

https://dspace.org/

# Evaluation of the NIH clinical collection to identify potential HIV-1 integrase inhibitors

**Abrahams, Shaakira**

**2014-09-09**
http://hdl.handle.net/10539/15448

Downloaded from a DSpace repository. https://dspace.org/

# Evaluation of the NIH Clinical Collection to identify potential HIV-1 integrase inhibitors

Shaakira Abrahams

A dissertation submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in fulfillment of the requirements for the degree of Master of Science in Medicine

January 2014

# *DECLARATION*

I, Shaakira Abrahams, declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Master of Science at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

_____

……day of ……, 2014

# *ABSTRACT*

HIV-1 integrase is an essential enzyme in the HIV replication cycle and is a validated target for antiretroviral drugs. Due to the inevitable emergence of drug resistance of HIV-1 strains to all currently approved FDA antiretroviral drugs, antivirals with new mechanisms of action are continuously investigated. As such, this study aimed to reposition existing drugs as HIV-1 integrase inhibitors by screening the NIH Clinical Collection compound library comprising 727 compounds. Recombinant integrase was expressed in bacterial cells, purified by nickel affinity chromatography, and used to set up a Scintillation Proximity Assay (SPA). The SPA was subsequently amended to an automated system to allow for rapid screening of compounds. The complete compound library was successfully screened using the newly established automated SPA. Overall, only two compounds were identified as HIV-1 IN inhibitors: cefixime trihydrate and a previously identified HIV integrase inhibitor, epigallocatechin gallate. These compounds exerted $IC_{50}$ values < 10μM in the automated SPA. Cefixime trihydrate was not toxic to mammalian cells ($CC_{50}$ > 200μM) while no appreciable antiretroviral activity was observed in *in vitro* phenotypic inhibition assays (23% inhibition of viral replication), thus concluding that this compound was non-selective. By contrast, epigallocatechin gallate was toxic to mammalian cells at the evaluated ranges ($CC_{50}$ = 23 + 1μM) and therefore could not be validated as an integrase inhibitor in *in vitro* phenotypic inhibition assays. Overall, this study resulted in the establishment of an automated SPA, the successful screening of 727 compounds, and the availability of a platform to expedite the future screening of potential HIV-1 integrase inhibitors.

# *CONFERENCE PROCEEDINGS*

Oral presentations of this study:

1. R. Hewer, A. Harrison, S. Abrahams, S. Mosebi, FH. Kriel, MA Papathanasopoulos, MQ Fish. Identification of novel inhibitors that target HIV-1 integration. Oral presented at 6th SA AIDS, Durban, RSA, 18-21 June 2013.

2. R.Hewer, A. Harrison, S. Abrahams, S. Mosebi, MQ Fish, FH Kriel, MA Papathanasopoulos. Novel technologies and useful inhibitors for HIV drug discovery. Oral presented at 3rd National Conference on Science and Technology, Luanda, Angola, 11 -13 September 2013.

# *ACKNOWLEDGEMENTS*

It is a pleasure to thank the following people who have contributed to this dissertation:

My supervisors, Dr. Raymond Hewer, Dr. Salerwe Mosebi and Prof. Maria Papathanasopoulos: I sincerely thank you for the opportunity, guidance and patience and for freely sharing your knowledge with me.
*If I have seen further than others, it is by standing on the shoulders of giants* - Sir Isaac Newton

The National Research Foundation (NRF) and GDARD for financial assistance throughout my studies and Mintek for funding my project.

My colleagues in the Biomed group, I appreciate your helpful discussions and assistance in the lab especially Muhammed Qasim Fish and Angela Harrison.

Thank you to my family for the constant support, love and encouragement. Mom, I cannot thank you enough. Your unconditional belief in me means everything.

# *TABLE OF CONTENTS*

DECLARATION ..... II

ABSTRACT ..... III

CONFERENCE PROCEEDINGS ..... IV

ACKNOWLEDGEMENTS ..... V

ABBREVIATIONS ..... XI

LIST OF FIGURES ..... XVIII

LIST OF TABLES ..... XXII

APPENDICES ..... XXIII

**CHAPTER 1: Introduction**

1.1. HIV/AIDS overview ..... 1

1.1.1. History of HIV/AIDS ..... 1

1.1.2. Epidemiology and phylogeny of HIV/AIDS ..... 1

1.1.3. The HIV-1 virion and its genomic organization ..... 3

1.1.4. The HIV lifecycle ..... 5

1.2. HIV treatment and drug discovery ..... 7

1.2.1. Current antiretroviral treatment ..... 7

1.2.2. Drug discovery and development ..... 9

1.2.2.1. Random screening approach ..... 10

1.2.2.2. Rational design approach ..... 11

1.2.2.3. High Throughput Screening and automation ..... 13

1.2.2.4. Drug repositioning ..... 14

1.2.2.4.1. The NIH Clinical Collection compound library ..... 15

1.2.3. HIV-1 integrase as a drug target ..... 16

1.2.3.1. The structure of HIV-1 integrase and its function ..... 16

1.2.3.2. The mode of action of the integrase strand transfer inhibitor, raltegravir ..... 19
1.2.3.3. Resistance mutations against raltegravir and other integrase inhibitors ..... 20
1.2.3.4. In vitro assays used to identify HIV integrase inhibitors ..... 21
1.2.3.5. HIV integrase assays amenable to automation ..... 21
1.2.3.6. Automation of an HIV integrase scintillation proximity assay ..... 22
1.2.4. Hypothesis ..... 23
1.2.5. Study aims and objectives ..... 23

## CHAPTER 2: Materials and Methods

2.1. Recombinant wild type HIV-1 subtype B integrase expression and purification ..... 25
2.1.1. HIV-1 Pol gene induction ..... 25
2.1.2. Bacterial cell lysis ..... 25
2.1.3. Nickel-Affinity column chromatography ..... 26
2.1.4. PD-10 sephadex column ..... 26
2.1.5. Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis ..... 26
2.1.6. Western blot analysis of the recombinant HIV-1 integrase ..... 27
2.2. Radiolabeling 3'-target DNA with $^{3}$H-dNTPs ..... 28
2.2.1. Lyophilizing $^{3}$H-dNTPs ..... 28
2.2.2. Klenow fragment 3'-target DNA labelling with 3H-dTTP and 3H-dCTP ..... 28
2.2.3. Purifying the radiolabeled target DNA ..... 29
2.3. Scintillation proximity assay (SPA) ..... 29
2.3.1. SPA optimization ..... 29
2.3.2. Preparation of the streptavidin coated SPA bead complex ..... 30
2.3.3. The optimized SPA ..... 30

2.3.4. Handling of the Hamilton STARlet Robotic System ..... 31
2.3.5. Amending the SPA onto the automated system ..... 31
2.3.6. Pooling of the NCC compound library ..... 31
2.3.7. Screening of the NIH Clinical Collection ..... 32
2.4. Cell-based assays ..... 32
2.4.1. Cultivation of mammalian cells ..... 32
2.4.2. Cytotoxicity assays ..... 33
2.4.3. HIV-1 phenotypic inhibition assay ..... 34
2.5. HIV-1 integrase ELISA ..... 34
2.6. Reverse transcriptase activity colorimetric assay ..... 35
2.7. Virtual screening of cefixime trihydrate and epigallocatechin gallate ..... 37
2.7.1. *In silico* evaluation of some ADME descriptors of cefixime trihydrate and epigallocatechin gallate ..... 37
2.7.2. Molecular modelling of cefixime trihydrate and epigallocatechin gallate ..... 37
2.8. Data analysis ..... 38

## CHAPTER 3: Results

3.1. Expression and purification of recombinant wild type HIV-1 subtype B integrase ..... 39
3.2. SPA screening ..... 42
3.2.1. Purification of the radiolabeled ($^{3}$H) target DNA ..... 42
3.2.2. Optimization of the HIV-1 integrase SPA ..... 43
3.2.3. Validating the optimized SPA with known integrase inhibitors ..... 46
3.2.4. HIV-1 integrase SPA amended to the automated system ..... 47
3.2.5. Screening of the NCC library ..... 48

3.2.6. Dose response studies of cefixime trihydrate and epigallocatechin gallate ..... 49
3.2.7. Integrase resistant mutation profiles of cefixime trihydrate and epigallocatechin gallate ..... 50
3.2.8. Screening of cefixime trihydrate derivatives ..... 51
3.2.9. Further development of identified HIT compounds ..... 52
3.2.9.1. Evaluating the cytotoxicity of epigallocatechin gallate, cefixime trihydrate and ampicillin ..... 52
3.2.9.2. Determining the effect of cefixime trihydrate, epigallocatechin gallate and ampicillin on reverse transcriptase activity ..... 53
3.2.10. Orthogonal screening of the identified HITS ..... 54
3.2.10.1. Validation of the cefixime trihydrate and epigallocatechin gallate through HIV-1 IN ELISA ..... 54
3.2.10.2. Further validation of cefixime trihydrate, epigallocatechin gallate and ampicillin through an HIV-1 phenotypic inhibition assay ..... 55
3.2.10.3. Determining the effect of cefixime trihydrate and epigallocatechin gallate in the presence of reducing agents on strand transfer inhibition ..... 57
3.3. Computational studies of cefixime trihydrate and epigallocatechin gallate ..... 58
3.3.1. Molecular docking of compounds with HIV-1 integrase ..... 58
3.3.1.1. Validating the docking protocol using control compounds ..... 58
3.3.1.2. Molecular docking of cefixime trihydrate and epigallocatechin gallate with HIV-1 integrase ..... 61
3.3.2. Evaluation of the absorption and distribution properties of cefixime trihydrate and epigallocatechin gallate ..... 63

## CHAPTER 4: Discussion

4.1. Successful expression and purification of HIV-1 subtype B integrase .. 65
4.2. The optimized SPA parameters ..... 66
4.3. HITS identified when screening the NCC compound library ..... 68

4.3.1. Cefixime trihydrate identified as a strand transfer inhibitor ..... 69

4.3.2. The possible binding mechanism of cefixime trihydrate..... 69

4.3.3. *In silico* studies predicted that cefixime trihydrate is not drug-like..... 71

4.3.4. Cefixime trihydrate was not active against integrase in orthogonal screening ..... 72

4.3.6. Epigallocatechin gallate is active against integrase and reverse transcriptase ..... 73

4.3.7. The galloyl moiety of epigallocatechin is involved in antiretroviral activity..... 73

4.3.8. Epigallocatechin gallate as a natural product is drug-like ..... 74

4.3.9. DTT influences the activity of epigallocatechin gallate and cefixime trihydrate..... 75

4.3.10. Epigallocatechin gallate is auto-oxidized at pH levels above 7 ..... 76

4.4. Conclusion..... 76

4.5. Future studies..... 77

**REFERENCES**

**APPENDIX A**

**APPENDIX B**

# *ABBREVIATIONS*

| | |
|---|---|
| °C | Degrees Celsius |
| •dimer | Dimer radical |
| •EGCG | EGCG radical |
| •O2 | Superoxide |
| 2D | Two dimensional |
| 3D | Three dimensional |
| $^{3}$H | Tritium/ tritiated |
| $^{3}$H-dCTP | deoxycytidine 5'- [5-$^{3}$H] triphosphate |
| $^{3}$H-dCTP | deoxythymidine 5'-[methyl-$^{3}$H] triphosphate |
| 7-ACA | 7-aminocephalosporanic acid |
| 7-ADCA | 7-aminodesacetoxycephalosporanic acid |
| α | Alpha |
| β | Beta |
| Å | Angstrom |
| µCi/ml | Microcurie per millilitre |
| µm | Micrometer |
| µM | Micromolar |
| µl | Microlitre |
| µg | Microgram |
| µg/ µl | Micrograms/Microlitre |
| aa | Amino acid |
| ABTS | 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid) |
| ADMET | Absorption, distribution, metabolism, excretion and toxicity |
| AIDS | Acquired immunodeficiency syndrome |

| | |
|---|---|
| Alpha-LISA | Amplified luminescent proximity homogenous assay -LISA |
| AMP | Ampicillin |
| anti-DIG-POD | anti-digoxigenin-peroxidase |
| APOBEC | antiviral host protein apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like |
| Asn | Asparagine |
| Asp | Aspartic acid |
| AZT | Azidothymidine |
| BBB | Blood brain barrier |
| B-Me | B-Mercaptoethanol |
| bp | Basepairs |
| BSA | Bovine serum albumin |
| Cat# | Catalogue number |
| $CC_{50}$ | Cytotoxic concentration 50 |
| CCD | Catalytic core domain |
| CDC | Centers for Disease Control |
| CEF | Cefixime trihydrate |
| CHAPS | 3-[(3-Cholamidopropyl)-dimethylammonio]-1-propanesulfonate |
| CIS | Cellular IN screening |
| CNS | Central nervous system |
| CPM | Counts per minute |
| CRF | Circulating recombinant form |
| CTD | C-terminal domain |
| CSS | Cephalothin Sodium Salt |
| CYP P40 | Cytochrome P40 |
| Cys | Cysteine |
| Da | Dalton |

| | |
|---|---|
| DC | Dendritic cells |
| $dH_2O$ | Distilled water |
| DKA | Diketo acid |
| DMSO | Dimethyl sulfoxide |
| DNA | Deoxyribonucleic acid |
| dNTP | Deoxynucleotide triphosphate |
| ds | Double stranded |
| DTG | Dolutegravir |
| DTP | Developmental Therapeutics Programme |
| DTT | Dithiothreitol |
| EDTA | Ethylenediaminetetraacetic acid |
| EGCG | Epigallocatechin gallate |
| ELISA | Enzyme-linked immunosorbant assay |
| ER | Endoplasmic reticulum |
| *E.coli* | *Eschericia coli* |
| EVG | Elvitegravir |
| $FCIC_{50}$ | Fold-change $IC_{50}$ |
| FCS | Fetal calf serum |
| FDA | Food and Drug Administration |
| FITC | Fluorescein isothiocyanate |
| FTC | Emtricitabine |
| Gln | Glutamine |
| GLT1 | Glutamate transporter |
| Glu | Glutamic acid |
| gp | Glycoprotein |
| g/Mol | Grams per mole |
| GRID | Gay-related immune deficiency |

| | |
|---|---|
| HAART | Highly Active Antiretroviral Therapy |
| HIA | Human intestinal absorption |
| His | Histidine |
| HIV | Human immunodeficiency virus |
| HIV-1 | HIV type 1 |
| HIV-2 | HIV type 2 |
| HMPO | 5-hydroxy-3-methylpyrimidin-4(3H)-one |
| HRP | Horseradish peroxidase |
| HTLV | Human T-lymphotropic virus |
| HTS | High throughput screening |
| Hz | Hertz |
| $IC_{50}$ | Inhibitory concentration 50 |
| IMAC | Immobilized metal affinity chromatography |
| IN | Integrase |
| INI | Integrase inhibitor |
| INSTI | Integrase strand transfer inhibitor |
| $IN_{WT}$ | Integrase wild type |
| IPTG | Isopropyl β-D-1-thiogalactopyranoside |
| KDa | kilo dalton |
| KS | Kaposi's sarcoma |
| LAV | Lymphadenopathy virus |
| LB | Luria Bertani |
| Leu | Leucine |
| mARSC | Microarray screening |
| $Mg^{2+}$ | Magnesium |
| ml | Millilitre |
| MLSCN | Molecular Libraries Screening Centers Network |

| | |
|---|---|
| MOI | Multiplicity of infection |
| mM | Millimolar |
| Mr | Relative molecular mass |
| mRNA | Messenger RNA |
| MTS | 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium |
| MW | molecular weight |
| NCC | NIH Clinical Collection |
| Nef | Negative factor |
| ng | Nanogram |
| Ni | Nickel |
| NIH | National Institutes of Health |
| nm | Nanometer |
| nM | Nanomolar |
| NMR | Nuclear magnetic resonance |
| NNRTI | Non-Nucleoside reverse transcriptase inhibitors |
| NRTI | Nucleoside reverse transcriptase inhibitors |
| NTD | N-terminal domain |
| NtRTI | Nucleotide reverse transcriptase inhibitors |
| OD | Optical density |
| PAGE | Polyacrylamide gel electrophoresis |
| PBS | Phosphate buffered saline |
| PCP | *Pneumocystis carinii* pneumonia |
| PFV | Prototype foamy virus |
| PI | Protease inhibitor |
| PIC | Preintegration complex |
| PMSF | Phenylmethanesulfonylfluoride |

| | |
|---|---|
| PR | Protease |
| *Pts* | *Pan troglodytes schweinfurthii* |
| *Ptt* | *Pan troglodytes troglodytes* |
| PVDF | Polyvinylidene difluoride |
| PVT | Polyvinyltoluene |
| QSAR | Quantitative structural activity relationships |
| RAL | Raltegravir |
| RC | Replication capacity |
| Rev | Regulatory factor |
| RNA | Ribonucleic acid |
| Ro5 | Rule of five |
| RPMI | Roswell Park Memorial Institute Medium |
| RRE | Rev Response Element |
| RT | Reverse transcriptase |
| SD | Standard deviation |
| SDF | Standard database format |
| SDS | Sodium dodecyl sulfate |
| SIV | Simian immunodeficiency virus |
| SIVcpz | SIV that naturally occur in chimpanzees |
| SIVcpz*Ptt* | SIVcpz *Pan troglodytes troglodytes* |
| SIVcpz*Pts* | *SIVcpz Pan troglodytes schweinfurthii* |
| SIVgor | SIV that naturally occur in gorillas |
| SIVsmm | SIV that naturally occur in sooty mangabey monkeys |
| SMN2 | Survival motor-neuron 2 |
| SPA | Scintillation proximity assay |
| SSC Buffer | Saline-sodium citrate buffer |
| ST | Strand transfer |

| | |
|---|---|
| TAR | Transactivation response element |
| Tat | Transcription transactivator |
| TDF | Tenofovir disoproxil fumarate |
| tDNA | Target DNA |
| Thr | Threonine |
| TMB | Tetramethylbenzidine |
| TP | Triphosphate |
| TR-FRET | Time resolved fluorescence resonance energy transfer |
| uHTS | ultra High Throughput Screening |
| URF | Unique recombinant form |
| USA | United States of America |
| Vif | Viral infectivity factor |
| Vpr | Viral protein R |
| Vpu | Viral protein U |
| VS | Virtual screening |
| WT | Wild type |
| x g | Relative centrifugal force |

# *LIST OF FIGURES*

***Figure 1.1:*** The structure and constituents of the HIV virus particle (virion). The illustration depicts the viral RNA genome and viral enzymes contained in a p24 viral core..... *4*

***Figure 1.2:*** Representation of the HIV-1 genome and some of the essential proteins that these genes express. ..... *5*

***Figure 1.3:*** Schematic representation of the seven steps in the HIV-1 lifecycle... *7*

***Figure 1.4:*** Representation of the stages involved in drug discovery and development depicting the approximate success rate of the compounds, the average cost contributed to each stage and the duration of each stage..... *10*

***Figure 1.5:*** A schematic summation of the random and rational approaches in early drug discovery that ultimately lead to lead optimization..... *12*

***Figure 1.6:*** The NIH Clinical Collection (NCC) small molecule therapeutic indication. ..... *16*

***Figure 1.7:*** The schematic representation of the HIV-1 structural domains[7] ..... *17*

***Figure 1.8:*** The structure of the HIV-1 integrase (IN) dimer..... *17*

***Figure 1.9:*** Schematic representation of HIV-1 integration catalyzed by integrase (IN). ..... *18*

***Figure 1.10:*** The binding of raltegravir (RAL) to the active site of integrase (IN). ..... *19*

***Figure 1.11:*** The principle of the HIV-1 integrase (IN) scintillation proximity assay (SPA)..... *23*

***Figure 2.12:*** A representative of the NCC 96-well compound plate illustrating the strategic orthogonal pooling of compounds..... *32*

***Figure 3.13:*** SDS-PAGE depicting the pINSD.His gene expression in E. coli BL21 after induction with 1mM IPTG..... *39*

***Figure 3.14:*** Elution profile depicting the elution of the His-tagged proteins from the Ni-affinity column.. ..... *40*

***Figure 3.15:*** SDS-PAGE depicting the purity and intensity of the recombinant expressed proteins.. ..... *41*

***Figure 3.16:*** Western blot probed with A - anti-His and B - anti-integrase primary antibodies verifying the expression of $IN_{WT}$.. ..... *42*

***Figure 3.17:*** The absorbance spectra of the $^{3}$H-tDNA extracted from the Nanodrop installation version 1.3.1 software obtained when quantifying the $^{3}$H-tDNA concentration using the Nanodrop spectrophotometer.. ..... *43*

***Figure 3.18:*** Comparison of the percentage integrase (IN) strand transfer (ST) inhibition caused by raltegravir (RAL) when conducting experiments with metal cofactors, $MgCl_2$ and $MnCl_2$.. ..... *44*

***Figure 3.19:*** Comparison of the integrase (IN) strand transfer (ST) inhibition by raltegravir (RAL) after the enzymatic reaction between IN and $^{3}$H-tDNA was terminated with different stop solutions: ..... *46*

***Figure 3.20:*** Representative sigmoidal curves of raltegravir (RAL), elvitegravir (EVG) and 118-D24 obtained when conducting dose-response studies using an integrase inhibition SPA..... *47*

***Figure 3.21:*** The sigmoidal curves depict the comparison between the raltegravir (RAL) dose-dependent response (0.001µM to 10µM) when using the automated SPA and the manual SPA. ..... *48*

***Figure 3.22***: Pie chart representation of the screened NCC pools according to their percentage integrase (IN) strand transfer (ST) inhibition. ..... *49*

***Figure 3.23:*** The sigmoidal curves demonstrate the dose-response studies of dolutegravir (DTG), raltegravir (RAL) and elvitegravir (EVG) controls against HIV integrase (IN) harbouring the Q148H/G140S mutation. ..... *50*

***Figure 3.24:*** The dose response curves of A: Cefixime trihydrate (CEF) and B: Epigallocatechin gallate (EGCG) exemplifying their $IC_{50}$ values when screening against HIV integrase wild type ($IN_{WT}$) and $IN_{WT}$ harbouring the Q148H/G140S mutation. ..... *51*

***Figure 3.25:*** The structures of cefixime trihydrate and its structurally related compounds. A - Cefixime trihydrate (CEF) B - 7-aminodesacetoxycephalosporanic acid (7-ADCA) C - 7-aminocephalosporanic acid (7-ACA) D - Cephalothin sodium salt (CSS) E – Ampicillin (AMP).. ..... *52*

***Figure 3.26:*** A representation of the dose-response studies of epigallocatechin gallate to determine the cytotoxicity of the compound in MT-4 cells. ..... *53*

***Figure 3.27:*** Representation of the absorbance values measured at 405nm when analysing the effect of cefixime trihydrate (CEF), epigallocatechin gallate (EGCG) and ampicillin (AMP) on reverse transcriptase (RT) activity.. ..... *54*

***Figure 3.28:*** The absorbance readings ($A_{620}$) measured when conducting an HIV-1 integrase (IN) ELISA to confirm the IN strand transfer (ST) inhibitory effect of cefixime trihydrate (CEF), epigallocatechin gallate (EGCG) and ampicillin (AMP). ..... *55*

***Figure 3.29***: The absorbance values of cefixime trihydrate (CEF), epigallocatechin gallate (EGCG), ampicillin (AMP) and the controls measured at 450nm when detecting the level of p24 expression in an HIV-1 phenotypic assay.. ..... *56*

***Figure 3.30:*** Sigmoidal curves representing the dose-dependent inhibition of p24 expression attributed to epigallocatechin gallate (EGCG) and the control 118-D-24 in HIV-1 infected MT-4 cells.. ..... *57*

***Figure 3.31***: The comparison of the activity of cefixime trihydrate (CEF) and epigallocatechin gallate (EGCG) against HIV-1 integrase (IN) in the presence of 5mM and 10mM DTT as well as 10mM B-Me. ..... *58*

***Figure 3.32***: Two-dimensional (2D) structural diagrams generated using Discovery Studio version 3.1 software, exemplify the interactions between the defined active site of the HIV-1 integrase (IN) and the control compounds A, B and C; raltegravir (RAL), elvitegravir (EVG) and dolutegravir (DTG), respectively.. ..... *60*

***Figure 3.33***: Two-dimensional (2D) structures demonstrating the predicted structural interactions between A - Cefixime trihydrate (CEF) and B - Epigallocatechin gallate (EGCG) and the defined HIV integrase (IN) active site... ..... *62*

# *LIST OF TABLES*

***Table 3.1:*** Comparison of the different reaction buffers tested in the SPA and its effect on the CPM signal, integrase strand transfer inhibition and the Z-factor of the assay.....*45*

***Table 3.2:*** The predicted solubility, absorption and drug-like score of cefixime trihydrate and epigallocatechin gallate.....*63*

***Table 3.3:*** The predicted human intestinal absorption, solubility and blood brain barrier penetration scores obtained from Discovery Studio version 3.1.....*64*

***Table 3.4:*** The predicted oral bioavailability of cefixime trihydrate and epigallocatechin gallate using Lipinski Rule of 5.....*64*

# *APPENDICES*

**Appendix A:** Human Ethics waiver letter.....108

**Appendix B:** Table I - Recapitulation of the data obtained from biological assays throughout the study for all the controls, identified HIT compounds and derivatives…….....109

# CHAPTER 1

# Introduction

## 1.1. HIV/AIDS overview

### 1.1.1. History of HIV/AIDS

The human immunodeficiency virus (HIV) is a retrovirus that attacks and destroys the immune system by infecting vital cells in the immune defence system such as T-lymphocytes, more specifically CD4+ T-lymphocytes, and ultimately progresses to acquired immunodeficiency syndrome (AIDS).[1–3] HIV/AIDS was recognised in the 1980's however the first confirmed mortality attributed to HIV/AIDS is believed to have occurred in 1959 when a Congolese man died from a mysterious illness.[4] In 1986, preserved blood samples obtained from this man revealed that he was infected with HIV.[4] AIDS was first recognised in 1981 in the United States of America (USA) when an outbreak of rare infections associated with the immune system such as Kaposi's sarcoma (KS) and *Pneumocystis carinii* pneumonia (PCP) emerged among gay men.[5–7] The Centers for Disease Control (CDC) reported this illness as gay-related immune deficiency (GRID).[5] In 1982, this disease was then observed in heterosexuals, individuals who received blood transfusions and drug addicts which indicated that HIV/AIDS was not restricted to homosexuals.[8] The CDC subsequently termed this deadly disease AIDS since the major opportunistic infections, PCP and KS, only occurs in individuals' with weak immune systems. In 1984, the Pasteur Institute isolated a retrovirus lymphadenopathy virus (LAV) believed to be the causative agent of AIDS.[1] However, in 1984 Gallo and co-workers[3] reported the isolation of human T-lymphotropic virus (HTLV-III) which was also responsible for HIV/AIDS. LAV and HTLV-III were identified as the same virus and was named HIV in 1986. At this point, HIV/AIDS was found to be disseminated across at least 33 countries.[9] The first HIV treatment, azidothymidine (AZT), was approved by the Food and Drug Administration (FDA) in 1987.[10]

### 1.1.2. Epidemiology and phylogeny of HIV/AIDS

In 2012, it was estimated that 35.3 million people were infected with HIV.[11] Sub-Saharan Africa is the region most affected by HIV with 25 million infected people.[11] Of these, 6.1 million HIV infected people are living in South Africa, the country with the highest number of HIV infections across the world.[11] Although

# *Introduction*

HIV incidence levels have reduced by 42% in South Africa, the epidemic persists.[11]

HIV is believed to have originated from multiple cross-species transmission of the simian immunodeficiency virus (SIV) infecting non-human primates in Sub-Saharan Africa.[12–15] HIV is characterized into HIV type 1 (HIV-1) and HIV type 2 (HIV-2).[16,17] HIV-1, the more infectious strain, is derived from SIV that naturally occur in chimpanzees (SIVcpz) and gorillas (SIVgor).[13,18] SIVcpz is classified into SIVcpz *Pan troglodytes troglodytes* (*Ptt)* and SIVcpz *Pan troglodytes schweinfurthii (Pts)* based on their sub-species lineages.[19] The SIVcpz*Ptt* and SIVcpz*Pts* lineages resulted in viruses that differ by 30-50% in the viral *gag, pol and env* sequences.[20] HIV-2 derived from SIV in sooty mangabey monkeys (SIVsmm).[19,21,22]

HIV belongs to the *Retroviridae* family and falls within the *Lentivirus* genus as it consists of single stranded positive-sense RNA that is reverse transcribed by a virally encoded reverse transcriptase (RT).[23,24] Phylogeny analysis of the HIV *Lentivirus* have elucidated the geographic location of the HIV-1 and HIV-2 transmission across species. HIV-1 comprises four groups that are a result of distinct cross-species transmissions: M (Major), O (Outlier), N (Non M or Non O) and P.[25,26] Group M is the most predominant group and is responsible for almost 90% of HIV infections globally.[27] Group O is less prevalent than Group M and is responsible for only 1% of HIV infections and is confined to Cameroon, Gabon and neighbouring countries.[28,29] Only 13 cases of Group N infection have thus far been reported and they were restricted to infected individuals in Cameroon.[30] Group P, the most recently identified group, was documented in only two infected individuals from Cameroon.[27,31] All four groups of HIV-1, derived from SIVcpz, as well as the SIVgor strain, cluster with the SIVcpz*Ptt* lineage which infers that this subspecies was the original reservoir for human and gorilla infections.[32] Group M and N are believed to have originated from chimpanzees in the south-eastern corner of Cameroon and south-central Cameroon, respectively.[18,33] Group P is derived from the SIVgor lineage whereas the origin of Group O is not known since it is not closely related to ape viruses and hence could originate from either chimpanzees or gorillas.[32] Group M comprise subtypes A (A1, A2, A3), B, C, D, F(F1, F2), G, H, J and K that represent different HIV lineages, geographical and phylogenetic associations.[34,35] The migration pathways of these subtypes have

been described. Subtype A and D originated in central Africa and established in eastern Africa.[36] Subtype B is believed to have originated from a single African strain that spread to Haiti and ultimately into America and Europe.[36] HIV-1 subtype C is responsible for most HIV infections and is predominant across Southern- and East Africa as well as India.[35,37] Subtype F is prevalent in central Africa, eastern Europe and South America whereas subtype G is prevalent in West- and East Africa and central Europe.[38] Subtype H has been observed in central Africa whereas subtype K has been observed in the Democratic Republic of Congo and Cameroon.[38] Subtype J is only prevalent in central America.[38] Inter-subtype recombination is the result of co-infection with at least two different subtypes. When this occurs in only one individual, it is known as a unique recombinant form (URF). Upon transmission of a URF in many individuals that are not epidemiologically related, it is known as a circulating recombinant form (CRF). There are currently 51 CRFs characterized within the HIV-1 group M.[34] HIV-2 is less prevalent than HIV-1 and is mainly confined to West-Africa.[39] Thus far, eight distinct HIV-2 lineages (A-H) have been identified that originated from independent host transmissions.[40,41] HIV-2 is less infectious than HIV-1 and the number of infection caused by HIV-2 is declining, presumably due to its low viral load.[42,43]

### *1.1.3. The HIV-1 virion and its genomic organization*

The HIV-1 virus particle (virion) is sphere shaped with a diameter of 90 to 100 nanometer (nm).[3,44] Figure 1.1 illustrates the three basic structures of the HIV virion: the viral envelope, HIV matrix proteins and the viral capsid. The viral envelope consists of a double lipid layer, formed upon budding of the capsid from the host cell, surrounded by 72 protrusions comprised of trimeric glycoproteins (Gp) 120 attached to a trimeric transmembrane Gp41 stem.[8,17,18] The matrix comprises p17 proteins and is found between the viral envelope and viral core. The conical viral core consists of p24 proteins that encapsulate the viral genetic material, two copies of non-covalently linked single RNA positive strands, and essential viral replication enzymes such as RT, integrase (IN) and protease (PR).[47,48]

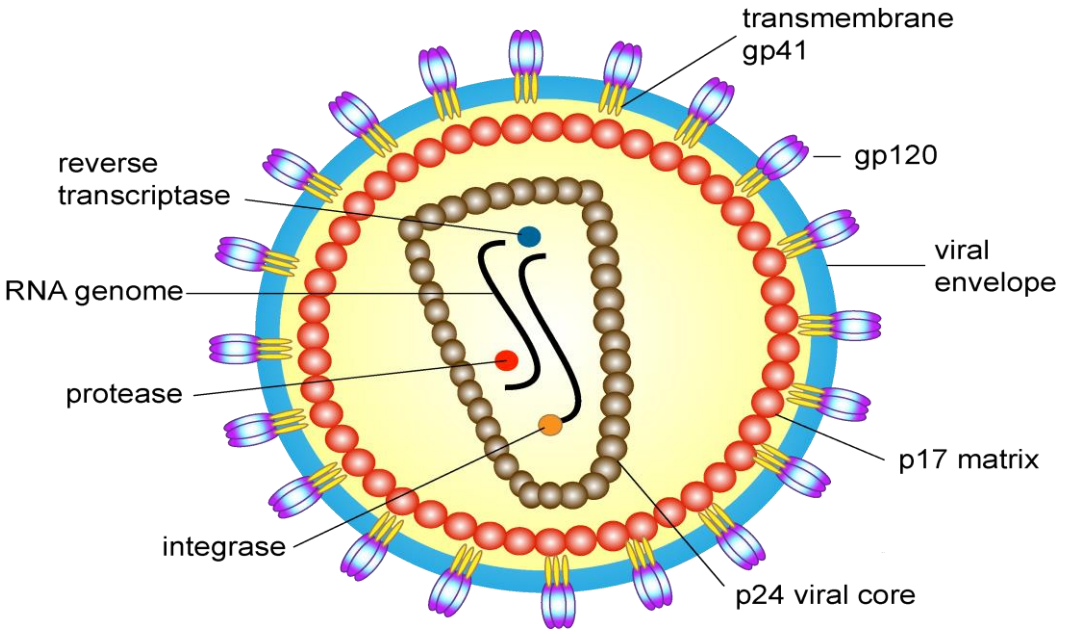



**Figure 1.1: The structure and constituents of the HIV virus particle (virion).** The illustration depicts the viral RNA genome and viral enzymes contained in a p24 viral core. The p24 viral core is then surrounded by the p17 matrix and the viral envelope. The viral envelope contains the transmembrane glycoprotein (gp) 41 stem attached to the surface gp120, arranged as trimers. Adapted and modified from www.stanford.edu/group/virus, accessed 1/09/2013.[49]

The HIV genome consists of only nine genes: *gag*, *pol, env, tat, rev, nef, vif, vpr* and *vpu* depicted in Figure 1.2. The *gag* and *env* genes encode structurally related proteins, Gag and envelope glycoproteins (gp160) that are required to produce new virus particles. The Gag polyprotein is cleaved by the viral PR into smaller structural proteins such as p17, p24 and the nucleocapsid (p9).[50] Cellular PR cleave gp160 into the transmembrane gp41 and the surface gp120.[51] The *pol* gene encodes polymerase enzymes RT, RNAse H, IN and PR that are expressed as a Gag-Pol fusion protein (Figure 1.2). The virally encoded PR then cleaves the Pol polypeptide from the Gag-Pol fusion protein where the Pol peptide is subsequently digested further into RT, RNAse H, IN and PR that are responsible for DNA synthesis, viral DNA integration into host DNA and the cleavage of Gag-Pol polyproteins during maturation of virions, respectively.[52–54] The regulatory genes *tat* and *rev* encodes the transcription transactivator (Tat) and regulatory factor (Rev), respectively and aid in regulating viral replication. Tat is an RNA binding protein that binds to the transactivation response element (TAR) at the 5' terminal end of HIV RNA thereby activating transcription.[55] Rev is a sequence based RNA binding protein that regulates HIV gene expression by binding to a

complex RNA secondary structure, termed the Rev Response Element (RRE), subsequently facilitating the migration of unspliced or incomplete spliced viral RNA from the nucleus to the cytoplasm.[56–58] Accessory genes *nef, vif, vpr* and *vpu* encode the virulence factors negative factor (Nef), viral infectivity factor (Vif), viral protein R (Vpr) and viral protein U (Vpu), respectively.[47,48,59,60] These proteins while not essential for viral replication *in vitro,* are however vital for viral replication *in vivo*. Nef ensures HIV infection by perturbing T-cell activation, increases virulence by down regulating CD4+ lymphocytes as well as stimulating the infectivity of the HIV virion.[61–63] Vif prevents the antiviral host protein apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) from entering the HIV virion by targeting it for cellular degradation.[64] HIV replication does not rely on Vif in most cells suggesting that these cells host a protein similar in function to Vif.[52] Vpr is involved in the nuclear localization of the HIV preintegration complex (PIC) in non-dividing cells such as macrophages.[65] Vpu plays a role in CD4+ down regulation and increases the release of HIV virions from the infected cell surface.[66,67]

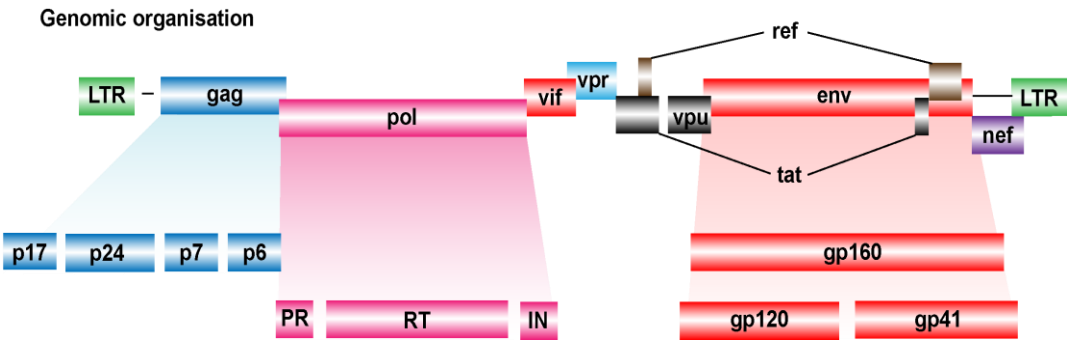



**Figure 1.2: Representation of the HIV-1 genome and some of the essential proteins that these genes express.** Adapted and modified from www.stanford.edu, accessed 1/09/2013.[49]

### *1.1.4. The HIV lifecycle*

The human immune cells involved in HIV-1 infection and replication include macrophages, dendritic cells (DC) and T-lymphocytes where activated CD4+ lymphocytes are the major target for HIV infection.[3,68] Upon HIV transmission, macrophages and DCs capture the cell free virus at the mucosal surface *in vivo* and present it to naive T-cells for infection and subsequent viral replication.[68–72] Since CD4+ T-lymphocytes are the major target for HIV infection, Figure 1.3 briefly describes the HIV-1 life cycle in CD4+ lymphocytes.

Briefly, the HIV-1 life cycle , demonstrated in Figure 1.3, is carried out in multiple steps that begin when the gp120 of the virion binds to the CD4+ receptor on CD4+ lymphocytes as well as one of the co-receptors, CCR5 or CXCR4, consecutively.[73,74] The binding of gp120 to the CD4+ receptors and co-receptors induces a conformational change that allows the gp41 to insert its hydrophobic terminus into the host cell membrane facilitating fusion of the host- and viral cell membranes (Step 2 in Figure 1.3). The HIV envelope subsequently propels the viral capsid into the cytoplasm of the host cell.[75–77] Upon entering the host cell, the p24 capsid is digested, subsequently releasing the viral genetic material, viral replication enzymes and associated proteins into the cytoplasm of the host cell. The viral genetic material, two single RNA positive strands, is reverse transcribed to viral cDNA by error-prone RT (Step 3 in Figure 1.3).[78] A tetramer of IN collates with Vpr, p6, p7 matrix, and host proteins such as BAF, Gemin2, HAT p300, HMGA1, HSP 60, Human EED protein, Importin 7, IN interactor 1, LEDGF/p75, and UNG2 thus forming the PIC.[79–88] The PIC translocates into the nucleus of the host cell where IN catalyzes the irreversible integration of the pro-viral DNA into the host DNA forming the provirus (Step 4 in Figure 1.3). The viral transcription factors transcribe the integrated DNA and the resulting mRNA is exported to the cytoplasm (Step 5 in Figure 1.3). The mRNA is subsequently translated to polypeptide chains that are cleaved by viral or host proteases to synthesize viral proteins in the endoplasmic reticulum (ER).[89] These viral proteins are transported to the surface of the cells and are embedded in the host cell membrane where these proteins collate with other viral proteins (Step 6 in Figure 1.3). The viral proteins cluster together with essential host proteins thereby forming an immature virion. The immature virion subsequently buds off from the cell surface and together with a portion of the host cell membrane forms the viral envelope. When the Gag polyprotein chains of the immature virion are fully cleaved by PR, it forms an infectious mature virion that has the ability to infect new host cells (Step 7 in Figure 1.3).[90]

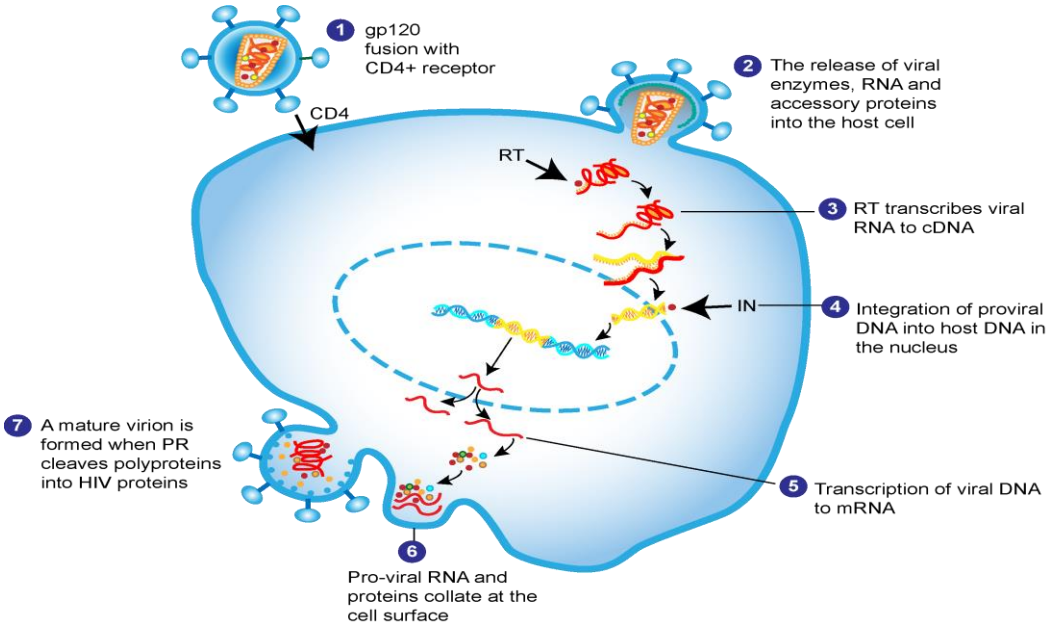



**Figure 1.3: Schematic representation of the seven steps in the HIV-1 lifecycle:** 1 - Binding and fusion of the gp120 on the surface of the HIV virion and the host CD4+ lymphocytes. 2 - Insertion of viral replication enzymes, RNA and accessory proteins into the host cell. 3 - Reverse transcription of viral RNA to DNA. 4 - Integration of pro-viral DNA into the host DNA in the nucleus. 5 - Transcription of pro-viral DNA to mRNA to synthesize viral proteins. 6 - Immature virions forms when viral proteins cluster together and buds off from the host cell membrane. 7 - Protease cleaves viral polyproteins into individual viral proteins and is known as mature virions. Illustration obtained and modified from www.niaid.nih.gov, accessed 12/07/2013.[91]

## 1.2. HIV treatment and drug discovery

### 1.2.1. Current antiretroviral treatment

Each stage of the viral life cycle represents a potential antiretroviral drug target. There are currently four drug classes in the HIV life cycle namely; viral entry inhibitors, RT inhibitors, IN inhibitors (INI) and PR inhibitors (PIs).[92,93] Highly Active Antiretroviral Therapy (HAART) is a combination of antiretroviral drugs that comprise at least two different drug classes that delay disease progression from HIV to AIDS by effectively suppressing the viral load and increasing the genetic barrier to the development of antiretroviral drug resistance.[94]

Antiretrovirals targeting RT encompass Nucleoside RT inhibitors (NRTI), nucleotide RT inhibitors (NtRTI) and non-nucleoside RT inhibitors (NNRTI). AZT was the first NRTI approved by the FDA in 1985 and functions as faulty DNA blocks that are incorporated into the newly synthesized viral DNA, resulting in

chain termination.[95,96] NRTIs and NtRTIs share a similar mechanism of action in that they are analogues of the naturally occurring deoxynucleotides required for DNA synthesis.[97] These NRTIs and NtRTIs analogues differ from the naturally occurring deoxynucleotides as the analogues do not possess a 3'-hydroxyl on the deoxyribose moiety required for the formation of a 5'-phosphodiester bond that facilitate DNA chain extension.[97–99] NRTIs and NtRTIs compete and replace the naturally occurring deoxynucleotides thereby impeding viral DNA synthesis. NRTIs require activation before they are incorporated into viral DNA.[99] These NRTIs are activated when cellular kinase enzymes phosphorylate the deoxyribose moiety on the NRTIs.[97] Current NRTIs available are lamivudine, AZT, emtricitabine, abacavir sulphate, stavudine, didanosine, dideoxyinosine, emtricitabine (FTC) and enteric coated didanosine.[100] NtRTIs evade the conversion step of nucleosides to nucleotides. Current NtRTIs available are tenofovir disoproxil fumarate (TDF) and adefovir.[100] NNRTIs bind to a site on RT distant from the binding site subsequently inactivating the enzyme allosterically.[99,101,102] Currently available NNRTIs are nevirapine, efavirenz, delavirdine, etravirine and rilpivirine.[100,103]

The discovery of PIs was pivotal as it was the beginning of combinational therapy, HAART. PIs are substrate analogues of the HIV aspartyl protease enzyme that bind to the active site of protease subsequently blocking the activity of the enzyme and therefore preventing the cleavage of viral proteins. These uncleaved viral proteins are then considered defective and therefore the virion cannot mature into an infectious particle.[104] FDA approved PIs include nelfinavir, saquinavir, darunavir, ritonavir, fosamprenavir calcium, indinavir and tipranavir.[100]

Entry inhibitors such as maraviroc, a co-receptor antagonist, inhibit viral entry into the host cell by blocking CCR5 on the host cell. FDA approved fusion inhibitor, enfuvirtide, prevents the fusion of the viral envelope and cell membrane by interacting with gp41 thereby inhibiting viral replication.[100,105,106]

The latest class of antiretrovirals to obtain FDA approval inhibit viral integration into the host genome by targeting IN.[107] As such, IN is considered to be one of the most favourable antiretroviral targets in the replication cycle of HIV. Raltegravir (RAL), an IN strand transfer inhibitor (INSTI), was the first INI that was FDA approved in 2007.[108,109] Elvitegravir (EVG) was FDA approved in 2012

and dolutegravir (DTG) is the most recent INSTI that was FDA approved in August, 2013.[110] RAL is further discussed in section 1.2.3.2.

A single tablet regimen, QUAD pill (Gilead Science, USA), was approved in 2012 that encompasses a fixed dose-combination of EVG, TDF, FTC and cobicistat.[111] EVG, TDF and FTC confers HIV antiretroviral activity by suppressing viral replication whereas cobicistat inhibits the EVG metabolizing enzyme cytochrome P450 3A consequently increasing the concentration of EVG in blood.[100,112,113]

Despite the successful development of antiretrovirals, obstacles such as the transmission of HIV-1 drug resistant viruses loom. Due to its high mutation rate, as well as antiretroviral drug pressure, the HIV-1 virus ultimately develops resistance to all currently available antiretrovirals.[114,115] As such, individuals who harbour these resistant strains have less antiretroviral treatment options resulting in an increase in mortality.[116] By using drug combinations such as those recommended in HAART, the dosage required for effective treatment decreases, the genetic barrier to developing drug resistance mutations increases, and therefore the emergence of drug resistant HIV-1 viral strains is delayed.[117] Resistant viral mutations against RAL are further discussed in section 1.2.2.5.

### *1.2.2. Drug discovery and development*

Drug discovery is the process whereby new therapeutics against a specific target is discovered. Figure 1.4 describes the lengthy and costly process of drug discovery and its low drug output rate. It is estimated that the entire drug discovery and development process ranges between 10-15 years with an approximate cost of up to $1.5 billion per successful drug.[118] It is estimated that only one drug candidate of a possible 5000 -10 000 tested will be FDA approved. Drug discovery and development stages include: drug discovery, preclinical trials, clinical trials, and FDA review and post-FDA approval.[119] As demonstrated in Figure 1.4, the typical steps involved in early drug discovery include: target identification and validation, development of a compound screening assay followed by the identification of a HIT compound through random- or rational approaches that ultimately leads to lead optimization.[117] A HIT compound is defined, in the context of this study, as a compound that confers inhibition against a specific target.

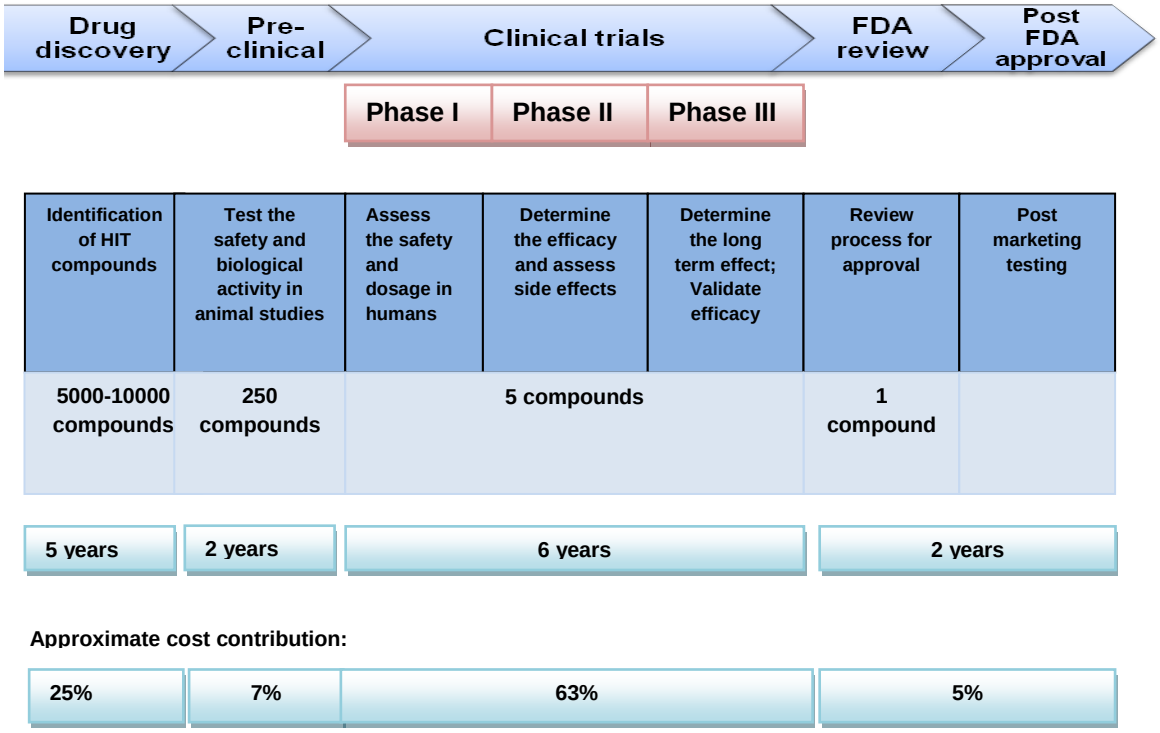



**Figure 1.4: Representation of the stages involved in drug discovery and development depicting the approximate success rate of the compounds, the average cost contributed to each stage and the duration of each stage.**

### *1.2.2.1. Random screening approach*

Random screening involves the biological screening of compound libraries and does not require any prior knowledge of the compounds that are screened against a specific target. Compound libraries are typically compiled through combinatorial chemistry, parallel synthesis, drug repositioning and natural product libraries as recapitulated in Figure 1.5.[120] Combinatorial chemistry involves the compilation of molecules related to a particular scaffold that is active against a specific target (biased library) or a collection of randomly synthesized compounds that are not limited to a single target (unbiased library).[121] Parallel synthesis involves the synthesis of compounds in parallel using spatially separated compartments.[122] Furthermore, drug repositioning is based on existing drugs that are redistributed in novel drug targets (further discussed in section 1.2.2.4).[123] The compounds screened in random biological assays are not restricted to synthesized molecules but also low-molecular-weight natural products derived from plants that comprise natural product libraries.[124] The screening of these compound libraries are often screened through High

Throughput Screening (HTS) where biological screening assays are amended to an automated system.[121]

Random screening has been documented in HIV research in Africa where screening of natural products was predominant.[125–127] Through random screening, a plant widely used in South Africa as traditional medicine, *Sutherlandia frutescens,* has been reported as a possible treatment for HIV/AIDS by increasing CD4+ counts, stimulating appetite and decreasing viral loads.[128] This plant extract is in phase II clinical trials and the results thereof are expected in 2013.[117] High running cost is often associated with HTS and since compounds are randomly screened without prior knowledge of its interaction with the target, this random approach results in a high failure rate.[117] As such, random approaches in early drug discovery are not conducive to cost effectiveness**.** However, random screening is the preferred approach when searching for HITS with novel mechanisms of action since these compounds are not designed against a specific target.[120]

#### *1.2.2.2. Rational design approach*

The rational design drug discovery approach is based on selecting compounds in available compound libraries that are likely to have activity against the specific target.[129] These compounds are selected due to prior investigation with their target of interest.[129] Rational design can be classified as ligand-based design or structure-based design summarized in Figure 1.5. Structure based design approach is where the 3 dimensional (3D) structure of the specific target is exploited in order to develop suitable molecules that would interact with the target.[130] The structure of the target can be determined biologically through X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy. The structure of the target may also be elucidated through computational methods such as threading and homology modelling.[131] Homology modelling relies on the known 3D structure of a homologous target that serves as a template for the prediction of the 3D structure of the new target.[131] When there is no known structure available of an homologous target, the 3D structure of the target of interest is predicted using its amino acid (aa) sequence by comparing it to a database comprising known folds.[131,132] *De novo* drug design is based on the structure and aa orientation of the active site. Ligands are then designed specific to the active site of the target.[133]

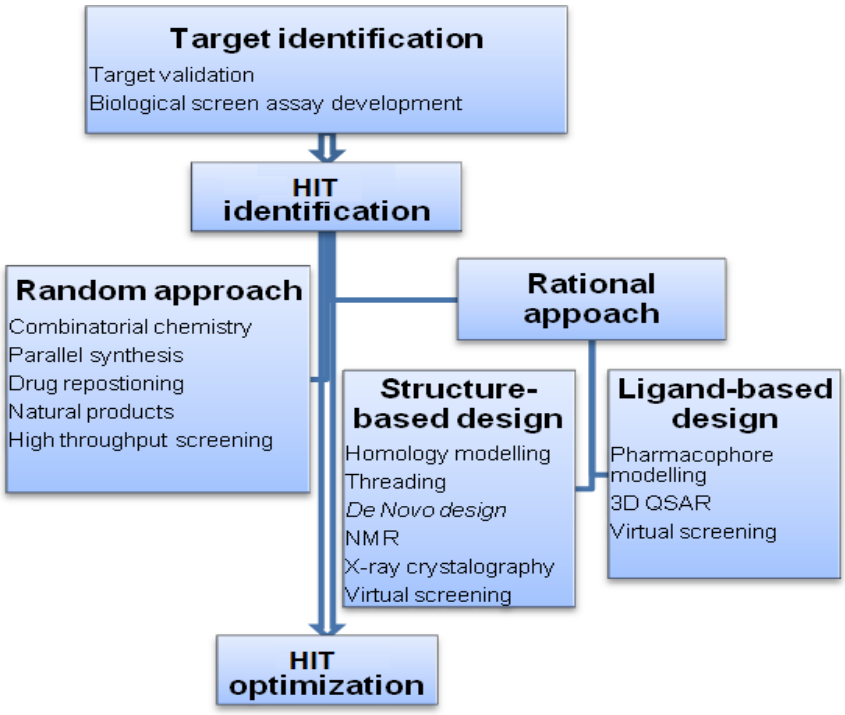



**Figure 1.5: A schematic summation of the random and rational approaches in early drug discovery that ultimately lead to HIT optimization.** Herein, some of the tools involved in random screening and rational design are listed that ultimately lead to HITcompounds that enter preclinical trials.

Ligand-based design involves screening of molecules that are similar or related to the pharmacophore associated with a specific target. Figure 1.5 briefly lists the approaches used in ligand-based drug design. Pharmacophore modelling and 3D quantitative structural activity relationships (QSAR) are the most common ligand-based design tools.[134] Both structure- and ligand-based drug designs are typically linked with virtual screening (VS) that aid in HIT identification.[135] VS has become an integral part of drug discovery as it predicts the interaction of ligands with a specific target through computational methods. Rational drug design is advantageous when linked to computational methods since it evaluates the interaction of compounds with a specific target through rapid and low cost simulation. An example of successful rational drug design include the discovery of PIs such as ritonavir and indinavir.[136]

Furthermore, rational drug design can be extended to predictive ADMET (absorption, distribution, metabolism, excretion and toxicity) studies to describe the pharmacological properties of compounds before they enter preclinical trials.[137] All ADMET descriptors of drugs can be tested *in silico*.[137] The absorption

properties of drugs refer to the ability of compounds to enter the bloodstream via human intestinal absorption (HIA) after oral administration which is dependent on the solubility and permeability of the drugs.[138,139] Chris Lipinski[140] pioneered the Lipinski rule of 5 (Ro5) paradigm that identify the parameters that contribute to poor oral bioavailability of synthetic compounds. The Lipinski rule of 5 states that drug properties should adhere to the following rules for it to be considered as drug like: the molecular weight (MW) of the compound should be less than 500 Daltons (Da); the water partition coefficient (cLogP) should be less than 5; there should be less than 5 hydrogen bonds donors; there should be less than 10 hydrogen bond acceptors.[140] The distribution of a compound determines ability of a compound to reach its specific target in the body such as the organs, muscles and crossing the blood brain barrier (BBB).[141] In addition, the distribution of compounds includes the ability of the compound to bind to plasma binding proteins in blood consequently delivering drugs to its specific target.[142] After oral bioavailability, most drugs are metabolized in the liver by CYP p450.[143] Important metabolic drug properties include metabolic stability, drug-drug interactions and toxicity influenced by the metabolism of a drug.[141,144] Drugs are then removed from the body through kidney excretion, biliary excretion or fecal excretion. Dose dependent studies of drugs are conducted to determine whether these drugs are toxic to cells in the body.[142]

#### *1.2.2.3. High Throughput Screening and automation*

HTS is a high-tech approach in drug discovery that has gained popularity in the last two decades due to its ability to screen large compound libraries against specific targets.[121] As mentioned in section 1.2.2.1, HTS is a technique widely used in random drug screening. HTS used on an automated platform aims to reduce reagent use, human error and extensive methodology thus accelerating drug development.[145] HTS laboratories have the ability to screen up to 10 000 compounds per day whereas ultra HTS (uHTS) laboratories are more advanced in their screening capabilities and ≥100 000 compounds can be screened per day.[146] In order to reduce cost, automation together with miniaturization has become popular. Miniaturization decreases the volume of the biological- and chemical reagents by using half-area well plates or plates with higher well densities. [147]

Automation has rapidly become the driving force of HTS in drug discovery and can be classified into three types: hand-held automation, unit automation and integral automation.[148] Hand-held automation includes devices that are held and controlled by a human operator. Such automation is widely used in laboratories in low throughput screening.[149] Unit automation is the most common type of automation that includes automated benchtop systems that do not require constant control by a laboratory operator. Unit automated devices are generally integrated into fully automated workstations termed integral automation devices. These workstations comprise automated liquid handling systems on an X, Y and Z axis that have the ability to aspirate and dispense liquid. In addition, these workstations contain a robotic arm that has the ability to transfer microplates and test tubes to integrated unit automated devices. Fully integral automated workstations are programmed with software that controls and schedules experiments as well as identify errors made during the experiment. These fully integral automated workstations can ease complex HTS screening by amending biological screening assays to an automated robotic platform. Biological screen assays are optimized and amended to an automated system ensuring stable assay signal and sensitivity.[120,150] Fully automated systems are advantageous as these systems provides higher data quality, rapid screening of small molecules and they are less laborious than manual assays.[121,151]

#### *1.2.2.4. Drug repositioning*

Drug repositioning is a method in drug discovery which involves the redistribution of FDA approved drugs or drugs that have been in clinical trials, into new disease models.[123] Drug repositioning is advantageous in drug development since the risk profile and adverse effects of the drugs are already known thereby reducing cost as well as development time by 3-5 years.[152] The average cost of a successful repositioned drug is estimated at $8.4 million compared to $1.5 billion when developing a new drug.[153] Compound libraries comprising of repositioned drugs include the National Institutes of Health (NIH) Clinical Collection (NCC), the John Hopkins clinical compound library and the Developmental Therapeutics Programme (DTP) approved oncology drug set.[154]

#### 1.2.2.4.1. The NIH Clinical Collection compound library

The NIH have implemented drug repositioning by assembling a compound library comprised of the NCC I and NCC II that contain 446 and 281 small molecules, respectively. The NCC forms part of the NIH roadmap Molecular Libraries Screening Centers Network (MLSCN) in a mission to expand the bioactivity data on these compounds and to re-supply compounds that are FDA approved or that have undergone clinical trials. The compounds in the NCC are selected based on their purity, solubility properties and commercial availability. The molecules in the NCC have been classified according to their therapeutic categories and are depicted in Figure 1.6. Since most of these molecules have been in clinical trials phase I to III, they have well characterized safety profiles and high drug-like properties, hence easing drug development. These molecules can be used as a basis for medicinal chemistry optimization and might possibly be useful for new drug targets such as HIV-1 IN.[154] The success of drug repositioning using the NCC has been reported in several studies.[155-158] Corcoran and co-workers[155] have identified a novel mechanism of action of histone deacetylase inhibitors in protein aggresome models. Stavrovskaya and co-workers[156] tested the NCC small molecules as mitochondrial permeability transition inhibitors and identified 28 small molecules that delayed mitochondrial permeability. Lunn and co-workers[157] conducted a HTS study to identify a compound that increase the production of survival motor-neuron 2 (SMN2) luciferase reporter protein in patients with paediatric neurodegenerative disease spinal muscular atrophy where Ibuprofen was found to selectively increase the production of SMN2 luciferase reporter protein. Rothstein and co-workers[158] randomly screened 1040 FDA approved drugs that included the small molecules from the NCC compound library to identify stimulators of the dominant astroglial protein, glutamate transporter (GLT1) that inactivates glutamate which is a neurotransmitter in the nervous system. They discovered that beta (β)-lactam antibiotics found in the NCC stimulated GLT1 expression which consequently conferred neuroprotection.

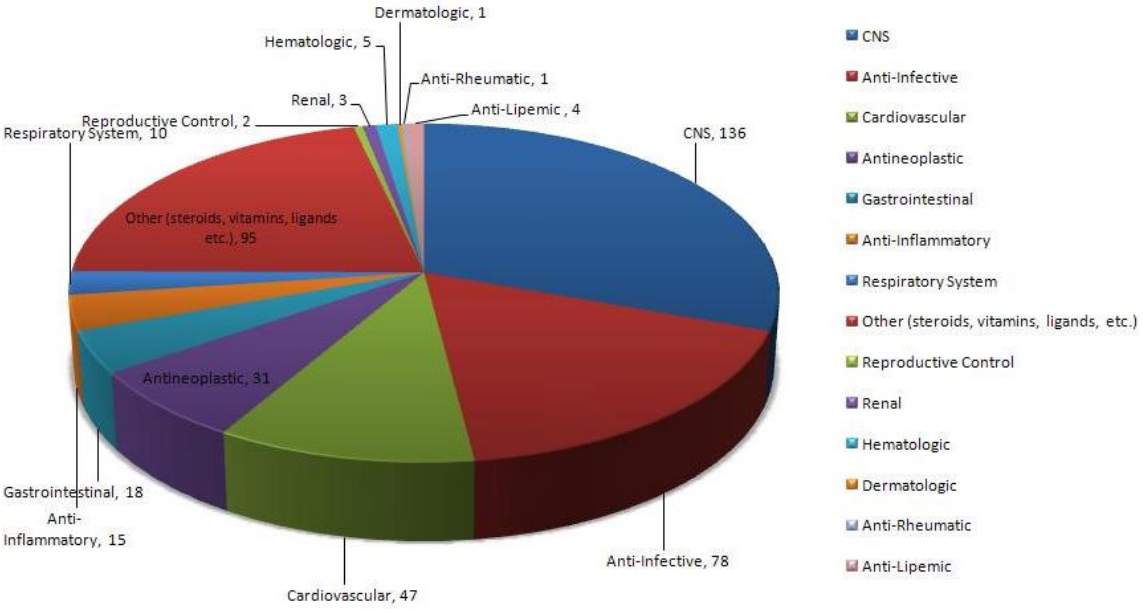



**Figure 1.6: The NIH Clinical Collection (NCC) small molecule therapeutic indication.** Drugs used in central nervous system (CNS) therapy (136) are predominant in the NCC while molecules used in renal, reproductive control, anti-lipemic, hematolic and dermatological therapy are minimal (< 10). Diagram extracted from www.nihclinicalcollection.com, accessed 20/10/2013.[154]

## *1.2.3. HIV-1 integrase as a drug target*

### *1.2.3.1. The structure of HIV-1 integrase and its function*

The HIV-1 IN enzyme is one of three enzymes essential for HIV replication that belongs to the transposase family of DNA transferases. Its function is to catalyse the integration of newly synthesized viral DNA into the host chromosomal DNA.[107,159] The 32 kilodalton (kDa) protein, depicted in Figure 1.7, is encoded from the 3'-terminus of the *pol* gene and is comprised of 288 aa sectioned into three functional domains: the N-terminal domain (NTD), the catalytic core domain (CCD) and the C-terminal domain (CTD).[160–162] The NTD (aa 1-50) consists of a dimeric helix-turn-helix that has a conserved histidine (His) and cysteine (Cis) motif (HHCC) coordinated by zinc.[163] The CCD (aa 51-212) is comprised of six β-strands surrounded by six α-helices and contains the $D_{64}D_{116}E_{152}$ (DDE) motif consisting of Asp64, Asp116 and Glu152 acidic residues essential to the catalytic activity of IN. The divalent metal cation $Mg^{2+}$, coordinates Glu152, Asp64 and Asp116 making it a fundamental cofactor in the catalytic domain.[107] These aa residues are conserved throughout the IN CCD domain of retroviruses.[164–166] The CTD (aa 213-288) is less conserved and has a SH3-like fold that facilitates non-

specific DNA binding.[163] Overall, all three domains of IN are involved in DNA binding and multimerization that is important for integration.[160]

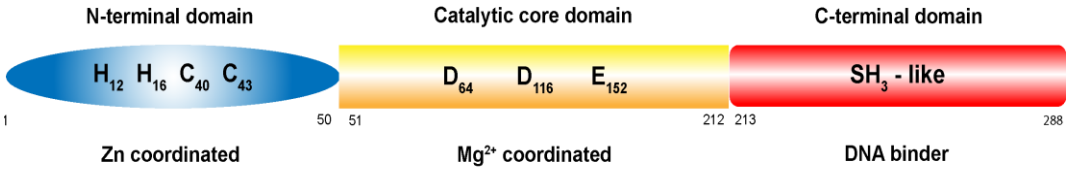



**Figure 1.7: The schematic representation of the HIV-1 structural domains**: the zinc coordinated N-terminal domain, the $Mg^{2+}$ coordinated catalytic core domain and the non-specific DNA binding C-terminal domain. Illustration obtained and modified from McCol and Chen, 2010.[167]

When zinc binds to the HHCC motif in the NTD, it promotes multimerization resulting in the formation of tetramers which is the functional state of IN.[168] Each tetramer consists of a dimer of dimers that is linked by the six α-helices of the CCD subsequently forming a flexible loop between the helices of each monomer that plays a vital role in the integration of DNA (Figure 1.8).[107] The positive charge residues in the area spanning from the end of the CCD on one monomer to the CTD of the other monomer plays a crucial role in viral DNA stabilization for subsequent integration.

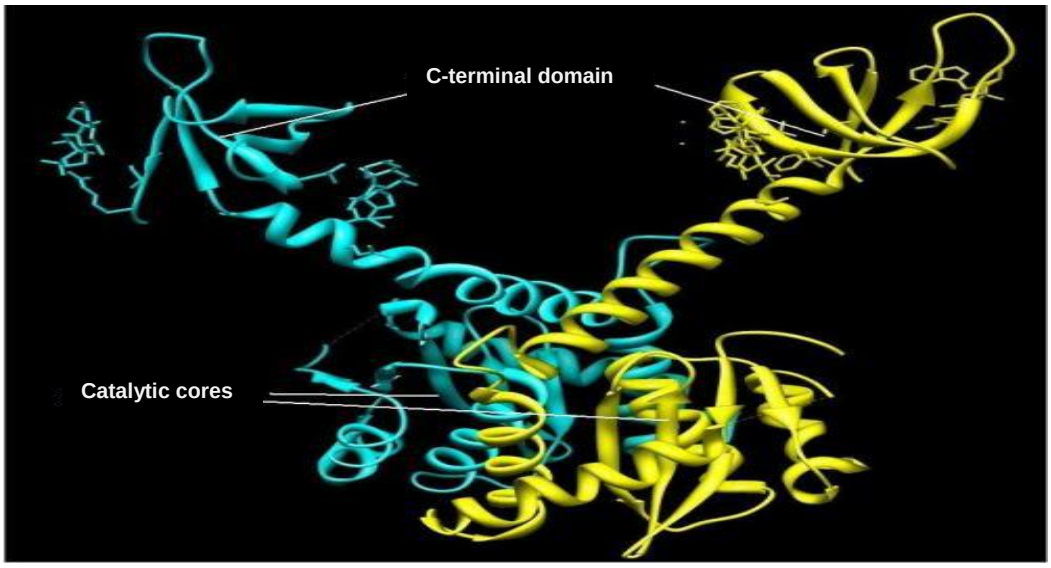



**Figure 1.8: The structure of the HIV-1 integrase (IN) dimer.** The IN dimer comprise the C-terminal domain and the catalytic core domain. Each monomer is denoted in a different colour. Structure obtained from mapdev.rutgers.edu, accessed 29/10/2013. [169]

IN catalyses irreversible viral integration in a two-step transesterification reaction: 3'-end processing and the ST reaction. Figure 1.9 briefly describes the integration of viral cDNA into host DNA. Firstly, the IN cleaves conserved dinucleotides, CA, at the linear 3'-viral DNA end in the cytoplasm resulting in CA overhangs subsequently exposing the hydroxyl groups (Step 1 in Figure 1.9). IN together with the processed viral DNA and other viral and host proteins forms a PIC and is transported into the nucleus where the ST reaction occurs (Step 2 in Figure 1.9). In the presence of $Mg^{2+}$, the exposed hydroxyl groups of the processed viral DNA attack the phosphodiester bonds on the host DNA at the site of insertion. Integration is completed when the two unpaired nucleotides at the 5'-end of the viral DNA are removed, the gaps between the host- and viral DNA are filled and the strands are covalently ligated by the host DNA repair proteins.[164,170–172] The host cell genome now contains the viral genetic material necessary to create progeny viruses.

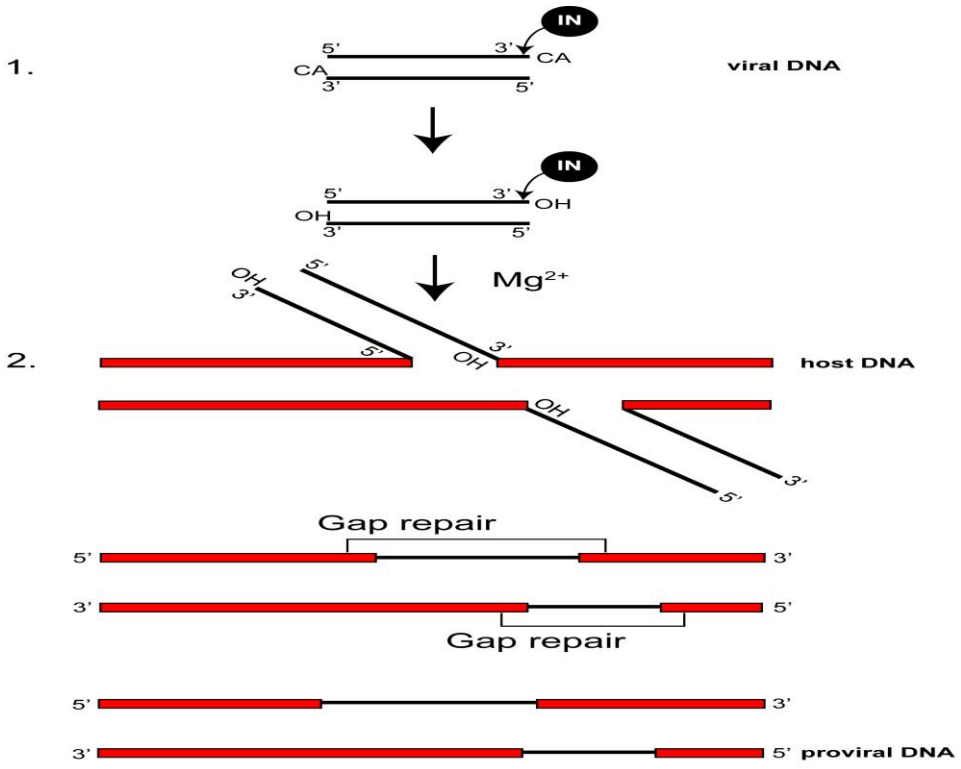



**Figure 1.9: Schematic representation of HIV-1 integration catalyzed by integrase (IN).** In the first step of integration, known as 3'-end processing, IN binds to the viral DNA subsequently cleaving the CA dinucleotides at the 3'-end thereby exposing hydroxyl groups. Secondly, the hydroxyl groups of the processed viral DNA attacks the host DNA resulting in integration. The DNA is ligated using host DNA repair proteins thereby forming a provirus. Adapted and modified from Suzuki and co-workers, 2012.[173]

### 1.2.3.2. The mode of action of the integrase strand transfer inhibitor, raltegravir

RAL is derived from a class of inhibitors that contain a distinct β diketo acid (DKA), 4-Aryl-2,4-diketobutanoic acid, moiety that demonstrate ST inhibition. The most potent β-DKA compound, L-731,988, exhibited a half maximal inhibitory concentration ($IC_{50}$) of 80 nanomolar (nM).[174] By substituting chemical moieties of DKAs, Merck Research Laboratories discovered that dihydroxypyrimidine carboxamide was an effective ST inhibitor selective for HIV IN only. RAL, a pyrimidine carboxamide, was the most potent derivative of the dihydroxypyrimidine carboxamides identified and demonstrated nanomolar activity with an $IC_{50}$ = 7nM.[175] RAL inhibits the ST reaction by chelating the divalent metal cations in the active site that modulate IN activity.[176] The binding of the RAL to the CCD of IN is depicted in Figure 1.10. The chemical structure of RAL possess a 5-hydroxy-3-methylpyrimidin-4(3H)-one (HMPO) chelator group that combines with an amide carbonyl oxygen atom providing the essential coplanar oxygen atoms that binds to the $Mg^{2+}$ in the DDE motif of the CCD.[108,177,178]

EVG and DTG are DKA derivatives with similar structural moieties as RAL and thus these derivatives exert ST inhibition in a similar mechanism of action as RAL. [110,179] RAL and EVG in the HAART regimen are currently at the forefront of HIV therapy that vastly reduces viral replication in HIV treatment-naïve patients as well as HIV treatment experienced patients. RAL and EVG are novel class inhibitors and are therefore active against viral strains resistant to NRTIs, NNRTIs, PIs and entry inhibitors.[180]

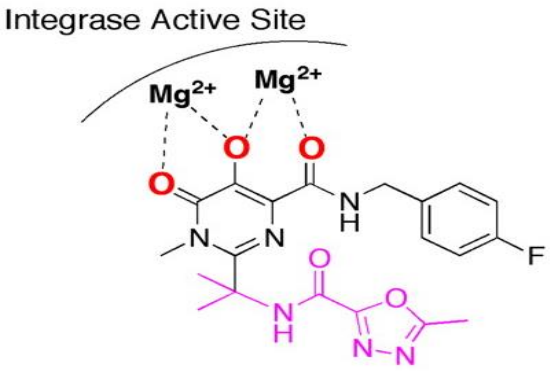



**Figure 1.10: The binding of raltegravir (RAL) to the active site of integrase (IN).** Coplanar oxygen groups of RAL bind to the $Mg^{2+}$ coordinated catalytic core domain of the HIV-1 IN binding site. Diagram extracted from Agrawal and co-workers.[181]

### 1.2.3.3. Resistance mutations against raltegravir and other integrase inhibitors

The emergence of viral resistant mutations compromise the efficacy of the antiretroviral agents.[182] Viral resistant mutations in relevant HIV-1 genes may confer cross-resistance to antiretroviral agents within a drug class and can therefore limit treatment options for treatment experienced patients.[182] To overcome the drug failure concern, novel drug classes or second generation antiretrovirals within a drug class should be developed.[182] Mutations in the viral genome, caused by the error-prone RT as well as drug pressure, attribute to viral resistance against antiretroviral agents. O'Neil[183] described the frequency of mutations due to RT as 0.85bp/genome/replication cycle. Multiple viral mutations emerge with increase in replication cycles resulting in a quasispecies of viruses that may hinder viral replication or confer resistance against antiretrovirals.[182] The extent of resistance can be quantified by analysing the properties of the virus such as the degree of resistance through fold-change $IC_{50}$ ($FCIC_{50}$) and the replication capacity (RC) of the antiretroviral drug resistant virus relative to wild type wild type.[182]

Recent reports have indicated that virological failure due to RAL treatment is attributed to resistance mutations in the IN catalytic domain.[184,185] Resistance against RAL is conferred in a stepwise process through mutations in one of three distinct pathways $IN_{N155H}$, $IN_{Q148H/R/K}$ or $IN_{Y143R/C/H}$.[186,187] $IN_{Q148H}$ is the predominant pathway in viral resistance against RAL followed by $IN_{N155H}$ and $IN_{Y143R/C/H}$.[188] The mutations observed emerge consecutively where studies conducted by Malet and co-workers[189] demonstrated that $IN_{N155H}$ is most likely the initial mutation that shifts to the $IN_{Q148H}$ mutation. The $IN_{Q148}$ pathway mutations have demonstrated stronger RAL resistance than $IN_{N155H}$ pathway mutations.[189] However, these primary mutations are often associated with impaired IN that subsequently reduce the RC of the virus. To deter this, primary mutations are usually coupled with secondary mutations that restore RC as well as confer strong resistance against RAL. $IN_{Q148}$ is usually associated with $IN_{G140S}$ and $IN_{N155H}$ is associated with $IN_{E92Q}$.[190] Delelis and co-workers[190] elucidated that $IN_{Q148H}$ as a single mutation confers RAL resistance but impairs RC. $IN_{G140S}$ as a single mutation does not confer resistance but restore RC when coupled to $IN_{Q148H}$. Therefore, $IN_{Q148H}$ in combination with $IN_{G140S}$ produces a virus that has a high RC and is greatly resistant to RAL.[190] Likewise, $IN_{N155H}$ in combination with $IN_{E92Q}$ or $IN_{T97A}$

increase both RC as well as resistance against RAL.[190] The primary mutation, $IN_{Y143C/R}$, reduces RAL susceptibility 5-20 fold however when coupled with $IN_{T97A}$, RAL susceptibility decreases by >100 fold.[191–193] Resistant mutations against RAL were also observed in patients failing EVG therapy indicating cross resistance mutations between RAL and EVG.[194,195] Furthermore, DTG conferred the mutations, $IN_{E92Q}$ and $IN_{Q148H/R}$, only in RAL-experienced patients.[196] $IN_{T124A}$ was observed in treatment-naive patients and conferred low-level resistance against DTG.[196] In addition the primary mutation $IN_{R263K}$ was observed in treatment-naive patients where viral replication decreased by 20% and a low-level of DTG resistance was observed (2-6 fold).[197] Secondary mutations $IN_{H51Y}$, $IN_{E138K}$ or $IN_{M50I}$ were observed that increased DTG resistance but decreased viral replication.[198]

#### *1.2.3.4. In vitro assays used to identify HIV integrase inhibitors*

Several assays have been developed to measure IN activity and identify inhibitors. Such assays include gel based assays; fluorescent resonance energy transfer (FRET)-based assays, microarray screening (mARSC) and *in vitro* microtiter plate assays that have been developed to recreate the integration process by using recombinantly expressed IN and double stranded (ds) DNA resembling viral DNA ends.[199–201] Most HIV-1 IN assays follow similar principles where viral dsDNA is immobilized onto a solid phase-support thereby quantitatively measuring IN activity by monitoring the fusion of the viral dsDNA with a labelled target DNA (tDNA).[145] These assays are used to screen HIV-1 IN activity in isolation and in the presence of possible inhibitors where ST and 3'-end processing can be measured consequently identifying or characterizing inhibitors.[202,203]

#### *1.2.3.5. HIV integrase assays amenable to automation*

Biochemical assays amended to HTS are desirable as it expedites the screening of small molecules and identification of INI. Such assays are responsible for the identification of several INI.[204,205] For an assay to be suitable for automation it should possess the following qualities: homogenous assay, sensitive detection method, preferably conducted in microplates, rapid and minimal steps and incubation times.[145] HIV-1 IN assays amenable to an automated system include time resolved FRET (TR-FRET), amplified luminescent proximity homogenous

assay (Alpha)-LISA, scintillation proximity assay (SPA) and more recently the cellular IN screening (CIS) assay.[202,206,207]

### *1.2.3.6. Automation of an HIV integrase scintillation proximity assay (SPA)*

Although several assay formats can be adapted to automated systems, the SPA is the method that is highly compatible with automation since it is a homogenous biochemical assay used for biological screening of a large scope of biochemical processes in a rapid and sensitive approach. Since the bound and free radiolabeled molecules can be detected, the assay does not require any centrifugation, physical separation and filtration steps therefore easing robotic complexity when used in automation.[208] The SPA is advantageous in automation with its high detection sensitivity, simplicity and fewer separating steps that reduce cost, time and potential errors.[208] The principle of the assay is based on a radiolabeled molecule when in close proximity to the SPA bead, activates a scintillant that emits light thereby monitoring the activity of a biomolecule.[208] The main components of the SPA are the radioactive isotopes and supportive material such as scintillation beads or scintillation microplates. The most common SPA beads are polyvinyltoluene (PVT) beads because they contain high efficiency scintillants within its matrix that are excited when in close proximity with a radioactive isotope. The PVT bead is coated in hydrophilic polyhydroxy film that masks hydrophobicity thus reducing non-specific binding.[145] The coated film covalently couples generic capture molecules such as avidin, streptavidin, glutathione as well as polyclonal secondary antibodies. The SPA has been adapted to radioactive isotopes such as tritium ($^{3}$H), $^{22}$P, $^{14}$C, $^{35}$S and $^{125}$I. However, $^{3}$H is the preferred radioisotope as the β-particles, released upon decay, has a short pathlength of 1.5 micrometer (μm) in aqueous solution that satisfies the energy transference distance criteria for SPA.[145]

The principle of this HIV IN SPA, illustrated in Figure 1.11, is based on the binding of the $^{3}$H host/tDNA ($^{3}$H-tDNA) to the IN-viral/donor DNA (dDNA) complex via a ST reaction bringing the $^{3}$H-tDNA in close proximity with the scintillant in the SPA beads. The β-particles emitted, when $^{3}$H decays, transfer energy to the scintillant resulting in light emission. If the ST reaction is inhibited by an inhibitory compound, the $^{3}$H-tDNA will not be captured by the IN-dDNA complex that is bound to the scintillation bead. The energy emitted by the $^{3}$H-tDNA will then

dissipate into the aqueous solution thus no light emission will occur. The intensity of radiation emitted is quantified through scintillation counting. The SPA in HIV-1 IN studies have been reported where Hu and co-workers[206] screened a large compound library of 1 million compounds using a SPA to identify possible IN inhibitors.

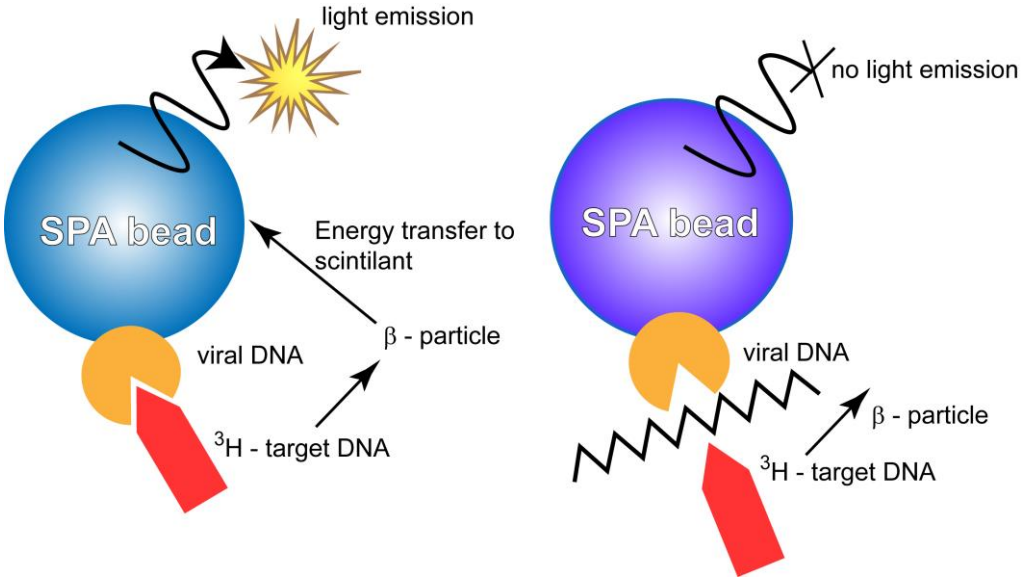



**Figure 1.11: The principle of the HIV-1 integrase (IN) scintillation proximity assay (SPA).** The assay is based on radiolabeled host target DNA ($^3$H-tDNA) that releases β-particles upon radioactive decay. When strand transfer (ST) occurs, the $^3$H-tDNA is in close proximity with the viral DNA attached to the SPA bead. The β-particles released activate the scintillant in the SPA bead consequently emitting light. When ST is inhibited, the $^3$H-tDNA is not in close proximity with the scintillant and thus its β-particles dissipate in the aqueous solution. Illustration extracted and modified from www.perkinelmer.com, accessed 31/07/2013.[209]

### *1.2.4. Hypothesis*

There is no published data on screening the NIH Clinical Collection I and II for potential HIV-1 IN inhibitors. We hypothesize that screening of the NIH Clinical Collection I and II using an automated HIV-1 IN SPA may identify a possible IN ST inhibitor.

### *1.2.5. Study aims and objectives*

The overall aim of this study is to adapt an HIV-1 IN assay onto an automated platform in order to screen a well-defined compound library for potential HIV-1 IN inhibitors. This will be achieved by completing the following objectives:

1. Expression and purification of recombinant HIV-1 IN proteins (subtype B).

2. Optimization and validation of the HIV-1 IN inhibition SPA
3. Adaption of the optimized HIV-1 IN inhibition SPA to an automated format
4. Biological evaluation of the NCC using the automated HIV-1 IN inhibition SPA.
5. Further development should HIT compounds be identified; including validation in secondary assays, mutation profiles, cell-based assays, ADMET evaluation and molecular docking.

# CHAPTER 2

## Materials and Methods

### 2.1. Recombinant wild type HIV-1 subtype B integrase expression and purification

The HIV-1 IN expression and purification protocol was extracted from Marchand and co-workers[210], Bushman and co-workers[211] and Fish[212] to which several parameters have been modified.

#### 2.1.1. HIV-1 Pol gene induction

Luria Bertani (LB, Laboratorios Conda, Spain) supplemented with 100 microgram per millilitre (µg/ml) ampicillin (AMP, Melford, UK) and 100µg/ml chloramphenicol (Dulfecha Biochemie, Spain) was inoculated with transformed *Eschericia (E) coli* BL 21 pINSD.His (NIH AIDS Research and Reference Reagents Programme; Catalogue number (Cat #) 2957) and incubated overnight shaking at 37°C. The overnight culture was diluted 100 fold in 350ml LB supplemented with 100µg/ml AMP and 100µg/ml chloramphenicol. The culture was grown in a shaking incubator at 37°C and the optical density (OD) thereof was intermittently read at 600nm using the Spectramax Plus384 spectrophometer (Molecular Devices, USA) until an OD of 0.6 was reached. This OD was indicative of the logarithmic phase of the bacterial cell growth. The culture was then induced with 1 millimolar (mM) isopropyl β-D-1-thiogalactopyranoside (IPTG, Thermo Fisher Scientific, USA) and incubated for an additional three hours. The pellets of the induced cell lysate were collected after centrifugation at 3220 x g for 30 minutes at 4°C.

#### 2.1.2. Bacterial cell lysis

The pellets collected after centrifugation of the induced cell lysate were resuspended in Lysis Buffer (10mM $MgCl_2$, 0.25% 3-[(3-Cholamidopropyl)-dimethylammonio]-1-propanesulfonate (CHAPS, Sigma Aldrich, USA), 1mM phenylmethanesulfonylfluoride (PMSF, Thermo Fisher Scientific, USA) and 200 micrograms per microlitre (µg/µl) DNAse (Sigma Aldrich, USA) made up in Binding Buffer, pH 7.2 (1M NaCl, 5mM imidazole, 20mM HEPES, 5% glycerol, 2mM β- Mercapthoethanol (B-Me, Sigma Aldrich, USA). The cell suspension was homogenised until the solution was uniform and stirred at 4 °C for 30 minutes. The homogenous cell solution was sonicated for 1 minute at 75Hz and 0.6 cycles using the Labsonic M ultrasonic processor (Sartorius, Germany). The lysate was then centrifuged at 15 000 x g at 4°C for 30 minutes from which the supernatant was collected.

### 2.1.3. Nickel-Affinity column chromatography

The Protino Nickel (Ni) NTA Affinity column (Macherey Nagel, Germany) used to purify the His-tagged recombinant HIV-1 IN was prepared on the ÄKTA PrimePlus (GE Healthcare, UK). The column was equilibrated with five column washes of Binding Buffer, described in section 2.1.2, at a flow rate of 3ml/minute. The supernatant containing protein was loaded onto the Ni-column through the super-loop at a flow rate of 1ml/minute and the flow through was subsequently collected. The column was washed with 20 column washes of Wash Buffer 1, pH 7.2 (1M NaCl, 20mM HEPES, 60mM imidazole, 10% glycerol, 2mM B-Me) at a flow rate of 3ml/minute followed by 20 column washes of Wash Buffer 2, pH 7.2 (1M NaCl, 20mM HEPES, 150mM imidazole, 10% glycerol, 2mM B-Me). The protein was eluted at a gradient with Wash Buffer 2 as the starting buffer and a set final target of 100% Elution Buffer, pH 7.2 (1M NaCl, 20mM HEPES, 600mM imidazole, 10% glycerol, 2mM B-Me). The elution fractions were collected at a flow rate of 1ml/minute and 3ml per fraction. The eluate containing the protein was pooled and concentrated under liquid nitrogen using an ultra-filtration vacuum system (Merck Millipore, USA) with a 10 000 Da filter membrane.

### 2.1.4. PD-10 Sephadex column

The concentrated protein sample in Elution Buffer was buffer exchanged into Storage Buffer, pH 7.2 (20mM HEPES, 1M NaCl, 4mM EDTA, 2mM dithiothreitol (DTT, Thermo Fisher Scientific, USA), and 50% glycerol) by desalting the protein using a PD-10 Sephadex column (GE Healthcare, UK). The concentration of the purified recombinant IN protein was quantified using the NanoDrop installation version 1.3.1 software on the NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA) and calculated according to its molar extinction coefficient of 50 460 using the equation: $A=\varepsilon.c.l$.

### 2.1.5. Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

Lysate samples were collected during the recombinant expression and purification of IN protocol, and analysed through Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE). The positive control was a recombinant HIV-1 IN protein received from the NIH AIDS Research and Reference Reagents Programme (Cat #9420).[213,214] Samples were diluted 1:1 with Sample Loading Buffer (1M Tris-HCL pH 6.8, 10% glycerol, 10% SDS,

0.05% bromophenol blue, 5% B-Me) and boiled for two minutes. Samples were loaded onto a Mini Protean TGX 10% SDS-PAGE precast gel (Bio-Rad, USA) at a volume of 10µl alongside a Protein Molecular Weight (Mr) Marker (Bio-Rad, USA). The gels were equilibrated using a 1x Tris-Glycine-SDS (TGS) buffer (25mM Tris, 150mM glycine, and 0.1% SDS) and separated at 120 volts. After completion of the run, the gel was stained with Coomassie (0.025% Coomassie Brilliant Blue, 40% methanol and 7% acetic acid) using a staining protocol extracted from Sasse and Gallagher.[215] Following an overnight shaking incubation at room temperature with Coomassie, the gel was destained with Destaining Solution I (40% methanol and 7% acetic acid in $dH_2O$) for approximately two hours followed by an additional destaining with Destaining Solution II (4% methanol and 7% acetic acid in $dH_2O$) for 3-4 hours. Gel images were captured using the gel capture software version 5.8 on the MiniBIS Pro Gel Imager (DNR Bio-Imaging Systems, Israel). The purity of the recombinant IN was determined using the GelQuant 1D gel analysis software version 4.1 (Bio-Imaging Systems, Israel).

### *2.1.6. Western blot analysis of the recombinant HIV-1 integrase*

To verify the expression of the HIV-1 IN protein, a Western blot was conducted using protocols extracted and modified from Bronstein and co-workers[216] and Gallagher and co-workers.[217] SDS-PAGE gels were incubated in Towbin Buffer (20% methanol in TGS Buffer) for 15 minutes at room temperature before transferring to a polyvinylidene difluoride (PVDF) membrane using the iBlot gel transfer stacks (Life Technologies, USA) according to the IBlot Transfer System instructions, (iBlot Transfer System, Life Science Technologies, USA). The PVDF membrane was probed with 1:1000 diluted anti-His.H8 (Abcam, UK) primary antibodies and 1:10 000 diluted rabbit anti-mouse horseradish peroxidase (HRP) linked IgG secondary antibody using the automatic Western blot system (Benchpro 4100, Life Technologies, USA). Briefly, the membrane was blocked with Blocking Buffer (5% Bovine Serum Albumin (BSA) in 1x TBST (420mM Tris pH 7.6, 137mM NaCl, 0.1% Tween 20)) and washed with 1x TBST before and after probing the membrane with primary- and secondary antibody. The recombinant IN protein was detected using Supersignal West Pico Chemiluminescent substrate (1:1 Luminol and peroxidase, Thermo Fisher Scientific, USA) and exposed to X-Ray film for 30 seconds. To further verify the

presence of recombinant HIV-1 IN, the membrane probed with anti-His.H8 was stripped and re-probed with 1:2000 diluted HIV-1 HXB2 IN antiserum (NIH AIDS Research and Reagents Programme; Cat #757)[218] that targets aa 22-34 on the IN protein and 1:10 000 goat anti-rabbit secondary antibody. The membrane was stripped by washing it with 1x TBST thereafter incubated for 15 minutes in Restore Western Blot Stripping Buffer (Thermo Fisher Scientific, USA) and washed 3x for 5 minutes with 1x TBST. This was followed by an overnight incubation at 4ºC with Blocking Buffer. The stripped membrane was re-probed according to the method described above.

## 2.2. Radiolabeling 3'-target DNA with $^{3}$H-dNTPs

### 2.2.1. Lyophilizing $^{3}$H-dNTPs

$^{3}$H-radionucleotides, deoxycytidine 5'- [5-$^{3}$H] triphosphate (dCTP) and deoxythymidine 5'-[methyl-$^{3}$H] triphosphate (dTTP) (American Radiolabeled Chemicals, USA), were lyophilized using the DNA 120 Savant Speed Vac Concentrator (Thermo Fisher Scientific, USA) for approximately 90 minutes until no liquid residue was visible. The lyophilized radionucleotides were reconstituted in 35µl $dH_2O$. The concentration of the $^{3}$H radionucleotides was quantified using the Nanodrop 2000 and calculated using the Beer-Lambert law: $A = \varepsilon Cl$ according to the molar extinction coefficients of $^{3}$H-dCTP and $^{3}$H-dTTP - 9.300 and 9.600, respectively.

### 2.2.2. Klenow fragment 3'-target DNA labelling with 3H-dTTP and 3H-dCTP

The complimentary target oligonucleotides to be radiolabeled, T 56-S (AAAAGGAGGAGAAGGAAAGGAGAGAGAGCGAATTAGCCCTTGGTC) and T 56-A (AAAAGGAGGAGAAGGAAAGGAGAGAGAGGACCAAGGGCTAATTCG) (Inqaba Biotech, South Africa) were mixed in a 1:1 ratio at a final concentration of 50 micromolar (µM) and annealed using the Eppendorf Mastercycler gradient PCR (Eppendorf, USA). The oligonucleotides were annealed under the following conditions: the reaction was heated at 95°C for two minutes then rapidly cooled to 68°C and held at 68°C for ten minutes. The reaction was then cooled down to storage temperature of 4°C over a 90 minute period. The 3'-end of the annealed linear tDNA was radiolabeled by inserting deoxynucleotides (dNTPs) at the 5'-overhangs using the Fermentas Klenow fragment DNA labelling kit (Thermo

Fisher Scientific, USA). The reaction mixture consisted of 5µM linear tDNA, 150 microcurie per millilitre (µCi/ml) $^{3}$H-dCTP, 150µCi/ml $^{3}$H-dTTP, 625µM mixed dNTPs, 1x kit Reaction Buffer, 10 units (U) Klenow fragment enzyme and was made up to a final volume of 20 microlitre (µl) with nuclease free water. The reaction mixture was incubated for four hours at 37°C and stopped by heating it at 75°C for ten minutes.

### 2.2.3. Purifying the radiolabeled target DNA

The QIAquick nucleotide removal kit (Qiagen, Germany) was used to purify unincorporated nucleotides from the $^{3}$H-tDNA. The high-salt concentration premixed kit Binding Buffer PNI was added to the $^{3}$H-tDNA mixture at a 10:1 ratio which was applied to a QIAquick spin column and centrifuged at 3800 x g for one minute. The flow through was discarded and the spin column was washed twice by adding 500µl premixed Buffer PE to the column and centrifuging it for one minute at 3800 x g. The spin column containing the $^{3}$H-tDNA was centrifuged for an additional one minute at 17 900 x g to remove residual ethanol from the premixed kit Buffer PE. The $^{3}$H-tDNA was eluted by adding 100µl $dH_2O$ to the center of the spin column membrane which was left to stand for one minute then centrifuged for one minute at 17 900 x g. The purified $^{3}$H-tDNA concentration was quantified using the NanoDrop 2000 and calculated using the MW of the tDNA (MW = 14136 grams per mole (g/Mol)).

## 2.3. Scintillation proximity assay (SPA)

The fundamental protocol for the SPA was derived from Grobler and co-workers[203] to which several parameters have been modified. All SPA experiments had the following controls: negative control (no IN), positive control (no test/inhibitory compound present) and 10µM RAL (Selleck Chemicals, USA). All test compounds were diluted in dimethyl sulfoxide (DMSO, Sigma Aldrich, USA).

### 2.3.1. SPA optimization

The SPA parameters that were modified include reaction buffers, divalent-metal ion cofactors ($MgCl_2$ or $MnCl_2$) and stop solutions. The reaction buffers were extracted from previously documented protocols described in Chow and co-workers[199] (20mM HEPES pH 7.5, 30mM NaCl, 10mM DTT and 0.5% IGEPAL (Nonidet-P40, Sigma Aldrich, USA), Dicker and co-workers[219] (13.3mM DTT,

32mM MOPS pH 7.0, 0.067% IGEPAL (Nonidet-P40) and 12.8% DMSO), Grobler and co-workers[203] (25mM HEPES pH 7.8, 50mM NaCl, 100µg/ml BSA and 5mM B-Me), and the Chow *et al*. buffer with 100µg/ml BSA. These buffers were further referred to as Buffer 1, Buffer 2, Buffer 3 and Buffer 4, respectively. The stop solutions and its constituents were extracted from Grobler and co-workers[203] and US patent no. 60/422,513.[220] The stop solutions tested were 62mM EDTA, 62mM EDTA with 63mM NaOH and 62mM EDTA with 63mM NaOH and 2.5M CsCl further refered to as stop solution 1,2 and 3, respectively.The optimized assay was validated by conducting dose-response experiments using RAL, EVG (Selleck Chemicals, USA) and 118-D24 (NIH AIDS Research and References Reagents Programme; Cat # 9957).[221,222]

### *2.3.2. Preparation of the streptavidin coated SPA bead complex*

A 10x Reaction Buffer was prepared containing 200mM HEPES (pH 7.5), 300mM NaCl, 50mM DTT and 0.5% IGEPAL (Nonidet-P40). PVT streptavidin coated scintillation beads (GE Healthcare, UK) were reconstituted in 1x Reaction Buffer at a final concentration of 10mg/ml. Pre-processed biotinylated dDNA (U5 top strand, Biotin 5'-ACCCTTTTAGTCAGTGTGGAAAATCTCTAGCA-3' and U5 complimentary bottom strand 5'-ACTCCTAGAGATTTTCCACACTGACTAAAAG-3') was added to the SPA bead suspension at a final concentration of 500nM and rocked at room temperature for one hour. The beads were washed twice with 1x Reaction Buffer by centrifugation at 1000 x g for five minutes. The SPA bead pellet was resuspended in 2x Reaction Buffer (with 20mM $MgCl_2$) to a final concentration of 2mg/ml. Recombinant HIV-1 $IN_{WT}$ was added to the amount of beads required for the experiment at a final concentration of 1µM and rocked for 30 minutes at room temperature.

### *2.3.3. The optimized SPA*

The SPA was carried out in 96-well microtiter plates (Perkin Elmer, USA) comprising of 50µl reactions per well. Each reaction consisted of 25µl of the prepared SPA bead-dDNA-IN complex, 50nM $^3$H-tDNA, 10µM test compound and was made up to 50µl with $dH_2O$. DMSO replaced the test compound in the negative- and positive control. This reaction mixture was incubated at 22°C for 30 minutes while shaking. Thereafter, ST reactions were initiated by adding $^3$H-tDNA to the reaction mixture at a final concentration of 50nM and incubated at 37°C

while shaking for 90 minutes. The enzymatic reaction was stopped with EDTA (without salt), a variable tested in section 2.3.1, at a final concentration of 62mM. The reaction product formation was subsequently measured in counts per minute (CPMs) using the Topcount NXT software version 2.54 on the Top Count scintillation counter NXT (Perkin Elmer, USA).

The optimized SPA was used to screen the NCC compound library (Evotec, USA) and to conduct dose-response studies of cefixime trihydrate (CEF) and epigallocatechin gallate (EGCG). The optimized SPA was also used to screen CEF analogues against $IN_{WT}$. The CEF derivatives include: 7-aminodesacetoxycephalosporanic acid (7-ADCA, Sigma Aldrich, USA), 7-aminocephalosporanic acid (7-ACA, Sigma Aldrich, USA) and cephalothin sodium salt (CSS, Sigma Aldrich, USA). Mutation studies were conducted using the SPA where $IN_{WT}$ was substituted with RAL resistant IN mutants, $IN_{Q148H}$ and $IN_{Q148H/G140S}$ that were previously prepared within our laboratories. DTG (Selleck Chemicals, USA) was used as a control.

### *2.3.4. Handling of the Hamilton STARlet Robotic System*

Extensive training was undertaken on the Hamilton STARlet Robotic System (Hamilton Robotics, Switzerland) to gain experience and become familiar with operating and independent handling of the system. Training involved handling of the robotic system, programming protocols onto the system using the Hamilton method editor software version 4.2.0 7270 as well as troubleshooting of possible errors on the system.

### *2.3.5. Amending the SPA onto the automated system*

A SPA protocol analogous to the optimized bench SPA was programmed on the automated system. To ascertain whether the automated SPA was on par with the manual SPA, dose-response experiments with RAL were conducted using the automated SPA and results thereof were compared with those obtained when conducting the manual SPA.

### *2.3.6. Pooling of the NCC compound library*

In order to reduce cost and time, the NCC compound library was strategically pooled in an orthogonal method where one compound was present in two wells with different compounds (a row and a column pool).[223] The Hamilton STARlet

robotic system was programmed to prepare the pooled compound dilution plates. Compounds were diluted to a final working concentration of 100µM in each pool. Figure 2.12 illustrates the assembly of pools comprised of eight or ten compounds. The eight-compound pools were comprised by combining an entire row of compounds from the NCC 96-well plates into a single well. The ten-compound pools were comprised by combining the compounds of each column from the NCC 96-well plates into a single well. A total of 90 eight-compound pools and 72 ten-compound pools were generated.

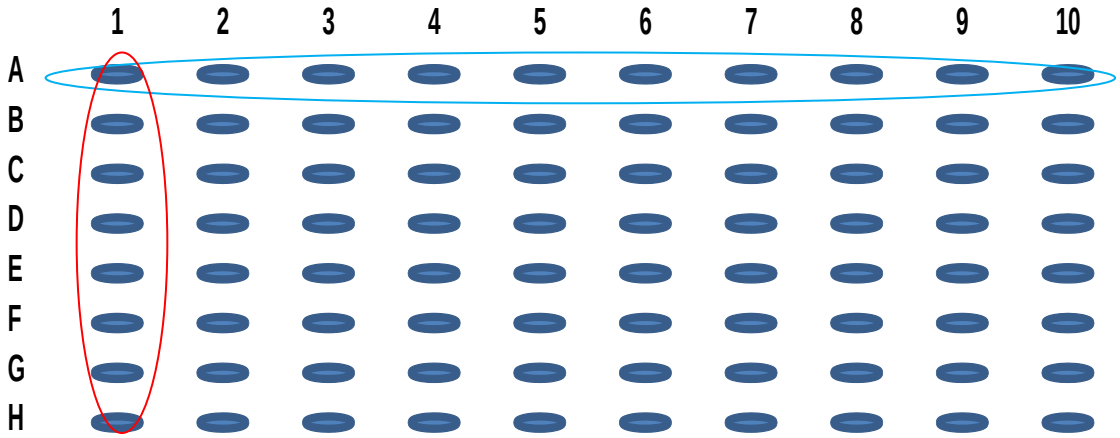



**Figure 2.12: A representative of the NCC 96-well compound plate illustrating the strategic orthogonal pooling of compounds.** Pools were comprised of either eight- or ten compounds. The red oval demarcates the eight-compound pools constituted by combining the compounds of each column into a single well. The blue oval demarcates the ten compounds in a row that were combined to constitute the ten-compound pools.

### *2.3.7. Screening of the NIH Clinical Collection*

The amended automated SPA was used to screen the prepared compound pools. Each automated experiment screened 18 pools (80 compounds). Overall, the entire NCC compound library consisting of 727 compounds was fully screened.

## *2.4. Cell-based assays*

### *2.4.1. Cultivation of mammalian cells*

MT-4 cells (NIH AIDS Research and Reference Reagents Programme; Cat #120) were thawed at 37ºC and harvested in 20% growth media (Roswell Park Memorial Institute Medium 1640 (RPMI, Gibco, USA) containing 205mM L-

glutamine (Thermo Fisher Scientific, USA) and supplemented with 20% inactivated fetal calf serum (FCS, Highveld Biologicals, SA), 20U/ml penicillin-streptomycin (Gibco, USA) and 20µg/µl gentamycin (Gibco, USA). Cells were continuously sub-cultured every 2-3 days by centrifugation of the cell suspension at 360 x g for ten minutes and the subsequent pellet was resuspended in 10% growth media (RPMI 1640 media supplemented with 10% inactivated FCS, 20µg/µl gentamycin and 20U/ml penicillin-streptomycin) for cell counting. Cells were counted by diluting the cell suspension 10x with 0.4% Trypan Blue stain (Gibco, USA) which cannot penetrate viable cells. The diluted cell sample was then added to the Countess Cell Counting Chamber Slide (Life Technologies, USA) and counted on the Countess Automated Cell Counter (Life Technologies, USA). Cells were then resuspended to 2 x $10^7$ cells/ml in 10% growth media and incubated at 37ºC with 5% $CO_2$.

### *2.4.2. Cytotoxicity assays*

The protocol for the cytotoxicity assay was adapted from Mphahlele and co-workers.[224] Following sub-culturing and counting of cells, MT-4 cells were seeded at a final concentration of 1 x $10^5$ cells/ml per well in 96-well plates. The cells were then incubated at 37ºC with 5% $CO_2$ for one hour to stabilize. Dose-response studies of CEF, EGCG and AMP were conducted using 2 fold serial dilutions with final concentrations ranging from 200µM to 1.56µM in 10% growth media. Auranofin (Biomol International, USA) was used as a control. These compound dilutions were added to the stabilized cells and incubated for four days at 37ºC with 5% $CO_2$. The viability of the cells was determined using the Celltiter 96 $AQ_{ueous}$ One Solution Cell proliferation assay (Promega, USA) that contains a tetrazolium dye 3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium (MTS).[225] The MTS solution was added to the cells at a volume of 10µl and incubated for two hours. Viable cells were quantified by measuring the reduction of the yellow tetrazolium dye in the MTS solution by metabolic active live cells to purple coloured soluble formazan. The absorbance of the formazan end-product was measured using the xMark spectrophotometer (Bio-Rad, USA) at 490nm.

### 2.4.3. HIV-1 phenotypic inhibition assay

MT-4 cells were infected with HIV-$1_{NL4-3}$ through spinoculation.[226] Briefly, following sub-culturing of the cells, the resuspended cells were seeded at a volume of 100µl at a working concentration of 2 x $10^5$ cells/ml in a 50ml conical tube. HIV-$1_{NL4-3}$ was added to the cell suspension at a 0.1 multiplicity of infection (MOI). Cells that were not treated with virus were used as the negative control. The cell suspensions were centrifuged for 90 minutes at 1186 x g. The supernatant was discarded and the pellet was washed thrice with 10% growth media and centrifugation at 1186 x g for 10 minutes. The cells were resuspended in 10% growth media at a concentration of 2 x $10^5$ cells/ml and were seeded in 96-well microplates at a final concentration of 1 x $10^5$ cells/well. The seeded cells were then incubated at 37ºC with 5% $CO_2$ for one hour. Compounds were serially diluted to final concentrations ranging from 200µM to 1.56µM in 10% growth medium and added to the cells following the one hour incubation. The treated HIV-1 infected cells were then incubated for four days at 37ºC with 5% $CO_2$. RAL and 118-D-24 were used as positive antiretroviral controls. To determine whether the EGCG, CEF and AMP had antiretroviral activity, a p24 detection assay was conducted using the Vironostika HIV-1 Ab/Ag ELISA kit (Biomerieux, France). Following the four day incubation, 15µl of the supernatant cells was transferred to p24 microplate wells that contained HRP-labelled anti-HIV-1 p24. The reaction mixture was then made up to 150µl with a premixed Specimen Dilution Buffer followed by a one hour incubation period at 37ºC. The reaction mixture was discarded and the wells were washed 7x with a 1x Wash Buffer (25x propriety mixed kit phosphate buffer concentrate diluted in phosphate buffered saline (PBS, 137mM NaCl, 2.7mM KCl, 10mM $Na_2HPO_4$ and 1.8mM $KH_2PO_4$) for 30 seconds per wash. A chromogenic substrate, tetramethylbenzidine (TMB) was then added to the wells. In the presence of the p24 antigen, the HRP enzyme linked to anti-HIV-1 p24 cleaves the TMB substrate subsequently producing a blue colour. The enzymatic reaction was terminated using 10% sulphuric acid and subsequently measured at 450nm using the xMark spectrophotometer.

## 2.5. HIV-1 integrase ELISA

An HIV-1 IN enzyme-linked immunosorbent assay was used as a secondary assay to validate the HITS identified through SPA screening. The protocol for this ELISA was extracted from Mphahlele and co-workers.[224] Stock biotinylated dDNA

(section 2.3.2) was diluted to 150nM with 1x ELISA Buffer 1, pH 7.2 (20mM HEPES, 75mM NaCl, 4µM $ZnCl_2$ and 2% glycerol). The diluted dDNA was added to streptavidin coated microplate wells (R & D Systems, USA) and incubated on a shaker at 25ºC for one hour. The wells were then washed thrice with PBS. Recombinant expressed HIV-1 $IN_{WT}$ (section 2.1) was diluted to a final concentration of 1µM in ELISA Buffer 2 (ELISA buffer 1 with 5mM DTT and 10mM $MgCl_2$). The diluted IN was then added to the microplate wells and incubated at 22ºC for 30 minutes. The negative control did not contain IN. The unbound IN was aspirated and the wells were washed 2x with 1x PBS for five minutes at 25ºC. The compound stock solutions were diluted to a working concentration of 10mM in DMSO. The test compounds were diluted to 100µM or 10µM in ELISA Buffer 2 and added to the wells followed by a 30 minute incubation period at 37ºC. DMSO substituted the test compounds in the negative- and positive control. Annealed tDNA (5'-TGACCAAGGGCTAATTCACT-3' fluorescein and 5'- AGTGAATTAGCCCCTTGGTCA-3' fluorescein) (Inqaba biotech, South Africa) was added to the wells at a final concentration of 250nM and the mixture was incubated for an additional hour at 37ºC. The tDNA and unbound compounds were aspirated and the wells were then washed 3x with 2x saline sodium citrate (SSC) buffer pH 7 (0.15M $Na_3C_6H_5O_7$ and 1.5M NaCl) for 10 minutes at 25ºC. Rabbit anti-fluorescein isothiocyanate (FITC): HRP antibody (AbD Serotech, UK) was diluted 1:10 000 in 1x PBS and added to each well. The wells containing the antibodies were incubated for 1-2 hours at 22ºC and were transferred to 4ºC for an overnight incubation. Following the overnight incubation, the antibodies were aspirated and the wells were washed 3x with 1x PBS for 10 minutes at 22ºC. Sureblue TMB microwell peroxidase substrate (KPL, USA) was added to each well and the reaction was incubated at 37ºC for 15-30 minutes. In the presence of an HRP labelled conjugate, the reaction turns blue which was measured at 620nm using the xMark spectrophotometer.

## *2.6. Reverse transcriptase activity colorimetric assay*

The effect of CEF, EGCG and AMP on RT activity was determined using the colorimetric Reverse Transcriptase Assay kit (Roche Diagnostics, Switzerland). The assay is based on a typical sandwich ELISA method and was carried out according to the kit instructions. All reagents used in the assay were provided with the kit. The reagents in the kit were prepared as follows: the nucleotide

mixture was prepared by diluting the nucleotide, digoxigenin (DIG)-dUTP, biotin-dUTP and dTTP, 10x with Incubation Buffer, pH 7.8 (50Mm Tris Buffer with 319mM KCl, 33mM $MgCl_2$ and 11mM DTT). The reaction mixture was prepared by adding the primer template to the 10x prepared nucleotides at a final concentration of 0.9 $A_{260}$/ml. The lyophilized anti-DIG-peroxidase (anti-DIG-POD) was prepared by reconstituting the antibody in $dH_2O$ to a final concentration of 20U. The working dilution of anti-DIG-POD was prepared by diluting the reconstituted antibody to a final concentration of 200mU/ml with Conjugate Dilution Buffer (sodium phosphate buffer pH 7.4 containing a blocking reagent). The 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid) (ABTS) substrate was prepared by dissolving one ABTS tablet in Substrate Buffer ($NaBO3•nH_2O$ and citric acid/phosphate buffer). Sample mixtures were prepared in separate tubes for each test compound at a final volume of 60µl. Each sample comprised 5 nanograms (ng) HIV-1 RT, 100µM or 10µM test compound, 20µl reaction mixture and the sample mixture was made up to 60µl with RT Lysis Buffer pH 7.8 (50mM Tris, 80mM KCl, 2.5mM DTT, 0.75mM EDTA and 0.5% Triton X-100). The test compound in the positive control was replaced with RT Lysis Buffer. The negative control did not contain RT enzyme and was replaced with RT Lysis Buffer. AZT-triphosphate (TP) (GeneCraft, UK) was used as a control compound. The sample mixture was then incubated at 37ºC for one hour. After the incubation period, the sample mixtures were transferred to the streptavidin coated microplate modules and then further incubated for one hour at 37ºC. The sample mixtures were then removed from the wells followed by washing of the wells with the premixed kit Washing Buffer 5x for 30 seconds per wash at room temperature. The anti-DIG-POD was added to each well at a working dilution of 200mU/ml and incubated for one hour at 37ºC. The wells were then washed 5x with the premixed kit Washing Buffer for 30 seconds per wash at room temperature. The ABST substrate was added to each well at a final volume of 200µl and incubated at room temperature for 15 minutes. The absorbance of the green/blue colour was measured at 405nm.

## 2.7. Virtual screening of cefixime trihydrate and epigallocatechin gallate

### 2.7.1. In silico evaluation of some ADME descriptors of cefixime trihydrate and epigallocatechin gallate

The structures of CEF and EGCG were obtained from the standard database format (SDF) file on www.nihclinicalcollection.com for ADME studies using Osiris Property Explorer and Discovery Studio Software.[154] The online software, Osiris Property Explorer (www.organic-chemistry.org/prog/peo/), was used to calculate drug-relevant properties such as the logarithm partition coefficient (cLogP), aqueous solubility (LogS) and drug score. The cLogP of a compound is a measure of the hydrophilicity of a compound. High cLogP values are indicative of low hydrophilicity and are associated with poor absorption and permeation. The LogS of a compound modulate the absorption and distribution of a compound where low solubility compounds are associated with poor absorption. The drug score of a compound is a single value that describes the overall potential of a compound as a drug. ADME descriptors such as absorption (HIA and aqueous solubility) and distribution (BBB penetration) properties of CEF and EGCG were predicted using the ADME Descriptor tool in the Discovery Studio version 3.1 software (Accelrys, USA). The cLogP of CEF and EGCG used in Lipinski Ro5 characterizations were retrieved from Osiris Property Explorer. The MW of the compounds and hydrogen donors and acceptors were retrieved from the SDF file on www.nihclinicalcollection.com.

### 2.7.2. Molecular modelling of cefixime trihydrate and epigallocatechin gallate

An HIV-1 IN monomer structure attached to viral DNA was previously prepared within our laboratory and was used in molecular docking simulations using Discovery Studio software version 3.1 (Accelrys, USA). The protein was automatically prepared for docking by inserting missing atoms, modifying missing loop regions and removing disorderly conformations. A binding site within a protein cavity defining a specific area where binding interactions may occur was created by selecting residues within a 7 angstrom (Å) radius surrounding the $Mg^{2+}$. The small molecules of the NCC library were prepared using the Prepare Ligand Tool that generated isomers and tautomers and removed duplicate ligand

structures and ligands with undesirable properties. The structures of the control compounds RAL, EVG and DTG were obtained from the protein data bank (www.rcsb.org/pdb) and prepared using the prepare ligand tool. The LibDock tool was used to generate multiple conformations of the prepared ligands using the BEST conformation method. These conformations were then docked by locating the hotspots in the receptor binding site consequently generating numerous poses per ligand conformation. Poses with the best LibDock scores were selected for further optimization. The grid based CDOCKER tool that utilizes a CHARMm force field was used to refine the poses of the ligand by keeping the receptor binding site rigid whilst the ligands are more flexible. Thereafter, the scores of the refined poses, obtained from the CDOCKER, were calculated to select the best representative score that is indicative of the receptor-ligand interactions.

## 2.8. Data analysis

Microsoft Excel 2010 was used to calculate all averages and standard deviations (SD) as well as to construct bar- and pie graphs. Percentage inhibition was calculated using the following equation:

$$\% \text{ inhibition} = \left(1 - \frac{test_I - neg}{pos - neg}\right) * 100$$

Where $test_I$ represents the endpoint value of the inhibitor, neg the endpoint value of the negative control and pos the endpoint value of the positive control. *p*-values were calculated to assess statistical differences within parameters and assay conditions. This was done using the online pairwise t-test (www.quantitativeskills.com/sisa/). Origin software version 6.1 (OriginLab, USA) was used to calculate the $IC_{50}$, cytotoxicity change at 50% cytotoxic concentration ($CC_{50}$) and the half maximal effective concentration ($EC_{50}$) of the compounds. Origin software was also used to construct sigmoidal curves and to calculate the Hill-Slope of the curves. The $FCIC_{50}$ values of inhibitors in mutation studies were calculated by dividing the $IC_{50}$ of an inhibitor against a mutant integrase by the $IC_{50}$ of an inhibitor against WT integrase. The Z-factor, used to describe the quality of the SPA, was calculated using an online calculator (www.screeningunit-fmp.net/tools/z-prime.php).

## 3.1. Expression and purification of recombinant wild type HIV-1 subtype B integrase

Recombinant HIV-1 $IN_{WT}$ was successfully expressed in *E.coli* after a three hour incubation at 37°C when induced with 1mM IPTG. Recombinant expressed $IN_{WT}$ was verified on an SDS-PAGE (Figure 3.13) and compared to a positive- and negative control. The positive control was a recombinant HIV-1 IN received from the NIH while the negative control comprised *E.coli* BL21 pINSD.His that was not induced with 1mM IPTG. The expressed $IN_{WT}$ was observed at its expected size of approximately 32kDa, which also corresponded with the positive HIV-1 IN control. No expressed $IN_{WT}$ protein was observed in the negative control.

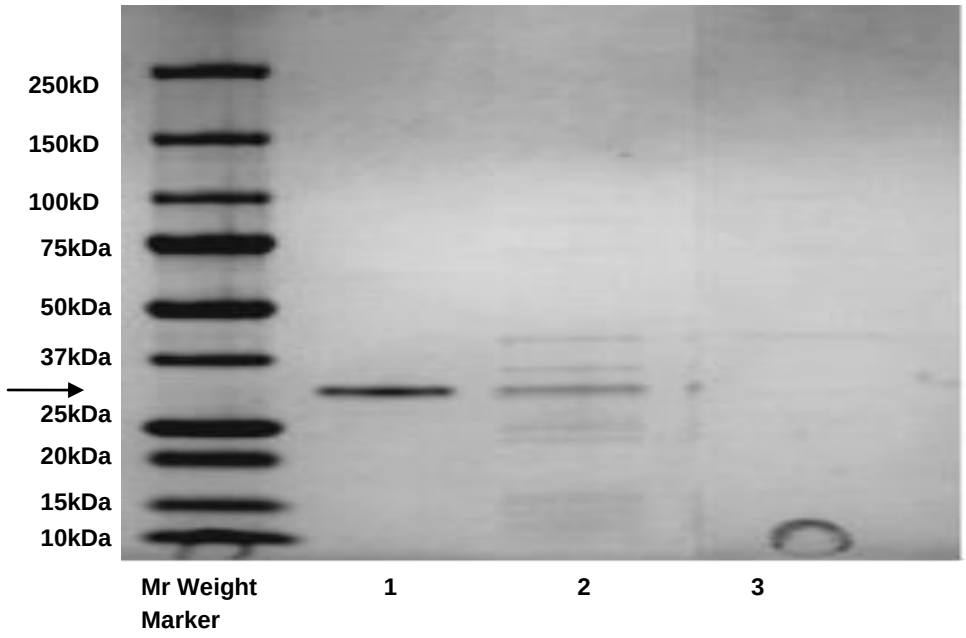



**Figure 3.13: SDS-PAGE depicting the pINSD.His gene expression in *E. coli* BL21 after induction with 1mM IPTG.** The samples were analysed alongside Protein Mr Weight Markers (Bio-Rad, USA) that aid in determining the size of the expressed recombinant HIV-1 integrase. The SDS-PAGE was visualized using a Coomassie stain and the image was captured on the MiniBIS Pro gel capture imager. Lane 1 contained the positive control; Lane 2 contained the recombinant proteins expressed after induction with 1mM IPTG where the arrow indicates the protein that was predominantly expressed at the estimated size of 32kDa; Lane 3 contained the negative control (*E.coli* BL21 pINSD.His cells that were not induced with 1mM IPTG) where no bands were prominently expressed. Molecular weights in kDa, are indicated on the left.

After clarification of the 1mM IPTG induced crude extract, the pellet was discarded because it contained insoluble proteins and was therefore not used for further purification (data not shown). The supernatant, expected to contain the soluble recombinant $IN_{WT,}$ was loaded onto the Ni-affinity column for purification. Figure 3.14 illustrates the elution profile of $IN_{WT}$ from the Ni-affinity column at a gradient where Wash Buffer 2 (150mM imidazole) was the starting buffer and Elution Buffer (600mM imidazole) was the final buffer. The peak observed between fraction 14 and 23 is representative of the eluted $IN_{WT.}$

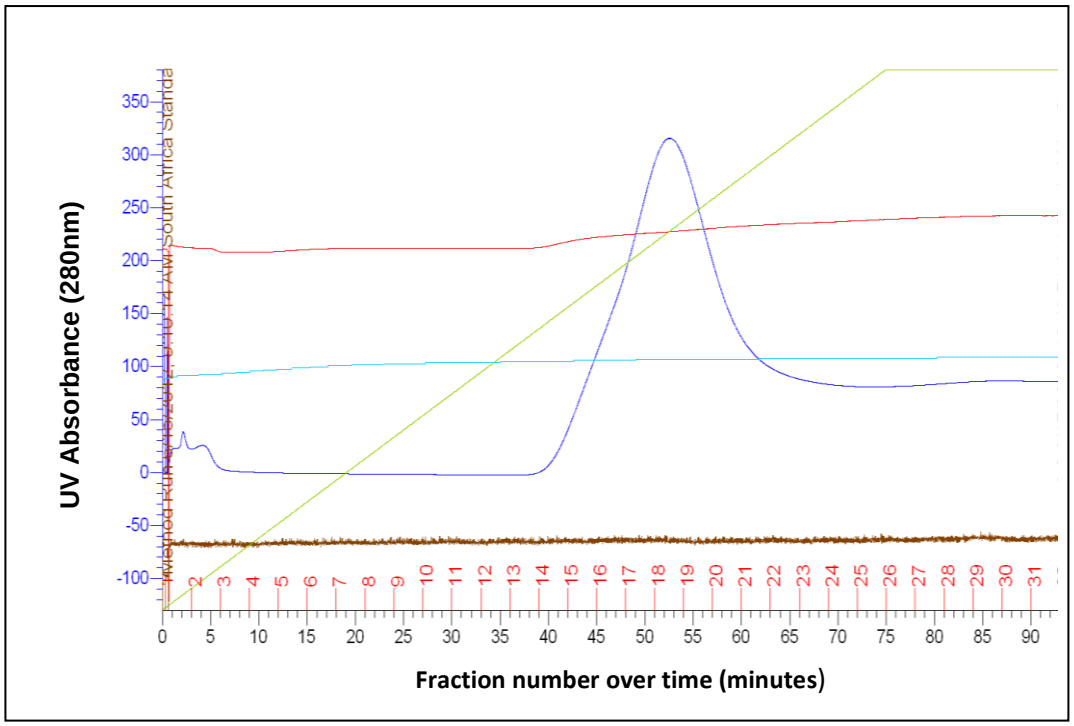
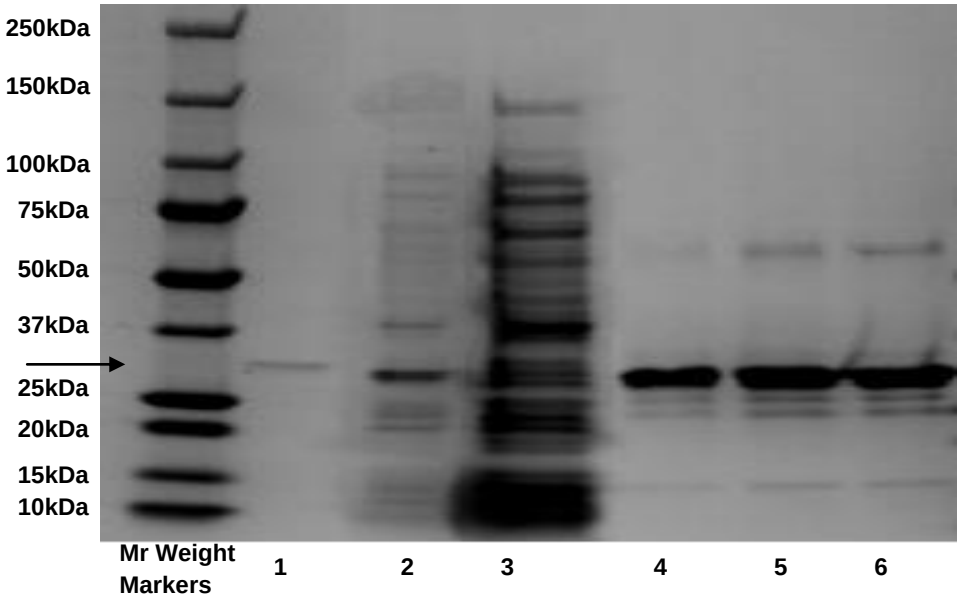



**Figure 3.14: Elution profile depicting the elution of the His-tagged proteins from the Ni-affinity column.** The blue, green, red, light blue and brown lines denote UV absorbance, imidazole concentration, conductivity, temperature and pressure, respectively. The elution profile, extracted from the ÄKTA Primeview 2 software, exemplified the elution of His-tagged proteins by measuring the UV absorbance of proteins per fraction (3ml/fraction) at a flow rate of 1ml/min for 3 minute intervals. A peak comprised of fractions 14 to 22 was observed during the elution period between 39- and 63 minutes.

The fractions (14-22), corresponding to the peak (Figure 3.14), were pooled and separated from other buffer components through ultra-filtration in order to increase the protein concentration. Recombinant expressed His-tagged protein samples were collected after each purification and concentration step to analyse the purity and intensity thereof through SDS-PAGE, as illustrated in Figure 3.15.

## *Results*

The recombinant expressed proteins were observed at approximately 32kDa which aligned with the positive control. The intensity of the bands depicted by the arrow in Figure 3.15, increased after Ni-affinity chromatography, ultra-filtration and PD-10 Sephadex size exclusion chromatography where each band demonstrated intensity of 145 298, 142 0994 and 196 839 after each step, respectively. The percentage purity of the recombinant proteins observed at approximately 32kDa was 73%, 70% and 77% after Ni-affinity chromatography, ultra filtration and size exclusion chromatography, respectively. Additional bands were observed between 50-75kDa.




| Lane | 4 | 5 | 6 |
|---|---|---|---|
| Intensity | 145 298 | 142 0994 | 196 839 |
| Purity (%) | 73 | 70 | 77 |

**Figure 3.15: SDS-PAGE depicting the purity and intensity of the recombinant expressed proteins.** Samples were analysed alongside Protein Mr Weight Markers and verified using a recombinant HIV-1 integrase (IN) as a positive control. The gel was stained with Coomassie stain and visualized on the gel capture imager. Lane 1 contained the positive control; Lane 2 contained the *E.coli* BL21 pINSD.His after induction with 1mM IPTG; Lane 3 contained the crude extract after centrifugation; Lane 4 contained the fractions eluted from the Ni-affinity column; Lane 5 contained the concentrated His-tagged proteins using the Millipore ultra-filtration vacuum; Lane 6 contained the His-tagged protein after it was buffer exchanged into Storage Buffer using the PD-10 Sephadex column. The arrow denotes that the over-expressed recombinant protein was at the size (32kDa) at which $IN_{WT}$ is expected. Additional protein bands were observed in lanes 4-6 between 50-75kDa. Molecular weights, in kDa are indicated on the left.

The presence of the recombinant His tagged protein was confirmed by a Western blot probed against the His-tag (using anti-His primary antibodies) (Figure 3.16 A) on the recombinant expressed $IN_{WT}$. The presence of recombinant $IN_{WT}$ was further verified by stripping and re-probing the PVDF membrane with HIV-1 HXB2 IN antiserum (Figure 3.16 B). The bands observed on the Western blot probed with anti-His primary antibodies correlated with the Western blot probed with IN antiserum. The bands detected also correlated with the HIV-1 IN positive control.

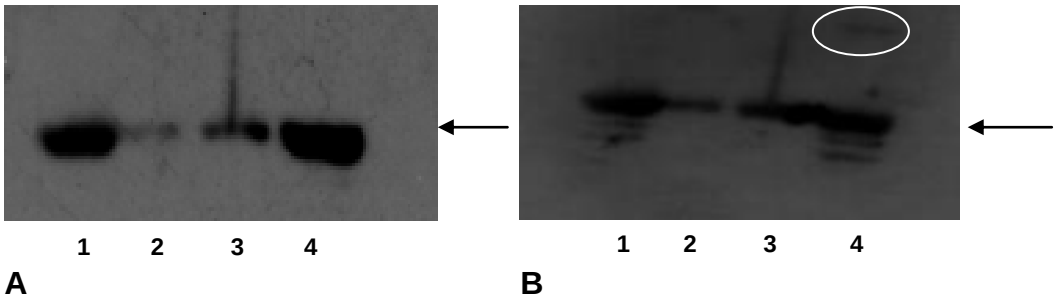



**Figure 3.16: Western blot probed with A - anti-His and B - anti-integrase primary antibodies verifying the expression of $IN_{WT}$.** The Western blot images were visualized by means of chemiluminescence. The presence of recombinant expressed $IN_{WT}$ was verified using a positive control HIV-1 integrase (IN). The presence of the recombinant expressed $IN_{WT}$ was detected using antibodies against the His-tag on the recombinant expressed $IN_{WT}$ as well as antibodies against the HIV-1 IN protein itself. Lane 1 contained the positive control; Lane 2 contained the pooled fractions (14-23) eluted from the Ni-affinity column; Lane 3 contained the concentrated $IN_{WT}$ through ultra-filtration; Lane 4 contained $IN_{WT}$ after it had been buffer exchanged into Storage Buffer using the PD-10 Sephadex column. An additional band was observed that is demarcated with a white circle while the arrows denote recombinant $IN_{WT}$ and positive control.

## *3.2. SPA screening*

### *3.2.1. Purification of the radiolabeled ($^3H$) target DNA*

Radiolabeled tDNA, $^3$H-tDNA, was purified to remove unincorporated nucleotides. The absorbance spectrum of the $^3$H-tDNA was extracted from the Nanodrop installation version 1.3.1 software where Figure 3.17 exemplifies the absorbance spectrum. A peak was observed at 260nm, the wavelength at which pyrimidines and purines of nucleic acids absorbs UV light. The purified DNA yielded a nucleic absorbance ratio of $A_{260}/A_{280} = 1.85$ which is in the pure DNA ratio range.

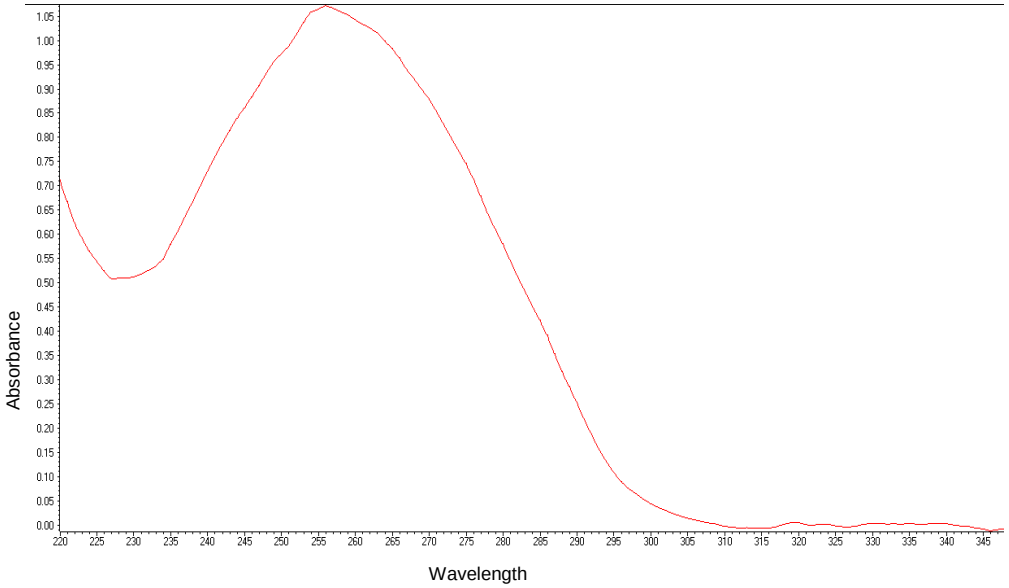



**Figure 3.17: The absorbance spectra of the $^{3}$H-tDNA extracted from the Nanodrop installation version 1.3.1 software obtained when quantifying the $^{3}$H-tDNA concentration using the Nanodrop spectrophotometer.** Absorbance was read at a wavelength ranging from 220nm-345nm where a peak was observed at 260nm with a maximum < 1.05.

### *3.2.2. Optimization of the HIV-1 integrase SPA*

The fundamental SPA protocol was extracted from Grobler and co-workers[203] where several parameters were modified such as the reaction buffer, metal cofactors and stop solutions. The optimal conditions were determined using the percentage ST inhibition attributed to RAL, the CPM signal and the Z-factor of the assay.

The Z-factor of an assay is usually within a $-1 < Z \leq 1$ range. When there is no data variation (SD = 0), and Z = 1, the assay is termed "ideal". When the separation band is large ($1 > Z \geq 0.5$), the assay is termed "excellent". The assay is termed "marginal" when the separation band is low ($0.5 > Z > 0$). When there is no separation band between the positive- and negative control, the assay is not useful for screening.[227]

Since the catalytic activity of HIV-1 IN is modulated by metal cofactors in the DDE motif of the IN active site, the optimal metal cofactor was determined by conducting experiments with $MgCl_2$ and $MnCl_2$. When comparing the CPM signal of the postive controls when conducting experiments using $MgCl_2$ and $MnCl_2$, $MgCl_2$ rendered a higher CPM signal (97.5 ± 0.7) compared to experiments conducted with $MnCl_2$ (78 ± 17.1). When comparing IN ST inhibition due to RAL,

experiments using $MgCl_2$ demonstrated a significantly higher inhibition (65 ± 15%, $p$ = 0.02) compared to experiments conducted with $MnCl_2$ (31 ± 8.5%). Figure 3.18 compares the IN ST inhibition attributed to RAL when experiments were conducted using $MgCl_2$ and $MnCl_2$. The Z-factor of the experiments, when using $MgCl_2$ and $MnCl_2$, were 0.952 and 0.497, respectively. $MgCl_2$ was considered as the optimal metal chelator as it rendered the highest IN inhibition and the higher CPM signals and had the highest Z-factor when compared to $MnCl_2$.

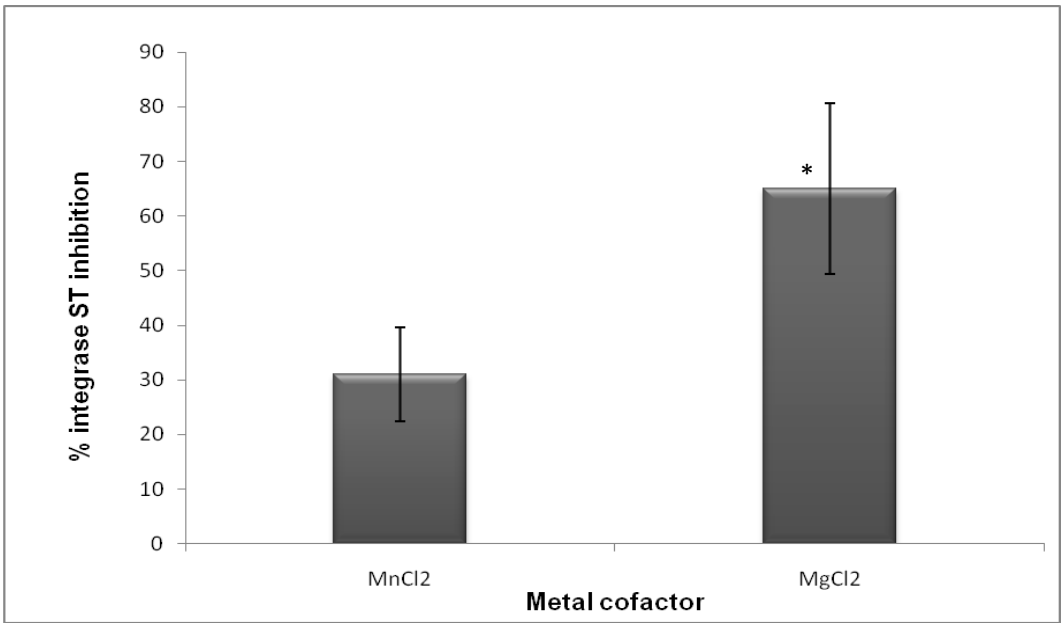



**Figure 3.18: Comparison of the percentage integrase (IN) strand transfer (ST) inhibition caused by raltegravir (RAL) when conducting experiments with metal cofactors, $MgCl_2$ and $MnCl_2$.** SPA experiments using $MgCl_2$ demonstrated 65 ± 15% IN ST by RAL where experiments using $MnCl_2$ only inhibited IN ST by 31 ± 8.5%. The asterisks indicate that the % IN ST in the presence of $MgCl_2$ is significantly different from that of $MnCl_2$. The error bars indicate the SD of the mean average of at least three experiments.

To increase the CPM signal of the SPA, several reaction buffers, obtained from literature, were tested. The highest CPM signal was observed when using Buffer 4 (278 ± 91.85), however the Z-factor (-0.277) was the lowest when compared to the Z-factor of the experiments conducted using Buffer 1, 2 and 3, respectively. The highest IN ST inhibition was observed when using Buffer 2 (85 ± 10%) which also rendered the highest Z-factor (0.445).

Table 3.1 compared the reaction buffers and its effect on the CPM signal, the IN ST inhibition attributed to RAL as well as the Z-factor of the assay. The highest

CPM signal was observed when using Buffer 4 (278 ± 91.85), however the Z-factor (-0.277) was the lowest when compared to the Z-factor of the experiments conducted using Buffer 1, 2 and 3, respectively. The highest IN ST inhibition was observed when using Buffer 2 (85 ± 10%) which also rendered the highest Z-factor (0.445).

**Table 3.1: Comparison of the different reaction buffers tested in the SPA and its effect on the CPM signal, integrase strand transfer inhibition and the Z-factor of the assay.**

| Buffer | CPM signal ± SD (Positive control) | CPM signal ± SD (Positive control) | Percentage integrase strand transfer inhibition ± SD (Raltegravir) | Z-factor |
|---|---|---|---|---|
| **1** | 83 ± 19 | 11 ± 1.41 | 65 ± 0.17 | 0.22 |
| **2** | 141 ± 49 | 13 ± 4.95 | 85 ± 10 | 0.45 |
| **3** | 92 ± 7.41 | 17 ± 9.59 | 78 ± 12.2 | 0.20 |
| **4** | 278 ± 91.85 | 21 ± 8.88 | 70 ± 4.3 | -0.28 |

The optimal stop solution was determined by terminating the reaction between IN and $^{3}$H-tDNA with different stop solutions. When comparing the CPM signal of the positive controls after the addition of the stop solutions, Stop Solution 3 demonstrated the lowest positive control CPM signal (305 ± 29) as opposed to reactions stopped with Stop Solution 2 that had the highest CPM signal (589 ± 39). The reaction terminated with Stop Solution 2 yielded a CPM signal of 484.5 ± 16) for the positive control. RAL inhibited IN ST by 88 ± 4%, 74 ± 2% and 56 ± 17% when the SPA enzymatic reactions were stopped with Stop Solution 1, 2 and 3, respectively (Figure 3.19). RAL demonstrated no appreciable IN ST inhibition (0.209 ± 0.03%) when the reaction was not stopped.

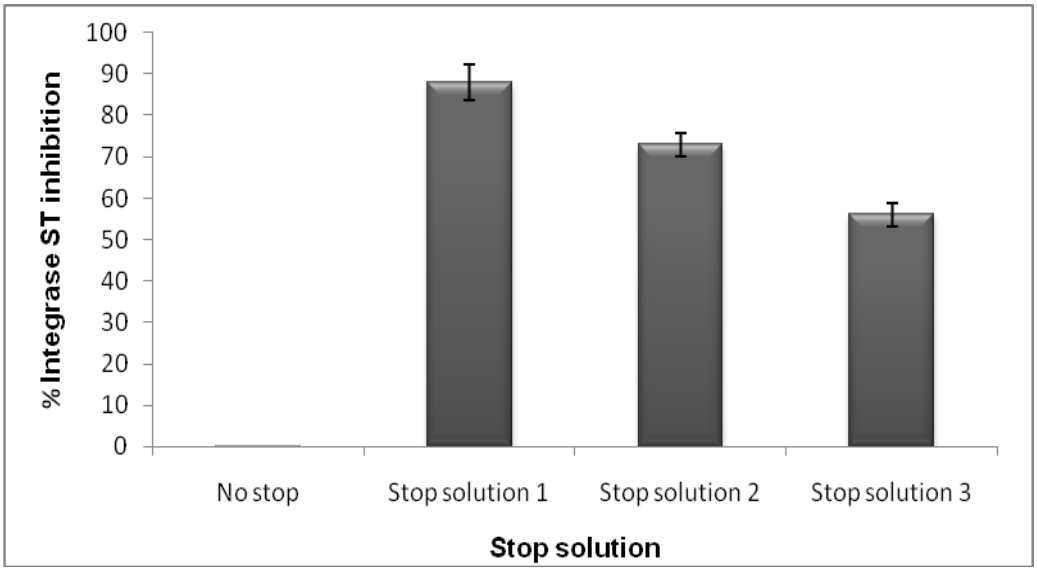



**Figure 3.19: Comparison of the integrase (IN) strand transfer (ST) inhibition by raltegravir (RAL) after the enzymatic reaction between IN and $^{3}$H-tDNA was terminated with different stop solutions:** Stop Solution 1 (62mM EDTA), Stop Solution 2 (62mM EDTA with 63mM NaOH) and Stop Solution 3 (62mM EDTA with 63mM NaOH and 2.5M CsCl). The highest IN ST inhibition was observed when stopping the reaction with Stop Solution 1 (88 ± 4%) followed by Stop Solution 2 (73 ± 3%) and Stop Solution 3 (56 ± 3%). The reaction that was not stopped demonstrated no appreciable IN ST inhibition (0.209 ± 0.03%). Error bars indicate the standard deviation within 5% of the mean of triplicate data.

### *3.2.3. Validating the optimized SPA with known integrase inhibitors*

The optimized SPA was validated using known controls and inhibitors such as RAL, EVG, 118-D24 and AZT. Dose response studies of RAL, EVG and 118-D24 were conducted subsequently determining their $IC_{50}$ values. Figure 3.20 represents the dose-dependent response of RAL, EVG and 118-D24. The $IC_{50}$ values obtained were 9.98 ± 0.83nM, 4 ± 1.04nM and 1.50 ± 0.81µM for RAL, EVG and 118-D24, respectively and were compared to literature to validate the authenticity of the optimized SPA. The results obtained correlated with documented $IC_{50}$ values of RAL, EVG and 118-D24 which have been reported as 7nM, 8.80nM and 1.53µM, respectively.[110,222,228] AZT was used as the negative control where no IN inhibition was observed. The Z-factor of this experiment, 0.93, further validated the quality of the SPA.

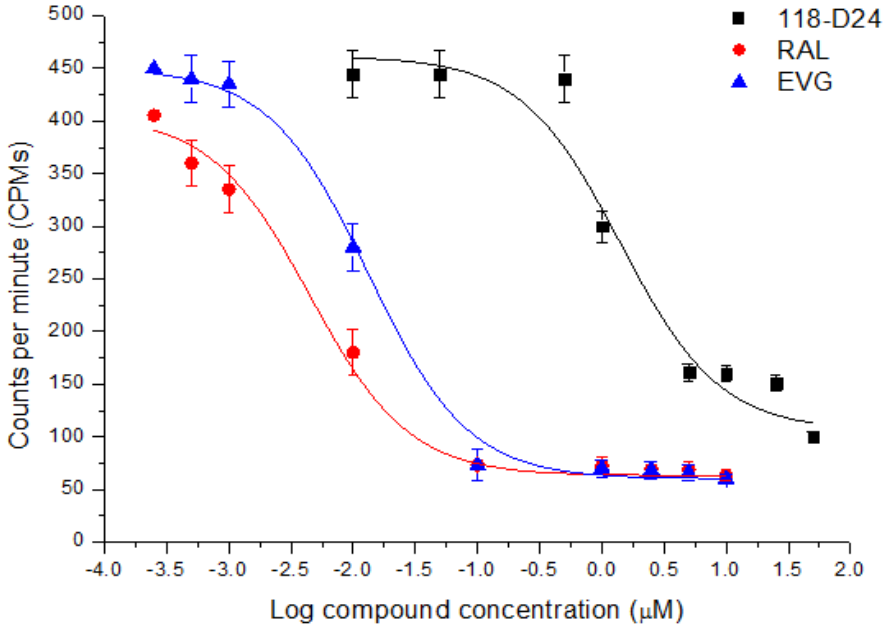



**Figure 3.20: Representative sigmoidal curves of raltegravir (RAL), elvitegravir (EVG) and 118-D24 obtained when conducting dose-response studies using an integrase inhibition SPA.** Each point indicates the normalized average mean of triplicate results and its standard deviation within 5% of the mean. The Hill-Slope variable describes the slope of the curve. The sigmoidal curve of 118-D24 demonstrates a standard sigmoid dose-response curve with a Hill-Slope close to -1. The slope of RAL and EVG are more shallow than the slope of 118-D24. The area under curve for RAL and EVG decreased with an increase in $IC_{50}$.

### *3.2.4. HIV-1 integrase SPA amended to the automated system*

A SPA protocol analogous to the optimized SPA was programmed onto the Hamilton Robotic STARlet System. The automated assay was validated by conducting a dose-response study using RAL and comparing the $IC_{50}$ of RAL obtained when using the automated assay to the RAL $IC_{50}$ obtained when using the manual assay. Figure 3.21 compares the sigmoidal curve of RAL obtained when conducting the automated SPA and when conducting the manual SPA. The $IC_{50}$ of RAL obtained when using the automated SPA was 7.30 ± 0.35nM which was not significantly different from the $IC_{50}$ of RAL when using the manual SPA ($p$ = 0.99). It was also observed that the Z-factor obtained when conducting the manual SPA (0.96) was slightly higher than the Z-factor obtained when conducting the automated SPA (0.88), however not significantly different ($p$ = 0.85).

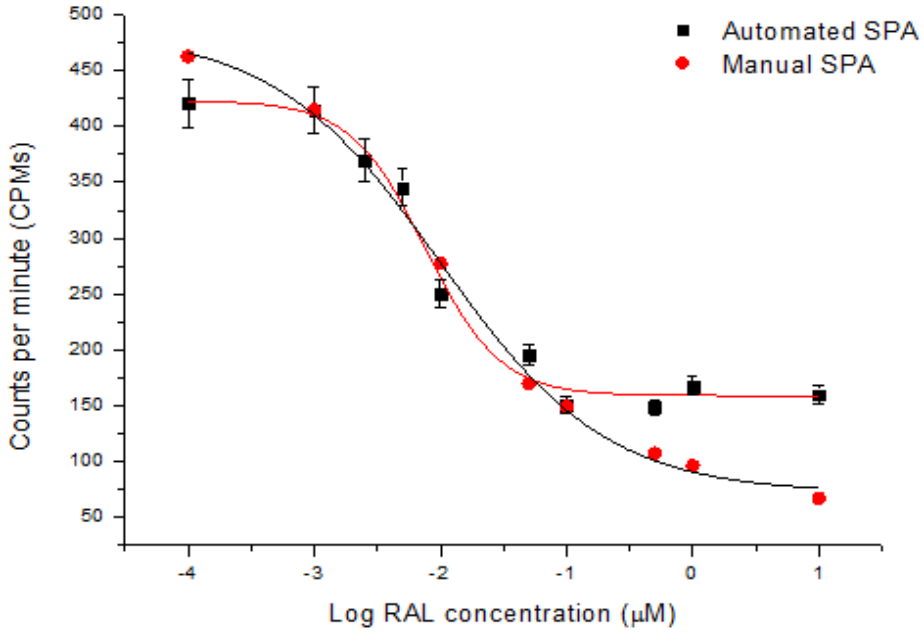



**Figure 3.21: The sigmoidal curves depict the comparison between the raltegravir (RAL) dose-dependent response (0.001µM to 10µM) when using the automated SPA and the manual SPA.** The $IC_{50}$ of RAL when using the automated SPA was 7.3 ± 0.35nM which correlated with the $IC_{50}$ = 9.98 ± 0.83nM ($p$ = 0.99) obtained when using the manual SPA. The Hill-Slope for the automated SPA was -0.6 whereas the manual SPA demonstrated a Hill-Slope of -1.46 indicating that the manual SPA curve is steeper with a higher $IC_{50}$ value than the sigmoidal curve for the automated SPA.

### *3.2.5. Screening of the NCC library*

Before the compounds in the NCC compound library were screened, they were diluted to a final working concentration of 100µM in their respective pools with DMSO. To validate whether the pooling method would be suitable in a SPA, 10µM of RAL was added to two random pools containing eight and ten compounds, respectively. Pools that included 10µM RAL were compared to pools without RAL to determine if RAL in the presence of eight or ten other compounds inhibited IN ST. RAL maintained its inhibitory effect by inhibiting IN ST by 89 ± 4% within the eight-compound pool and 90 ± 3% within the ten-compound pool.

The entire NCC was evaluated by screening one NCC plate, containing 80 compounds (18 pools), per experiment. A total of 171 pools comprising 727 compounds in the NCC compound library were tested. Figure 3.22 is a summation of the percentage IN ST inhibition of each compound pool. Amongst the 171 pools, 136 pools demonstrated 0-10% IN ST inhibition, six pools

demonstrated 10-50% IN ST, three pools demonstrated 50-80% IN ST inhibition and one pool yielded IN ST inhibition > 80%. The remaining 25 pools were undefined due to the colour of the compounds that may have influenced the CPM signal thus resulting in false positives.

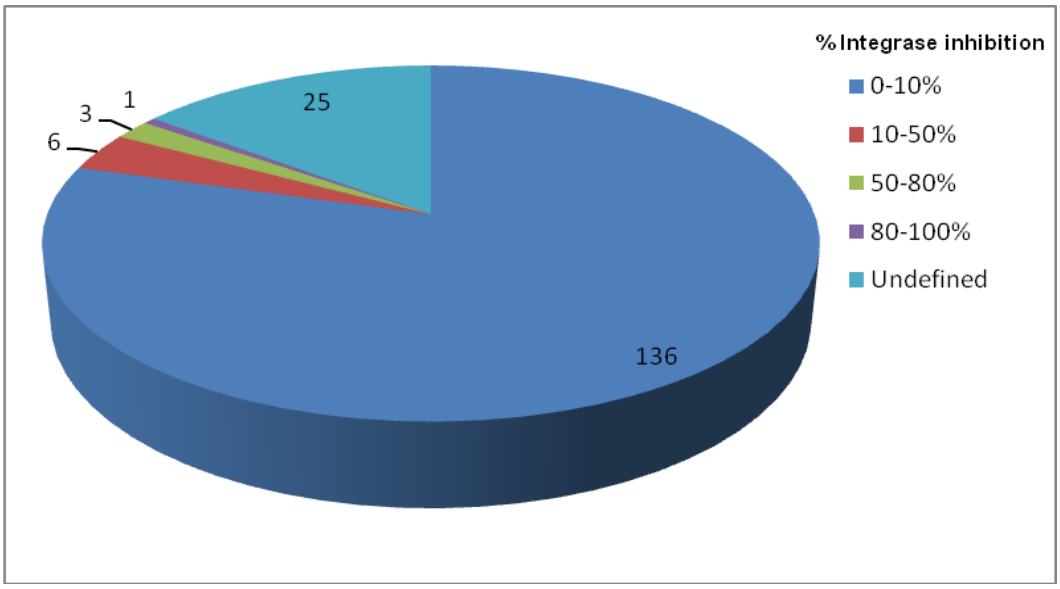



**Figure 3.22: Pie chart representation of the screened NCC pools according to their percentage integrase (IN) strand transfer (ST) inhibition.** A total of 171 pooled compounds were screened in an orthogonal method where each compound was present in at least two pools. Amongst the 171 pools, 136 had approximately 0-10% IN ST inhibition compared to only one pool that demonstrated IN ST inhibition exceeding 80%. Furthermore, only six pools showed IN ST inhibition between 10-50% and three pools demonstrated IN ST inhibition between 50-80%. Compounds that influenced the CPM signal through quenching were deemed undefined.

The compounds comprising pools that demonstrated IN ST inhibition > 50%, were tested individually on the SPA to identify the compound responsible for the IN ST inhibition. A Cephalosporin Beta (β)-lactam antibiotic, CEF, and an antioxidant found in Green Tea, EGCG, were identified as the compounds that attributed to the IN ST inhibition.

### *3.2.6. Dose response studies of cefixime trihydrate and epigallocatechin gallate*

Dose-response studies of CEF and EGCG were conducted to determine the concentration of compound required for 50% IN ST inhibition. $IC_{50}$ curves were constructed using the optimized SPA and the $IC_{50}$ values of CEF and EGCG

against $IN_{WT}$ were determined. The $IC_{50}$ values were 6.03 ± 1.29µM and 9.57 ± 1.62µM for CEF and EGCG, respectively, as depicted in Figure 3.24.

### 3.2.7. Integrase resistant mutation profiles of cefixime trihydrate and epigallocatechin gallate

The activity of $IN_{WT}$ and RAL induced mutants, $IN_{Q148H}$ and $IN_{Q148H/G140S}$, was determined using the optimized SPA. The activity of the enzyme was determined using the CPM signal of the positive control for $IN_{WT}$, $IN_{Q148H}$ and $IN_{Q148H/G140S}$. The activity of $IN_{Q148H}$ decreased 6 fold (CPMs for $IN_{WT}$ = 220 and $IN_{Q148H}$ = 34 with background CPM = 18) whereas the activity of $IN_{Q148H/G140S}$ decreased 2.6 fold ($IN_{WT}$ = 311 and $IN_{Q148H/G140S}$ = 117 with background CPM = 15). RAL, EVG and DTG were used as controls against $IN_{Q148H/G140S}$. The sigmoidal curves of the RAL, EVG and DTG dose-response studies are demonstrated in Figure 3.23. $IN_{Q148H/G140S}$ demonstrated resistance to the known inhibitors with $IC_{50}$ values of 1.5 ± 0.50µM and 2 ± 0.20µM obtained with RAL and EVG, respectively. These findings were comparable to previously documented $IC_{50}$ values of 1.96µM and > 1 obtained for RAL and EVG respectively against this integrase mutant.[229,230] The $FCIC_{50}$ of RAL and EVG was 214 and 200, respectively. $IN_{Q148H/G140S}$ mutation demonstrated no appreciable resistance to DTG ($IC_{50}$ = 3 ± 1nM).

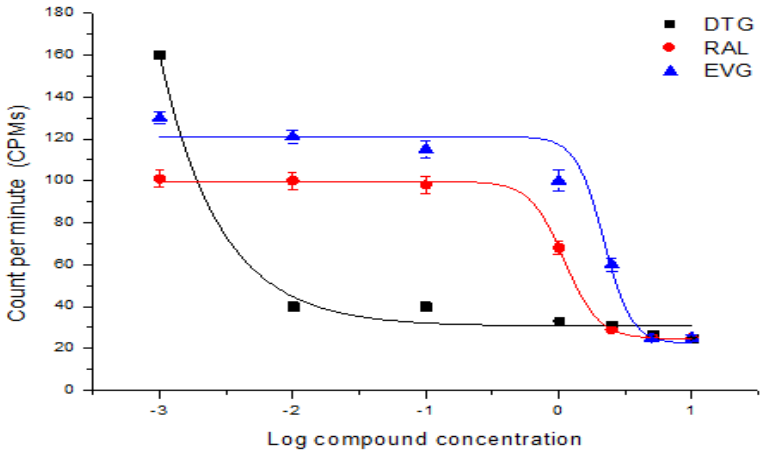
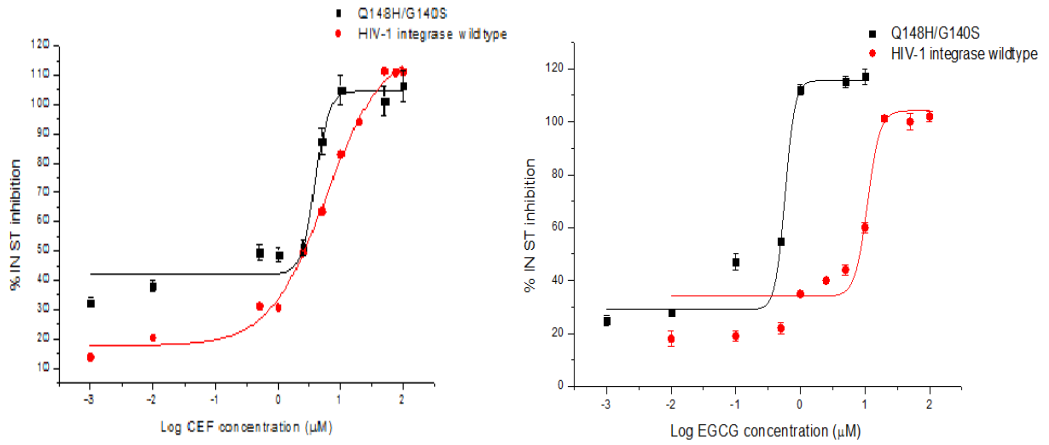



**Figure 3.23: The sigmoidal curves demonstrate the dose-response studies of dolutegravir (DTG), raltegravir (RAL) and elvitegravir (EVG) controls against HIV integrase (IN) harbouring the Q148H/G140S mutation**. DTG did not exhibit a sigmoidal curve and generated a steep Hill-Slope of -4. EVG demonstrated a standard sigmoid dose-response curve with a Hill-Slope of -1. RAL demonstrated a steeper curve than EVG with a Hill-Slope of -3.

Figure 3.24 represents the dose-response studies of CEF and EGCG conducted to determine the $IC_{50}$ of the HITS against $IN_{Q148H/G140S}$ mutation. The $IC_{50}$ for CEF and EGCG were 5.01 ± 0.3µM and 0.619 ± 0.5µM respectively, with $FCIC_{50}$ values of 0.8 and 0.065.




**A** **B**

**Figure 3.24: The dose response curves of A: Cefixime trihydrate (CEF) and B: Epigallocatechin gallate (EGCG) exemplifying their $IC_{50}$ values when screening against HIV integrase wild type ($IN_{WT}$) and $IN_{WT}$ harbouring the Q148H/G140S mutation.** The $IC_{50}$ of CEF and EGCG screened against $IN_{WT}$ were 6.03 ± 1.29µM and 9.57 ± 1.62µM, respectively. The $IC_{50}$ values obtained when screening CEF and EGCG against the $IN_{Q148H/G140S}$ were 5.01 ± 0.3µM and 0.619±0.5µM, respectively. The curves represent experimental data in triplicate with error bars within a standard deviation of 5% of the mean. The slope of both CEF and EGCG against $IN_{Q148H/G140S}$ are shallower than the slopes of CEF and EGCG against HIV-1 $IN_{WT}$. The $IC_{50}$ values of CEF and EGCG against $IN_{Q148H/G140S}$ and $IN_{WT}$ increase as the area under the curve decrease.

### *3.2.8. Screening of cefixime trihydrate derivatives*

The CEF derivatives, 7-ADCA, 7-ACA and CSS containing the distinct β-lactam ring (Figure 3.25), were screened against $IN_{WT}$ using the SPA to identify the moiety on the CEF structure that attributed to the IN ST inhibition. 7-ADCA, 7-ACA and CSS at 10µM yielded 2.45 ± 1.08%, 1.9 ± 1.46% and 5.28 ± 12.10% IN ST inhibition, respectively. In addition, a closely related β-lactam antibiotic, AMP (Figure 3.25), was screened against $IN_{WT}$ and demonstrated 66.5 ± 1.44% IN ST inhibition at 10µM.

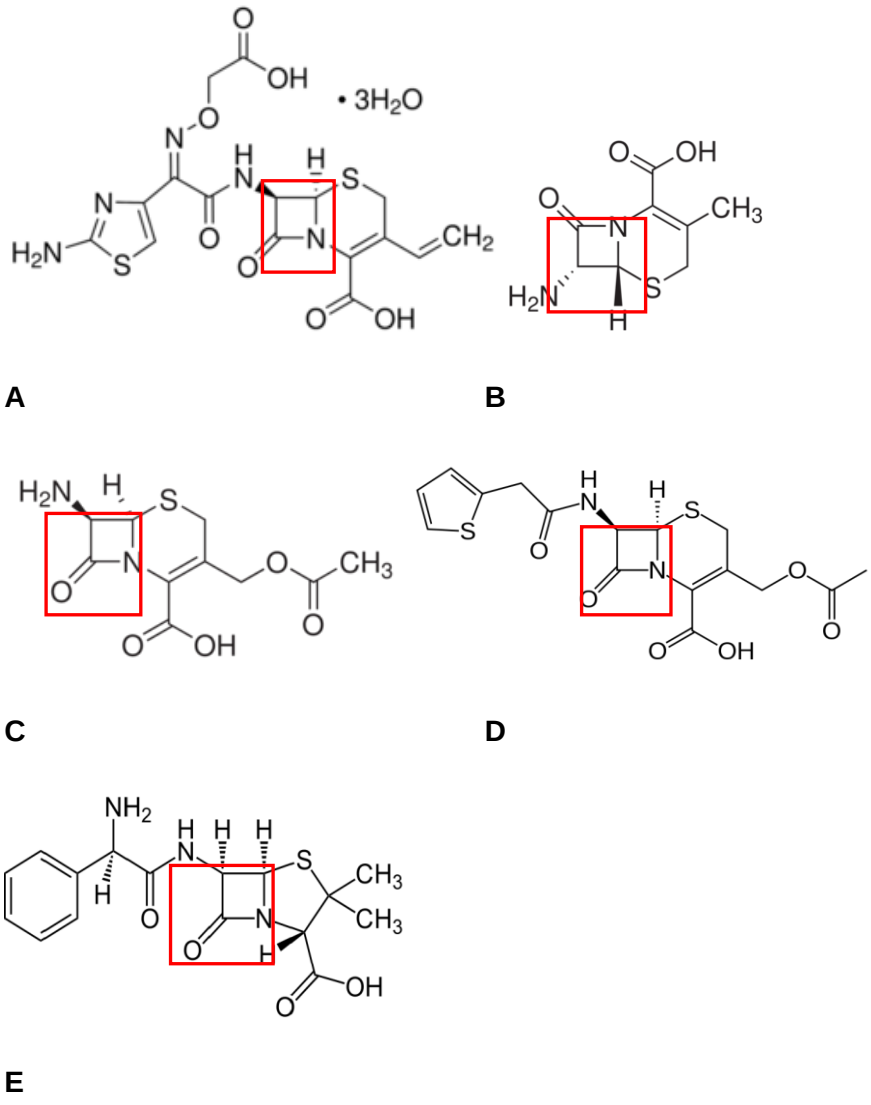



**Figure 3.25: The structures of cefixime trihydrate and its structurally related compounds**. A - Cefixime trihydrate (CEF) B - 7-aminodesacetoxycephalosporanic acid (7-ADCA) C - 7-aminocephalosporanic acid (7-ACA) D - Cephalothin sodium salt (CSS) E – Ampicillin (AMP). The β-lactam rings, consistent in all structures, are demarcated with the red blocks. Structures obtained from www.chemspider.com, accessed 19/09/2013.

### *3.2.9. Further development of identified HIT compounds*

#### *3.2.9.1. Evaluating the cytotoxicity of epigallocatechin gallate, cefixime trihydrate and ampicillin*

The cytotoxicity of CEF, EGCG and AMP in MT-4 cells was tested and $CC_{50}$ values were determined. CEF and AMP did not demonstrate toxicity ($CC_{50}$ > 200μM) within the range evaluated. The dose-response sigmoidal curve of EGCG is demonstrated in Figure 3.26. EGCG yielded a $CC_{50}$ of 23 ± 1μM whereas

previous studies reported that EGCG did not demonstrate cytotoxicity at a concentration < 100µM in peripheral blood mononuclear cells.[231] The control, auranofin, had a $CC_{50}$ = 0.57 ± 0.16µM which correlated with a previously reported $CC_{50}$ value for auranofin ($CC_{50}$ < 1.652µM).[232]

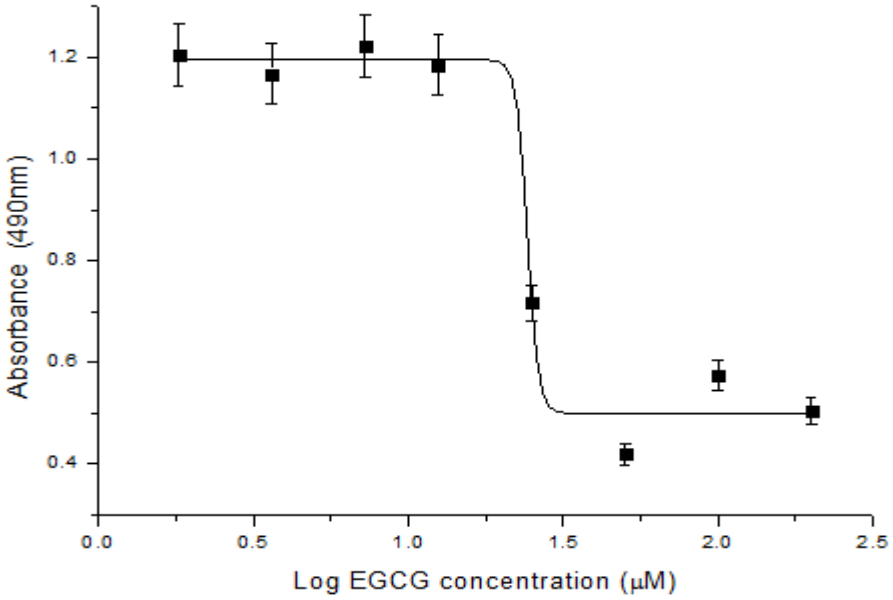



**Figure 3.26: A representation of the dose-response studies of epigallocatechin gallate to determine the cytotoxicity of the compound in MT-4 cells.** The viability of the cells was quantified through an MTS assay where the absorbance of the reduced formazan product was read at 490nm. The sigmoidal curve demonstrates the concentration that reduces cell viability by 50% ($CC_{50}$ = 23 ± 1µM). A steep curve is observed with a Hill-Slope of -24. Each datapoint is the average of triplicate experiments with errors bars indicating the standard deviation of the average mean within 5%.

### *3.2.9.2. Determining the effect of cefixime trihydrate, epigallocatechin gallate and ampicillin on reverse transcriptase activity*

CEF and EGCG were screened against RT to determine whether the compounds had an effect on RT activity. The absorbance value of the ABTS-peroxidase end-product was indicative of RT activity. Figure 3.27 represents the absorbance value of each compound tested as well as the absorbance values of the controls. EGCG inhibited RT activity by 56% and 25% at 100µM and 10µM, respectively. CEF inhibited RT activity by 14% and 9% at 100µM and 10µM, respectively. AMP inhibited RT activity by 8% and 1% at 100µM and 10µM, respectively. It is evident that the effect of EGCG on RT activity was dose-dependent.

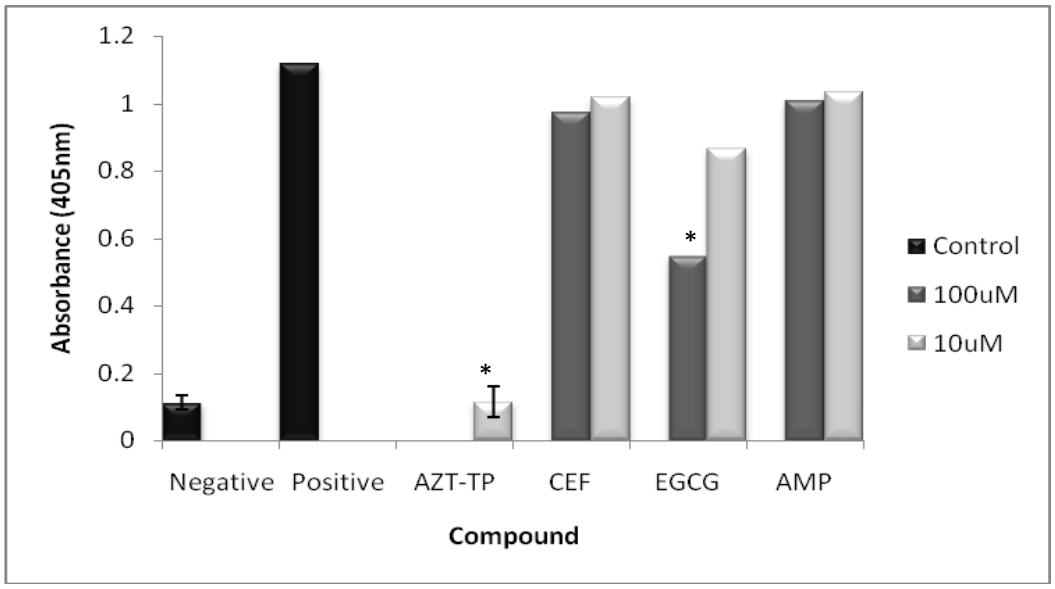



**Figure 3.27: Representation of the absorbance values measured at 405nm when analysing the effect of cefixime trihydrate (CEF), epigallocatechin gallate (EGCG) and ampicillin (AMP) on reverse transcriptase (RT) activity.** The absorbance values of CEF, AMP and EGCG (10µM) was not significantly lower than the absorbance value of the positive control ($p > 0.05$). The absorbance value of EGCG at 100µM was significantly lower than the absorbance value of the positive control ($p = 0.023$) indicating inhibition of RT activity. The AZT-tp control absorbance value was significantly lower than the positive control ($p = 0.002$). The asterisks (*) signify statistical differences.

### *3.2.10. Orthogonal screening of the identified HITS*

#### *3.2.10.1. Validation of the cefixime trihydrate and epigallocatechin gallate through HIV-1 IN ELISA*

The effect of CEF, EGCG and AMP against $IN_{WT}$ ST activity was confirmed through an HIV-1 IN ELISA. Figure 3.28 represents the absorbance values obtained for CEF, EGCG, AMP and RAL. CEF demonstrated 21% and 15% ST inhibition at 100µM and 10µM, respectively. EGCG exhibited 29% and 26% ST inhibition at 100µM and 10µM, respectively. AMP inhibited ST activity by 53% and 27% at 100µM and 10µM, respectively. RAL at 10µM was used as a control and exhibited 97% IN ST inhibition.

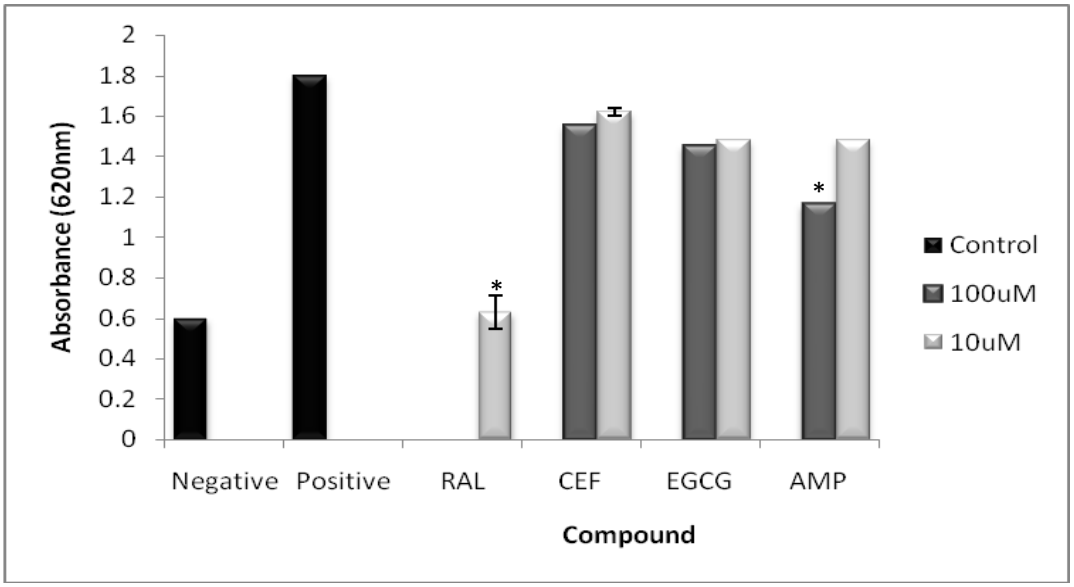



**Figure 3.28: The absorbance readings ($A_{620}$) measured when conducting an HIV-1 integrase (IN) ELISA to confirm the IN strand transfer (ST) inhibitory effect of cefixime trihydrate (CEF), epigallocatechin gallate (EGCG) and ampicillin (AMP).** The $A_{620}$ measured was directly proportional to the ST activity. Although the absorbance values of CEF, EGCG at 100µM and 10µM were lower than the positive control, they were not significant ($p > 0.05$). The AMP absorbance value at 100µM was significantly lower than the positive control ($p = 0.01$). The absorbance values observed for AMP was dose-dependent since the absorbance values decreased with an increase in compound concentration. The raltegravir control absorbance values were significantly lower than the positive control ($p = 0.0004$). The asterisks (*) represent statistical differences.

### *3.2.10.2. Further validation of cefixime trihydrate, epigallocatechin gallate and ampicillin through an HIV-1 phenotypic inhibition assay*

Phenotypic assays were conducted to further validate the effect of CEF, EGCG and AMP against HIV-1. The activity of the virus was determined by detecting the p24 expression within infected MT-4 cells. Figure 3.29 represents the absorbance values measured for each compound tested as well as the controls. EGCG at 200µM demonstrated viral activity inhibition > 100% whereas CEF and AMP at 200µM demonstrated viral activity inhibition of 23% and 51%, respectively. RAL inhibited viral actvity by 74% at 100nM and 118-D-24 at 100µM inhibited viral activity by 97%.

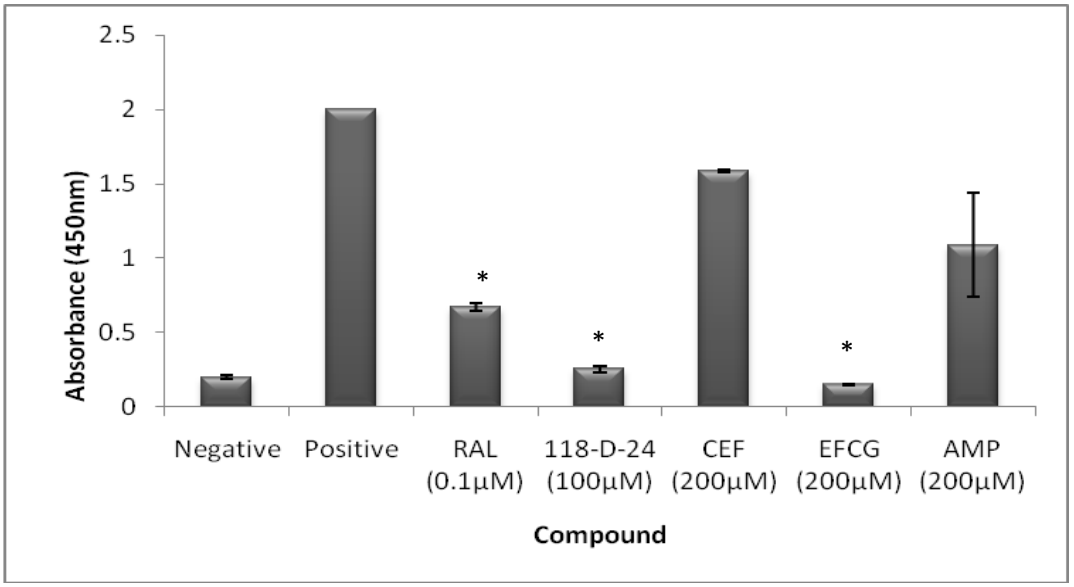



**Figure 3.29: The absorbance values of cefixime trihydrate (CEF), epigallocatechin gallate (EGCG), ampicillin (AMP) and the controls measured at 450nm when detecting the level of p24 expression in an HIV-1 phenotypic assay**. The absorbance value of EGCG, AMP, 118-D-24 and RAL were significantly lower than the positive control ($p < 0.05$). The asterisks (*) indicate statistical differences.

Dose-dependent studies of EGCG and the control, 118-D-24, yielded $EC_{50}$ values of 24 ± 3µM and 9 ± 4µM, respectively. Previous phenotypic studies have reported $EC_{50}$ values of 730nM and 2.1µM for EGCG and 118-D24, respectively.[222,233] Figure 3.30 exemplifies dose-dependent response curves of EGCG and 118-D-24.

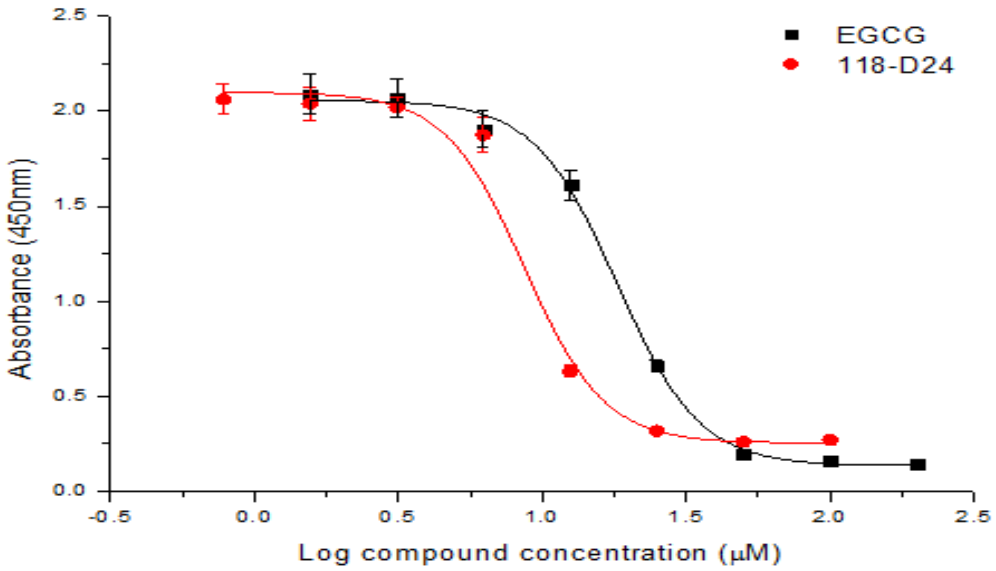



**Figure 3.30: Sigmoidal curves representing the dose-dependent inhibition of p24 expression attributed to epigallocatechin gallate (EGCG) and the control 118-D-24 in HIV-1 infected MT-4 cells.** EGCG exhibited an $EC_{50}$ = 24 ± 3µM and 118-D24 an $EC_{50}$ = 9 ± 4µM. The Hill-Slope of 118-D24 demonstrates a standard sigmoid dose-response curve with a Hill-Slope of close to -1. The slope of the EGCG curve was steeper than the 118-D-24 slope with a Hill-Slope of -2.

### *3.2.10.3. Determining the effect of cefixime trihydrate and epigallocatechin gallate in the presence of reducing agents on strand transfer inhibition*

The effect of reducing agents such as DTT and B-Me was tested using the HIV-1 IN ELISA when screening CEF and EGCG. The DTT concentration was increased to 10mM as opposed to the optimal 5mM DTT used in this ELISA. In addition, the effect of a weaker reducing agent than DTT, B-Me at 10mM, was tested when screening CEF and EGCG against HIV-1 IN. The comparison of ST inhibition attributed to CEF and EGCG in the presence of 5mM DTT, 10mM DTT and 10mM B-Me is depicted in Figure 3.31. An increase in IN ST inhibition was observed with an increase in DTT concentration when screening CEF and EGCG against HIV-1 IN. CEF demonstrated 53% in the presence of 10mM DTT whereas 66% IN ST inhibition was observed for EGCG. CEF and EGCG in the presence of 10mM B-Me yielded no appreciable activity against HIV-1 IN with 16% and < 0% ST inhibition in comparison to the positive control.

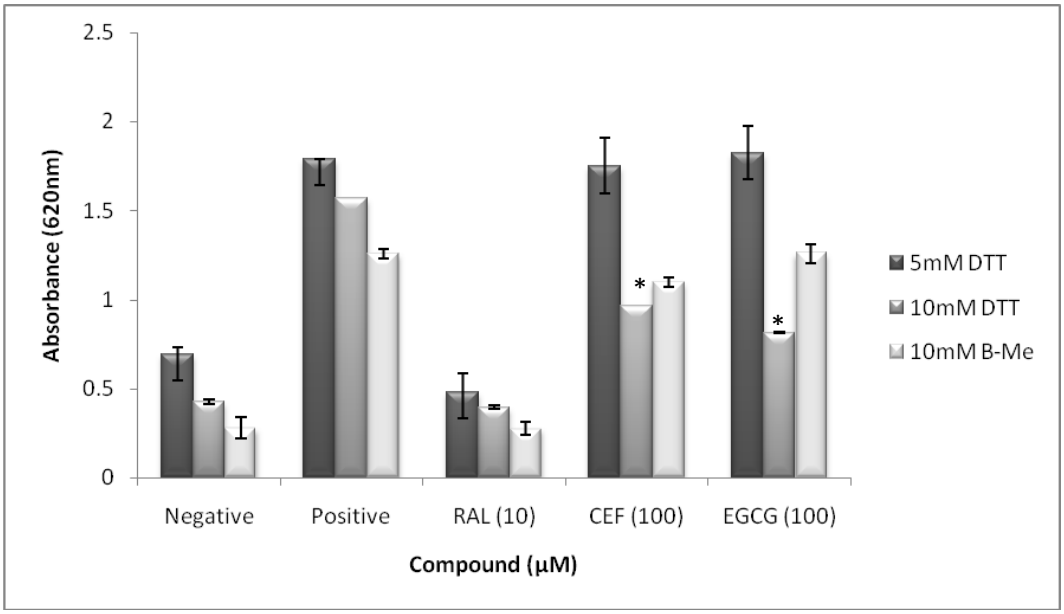



**Figure 3.31: The comparison of the activity of cefixime trihydrate (CEF) and epigallocatechin gallate (EGCG) against HIV-1 integrase (IN) in the presence of 5mM and 10mM DTT as well as 10mM B-Me.** The absorbance values measured at 620nm when conducting an HIV-1 IN ELISA was indicative of the activity of CEF and EGCG against HIV-1 IN. T-test analysis indicated that there were no significant differences between the positive controls obtained in the presence of the 5mM DTT, 10mM DTT and 10mM B-Me ($p > 0.005$). The absorbance value of CEF in the presence of 10mM DTT and 10mM B-Me was significantly lower than the absorbance value of CEF in the presence of 5mM DTT ($p = 0.0001$). Although CEF in the presence of 10mM B-Me was significantly lower than CEF in the presence of 5mM DTT, it was not significantly lower than the positive control (10mM B-Me). EGCG in the presence of 10mM DTT was significantly lower than the positive control (10mM DTT) and the absorbance value of EGCG in the presence of 5mM DTT ($p < 0.005$). The error bars indicate the standard deviation of the average n = 3.

## 3.3. Computational studies of cefixime trihydrate and epigallocatechin gallate

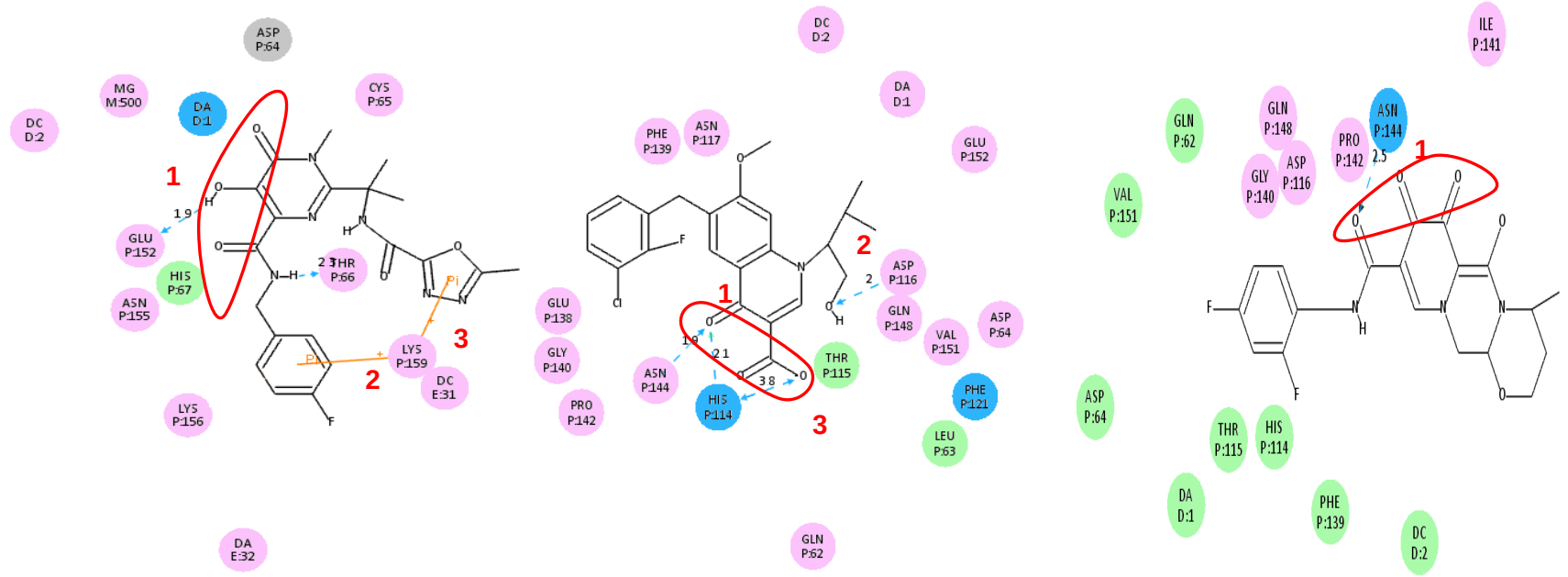
### 3.3.1. Molecular docking of compounds with HIV-1 integrase

#### 3.3.1.1. Validating the docking protocol using control compounds

To validate the ligand-receptor docking method, structurally-related INIs such as RAL, EVG and DTG were used in the docking studies since they have similar interactions with the HIV-1 IN active site. The best docking poses for RAL, EVG and DTG were selected based on their docking scores such as binding energy. Compounds with the lowest binding energy are deemed as the most favourable

receptor-ligand interactions.[234] The binding energy for RAL, EVG and DTG were -89.92 kilocalorie per mole (kCal/Mol), -39.89kCal/Mol and -54.45kCal/Mol, respectively. Figure 3.32 A illustrates the interaction of RAL with the defined active site of the HIV-1 IN with viral DNA. The hydroxyl group that forms part of the heteroatom triad of RAL forms a hydrogen bond with the side chain of the Glu 152 residue that is present in the DDE motif of the IN active site (1.9Å). A secondary hydrogen bond is formed between an imine group (1) and the hydroxyl group of the threonine (Thr) 66 side chain (2.3Å). Furthermore, the lysine (Lys) 159 residues present in the active site of IN formed two Pi interactions with the benzene ring (2) and the amine ring group (3) of RAL.

EVG and DTG (Figure 3.32 B and C, respectively) interact with the active site mainly through a heteroatom triad as with RAL. The O (1) of the heteroatom oxygen triad of EVG forms two hydrogen bonds with the amino group on the side chain of asparagines (Asn) 144 (1.9Å) and with the imidazole group present on the side chain of histidine (His) 114 (2.1Å), respectively. A hydrogen bond interaction is observed between the amino group on the main chain of aspartic acid (Asp) 116 and the hydroxyl group (2) on EVG (2Å). In addition, charge interactions occur between the positively charged side chain of His 144 and the deprotonated O (3) of the carboxylic acid in the heteroatom oxygen triad (3.8Å). The interaction between DTG (Figure 3.32 C) and the IN monomer occurs via a hydrogen bond between the O (1) of the heteroatom oxygen triad present on DTG and the amino group on the side chain of Asn 144.

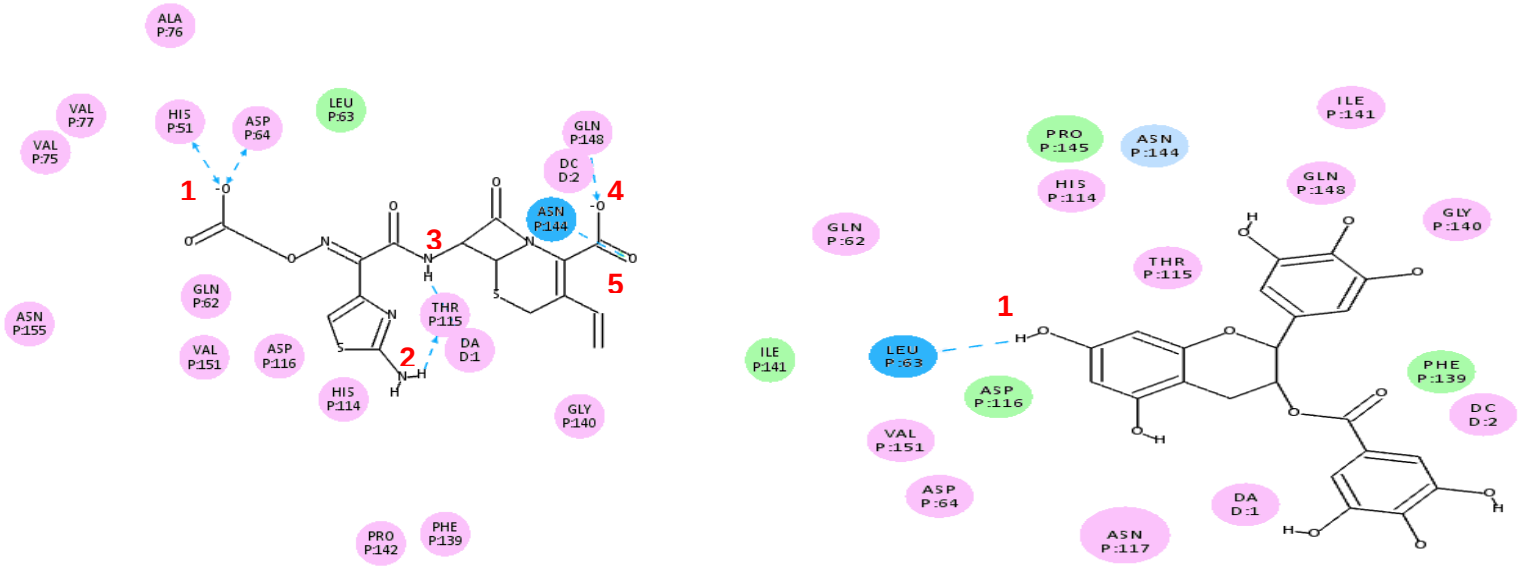



**A** **B** **C**

**Figure 3.32: Two-dimensional (2D) structural diagrams generated using Discovery Studio version 3.1 software, exemplify the interactions between the defined active site of the HIV-1 integrase (IN) and the control compounds A, B and C; raltegravir (RAL), elvitegravir (EVG) and dolutegravir (DTG), respectively.** Residues involved in possible van der Waals interactions are represented as green circles. Pink and blue circles denote residues involved in hydrogen bonds, charge- and polar interactions, respectively. The blue arrows denote interactions between the HIV IN active site residues and the control compounds. The heteroatom oxygen triad is demarcated with red. The molecules on the compounds involved in interactions are denoted numerically.

### 3.3.1.2. Molecular docking of cefixime trihydrate and epigallocatechin gallate with HIV-1 integrase

Interactions between the defined HIV-1 IN active site and the HIT compounds identified when screening the NCC library were predicted via molecular modelling. CEF and EGCG were identified as INIs through SPA screening and their structures were subsequently obtained from www.nihclinicalcollection.com. Overall, 15 poses were successfully docked for CEF and 16 poses for EGCG through CDOCKER. These poses were ranked according to their binding energies and CDOCKER energies. Albeit many poses docked successfully, not all were favourable. The binding energy of CEF and EGCG were 183kCal/Mol and -74.57kCal/Mol, respectively. However, the CDOCKER scoring function rendered a more favourable result with CEF, -439.55kCal/Mol as opposed to the CDOCKER energy of EGCG, -29.37kCal/Mol. The 2D structures depicted in Figure 3.33 describe the interactions between the CEF (A) and EGCG (B) and the defined IN active site, respectively. Interactions between the residues in the IN active site and CEF mainly occur through hydrogen bonds within a 3Å bond distance (Figure 3.33 A). The deprotonated carboxylic acid (1) forms two interactions with the aa residues of the IN active site: 1 - a hydrogen bond with the hydroxyl group on the side chain of the Asp 64 residue present in the IN active site (2.7Å) and 2 - charge interactions with the positively charged side chain of His 51 (2.1Å). The highly reactive aromatic system is involved in the hydrogen bond interactions between the H31 of the amino group (2) attached to the amino-thiazoyl moiety and the carboxylic acid on the main chain of Thr 115 (2.1Å). The carboxylic acid on the main chain of Thr 115 forms an additional hydrogen bond with H36 that comprise the central imino group (3) (2.2Å). Furthermore, the deprotonated carboxylic acid at the terminal end of the chain (4) forms a hydrogen bond with the amino group on the side chain of glutamine (Gln) 148 (2.2Å). The O (5) of the carboxylic acid forms a hydrogen bond with the amino group on the side chain of Asn 144 (2.2Å). Figure 3.33 B illustrates the interaction between the aa residues present in the IN active site and EGCG. The hydroxyl group (1) attached to the highly reactive benzopyran-diol interacts with the carboxylic acid on the main chain of leucine (Leu) 63 (2.3Å) via a hydrogen bond interaction.




**A** **B**

**Figure 3.33: Two-dimensional (2D) structures demonstarting the predicted structural interactions between A - Cefixime trihydrate (CEF) and B - Epigallocatechin gallate (EGCG) and the defined HIV integrase (IN) active site. The 2D structural diagrams were constructed using the Discovery Studio version 3.1 receptor-ligand docking tool.** All possible interacting amino acids (aa) of the active sites are indicated through the coloured circles. The pink and blue circles indicate the aa that are involved in possible hydrogen bonds, charge- or polar interactions, respectively. Aa residues with possible van der Waals interactions are represented by the green circles. The blue arrows denote possible bond formations between aa residues and the molecules comprising CEF and EGCG. The molecules of the compounds involved in interactions are numerically denoted.

### 3.3.2. Evaluation of the absorption and distribution properties of cefixime trihydrate and epigallocatechin gallate

Osiris Property Explorer predicted the absorption, solubility and drug-like score of CEF and EGCG. Table 3.2 exhibit the predicted drug property values of CEF and EGCG. A good absorbed drug has a cLogP between -0.4 to 5.6 where CEF and EGCG exibited cLogP values < 5 thus indicating good absorption. The solubility of CEF and EGCG was -2.8 and -2.16, respectively, indicating moderate solubility of the compounds. LogS values greater than -4 are typically an indication of good solubility. The drug-like scores of CEF and EGCG were 0.76 and 0.69, respectively. A drug-like score ranges from 0-1, where 1 indicates a good drug candidate while 0 indicates a poor drug candidate.

**Table 3.2: The predicted solubility, absorption and drug-like score of cefixime trihydrate and epigallocatechin gallate using Osiris Property Explorer.**

| | Cefixime trihydrate | Epigallocatechin gallate |
|---|---|---|
| **CLogP** | -0.65 | 2.65 |
| **Solubility** | -2.8 | -2.16 |
| **Drug score** | 0.76 | 0.69 |

In addition, absorption and distribution properties of CEF and EGCG were predicted using Discovery Studio version 3.1 software (Table 3.3). The predictive aqueous solubility model used predicts the solubility of compounds in water at 25°C which is expressed as logSw. The predicted logSw for CEF was -3.584 and the predicted logSw for EGCG was -6.104. These logSw values indicate that CEF is drug-like and highly soluble. The logSw value predicted for EGCG indicated that the drug-likeness of the compound was low but may still be soluble. The predicted HIA of CEF and EGCG were expressed according to their AlogP levels. Both compounds exerted an HIA level of 3 with -2 ≥ AlogP98 ≥ 7. This indicates that CEF and EGCG are poorly absorbed after oral administration. The ADMET BBB model predicted the penetration of the compounds through the BBB after oral administration. The categorical BBB level of both CEF and EGCG were 4. This indicates that the BBB penetration predication of CEF and EGCG could not be defined because the compounds are not within the 99% confidence ellipse.

**Table 3.3: The predicted human intestinal absorption, solubility and blood brain barrier penetration scores for the two HIT compounds obtained from Discovery Studio version 3.1**

| | Cefixime trihydrate | Epigallocatechin gallate |
|---|---|---|
| **Human intestinal absorption** | 3 | 3 |
| **Solubility** | -3.584 | -6.104 |
| **Blood brain barrier penetration** | 4 | 4 |

The data obtained from Osiris Property Explorer and Discovery Studio version 3.1 was used to predict the oral bioavailability of CEF and EGCG using the Lipinski Ro5 described in section 1.2.2.2. Data in Table 3.4 indicated that both CEF and EGCG do not fully adhere to these rules.

**Table 3.4: The predicted oral bioavailability of cefixime trihydrate and epigallocatechin gallate using Lipinski Rule of 5.**

| | Cefixime trihydrate | Epigallocatechin gallate |
|---|---|---|
| **MW < 500 Daltons** | X | √ |
| **cLogP < 5** | √ | √ |
| **Hydrogen bond donors < 5** | √ | X |
| **Hydrogen bond acceptors < 10** | X | X |

# CHAPTER 4
## Discussion

The inevitable emergence of HIV-1 drug resistance mutations against the current antiretroviral agents has propagated investigations to identify novel antiretrovirals against enzymes and viral- or host proteins essential for HIV-1 replication. This study focuses on repurposing available drugs as HIV-1 INIs. By using recombinantly expressed HIV-1 $IN_{WT}$ in an optimized SPA amended to an automated system, a small compound library, NCC, was screened to identify potential HIV-1 INIs. The expression and purification of HIV-1 $IN_{WT}$ was confirmed through SDS-PAGE analysis as well as Western blot analysis. Binding and activity against the known INIs RAL, EVG and DTG confirmed that the purified recombinant IN was functional. Overall, two compounds were identified when screening the NCC compound library that yielded IN ST inhibition over 60%. These compounds were CEF and EGCG. Mutation studies indicated that the activity of CEF and EGCG were not affected by the viral IN mutations $IN_{Q148H/G140S}$. The binding mode of CEF and EGCG were predicted through molecular docking studies as well as through the screening of CEF derivatives. *In silico* studies indicated that CEF and EGCG were drug-like. The activity of CEF and EGCG could, however, not be confirmed through orthogonal assays.

### 4.1. Successful expression and purification of HIV-1 subtype B integrase

The HIV-1 NL4-3 IN sequence was cloned into pET15B which was subsequently transformed into *E. coli* BL21 (DE3) bacterial expression cells and induced with IPTG. SDS-PAGE gels indicated that the protein saliently expressed was HIV-1 IN at approximately 32kDa.[160] The His-tagged HIV-1 IN was purified through Ni-affinity chromatography where the electron rich His-tag has a high affinity for the transitional metal, $Ni^{2+}$. The single purification step has previously demonstrated protein purification of 95% expressed in *E.coli* [235] whereas the recombinant HIV-1 IN in this study demonstrated 73% purity. Previous studies have indicated that the His-tag present on the recombinant HIV-1 IN did not affect the activity of the enzyme[211] and therefore the His-tag was not cleaved in this study. Western blot analysis (Figure 3.16) confirmed the recombinant expression of HIV-1 IN with primary antibodies against the His-tag of the expressed HIV-1 IN as well as antibodies against the aa (23-34) sequence of HIV-1 IN. Additional bands were

observed at approximately 60kDa on a reducing SDS-PAGE gel which suggests dimerization of the HIV-1 IN enzyme.[8,9] The Western blot analysis demonstrated an additional band higher than the prominently visible band when probed against the HIV-1 IN aa 23-34. However, the occurrence of dimerization on a reducing gel is unlikely since a reducing agent is added to perturb cysteine bonds involved in multimerization of HIV-1 IN enzymes. Since the additional band was also observed on a Western blot it can be assumed that dimerization may have occurred due to incomplete reduction of cysteine bonds. HIV-1 $IN_{WT}$ is prone to aggregation and together with the fact that a high concentration of the protein was loaded on to the gel (75µM), the 5% B-ME used to reduce the protein may not have been sufficient. Parameters that influence dimer formation such as the reducing agent concentration, protein concentration and boiling duration of the sample can be tested. In addition, mass spectroscopy studies can be conducted to validate whether the additional band is indeed an HIV-1 IN dimer. Sufficient recombinant HIV-1 IN was expressed and purified for optimization of the SPA, the screening of the NCC compound library and HIV-1 IN ELISAs to minimize variability between experiments.

## *4.2. The optimized SPA parameters*

In order to screen the NCC compound library, a suitable assay had to be developed that was able to screen multiple compounds simultaneously using an automated system. As described in section 1.2.3.5, the SPA was deemed the most suitable assay for this study and was therefore used to monitor IN ST activity. A protocol extracted from Grobler and co-workers[203] was altered by modifying several variables in the assay such as buffer reagents, divalent metal cations, reducing agents and detergent concentrations which can compromise the efficacy of inhibitory compounds.[236] The optimal conditions for the SPA were determined based on CPM signal, IN ST inhibition attributed to RAL and most importantly the Z-factor of the assay.

Previous studies have demonstrated that the assembly of the IN viral DNA complex as well as 3'-end processing and IN ST requires a divalent metal cation, either $Mg^{2+}$ or $Mn^{2+}$, in order for it to exert its function.[237,238] Although, Grobler and co-workers[178] elucidated that $Mg^{2+}$ was physiologically more relevant *in vivo*, $Mn^{2+}$ is more compatible in *in vitro* experiments such as the SPA as IN activity is more robust in the presence of $Mn^{2+}$.[178] The $Mg^{2+}$ was used in further SPA

experiments in this study, since this divalent metal cation is physiologically more relevant and RAL exerted a higher IN ST inhibition in the presence of $Mg^{2+}$.

In order to increase the CPM signal the stop solutions and buffer compositions were tested. When testing the buffer compositions, Buffer 4 (Chow[199] buffer with 100μg/ml BSA) demonstrated the highest CPM signal however the Z-factor was below zero which indicated that there was no separation band between the positive- and negative control and therefore the assay was deemed futile. This low Z-factor might be attributed to background signal indicating that the IN ST inhibition observed (70 ± 4.3%) when using Buffer 4 was possibly a false positive. For these reasons, Buffer 2 (Chow[199]) was used as the optimal assay buffer since this buffer demonstrated the second highest CPM signal (141 ± 49) and a marginal Z-factor of 0.4. However, the Z-factor indicated that the separation band between the positive-and negative control was low which suggested background signal. To further reduce background signal, various stop solutions were tested.

Stop solution was added to the assay to terminate the enzymatic reaction and to float scintillation beads to the suface thus bringing the beads in closer proximity to the photomultiplier tube detectors thereby reducing background signal. EDTA at a final concentration of 62mM was found to be the optimal stop solution; likely due to the fact that EDTA is a $Mg^{2+}$ chelator and therefore effectively terminated the activity of the enzyme.

The fully optimised SPA was deemed an excellent assay with Z = 0.8 according to the screen assay quality category described by Zhang and co-workers.[227] The assay was then further validated through dose-dependent studies of known controls. The SPA yielded $IC_{50}$ values of 7.3 ± 0.3nM, 4.0 ± 1nM and 1.5 ± 0.8μM for RAL, EVG and 118-D24, respectively, which was statistically comparable ($p$ > 0.005) to the $IC_{50}$ values previously documented in literature (described in section 3.2.3). This then indicated that the optimised SPA was validated and could be used in subsequent HIV-1 IN activity screening experiments.

In order to screen the entire NCC compound library, a SPA protocol analogous to the manual SPA protocol was amended to an automated system. The success of amending the manual SPA to an automated system was determined by comparing the $IC_{50}$ values of RAL obtained from both the automated- and manual SPA. T-test analysis confirmed that the $IC_{50}$ of RAL obtained when conducting

the manual SPA was statistically comparable ($p$ = 0.99) to the $IC_{50}$ of RAL when conducting the automated SPA indicating that this protocol was adequate for use on the automated system. The optimized SPA was used to screen the NCC compound library and subsequent compound derivative experiments. The SPA was then further used for mutant screening by replacing the HIV-1 $IN_{WT}$ with the $IN_{Q148H/G140S}$. The SPA was validated by screening known controls such as RAL, EVG and DTG against the $IN_{Q148H/G140S}$ mutant. Resistant profiles of RAL, EVG and DTG obtained through SPA screening against $IN_{Q148H/G140S}$ correlated with previous reports as indicated in section 3.2.7 thus validating the SPA as a suitable assay for HIV-1 IN mutant screening. Since the $IN_{Q148H/G140S}$ mutation decreased the activity of the enzyme, the concentration of the IN mutant was increased in order to increase the CPM signal that would yield a satisfactory Z-factor.

For mutation studies, the activity of the mutant IN, $IN_{Q148H/G140S}$, was tested and compared to the activity of $IN_{WT}$. The $FCIC_{50}$ of $IN_{Q148H/G140S}$ against control compounds was tested to determine the degree of resistance of the mutation. The activity of $IN_{Q148H}$ decreased 6 fold when compared to the activity $IN_{WT}$ which was in accordance with previous studies.[229] The decrease in enzyme activity may be due to the $IN_{Q148H}$ mutation that is present in the DDE motif of the HIV-1 IN active site and decreases the amount of viral DNA integrated into the host DNA consequently impairing the integration process.[176] When pairing the primary mutation, $IN_{Q148H}$, with the secondary mutation, $IN_{G140S}$, the integration efficiency was restored and the activity of HIV-1 IN decreased by only 2 fold when compared to HIV-1 $IN_{WT.}$ This data correlated with the results observed in the study conducted by Delelis and co-workers.[229]

## *4.3. HITS identified when screening the NCC compound library*

This study focused on compounds exhibiting appreciable IN ST inhibition when screening the NCC compound library. Two compounds (out of a total of 727) that demonstrated IN ST inhibition > 60% were identified as a β-lactam antibiotic, CEF, and the Green Tea antioxidant, EGCG. Pooled compounds that contained a coloured compound within the group of compounds were considered false

positives. Due to quenching, these coloured compounds absorb the light emitted from the scintillant in the PVT SPA bead thereby decreasing the energy emitted. As such, these coloured samples appear to inhibit IN activity but may be artifacts.

### 4.3.1. Cefixime trihydrate identified as a strand transfer inhibitor

The notion of an antibiotic inhibiting the activity of a virus is uncommon, however, the β-lactam antibiotic, CEF, demonstrated IN ST inhibition with an $IC_{50}$ = 6.0 ± 1.2μM. CEF is a third generation cephalosporin antibiotic active against a wide range of gram-negative bacteria that is resistant to β-lactam antibiotics.[239] CEF is renowned for its high affinity for its target, penicillin-binding protein (PBP) located in the bacterial cell wall that catalyses cell wall biosynthesis.[240] The involvement of CEF in HIV antiretroviral activity has not been documented as yet, however cephalosporin oligonucleotides as well as monocyclic β-lactams have been reported as HIV PIs.[241] To determine whether CEF was specific against HIV-1 IN, RT inhibition studies were conducted. CEF demonstrated no appreciable RT inhibition at both 10μM and 100μM which indicated that CEF may be specific to HIV-IN. Studies testing CEF against HIV-1 PR and other HIV-1 drug targets would confirm whether CEF is a specific HIV-1 IN inhibitor.

### 4.3.2. The possible binding mechanism of cefixime trihydrate

The activity of CEF against $IN_{Q148H/G140S}$ was investigated since this mutation pathway is one of the main RAL resistance pathways. CEF maintained its activity against HIV-1 $IN_{Q148H/G140S}$ ($FCIC_{50}$ = 0.8) indicating that CEF does not share a genetic barrier with the RAL resistant mutation. This data suggests three theories: 1 - CEF may not bind to the active site of IN and its mechanism of action may differ from the mechanism of action of RAL and EVG, 2 - CEF may bind to viral DNA thereby disrupting the viral DNA-IN interaction which in turn interrupts the integration process, and 3 - CEF may influence the IN enzyme or the assay in a non-specific manner. Further studies to elucidate the mechanism of action of CEF against HIV-1 IN have to be conducted.

Derivatives of CEF; 7-ACA, 7-ADCA and CSS were screened against HIV-1 IN to determine the moiety responsible for the HIV-1 IN ST inhibition. CSS, 7-ACA and 7-ADCA are structurally related to CEF where a β-lactam ring structure attached to a dihydrothiazine ring is consistent throughout the structures of these compounds (Figure 3.25).[242] Cephalosporin acylase hydrolyses cephalosporin C,

consequently producing 7-ACA and 7-ADCA which are intermediates in the production of semi-synthetic cephalosporin antibiotics such as CEF.[243] No appreciable IN ST inhibition was observed with the CEF derivatives at 10μM and therefore it can be assumed that the β-lactam moiety alone is not responsible for the IN ST inhibition. AMP is a β-lactam antibiotic derived from penicillin and follows the same mechanism of action as CEF in preventing the synthesis of gram-negative and gram-positive bacterial cell wall.[240] Both CEF and AMP possess β-lactam rings however they differ structurally where the dihydrothiazine ring on CEF is substituted with a thiazolidine ring on AMP. As such, the effect of AMP on IN ST activity was investigated. AMP demonstrated IN ST inhibition by 66.5 ± 1.4%. Since our findings suggests that the β-lactam ring does not attribute to IN ST inhibition, inhibition may likely be due to interactions between the acyl side chain present on CEF and AMP or due to the terminal carboxylic acid group present on both CEF and AMP.

Molecular modelling studies were conducted to predict the possible binding of CEF to the IN active site. RAL, EVG and DTG were docked into the modelled IN active site to validate the docking method. As described in literature, the coplanar, oxygen heteroatom triad was pivotal in the interaction with the IN active site and its subsequent chelation of $Mg^{2+}$ (described in Section 3.3.2.1).[244] Molecular modelling data (depicted in Figure 3.33 A) demonstrated hydrogen bond interactions between the terminal carboxylic acid present on CEF and the Gln 148 residue attached to viral DNA. Since this carboxylic acid is present on both CEF and AMP, it can be conjectured that this moiety is responsible for IN ST inhibition as hypothesized above. However, the Gln 148 is attached to viral DNA indicating that CEF and AMP may disrupt the viral DNA-IN as previously mentioned. The theory of CEF as a DNA-binder is supported by the distinct 7-α-iminomethoxycarboxy group of CEF that interacts via hydrogen bond interactions with the Thr 115 bound to viral DNA. CEF as a DNA binder has not been reported previously, and its mechanism of action solely relies on CEF binding to a transpeptidase enzyme. An additional hydrogen interaction was observed between CEF and the Asp 64 present in the DDE motif of the IN active site indicating that CEF might also perturb $Mg^{2+}$. This predicted interaction is negated by resistant mutation profile results which suggested that CEF may not bind in the DDE motif of IN. The exact moiety responsible for IN ST inhibition is still

unclear however the predicted interactions and screening of CEF derivatives provide an insightful foundation.

### 4.3.3. In silico studies predicted that cefixime trihydrate is not drug-like

The focus of this study was to identify an INI through drug-repositioning as it would expedite the preclinical drug development phase since the compounds that were tested are in clinical trials or have already been FDA approved. CEF is an FDA approved drug used in bacterial infection treatment and as such it is expected to be drug-like. The drug-score of the compound was determined by evaluating the cLogP, aqueous solubility, MW and the overall drug-likeness of the compound. Prediction studies using Osiris Property Explorer exhibited a very low cLogP value (-0.65) which indicated high hydrophilicity. According to Osiris Property Explorer, a compound that is highly hydrophilic is easily absorbed, however, amendments to Lipinski Ro5 states that the cLogP of a compound should be between -0.4 and 5.6. Since the cLogP of CEF is not within the acceptable partition coefficient range, CEF is predicted as too hydrophilic and may not be able to cross the lipid membrane. CEF was predicted to be moderately soluble.

A successful compound requires a solubility level that is concomitant with good absorption which is required for the systematic transportation of the compound. Discovery Studio software predictions confirmed that CEF is poorly absorbed after oral administration. Discovery Studio software utilises a model developed by Egan and co-workers[245] that consists of a 182 compound training set. This training set includes 95% and 99% confidence ellipses generated by the ALogP98 and 2D polar surface area (2D_PSA) planes. Well-absorbed compounds are within these confidence ellipses where the absorption ability of compounds decreases outside the 95% ellipse. The absorption of CEF could not be predicted since CEF was not within the 95% confidence ellipse and was therefore deemed as a poor absorption compound.

In addition, the oral bioavailability of CEF was predicted using Lipinski Ro5. By assessing the Lipinski Ro5 (Table 3.4), two of the four parameters were out of range (MW > 500 Daltons and hydrogen bond acceptors > 10) which indicated that CEF may be poorly absorbed or the permeability of the compound may be

low. Low MW of compounds are desirable in developmental stages of drugs since modifications during the developmental stages would likely result in an increase in MW. However, since CEF is an FDA approved drug where the MW > 500 Daltons, it is likely that MW increased due to modifications during the development stages to optimize the drug-likeness of CEF. A high MW has been associated with low absorption, which corroborated the absorption prediction obtained using the Discovery Studio software and Osiris Property Explorer. In addition, *in vivo* testing of CEF confirmed the absorption predictions obtained through the Discovery Studio software which indicated that CEF is only 40-50% absorbed after oral absorption. Distribution studies in terms of BBB penetration after oral administration could not be determined because CEF was not within the 99% confidence ellipse. This suggests that the ability of CEF to penetrate the BBB is low according to the BBB penetration criteria levels developed by Egan and Lauri.[245] Therefore, through prediction studies, it can be concluded that CEF is not drug-like.

### *4.3.4. Cefixime trihydrate was not active against integrase in orthogonal screening*

To validate whether CEF was a true IN ST inhibitor, it was screened in an orthogonal assay; the HIV-1 IN ELISA. CEF demonstrated no appreciable IN ST inhibition at 10µM and 100µM thereby contradicting the results obtained through the SPA. This was corroborated through HIV-1 phenotypic inhibition studies (Figure 3.29) suggesting that CEF does not directly inhibit IN activity. The discrepancy between the results obtained through SPA screening and the results obtained from the HIV-1 IN ELISA and phenotypic inhibition assay may be attributed to non-specific binding of CEF to IN when using the SPA. Cephalosporin C, that contains a β-lactam ring, has been reported as a non-specific inhibitor against NADH dehydrogenase and murG enzymes when using a SPA.[246] Cephalosporin C was considered a perturbing agent that causes enzyme inactivation. Since cephalosporin C and CEF are structurally related, CEF may perturb the IN enzyme in a similar manner and therefore CEF appeared to have inhibited IN activity. AMP inhibited IN ST reaction by 12% at 10µM and 51% at 100µM, while the antiretroviral activity of AMP in the phenotypic inhibition assay was weaker but still apparent (51% IN ST inhibition at 200µM). The antiretroviral activity of AMP was not attributed to cytotoxic effects since AMP was not toxic at

the evaluated concentrations (section 3.2.9.1). This data indicate that AMP may play a role the inhibition of HIV-1 IN. The mechanism and binding mode of AMP to HIV-1 IN is yet to be elucidated.

### 4.3.6. Epigallocatechin gallate is active against integrase and reverse transcriptase

EGCG is potent antioxidant and forms part of the catechin group derived from Green Tea, *Camillia sinensis.*[247] EGCG is the most abundant catechin in Green Tea and accounts for 50-80% of the catechin content.[248] Previous studies have reported the involvement of EGCG in many therapeutic applications such as cancer, bacterial, viral- and fungal infections, diabetes, Parkinson's disease, Alzheimer's disease, stroke, inflammation, obesity and HIV.[249–252] In 2011, EGCG phase 1 clinical trials in HIV-1 patients were initiated.[253] The clinical trials were due for completion in July 2013 and the outcome thereof is still pending. (www.clinicaltrials.gov, accessed 1/10/2013). Interestingly, by independently screening the NCC compound library, EGCG was identified as IN ST inhibitor (6.0 ± 1.2μM). EGCG has been previously reported as an HIV-1 INSTI with $IC_{50}$ = 960nM.[254] However, EGCG activity was not limited to IN where RT inhibition studies demonstrated that EGCG inhibited RT by 56% (Section 3.2.9.2). This result was expected since previous studies have reported that EGCG was active against several targets in the HIV-1 replication cycle such as the destruction of the virion, blocking of the gp120-CD4 interaction, as well as inhibition of HIV replication enzymes such as RT, PR and IN.[254–257]

### 4.3.7. The galloyl moiety of epigallocatechin is involved in antiretroviral activity

EGCG was tested against the HIV-1 mutant $IN_{Q148H/G140S}$ and yielded a $FCIC_{50}$ = 0.06. This suggests that EGCG may bind to IN using a different binding site to which RAL binds as mentioned in section 4.3.2. The binding mechanism of action has not yet been defined however previous studies demonstrated that the galloyl moieties of catechins are pivotal in HIV-1 inhibition.[254] EGCG is an ester of gallic acid and is comprised of a galloyl moiety (D ring) attached to a catechin backbone.[258] Previous studies have reported that galloyl moieties are involved in HIV-1 IN ST inhibition as well as RT inhibition.[233,254] According to Yang and co-workers[259], the electron rich polyphenolic structure of the galloyl moiety may

interact via hydrogen bond interactions with IN at a binding site that is still unknown.[259] Molecular docking results in this study predicted that hydrogen bond interactions occurred between the hydroxyl group on the benzene ring of EGCG and the carboxylic acid group of Leu 68 in the catalytic core domain on IN. This predicted interaction suggests that EGCG may influence the activity of IN by binding to a residue within the active site of IN. The predicted interaction between EGCG and HIV-1 IN does not involve the galloyl moiety present on EGCG as identified through biological evaluation. This could be due to the stringent parameters used to dock the compound in the IN active site resulting in fewer poses docked or due to the active site on the prepared protein model that was not well defined.

### *4.3.8. Epigallocatechin gallate as a natural product is drug-like*

According to the Osiris Property Explorer predictions (Section 3.3.2), EGCG is a moderately soluble (-2.16) and well absorbed drug (2.6) and is therefore drug-like (0.69). As stated before, solubility of a compound is usually indicative of good absorption. However, Discovery Studio software predicted EGCG was a very low absorbed drug with a low solubility and therefore was not drug-like. The disparity between the solubility property and dug-likeness of EGCG obtained from Osiris property explorer and Discovery Studio may be due to the prediction models used. Osiris Property Explorer utilises the increment system that is based solely on atom type contributions, while Discovery Studio software utilises a predictive model that was generated using a training set comprising compound classes such as alkanes, alkenes, alkynes, halogens and amines to name a few.[260] The Discovery studio software is more reliable than the Osiris property explorer since it does not include compounds that are not within the 99% and 95% confidence ellipse. As with CEF, the BBB penetration could not be predicted since the compound was not within the 95% and 99% confidence ellipsoids.

By analysing EGCG according to Lipinski Ro5, EGCG does not fully adhere to the rules since the hydrogen bond acceptor is < 10 and hydrogen bond donor is < 5. However, EGCG is a natural compound and therefore does not require stringent compliance to the Lipinski Ro5. Natural products have been de-prioritized or in some cases eliminated from the drug discovery process due to the fact that some natural products do not comply to Lipinski Ro5 yet they can be excellent drugs.[261] Although the Lipinski Ro5 may be evaded for natural products,

a low LogP property should be maintained as natural products are more readily absorbed than synthetic compounds. The LogP value of EGCG predicted was not ideal however it complied with Lipinski Ro5 which states that the logP < 5. On the premise of the compliancy of the LogP value of EGCG with Lipinski Ro5, it can be assumed that EGCG is drug-like.

### *4.3.9. DTT influences the activity of epigallocatechin gallate and cefixime trihydrate*

As with CEF, a secondary assay was conducted to confirm the HIV-1 IN ST inhibitory effect of EGCG. Results from the HIV-1 IN ELISA exhibited no appreciable IN ST inhibition (Figure 3.28). It was evident that there was a component present in the SPA experiment that influenced the behaviour of both EGCG and CEF (Section 4.3.2) since both of these compounds did not evoke appreciable activity against HIV-1 IN in an orthogonal assay. Following this observation, it was then hypothesised that DTT in the presence of EGCG and CEF played a role in HIV-1 IN activity. Subsequently, DTT was used as a reducing agent in both biological assays to enhance IN activity *in vitro* by inhibiting the oxidation of free sulfhydryl residues. The concentration of DTT differed between the two biological assays where the HIV-1 IN ELISA reaction buffer contained 5mM DTT and the SPA reaction buffer contained 10mM DTT. The DTT concentration was increased to 10mM in the HIV-1 IN ELISA to determine whether this variable would influence the effect of EGCG and CEF on HIV-1 IN activity. A weaker reducing agent, B-Me (10mM), was used as an additional parameter to confirm whether an increased reducing agent concentration or the reducing agent itself had an influence on the behaviour of EGCG and CEF. The reducing agent, DTT at a high concentration, has previously been reported to weaken interactions between HIV-1 IN monomers.[262] EGCG and CEF in the presence of 10mM DTT yielded higher IN ST inhibition while EGCG and CEF in the presence of 10mM B-Me exerted no appreciable IN ST inhibition. This data suggests that EGCG and CEF in the presence of an increased DTT concentration inhibit IN ST. The controls (negative, positive and RAL) in the presence of 10mM DTT and 10mM B-ME were not significantly different ($p > 0.05$) from the controls in the presence of 5mM DTT (Figure 3.31). This indicated that an increased DTT concentration only affected the activity of HIV-1 IN in the presence of EGCG and CEF.

Studies have reported that a high concentration of DTT disrupts IN dimers where the thiol groups of DTT targets the CCD of IN and therefore decrease the activity of IN.[262] DTT at a higher concentration may interact with EGCG and CEF in such a manner that enables the compounds to more efficiently disrupt IN dimers. B-ME is considered a weaker reducing agent than DTT because B-ME only comprises one thiol group whereas DTT comprise two thiol groups. Since thiol groups play a vital role in IN dimer disruptions, the weaker B-Me did not interrupt IN monomer interactions and therefore no appreciable IN ST inhibition was observed.[262,263] Altogether, the discrepancy between the results obtained from the SPA and the HIV-1 IN ELISA may be attributed to the DTT concentration difference.

### *4.3.10. Epigallocatechin gallate is auto-oxidized at pH levels above 7*

Since EGCG activity could not be validated through an HIV-1 IN ELISA (Section 4.3.9), the activity of EGCG was investigated in HIV infected mammalian cells. An $EC_{50}$ = 23.5 ± 2µM was observed however cytotoxicity results indicated that EGCG was toxic to MT-4 cells within this concentration range (Figure 3.26). This suggests that the antiretroviral activity caused by EGCG may be attributed to cytotoxicity. According to previous reports, polyphenolic compounds such as EGCG are susceptible to auto-oxidation in cell culture medium such as DMEM and RPMI 1640 due to the pH of the cell culture medium (pH 7.2).[264–266] Hou and co-workers[267] proposed a possible mechanism of action of the auto-oxidation of EGCG. When EGCG undergoes auto-oxidation, it consumes $O_2$ consequently producing superoxide ($\bullet O_2^-$) and EGCG radicals (•EGCG). The $\bullet O_2^-$ is converted to $H_2O_2$ and $\bullet O_2^-$ and in the presence of metal ions, unpaired electrons binds to EGCG and is localized at the B ring thus forming •EGCG and $H_2O_2$. The •EGCG reacts with another •EGCG producing an EGCG dimer and $\bullet O_2^-$. The EGCG dimers can be further oxidized consequently forming other compounds. Due to the extracellular $H_2O_2$ produced when EGCG undergoes auto-oxidation, the cells may have encountered oxidative stress leading to cell death.[268]

## *4.4. Conclusion*

Overall, 727 compounds in the NCC compound library were screened with the aim of repositioning existing drugs as INIs. CEF and EGCG were identified as INI through SPA screening and were therefore further investigated. Computational studies confirmed the drug-like properties of CEF and EGCG as indicated by

previous studies. Molecular modelling studies predicted that CEF and EGCG interacted with HIV-1 IN via hydrogen bond interactions. No cross-resistance of CEF and EGCG was observed with RAL viral mutations which indicated that these compounds do not share a binding mechanism of action with RAL. Orthogonal screening indicated that the activity of both CEF and EGCG was dependent on the DTT concentration. The IN ST inhibitory effect of CEF was not significant in *in vitro* phenotypic inhibition screening and therefore it can be concluded that CEF was an IN dimer disruptor in the presence of 10mM DTT. Meanwhile, the antiretroviral effect of EGCG on HIV could not be validated since EGCG demonstrated cytotoxicity. This study could not validate whether EGCG was a true INI however, our results corroborate the findings from previous studies showing that EGCG is an INI with an unknown mechanism of action.

## *4.5. Future studies*

The findings from the present study can be advanced by conducting experiments for:

1) Generating a quench curve on the scintillation counter in order to screen the remaining NCC compounds that posses colour as potential INSTIs.

2) Evaluating the resistance mutation profiles of CEF and EGCG against the remaining RAL resistance pathways (N155H, N155H/E92Q or T97A, Y143C/R, Y143C/R coupled with T97A) to determine whether IN mutations responsible for resistance to RAL confer cross resistance to CEF and EGCG.

3) Further optimization of the manual SPA to evade the identification of non-specific HITS.

4) Conducting dimerization assays to determine the multimerization of IN in the presence of EGCG and CEF with a range of DTT concentrations.

5) Identifying the structural moieties of CEF, AMP and EGCG involved in IN binding as well as the residues to which these compounds bind and the subsequent mechanism of action.

6) Screening CEF against other HIV-1 drug targets to determine whether CEF is specific to IN.

# *REFERENCES*


1. Barré-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science. 1983 May 20;220(4599):868–71.

2. Durack D. Opportunistic infections and Kaposi's sarcoma in homosexual men. N Engl J Med. 1981;305(24):1465–1467.

3. Gallo RC, Salahuddin SZ, Popovic M, Shearer GM, Kaplan M, Haynes BF, et al. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. Science. 1984 May 4;224(4648):500–3.

4. Zhu T, Korber BT, Nahmias AJ, Hooper E, Sharp PM, Ho DD. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. Nature. 1998 Feb 5;391(6667):594–7.

5. Centers for Disease Control (CDC). Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men--New York City and California. MMWR Morb Mortal Wkly Rep. 1981 Jul 3;30(25):305–8.

6. Gottlieb MS, Schroff R, Schanker HM, Weisman JD, Fan PT, Wolf RA, et al. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. N Engl J Med. 1981 Dec 10;305(24):1425–31.

7. Popovic M, Sarngadharan MG, Read E, Gallo RC. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science. 1984 May 4;224(4648):497–500.

8. Masur H, Michelis MA, Greene JB, Onorato I, Stouwe RA, Holzman RS, et al. An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction. N Engl J Med. 1981 Dec 10;305(24):1431–8.

9. Mann JM, Gottlieb MS, Mildvan D, Jeffries DJ, Pinching AJ. AIDS: A worldwide pandemic. New Jersey: John Wiley & Sons; 1989.

10. Klecker RW, Collins JM, Yarchoan R, Thomas R, Jenkins JF, Broder S, et al. Plasma and cerebrospinal fluid pharmacokinetics of 3′-azido-3′-deoxythymidine: a novel pyrimidine analog with potential application for the treatment of patients with AIDS and related diseases. Clin Pharmacol Ther. 1987 Apr;41(4):407–12.

11. UNAIDS global report 2013 [Internet]. UNAIDS. [cited 2013 Oct 4]. Available from: http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2013/gr2013/UNAIDS_Global_Report_2013_en.pdf

12. Peeters M, Honoré C, Huet T, Bedjabaga L, Ossari S, Bussi P, et al. Isolation and partial characterization of an HIV-related virus occurring naturally in chimpanzees in Gabon. AIDS Lond Engl. 1989 Oct;3(10):625–30.

13. Huet T, Cheynier R, Meyerhans A, Roelants G, Wain-Hobson S. Genetic organization of a chimpanzee lentivirus related to HIV-1. Nature. 1990 May 24;345(6273):356–9.

14. Sharp PM, Robertson DL, Gao F, Hahn BH. Origins and diversity of human immunodeficiency viruses. AIDS. 1994;8:S27–S28.

15. Fabre P-H, Rodrigues A, Douzery EJP. Patterns of macroevolution among Primates inferred from a supermatrix of mitochondrial and nuclear DNA. Mol Phylogenet Evol. 2009 Dec;53(3):808–25.

16. Chakrabarti L, Guyader M, Alizon M, Daniel MD, Desrosiers RC, Tiollais P, et al. Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses. Nature. 1987 Aug 6;328(6130):543–7.

17. Guyader M, Emerman M, Sonigo P, Clavel F, Montagnier L, Alizon M. Genome organization and transactivation of the human immunodeficiency virus type 2. Nature. 1987 Apr 16;326(6114):662–9.

18. Van Heuverswyn F, Li Y, Bailes E, Neel C, Lafay B, Keele BF, et al. Genetic diversity and phylogeographic clustering of SIVcpzPtt in wild chimpanzees in Cameroon. Virology. 2007 Nov 10;368(1):155–71.

19. Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, et al. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature. 1999 Feb 4;397(6718):436–41.

20. Vanden Haesevelde MM, Peeters M, Jannes G, Janssens W, van der Groen G, Sharp PM, et al. Sequence analysis of a highly divergent HIV-1-related lentivirus isolated from a wild captured chimpanzee. Virology. 1996 Jul 15;221(2):346–50.

21. Santiago ML, Bibollet-Ruche F, Bailes E, Kamenya S, Muller MN, Lukasik M, et al. Amplification of a Complete Simian Immunodeficiency Virus Genome from Fecal RNA of a Wild Chimpanzee. J Virol. 2003 Feb 1;77(3):2233–42.

22. Hirsch VM, Olmsted RA, Murphey-Corb M, Purcell RH, Johnson PR. An African primate lentivirus (SIVsm) closely related to HIV-2. Nature. 1989 Jun 1;339(6223):389–92.

23. Gelderblom HR, Özel M, Hausmann EHS, Winkel T, Pauli G, Koch MA. Fine structure of human immunodeficiency virus (HIV), immunolocalization of structural proteins and virus-cell relation. Micron Microsc Acta. 1988;19(1):41–60.

24. Smith JA, Daniel R. Following the path of the virus: the exploitation of host DNA repair mechanisms by retroviruses. ACS Chem Biol. 2006 May 23;1(4):217–26.

25. De Leys R, Vanderborght B, Vanden Haesevelde M, Heyndrickx L, van Geel A, Wauters C, et al. Isolation and partial characterization of an unusual human immunodeficiency retrovirus from two persons of west-central African origin. J Virol. 1990 Mar;64(3):1207–16.

26. Gürtler LG, Hauser PH, Eberle J, von Brunn A, Knapp S, Zekeng L, et al. A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. J Virol. 1994 Mar;68(3):1581–5.

27. Plantier J-C, Leoz M, Dickerson JE, De Oliveira F, Cordonnier F, Lemée V, et al. A new human immunodeficiency virus derived from gorillas. Nat Med. 2009 Aug;15(8):871–2.

28. Mauclère P, Loussert-Ajaka I, Damond F, Fagot P, Souquières S, Monny Lobe M, et al. Serological and virological characterization of HIV-1 group O infection in Cameroon. AIDS Lond Engl. 1997 Mar 15;11(4):445–53.

29. Peeters M, Gueye A, Mboup S, Bibollet-Ruche F, Ekaza E, Mulanga C, et al. Geographical distribution of HIV-1 group O viruses in Africa. AIDS Lond Engl. 1997 Mar 15;11(4):493–8.

30. Vallari A, Bodelle P, Ngansop C, Makamche F, Ndembi N, Mbanya D, et al. Four new HIV-1 group N isolates from Cameroon: prevalence continues to be low. AIDS Res Hum Retroviruses. 2010 Jan;26(1):109–15.

31. Vallari A, Holzmayer V, Harris B, Yamaguchi J, Ngansop C, Makamche F, et al. Confirmation of putative HIV-1 group P in Cameroon. J Virol. 2011 Feb;85(3):1403–7.

32. Sharp PM, Hahn BH. Origins of HIV and the AIDS Pandemic. Cold Spring Harb Perspect Med [Internet]. 2011 Sep 1 [cited 2013 Sep 12];1(1). Available from: http://perspectivesinmedicine.cshlp.org/content/1/1/a006841

33. Keele BF, Van Heuverswyn F, Li Y, Bailes E, Takehisa J, Santiago ML, et al. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. Science. 2006 Jul 28;313(5786):523–6.

34. HIV and SIV nomenclature [Internet]. [cited 2013 Jul 12]. Available from: http://www.hiv.lanl.gov/content/sequence/HelpDocs/subtypes-more.html

35. Avert: HIV types, subtypes, groups and strains [Internet]. [cited 2013 Jul 12]. Available from: http://www.avert.org/hiv-types.htm

36. Gilbert MTP, Rambaut A, Wlasiuk G, Spira TJ, Pitchenik AE, Worobey M. The emergence of HIV/AIDS in the Americas and beyond. Proc Natl Acad Sci U S A. 2007 Nov 20;104(47):18566–70.

37. Goudsmit J. Viral Sex: The Nature of AIDS. Oxford University Press; 1998. 292 p.

38. McCutchan FE. Global epidemiology of HIV. J Med Virol. 2006;78 Suppl 1:S7–S12.

39. De Silva TI, Cotten M, Rowland-Jones SL. HIV-2: the forgotten AIDS virus. Trends Microbiol. 2008 Dec;16(12):588–95.

40. Peeters M, Toure-Kane C, Nkengasong JN. Genetic diversity of HIV in Africa: impact on diagnosis, treatment, vaccine development and trials. AIDS Lond Engl. 2003 Dec 5;17(18):2547–60.

41. Damond F, Descamps D, Farfara I, Telles JN, Puyeo S, Campa P, et al. Quantification of proviral load of human immunodeficiency virus type 2 subtypes A and B using real-time PCR. J Clin Microbiol. 2001 Dec;39(12):4264–8.

42. Hamel DJ, Sankalé J-L, Eisen G, Meloni ST, Mullins C, Gueye-Ndiaye A, et al. Twenty years of prospective molecular epidemiology in Senegal: changes in HIV diversity. AIDS Res Hum Retroviruses. 2007 Oct;23(10):1189–96.

43. Van der Loeff MFS, Awasana AA, Sarge-Njie R, van der Sande M, Jaye A, Sabally S, et al. Sixteen years of HIV surveillance in a West African research clinic reveals divergent epidemic trends of HIV-1 and HIV-2. Int J Epidemiol. 2006 Oct;35(5):1322–8.

44. Lytle CD, Tondreau SC, Truscott W, Budacz AP, Kuester RK, Venegas L, et al. Filtration sizes of human immunodeficiency virus type 1 and surrogate viruses used to test barrier materials. Appl Environ Microbiol. 1992 Feb;58(2):747–9.

45. Ozel M, Pauli G, Gelderblom HR. The organization of the envelope projections on the surface of HIV. Arch Virol. 1988;100(3-4):255–66.

46. Hallenberger S, Moulard M, Sordel M, Klenk HD, Garten W. The role of eukaryotic subtilisin-like endoproteases for the activation of human immunodeficiency virus glycoproteins in natural host cells. J Virol. 1997 Feb;71(2):1036–45.

47. Gelderblom HR. Assembly and morphology of HIV: potential effect of structure on viral function. AIDS Lond Engl. 1991 Jun;5(6):617–37.

48. Gelderblom DHR, Özel M, Pauli G. Morphogenesis and morphology of HIV structure-function relations. Arch Virol. 1989 Mar 1;106(1-2):1–13.

49. Human immuno deficiency: A global pandemic [Internet]. Stanford University. [cited 2013 Jan 10]. Available from: http://www.stanford.edu/group/virus/retro/2005gongishmail/HIV.html

50. Göttlinger HG, Sodroski JG, Haseltine WA. Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. Proc Natl Acad Sci U S A. 1989 Aug;86(15):5781–5.

51. Bernstein HB, Tucker SP, Kar SR, McPherson SA, McPherson DT, Dubay JW, et al. Oligomerization of the hydrophobic heptad repeat of gp41. J Virol. 1995 May;69(5):2745–50.

52. Strebel K, Daugherty D, Clouse K, Cohen D, Folks T, Martin MA. The HIV “A” (sor) gene product is essential for virus infectivity. Nature. 1987 Aug 20;328(6132):728–30.

53. Sherman MP, Greene WC. Slipping through the door: HIV entry into the nucleus. Microbes Infect Inst Pasteur. 2002 Jan;4(1):67–73.

54. Jacks T, Power MD, Masiarz FR, Luciw PA, Barr PJ, Varmus HE. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. Nature. 1988 Jan 21;331(6153):280–3.

55. Ruben S, Perkins A, Purcell R, Joung K, Sia R, Burghoff R, et al. Structural and functional characterization of human immunodeficiency virus tat protein. J Virol. 1989 Jan;63(1):1–8.

56. Zapp ML, Green MR. Sequence-specific RNA binding by the HIV-1 Rev protein. Nature. 1989 Dec 7;342(6250):714–6.

57. Malim MH, Hauber J, Le SY, Maizel JV, Cullen BR. The HIV-1 rev trans-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. Nature. 1989 Mar 16;338(6212):254–7.

58. Felber BK, Drysdale CM, Pavlakis GN. Feedback regulation of human immunodeficiency virus type 1 expression by the Rev protein. J Virol. 1990 Aug;64(8):3734–41.

59. Seelamgari A, Maddukuri A, Berro R, de la Fuente C, Kehn K, Deng L, et al. Role of viral regulatory and accessory proteins in HIV-1 replication. Front Biosci J Virtual Libr. 2004 Sep 1;9:2388–413.

60. Glushakova S, Münch J, Carl S, Greenough TC, Sullivan JL, Margolis L, et al. CD4 down-modulation by human immunodeficiency virus type 1 Nef correlates with the efficiency of viral replication and with CD4(+) T-cell depletion in human lymphoid tissue ex vivo. J Virol. 2001 Nov;75(21):10113–7.

61. Garcia JV, Miller AD. Downregulation of cell surface CD4 by nef. Res Virol. 1992 Feb;143(1):52–5.

62. Luria S, Chambers I, Berg P. Expression of the type 1 human immunodeficiency virus Nef protein in T cells prevents antigen receptor-mediated induction of interleukin 2 mRNA. Proc Natl Acad Sci U S A. 1991 Jun 15;88(12):5326–30.

63. Miller MD, Warmerdam MT, Gaston I, Greene WC, Feinberg MB. The human immunodeficiency virus-1 nef gene product: a positive factor for viral infection and replication in primary lymphocytes and macrophages. J Exp Med. 1994 Jan 1;179(1):101–13.

64. Stanley BJ, Ehrlich ES, Short L, Yu Y, Xiao Z, Yu X-F, et al. Structural insight into the human immunodeficiency virus Vif SOCS box and its role in human E3 ubiquitin ligase assembly. J Virol. 2008 Sep;82(17):8656–63.

65. Heinzinger NK, Bukinsky MI, Haggerty SA, Ragland AM, Kewalramani V, Lee MA, et al. The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells. Proc Natl Acad Sci U S A. 1994 Jul 19;91(15):7311–5.

66. Willey RL, Maldarelli F, Martin MA, Strebel K. Human immunodeficiency virus type 1 Vpu protein induces rapid degradation of CD4. J Virol. 1992 Dec;66(12):7193–200.

67. Klimkait T, Strebel K, Hoggan MD, Martin MA, Orenstein JM. The human immunodeficiency virus type 1-specific protein vpu is required for efficient virus maturation and release. J Virol. 1990 Feb;64(2):621–9.

68. Haase AT. Population biology of HIV-1 infection: viral and CD4+ T cell demographics and dynamics in lymphatic tissues. Annu Rev Immunol. 1999;17:625–56.

69. Piguet V, Steinman RM. The interaction of HIV with dendritic cells: outcomes and pathways. Trends Immunol. 2007 Nov;28(11):503–10.

70. Trapp S, Turville SG, Robbiani M. Slamming the door on unwanted guests: why preemptive strikes at the mucosa may be the best strategy against HIV. J Leukoc Biol. 2006 Nov;80(5):1076–83.

71. Aziz S, Fackler OT, Meyerhans A, Müller-Lantzsch N, Zeitz M, Schneider T. Replication of M-tropic HIV-1 in activated human intestinal lamina propria lymphocytes is the main reason for increased virus load in the intestinal mucosa. J Acquir Immune Defic Syndr 1999. 2005 Jan 1;38(1):23–30.

72. Groot F, Welsch S, Sattentau QJ. Efficient HIV-1 transmission from macrophages to T cells across transient virological synapses. Blood. 2008 May 1;111(9):4660–3.

73. Samson M, Libert F, Doranz BJ, Rucker J, Liesnard C, Farber CM, et al. Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature. 1996 Aug 22;382(6593):722–5.

74. Chen L, Kwon YD, Zhou T, Wu X, O'Dell S, Cavacini L, et al. Structural basis of immune evasion at the site of CD4 attachment on HIV-1 gp120. Science. 2009 Nov 20;326(5956):1123–7.

75. Wyatt R, Kwong PD, Desjardins E, Sweet RW, Robinson J, Hendrickson WA, et al. The antigenic structure of the HIV gp120 envelope glycoprotein. Nature. 1998 Jun 18;393(6686):705–11.

76. Liu J, Bartesaghi A, Borgnia MJ, Sapiro G, Subramaniam S. Molecular architecture of native HIV-1 gp 120 trimers. Nature. 2008 Sep 4;455(7209):109–13.

77. Goto T, Nakai M, Ikuta K. The life-cycle of human immunodeficiency virus type 1. Micron Oxf Engl 1993. 1998 Jun;29(2-3):123–38.

78. Haseltine WA. Molecular biology of the human immunodeficiency virus type 1. FASEB J. 1991 Jul 1;5(10):2349–60.

79. Raghavendra NK, Shkriabai N, Graham RL, Hess S, Kvaratskhelia M, Wu L. Identification of host proteins associated with HIV-1 preintegration complexes isolated from infected CD4+ cells. Retrovirology. 2010 Aug 11;7(1):1-7.

80. Hamamoto S, Nishitsuji H, Amagasa T, Kannagi M, Masuda T. Identification of a novel human immunodeficiency virus type 1 integrase interactor, Gemin2, that facilitates efficient viral cDNA synthesis in Vivo. J Virol. 2006 Jun;80(12):5670–7.

81. Cereseto A, Manganaro L, Gutierrez MI, Terreni M, Fittipaldi A, Lusic M, et al. Acetylation of HIV-1 integrase by p300 regulates viral integration. EMBO J. 2005 Sep 7;24(17):3070–81.

82. Farnet CM, Bushman FD. HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes in vitro. Cell. 1997 Feb 21;88(4):483–92.

83. Parissi V, Calmels C, De Soultrait VR, Caumont A, Fournier M, Chaignepain S, et al. Functional interactions of human immunodeficiency virus type 1 integrase with human and yeast HSP60. J Virol. 2001 Dec;75(23):11344–53.

84. Violot S, Hong SS, Rakotobe D, Petit C, Gay B, Moreau K, et al. The human polycomb group eed protein interacts with the integrase of human immunodeficiency virus type 1. J Virol. 2003 Dec;77(23):12507–22.

85. Ao Z, Huang G, Yao H, Xu Z, Labine M, Cochrane AW, et al. Interaction of human immunodeficiency virus type 1 integrase with cellular nuclear import receptor importin 7 and its impact on viral replication. J Biol Chem. 2007 May 4;282(18):13456–67.

86. Kalpana GV, Marmon S, Wang W, Crabtree GR, Goff SP. Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. Science. 1994 Dec 23;266(5193):2002–6.

87. Cherepanov P, Maertens G, Proost P, Devreese B, Van Beeumen J, Engelborghs Y, et al. HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. J Biol Chem. 2003 Jan 3;278(1):372–81.

88. Willetts KE, Rey F, Agostini I, Navarro JM, Baudat Y, Vigne R, et al. DNA repair enzyme uracil DNA glycosylase is specifically incorporated into human immunodeficiency virus type 1 viral particles through a Vpr-independent mechanism. J Virol. 1999 Feb;73(2):1682–8.

89. Miller RJ, Cairns JS, Bridges S, Sarver N. Human immunodeficiency virus and AIDS: insights from animal lentiviruses. J Virol. 2000 Aug;74(16):7187–95.

90. Kolegraff K, Bostik P, Ansari AA. Characterization and role of lentivirus-associated host proteins. Exp Biol Med Maywood NJ. 2006 Mar;231(3):252–63.

91. HIV replication cycle [Internet]. [cited 2013 Jul 14]. Available from: http://www.niaid.nih.gov/topics/HIVAIDS/Understanding/Biology/pages/hivreplicationcycle.aspx

92. Reeves JD, Piefer AJ. Emerging drug targets for antiretroviral therapy. Drugs. 2005;65(13):1747–66.

93. Lutzke RA, Plasterk RH. HIV integrase: a target for drug discovery. Genes Funct. 1997 Dec;1(5-6):289–307.

94. De Clercq E. Anti-HIV drugs: 25 compounds approved within 25 years after the discovery of HIV. Int J Antimicrob Agents. 2009 Apr;33(4):307–20.

95. Broder S. The development of antiretroviral therapy and its impact on the HIV-1/AIDS pandemic. Antiviral Res. 2010 Jan;85(1):1–18.

96. Yarchoan R, Weinhold K, Lyerly HK, Gelmann E, Blum R, Shearer G, et al. Administration of 3'-azido-3'-deoxythymidine, an inhibitor of htlv-iii/lav replication, to patients with aids or aids-related complex. The Lancet. 1986 Mar 15;327(8481):575–80.

97. Goody RS, Müller B, Restle T. Factors contributing to the inhibition of HIV reverse transcriptase by chain-terminating nucleotides in vitro and in vivo. FEBS Lett. 1991 Oct 7;291(1):1–5.

98. Tantillo C, Ding J, Jacobo-Molina A, Nanni RG, Boyer PL, Hughes SH, et al. Locations of anti-aids drug binding sites and resistance mutations in the three-dimensional structure of hiv-1 reverse transcriptase: implications for mechanisms of drug inhibition and resistance. J Mol Biol. 1994 Oct 27;243(3):369–87.

99. Spence RA, Kati WM, Anderson KS, Johnson KA. Mechanism of inhibition of HIV-1 reverse transcriptase by nonnucleoside inhibitors. Science. 1995 Feb 17;267(5200):988–93.

100. HIV and AIDS Activities - Antiretroviral drugs used in the treatment of HIV infection [Internet]. [cited 2013 Jul 16]. Available from: http://www.fda.gov/ForConsumers/byAudience/ForPatientAdvocates/HIVandAIDSActivities/ucm118915.htm

101. Esnouf R, Ren J, Ross C, Jones Y, Stammers D, Stuart D. Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors. Nat Struct Mol Biol. 1995 Apr;2(4):303–8.

102. De Clercq E. The role of non-nucleoside reverse transcriptase inhibitors (NNRTIs) in the therapy of HIV-1 infection. Antiviral Res. 1998 Jun 1;38(3):153–79.

103. Pauwels R. New non-nucleoside reverse transcriptase inhibitors (NNRTIs) in development for the treatment of HIV infections. Curr Opin Pharmacol. 2004 Oct;4(5):437–46.

104. Wensing AMJ, van Maarseveen NM, Nijhuis M. Fifteen years of HIV protease inhibitors: raising the barrier to resistance. Antiviral Res. 2010 Jan;85(1):59–74.

105. Lieberman-Blum SS, Fung HB, Bandres JC. Maraviroc: a CCR5-receptor antagonist for the treatment of HIV-1 infection. Clin Ther. 2008 Jul;30(7):1228–50.

106. Bai Y, Xue H, Wang K, Cai L, Qiu J, Bi S, et al. Covalent fusion inhibitors targeting HIV-1 gp41 deep pocket. Amino Acids. 2013 Feb;44(2):701–13.

107. D' Angelo J, Mouscadet JF, Desmaele D, Zouhiri F, Leh H. HIV-1 integrase: the next target for AIDS therapy? Pathol Biol. 2001;49(3):237–46.

108. Summa V, Petrocchi A, Bonelli F, Crescenzi B, Donghi M, Ferrara M, et al. Discovery of raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection. J Med Chem. 2008 Sep 25;51(18):5843–55.

109. HIV and AIDS Activities - FDA approval of Isentress (raltegravir) [Internet]. [cited 2013 Jul 17]. Available from: http://www.fda.gov/ForConsumers/ByAudience/ForPatientAdvocates/HIVandAIDSActivities/ucm124040.htm

110. Serrao E, Odde S, Ramkumar K, Neamati N. Raltegravir, elvitegravir, and metoogravir: the birth of "me-too" HIV-1 integrase inhibitors. Retrovirology. 2009 Mar 5;6(1):25.

111. Elion R, Gathe J, Rashbaum B, Shalit P, Hawkins T, Liu H, et al. The single-tablet regimen elvitegravir/cobicistat/emtricitabine/tenofovir disoproxil fumarate (evg/cobi/ftc/tdf; “quad”) maintains a high rate of virologic suppression, and cobicistat (cobi) is an effective pharmacoenhancer through 48 weeks. 50th Interscience conference on antimicrobial agents and chemotherapy. Washington DC, USA; 2010.

112. Xu L, Liu H, Murray BP, Callebaut C, Lee MS, Hong A, et al. Cobicistat (GS-9350): a potent and selective inhibitor of human cyp3a as a novel pharmacoenhancer. ACS Med Chem Lett. 2010 Aug 12;1(5):209–13.

113. Cohen C, Elion R, Ruane P, Shamblaw D, DeJesus E, Rashbaum B, et al. Randomized, phase 2 evaluation of two single-tablet regimens elvitegravir/cobicistat/emtricitabine/tenofovir disoproxil fumarate versus efavirenz/emtricitabine/tenofovir disoproxil fumarate for the initial treatment of HIV infection: AIDS. 2011 Mar;25(6):F7–F12.

114. El Annaz H, Recordon-Pinson P, Tagajdid R, Doblali T, Belefquih B, Oumakhir S, et al. Drug resistance mutations in HIV type 1 isolates from patients failing antiretroviral therapy in morocco. AIDS Res Hum Retroviruses. 2012 Aug;28(8):944–8.

115. Mansky LM. HIV mutagenesis and the evolution of antiretroviral drug resistance. Drug Resist Updat Rev Comment Antimicrob Anticancer Chemother. 2002 Dec;5(6):219–23.

116. Palella FJ Jr, Armon C, Buchacz K, Cole SR, Chmiel JS, Novak RM, et al. The association of HIV susceptibility testing with survival among HIV-

infected patients receiving antiretroviral therapy: a cohort study. Ann Intern Med. 2009 Jul 21;151(2):73–84.

117. Hewer R, Kriel FH, Coates J. Drug discovery in Africa - impacts of genomics, natural products, traditional medicines, insights [Internet]. Chibale Kelly, Davies-Coleman Mike, Masimirembwa Collen, editors. Germany; 2012 [cited 2013 Oct 7]. 325-347. Available from: http://www.springer.com/chemistry/book/978-3-642-28174-7

118. The truly staggering cost of inventing new drugs [Internet]. Forbes. [cited 2013 Oct 7]. Available from: http://www.forbes.com/sites/matthewherper/2012/02/10/the-truly-staggering-cost-of-inventing-new-drugs/

119. Hughes J, Rees S, Kalindjian S, Philpott K. Principles of early drug discovery. Br J Pharmacol. 2011 Mar;162(6):1239–49.

120. Hughes JP, Rees S, Kalindjian SB, Philpott KL. Principles of early drug discovery. Br J Pharmacol. 2011 Mar;162(6):1239–49.

121. Huser J. High Throughput Screening in Drug Discovery. Wiley - VCH: Germany. 2006. 35:343

122. Obrecht D, Villalgordo JM. Solid-supported Combinatorial and Parallel Synthesis of Small-molecular-weight Compound Libraries. Elsevier; UK. 1998. 1(17):360.

123. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. Nat Rev Drug Discov. 2004 Aug;3(8):673–83.

124. Abel U, Koch C, Speitling M, Hansske FG. Modern methods to produce natural-product libraries. Curr Opin Chem Biol. 2002 Aug;6(4):453–8.

125. Newman DJ, Cragg GM. Natural products as sources of new drugs over the last 25 years. J Nat Prod. 2007 Mar 1;70(3):461–77.

126. Chin Y-W, Balunas MJ, Chai HB, Kinghorn AD. Drug discovery from natural sources. AAPS J. 2006 Jun 1;8(2):E239–E253.

127. Molinari G. Natural products in drug discovery: present status and perspectives. Pharmaceutical Biotechnology [Internet]. Springer New York; 2009 [cited 2013 Oct 7]. 13–27. Available from: http://link.springer.com/chapter/10.1007/978-1-4419-1132-2_2

128. Stokes T. AIDS herbal therapy. Trends Plant Sci. 2002 Feb;7(2):57–57.

129. Boppana K, Dubey PK, Jagarlapudi SARP, Vadivelan S, Rambabu G. Knowledge based identification of MAO-B selective inhibitors using pharmacophore and structure based virtual screening models. Eur J Med Chem. 2009 Sep;44(9):3584–90.

130. Ghosh S, Nie A, An J, Huang Z. Structure-based virtual screening of chemical libraries for drug discovery. Curr Opin Chem Biol. 2006 Jun;10(3):194–202.

131. Kroemer RT. Structure-based drug design: docking and scoring. Curr Protein Pept Sci. 2007 Aug;8(4):312–28.

132. Cavasotto CN, Phatak SS. Homology modeling in drug discovery: current trends and applications. Drug Discov Today. 2009 Jul;14(13-14):676–83.

133. Reviews in computational chemistry [Internet]. [cited 2013 Oct 7]. Available from: http://www.scribd.com/doc/147904717/Reviews-in-Computational-Chemistry

134. Acharya C, Coop A, Polli JE, Mackerell AD Jr. Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. Curr Comput Aided Drug Des. 2011 Mar;7(1):10–22.

135. Zhang S. Computer-aided drug discovery and development. In: Satyanarayanajois SD, editor. Drug Design and Discovery [Internet]. Humana Press; 2011 [cited 2013 Jul 22]. p. 23–38. Available from: http://link.springer.com/protocol/10.1007/978-1-61779-012-6_2

136. Golan DE. Principles of pharmacology: the pathophysiologic basis of drug therapy. Lippincott Williams & Wilkins; USA. 2008. 2:1012.

137. Eddershaw PJ, Beresford AP, Bayliss MK. ADME/PK as part of a rational approach to drug discovery. Drug Discov Today. 2000 Sep 1;5(9):409–14.

138. Borchardt RT, Hidalgo IJ, Hillgren KM, Hu M. Pharmaceutical applications of cell culture: an overview. In: Wilson G, Davis SS, Illum L, Zweibaum A, editors. Pharmaceutical applications of cell and tissue culture to drug transport [Internet]. Springer New York; 1992 [cited 2013 Sep 16]. 1–14. Available from: http://link.springer.com/chapter/10.1007/978-1-4757-0286-6_1

139. Hildago I. Characterization of the human colon carcinoma cell line (Caco-2) as a model system for intestinal epithelial permeability. Gastroenterology. 1989;(96):736–49.

140. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev. 1997 Jan 15;23(1–3):3–25.

141. Li AP. Screening for human ADME/Tox drug properties in drug discovery. Drug Discov Today. 2001 Apr 1;6(7):357–66.

142. Coleman RA, Bowen WP, Baines IA, Woodrooffe AJ, Brown AM. Use of human tissue in ADME and safety profiling of development candidates. Drug Discov Today. 2001 Nov 1;6(21):1116–26.

143. Wrighton SA, Stevens JC. The human hepatic cytochromes p450 involved in drug metabolism. Crit Rev Toxicol. 1992 Jan;22(1):1–21.

144. Li AP. Primary hepatocyte cultures as an in vitro experimental model for the evaluation of pharmacokinetic drug–drug interactions. Advances in Pharmacology [Internet]. Academic Press; 1997 [cited 2013 Sep 16]. 103–30. Available from: http://www.sciencedirect.com/science/article/pii/S1054358908602033

145. Martis EA, Radhakrishnan R, Badve RR. High-throughput screening: the hits and leads of drug discovery-an overview. J Appl Pharm Sci. 2011;1(1):2–10.

146. Sundberg SA. High-throughput and ultra-high-throughput screening: solution-and cell-based approaches. Curr Opin Biotechnol. 2000;11(1):47–53.

147. Wölcke J, Ullmann D. Miniaturized HTS technologies – uHTS. Drug Discov Today. 2001 Jun 15;6(12):637–46.

148. Janzen W. The impact automation on drug discovery [Internet]. Netsci.org. [cited 2013 Sep 15]. Available from: http://www.netsci.org/Science/Screening/feature08.html

149. Houston JG, Banks M. The chemical-biological interface: developments in automated and miniaturised screening technology. Curr Opin Biotechnol. 1997 Dec;8(6):734–40.

150. Harding D, Banks M, Fogarty S, Binnie A. Development of an automated high-throughput screening system: a case history. Drug Discov Today. 1997 Sep;2(9):385–90.

151. Chapman T. Lab automation and robotics: automation on the move. Nature. 2003 Feb 6;421(6923):661–6.

152. Tartaglia LA. Complementary new approaches enable repositioning of failed drug candidates. Expert Opin Investig Drugs. 2006 Nov;15(11):1295–8.

153. Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, et al. How to improve R&;D productivity: the pharmaceutical industry's grand challenge. Nat Rev Drug Discov [Internet]. 2010 Nov 19 [cited 2013 Oct 29]; Available from: http://www.nature.com/doifinder/10.1038/nrd3078

154. NIH Clinical Collection [Internet]. [cited 2013 Apr 2]. Available from: www.nihclinicalcollection.com

155. Corcoran LJ, Mitchison TJ, Liu Q. A novel action of histone deacetylase inhibitors in a protein aggresome disease model. Curr Biol CB. 2004 Mar 23;14(6):488–92.

156. Stavrovskaya IG, Narayanan MV, Zhang W, Krasnikov BF, Heemskerk J, Young SS, et al. Clinically approved heterocyclics act on a mitochondrial target and reduce stroke-induced pathology. J Exp Med. 2004 Jul 19;200(2):211–22.

157. Lunn MR, Root DE, Martino AM, Flaherty SP, Kelley BP, Coovert DD, et al. Indoprofen upregulates the survival motor neuron protein through a cyclooxygenase-independent mechanism. Chem Biol. 2004 Nov;11(11):1489–93.

158. Rothstein JD, Patel S, Regan MR, Haenggeli C, Huang YH, Bergles DE, et al. Beta-lactam antibiotics offer neuroprotection by increasing glutamate transporter expression. Nature. 2005 Jan 6;433(7021):73–7.

159. Rice PA, Baker TA. Comparative architecture of transposase and integrase complexes. Nat Struct Mol Biol. 2001 Apr;8(4):302–7.

160. Engelman A, Bushman FD, Craigie R. Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. EMBO J. 1993 Aug;12(8):3269–75.

161. Engelman A, Craigie R. Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro. J Virol. 1992 Nov 1;66(11):6361–9.

162. Van Gent DC, Vink C, Groeneger AA, Plasterk RH. Complementation between HIV integrase proteins mutated in different domains. EMBO J. 1993 Aug;12(8):3261–7.

163. Esposito D, Craigie R. HIV integrase structure and function. Adv Virus Res. 1999;52:319–33.

164. Kulkosky J, Jones KS, Katz RA, Mack JP, Skalka AM. Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. Mol Cell Biol. 1992 May;12(5):2331–8.

165. Polard P, Chandler M. Bacterial transposases and retroviral integrases. Mol Microbiol. 1995 Jan;15(1):13–23.

166. Rowland SJ, Sherratt DJ, Stark WM, Boocock MR. Tn552 transposase purification and in vitro activities. EMBO J. 1995 Jan 3;14(1):196–205.

167. McColl DJ, Chen X. Strand transfer inhibitors of HIV-1 integrase: Bringing IN a new era of antiretroviral therapy. Antiviral Res. 2010 Jan;85(1):101–18.

168. Lee SP, Xiao J, Knutson JR, Lewis MS, Han MK. Zn2+ promotes the self-association of human immunodeficiency virus type-1 integrase in vitro. Biochemistry (Mosc). 1997 Jan 7;36(1):173–80.

169. Mathew K. HIV-1 Integrase: Structure and Function [Internet]. Molecular Anatomy Project. 2011 [cited 2013 Oct 29]. Available from: http://maptest.rutgers.edu/drupal/?q=node/436

170. Yoder KE, Bushman FD. Repair of Gaps in Retroviral DNA Integration Intermediates. J Virol. 2000 Dec 1;74(23):11191–200.

171. Brin E, Yi J, Skalka A, Leis J. Modeling the late steps in HIV-1 retroviral integrase-catalyzed DNA integration. J Biol Chem. 2000;275(50):39287.

172. Engelman A, Mizuuchi K, Craigie R. HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. Cell. 1991 Dec 20;67(6):1211–21.

173. Suzuki Y, Suzuki Y. Role of host-encoded proteins in restriction of retroviral integration. Front Virol. 2012;3:227.

174. Espeseth AS, Felock P, Wolfe A, Witmer M, Grobler J, Anthony N, et al. HIV-1 integrase inhibitors that compete with the target DNA substrate define a unique strand transfer conformation for integrase. Proc Natl Acad Sci. 2000;97(21):11244–9.

175. Summa V, Petrocchi A, Matassa VG, Gardelli C, Muraglia E, Rowley M, et al. 4,5-dihydroxypyrimidine carboxamides and N-alkyl-5-hydroxypyrimidinone carboxamides are potent, selective HIV integrase

inhibitors with good pharmacokinetic profiles in preclinical species. J Med Chem. 2006 Nov 16;49(23):6646–9.

176. Hazuda DJ, Felock P, Witmer M, Wolfe A, Stillmock K, Grobler JA, et al. Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. Science. 2000 Jan 28;287(5453):646–50.

177. Marchand C, Maddali K, Metifiot M, Pommier Y. HIV-1 IN Inhibitors: 2010 Update and Perspectives. Curr Top Med Chem. 2009;9(11):1016–37.

178. Grobler JA, Stillmock K, Hu B, Witmer M, Felock P, Espeseth AS, et al. Diketo acid inhibitor mechanism and HIV-1 integrase: implications for metal binding in the active site of phosphotransferase enzymes. Proc Natl Acad Sci. 2002;99(10):6661–6.

179. Kobayashi M, Yoshinaga T, Seki T, Wakasa-Morimoto C, Brown KW, Ferris R, et al. In Vitro Antiretroviral Properties of S/GSK1349572, a Next-Generation HIV Integrase Inhibitor. Antimicrob Agents Chemother. 2010 Nov 29;55(2):813–21.

180. Lee DJ, Robinson WE. Human Immunodeficiency Virus Type 1 (HIV-1) integrase: resistance to diketo acid integrase inhibitors impairs hiv-1 replication and integration and confers cross-resistance to l-chicoric Acid. J Virol. 2004 Jun 1;78(11):5835–47.

181. Agrawal A, DeSoto J, Fullagar JL, Maddali K, Rostami S, Richman DD, et al. Probing chelation motifs in HIV integrase inhibitors. Proc Natl Acad Sci. 2012 Feb 14;109(7):2251–6.

182. Witmer M, Danovich R. Selection and analysis of HIV-1 integrase strand transfer inhibitor resistant mutant viruses. Methods. 2009 Apr;47(4):277–82.

183. O'Neil PK. Mutation and selection: shaping the hiv-1 genome [Internet]. The University of Utah; 2003 [cited 2013 Oct 8]. Available from: http://content.lib.utah.edu/utils/getfile/collection/etd2/id/94/filename/2089.pdf

184. Gallien S, Delaugerre C, Charreau I, Braun J, Boulet T, Barrail-tran A, et al. Emerging integrase inhibitor resistance mutations in raltegravir-treated HIV-1-infected patients with low-level viremia. AIDS Lond Engl. 2011 Mar 13;25(5):665–9.

185. Baldanti F, Paolucci S, Gulminetti R, Brandolini M, Barbarini G, Maserati R. Early emergence of raltegravir resistance mutations in patients receiving HAART salvage regimens. J Med Virol. 2010 Jan;82(1):116–22.

186. Fikkert V, Hombrouck A, Van Remoortel B, De Maeyer M, Pannecouque C, De Clercq E, et al. Multiple mutations in human immunodeficiency virus-1 integrase confer resistance to the clinical trial drug S-1360. AIDS Lond Engl. 2004 Oct 21;18(15):2019–28.

187. Roquebert B, Blum L, Collin G, Damond F, Peytavin G, Leleu J, et al. Selection of the Q148R integrase inhibitor resistance mutation in a failing raltegravir containing regimen. AIDS Lond Engl. 2008 Oct 1;22(15):2045–6.

188. Fransen S, Gupta S, Danovich R, Hazuda D, Miller M, Witmer M, et al. Loss of raltegravir susceptibility by human immunodeficiency virus type 1 is conferred via multiple nonoverlapping genetic pathways. J Virol. 2009 Nov;83(22):11440–6.

189. Malet I, Delelis O, Valantin M-A, Montes B, Soulie C, Wirden M, et al. Mutations associated with failure of raltegravir treatment affect integrase sensitivity to the inhibitor in vitro. Antimicrob Agents Chemother. 2008 Apr;52(4):1351–8.

190. Delelis O, Malet I, Na L, Tchertanov L, Calvez V, Marcelin A-G, et al. The G140S mutation in HIV integrases from raltegravir-resistant patients rescues catalytic defect due to the resistance Q148H mutation. Nucleic Acids Res. 2008 Dec 18;37(4):1193–201.

191. Blanco J-L, Varghese V, Rhee S-Y, Gatell JM, Shafer RW. HIV-1 integrase inhibitor resistance and its clinical implications. J Infect Dis. 2011 May 1;203(9):1204–14.

192. Fransen S, Gupta S, Frantzell A, Petropoulos CJ, Huang W. Substitutions at amino acid positions 143, 148, and 155 of HIV-1 integrase define distinct genetic barriers to raltegravir resistance in vivo. J Virol. 2012 Jul;86(13):7249–55.

193. Geretti AM, Armenia D, Ceccherini-Silberstein F. Emerging patterns and implications of HIV-1 integrase inhibitor resistance. Curr Opin Infect Dis. 2012 Dec;25(6):677–86.

194. Ramanathan DS, Mathias AA, German P, Kearney BP. Clinical pharmacokinetic and pharmacodynamic profile of the hiv integrase inhibitor elvitegravir. Clin Pharmacokinet. 2011 Apr 1;50(4):229–44.

195. Goethals O, Clayton R, Van Ginderen M, Vereycken I, Wagemans E, Geluykens P, et al. Resistance mutations in human immunodeficiency virus type 1 integrase selected with elvitegravir confer reduced susceptibility to a wide range of integrase inhibitors. J Virol. 2008 Nov;82(21):10366–74.

196. Garrido C, Soriano V, Geretti AM, Zahonero N, Garcia S, Booth C, et al. Resistance associated mutations to dolutegravir (S/GSK1349572) in HIV-infected patients--impact of HIV subtypes and prior raltegravir experience. Antiviral Res. 2011 Jun;90(3):164–7.

197. Quashie PK, Mesplède T, Han Y-S, Oliveira M, Singhroy DN, Fujiwara T, et al. Characterization of the R263K mutation in HIV-1 integrase that confers low-level resistance to the second-generation integrase strand transfer inhibitor dolutegravir. J Virol. 2012 Mar;86(5):2696–705.

198. Mesplede T, Quashie PK, Osman N, Han Y, Singhroy DN, Lie Y, et al. Viral fitness cost prevents HIV-1 from evading dolutegravir drug pressure. Retrovirology. 2013 Feb 22;10:22.

199. Chow S.A. In vitro assays for activities of retroviral integrase. Methods: Companion Methods Enzym. 1997;12(4):306–17.

200. Hazuda DJ, Hastings JC, Wolfe AL, Emini EA. A novel assay for the DNA strand-transfer reaction of HIV-1 integrase. Nucleic Acids Res. 1994;22(6):1121.

201. Xuei X, David CA, Middleton TR, Lim B, Pithawalla R, Chen C-M, et al. Use of sam2® biotin capture membrane in microarrayed compound screening (μarcs) format for nucleic acid polymerization assays. J Biomol Screen. 2003 Jun 1;8(3):273–82.

202. Wang Y, Klock H, Yin H, Wolff K, Bieza K, Niswonger K, et al. Homogeneous high-throughput screening assays for HIV-1 integrase 3beta-processing and strand transfer activities. J Biomol Screen. 2005 Aug;10(5):456–62.

203. Grobler JA, Stillmock KA, Hazuda DJ. Scintillation proximity assays for mechanistic and pharmacological analyses of HIV-1 integration. Methods San Diego Calif. 2009 Apr;47(4):249–53.

204. Savarino A. A historical sketch of the discovery and development of HIV-1 integrase inhibitors. Expert Opin Investig Drugs. 2006 Dec;15(12):1507–22.

205. Christ F, Voet A, Marchand A, Nicolet S, Desimmie BA, Marchand D, et al. Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication. Nat Chem Biol. 2010 Jun;6(6):442–8.

206. Van Loock M, Meersseman G, Van Acker K, Van Den Eynde C, Jochmans D, Van Schoubroeck B, et al. A novel high-throughput cellular screening assay for the discovery of HIV-1 integrase inhibitors. J Virol Methods. 2012 Feb;179(2):396–401.

207. Boissenneault M, Pinard G, Rouleau N, Sevigny M. Rapid, homogeneous and robust HIV-p24 detection assay using the AlphaLISA [Internet]. [cited 2012 Jan 30]. Available from: http://www.perkinelmer.com/CMSResources/Images/44-73783PST_AlphaLISAHIVp24Detection.pdf

208. Wu S, Liu PB. Application of scintillation proximity assay in drug discovery. BioDrugs. 2005 Nov 1;19(6):383–92.

209. Scintillation proximity assays | PerkinElmer [Internet]. PerkinElmer Inc. [cited 2013 Jul 31]. Available from: http://www.perkinelmer.com/resources/technicalresources/applicationsupportknowledgebase/radiometric/proximity.xhtml

210. Marchand C, Johnson AA, Karki RG, Pais GC, Zhang X, Cowansage K, et al. Metal-dependent inhibition of HIV-1 integrase by β-diketo acids and resistance of the soluble double-mutant (F185K/C280S). Mol Pharmacol. 2003;64(3):600–9.

211. Bushman FD, Engelman A, Palmer I, Wingfield P, Craigie R. Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding. Proc Natl Acad Sci. 1993 Apr 15;90(8):3428–32.

212. Qasim M. Development of an HIV-1 integrase enzyme strand transfer assay [Internet]. 2011 [cited 2013 Oct 17]. Available from: http://wiredspace.wits.ac.za/handle/10539/10834

213. Li M, Craigie R. Processing of viral DNA ends channels the HIV-1 integration reaction to concerted integration. J Biol Chem. 2005 Aug 12;280(32):29334–9.

214. Bukrinsky MI, Sharova N, McDonald TL, Pushkarskaya T, Tarpley WG, Stevenson M. Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type 1 with viral nucleic acids following acute infection. Proc Natl Acad Sci U S A. 1993 Jul 1;90(13):6125–9.

215. Sasse J, Gallagher SR. Staining proteins in gels. Curr Protoc Mol Biol. Frederick M Ausubel Al, editors. 2003 Aug;Chapter 10:Unit 10.6.

216. Bronstein I, Voyta JC, Murphy OJ, Bresnick L, Kricka LJ. Improved chemiluminescent western blotting procedure. BioTechniques. 1992 May;12(5):748–53.

217. Gallagher S, Winston SE, Fuller SA, Hurrell JGR. Immunoblotting and immunodetection. Curr Protoc Mol Biol. Frederick M Ausubel Al, editors. 2008 Jul;Chapter 10:Unit 10.8.

218. Grandgenett DP, Goodarzi G. Folding of the multidomain human immunodeficiency virus type-I integrase. Protein Sci. 1994;3(6):888–97.

219. Dicker IB, Terry B, Lin Z, Li Z, Bollini S, Samanta HK, et al. Biochemical analysis of HIV-1 integrase variants resistant to strand transfer inhibitors. J Biol Chem. 2008 Jun 4;283(35):23599–609.

220. Hu Q, Kuki A, Nowlin DM, Plewe MB, Wang H, Zhang J. Hiv-integrase inhibitors, pharmaceutical compositions, and methods for their use. WO2004039803 A2, 2004.

221. Svarovskaia ES, Barr R, Zhang X, Pais GCG, Marchand C, Pommier Y, et al. Azido-containing diketo acid derivatives inhibit human immunodeficiency virus type 1 integrase in vivo and influence the frequency of deletions at two-long-terminal-repeat-circle junctions. J Virol. 2004 Apr;78(7):3210–22.

222. Zhang X, Pais GC., Svarovskaia ES, Marchand C, Johnson AA, Karki RG, et al. Azido-containing aryl β-Diketo acid HIV-1 integrase inhibitors. Bioorg Med Chem Lett. 2003 Mar 24;13(6):1215–9.

223. Devlin JJ, Liang A, Trinh L, Polokoff MA, Senator D, Zheng W, et al. High capacity screening of pooled compounds: identification of the active compound without re-assay of pool members. Drug Dev Res. 1996;37(2):80–5.

224. Mphahlele M, Papathanasopoulos M, Cinellu MA, Coyanis M, Mosebi S, Traut T, et al. Modification of HIV-1 reverse transcriptase and integrase activity by gold(III) complexes in direct biochemical assays. Bioorg Med Chem. 2012 Jan 1;20(1):401–7.

225. Cory AH, Owen TC, Barltrop JA, Cory JG. Use of an aqueous soluble tetrazolium/formazan assay for cell growth assays in culture. Cancer Commun. 1991 Jul;3(7):207–12.

226. O'Doherty U, Swiggard WJ, Malim MH. Human immunodeficiency virus type 1 spinoculation enhances infection through virus binding. J Virol. 2000 Nov;74(21):10074–80.

227. Zhang J-H. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. J Biomol Screen. 1999 Apr 1;4(2):67–73.

228. Temesgen Z, Siraj DS. Raltegravir: first in class HIV integrase inhibitor. Ther Clin Risk Manag. 2008 Apr;4(2):493–500.

229. Delelis O, Carayon K, Saïb A, Deprez E, Mouscadet J-F. Integrase and integration: biochemical activities of HIV-1 integrase. Retrovirology. 2008;5(1):114.

230. Hu Z, Kuritzkes DR. Effect of raltegravir resistance mutations in hiv-1 integrase on viral fitness. J Acquir Immune Defic Syndr 1999. 2010 Oct 1;55(2):148–55.

231. Nance CL, Siwak EB, Shearer WT. Preclinical development of the green tea catechin, epigallocatechin gallate, as an hiv-1 therapy. J Allergy Clin Immunol. 2009 Feb;123(2):459–65.

232. Mphahlele MK. The evaluation of gold-based compounds as potential inhibitors of HIV-1 replication. [Internet]. 2012 [cited 2013 Oct 17]. Available from: http://wiredspace.wits.ac.za/handle/10539/11030

233. Tillekeratne LM., Sherette A, Grossman P, Hupe L, Hupe D, Hudson R. Simplified catechin-gallate inhibitors of HIV-1 reverse transcriptase. Bioorg Med Chem Lett. 2001 Oct 22;11(20):2763–7.

234. Sadiq SK, Wright DW, Kenway OA, Coveney PV. Accurate ensemble molecular dynamics binding free energy ranking of multidrug-resistant HIV-1 proteases. J Chem Inf Model. 2010 May 24;50(5):890–905.

235. Bornhorst JA, Falke JJ. Purification of proteins using polyhistidine affinity tags. Methods Enzymol. 2000;326:245–54.

236. Schröter A, Tränkle C, Mohr K. Modes of allosteric interactions with free and [3H]N-methylscopolamine-occupied muscarinic M2 receptors as deduced from buffer-dependent potency shifts. Naunyn Schmiedebergs Arch Pharmacol. 2000 Dec;362(6):512–9.

237. Ellison V, Brown PO. A stable complex between integrase and viral DNA ends mediates human immunodeficiency virus integration in vitro. Proc Natl Acad Sci U S A. 1994 Jul 19;91(15):7316–20.

238. Wolfe AL, Felock PJ, Hastings JC, Blau CU, Hazuda DJ. The role of manganese in promoting multimerization and assembly of human immunodeficiency virus type 1 integrase as a catalytically active complex on immobilized long terminal repeat substrates. J Virol. 1996 Mar;70(3):1424–32.

239. Shigi Y, Matsumoto Y, Kaizu M, Fujishita Y, Kojo H. Mechanism of action of the new orally active cephalosporin FK027. J Antibiot (Tokyo). 1984 Jul;37(7):790–6.

240. Waxman DJ, Strominger JL. Penicillin-binding proteins and the mechanism of action of beta-lactam antibiotics. Annu Rev Biochem. 1983;52(1):825–69.

241. Sperka T, Pitlik J, Bagossi P, Tözsér J. Beta-lactam compounds as apparently uncompetitive inhibitors of HIV-1 protease. Bioorg Med Chem Lett. 2005 Jun;15(12):3086–90.

242. Sakane K, Kawabata K, Inamoto Y, Yamanaka H, Takaya T. Research and development of new oral cephems, cefixime and cefdinir. Yakugaku Zasshi. 1993 Sep;113(9):605–26.

243. Sudhakaran VK, Deshpande BS, Ambedkar SS, Shewale JG. Molecular aspects of penicillin and cephalosporin acylases. Process Biochem. 1992 May;27(3):131–43.

244. Hare S, Vos AM, Clayton RF, Thuring JW, Cummings MD, Cherepanov P. Molecular mechanisms of retroviral integrase inhibition and the evolution of viral resistance. Proc Natl Acad Sci. 2010 Nov 16;107(46):20057–62.

245. Egan WJ, Lauri G. Prediction of intestinal permeability. Adv Drug Deliv Rev. 2002 Mar 31;54(3):273–89.

246. Ravishankar S, Kumar VP, Chandrakala B, Jha RK, Solapure SM, de Sousa SM. Scintillation proximity assay for inhibitors of escherichia coli murg and, optionally, mray. Antimicrob Agents Chemother. 2005 Apr;49(4):1410–8.

247. Hamilton-Miller JM. Antimicrobial properties of tea (Camellia sinensis L.). Antimicrob Agents Chemother. 1995 Nov;39(11):2375–7.

248. Perva-Uzunalić A, Škerget M, Knez Ž, Weinreich B, Otto F, Grüner S. Extraction of active ingredients from green tea (Camellia sinensis): Extraction efficiency of major catechins and caffeine. Food Chem. 2006 Jun;96(4):597–605.

249. Khan N, Afaq F, Saleem M, Ahmad N, Mukhtar H. Targeting multiple signaling pathways by green tea polyphenol (-)-epigallocatechin-3-gallate. Cancer Res. 2006 Mar 1;66(5):2500–5.

250. Higdon JV, Frei B. Tea catechins and polyphenols: health effects, metabolism, and antioxidant functions. Crit Rev Food Sci Nutr. 2003;43(1):89–143.

251. Shankar S, Ganapathy S, Hingorani SR, Srivastava RK. EGCG inhibits growth, invasion, angiogenesis and metastasis of pancreatic cancer. Front Biosci J Virtual Libr. 2008;13:440–52.

252. S MP, C P. Observations on the inhibition of HIV-1 reverse transcriptase by catechins. [Internet]. 1992 [cited 2013 Oct 1]. Available from: http://www.biochemj.org/bj/288/bj2880717.htm

253. Mereles D, Hunstein W. Epigallocatechin-3-gallate (EGCG) for clinical trials: more pitfalls than promises? Int J Mol Sci. 2011 Aug 31;12(12):5592–603.

254. Jiang F, Chen W, Yi K, Wu Z, Si Y, Han W, et al. The evaluation of catechins that contain a galloyl moiety as potential HIV-1 integrase inhibitors. Clin Immunol. 2010 Dec;137(3):347–56.

255. Yamaguchi K, Honda M, Ikigai H, Hara Y, Shimamura T. Inhibitory effects of (−)-epigallocatechin gallate on the life cycle of human immunodeficiency virus type 1 (HIV-1). Antiviral Res. 2002 Jan;53(1):19–34.

256. Ono K, Nakane H, Fukushima M, Chermann JC, Barré-Sinoussi F. Differential inhibitory effects of various flavonoids on the activities of reverse transcriptase and cellular DNA and RNA polymerases. Eur J Biochem FEBS. 1990 Jul 5;190(3):469–76.

257. Kawai K, Tsuno NH, Kitayama J, Okaji Y, Yazawa K, Asakage M, et al. Epigallocatechin gallate, the main component of tea polyphenol, binds to CD4 and interferes with gp120 binding. J Allergy Clin Immunol. 2003 Nov;112(5):951–7.

258. Poon G. Analysis of catechins in tea extracts by liquid chromatography–electrospray ionization mass spectrometry. J Chromatogr A. 1998 Jan 23;794(1–2):63–74.

259. Yang CS, Wang X, Lu G, Picinich SC. Cancer prevention by tea: animal studies, molecular mechanisms and human relevance. Nat Rev Cancer. 2009 Jun;9(6):429–39.

260. Cheng A, Merz KM. Prediction of aqueous solubility of a diverse set of compounds using quantitative structure−property relationships. J Med Chem. 2003 Aug 1;46(17):3572–80.

261. Ganesan A. The impact of natural products upon modern drug discovery. Curr Opin Chem Biol. 2008 Jun;12(3):306–17.

262. Owens J. Chris Lipinski discusses life and chemistry after the Rule of Five. Drug Discovery Today. 2003;8(1):12–6.

263. Tsiang M, Jones GS, Hung M, Samuel D, Novikov N, Mukund S, et al. Dithiothreitol causes hiv-1 integrase dimer dissociation while agents

interacting with the integrase dimer interface promote dimer formation. Biochemistry (mosc). 2011 Mar 15;50(10):1567–81.

264. Verduyn C, Van Kleef R, Frank J, Schreuder H, Van Dijken JP, Scheffers WA. Properties of the NAD(P)H-dependent xylose reductase from the xylose-fermenting yeast Pichia stipitis. Biochem J. 1985 Mar 15;226(3):669–77.

265. Long LH, Clement MV, Halliwell B. Artifacts in Cell Culture: Rapid generation of hydrogen peroxide on addition of (−)-epigallocatechin, (−)-epigallocatechin gallate, (+)-catechin, and quercetin to commonly used cell culture media. Biochem Biophys Res Commun. 2000 Jun 24;273(1):50–3.

266. Long LH, Kirkland D, Whitwell J, Halliwell B. Different cytotoxic and clastogenic effects of epigallocatechin gallate in various cell-culture media due to variable rates of its oxidation in the culture medium. Mutat Res Toxicol Environ Mutagen. 2007 Dec 1;634(1–2):177–83.

267. Chai PC, Long LH, Halliwell B. Contribution of hydrogen peroxide to the cytotoxicity of green tea and red wines. Biochem Biophys Res Commun. 2003 May 16;304(4):650–4.

268. Hou Z, Sang S, You H, Lee M-J, Hong J, Chin K-V, et al. Mechanism of action of (−)-epigallocatechin-3-gallate: auto-oxidation–dependent inactivation of epidermal growth factor receptor and direct effects on growth inhibition in human esophageal cancer KYSE 150 cells. Cancer Res. 2005 Sep 1;65(17):8049–56.

269. Zieger MAJ, Glofcheski DJ, Lepock JR, Kruuv J. Factors influencing survival of mammalian cells exposed to hypothermia: V. Effects of hepes, free radicals, and H2O2 under light and dark conditions. Cryobiology. 1991 Feb;28(1):8–17.

# APPENDIX A

University of the Witwatersrand, Johannesburg

Human Research Ethics Committee (Medical)
(formerly Committee for Research on Human Subjects (Medical))

Secretariat: Research Office, Room SH10005, 10th floor, Senate House • Telephone: +27 11 717-1234 • Fax: +27 11 339-5708
Private Bag 3, Wits 2050, South Africa

Ref: W-CJ-150213-2 15/02/2013

***TO WHOM IT MAY CONCERN:***

**Waiver:** This certifies that the following research does not require clearance from the Human Research Ethics Committee (Medical).

**Investigator:** **Miss S. Abrahams (student no 500916).**

**Project title:** **Evaluation of the NIH Clinical Collection for potential HIV-1 integrase inhibitors.**

**Reason:** This is a wholly laboratory study with use of the NIH Clinical Collection of compounds. No humans are involved.

UNIVERSITY OF THE WITWATERSRAND
PROF PE CLEATON - JONES
HREC (MEDICAL)
2013 -02- 15
JOHANNESBURG

Professor Peter Cleaton-Jones
Chair: Human Research Ethics Committee (Medical)

copy: Anisa Keshav / Zanele Ndlovu Research Office, Senate House, Wits

**Table I: Recapitulation of the data obtained from biological assays throughout the study for all the controls, identified HIT compounds and derivatives.**

| | SPA | | | | HIV-1 IN ELISA | | | MT4-cell based assays | | | RT inhibition colometric assay |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $IC_{50}$ $IN_{WT}$ | $IC_{50}$ Q148H/G140S | $FCIC_{50}$ | %IN ST inhibition at 10µM compound | %IN ST inhibition | %IN ST inhibition (10mM DTT) | %IN ST inhibition (10mM B-Me) | $CC_{50}$ (µM) | %Antiviral activity | $EC_{50}$ (µM) | %RT inhibition |
| **RAL** | 9.98 ± 0.83nM | 1.5 ± 0.50µM | 214 | 92 ± 5 | > 100 (at 10µM) | > 100 (at 10µM) | > 100 (at 10µM) | > 200 | >74 (at 100nM) | | |
| **EVG** | 4 ± 1.04nM | 2 ± 0.20µM | 200 | 88 ± 7.23 | | | | | | | |
| **DTG** | | 3 ± 1nM | | | | | | | | | |
| **118-D24** | 1.5 ± 0.81µM | | | 77 ± 3.70 | | | | | 97 ± 0.92 (at 100µM) | 9 ± 4 | |
| **AZT-TP** | | | | < 0 | | | | | | | > 100 (at 10µM) |
| **CEF** | 6.03 ± 1.29µM | 5.01 ± 0.30µM | 0.8 | 83 ± 1.20 | 21 ± 0.04 (at 100µM) | 53 ± 0.16 (at 100µM) | 16 ± 1.50 (at 100µM) | > 200 | 23 ± 0.50 (at 200µM) | | 14 ± 0.06 (at 100µM) |
| **EGCG** | 9.57 ± 1.62µM | 0.619 ± 0.50µM | 0.065 | 75 ± 4 | 29 ± 0.21 (at 100µM) | 66 ± 0.89 (at 100µM) | < 0 (at 100µM) | 23 ± 1 | > 100 (at 200µM) | 24 ± 3 | 56 ± 0.19 (at 100µM) |
| **7-ADCA** | | | | 2.45 ± 1.08 | | | | | | | |
| **7-ACA** | | | | 1.9 ± 1.46 | | | | | | | |
| **CSS** | | | | 5.28 ± 12.10 | | | | | | | |
| **AMP** | | | | 66.5 ± 1.44 | 53 ± 1.30(at 100µM) | | | > 200 | 51 ± 6.52 (at 200µM) | | 14 ± 0.09(at 100µM) |
| **Auranofin** | | | | | | | | 0.57 ± 0.16 | | | |