CHAPTER 5
COMPOUND EFFECTS ON VIRAL ENZYMES

SUMMARY

Background: A lot of success (in clinical use) has been achieved with inhibitors that target HIV enzymes, RT, PR and IN. To investigate the effect of the gold-based compounds on these viral targets, direct enzyme assays and computer aided *in silico* analysis were performed. In the direct enzyme RT and PR bioassays, the eleven new compounds (the three bimetallic phosphine thiolate complexes in class III and the thiosemicarbazone-based complexes and complementary ligands in class V) were tested while for the IN assay a preliminary study was performed for all twenty seven compounds. Compounds analysed in the *in silico* tests were mostly those which had demonstrated ≥50% inhibition in the direct enzyme inhibitory assays.

Materials and Methods: The direct enzyme assays for RT were performed using a colorimetric kit and recombinant RT enzyme by ELISA while the PR assay was done using an HIV protease substrate and recombinant PR enzyme in a fluorogenic substrate assay. For the IN assay, ELISAs were performed using two different kits; a dual kit consisting of both 3' processing and strand transfer components and a second strand transfer specific kit. *In silico* studies were performed using the CDOCKER protocol in Discovery Studio®.

Results and Discussion: None of the eleven compounds from class III and V inhibited RT while PFK7 inhibited PR by 55.5% at a concentration (toxic to cells) of 100 µM (p=0.03). In a preliminary screen four complexes (EK231, PFK7, PFK8 and PFK174) inhibited IN by ≥ 50% but not upon subsequent repeats. All twenty seven compounds were tested using the dual IN ELISA kit. For repeats, attempts were made to access additional kits but the manufacturer reported difficulties in developing kit components. The kit was again available several months later and the initial data was not reproducible. In the *in silico* docking studies favourable binding free energy predictions were obtained for five of the gold complexes in the RNase H site of RT, none for the PR site and five for the lens epithelium derived growth factor binding site of IN. Although favourable enthalpic contributions were noted for the RNase H site, size-shape complementarity and thus good binding affinity was lacking; the flatness of the RNase H binding pocket and the fact that the complexes lacked metal chelating groups (required by substrates binding to this site) may be responsible for this. PFK7 was predicted to interact more favourably with hotspot residues in the LEDGF binding site of IN and with better complementarity than the corresponding ligand or its analogues. Structure activity relationships were observed in the *in silico* studies for both the RNase H and the LEDGF site.

Conclusions: The binding affinities from the *in silico* predictions studies did not appear to represent stable interactions. This finding appeared to support findings from the direct

enzymes assays where inhibition was noted at toxic concentrations e.g. PFK7's inhibition of PR or the inconsistent results seen for IN. Although the RNase H site of RT and the LEDGF site of IN were favoured in the overall predictions, the favoured compounds would not in the current form inhibit or bind to these enzymes appreciably and will require structural optimisation through rational drug design to increase activity.

Keywords: enzyme inhibition, RT, PR, IN, direct enzyme bioassays, docking, mechanism of inhibition.

5.1 INTRODUCTION

The virally encoded RT, PR and IN are crucial enzymes in the life cycle of HIV and have been successfully targeted for ARV therapy. The combination of drugs that inhibit these enzymes especially RT and PR were the earliest to be used in HAART and lately IN inhibitors have been approved for clinical use and for inclusion in HAART or as salvage therapy for patients who have already developed resistance to existing combinations (McColl and Chen, 2010). The development of resistance by HIV to these drugs is the driving force for research towards the identification of novel inhibitors of these enzymes while efforts to identify new viral targets are also being pursued. A total of twelve RT inhibitors including eight NRTIs and four NNRTIs, have been approved for clinical use (de Bethune, 2010), nine PR inhibitors have been approved (Wensing *et al.*, 2010) and the most recent addition in terms of class being the IN inhibitor raltegravir which was approved by the US FDA in 2007 (McColl and Chen, 2010). HAART has led to significant declines in morbidity and mortality associated with HIV infection especially in countries where ARV medications are widely accessible (Bartlett *et al.*, 2007). Despite all the progress that has been made in terms of delaying disease progression, prolonging survival and improving the quality of life of patients (Antiretroviral Therapy Cohort collaboration, 2008), ARV therapy still fails to suppress HIV completely for both existing and even the newer classes of drugs (Marcelin *et al.*, 2009). The failure is mostly associated with the development of resistant viral strains leading to the accumulation of mutant forms (Ceccherini-Silberstein *et al.*, 2007, Cozzi-Lepri *et al.*, 2005, Clavel *et al.*, 2004, Hanna *et al.*, 2000). A troubling concern is the fact that there are few treatment options and strategies in the case of drug failure and/or cross resistance to the same class of compounds e.g. NNRTIs (Johnson *et al.*, 2005). Cross resistance has also been shown for the recently approved IN inhibitor raltegravir, and eltragravir (an IN inhibitor still in clinical trials, Marinello *et al.*, 2008). This rapid rate of drug resistance development by the virus and the associated cross resistance to drugs within the same class makes the quest for identifying novel therapy imperative while at the same time pursuing research on novel targets and vaccines. The toxicity associated with HAART and uncomfortable side effects (Yeni, 2006, Montessori *et al.*, 2004, Montaner *et al.*, 2003) are also some of the reasons driving the search for new and hopefully safer drugs.

In this study, enzyme assays were used in high throughput screening (96 well plate format) methods for identifying compounds that could directly inhibit RT, PR and IN experimentally in bioassays. However, because data from such experimental studies does not say much about the type of binding interactions made by the compound with the enzyme, computer aided studies were used to decipher this information (specifically molecular modelling or docking). Molecular modelling as defined by Richon, (1994), is the science or art of representing molecular structures numerically and simulating their behaviour with the equations of quantum and classical physics. It is important to note that both experimental and computational techniques have important roles in drug development and represent complementary approaches (Kapetanovic, 2008). The popularity of computer aided drug design goes beyond mechanistic exploration. It has been used in target identification and validation, in streamlining the drug discovery and development process, for optimisation of experimental findings as well as to eliminate compounds with undesirable characteristics using *in silico* filters (Kapetanovic, 2008, Tang *et al.*, 2006). Although recent trends in rational drug design studies begin with *in silico* analysis before synthesis and biological testing as proposed by Tarbit and Berman (1998), it should be noted that *in silico* studies are mainly predictions and must be validated experimentally. The approach in this study was synthesis based on the history of gold and complementary ligands as possible anti-HIV agents and on literature accounts on drug-likeness of the various ligands e.g. lipophilicity of the phosphine ligands (Shaw *et al.*, 1994) and the anti-viral activity of Tscs-base ligands (Easmon *et al.*, 1992, Spector and Jones, 1985). This was followed by HTS *in vitro* analysis and the *in silico* work was performed as a complementary approach to corroborate bioassay findings.

Gold compounds have previously been reported to inhibit HIV RT (Fonteh *et al.*, 2009, Fonteh and Meyer 2008, Sun *et al.*, 2004, Tepperman *et al.*, 1994, Okada *et al.*, 1993, Blough *et al.*, 1989) and to interact with proteins by undergoing ligand exchange reactions with sulfhydryl groups of cysteine residues (Shaw III, 1999, Sadler and Guo, 1998). The chrysotherapeutic effect of gold compounds was also reported in the inhibition of the lysosomal cysteine PRs (a family of proteases responsible for joint destruction in rheumatoid arthritis) through ligand exchange reactions with sulfhydryl group of cysteine (Gunatilleke *et al.*, 2008, Chircorian and Barrios 2004), making these compounds possible PR inhibitors. Although HIV PR is an aspartic PR, it was tempting to speculate that the presence of cysteine residues in the dimerisation interface of HIV PR (Zutshi and Chmielewski, 2000) could result in interactions with the gold ion. Inhibition of IN has not been reported for gold compounds but because the integration process involves viral cDNA, it is possible that gold complexes could intercalate with it since gold compounds have been reported to interact with DNA (Mirabelli *et al.*, 2002). This interaction could possibly result in the inhibition of IN activity.

Overall, the expectation for all three viral enzymes in addition to possible ligand exchange reactions and DNA intercalation was that since complexation has been reported to

lead to more stable compounds which stay at enzyme active sites longer and to increased drug efficacy (Navarro, 2009, Beraldo and Gambino, 2004), the gold complexes should be better inhibitors compared to the ligands.

In the next sections, findings obtained for the effect of the compounds on RT, PR and IN from both the direct enzyme and *in silico* binding predictions will be provided.

5.2 MATERIALS AND METHODS

Some of the compounds currently being investigated in this project have previously inhibited HIV RT and PR in direct enzyme assays (Fonteh *et al.*, 2009, Fonteh and Meyer 2009). The RT inhibitors were tested again as controls in the bioassays and *in silico* predictions of the binding modes with this enzyme were also done using molecular modelling. Compounds that inhibited PR in the direct enzyme bioassays were also analysed to predict potential binding modes using molecular modelling. The eleven new compounds that had not been tested before in both the RT and PR direct enzyme assays were also screened for the inhibition of these enzymes prior to molecular modelling of successful candidates. IN bioassays were initiated for all the compounds and repeated for compounds that showed promise in the pre-screen followed by molecular modelling.

5.2.1 Direct Enzyme-Based Assays

The direct enzymes assays included either sandwich ELISAs (RT and IN) or a fluorogenic substrate assay (PR). These are assays in which the compounds were allowed to interact with the enzymes in the presence of substrate and enzyme activity monitored in endpoint analysis either by determining absorbance or fluorescence.

5.2.1.1 RT inhibition assay

The Reverse Transcriptase Colorimetric kit (Roche Diagnostics, Mannheim, Germany) was used (assay principle pictured in Figure 5.1). This assay gives a quantitative measure of the RT activity and takes advantage of the ability of RT to synthesise DNA, starting from the template/primer hybrid poly (A) x oligo $(dT)_{15}$. Digoxigenin and biotin labelled nucleotides are incorporated into the same DNA molecule as it is freshly synthesized by RT. The detection and quantification of newly synthesized DNA as a parameter for RT activity follows a sandwich ELISA protocol where biotin-labelled DNA binds to the surface of streptavidin-coated microplate modules. An antibody to digoxigenin, conjugated to peroxidase (anti-DIG- POD) is then added and binds to the digoxigenin-labelled nucleotides. Finally, a peroxidase substrate, 2, 2'-azino-di-[3-ethylbenzthiazoline sulfonate (6)] diammonium salt crystals (ABTS), is added. The peroxidase enzyme catalyzes the cleavage of the substrate to produce a coloured reaction product. The absorbance of the samples is determined using a microplate (ELISA) reader, and is directly correlated to the level of RT activity in the sample.
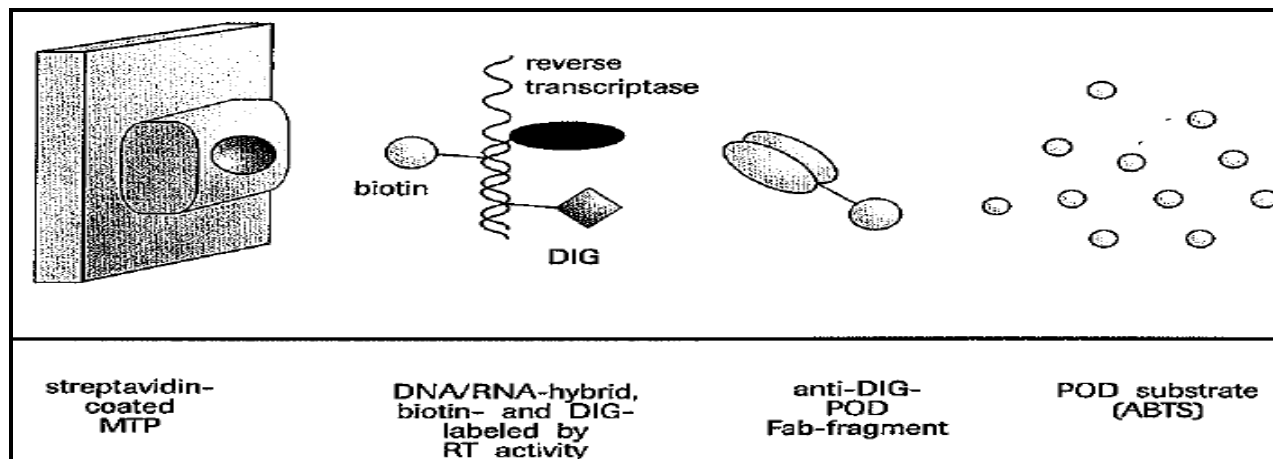
Figure 5.1: Reverse transcriptase colorimetric test principle. The test follows a sandwich protocol. The figure was adapted from the Roche Diagnostics colorimetric reverse transcriptase assay brochure (Version 13, 2010).

Procedure: The assay was performed according to the manufacturer's instructions and as previously described (Fonteh and Meyer, 2008). Briefly, 20 µL (0.2U) of purified recombinant HIV RT (Merck, Darmstadt-Germany) and 20 µL of reaction mixture consisting of reconstituted template-template/primer hybrid poly (A).Oligo (dT)$_{15}$ and diluted nucleotide (Tris-HCl, 50 mM, pH 7.8, with DIG-dUTP, biotin-dUTP and dTTP) were transferred to microfuge tubes containing 20 µL of pre-determined concentrations of the compounds dissolved in DMSO and diluted with lysis buffer. This was followed by 1 h of incubation at 37 ºC. The samples were then transferred to appropriate wells of a streptavidin-coated plate followed by another hour of incubation at 37 ºC. The plate was washed 5 times with 250 µL of wash buffer and blotted on paper towels to completely remove buffer before adding 200 µL of anti-DIG-POD working solution (200 mU/mL). A further 1 h incubation at 37 ºC followed by 5 rinses using wash buffer was performed. An ABTS substrate solution (200 µL) was transferred into all wells of the plate and the plate(s) incubated at room temperature (15-25 ºC) until sufficient green colour development for photometric detection was attained (approximately 15 min). Controls included RT only with an equivalent amount of DMSO used in the test samples (1.5%, v/v) while a positive control was a plant extract (designated known inhibitor or KI) for which anti-HIV data exists. The plate was read on a Multiskan Ascent® plate reader (Labsystems, Helsinki, Finland) at 405 nm and a reference wavelength of 492 nm. Data analysis was performed using Microsoft® Office Excel® 2007(Microsoft Corporation, Washington, USA) with inhibition expressed as percentages and calculated based on the formula: 100 – [(Test reagent absorbance-Blank absorbance/ untreated control absorbance-blank absorbance) x100)].

5.2.1.2 PR inhibition assay

This assay makes use of a fluorogenic HIV PR substrate 1 with structure: Arg-glu-(EDANS)-Ser-Gln-Asn-Tyr-Pro-Ile-Val-Gln-Lys-(DABCYL)-Arg (Sigma Aldrich, Missouri USA). This substrate is a synthetic peptide that contains a cleavage site (Tyr-Pro) for HIV PR as well as two covalently modified amino acids for the detection of cleavage (Matayoshi *et al.*, 1990).

One of the modifications involves the attachment of the fluorophore 5-(2-aminoethylamino)-1-naphthalene sulfonate (EDANS) to the glutamic residue. The other modification is the addition of an acceptor chromophore 4'-dimethylaminoazobenzene-4-caboxylate (DABCYL) to the lysine residue. The modified amino acids are on opposite sides of the cleavage site. Spatial orientation and overlap of the DABCYL absorbance with the EDANS emission permits resonance energy transfer between the two moieties and quenching of the EDANS fluorescence at 490 nm occurs. However, when HIV PR cleaves the peptide, the DABCYL group is no longer proximal to the fluorophore and emission at 490 nm cannot be detected. A compound that inhibits HIV PR therefore prevents this cleavage thus allowing quenching to occur such that the EDANS fluorescence signal is diminished.

Procedure: The assay was performed according to procedures by Lam *et al.*, (2000) using a 1 mM stock of HIV PR substrate 1 dissolved in DMSO and diluted to 20 µM with assay buffer (0.1M sodium acetate, 1 M NaCl, 1 mM EDTA, 1mM DTT and 1 mg/mL BSA, pH 4.7). An aliquot of the substrate (20 µM, 49 µL) and 1 µL of HIV PR solution (1 µg/mL; Bachem, Switzerland) were added directly into Costar® black 96 well fluorescence assay plates (Corning Incorporated, New York, USA) in the presence or absence (untreated control) of the compounds to a final reaction volume of 100 µL. This mixture was incubated at 37 ⁰C for 1 h. Ten microlitres of acetyl pepstatin designated AP (Bachem BioScience Inc. PA, USA) at a concentration of 10 µg/mL was used as a positive control for inhibition of HIV PR while a blank treatment consisted of assay buffer and the substrate only. The fluorescence intensity was measured at an excitation wavelength of 355 nm and an emission wavelength of 460 nm using a Fluoroskan Ascent® plate reader (Labsystems, Helsinki, Finland). The data was analysed using Microsoft® Office Excel® 2007 (Microsoft Corporation, Washington, USA) and the percentage inhibition calculated based on the formula: 100 - [(Test reagent RFU-blank RFU / untreated control RFU-blank RFU) x100)] where RFU = relative fluorescence units.

5.2.1.3 IN inhibition assay

Two different direct enzyme assay kits were used for determining the effect of the compounds on HIV IN and thus on the integration process. The Xpress Bio HIV IN kit (Thurmont, Maryland, USA) which contains unprocessed viral DNA (due to the presence of dinucleotides at the 3' ends) such that both the 3'P and ST reactions can be executed (referred to as dual IN kit) and the Auro pure kit (Mintek, Johannesburg, South Africa) which was specifically designed for detecting ST IN inhibitors. Both assays mimic the integration process except for the fact that the target DNA in the Auro pure kit is modified to exclude the 3'P reaction (contains processed viral DNA with the dinucleotides at the 3' ends excluded) making it specific for determining ST IN inhibitors. The motivation for specifically targeting ST inhibition is because IN drugs currently in clinical use are the ST inhibitors unlike 3'P inhibitors (3'PIs) which have shown activity *in vitro* but not *in vivo* (Mouscadet *et al.*, 2010).

Procedure: The dual IN inhibitor kit (Thurmont, Maryland, USA) assay was performed according to the manufacturers' instructions. Streptavidin-coated 96 well plates were further coated with a double stranded HIV LTR U5 donor substrate oligonucleotide or donor DNA containing an end-labelled biotin for 1 h. This was followed by 3 washes and then by blocking with blocking buffer for 1 h. After 3 additional wash steps, full-length recombinant IN protein (200 nm, purified from bacteria) was loaded onto the oligo substrate and the plate incubated for 30 min at 37 ºC. Non-toxic concentrations of the compounds were added in triplicate to the plate after 3 further washes and the plate was incubated for 5 min at room temperature. A different double-stranded target substrate oligo containing 3'-end modifications was added directly to the plate containing the compounds. The IN enzyme cleaves the terminal two bases from the exposed 3'-end of the HIV LTR donor substrate (3'P) and then catalyzes a ST reaction to integrate the donor substrate into the target substrate. The products of the reactions were detected colorimetrically using a horse radish peroxidase (HRP)-labelled antibody directed against the target substrate 3'-end modification and a tetramethylbenzendine (TMB) peroxidase substrate. A blank treatment without enzyme or test compounds and a control containing 10% (v/v) sodium azide (NaA$_3$) as positive inhibitor of IN activity were included in the analysis. The plate was read at 450 nm using a Multiskan Ascent® plate reader (Labsystems, Helsinki, Finland).

The Auro Pure kit (Mintek, Johannesburg, South Africa) assay was performed according to the manufacturer's specifications at the AuTEK Biomed Laboratory (Mintek, Johannesburg, South Africa). Briefly, biotin labelled donor DNA or donor substrate (processed) was added to streptavidin coated microwell strips and incubated for 1 h at 22 ºC. The plates were washed 3 times with wash buffer followed by the addition of bacterially expressed recombinant HIV IN enzyme (1 µM). After 30 min of incubating at 22 ºC and 2 wash steps, the test compounds and controls were added to the plate and allowed to interact for a further 30 min. The target DNA or target substrate was then added to the mix and after another hour of incubation at 37 ºC, the plates were washed 3 times and a detection antibody added to the wells and incubated (2 h, 25 ºC). Three last washes were performed and a substrate reagent added into each well. The plate was sealed and incubated at 37 ºC for 1 h followed by absorbance measurements at 620 nm on an xMark^{TM} Microplate Absorbance Spectrophotometer (Bio-Rad Laboratories Inc., California, USA). Inhibition percentage calculations for both IN assays were done using Microsoft® Office Excel® 2007(Microsoft Corporation, Washington, USA) and the formula: 100-[(Test absorbance-blank absorbance/control abs-blank-absorbance) x100].

5.2.2. Molecular Modelling to Predict Potential Binding Sites

In this section, the term receptor will be used to represent the target enzyme also known as "receiving" molecule or protein i.e. RT, PR or IN. Ligand (a computational

terminology) will be used to refer to both the gold complexes and precursors (free or uncomplexed compounds) and is defined as the complementary partner molecule, which binds to the receptor.

To perform a direct or protein-based docking study where the active site is known, unlike indirect or ligand-based docking where the active site is not known (Vaidyanathan *et al.*, 2009), one of the first requirements is to obtain 3D crystal structures of the receptor which must have been solved by x-ray crystallography or NMR (Raha *et al.*, 2007). The next important requirement is a database of the ligands to serve as inputs in the docking program. The crystal structures of RT, PR and IN were obtained from the protein data bank (http://pubchem.ncbi.nlm.nih.gov/) and the ligands were compounds, which had inhibited these enzymes in direct enzyme assays. After docking (also known as searching), scoring functions are usually employed to help in predicting the binding affinity of the ligands to the receptor (Wu *et al.*, 2003).

Molecular modelling studies were done using Discovery Studio® version 2.5.5 (Accelrys®, California, USA). All simulations were performed on an Intel® Core™ 2 Duo CPU 2.2 GHz processor, 1.97 GB RAM with Windows XP professional version 2002 operating system. These specifications were the minimum recommended for DS® installation and performance. Higher specifications should lead to lower computational costs and thus faster run rates. The next subsections will outline the steps that were performed in order to simulate the interactions of the ligands with the receptors and to determine binding affinity.

5.2.2.1 Ligand preparation

Ligand preparation involved obtaining molecular structures, which had to be in the structural data file (.sdf) or molecular file (molfile) chemical format and these were obtained using the ChemDraw software (CambridgeSoft, PerkinElmer Inc., USA). Considering that molecular modelling greatly depends on molecular properties, for modelling to be useful, it must readily and reliably reproduce properties that resemble those of experimentally obtained data (Comba and Hambley, 1995). One of the commonly applied models for determining molecular properties is molecular mechanics (MM), which calculates the structure and strain (deformation of a molecule resulting from stresses) of a molecule based on known reference structures and properties to give it a geometry with minimum strain energy. An example of such a model is the Chemistry at Harvard Macromolecular mechanics or CHARMm force field.

Unlike in the ADMET prediction studies where the only required preparative phase was the application of CHARMm force fields and minimization of the ligands, for the docking studies, a further ligand preparative step was involved. This is because the CHARMm force field in the docking algorithm that was used in Discovery Studio® does not have force fields assigned for gold or other transition metals. This complication arises from the fact that these transition metal ions have partially filled d-orbitals (Comba and Hambley, 1995, Hay, 1993).

These partially filled orbitals result in the diverse structures of coordination compounds with a large variety of possible coordination numbers and geometries (Comba and Hambley, 1995, Hay, 1993). Developing reference structural and strain energy values for metal complexes is a daunting task (Comba *et al.*, 2006) and is therefore not feasible. Instead, a combined quantum mechanics and molecular mechanics (QUANTUMm also designated QM/MM) computation was performed to calculate all-atom force fields for each ligand. In this approach, some atoms were treated by classical MM (CHARMm) and others treated by QM (DMol3) thereby combining the advantages of the two approaches. QM is significant in describing chemical interactions that involved the breaking and formation of covalent bonds and unlike MM, does not assume that the nature of the bonding does not change with the structure and is thus applicable to metal complexes (Höltje *et al.*, 2003, Comba and Hambley, 1995). For the calculation, the gold atom was ionised by deleting the covalent bonds, then assigning formal charges, followed by constraining bond angle distances thus simulating a covalent bond while calculating the force fields. CHARMm normally omits from the nonbonded lists any interactions that include only fixed atoms and would therefore not proceed with docking if these unknown atoms were not fixed through the addition of constraints, restraints and QM/MM minimisation. The QM/MM minimisation also gives the ligands better geometries and removes steric overlap that could produce bad contacts before the dynamics (heating and cooling) process that is involved in docking. For all QM/MM calculations the parameters sets included a QM/MM boundary set to a nonbond list radius of 10.0-12.0 Å, where the electronic embedding method was to neglect boundary charges. The type of DFT exchange-correlation potential chosen was the gradient-corrected (PBE) potentials. The atomic spin state were treated such that the spin was restricted if the number of electrons in the system was even. A medium quality of the Dmol3 calculation was set. After minimisation, the constraints and restraints on the gold and bonded atoms e.g. P-Au-Cl were removed, the bonds and the formal charges restored. Although constraints followed by QM/MM calculations could have been applied to gold and bonded atoms for the compounds in the ADMET studies, it was not necessary since CHARMm and Dmol3 force fields are not included in the ADMET prediction protocol.

5.2.2.2 Receptor preparation

Published crystallographic structures of the respective receptors (HIV RT, PR and IN) in complex with active site identification ligands were obtained from the PDB. The receptors were either wild type or mutant forms. Where possible, subtype C crystal structures were used to increase the relevance and specificity of potential inhibitor(s) to the South African/Sub Saharan African context where subtype C is prevalent. Crystal structures of subtype C receptors were unfortunately not readily available just as was the case for the recombinant enzymes used in the direct enzyme assays since research has focused largely on the subtype B viral strain. In the case of RT, It was important to determine whether the compounds

inhibited the enzyme by binding to the RNase H site or the NNRTI site. Binding at the NRTI site was not sought since these compounds are not dNTP analogues (e.g. zidovudine and ddI, Figure 2.8). Therefore, only crystal structures of receptors complexed to known inhibitors of the RNase H and NNRTIs sites were obtained. Docking was also done on a third site of RT recently shown by Su *et al.*, (2010) to be an allosteric inhibitory site close to the NNRTI pocket. The PDB identification codes for the RT receptors are 3LP2 (Su *et al.*, 2010) for the NNRTI allosteric site and 3LP3 (Su *et al.*, 2010) in complex with naphthyridinone-containing RNase H inhibitor site while the 2WON PDB structure (Corbau *et al.*, 2010) was used for predicting binding to the NNRTI site. The RT crystal structures were of wild type strains. In the case of HIV PR, two crystal structures were used; 1HXW in complex with ritonavir (Kempf *et al.*, 1995) coding for a subtype B strain and 2R5P (Coman *et al.*, 2008) coding for a subtype C viral strain. The crystal structures employed for predicting binding interactions with IN were the 2B4J structure which is the crystal structure of the CCD of IN complexed to LEDGF/p75 (Cherepenov *et al.*, 2005) at the dimer interface of the enzyme. This site has been used by Christ *et al.*, (2010) to determine small molecule inhibitors of protein-protein (IN-LEDGF) interactions. Other sites that were used for predicting IN inhibition were the ISQ4 site complexed to the 5-CITEP inhibitor found in the Asp64, Asp116 and Glu152 motif of IN (Goldgur *et al.*, 1999) and the binding site for sucrose identified in 3L3V (Wielens *et al.*, 2010) both in the CCD of IN. A summary of all the receptors used is provided in Table 5.1. The PDB identification codes, resolution and information on whether the structure is a mutant or wild type are provided. The smaller the resolution, the better the crystal structure is. Resolutions of 2.8 Å and below were considered sufficient for this study.

Table 5.1: A summary of the protein data bank crystal structures used for molecular modelling. Three crystal structures were used for RT, one with an allosteric binding site, another with an RNase H and a third with a polymerase or NNRTI binding site, two for PR consisting of a clade B and a C variant, and three for the IN consisting of an allosteric binding site in complex with sucrose, one in the LEDGF binding site at the dimer interface of IN and a third in the DNA binding site (3'P and ST site).

Protein	PDB code	Notes	Wild type/mutant	Resolution (Å)
HIV RT	3LP2	Allosteric site near NNRTI site	Wild type	2.8
	3LP3	RNase H site	Wild type	2.8
	2WON	NNRTI site	Wild type	2.8
HIV PR	1HXW	Subtype B	Wild type	1.8
	2R5P	Subtype C	Mutant (Q7K, L33I, L63I)	2.3
HIV IN	3L3V	CCD in complex with sucrose (allosteric site to LEDGF site)	Mutant (C56S, W131D, W139D, F185H)	2
	2B4J	CCD bound to LEDGF (at dimer interface)	Mutant (F185K)	2.02
	ISQ4	CCD (DNA binding site)	Mutant F185K, W131E	2.1

Binding site sphere definition: Before the docking simulations could be initiated, a protein clean process was performed and a binding site sphere defined on the receptor. The protein clean process involves the addition of hydrogen atoms to the amino acid residues of the

receptor and the removal of unnecessary groups e.g. water molecules. Decisions on maintaining groups such as cofactors, active site crystal water molecules, and catalytic metal ions had to be made. Ultimately the importance of these molecules or ions in the activity of the enzyme had to be considered and usually such groups were maintained. Docking with active site waters for example has been shown to result in increased accuracy of docking (Höltje *et al.*, 2003). A ribbon structure of the receptor was then displayed followed by the application of a CHARMm force field and partial charges (charges on polar molecules due to differences in electronegativity), Momany and Rone (1992). The co-crystallised ligand or active site identification ligand was used in defining the binding site sphere which is an area around the bound inhibitor in the three axis direction (xyz) with a minimum radius of 5 Å (see Figure 5.2 for a typical binding site sphere). Once the active site sphere had been defined, the co-crystallised ligand within the sphere was removed by highlighting and deleting using keyboard commands.
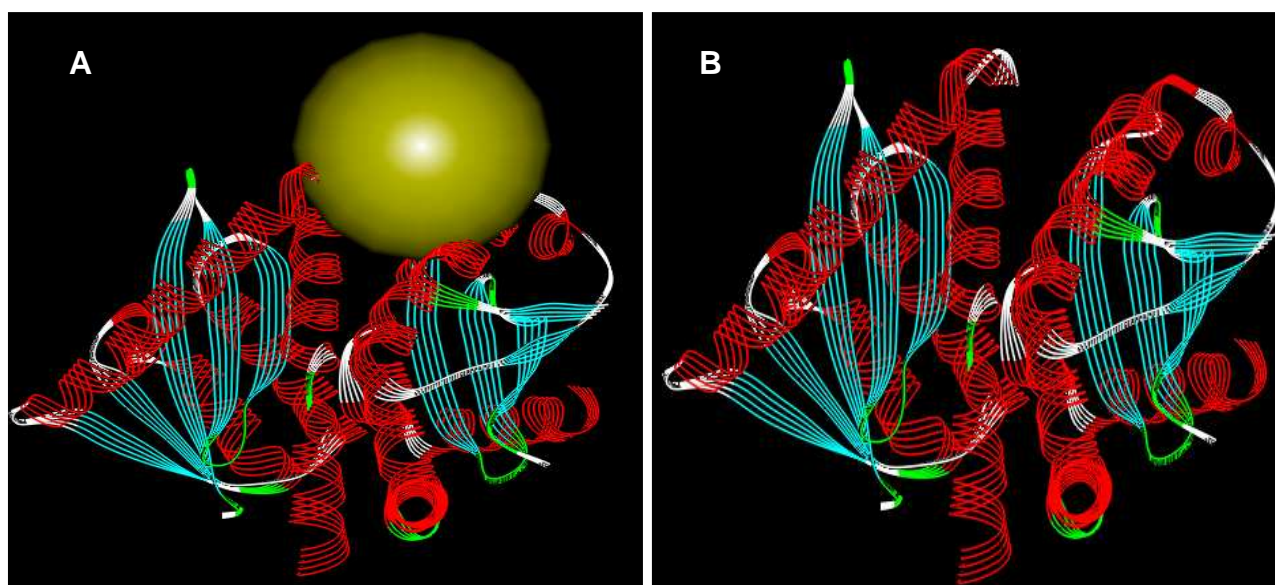
Figure 5.2: Binding site sphere (yellow ball in A) in the catalytic core domain of IN (2B4J). In (B) sphere has been removed. The sphere radius is 11 and the xyz coordinates with respect to the active site are 12.08, -18.587 and -11.632. This sphere was generated within one of the catalytic dimer interfaces specifically in the IN-LEDGF binding site (also shown in figure 2.17). Here the IN binding domain (depicted in Figure 2.17) of LEDGF has been removed and replaced with the binding site sphere in preparation for docking. This figure was obtained from Discovery Studio (Accelrys®, California, USA).

5.2.2.3 Docking with CDOCKER

The next step after ligand and receptor preparation was for docking to be initiated. This was done using CDOCKER (CHARMm-based DOCKER, Wu *et al.*, 2003), a grid-based molecular docking algorithm that employs CHARMm molecular dynamics in DS$^®$ (Accelrys®, California, USA). It allows one to run refinement docking of any number of ligands with a single protein receptor. The receptor is held rigid while the minimized ligands are allowed to flex during the refinement. Ligand placement in the active site was specified using the binding site sphere and the grid method helps to reduce computation time while facilitating molecular dynamics interactions between receptor and ligand atoms. The calculated QM/MM force fields

for the ligands (with the gold atom accounted for) were used in the docking and not the default CHARMm-based force fields. Random ligand conformations (poses) are generated from the initial ligand structure through high temperature molecular dynamics, followed by random rotations. The random conformations are refined by grid-based simulated annealing which involves intermittent cooling by decreasing simulation temperature. The final energy minimization during docking was set as "off" and a separate minimization step (below) was performed instead.

5.2.2.4 Ligand minimization

A separate minimisation step was launched which minimizes the series of ligand poses (a pose being a unique target bound orientation and conformation of the ligand after docking) using CHARMm and helps to filter the poses and ensure diversity while improving on the docking accuracy (Krovat *et al.*, 2005, Wu *et al.*, 2003). Generally, minimization (performed before and after docking) reduces the energy of a structure through geometry optimization. The minimization of the poses was done in the presence of the receptor. In the protocol, the receptor is held rigid and residues with atoms inside the minimization sphere of flexible atoms are allowed to be move. This strategy helps minimise computational costs (time) involved when the entire receptor is flexible (Krovat *et al.*, 2005). The protocol used was customized to include an implicit solvent model (generalized born with molecular volume) which allows for the calculation of the binding free energy between the receptor and ligand while mimicking solvent effect (Mohan *et al.*, 2005). This is because protein surface interactions occur in aqueous solution making the description of solvation forces and energies due to solute-solvent effect critical in modelling (Sun and Latour, 2006).

5.2.2.5 Energy calculations and analysis of minimized poses (Scoring)

Docking is usually performed together with scoring functions to predict the binding affinity of the ligands (Wu *et al.*, 2003). This was done by calculating the energy of binding represented as binding free energy and given in kcal/mol followed by analysis of the docked poses. The "calculate binding energy" protocol was used and the calculated QM/MM force fields employed for the ligands. The protocol estimates binding free energy between each ligand pose and the receptor using CHARMm implicit solvation models. The free energy of binding for a receptor-ligand complex is calculated from subtracting the energy of the ligand and that of the receptor from the energy of the complex and is given by the formula:

Energy of binding = energy of complex - energy of ligand - energy of receptor.

Energy calculations were done for each pose, and the binding free energies ranked from the lowest to highest with the former representing the most favourable receptor-ligand interaction in terms of binding affinity (Muegge and Rarey, 2001).

Further scoring analysis was done using, the "analyse ligand poses" protocol in DS®. It helps in determining heat maps, short distances between atoms, electrostatic interactions and generally aids in further interpreting the binding modes between the ligand and the receptor as realistic or not. The pose predicted to have the most favourable binding free energy with the receptor was then mapped using 2D drawings, molecular surface diagrams (depicting hydrophobicity) followed by molecular graphics generation.

Docking controls: To ensure that the ligand orientations and positions predicted by the docking studies were likely to represent valid and reasonable potential binding modes, the active site identification ligands (co-crystallised ligand) were docked using the customised CDOCKER docking parameters and the prepared sphere selections for each of the binding sites. The interactions were considered acceptable when the predicted binding orientations were comparable to the expected orientation and position of the inhibitor observed in the crystal structures.

5.2.2.6 Summary of methods used

A flow diagram of the methods that were used in determining the interactions of the compounds with HIV enzymes is depicted in Figure 5.3. Both "wet lab" and *in silico* (docking) methods were followed.
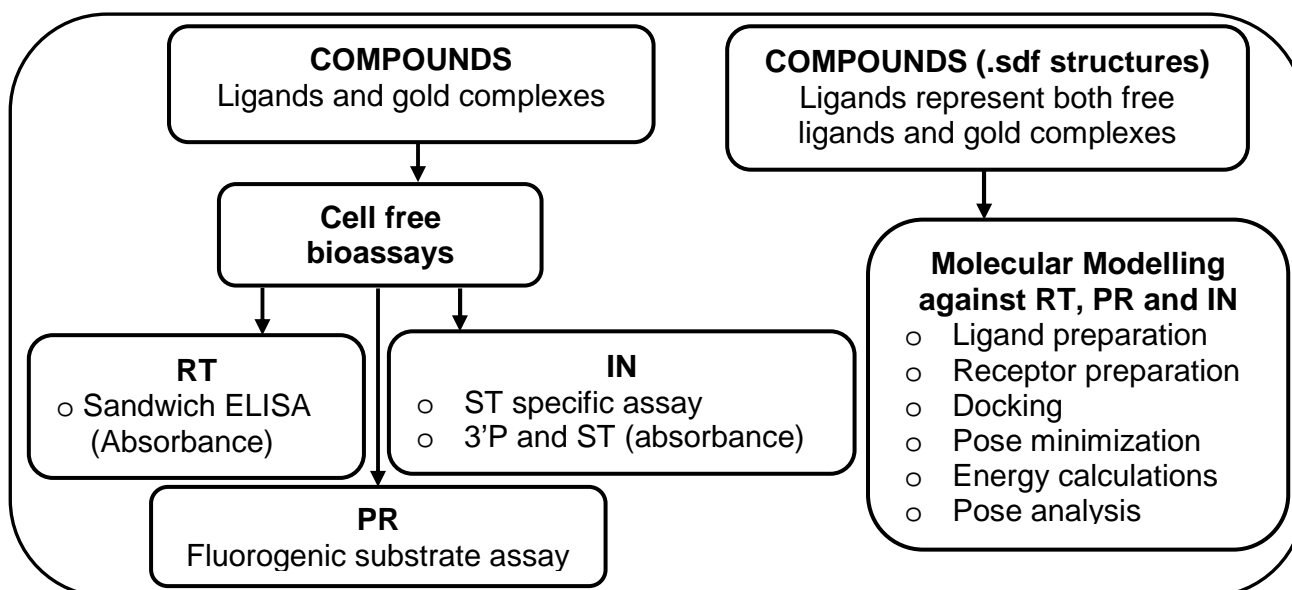

Figure 5.3: Summary of methods used in determining the effect of the compounds on viral enzymes.

5.3 RESULTS AND DISCUSSION

5.3.1 Direct Enzyme Assays

5.3.1.1 HIV RT and PR activity

Compounds tested for RT and PR inhibition included the Tscs-based compounds; PFK5, PFK7, PFK6, PFK8, PFK39, PFK41, PFK38 and PFK43 and the gold(I) phosphine thiolate complexes; PFK174, PFK189 and PFK190. No inhibition of RT was recorded at 25 and 100 μM (only 25 μM is shown, Table 5.2) and only one compound significantly inhibited

HIV PR activity by >50% (Table 5.2). This was the gold(III) thiosemicarbazonate complex, PFK7, which inhibited the enzyme by 55% at 100 µM (p = 0.03). Its free ligand PFK5 had no effect on PR's activity. Unfortunately 100 µM (which is the same concentration at which compounds tested in the prior study inhibited PR, Fonteh and Meyer, 2009) is toxic to all the cell types for which cytotoxicity measurements were performed suggesting poor specificity. PFK7 is cytostatic (at 5 and 10 µM, Figure 4.7B) and was not really expected to have a direct anti-viral effect (for either RT or PR) since cytostatic agents target cellular components (Lori *et al.*, 2005). While the higher concentration of PFK7 appears to have a direct anti-viral effect, this ability is not of any value since this concentration was also very toxic (the CC_{50} of this compound was 5.6 in PBMCs and 1.7 in PM1 cells, Table 4.2). However the compound can interfere with viral replication through its effect on the host cell at non-toxic but cytostatic concentrations (Figure 4.6, 4.7B, 4.8A).

Table 5.2: The effect of the compounds on HIV RT and PR activity. None of the compounds inhibited RT's activity while PFK7 inhibited PR at 100 µM by 55.5%. KI represents a known inhibitor which inhibited RT by 94.7% and PR by 100 %. PFK7 inhibited PR by 55% at 100 µM (p=0.03). Ligands are shaded in grey.

HIV RT Inhibition		HIV PR Inhibition	
Compound	% inhibition at 25 µM	Concentration (µM)	% inhibition
KI	94.7	16	100.2
$HAuCl_4.4H_2O$	-14.0		30.43
PFK189	23.7		35, 49
PFK190	10.5		34, 48.7
PFK5	9.55	2.5 , 100	14, 15.4,
PFK7	14.7	2.5, 25, 100	-5, 5.8, **55.2**
PFK6	9.36	5, 100	-8.8, 0.14
PFK8	5.35	5 , 25, 100	1.3, 5.3, 16.7
PFK39	7.86		ND
PFK41	6.39	0.2 , 100	-12., 12.9
PFK38	-8.68		ND
PFK43	11.26	0.04, 100	-9, 2.8

Eight gold complexes (TTC3, TTC10, TTC17, TTC24, EK207, EK219, EK231 and KFK154b) previously shown to inhibit RT at a concentration range of 6.25 to 250 µM (Fonteh and Meyer, 2009, Fonteh *et al.*, 2009, Fonteh and Meyer, 2008) were re-tested as controls in the present study. None of the compounds inhibited RT's activity at 25 and 100 µM (only 25 µM is shown, Table A5.1). These unexpected results triggered a search for reasons and explanations. The most serious of these was the possibility that degradation products not visible on NMR spectra (when stability studies were performed, chapter 2 in section 3.4.1) were present in addition to storage, compound age and dissolution related concerns. All the compounds previously published as active were freshly synthesised and tested in the same manner. These compounds were now re-tested after three years of storage at -20 ºC in powder form and sometimes dissolved in DMSO (used within a week, for KFK154b, storage

was at room temperature for a year and a half prior to testing). These storage conditions were obviously not optimal and the inherent hygroscopic abilities observed for some of these complexes (e.g. TTC3 and possibly its analogues, Table 3.7) could have potentiated solubility issues. Newly synthesised compounds (eleven in total) did not have similar storage issues and may simply have no RT/PR inhibitory abilities. When taking into account docking predictions which resulted in limited active site binding (full discussion in section 5.3.2) as well as the low ADMET solubility predictions for some of these previously active compounds, the absence of activity is somewhat confirmed. Possible additional reasons for the loss of RT activity are provided in the appendix (section 8.4)

5.3.1.2 HIV IN activity

Inhibition of HIV IN activity was performed using two different kits, one involving both 3'P and ST inhibitor steps and the other specific for ST inhibitors. In a preliminary assay using the dual IN kit, all 27 compounds including the gold starting material ($HAuCl_4.4H_2O$) were tested in triplicate at one concentration (Figure A5.1A). Four compounds inhibited HIV IN by >50% at non-toxic concentrations when this assay was done and included the BPH gold(I) complex, EK231 (50.8%), the gold(III) thiosemicarbazonate complexes PFK7 (54.5%) and PFK8 (58.6%) and the gold(I) phosphine thiolate complex PFK174 (78%). Ligands PFK5 and PFK6 inhibited the enzyme in this assay by 47.5 and 47% respectively suggesting that the ligands contributed in the inhibition. New kits were purchased to perform repeat experiments and the problems mentioned in the abstract were experienced. In the repeat experiments, percentage inhibitions were mostly negative values at three different concentrations (2.5, 5, 10, and 25 µM) and were all rounded off to 0% (Figure A5.1B). The assay was performed 3 times with similar results each time. The positive control inhibited the enzyme by 99.9±0.3%. The reason for this change in results between kits is not clear and might have been poor performance of the kits. For both the initial and subsequent kits, the positive control worked very well. The company from which the kit was obtained had manufacturing problems for some time and we had to wait for up to 8 months to be able to repeat the assay. Communication with the manufacturers did not lead to clarification of the discrepancies and the assumption was that the manufacturing problems might have been the cause. While these concerns are valid, one could postulate that some of the same reasons that were attributed to the compounds that inhibited RT previously and which subsequently lost the inhibitory ability may be applicable here (particularly poor aqueous solubility, aging, the presence of degradation products not detectable by NMR and solvent effects, see section 8.4.1). The first concern (poor aqueous solubility) does not apply to the gold(III) bisthiosmicarbazonate complexes since these compounds had good prediction values in the ADMET study and in the experimental shake flask method but are applicable to complexes EK231 and PFK174 (Table 3.8A, section 3.4.3). The eleven additional compounds had only been stored for a year and a

half (in the course of analysis). While solvent effects could have played a part, it is important to note that cytostasis was not affected. The same sample of PFK7 which inhibited IN in the pre-screen was also cytostatic at the time and remained cytostatic when additional repeats were recently (close to the time of writing) performed.

In the ST specific assay, representative compounds (consisting of those that had initially inhibited IN in the dual inhibitor assay pre-screen) from each class i.e. TTL24, TTC24, EK231, MCZS1, KFK154b, PFK5, PFK7, PFK8, HAuCl$_4$.4H$_2$O$_4$ and a positive control for IN inhibition (NaA$_3$) were tested (Figure 5.4). None of the compounds inhibited the ST activity of HIV IN including those that had inhibited IN by > 50% in the dual IN inhibitor assay pre-screen. NaA$_3$ (6%) inhibited the enzyme by 77.5%. Although both 3'P and ST inhibitors have been reported in *in vitro* tests, only the latter have been successful *in vivo* and have subsequently been approved for clinical use (Mouscadet *et al.*, 2010, Chirch *et al.*, 2009) making their identification important. The fact that none of the compounds that had inhibited IN in the dual assay did so in the ST specific study suggests that inhibition of the enzyme in the former could have been at the 3'P step. Considering that subsequent testing in the dual assay resulted in no inhibition, it is likely that the compounds were neither 3'P nor ST inhibitors. Failure of 3'P inhibitors to inhibit *in vivo* was the reason why the inconsistent data was not investigated further. The ST specific data was therefore considered more significant and the conclusion was that no IN inhibition was exhibited by the compounds.
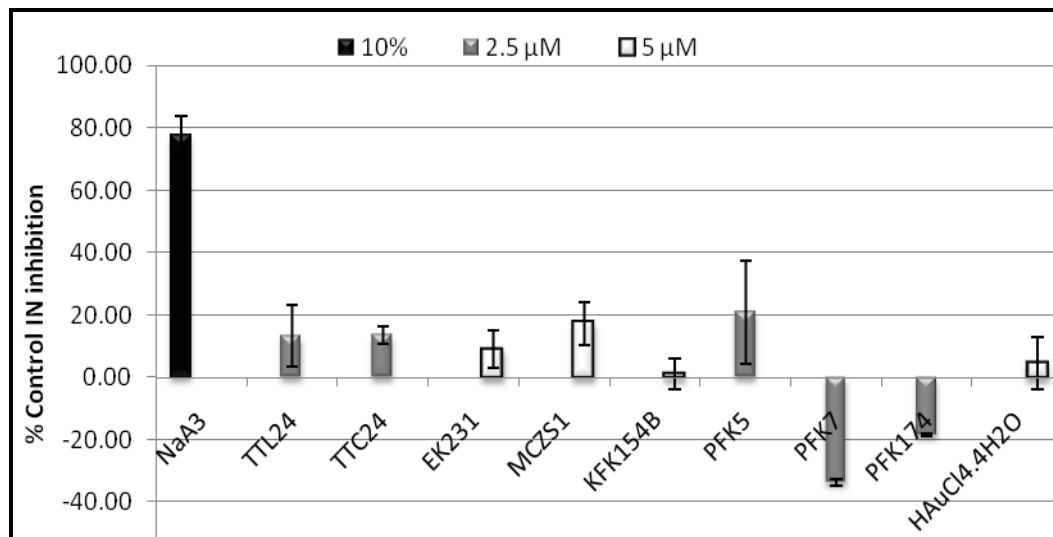
Figure 5.4: HIV IN inhibitory activity of representative compounds from different classes. The effect of the compounds on the enzyme was determined using the ST specific IN assay kit. None of the compounds significantly inhibited the enzyme (all p values were > 0.05). NaA$_3$ was used as a positive control for IN inhibition and inhibited the enzyme 77.5%. Concentrations tested here are those which resulted in >60% viability in the viability assays.

5.3.2. Molecular Modelling for Predicting Binding Interactions with Enzyme Active Site.

The selection criterion for the ligands (gold complexes and free ligands) chosen for molecular modelling studies was the fact that the latter had inhibited the enzyme in the respective direct enzyme bioassays. In the case of RT and PR these were compounds reported by Fonteh *et al.*, (2009), Fonteh and Meyer (2009) and herein (Table 5.2) as having

inhibitory activity while for IN, those which inhibited in the preliminary dual inhibitor assay (Figure A5.1) were tested. The respective ligands for each receptor are shown in Table 5.3. The phosphine gold(I) complexes; TTC3, TTC10, TTC17, TTC24 and the BPH gold(I) complexes EK219 and EK231 inhibited RT while one phosphine gold(I) complex TTC24, the BPH gold(I) complex, EK208, two phosphine gold(I) thiolate-based complexes (MCZS1 and MCZS3) and a gold(III) thiosemicarbazonate complex (PFK7) inhibited HIV PR. Four Tscs-based ligands (PFK5, PFK7, PFK6, PFK8) and a phosphine gold(I) thiolate complex PFK174 were also tested based on preliminary findings which showed that these ligands could inhibit IN by at least 48% when the xPressbio dual inhibition assay kit (Thurmont, MD, USA) was used (data is in the appendix Figure A5.1). Although subsequent screening with the same dual assay kit resulted in 0% inhibition (Figure A5.1B), it was important that these findings be further confirmed by an alternative method such as molecular modelling. This was also the rationale for docking the compounds that had previously shown RT inhibitory ability but which had later lost this property upon re-test (section 8.4.1, Table A5.1). This is because for compounds with favourable interactions in the *in silico* study, structural modifications can be recommended to enhance drug-likeness through rational drug design. In addition the *in silico* studies obviate the need for investing in new synthesis while at the same time determining if there was a correlation between experimental data and docking findings.

Table 5.3: Summary of compounds that inhibited HIV RT, PR and IN in direct enzyme bioassays. Compounds which inhibited previously (Fonteh and Meyer 2009, Fonteh *et al.*, 2009) and those for which inhibition was observed here (grey) are represented. Except for PFK5 and PFK6 with inhibitions of 45≥50%, the rest of the complexes inhibited by >50%.

RT inhibitory ability		PR inhibitory ability		IN inhibitory ability	
TTC3	EK231	TTC24	MCZS3	EK231	PFK7
TTC10	KFK154b	EK208	PFK7	PFK174	PFK6*
TTC17	EK207	MCZS1	KFK154b	PFK5*	PFK8
TTC24	EK219				

The binding of a ligand to a receptor depends on ionic interactions, hydrogen bonds, hydrophobic interactions, van der Waals and dipole interactions that can be established between the two (Sahu *et al.*, 2008). Other defining conditions for the interaction of a ligand with a receptor are its three dimensional characteristics which include size, stereochemical orientation of functional groups as well as physical and electrochemical properties. After the docking process, various numbers of unique conformations were generated for each ligand and this occupancy rate indicates ligand flexibility (Purohit *et al.*, 2008). In addition to giving some perception of the flexibility of the different ligands in each receptor site, differences in the variation between classes of ligands for a particular receptor site compared to the variation within a class could be observed. The variation within classes was usually less compared to that between classes probably because of the structure similarity of the latter. The more flexible a ligand is, the greater the ensemble of ligand-receptor conformations that it can have and the more likely it is to have good overall size-shape complementary and thus more

favourable enthalpic contributions to the binding free energy. Such inhibitors have a higher chance of remaining active in the event of a mutation (although with lower efficacy) since an alternative binding conformation is easily attainable. Flexibility is also linked to the number of rotatable bonds that a ligand has. A summary of all the docked poses for each of the receptor sites for successfully docked compounds is shown in the appendix (Table A5.2).

Further investigation of the binding affinity of the ligands to the receptors was based on the binding free energy. Generally the more negative the binding energy, the greater the affinity between an inhibitor and its receptor (Sadiq *et al*., 2010).

Not all compounds for which docking was initiated was successful. In some cases, either the binding free energies predicted were too high, depicting unfavourable interactions while in others, refined poses were not possible probably because of very poor stereochemical orientation or size-shape complementarity. While such a problem could be addressed by scaling the receptor site to accommodate the ligand, it is usually advisable to maintain the default parameters (Friesner *et al*., 2004), which was the case in this study. A summary of the most favourable binding free energies of the successfully docked ligands in the various receptor active sites are shown in Table 5.4. Binding free energies below 100 kcal/mol are shown for the gold complexes with the exception of gold complex TTC3 and free ligands TL17 and TTL24 (Table 5.4) whose energies are provided for comparison purposes. Although amino acid atoms within 4 Å of the ligand were identified as those interacting with the latter, bond distances of relevant dipole interactions beyond 4 Å (particularly those of cation-pi and pi-pi interactions) were also identified and are represented in Table 5.4. A majority of the binding free energy values shown in Table 5.4 were not negative suggesting that the reactions were not spontaneous or self driven. The ranking and further discussions are meant to serve as aids in describing the affinity of the compounds with the various receptors with the intention that these might serve as guides for optimising SAR for these compounds.

A table representing the twenty amino acids is shown in the appendix to aid in their identification. The three and one letter identification codes as well as classification based on hydrophobicity (non polar), polarity, polar acidic and polar basic properties are represented (Table A5.3). In the following subsections the interactions and binding affinities of the ligands with the various active site amino acid residues will be elaborated on.

5.3.2.1 Binding modes between ligands and RT sites

Docking was done for three different sites of RT namely the RNase H site (3LP3), a second site close to the NNRTI site (3LP2) and the NNRTI site (2WON). Very unfavourable binding energies were predicted for the NNRTI site suggesting poor binding affinity and for this reason no further analysis were done. This finding was not surprising considering that traditional structure-based design of NNRTIs has generally been complicated by the fact that

RT has considerable conformational flexibility (Hsiou *et al.*, 1996, Kroeger *et al.*, 1995, Jäger *et al.*, 1994).

Table 5.4: Summary of predicted binding free energy values and relevant bond distances obtained after molecular modelling. The lowest binding energies for the respective ligands are shown and represent the most favourable binding poses for each receptor site. The most favourable binding predictions for RT were with the RNase H site (3LP3) while those for IN were with the LEDGF binding site (2B4J). The free ligands are colour coded in a darker grey. The superscript ([a]) represents a pi-pi stacking interaction.

	HIV RT						
	3LP2			**3LP3+** Mn^{2+}		**3LP3-** Mn^{2+}	
Compound	**Energy**	**H-bond**	**Cation-pi**	**Energy**	**Cation/pi-pi**	**Energy**	**Cation-pi**
TTC3	243			18.7		-9.4	
TTL10	84.2			76.6			
TTC10	87.7		3.5, 3.8	10.9	6	0.21	2.8
TTL17	299						
TTC17	85.9			25.7		5.1	
TTL24	182			78.6			
TTC24	57.2	2.3, 1.9	3.8, 5.4	12.3	4.6[a]	0.52	
KFK154B	86.3			10.4			
	HIV IN						
	3L3V			**2B4J**			
	Energy		**H-bonds**	**Energy**			**H-bonds**
PFK5	40.1		2.9,1.7, 2.4, 2.2	8.9			2.2, 2.5
PFK7	42.3		2.4, 2	13.2			
PFK8	42.1			15.7			
PFK41	40.4			18.2			2.4, 1.7
PFK174				7.2			

Figure 5.5 represents annotated structures of ligands TTC3 and TTC24 which will subsequently be important in describing binding predictions with the RNase H and the 3LP2 receptor sites for which much more favourable binding free energies were obtained. The phenethyl amine portion of TTC3 also present in TTC17 (a), the *N,N*-dimethyl-ethane-1,2-diamine group of TTC24 also present in TTC10 and (b) the diphenylphosphanyl-benzyl portion (c) common to all the ligands in this group are shown as inserts.
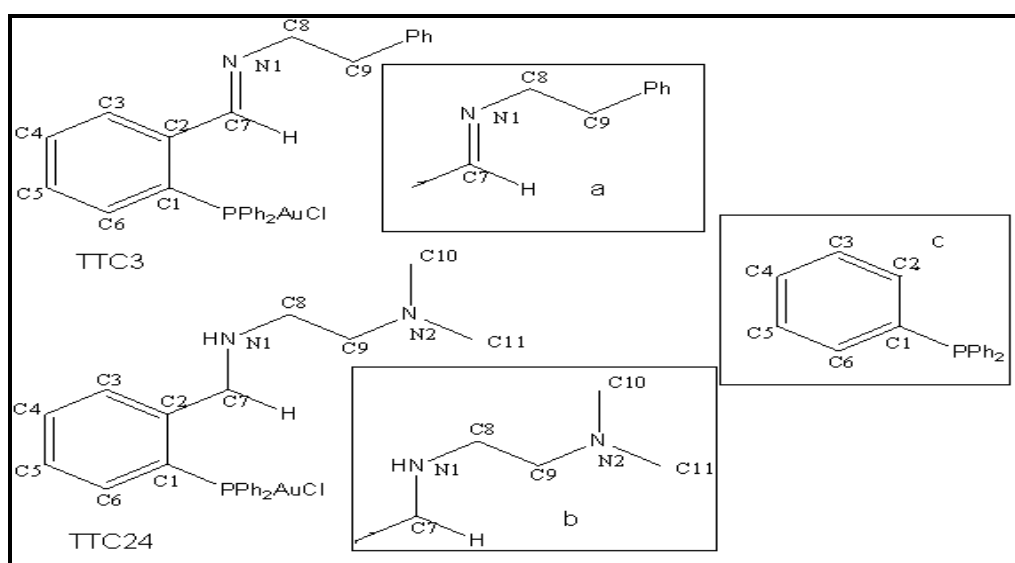

Figure 5.5: Annotated structures of TTC3 and TTC24 and important groups. The figure also shows the phenethyl-amine group of TTC3 (a), the *N,N*-dimethyl-ethane-1,2-diamine group of TTC24 (b) and the diphenylphosphanyl-benzyl portion present in all the ligands and complexes in this class as inserts.

Predicted binding interactions with the RNase H site: The RNase H site contains two metal ions (Mn^{2+}) that are ligated to active site carboxylate residues of Asp443, Glu478, Asp498 and Asp549 which are both needed in binding the substrates of this site and catalyzing the phosphodiester bond hydrolysis (Jochmans, 2008). The crystal structure used for this study (3LP3) was recently derived by Su *et al.*, (2010) into which the authors predicted the binding interactions of the metal binding naphthyridinone compounds (compounds that contain groups that can bind to the active site metal ions) designated MK1, MK2 and MK3. Docking in this site resulted in lower binding free energies compared to the 3LP2 site (Table 5.4). The lowest binding free energy of 10.4 kcal/mol was noted for the gold(III) pyrazolyl complex (KFK154b). This was followed by the gold(I) phosphine chloride complexes in the order TTC10<TTC24<TTC3<TTC17 with binding free energies of 10.9, 12.3, 17.8, and 25.7 respectively (Table 5.4). The order here did not correspond with that noted for the bioassays, where TTC24 was the most potent inhibitor followed by TTC10 and then TTC17 and TTC3 (Table 5.3). However, preference for the *N,N*-dimethyl-ethane-1,2-diamine group present in TTC10 and TTC24 over the phenethyl-amine group in TTC17 and TTC3 was observed.

Predicted interactions with the RNase H site in the presence of Mn^{2+} **ions:** In Figure 5.6 A and B, the interactions of TTC10 and TTC24 with the RNase H site in the presence of the Mn^{2+} ions are shown. The receptor in Figure 5.6 and those in the rest of this report are coloured according to the Kyte and Doolittle (1982) hydrophobicity profile.

A cation-pi (6 Å distance) interaction was predicted between the phosphate group of the 2-diphenylphosphanyl-benzyl portion of TTC10 and the imidazole ring of His539 (Figure 5.6A). Hydrophobic interactions were also predicted in the binding between the ligand and Ala538. The significance of the interaction with Ala538 is that it could confer specificity in the binding affinity of the compound since in human RNase H1 this group is replaced by Gly538 (Su *et al.*, 2010). In addition, Ala538 has been reported to interact with Asp549 which forms a critical hydrogen bond with water facilitating RNase H function (Di Grandi *et al.*, 2010). The predicted cation-pi interactions with His539 may therefore play a significant role in inhibiting RNase H activity. A molecular surface diagram of the receptors' hydrophobicity depicting the binding interactions of TTC10 with this site is shown in Figure A5.2, which unfortunately did not show great size-shape complementarity. Unlike the ribbon structure of the receptor shown in Figure 5.6 where it is easier to visualise H-bonds and other interactions (in addition to the hydrophobic interactions) unlike in a molecular surface map, which presents a different view of the overall hydrophobicity and hydrophilicity of the ligand-complex.
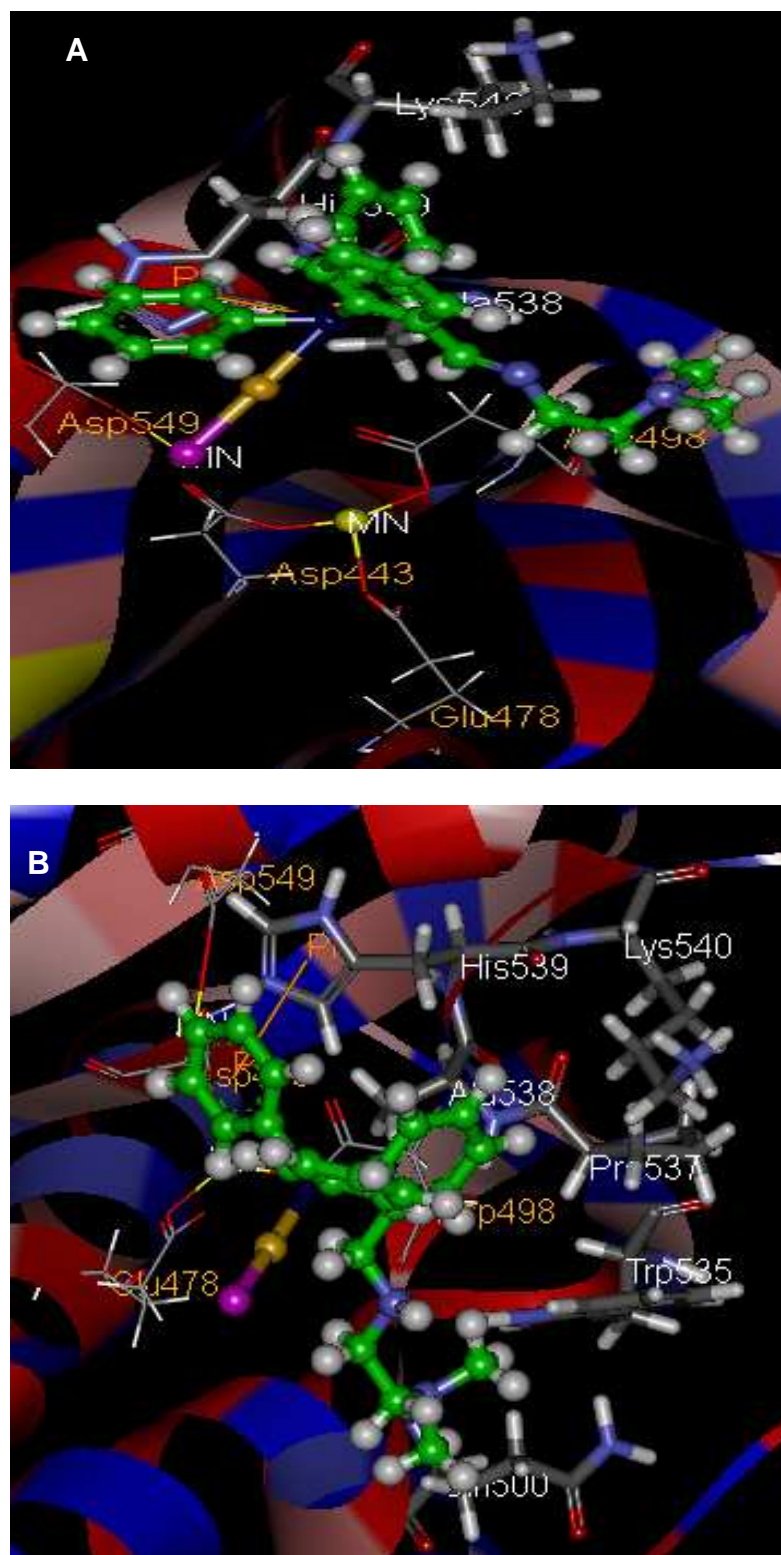
Figure 5.6: Predicted binding predictions of TTC10 and TTC24 (green ball and stick models) to the RNase H site in the presence of Mn^{2+} (yellow balls). Prominent amino acids are shown as stick models. Both compounds did not interact with the Mn^{2+} ions present in this site but formed crucial interactions with His539. A cation-pi interaction was formed between the phosphate group of TTC10 and the imidazole ring of His539 (A). Hydrophobic interactions were also observed between Ala538 and the side chain (CH2)3 group of Lys540. (B) TTC24 on the other hand formed a crucial pi-pi stacking interaction between one of its phenyl groups and the imidazole ring of His539. Hydrophobic interactions were also observed between Ala538, Pro537, Trp535 and the side chain group of Lys540 (CH2)3 and Gln500 (CH2)2. Active site carboxylate residues of Asp443, Glu478, Asp498 and Asp549 which are coordinated to the Mn^{2+} ions are represented with line diagrams and labelled in orange while the active site residues with which the ligands were predicted to interact are labelled white and represented as stick diagrams. Red balls or sticks = O, blue = N, white = H, orange=Au, purple = Cl, grey = C on receptor and green on ligand, P = dark blue.

Predicted interactions of TTC24 with this site (Figure 5.6B) included a pi-pi stacking interaction (4.6Å) between one of the phenyl rings attached to the phosphanyl-benzyl moiety (insert C, Figure 5.5) of the compound with the imidazole ring of His539 and hydrophobic interactions with Pro537, Trp535 and side chain residues of Gln500. In the same manner like TTC10, TTC24 was also predicted to interact with Lys540 through hydrophobic interactions with side chain $(CH_2)_3$ groups. It appears the *N,N*-dimethyl-ethane-1,2-diamine group present in TTC10 and TTC24 confers better interactions between these ligands and the RNase H site than the phenethyl-amine moiety of TTC3 and TTC17. These findings are obviously related to structure and suggest a SAR. In terms of H-bond donors, TTC10, TTC17 and TTC24 each have one while TTC3 has none and in terms of H-bond acceptors, TTC10 and TTC24 each have two while TTC3 and TTC17 each have one. The H-bond donors and acceptors are contributed by either the *N,N*-dimethyl-ethane-1,2-diamine of TTC10 and TTC24 or from the phenethyl-amine moiety of TTC3 and TTC17. In the experimental data reported by Fonteh and Meyer (2009), a similar trend with respect to inhibition of RT was observed. TTC24 which has the highest number of H-bond acceptors and donors (total of 3, Table 3.6) inhibited RT by > 74 % at 6.25 μM, TTC10 with two H-bond acceptors by > 34 % at 6.25 μM while complexes TTC3 and TTC17 resulted in <2% inhibition at 6.25 μM.

In the validation docking with the naphthyridinone-containing compound, MK3, a crucial pi-pi (5.5 Å) stacking interaction between one of its phenyl groups and the imidazole ring of His539 in addition to three H-bonds contacts were also predicted in conformity with Su *et al*'s findings. The interactions between His539 therefore appear to be very important in the binding of these ligands with the RNase H site.

Predicted docking interactions for the free ligands of TTC10 and TTC24 i.e. TTL10 and TTL24 respectively (which were tested as controls to determine the significance of complexation) resulted in binding free energies, which were seven fold higher than for the corresponding complexes. The predicted binding free energies of TTL10 and TTL24 were 76.6 and 78.6 kcal/mol compared to 10.9 and 12.3 kcal/mol for TTC10 and TTC24 respectively (Table 5.4). In the bioassays (early screening done in 2007 and reported in Fonteh and Meyer 2009), these ligands had no inhibitory activity and it is therefore not surprising that while the complexes were predicted to interact more favourably with the RNase H site of RT, the free ligands did not. These differences suggests that the gold complexes are more stable in binding to the RNase H site than the free ligands, a finding which supports the concept of increased stability and activity when organic compounds are complexed with metals (Navarro, 2009).

The binding affinity predictions for the gold(III) pyrazolyl compound KFK154b which had also previously inhibited HIV RT in cell free assays (Fonteh *et al*., 2009) resulted in energetically feasible binding predictions with the RNase H site (binding free energy of 10. 4 kcal/mol). Different portions of the compound i.e. the tetra-chloro gold portion, the bis-(3,5-

dimethylpyrazolyl)methane portion and the Cl$^-$ ion interacted with different sites of the receptor. The interaction of the Cl$^-$ and the tetra-chloro gold portion were outside the sphere that was defined as the active site (Figure A5.3). No conclusions could be made on the binding interactions of this compound with the RNase H site as a result.

Although favourable binding predictions were observed for the gold(I) phosphine chloride complexes (TTC10, and TTC24) in the RNase H site, there was unfortunately poor size-shape complementarity. The addition of chemical substituents that have metal binding groups (e.g. carbonyl groups) could aid in sandwiching these ligands better in the active site pocket while increasing the binding affinity and thus efficacy. A poor fit to the RNase H site has also been attributed to the flatness or absence of a deep pocket which makes it difficult for the development of RNase H inhibitors (Himmel *et al.*, 2009, Davies *et al.*, 1991). Another limitation to the interaction of the complexes with this site could have stemmed from the fact that metal-based docking parameters have not been incorporated into DS$^®$ and other docking software. As a result, expected covalent interactions that could have occurred between gold and the receptor (particularly with sulfhydryl groups of cysteine residues) were not possible. Unfortunately, this could not be assessed since docking algorithms were designed for organic molecules which form non covalent bonds such as hydrogen bonding and van der Waal forces unlike metallodrugs which form covalent bonds and ionic forces (Navarro, 2009).

Predicted binding interactions with the RNase H site in the absence of Mn^{2+}: Docking to the 3LP3 site was also performed in the absence of Mn^{2+} ions for TTC3, TTC10, TTC17 and TTC24 in an attempt to determine the influence of these metal ions on the binding. The predictions suggested better binding affinities than in the presence of Mn^{2+} in the order of TTC3<TTC10<TTC24<TTC17 with binding free energies of -9.4, 0.2, 0.5 and 5.2 kcal.mol respectively (Table 5.4). Surprisingly TTC3 had the highest affinity for this site compared to TTC10 or TTC24 which were the most favoured both in the bioassays and for this same site in the presence of Mn^{2+} (Table 5.4). It appears that there were repulsive forces between the ligands and the receptor when docking was done in the presence of Mn^{2+} (Figure 5.6) possibly because the ligands do not have metal binding groups that could interact with the active site Mn^{2+} ions. In the absence of the metal ions, there was better size-shape complementarity (Figure 5.7) with the ligands interacting with active site residues that were otherwise not within 4 Å of the ligand when docking was done in the presence of Mn^{2+} (Figure 5.6).
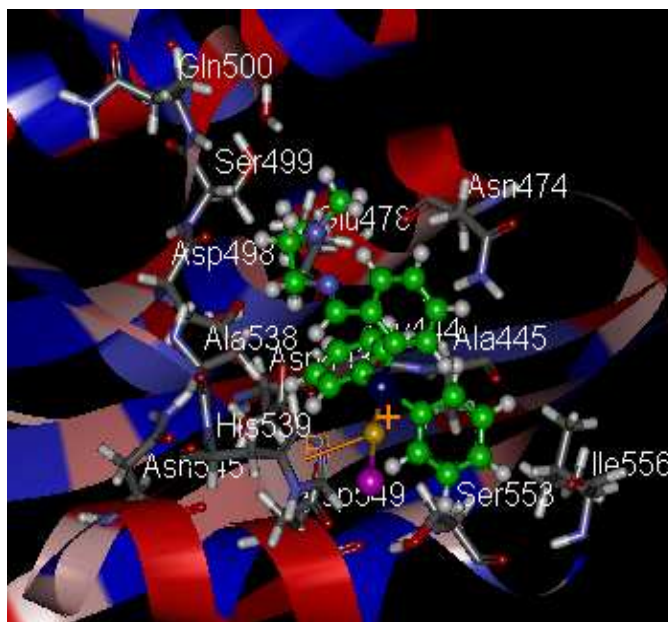
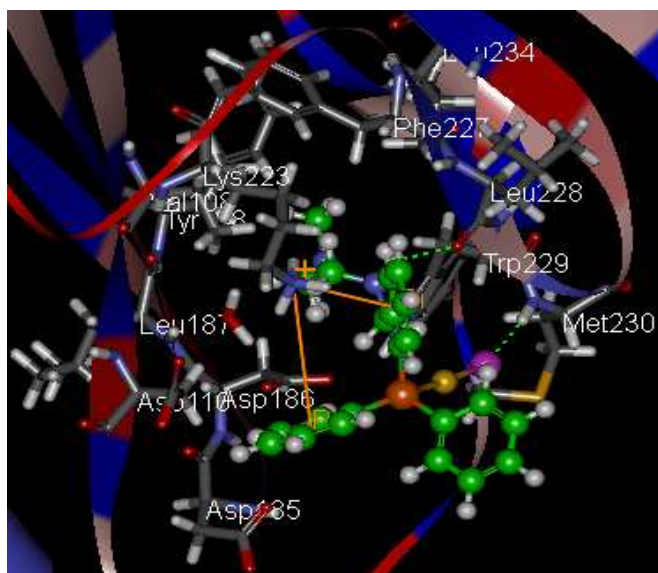**Figure 5.7: Predicted binding interactions of TTC10 to the RNase H site in the absence of Mn^{2+}.** The ligand fits more snugly into the active site making contact with many more residues including those normally bonded to Mn^{2+} (Asp443, Glu478, Asp498 and Asp549) it also interacted with hydrophophic residues Ala445, Ala538, Ile556 and polar residues Gly444, Asn474, Ser 499, Gln500, Asn545, Ser 553 and two water molecules. Red balls or sticks=O, blue=N, white =H, orange=Au, purple=Cl, grey=C on receptor and green ball on ligand, P=dark blue

Although favourable interactions were observed in the absence of Mn^{2+}, the two metal ions are important in the activity of RNase H and both are needed in binding the substrates of this site and catalyzing phosphodiester bond hydrolysis (Jochmans, 2008). The data however corroborates the fact that repulsive forces were present when docking with Mn^{2+} was done because the ligands lacked metal binding groups. In addition, the observed interactions of the ligands with the RNase H site in the absence of Mn^{2+} could aid in the design of compounds that would interact more favourably with this site in the presence of Mn^{2+} by including groups which should hopefully interact with the additional amino acid contacts present when docking was done in the absence of Mn^{2+}. Compound TTC10 for example (Figure 5.7) was predicted to form a cation-pi interaction (2.8 Å) between gold and the imidazole ring of His539 and in addition to interacting with the amino acids normally coordinated to Mn^{2+} (i.e. Asp443, Glu478, Asp498 and Asp549), also made hydrophobic contacts with residues Ala445, Ala538, Ile556 and polar residues Gly444, Asn474, Ser 499, Gln500, Asn545, Ser 553 and two water molecules (all residues within 4 Å of the ligand). These interactions were absent when docking was done in the presence of Mn^{2+} (Figure 5.6). The lower binding energy could therefore be attributed to a better fit which also correlated with a decrease in the cation-pi distance from 6 Å in the presence of the Mn^{2+} ions to 2.8 Å in their absence. The data suggests that the binding of these ligands to this site did not require coordination to the metal ion when comparing the binding free energies predicted in the absence or presence of the metal ions (Table 5.4). In the former case the probable reason lack metal chelating moieties such as carbonyl groups in the ligands. Unfortunately these metal ions are required for the activity of the enzyme making

their chelation necessary for the inhibition of RNase H (Kirschberg *et al.*, 2009). A modification of the compounds to contain metal chelating groups so as to enhance the activity of the gold-containing ligands might prove promising for improving the binding predictions and possibly reproducibility and efficacy in bioassays.

Predicted interactions with the NNRTI allosteric site (3LP2): Eight ligands were successfully docked into the 3LP2 site. These were the phosphine gold(I) complexes (TTC3, TTC10, TTC17 and TTC24, free ligands of TTL10, TTL17, TTL24) and the gold(III) pyrazolyl complex, KFK154b (Table 5.4). The binding free energies predicted were not in the single digit or negative range but some very interesting interactions that could be optimised were observed. TTC24 will be used as a model (Figure 5.8) for elaborating the binding predictions that were seen for the phosphine gold(I) complexes since its interactions resulted in the lowest binding free energy of 57 kcal/mol (Table 5.4). The predicted interactions of the compounds with this site showed SAR that correlated with the biological data (Fonteh and Meyer 2009). TTC24 inhibited RT the most (prior to loss of activity, Table 5.2) in the direct enzyme assays with a >50% inhibition at 6.25 µM (Table 5.3).

Figure 5.8: Predicted binding interactions of TTC24 with the site close to the polymerase/NNRTI site (3LP2). The interactions of TTC24 with this site showed the *N,N*-dimethyl-ethane-1,2-diamine group being inserted in a hydrophobic pocket and two cation-pi interactions occurring between two of the phenyl groups of TTC24 and Lys223. Two H-bonds were seen between one of the amine groups of the ligand and Leu228 (2.3 Å) and the other between the Cl$^-$ ion and the backbone NH$_3^+$ group of Met230 (1.9 Å). One of the phenyl groups not involved in cation-pi interactions is solvent exposed. Red balls or sticks=O, blue=N, white =H, orange=Au, purple=Cl, grey=C on receptor and green balls on ligand, P=dark blue. H-bonds are shown as green dotted lines. Cation-pi interactions=orange lines

The *N,N*-dimethyl-ethane-1,2-diamine portion of TTC24 (Figure 5.5) led the inhibitors into a predominantly hydrophobic pocket lined by numerous residues which include Val108, Asp110, Asp185, Asp186, Leu187, Tyr188, Lys223, Phe227, Leu228, Trp229, Met230, Leu234 and a water molecule (Figure 5.8). In addition to two cation-pi interactions (3.8 and 5.4 Å respectively) predicted between the NH$_3^+$ group of Lys223 and two of the phenyl groups (attached to C1 of the diphenylphosphanyl benzyl portion, Figure 5.5) of TTC24, two H-bonds

interactions were also predicted between the H-bond donor (N1, Figure 5.5) and Leu228 (2.3 Å) and the other between the Cl$^-$ ion and the backbone NH$_3^+$ group of Met230 (1.9 Å). Two cation-pi interactions were predicted between two of the phenyl groups of TTC10 and Lys223 (3.5 and 3.8 Å respectively). TTC10 was also predicted to interact with similar residues lining the binding pocket as seen for TTC24 (with the exception of Tyr181, Gln182, Phe226, Gly231, Gln242). The absence of the two H-bond interactions between TTC10 and the receptor is probably responsible for the higher binding free energy (87.7 kcal/mol) making TTC24 (57.kcal/mol) a better inhibitor. Although cation-pi interactions are known to be strong non-covalent bonds (Dougherty, 1996) which could have easily increased the binding affinity of TTC24 to this site, one of the two phenyl rings attached to phosphorous (at the PPh2 position, Figure 5.5) not involved in cation-pi interactions appeared to be predominantly solvent exposed and could be the contributing factor for the high binding free energy. This solvent exposed phenyl group possibly contributed unfavourably to the overall entropy of binding and hence reduced the stability of binding. The interaction of TTC17 with this site is similar to that of TTC10 but better than that of TTC3 which had the highest binding free energy prediction of 243 kcal/mol (Table 5.4). The only difference between compounds TTC17 and TTC3 is the presence of an H-bond donor in TTC17.

The overall binding interactions of the compounds with this site presented similar orientations as reported by Su *et al.*, (2010) for the diethylaminophenoxy group of the naphthyridinone-containing inhibitors. Although the interactions were not spontaneous nor in the single digit range in terms of energy rankings, these findings suggests that modifying the dimethyl-ethane-1,2-diamine portion of TTC10 and TTC24 could result in better binding predictions for these ligands. In addition, the solvent exposed phenyl ring might need to be replaced with a smaller group to reduce solvent effects. These findings also provide useful information that can aid in the design of new gold-based complexes targeting this site and should be of interest to medicinal chemists involved in rational drug design.

It has not been determined whether the 3LP2 site is a biologically relevant inhibitory site but the binding of one of the naphthyridinone compounds that was probed by Su *et al.*, (2010) was able to sufficiently bind and displace nevirapine from its NNRTI pocket making it a potential allosteric inhibition site.

5.3.2.2 Binding modes of ligands to the HIV PR site

Unlike RT for which structure-based drug design took a longer time to materialize because the crystal structural information database was lacking, structure-based design for PR inhibitors has been central to the development of many of the drugs that target this enzyme (Ren and Stammers, 2005). Docking for the HIV PR site was performed on two crystal structures; a wild type with pdb code, 1HXW, which like the rest of the crystal structures used in this study, is a subtype B strain and another coded 2R5P which is a subtype

C strain. Very poor binding free energies were predicted for the successfully docked ligands. These were compounds that had previously inhibited PR activity in direct enzyme bioassays at a high 100 µM concentration (Table 5.3). The poor binding free energies predicted in the *in silico* studies are thus suggestive of the fact that the inhibitions observed at 100 µM (in the direct enzyme assays) were non specific especially because these concentrations were also toxic to cells. Based on these findings, no further analyses of the docked poses were done.

### 5.3.2.3 Binding modes of the ligands with HIV IN sites

Docking studies for IN were done for the compounds that had inhibited HIV IN in the dual 3'P and ST assay pre-screen (Figure A5.1A). The ligands analysed included the Tscs compounds PFK5, PFK7, PFK8 and PFK41 and the gold(I) phosphine thiolate compound, PFK174. Docking was done on the LEDGF binding site (pdb coded 2B4J). LEDGF functions in targeting IN to the chromosome of infected cells and enhances the integration process (Maertens *et al.*, 2003). Cherepanov *et al.*, (2005), speculated that the binding of small molecule inhibitors to the LEDGF binding site is likely to induce defects in HIV replication similar to those seen in mutant viruses. Docking was also done on the ISQ4 site complexed to 5-CITEP (Goldgur *et al.*, 1999) and the 3L3V site in complex with sucrose (Wielens *et al.*, 2010) both in the CCD of IN. The sucrose binding site located 10 Å from the LEDGF binding site was identified by Wielens *et al.*, (2010) as an allosteric inhibitory binding site that can be exploited for developing inhibitors that target LEDGF. Other investigators (Du *et al.*, 2008, Shkriabai *et al.*, 2004) have also identified this site as a putative IN binding site.

Binding predictions for the 1SQ4 site were enthalpically unfavourable suggesting very poor complementarity and hence binding affinity. This site contains only one of the two Mg^{2+} ions that should ideally be found in the enzyme (Cox and Nair, 2006, Bujacz *et al.*, 1997) and because the ligands do not contain metal chelating moieties (also noted for the RNase H site), unfavourable repulsive forces prevailed. This finding is not surprising given that this site is also the DNA binding site since no significant inhibition was observed in the direct enzyme assays that mimicked the integration process. This was the case in the dual inhibitor assay and ST specific assay, except for the once off findings in the dual assay pre-screen (Figure A5.1). Inhibitors of 3'P are known to target the unbound enzyme while IN ST inhibitors target the enzyme/DNA complex in cell-based assays (Hazuda *et al.*, 2000). With regards to docking, there is supporting evidence (Johnson *et al.*, 2006) that suggest that 3' processor inhibitors dock at the HIV DNA site of the enzyme while IN ST inhibitors occupy the position of acceptor DNA (Johnson *et al.*, 2007, Pommier *et al.*, 2005). Based on these theories (which was corroborated by both the bioassays and docking studies on the 1SQ4 sites), one could conclude that the ligands were neither 3' processors nor ST inhibitors.

Binding interactions of the ligands with the LEDGF and the sucrose binding sites on the other hand resulted in significantly lower binding free energy predictions. Interactions with the

LEDGF site were favoured over those of the sucrose binding site as seen from the binding free energies in Table 5.4.

Binding of the ligands to the LEDGF site was in the order of PFK174<PFK5<PFK7<PFK8<PFK41 with corresponding binding free energies of 7.2<8.9<13.2<15.2<18.2 kcal/mol respectively. Only the interactions with the Tscs-based ligands will be discussed since PFK714 is one of the three complexes whose structure cannot be described in detail. The annotated structures of PFK5 and PFK7 depicted in Figure 5.9 will aid in the description of predicted interactions with the LEDGF and the sucrose binding sites.
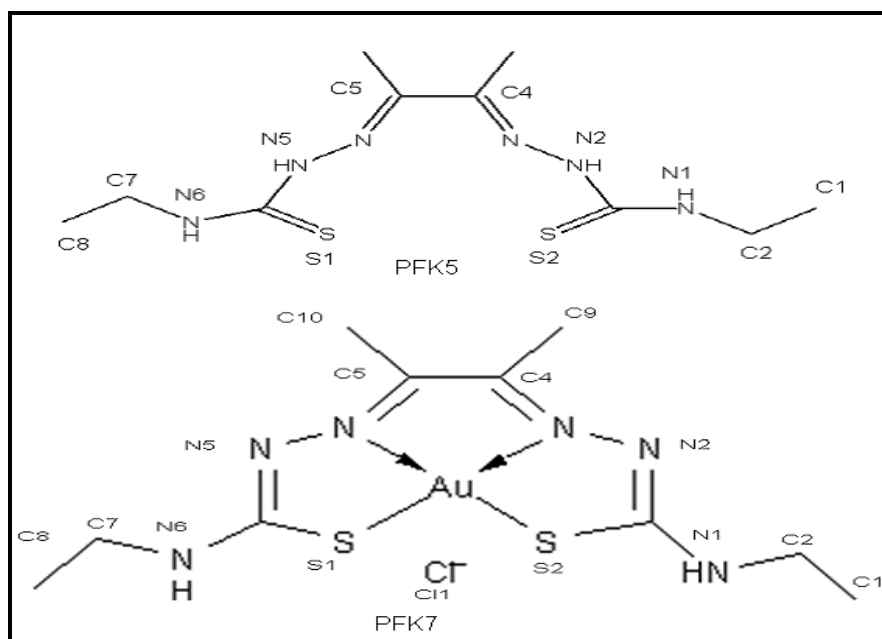

Figure 5.9: Annotated structures of ligands PFK5 and PFK7. Numbers are assigned to the various atoms to facilitate description of ligand-receptor interactions
.

Predicted interactions with LEDGF binding site: The predicted binding of PFK5 with this site which resulted in a binding free energy of 8.9 kcal/mol for the most favoured pose consisted of two H-bond interactions; one between the backbone NH_3^+ group of Glu170 in chain B (flat ribbon) and a sulphate ion (S2) of PFK5 (2.2 Å) and the other between the backbone carbonyl group of Gln168 of chain B and an H-bond donor (N2) of PFK5 (2.5 Å, Figure 5.10A). One of the sulphate atoms (S1) of PFK5 was however not satisfied since it was not involved in H-bonding.

Binding predictions for the corresponding gold complex of PFK5 (PFK7) resulted in better size-shape complementarity (Figure 5.10B) than the free ligand (PFK5) but with a slightly higher binding free energy (13.2 kcal/mol). Mostly hydrophobic interactions where predicted with both receptor chain A consisting of Ala98, Leu102, Ala128, Ala129, Trp131, Trp132 and Met178 of Chain B (Figure 5.10B). Other amino acid residues (polar residues) within close proximity (4 Å) of the side chain methyl groups of PFK7 were Gln95, Thr125, Gln168, His171 and Thr174. The modification of PFK7 to contain polar groups at points close to these amino acid residues may improve binding affinity for this site.
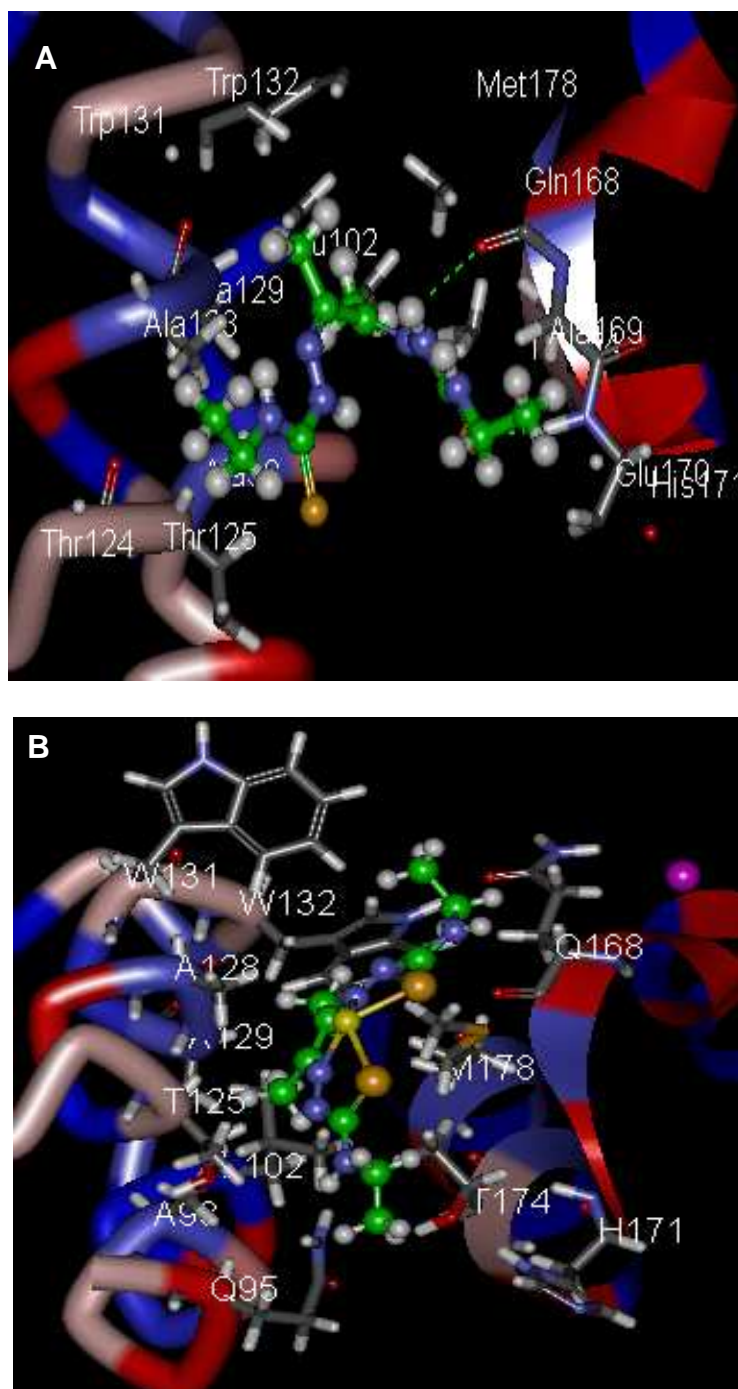
Figure 5.10: Predicted interactions of ligands PFK5 and PFK7 with the LEDGF binding site. Chain A of IN is represented by a tube model while chain B is represented by a flat ribbon model. Better size-shape complementarity was observed for PFK7 (B) with this site than with the free ligand, PFK5 (A). The complex formed mostly hydrophobic interactions with residues Ala98, Leu102, Ala128, Ala129, Trp131, Trp132 of chain A and Met178 of Chain B. PFK5 makes two H-bond contacts with the receptor. Red balls or sticks=O, blue=N, white =H, yellow=Au, orange = S, purple=Cl, grey=C on receptor and green ball on ligand, H-bonds are shown as green dotted lines.

The contacts made by PFK7 with this site are also LEDGF hotspot residues (Wielens *et al.*, 2010). The chloride ion of PFK7 (purple ball in Figure 5.10B) was free floating and made no contact with the receptor. Although PFK5 formed two H-bonds with the receptor and was predicted to have a slightly lower binding free energy than PFK7, it unfortunately had very poor size-shape fit compared to PFK7 which fitted more complementarily with this site. Molecular surface diagrams of the receptor with PFK5 and PFK7 depicting hydrophobic interactions are shown in Figure A5.4.

The interaction of PFK8 with this site resulted in a binding free energy of 15.2 kcal/mol. This compound however did not fit into the active site as snugly as PFK7. The probable reason for this poor fit is because unlike PFK7 this compound has two CH2 groups less (Table 3.4) and could therefore not make appropriate hydrophobic interactions similar to those observed for PFK7. This observation suggests a structure activity relationship.

PFK41, which differs from PFK7 by having two CH3 groups less also did not interact with this site as favourably as PFK7. The most favourable pose was predicted to have a binding free energy of 18.2 kcal/mol and made two H-bond contacts, one with Lys173 and the Cl$^-$ ion (2.4Å) and the other between Gln168 and H-bond donor (N1) of PF41 (1.7 Å). These interactions appeared not to compensate for the poor fit hence the higher binding free energy. The predicted interactions of these gold(III) Tscs-based compounds with the 2B4J or LEDGF binding site demonstrated SAR. IN on its own does not exhibit the same integration activity observed for the IN/LEDGF complex (Michel *et al.*, 2009) making compounds which bind and alter LEDGF interactions potential IN inhibitors.

Prediction interactions with the sucrose binding site: Binding free energies predicted for the sucrose binding site were in the order: 40.1<40.4<42.1<42.3 for PFK5, PFK41, PFK8 and PFK7 respectively. Since these enthalpic contributions were poor, and generally presented similar values for the various ligands, only those of PFK5 and PFK7 (Figure 5.11) as representatives are discussed.

The lowest binding free energy for this site (40.1 kcal/mol) was for PFK5 which formed four H-bonds; one between Lys103 and one of the sulphate ions (S1) of PFK5 (2.9 Å), another between H of NH_3^+ group of Lys173 and N3 of PFK5 (1.7 Å), a third with another H of NH_3^+ group of Lys173 and N3 of PFK5 (2.4 Å), and a fourth with the carboxylate group of Glu96 and the H-bond donor (N1) of PFK5 (2.2 Å) shown in Figure 5.11A. PFK7 on the other hand with a binding free energy of 42.3 kcal/mol was predicted to form two H-bonds with this receptor (Figure 5.11B). One was between the backbone carbonyl group of Val88 and H-bond donor N6 (2 Å) of PFK7 and the other between the floating Cl$^-$ and Gly94 (2.4 Å). The H-bond between the Cl$^-$ and the receptor as well as other predicted interactions with the receptor (Thr93, Gly94, Asn120 and Ser123) were outside the defined sphere for docking. This site may represent a putative binding site.

There appeared to be some fit for PFK7 in this site which is comparable to that observed for the LEDGF site (Figure 5.10) but was clearly not as compatible for this ligand as it was for the LEDGF binding. This is further supported by the differences that were seen in the predicted binding free energies and from the molecular surface diagrams (Figure A5.4B and A4.5 respectively). These findings suggest that the compounds do not have allosteric binding ability and will not displace LEDGF from its binding pocket which is 10 Å away.
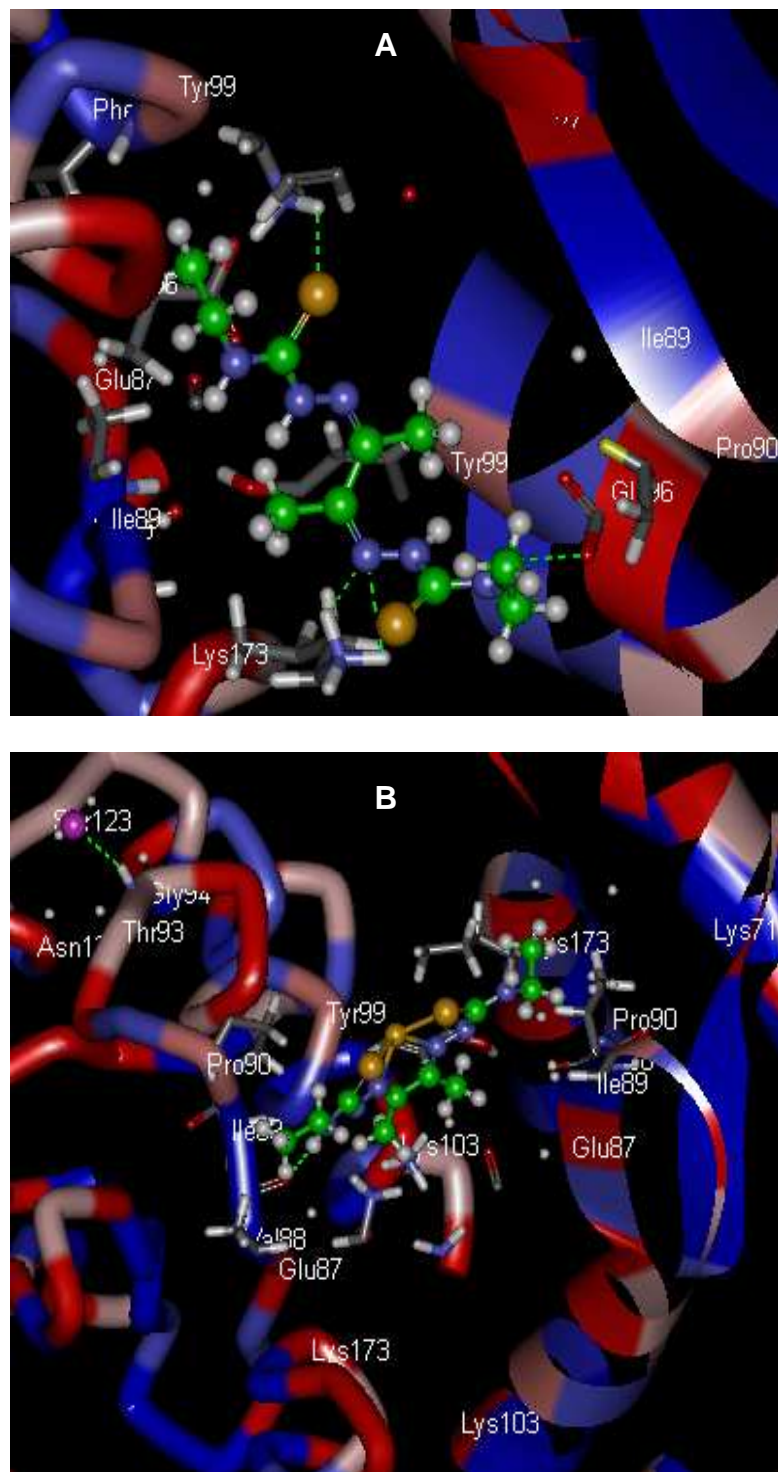
Figure 5.11: Predicted binding interactions of ligands with the sucrose binding site of IN. Prediction interactions of PFK5 and PFK7 are shown. Four H-bonds interactions were predicted between PFK5 and the receptor (A) and two between PFK7 and the receptor (B). The interactions of the ligands with this site showed poor complementarity and thus poor allosteric binding effect. Red balls or sticks=O, blue=N, white =H, yellow=Au, orange= S, purple=Cl, grey=C on receptor and green on ligand, H-bonds=dotted green lines.

The type of inhibition exhibited by IN inhibitors could either be 3'P or ST specific. New targets such as the IN cofactor or LEDGF binding site (Adamson and Freed, 2010), have been identified. In a pre-screen with a dual inhibitor kit, four compounds exhibited >50% inhibition of IN (Figure A5.1A) but this was absent upon subsequent testing (Figure A5.1B) while in the ST specific assay, no significant inhibition was observed. Even though both 3'P and ST inhibitors have been reported in *in vitro* tests, only ST ones have been successful in *vivo* and

subsequently approved for clinical use (Mouscadet *et al.*, 2010, Chirch *et al.*, 2009). Therefore, if the observed inhibition in the pre-screen was due to 3'P, then there is unfortunately no IN therapeutic potential for these compounds since *in vivo* inhibition is unlikely.

With respect to the docking data, the compounds were neither 3'P inhibitors or ST inhibitors, a finding supported by literature (Johnson *et al.*, 2007, Pommier *et al.*, 2005) since interactions with the DNA binding site (1SQ4) were unfavourable. The gold(III) Tscs-based complexes however displayed favourable interactions with the LEDGF binding site of IN in the virtual screening prediction studies. These findings must be confirmed experimentally by using assays specific for determining the effect of these ligands on IN-LEDGF interactions.

5.4 CONCLUSION

Direct enzyme bioassays for RT, PR and IN were performed. Compounds which demonstrated inhibition of these enzymes both in this study and in previous studies (Fonteh and Meyer 2009, Fonteh *et al.*, 2009) were further analysed using complementary *in silico* molecular modelling techniques for the respective receptor binding sites.

In the direct enzyme assays, none of the eleven new compounds inhibited RT while one (PFK7) inhibited PR but at a toxic concentration of 100 µM. Compounds with previous anti-RT activity when tested as controls three years later appeared to have lost this ability (Table A5.1). This loss of activity in the direct enzyme assay was thought to have resulted from one or more of a number of limitations; poor aqueous solubility seen in the ADMET studies and during wet lab studies for some of the compounds (Table 3.8A) and compound age (activity was noted earlier when compounds were freshly prepared soon after synthesis but absent after three years of storage) and the presence of degradation products not detectable by NMR. In addition, the poor complementarity in binding to the RNase H site due to the lack of metal chelating groups was also thought to be one of the possible reasons. The poor stereochemical orientation of ligands with the active site and the high flexibility associated with protein molecules (Mohan *et al.*, 2005, Höltje *et al.*, 2003) meant the ligands could easily be dislodged. This latter possibility together with the mentioned poor aqueous solubility limitation, and the possibility of the presence of degradation products (not detectable by NMR) makes these compounds poor RT inhibitors and poor drug-like candidates in the current form. The inconsistencies in the bioassay findings were therefore not surprising. The inclusion of metal chelating groups (e.g. carbonyl groups) which can bind to the metal ions found in the active sites of RNase H might prove useful in the binding and inhibition of this important viral enzyme and in enhancing the activity of these ligands in the bioassays. Overall, there was SAR in docking studies which appeared to correspond with RT bioassay findings (Fonteh and Meyer, 2009) where the *N,N*-dimethyl-ethane-1,2-diamine containing ligands

(TTC10 and TTC24) were favoured over the phenethyl-amine containing ligands (TTC3 and TTC17, Table 5.4).

Binding affinity to both a subtype B and C variant of HIV PR for compounds which had previously inhibited the enzyme at 100 µM (Table 5.3) was very poor and in some cases refined poses could not be obtained (Table A5.2 in the appendix). It is thought that the inhibition of the enzyme by these compounds was not specific especially considering the high and toxic concentrations (the CC_{50} of the complexes were mostly below 20 µM, Table 4.2) at which enzyme inhibition was detected. It was therefore not surprising that the binding affinity in the molecular modelling studies was very low.

Binding affinity predictions for HIV IN for the Tscs compounds showed favourable binding interactions with the LEDGF binding site (2B4J) but not with the allosteric sucrose binding site (3L3V) or the DNA binding site (1SQ4). The poor binding to the 1SQ4 site confirms the bioassay studies where the compounds were shown to be neither 3'P nor ST specific inhibitors

The interactions of PFK5 and its corresponding complex, PFK7, with the LEDGF binding site were the most favoured with PFK7 making contact with LEDGF hotspots and with better complementarily than the corresponding free ligand (PFK5, Figure 5.10).

Gold complexes have been reported to undergo ligand exchange reactions with sulfhydryl groups of cysteine residues present in the active site of receptors (Shaw III, 1999, Sadler and Guo, 1998). This was not observed in the *in silico* docking assays performed here possibly because the docking program has not been parameterised to include metals in its atom base such that binding interactions of gold complexes with the receptor could not be simulated. Complexation however appeared to confer stability and led to better binding affinity predictions than those seen for the free ligands (Table 5.4). Differences in the binding free energies of free ligands TTL10 and TTL24 which were 76.6 and 78.6 kcal/mol respectively compared to those of corresponding gold complexes, TTC10 and TTC24 which were 10.9 and 12.3 kcal/mol respectively for the RNase H binding site (Table 5.4) are some examples.

Although favourable interactions were observed for the ligands with the RNase H site of RT and the LEDGF binding site of IN, the binding orientations were poor especially with respect to the RNase H site. The binding predictions of the ligands with these sites can however provide crucial information for the design of gold-based compounds that could potentially attain better inhibition in bioassays and *in silico* with energetically favourable interactions.