5 nm, scanning at 50 nm min⁻¹. Data are an average of 3 accumulations. The data were smoothed using an FFT filter with a cut-off frequency of 0.25.

6.3.3. UV-Vis titrations

All UV-Vis titration measurements were conducted on a Cary 300 UV spectrophotometer. A quartz cell with a 1 cm path length was used for all absorbance studies. Compound stock solutions were 5.3 μM and contained 10 mM phosphate buffer (pH 7). The DNA (poly (dA-dT)₂ oligonucleotide) at increasing ratios (5.3 μM to 15 μM over 10 additions) was then titrated into the compound solution and the corresponding absorption spectra were recorded under the same conditions. All concentrations were determined using the appropriate extinction coefficients. Due to high degree of spectral overlap, subtraction plots of the UV titration of each compound with increasing aliquots of poly (dA-dT)₂ were obtained. This allowed for the calculation of the binding constants ($\times 10^5, \text{M}^{-1}$) using Bensi-Hildebrand plot analysis.

6.4. Biochemical & Biological methods

6.4.1. Resazurin Microtiter Plate Assay (REMA)

A stock of *M. bovis* BCG or *M. smegmatis* was inoculated and incubated at 37 °C for 14 days or 3 days respectively. This cloudy suspension was then subcultured and further diluted until an optical density (OD₆₀₀) ~0.1 was reached. This concentration of cells was used to conduct the assay. 100 µl of 7H9 or LB broth media was first added to each well of a 96-well plate. For the test wells, 100 µl of the test compound solution was added to the top three wells of each replicate. A 1:2 serial dilution was then carried out over 6 wells. 100 µl of bacteria (OD₆₀₀ ~0.1) was then added to each test well to make a final volume of 200 µl in each well. For the positive control wells, 100 µl of bacteria was added to each well. For the negative control wells, a further 100 µl of media and 100 µl of bacteria was added to each well, again giving a final volume of 200 µl. Sterile water was then added to the empty surrounding wells to prevent evaporation. Additionally the plate was wrapped in parafilm and incubated for at 37°C for 1 week in the case of *M. bovis* BCG and 2 days for *M. smegmatis*. 20 µl of resazurin (2.5 g/L) was added to each well containing any of the described samples and further Incubated at 37°C for 1-2 days. Using a plate reader, the fluorescence of the each well was measured at 590 nm. Using the controls as a measure of percentage growth the cell viability over the respective drug concentration range was recorded and graphed using Graphpad Prism5. Using the "Gompertz equation for MIC determination" the MIC of each compound was determined.

6.4.2. Drug resistant mutant generation

6.4.2.1. Liquid media continual passaging

A culture of *M. smegmatis* was adjusted to an optical density of 0.1. 100 µl of LB broth media was added to each well of a 96-well plate. In the first row of wells across 6 columns was added 100 µl of **105** stock solution. 1:2 serial dilutions were then carried out over six rows. 100 µl of *M. smegmatis* (OD₆₀₀ ~0.1) was then added to each test well to make a final volume of 200 µl in each well. Sterile water was then added to the empty surrounding wells to prevent evaporation. Additionally the plate was wrapped in parafilm and incubated for at 37°C for 2

days. To three of the six test columns was added 20 µl of resazurin (2.5 g/L) and was left to incubate for an additional 24 hours. Bacteria at the visual MIC point (lowest concentration of drug that prevented a colour change) from the non resazurin containing wells were inoculated in 5 mL of LB broth and allowed to grow for 5 days. REMA analysis with this culture was then conducted to any determine changes in the MIC. This procedure was repeated with increasing stock concentrations of **105** over 7 passages.

6.4.2.2. Solid media single shot experiment

A culture of *M. smegmatis* was adjusted to an optical density (OD₆₀₀) of 0.1 and a volume of 100 µl (~3x10⁶ CFU) was spread on LB agar containing 180 µM (3xMIC), 240 µM (4xMIC), 300 µM (5xMIC), 420 µM (7xMIC), 600 µM (10xMIC) and 1200 µM (20xMIC). Each plate was left to incubate for 3 days at 37°C. Colonies appearing on these plates were marked as possible resistant mutants only if they grew again on LB agar broth containing the same concentration of **105**.

6.4.3. Whole genome sequencing

DNA extraction and sequencing was performed by the group of Prof. Stephen Gordon from University College Dublin. *M. smegmatis* cells were lysed using a bead beater (Magna Lyser, Roche) and the DNA was extracted by sodium acetate/ethanol precipitation. Crude DNA extracts were then purified with Ampure XP magnetic purification beads (Beckman Coulter, UK) and resuspended in Tris-EDTA buffer (TE; Fisher Scientific, UK). DNA extracts were assessed for DNA quantity and quality using a Nanodrop spectrophotometer (NDW1000). DNA content was quantified with a Qubit dsDNA BR Assay kit (Invitrogen) following the manufacturer's instructions. DNA extracts were stored at -20oC pending library preparation and MIRU-VNTR analyses. The library preparation of each sample used 50 ng of DNA. Library preparations were carried out using Nextera DNA Flex kits (Illumina, USA) as per the manufacturer's protocol with a minor modification of using PCR strip-tubes instead of 96-well plates. Library DNA content was quantified using a Qubit dsDNA HS Assay kit (Invitrogen) Assay Kit, and fragment sizes were checked using the DNA chip system on a Bioanalyser 2100 (Agilent) prior to pooling. The standardised pool was adjusted to a recommended concentration (2nM) before denaturing and sequencing. Sequencing was carried out on a

NextSeq500 machine (Illumina, USA) at the Genomics Core, UCD Conway Institute, generating 151bp paired end reads.

6.4.4. Cytotoxicity analysis against THP-1 cells

Cytotoxicity analysis of compounds **105**, **106**, **109**, **66a**, **68a**, **69a**, **71a**, **72a** and **73a** as described in Chapter 4.6 was conducted by the group of Prof. Lydia Tabernero from the University of Manchester. THP-1 cells (8×10^3 cells/well) were incubated overnight with PMA (100ng/ml) in a 96 well plate. After incubation, cells were washed and incubated with the test compounds for 4h at a concentration of 1, 10, 50 and 100 μ M. Controls were incubated with 0.1% DMSO. After 24h, cells were washed and incubated with fresh compounds and 48h after (72h total), cells were washed and incubated with MTT stock (5mg/ml) for 2h. Supernatants were removed and DMSO-glycine buffer (8:1) was added to produce a colour change. Finally, plates were read for absorbance at 570 nm.

6.4.5. *In vitro* anti-parasitic analysis of novel DNA Minor Groove binders

These experiments were carried out by Marcel Kaiser at Prof. Reto Brun's laboratory in the Swiss Tropical and Public Health Institute (Basel, Switzerland). The cell lines used were *Trypanosoma brucei rhodesiense* (*T.b.r*) STIB 900, *Plasmodium falciparum* (*P. falciparum*) NF54 and *Leishmania donovani* (*L. don.*) axenic amastigotes and L6 cells.

Activity against Trypanosoma brucei rhodesiense STIB900: This stock was isolated in 1982 from a human patient in Tanzania and after several mouse passages cloned and adapted to axenic culture conditions.¹⁵ Minimum Essential Medium (50 μ L) supplemented with 25 mM HEPES, 1 g/L additional glucose, 1% MEM non-essential amino acids (100x), 0.2 mM 2-mercaptoethanol, 1 mM Na-pyruvate and 15% heat inactivated horse serum was added to each well of a 96-well microtiter plate. Serial drug dilutions of eleven 3-fold dilution steps covering a range from 100 to 0.002 μ g mL⁻¹ were prepared. Then 4×10^3 bloodstream forms of *T. b. rhodesiense* STIB 900 in 50 μ L was added to each well and the plate incubated at 37 °C under a 5% CO₂ atmosphere for 70 hours. 10 μ L alamarBlue (resazurin, 12.5 mg in 100 ml double-distilled water) was then added to each well and incubation continued for a further 2–4 hours.¹⁶ Then the plates were read with a Spectramax Gemini XS microplate fluorometer (Molecular Devices Cooperation, Sunnyvale, CA, USA) using an excitation wave length of 536

nm and an emission wave length of 588 nm. The IC₅₀ values were calculated by linear regression¹⁷ from the sigmoidal dose inhibition curves using SoftmaxPro software (Molecular Devices Cooperation, Sunnyvale, CA, USA).

Activity against L. donovani axenic amastigotes: Amastigotes of *L. donovani* strain MHOM/ET/67/L82 are grown in axenic culture at 37 °C in SM medium¹⁸ at pH 5.4 supplemented with 10% heat-inactivated fetal bovine serum under an atmosphere of 5% CO₂ in air. One hundred microlitres of culture medium with 10⁵ amastigotes from axenic culture with or without a serial drug dilution are seeded in 96-well microtitre plates. Serial drug dilutions of eleven 3-fold dilution steps covering a range from 100 to 0.002 µg/ml are prepared. After 70 h of incubation the plates are inspected under an inverted microscope to assure growth of the controls and sterile conditions. 10 µl of resazurin (12.5 mg resazurin dissolved in 100 ml distilled water) are then added to each well and the plates incubated for another 2 h. Then the plates are read with a Spectramax Gemini XS microplate fluorometer (Molecular Devices Cooperation, Sunnyvale, CA, USA) using an excitation wave length of 536 nm and an emission wave length of 588 nm. From the sigmoidal inhibition curves the IC₅₀ values are calculated by linear regression¹⁷ and 4-parameter logistic regression using SoftmaxPro software (Molecular Devices Cooperation, Sunnyvale, CA, USA).

Activity against P. falciparum: *In vitro* activity against erythrocytic stages of *P. falciparum* was determined using a ³H-hypoxanthine incorporation assay, using the drug sensitive NF54 strain (Schiphol Airport, The Netherlands) and the standard drug chloroquine (Sigma C6628). Compounds were dissolved in DMSO at 10 mg mL⁻¹ and added to parasite cultures incubated in RPMI 1640 medium without hypoxanthine, supplemented with HEPES (5.94 g/l), NaHCO₃ (2.1 g L⁻¹), neomycin (100 U mL⁻¹), Albumax^R (5 g L⁻¹) and washed human red cells A+ at 2.5% haematocrit (0.3% parasitaemia). Serial drug dilutions of eleven 3-fold dilution steps covering a range from 100 to 0.002 µg mL⁻¹ were prepared. The 96-well plates were incubated in a humidified atmosphere at 37 °C; 4% CO₂, 3% O₂, 93% N₂. After 48 hours 50 µL of ³H-hypoxanthine (=0.5 µCi) was added to each well of the plate. The plates were incubated for a further 24 hours under the same conditions. The plates were then harvested with a Betaplate™ cell harvester (Wallac, Zurich, Switzerland), and the red blood cells transferred onto a glass fibre filter then washed with distilled water. The dried filters were inserted into a plastic foil with 10 mL of scintillation fluid, and counted in a Betaplate™ liquid scintillation

counter (Wallac, Zurich, Switzerland). IC₅₀ values were calculated from sigmoidal inhibition curves by linear regression¹⁷ using Microsoft Excel.

In vitro cytotoxicity with L-6 cells: Assays were performed in 96-well microtiter plates, each well containing 100 µL of RPMI 1640 medium supplemented with 1% L-glutamine (200 mM) and 10% fetal bovine serum, and 4000 L6 cells (a primary cell line derived from rat skeletal myoblasts).^{19,20} Serial drug dilutions of eleven 3-fold dilution steps covering a range from 100 to 0.002 µg mL⁻¹ were prepared. After 70 hours of incubation the plates were inspected under an inverted microscope to assure growth of the controls and sterile conditions. 10 µL of alamarBlue was then added to each well and the plates incubated for another 2 hours. Then the plates were read with a Spectramax Gemini XS microplate fluorometer (Molecular Devices Cooperation, Sunnyvale, CA, USA) using an excitation wave length of 536 nm and an emission wave length of 588 nm. The IC₅₀ values were calculated by linear regression¹⁷ from the sigmoidal dose inhibition curves using SoftmaxPro software (Molecular Devices Cooperation, Sunnyvale, CA, USA). Podophyllotoxin was the reference drug used.

6.5. References

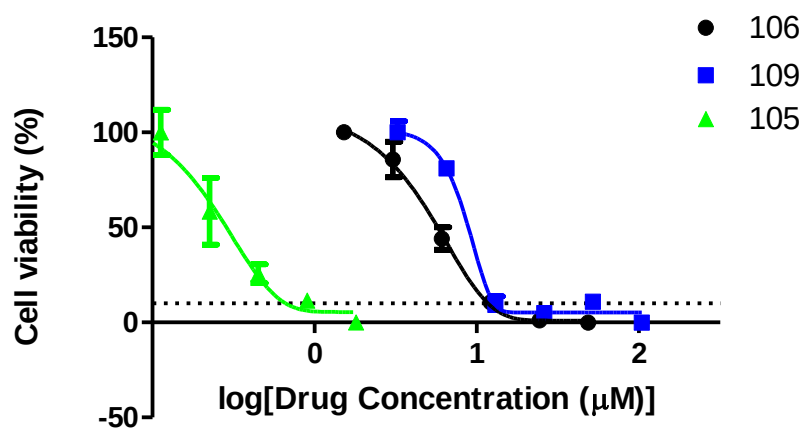
1. Nagle, P. S.; Rodriguez, F.; Kahvedžić, A.; Quinn, S. J.; Rozas, I.; *J. Med. Chem.*, **2009**, *52*, 7113–7121.
2. Dardonville, C.; Brun, R.; *J. Med. Chem.*, **2004**, *47*, 2296-2307.
3. Wodtke, R.; Steinberg, J.; Köckerling, M.; Löser, R.; Mamat, C.; *RSC Adv.* **2018**, *8*, 40921–40933.
4. Gaussian 16, Revision C.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian, Inc., Wallingford CT, **2016**.
5. A. D. Becke, *J. Chem. Phys.*, **1993**, *98*, 1372–1377.
6. Zhao, Y.; Truhlar, D. G., *Theor. Chem. Acc.* **2008**, *120*, 215–241.
7. Hehre, W. J.; Ditchfield, R.; Pople, J. A., *J. Chem. Phys.*, **1972**, *56*, 2257–2262.
8. Marenich, A. V.; Cramer, C. J.; Truhlar, D. G., *J. Phys. Chem. B.*, **2009**, *113*, 6378–6396.
9. Tomasi, J.; Mennucci, B.; Cammi, R., *Chem. Rev.*, **2005**, *105*, 2999–3093.
10. <https://www.rcsb.org/>
11. Trott, O and Olson, A. J., *Journal of Computational Chemistry*, **2010**, *31*, 455-461
12. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.
13. Humphrey, W.; Dalke, A.; Schulten, K., "VMD - Visual Molecular Dynamics", *J. Molec. Graphics*, **1996**, *14*, 33-38.
14. Schrödinger Release 2021-3: Maestro, Schrödinger, LLC, New York, NY, 2021.
15. Baltz, T.; Baltz, D.; Giroud C. and Crockett, J. *EMBO. J.*, **1985**, *4*, 1273.
16. Rätz, B.; Iten, M.; Grether-Buhler, Y.; Kaminsky R.; Brun, R., *Acta Trop.*, **1997**, *68*, 139.

17. Huber, W.; Koella, J. C., *Acta Trop.*, **1993**, 55, 257.
18. Cunningham, I., *J Protozool*, **1997**, 24, 325-329
19. Page B.; Page M. ; Noel, C., *Int. J. Oncol.*, **1993**, 3, 473.
20. Ahmed, S.A.; Gogal R.M.; Walsh, J.E., *J. Immunol. Methods*, **1994**, 170, 211.

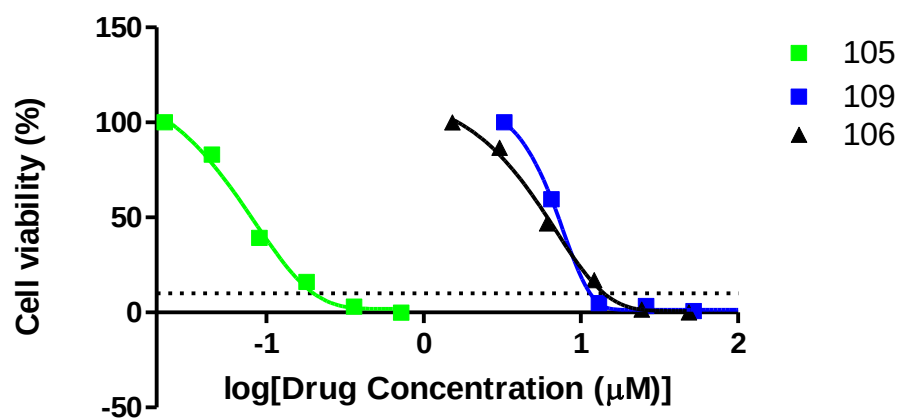
Appendix

Comparison of the anti-mycobacterial activity of the TFA and HCl salts of previously made in-house family of compounds

M. bovis BCG Tokyo



M. bovis BCG Denmark



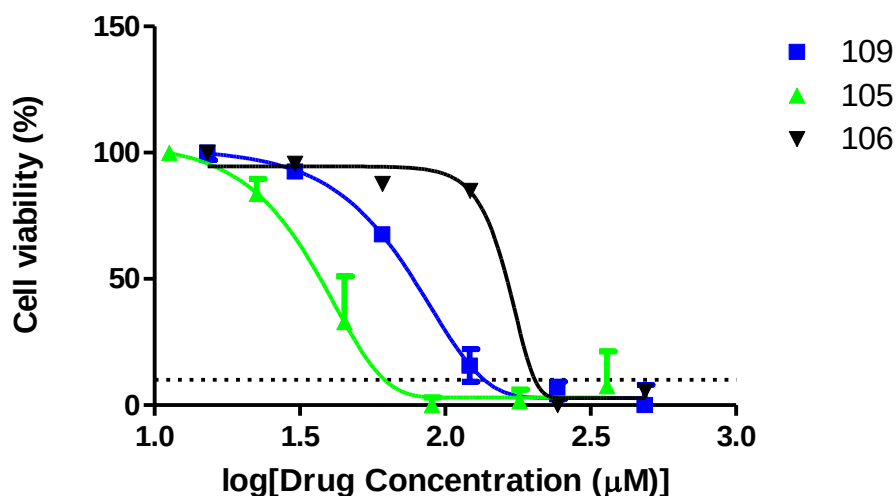
M. smegmatis

Figure 1. Cell viability curves generated from REMA analysis of HCl salts of **105**, **109** and **106** against *M. bovis* BCG Tokyo, *M. bovis* BCG Denamark and *M. smegmatis*.

Table 1. Minimum inhibitory concentrations (MIC) of TFA and HCl salts of previously made in-house family of compounds against mycobacterial strains.

Cmp	<i>M. bovis</i> BCG Tokyo MIC (μM)	<i>M. bovis</i> BCG Den. MIC (μM)	<i>M. smeg.</i> MIC (μM)
106 (HCl)	12.33 ±0.2	15.36 ±0.1	205.2 ±0.9
106 (TFA)	12.5 ±0.185	15.33 ±0.97	205.9 ±0.4
105 (HCl)	0.58 ±0.5	0.1933 ±0.11	62.3 ±8
105 (TFA)	1 ±0.5	0.38 ±0.27	60.7 ±9.7
109 (HCl)	12.44 ±2.3	11.76 ±3.4	138.3 ±10.2
109 (TFA)	12.65 ±2.3	13.3 ±6.2	142.6 ±20.8

Cells were used at an OD₆₀₀ of 0.1 (~3 × 10⁷ CFU/ mL) in a 96-well plate and treated with a serial dilution of drug compound dissolved in H₂O or DMSO. Once treated, cells were incubated for 3 days for *M.smeg* or 5 days for BCG strains at 37 °C then treated with resazurin and left to incubate for 24 or 48 h respectively. The resulting fluorescence was read using a plate reader from which percentage viability was calculated. MIC was defined by the drug concentration at 10% cell viability and was calculated using the "Gompertz equation for MIC determination" through GraphPad

Whole genome sequencing background mutations data

Table 2. *M. smegmatis* background variants identified from **105** continual passaging and solid media experiments

Position	SNP/INDEL	Gene	Protein	Mutation	Type
4160153	A>G	MSMEI_RS20020	LLM class Flavin dependent oxidoreductase	427F	syn
4031782	T>C	MSMEI_RS19450	GntR family transcriptional reg.	211V	syn
3264632	C>G	treY	Malto-oligosyltrehalose syn.	259G	syn
3107182 3107175	G>A	MSMEI_RS14930	3-dehydroquinate synthase	284G>284D 282G>282R	mis
2825972	A>G	MSMEI_RS13540	YihY/virulence factor BrBkB	251T	syn
2118639	T>G	MSMEI_RS100600	NAD(P)/FAD-dependent oxidoreductase	316F>316V	mis
1666994	G>T	MSMEI_RSO7790	Glutamate decarboxylase	370A>370S	mis
1196882	T>C	MSMEI_RS05605	Polyprenyl synthetase	145F>145L	mis
820242	G>A	ClpB	ATP-dependent chaperone	331E>331K	mis
4215932	G>T	MSMEI_RS20275	IS1380 family transposase	163A>163S	mis
4215937	C>A			164G	syn
4215958	A>C			171G	syn
4216171	C>A			242R	Syn
4216183	C>A			246A	syn
225872	A>AC	MSMEI_RS01000	Fructosamine kinase	Insertion	Frameshift
713730	C>CT	MSMEI_RS03115	Sulfatase	Insertion	Frameshift

860163	AC>ACC	MSMEI_RS03830	LLM class-Flavin dependent oxidoreductase	Insertion	Frameshift
1903851	GC>GCC	MSMEI_RS09025	GDP-mannose pyrophosphorylase	deletion	Frameshift
2110478	TGG>TG	MSMEI_RS10020	SDR family oxidoreductase	deletion	Frameshift
2128071	GAA>GA	MSMEI_RS10110	Enoyl-CoA hydratase	deletion	Frameshift
2427116	CCACAC>CCAC	MSMEI_RS11525	Lycopene cyclase	deletion	Frameshift
2450462	GAA>GA	MSMEI_RS11635	6-phosphofructokinase	deletion	Frameshift
2597300	AG>AGG	MSMEI_RS12350	DNA protecting protein DprA	Insertion	Frameshift
2602233	TG>TGG	MSMEI_RS12370	HNH endonuclease	Insertion	Frameshift
2818251	TCCCC>TCCCC	MSMEI_RS13500	PAC2 family protein	deletion	Frameshift
3034240	CGGGGG>CGGGGGG	MSMEI_RS14605	Flavin reductase family protein	Insertion	Frameshift
3054639	GT>GTT	MSMEI_RS14705	Integral membrane protein	Insertion	Frameshift
3487487	AGG>AG	MSMEI_RS16815	MerR family transcriptional regulator	deletion	Frameshift
3813499	CG>CGG	MSMEI_RS18420	Channel forming protein	Insertion	Frameshift
3942474	GAA>GA	MSMEI_RS35230	MoxR family ATPase	deletion	Frameshift
4023519	GAA>GA	MSMEI_RS19420	3-oxoacyl-ACP synthase	deletion	Frameshift
4031824	CTT>CT	MSMEI_RS19450	Gntr family transcriptional regulator	deletion	Frameshift

4043999	GCC>GC	MSMEI_RS19510	Hydantoinase	deletion	Frameshift
4054087	GAAA>GAA	MSMEI_RS19565	CoA transferase	deletion	Frameshift
4134485	TGGG>TGG	MSMEI_RS19915	Aspartate aminotransferase	deletion	Frameshift
4299364	CT>CTT	MSMEI_RS20660	Cytochrome P450	Insertion	Inframe deletion
4717190	ATT>ATTT ACC>ACCC	MSMEI_RS22635	Dihydrofolate synthase	Insertion Insertion	Frameshift
4717482	CAA>CA	MSMEI_RS22640	Valine tRNA ligase	deletion	Frameshift
5214694	AGG>AG	MSMEI_RS35040	VOC family protein	deletion	Frameshift
5495482	AC>ACC	MSMEI_RS35510	DUF501 domain-containing protein	Insertion	Frameshift
5500445	TGGGG>TGGG	MSMEI_RS26495	Peptidase M75 family protein	deletion	Frameshift
5684083	GT>GTT	MSMEI_RS27375	Rieske 2Fe-2S domain containing protein	Insertion	Frameshift
5875661	GCCCC>GCCCC	MSMEI_RS28415	ATP-binding cassette domain containing protein	deletion	Frameshift