ALLOSTERIC EFFECTS OF CHICORIC ACID ON HUMAN IMMUNODEFICIENCY VIRUS TYPE 1 INTEGRASE

Muhammad Qasim Fish

A dissertation submitted to the Faculty of Health Sciences, University of the Witwatersrand, in fulfilment of the requirements for the degree of Doctor of Philosophy in Medicine

12/6/2018

Declaration

I, Muhammad Qasim Fish affirm that the work presented in this dissertation is my own. It is being submitted for the degree of Doctor of Philosophy in Medicine in the University of Witwatersrand, Johannesburg. This work has not been submitted before for any degree or examination at this or any other University.

12 day of the month of June, 2018, Johannesburg

Dedication

For my mom, Farieda, and dad, Ganief.

The ones who have always given me unconditional support.

I love you.

Publications and conference proceedings

Poster presentation:

M. Fish, M. Papathanasopoulos, S. Mosebi, A. Harrison and R. Hewer. MAM-14/7th International Symposium on Macro- and Supramolecular Architectures and Materials. Emperors Palace, Johannesburg - South Africa, November 23 - 27, 2014

Abstract

Human immunodeficiency virus (HIV) integrase (IN) is an essential viral protein involved in the integration of the viral DNA into the host genome. Although having a specific catalytic function, it is apparent from mutagenesis studies that IN is pleiotropic and affects the viral life cycle at multiple points other than integration. Compounds that bind to allosteric sites on IN typically disrupt its ordered multimerization and stalls the viral life cycle at various points. Chicoric acid (CA) is a well-known IN inhibitor, however, its mechanism does not follow a conventional active site inhibition, yet it presents with antiviral activity. We thus, hypothesised that CA has an allosteric inhibitory mechanism. To test the hypothesis we aimed to determine alloteric effects of CA on HIV IN.

Site directed mutagenesis was used to develop IN mutants resistant to conventional IN inhibitors and these proteins were purified. These were used in an enzyme-linked immunosorbent assay (ELISA) for comparative resistance profiling of CA and raltegravir (RAL). The ELISA also compared magnesium (Mg^{2+}) and manganese (Mn^{2+}) dependent differences on CA inhibition and to determine the importance of order of addition of assay components. An AlphaScreen assay was developed to test for the disruption of the IN/LEDGF interaction. Crosslinking assays and size exclusion chromatography (SEC) was performed to determine the multimeric state of IN in the presence of CA. Surface Plasmon resonance (SPR) was used to confirm binding of CA to IN and determine the kinetics. *In silico* docking onto the IN catalytic core (CCD) structures was used to identify a possible binding mode of CA at the allosteric binding site.

Resistance profiling depicted a clear distinction between CA and RAL. IN resistance mutants: IN_{Q148H}, IN_{N155H}, $IN_{G140S/Q148H}$ and $IN_{E92Q/N155H}$, showed a fold change in IC_{50} ($FCIC_{50}$) of 3.42, 1.09, 1.43 and 2.90 for CA respectively. While the same mutants showed an $FCIC_{50}$ of 947.99, 392.44, 1262.41 and 583.92 for RAL respectively. Additionally, cooperativity trends between the compound's inhibition profiles were different. It was concluded that CA and RAL have different resistance profiles. Metal dependent differences show that the IN soluble mutant is resistant to CA inhibition only in the presence of Mg^{2+}. This indicates that metal specific

structural differences may play a role in resistance. Order of addition indicated that if DNA was present after CA incubation with IN the inhibition was cooperative, while if DNA was added before incubation with CA the inhibition was non-cooperative. This posits that CA may bind to a single site, possibly only the allosteric site, when DNA is already present in the IN active site. When DNA is absent CA binds to multiple sites, possibly both the active site and allosteric sites. The AlphaScreen indicated that the IN/LEDGF interaction is disrupted by CA with IC_{50} of 237nM (±27nM). Also the disruption is dependent on the order of addition of assay components. Multimerization of IN was increased in the presence of CA as shown by crosslinking assays as well as SEC. Not only did CA induce multimerization of free IN, a non-reducing gel electrophoresis indicated that virus assembled in the presence of CA had increased multimerization of IN, similar to a control. SPR indicated that CA binds to the full length IN with similar kinetics compared to the catalytic core domain (IN_{CCD}) indicating that this domain likely contains the CA binding site. The kinetics indicated a slow k_{on} and k_{off} rate for CA binding. These slow binding kinetics are indicative of a high barrier to resistance. Finally, *in sillico* molecular docking of CA to the active site as well as the allosteric site indicated that it potentially has a dual binding mode on the IN apo-enzyme.

The results show that CA has allosteric effects on IN and may provide a good pharmacophore for further development of allosteric IN inhibitors.

Acknowledgements

I thank Allah (SWT) for allowing me to complete this work.

I would also like to thank the following people:

- Dr Salerwe Mosebi for the major effort in supervision and technical assistance. I would not have achieved much in the way of experimental work without your guidance.
- Professor Maria Papathanasopoulos for supervision, allowing the use of laboratory infrastructure and support during the project especially towards the end.
- Dr Raymond Hewer for supervision, you are the integrase visionary and without you I doubt that any of our work would have reached this point.
- Dr Telisha Traut for the help with the computer modelling.
- Dr Jason Thompho Rashamuse for the synthesis of ACA-1.
- Dr Nichole Cerutti for help with the SPR analysis.
- Mintek for the bulk of the funding.
- The NRF Thuthuka for funding.
- The CMDD team.
- Mom and Dad, I have to mention them again, my parents who have worried over this thesis more than I have. Not to say I haven't stressed over this, so you can just imagine what they were going through.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

3’-P	3’-end processing
AIDS	Acquired immune deficiency syndrome
ALLINIs	Allosteric integrase inhibitors
NTD	Amino terminal domain
AMP	Ampicillin
ARV	Antiretroviral
CS	Capsid
CTD	Carboxyl terminal domain
CCD	Catalytic core domain
CCR5	CC-chemokine receptor 5
CPSF6	Cellular protein cleavage and polyadenylation specificity factor 6
CA	Chicoric acid
CAM	Chloramphenicol
CD4	Cluster of differentiation 4
CXCR4	CXC-chemokine receptor 4
CypA	Cyclophilin A
DCQA	Dicaffeoylquinic acid
DKAs	Diketo acids
Env	Envelope
gp	Glycoprotein
ELISA	Enzyme linked immunosorbent assay

FRET	Förster resonance energy transfer
HAART	Highly active antiretroviral therapy
HIV-1	Human immunodeficiency virus type 1
HIV-2	Human immunodeficiency virus type 2
INBIs	Integrase DNA binding inhibitors
S6	Integrase interactor 1 transdominant mutant
INI1	Integrase interactor 1
INSTI	Integrase strand transfer inhibitor
IN	Integrase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LEDGF	Lens epithelium derived growth factor
LTR	Long terminal repeats
Mg^{2+}	Magnesium
MLV	Maloney leukaemia virus
Mn^{2+}	Manganese
MOA	Mechanism of action
NLS	Nuclear localization signal
NUP	Nuclear pore proteins
OD	Optical density
PCR	Polymerase chain reaction
PIC	Pre-integration complex
PR	Protease

PFV	Prototype foamy virus
RAL	Raltegravir
RT	Reverse transcriptase
RTC	Reverse transcription complex
RSV	Rous sarcoma virus
SIV	Simian immunodeficiency virus
SEC	Size exclusion chromatography
SPR	Surface plasmon resonance
SSC	Stable synaptic complex
ST	Strand transfer
TNPO3	Transportin 3
TTBS	Tween-Tris Buffered Saline
vDNA	Viral DNA
Vpr	Viral protein R
WT	Wild type
Zn^{2+}	Zinc

List of symbols

α	Alpha
Å	Angstrom
≈	Approximately
β	Beta
L	Litre
m	Milli
µ	Micro
n	Nano
π	Pi
±	Plus/minus

Chapter 1: Introduction

1.1. The human immunodeficiency virus and acquired immune deficiency syndrome: A brief history and the impact on South Africa

Human immunodeficiency virus (HIV) infection is a chronic condition that has been a major medical focus for more than three decades [1]. It is believed to originally be a form of non-human primate virus and zoonosis, resulting in human infection [2, 3], likely began through bush-meat hunting in Kinshasa in the 1920s [3, 4]. There is a link between initial viral spread and the prevalence in genital ulcer disease in Central Africa [5] and this was likely exacerbated by a growing trade route network and sex industry in the region [6].

The earliest collected patient sample that tested HIV positive was from a man in the Congo in 1959 [7] and other HIV cases were detected in samples collected in the 1960s [4] and 1970s [8]. These however were only tested post 1983, after HIV identification. Global awareness had only been piqued by the 1980s, this due to the growing cases of acquired immune deficiency syndrome (AIDS) in the western world, initially in the high-risk male homosexual group [9]. The origin of HIV transmission in the Americas was traced back to commutes by Haitian business expats that worked in Central Africa and likely introduced HIV into America a decade prior to 1980 [10]. Since then HIV had been extensively characterised and targeted for medical treatments [11]. The virus was isolated and identified between 1983 and 1984 [12, 13], followed by its genome sequencing in 1985 [14, 15]. These seminal works led to the extensive characterization of HIV clades and elucidation of its replication cycle. Type 1 HIV (HIV-1) originated in Central Africa from a variation in a chimpanzee associated simian immunodeficiency virus (SIV) [2]. It consists of several Groups and has a wide geographical spread, where HIV-1 Group M subtype B is globally well dispersed and subtype C by the numbers is mainly responsible for the pandemic. HIV Type 2 (HIV-2), originally from sooty mangabeys, has a geographical spread mostly restricted to West Africa, however, intercontinental cases are increasingly common [16].

Eastern and southern Africa is the world's worst affected region documented by UNAIDS, and in 2016 it accounted for 52% of the 36.7 million people living with HIV/AIDS globally [17]. HIV is a primary health concern in South Africa, which has the largest single HIV-1 positive population in the world as of 2016, with 7.1 million people living with HIV and a prevalence of 18.9% in adults aged 15-49 [17]. Various factors are implicated in the spread of HIV-1 in South Africa, and are mainly related to social inequalities, lack of education and bad governance [18]. Ideologies of prominent political figures aggravated the situation in South Africa by limiting the availability of clinically proven treatments, which lead to significant socio-economic losses, and instead favouring the practice of ineffective life-style changes [19]. Also, other contributing factors include the region's cultural beliefs and misinformation regarding HIV/AIDS [20], with incorrect public perceptions of HIV increasing between 2005 and 2008 [21]. Strides in the fight against HIV have been made recently, where, as of 2016, there has been 56% coverage of antiretroviral (ARV) treatments for those in need, up from 20% in 2010 [17] and there has been a steep decline in AIDS related deaths between 2010 to 2013 [22]. Although improvements have been made, HIV-1 infection remains a significant problem in South Africa.

1.2. Human immunodeficiency viral life cycle

1.2.1. Host cell selectivity

HIV (Figure 1.1) targets cluster of differentiation 4 (CD4) positive cells during its replication cycle [23, 24], explaining why immune cells are depleted and how HIV leads to AIDS. The cell types include $CD4^+$ T-cells [23], dendritic cells [25], macrophages and monocytes [26]. Like other lentiviruses, HIV can infect and replicate in cells mainly independent of the cell cycle [27, 28]. Infection here refers to the absorption, fusion and entry of the HIV particle into the cell cytoplasm and integration of its genome in the host chromatin [29]. Replication then refers to the production and budding of new virus after infection [29]. Productive infection develops once replicative competent viral progeny are produced. HIV productively infects a wide range of cells at various points in the cell cycle [27, 30, 31], however, its optimal replication is during cell proliferation in activated $CD4^+$ T-cells [30].

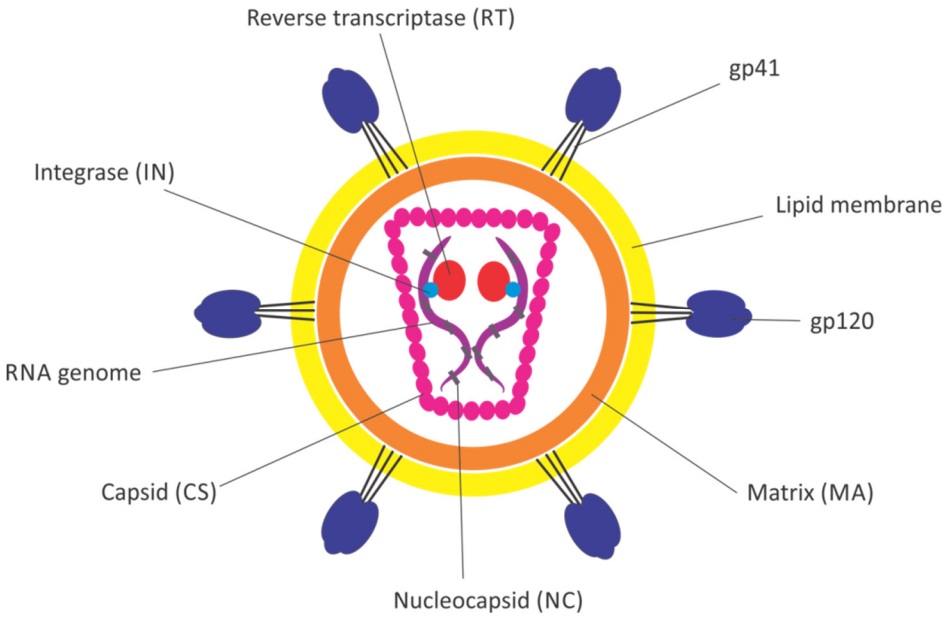


Figure 1.1: A schematic view of the general HIV particle with relevant structures labelled. Similar representations of HIV are used throughout the review. This is neither an accurate representation of scale nor stoichiometric ratio of viral proteins.

The transmission of HIV is mainly through sexual contact with infected persons, and upon exposure of the mucosal membranes at the site of infection, Langerhans cells presumably absorb HIV and presents it to activated memory $CD4^+$ T cells in associated lymph nodes, with evidence indicated *ex-vivo* [32]. An animal model indicated that non-proliferating memory T-cells are depleted during infection with SIV [33] which may relate similarly to HIV.

1.2.2. Steps in viral infection

For the initial infection of non-proliferating cells (Figure 1.2), HIV requires machinery to cross the intact nuclear membrane, and this process is determined by interconnected processes beginning with viral entry into the host. The virus recognises the host through interaction of the envelope glycoprotein (gp)120 to the host CD4 receptor. Upon binding, CD4 undergoes conformational changes and pulls the viral particle closer to the membrane [34, 35]. Thereafter, the gp120 interacts with a tropism specific co-receptor, either CXC-chemokine receptor 4 (CXCR4) [36] or CC-chemokine receptor 5 (CCR5) [37], which leads to specific conformational changes in gp120. This exposes gp41, which subsequently anchors into the cell membrane, and leads to membrane fusion and entry of the virion core into the cytoplasm [29, 35]. Un-coating of the capsid commences in a stepwise manner preceding viral DNA (vDNA) synthesis [38-40], likely allowing host factors and nucleic acids access to the viral reverse transcriptase (RT) thereby initiating reverse transcription [40]. Un-coating has been identified as a determinant in selection of a nuclear localization pathway and targeted proviral integration [41]. After finalising reverse transcription, the viral machinery termed the pre-integration complex (PIC) is formed and travels across the nuclear membrane via interactions with nuclear pore proteins (NUPs) [41, 42], ultimately integrating its vDNA payload into the chromatin, yielding the provirus [43]. Once the host cell is activated the provirus is transcribed. Viral poly-protein precursors are translated from viral genomic RNA (vRNA), while envelope glycoprotein (Env) and regulatory proteins are translated from messenger RNA [44]. The vRNA is packaged along with un-processed polyproteins at the cell membrane through interaction with nucleocapsid (NC). Some viral regulatory proteins, such as viral protein R (Vpr), are packaged into the virion via interactions with viral protein partners.

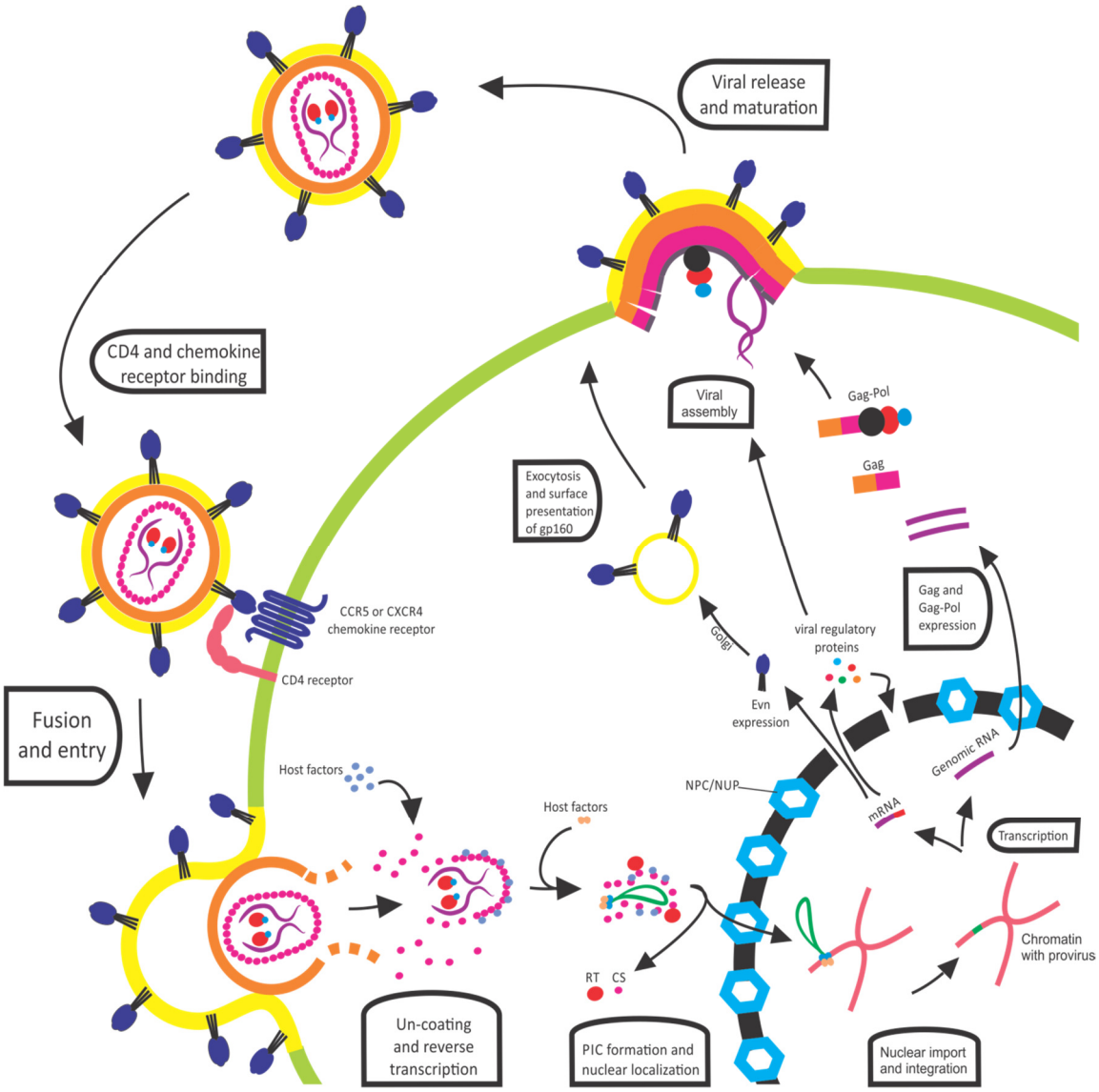


Figure 1.2: Life cycle diagram representing the infection of CD4⁺ T cells by HIV-1. Upon initial infection, HIV binds to CD4 chemokine receptors via its gp120. This causes protein conformational changes leading to fusion and viral entry into the cytoplasm. CS un-coating commences and is followed by reverse transcription, and both processes rely on host factor recruitment. Once the cDNA is transcribed, the PIC is formed, and un-coating completes at the nuclear periphery. The PIC gains entry into the nucleus through interactions with host proteins, and upon nuclear entry, is directed towards transcriptionally active sites on the chromatin. The viral cDNA is integrated into the host genome via catalytic action of IN, and yields the provirus. Upon activation of the cell, the provirus is transcribed into genomic RNA and mRNA, which lead to the translation of polyproteins and Env/regulatory proteins respectively. Gag, Gag-Pol and Env accumulate at the cell membrane, and viral genomic RNA is likewise transported via its interaction with NC in the context of Gag. Some regulatory proteins are packaged into the virion through protein-protein interactions. After budding, PR is activated and leads to the processing of the viral core proteins, leading to viral maturation, and yielding an infectious viral particle and the cycle is repeated.

The major un-processed precursors, Gag-Pol and Gag, are cleaved by viral protease (PR) into the virion and core structural proteins which completes after budding of the virus to yield a mature virion [44].

1.2.3. Un-coating, reverse transcription and nuclear localization

After HIV-1 entry into the host cell, a series of events occur in preparation to reverse transcription, and is termed generally as un-coating. It is a complex process, involving many viral-host interactions and an exact HIV-1 model is unknown, however pathways can be implied from existing information [40]. The viral capsid (CS) protein is the crucial determinant in un-coating. The two domain CS structure is primarily alpha helical and naturally conform in a lattice like configuration of intra N-terminal interactions and inter C terminal interactions of CS hexamers and pentamers to form the closed conical viral core [45, 46]. The timing of un-coating during the infection of non-dividing cells is of vital importance. Mutants of CS, that have hyper-stability of the core, are deficient in reverse transcription and stall viral replication in the cytosol [47]. Similarly, mutants with an unstable core that are uncoated distal to the nuclear periphery, also have reduced reverse transcription [47]. Un-coating therefore needs to be highly controlled, and this depends on the recruitment of host proteins [40].

The RNA genome of the HIV-1 virus undergoes reverse transcription in sequential steps and the enzyme mechanistically responsible is RT [48]. RT functions as a heterodimer with p66 and p51 related subunits [49]. Both RNA to DNA polymerase and RNase H functions exist in its p66 subunit, while the p51 has a structural importance. The reverse transcription reaction is primed with $tRNA_3^{Lys}$, a host tRNA that recognises the HIV long terminal repeat (LTR) sequences, and ultimately produces the viral cDNA from the RNA genome [48]. Viral and host factors interact with RT directly or indirectly to form the reverse transcription complex (RTC). Reverse transcription commences early after entry into the host cell, and is dependent on un-coating [47].

A more current view on HIV nuclear localization implicates viral un-coating as the determinant factor [39-41, 50-52]. The two domain CS protein structure is primarily alpha helical and naturally conforms to a lattice like configuration of CS oligomers, mainly hexamers and fewer pentamers at the ends, to form the closed conical viral core [45, 46]. An exact model of un-coating is as yet unknown with many different views regarding the details. Un-coating appears to commence just before reverse transcription and accompanies the conversion of the reverse transcription complex (RTC) to the PIC, as evident by the lack of CS associated with the PIC [39, 40, 53]. The exact position of complete un-coating is believed to occur at the nuclear periphery [39-41]. The timing of un-coating is important and mutations, that either stabilize or destabilize the CS core, are deficient in reverse transcription [47].

Cyclophilin A (CypA) is a host factor that is packaged into the virion via interaction with Gag [54] and binds directly to CS [55]. CypA is assumed to play a role in the stability of the core by preventing premature disassembly [56]. Once the core reaches the nuclear periphery, host factors such as transportin 3 (TNPO3) binds to and destabilises CS hexamers in opposition to CypA [56]. Remarkably, knockdown of TNPO3, has an impact on nuclear localization [57], chromatin target site selection [41, 58] and integration [59]. TNPO3 was reported to interact with IN [57], however, its interaction appeared to rather map to CS [60]. Mutational disruptions of the IN/TNPO3 did not primarily affect nuclear localization [61]. The pathway of nuclear import has been shown to be selected by CS and mediated by a CypA/NUP358 interchange mechanism, and this pathway selection allows for the specific karyopherin interaction, such as TNPO3, with the IN/vDNA complex [41].

1.2.4. Proteolytic processing and viral assembly

The late stage construction of the mature infectious virion is a complex multistep process that involves the transport of the Gag/Pr55 and Gag-Pol/Pr160 polyproteins to the cell membrane, assembly of the immature virion, budding and finally maturation of the viral core [44]. The Gag and Gag-Pol mediate all processes of virion assembly via a tightly controlled mechanism involving specific conformational changes. The Gag-Pol is produced by a frameshift at p6 and contains the PR, RT, RNAseH and IN protein sequences. Two copies of the vRNA genome are

packaged along with the host $tRNA_3^{Lys}$ primer to initiate reverse transcription in subsequent infections. Env proteins are concentrated and arranged at the budding site on the cell membrane by the Gag polyproteins [44]. HIV-1 IN is processed from the CTD of the Gag-Pol polyprotein during viral maturation [62].

1.2.5. Integrase is pleiotropic and involved in various points in the viral life cycle

Extensive mutagenesis studies have revealed that IN is involved in distinct stages of HIV acute infection involving un-coating, reverse transcription, nuclear localization, integration, viral budding/release, proteolytic processing and maturation [63-70]. Many IN mutations led to non-infectious virions, yet this could be corrected by mixing different pools of mutated virus, which led to the classification of IN mutations as Class I or Class II based on their complementarity (Figure 1.3) [63-65]. Incidentally, the Class I mutants mainly result in the disruption of integration specific catalytic activities [64], while Class II mutants cause IN structural defects which stalls other steps in the viral life cycle [63], and when these viral mutants are co-infected, or in other words complemented to each other, their infectivity is restored. This restoration of activity is due to the expression of both types of mutant virions in the same system, where functional groups of one mutant will complement the deficiency of the other mutant. This complementation thereby restores infectivity. This indicates that IN has a significant role distinct from its catalytic activities and likely revolves around its ordered structure.

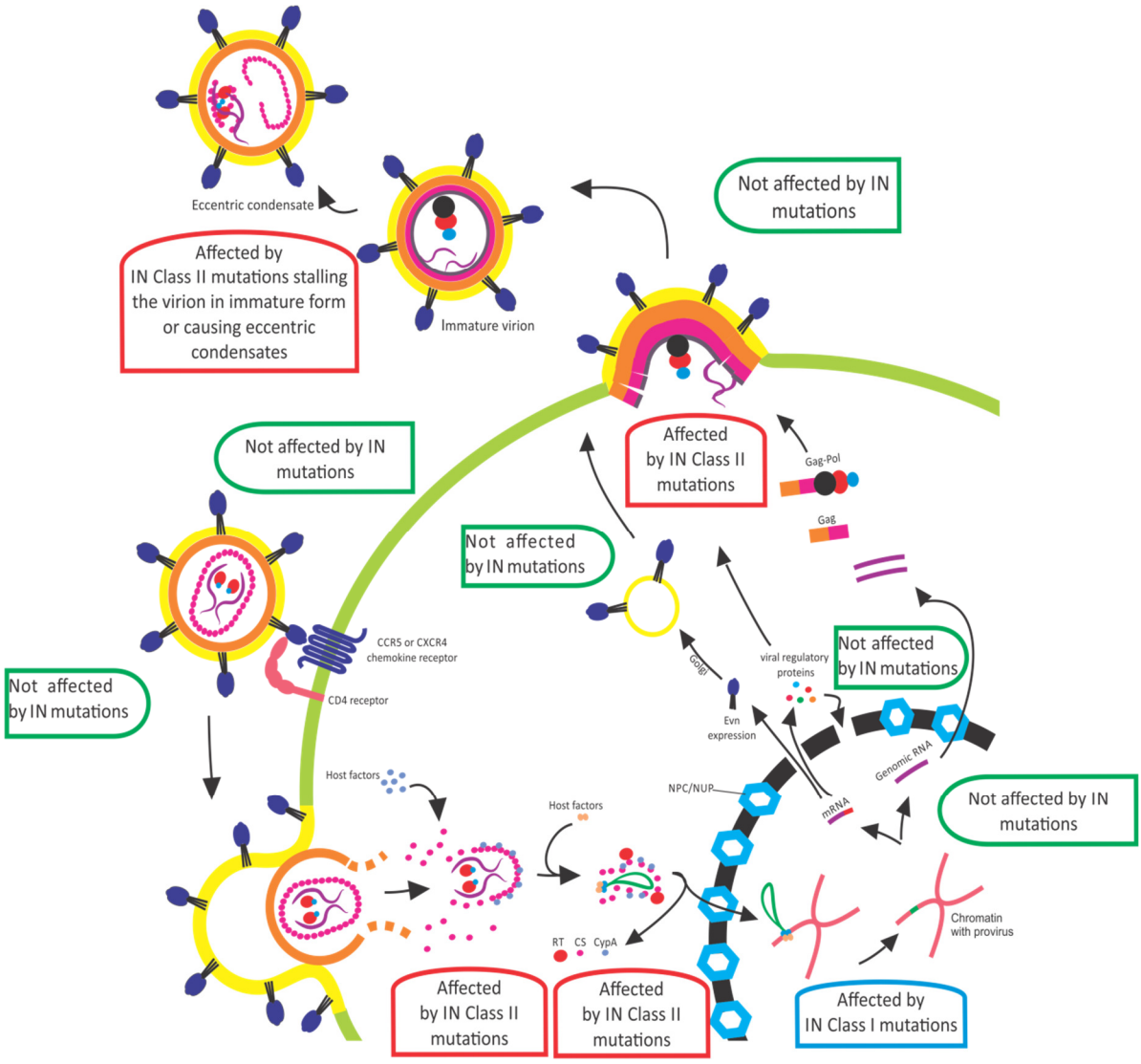


Figure 1.3: The effects of IN mutations on the life cycle of HIV. The interpretation of this figure is best when directly compared to Figure **1.2**, where the steps of the life cycle are explained as either unaffected by (green labels) or affected by IN (red and blue labels). IN is pleiotropic and influences the HIV life cycle at various points as indicated. RT, un-coating, nuclear localization, viral assembly and maturation are all affected by IN Class II mutations (red labels). IN Class I mutations (blue label) only affects catalytic processes leading to defects at integration.

1.3. Integrase

1.3.1. Phylogenetic classification

The retroviral IN superfamily comprises proteins that share a structural RNase H fold similarity [71, 72]. Within this group there are transposons and integrases that utilize the acidic residues aspartate, aspartate and glutamate triad at their catalytic site, and hence, termed the DDE family [72]. HIV-1 IN is a member of the DDE family, with acidic residues D64, D116 and E152 at the catalytic active site [73, 74]. This allows IN to carry out the critical functions of viral genome processing and integration into the host chromatin [73, 75].

1.3.2. Structural organization

HIV-1 IN is a 32kDa protein [76], with a 288 amino acid length that has three distinct domains [77, 78]. The amino terminal domain (NTD), between amino acids 1-50, mainly consists of alpha helices [79] and presents as a putative H12, H16, C40, C43 (HHCC) zinc (Zn^{2+}) binding motif [77, 80] (Figure 1.4). The NTD coordination of Zn^{2+} is structurally important for tetrameric formation [80]. The NTD's involvement has also been suggested in the correct positioning of the vDNA ends at the catalytic site, due to severe loss of catalytic activity in NTD deletion mutants [81]. Yet, the NTD has not been implicated in direct vDNA binding, indicating that it has an important allosteric role in forming the IN/vDNA complex. The NTD is connected to the catalytic core domain (CCD) via a flexible linker sequence [79]. The CCD, between amino acids 52-212, contains five β-sheets flanked by α-helices in a characteristic RNase H fold [73, 79, 82, 83] (Figure 1.4). It has a large hydrophobic protein interface, and isolated CCD domains form well defined dimers as shown by crystal structures [73, 79, 82, 83]. Crystallization of the CCD was only possible after the identification of the F185K mutation, which increased protein solubility [73, 84], and is the only domain to be crystalized on its own. It contains the active site DDE triad and is responsible for vDNA binding and performing catalytic functions of IN [73, 75]. It coordinates divalent metal ions, either magnesium (Mg^{2+}) or manganese (Mn^{2+}) [85], at the active site which is used during the transesterification reactions [60]. Due to its abundance, the Mg^{2+} is the putative coordinating metal ion used physiologically. Part of the active site comprises a highly disordered loop that does not resolve in crystal structures, and this flexibility is critical for catalytic functions [86]. The dimerization

interface of CCD is proven significant in protein-protein interactions with certain host factors, and for correct structural coordination [87, 88]. Thus the CCD is the critical target domain of IN, important in structural as well as catalytic functions. The CCD is connected to the carboxyl terminal domain (CTD) via a α-helix spanning residues 195-220 [89]. The CTD, between positions 220-288, has a β-barrel of five β-sheets in a SH3 fold [89] (Figure 1.4). The CTD is involved in non-specific DNA binding, and possibly tethers IN to chromatin [78, 89].

IN is functional in tetrameric or higher order oligomeric form [91-93], and once bound to vDNA is known as an intasome [94]. The structure of a functional HIV IN intasome has not yet been resolved, and the available HIV IN domain structures offer an incomplete view of a functional HIV intasome. The arrangement of the HIV CCD dimer, as seen in crystal structures with active sites at opposite poles of the molecule [82], is insufficient in explaining a five base stagger generated by IN's transesterification reactions [91]. For the homology models to be in agreement with a 5 base stagger outcome, the HIV intasome arrangement requires an oligomeric form. Many models have been proposed, and it is evident from their inter-variability that the HIV intasome prediction is a challenging task, especially the arrangement of the active sites in a conformation that supports the five base stagger [95]. The crystallization of the prototype foamy virus (PFV) intasome was a major breakthrough and provides a base on which the HIV intasome can be modelled [94, 96]. The functional unit of the PFV intasome is a dimer of dimers with two peripheral subunits and a single core subunit (Figure 1.5 A). This arrangement alleviates the problems experienced when modelling the HIV intasome from domain crystal structures (Figure 1.5 B). The HIV CCD crystal structures indicated a prevalent CCD-CCD dimer [73, 79, 82, 83]. This interaction is noticed between a peripheral monomer and a core monomer in the PFV intasome, and forms two dimers. These two dimers are joined together by intermolecular contacts that are shared between two juxtaposed monomers of the core subunit. The core subunit is responsible for the catalytic functions and its individual domains play ambiguous roles in DNA binding. This arrangement places the NTD in close proximity to the vDNA ends at the catalytic site, indicating precedence to its role in catalytic functions and tetramerization. Two CTD domains of the core subunit, while also having some contact with the vDNA ends at the active site, are diametrically positioned between the CCD and NTD of the core monomers in a CCD-CTD-NTD arrangement.

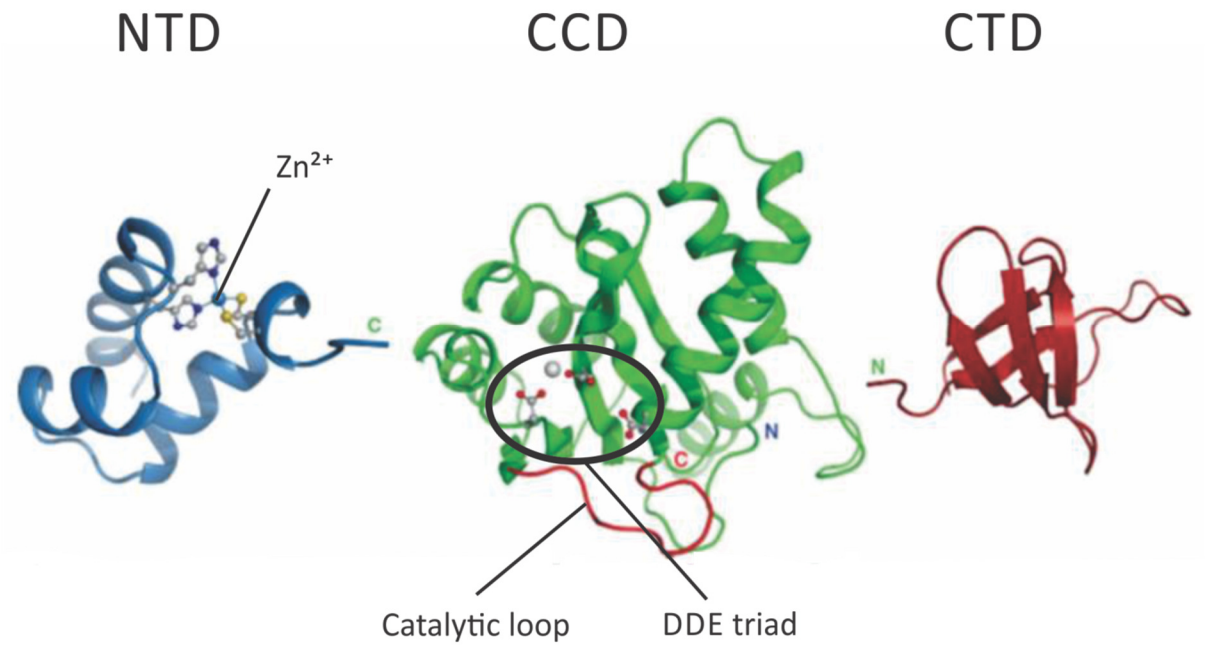


Figure 1.4: The three domains of IN. The alpha helical NTD (Blue) coordinates a Zn^{2+} through the HHCC motif. The CCD (Green), a mixture of Beta sheets and alpha helices, contains the active site with the DDE triad and the catalytic loop. The CTD (Red) presents with a distinctive SH3 beta barrel fold. Picture obtained and modified from Jaskolski *et al*. 2011 [90].

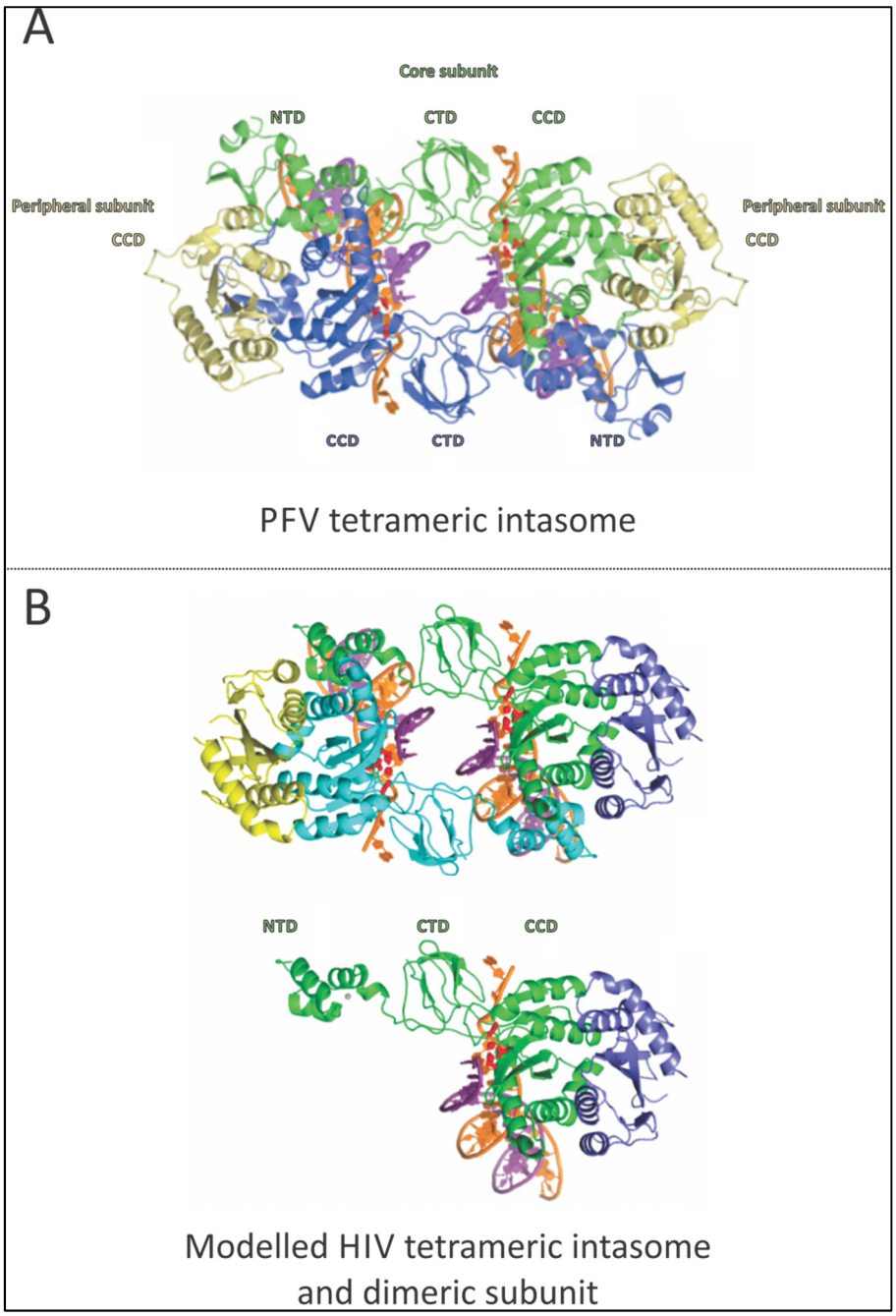


Figure 1.5: The tetrameric intasome of IN. A) The crystal structure of the PFV intasome (copied from [94]). The core subunits (green and blue) have high order and resolution which allowed for the diffraction of IN domains, and indicated the CCD-CTD-NTD arrangement. The peripheral subunits form a CCD-CCD dimer with each core subunit, and these form part of the diffractable structure, however the CTD and NTD of the periphery are disordered and not resolved. All domains appear to be involved in some interaction with the vDNA (purple and orange helix). B) A model of the HIV intasome and single site dimer based on the PFV structure (copied from [97]).

This is ideal for the placement of the catalytic site, which is flanked on two sides with the CTDs of the core subunit, and can potentially bring the active site into contact with host chromatin, enabling integration. The NTD and CTD of the peripheral subunit are not defined in the PFV crystal structure, indicating their flexibility and extrication from the defined intasome, where they are possibly important in target binding or interaction with host factors. To date, this model is a best fit to explaining the mechanistic behaviour of various IN proteins, and the HIV intasome has been modelled according to the PFV organization (Figure 1.5 for one example), and these models have passed some validation [97, 98]. Yet due to differences between PFV and HIV IN, unfavourable residue deletions in the CCD-CTD linker are required for model fitting [97], or reveal a helix-loop-helix linker that is absent from crystal structures [98]. The recent elucidation of the octameric Rous sarcoma virus (RSV) intasome reveals the diverse conformational possibilities of the multimeric IN intasome [99]. The RSV single subunits have the same CCD-CTD-NTD arrangements as seen with PFV, though, important CTD-CTD and CTD-NTD interactions between different subunits are additionally revealed. The minimal structural organization required for concerted integration by the HIV intasome is a tetramer [93, 100]. Yet, tetramers as well as possible octamers are found in HIV virions [92]. A definitive HIV intasome model remains elusive at present, and although validations based on the tetrameric models of the HIV intasome have been immensely informative, obtaining the HIV IN intasome structure remains a relevant endeavour.

The intasome structural models provide an ostensible view of the involvement of all IN domains for interactions with vDNA, and points to an important CCD-NTD interaction for tetramer organization [97]. The CCD residues Q164 and R187 are predicted to contact the NTD backbone through K14 and Y15, respectively, while K186 and K188 form salt bridges with E11 and D25, respectively (Figure 1.6). An earlier tetrameric crystal structure of a two domain IN_{1-212} fragment described a similar interaction of CCD-NTD contacts [79]. The phenotype induced by the F185K, and K186Q mutation indicated that these residues are important for multimerization [84, 101]. The Zn^{2+} coordination of the HHCC motif is required for tetramer formation [80, 102, 103], indicating an important role of the metal for IN structure by promoting the NTD-CCD interaction.

Like other transposons, IN requires the interaction of divalent metal ions at its active site to carry out catalytic functions [104]. Yet, the coordination of Mn^{2+} and/or Mg^{2+} also affects the assembly of the intasome by promoting multimerization [105, 106] and leads to metal induced aggregation effects [102, 106]. Structural changes involving the NTD and CTD domains of IN were apparent in the presence of Mn^{2+} [106, 107]. A profound conformational change is also experienced with the isolated CCD in solution, where a significant increase in resolution of nuclear magnetic resonance spectra was observed in the presence of the divalent metal [108]. Site specific profiling indicated changes in the NTD at K14, at the CCD between residues 161-173, and also in the CTD between residues 259-284 [107]. These studies proposed a CCD-CTD interaction that is required for catalysis. Interestingly, the HIV intasome model [97] was used for validations of the CCD-CTD interaction [109] (Figure 1.6), which indicated important contacts between the catalytic loop residue I141 and the CTD's V259.

Consequently, the coordination of divalent metal ions allows for the formation of catalytically competent conformations of IN. However, metal induced conformational changes differ between Mn^{2+} and Mg^{2+} themselves. The cooperative binding of IN to vDNA prefers Mn^{2+} over Mg^{2+}, when Zn^{2+} is absent [113], and the opposite is true when proteins were purified in the presence of Zn^{+2} [114]. This suggests differing conformational effects between Mn^{2+} and Mg^{2+}, where Mn^{2+} seems to not require the NTD's co-ordination of Zn^{2+}.

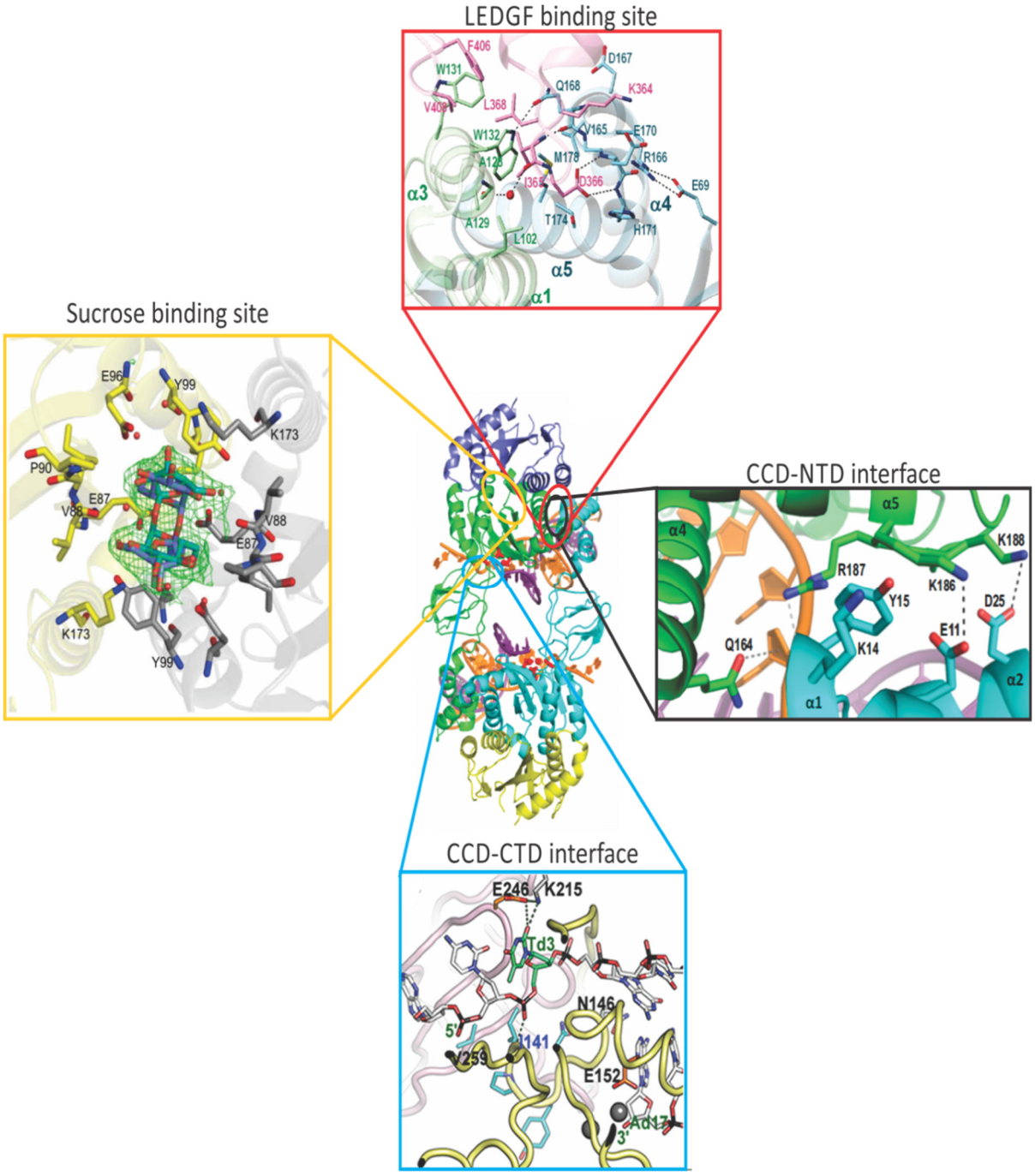


Figure 1.6: Important structural sites of HIV-1 IN intasome (centre and copied from [97]). LEDGF binds IN at its specific site shown in the red block insert (copied from [110]), where the notable interaction depicts LEDGF's (pink residues) D366 in a direct contact with IN's E170 and H171 (blue residues). The black box insert (copied from [97]), shows that the intasome model of the IN tetramer predicts that the CCD (green residues) contacts the NTD (blue residues). The blue box insert predicts an interaction between CTD (yellow ribbon/line) the CCD (pink ribbon/line) (copied from [109]). The yellow box insert shows important allosteric site that was discovered through its interaction with sucrose (green cage) (copied from [111]), and this symmetrical region is required for viral replication [112].

1.3.3. IN catalytic functions

Concerted integration is required for HIV-1 replication, and entails two sequential one-step transesterification reactions catalysed by IN which initiates in the cytoplasm and ends in the nucleus (Figure 1.7) [43, 75, 115]. There are various oligomeric forms of free IN in solution [106] as well as in virions [116], yet when the IN enzyme recognises the vDNA ends at the highly conserved 3'-CA nucleotides, it forms distinct multimers [100, 117], and is described by some as a stable synaptic complex (SSC) [91]. It then cleaves the adjacent terminal GT by nucleophilic attack of the phosphodiester bond using a water hydroxyl as donor, catalysed by the metal coordinating active site of IN. This is known as 3-prime end processing (3'-P) and exposes the hydroxyl group of the 3'-CA on both the vDNA ends [43], and the free 3'-OH is positioned close to the active site, ready for the next catalytic step. The 3'-P completes in the cytoplasm and the SSC migrates to the nucleus. The next step is strand transfer (ST) which is also a transesterification reaction, and commences after the SSC contacts the host genome to form the ST complex (STC). Here both free 3'-hydroxyls at the vDNA ends are used as donors for the sequential nucleophilic attack of the chromatin double stranded DNA [91]. The ST reaction ligates the vDNA into host chromatin and yields the provirus with a characteristic five base stagger and 5'-CA overhang [91]. This stagger indicates an important structural dynamic of the IN/DNA complex and has been used to validate structural models. The IN remains bound to the proviral/chromatin junction and is eventually removed by DNA repair machinery, which in addition, is believed to fill the five base staggers and remove the 3'-overhang at the flanks of the provirus [115].

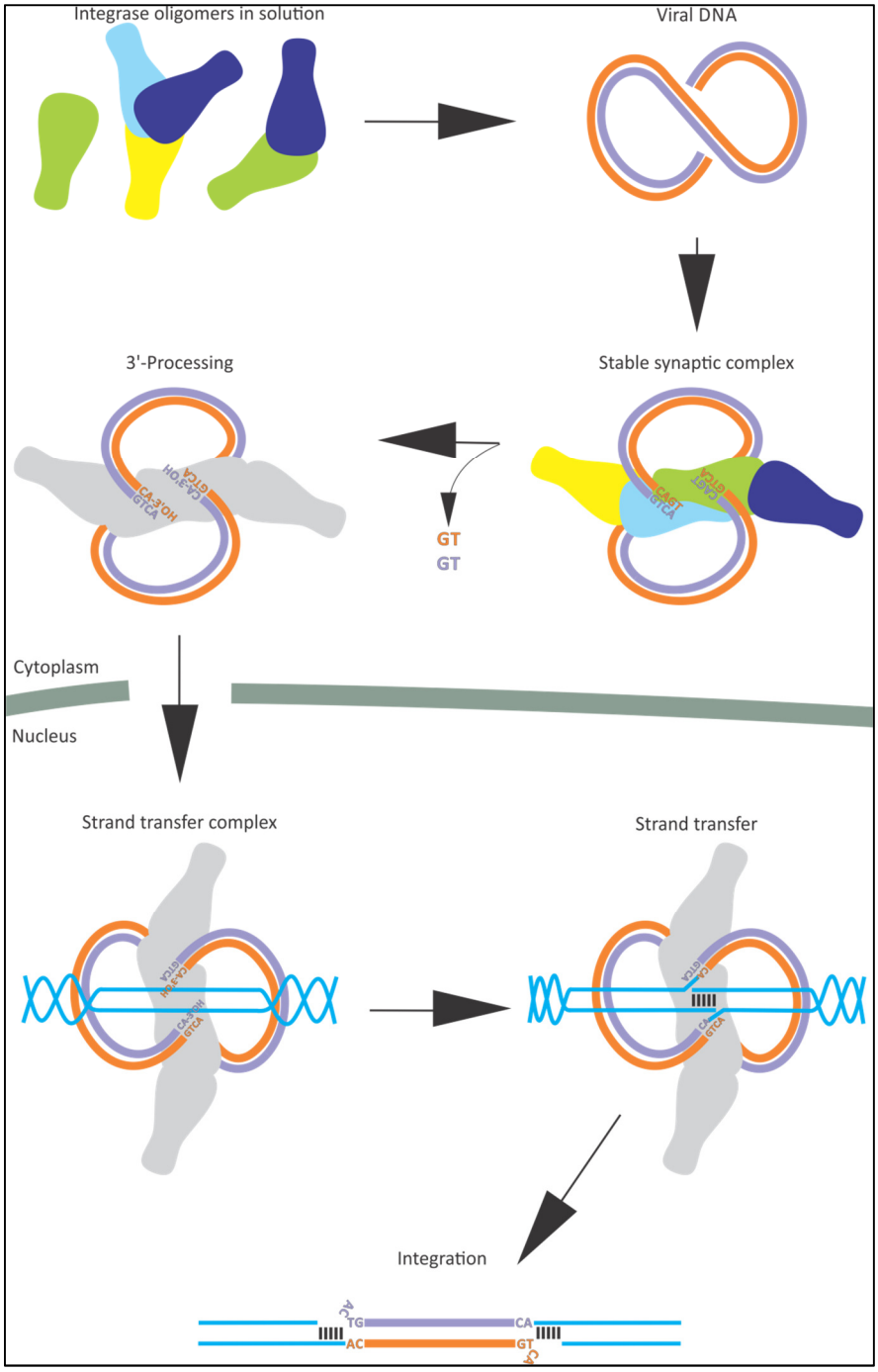


Figure 1.7: Concerted integration of HIV. IN oligomers are ordered into a stable synaptic complex (SSC) by recognizing 3' vDNA ends. A terminal 3'-GT is removed by phosphodiester cleavage, exposing the 3'-hydoxyl, and is known as 3-prime end processing (3'-P). Once in the nucleus the SSC contacts the host genome and is now known as the stand transfer complex (STC). A second transesterification reaction, using the 3'-hydoxyl, joins the vDNA with the host genome, and this is known as stand transfer (ST). ST produces a 5 base stager with a terminal 5'-CA overhang that is ultimately corrected by host DNA repair mechanisms, thus completing integration.

1.4. Viral integration

IN deletion mutants indicated that integration is not necessary for viral protein production but required for productive infection in cell culture [118]. Early mutational studies of the catalytic core and other conserved residues indicated IN's catalytic as well as pleiotropic nature during HIV replication [67, 119, 120]. These mutations can be classified according to the stage of the viral life cycle that it influences [63], where Class I mutations are directly responsible for integration defects as evident by the accumulation of 2LTR circles (Figure 1.8) compared to wild type (WT), and mainly involve mutations around the active site such as Q62, D64, D116, E152, N144, N155, Q148, K156, K159 [64, 70]. The most clinically successful IN inhibitors to date are the integrase strand transfer inhibitors (INSTI). These compounds specifically target the active site and are affected by changes of these Class I amino acids. To show the importance of the catalytic function of IN and its implications in the viral life cycle, the mechanism of action (MOA) of raltegravir (RAL) can be explained.

RAL was the first FDA approved INSTI for the clinical treatment of HIV-1 [121]. Biochemical and structural and modelling data, including the crystallization of PFV IN [94, 96], help explain RAL's MOA (Figure 1.9). This information represents a fragmentary perspective on RAL MOA, as crystal structures with RAL bound at the active site of HIV IN remain elusive, and is probably due to INSTIs distinct binding requirement to an IN/vDNA complex [122, 123]. The PFV structures have been used as a surrogate to understanding RAL's MOA and have shed light in this regard [96].

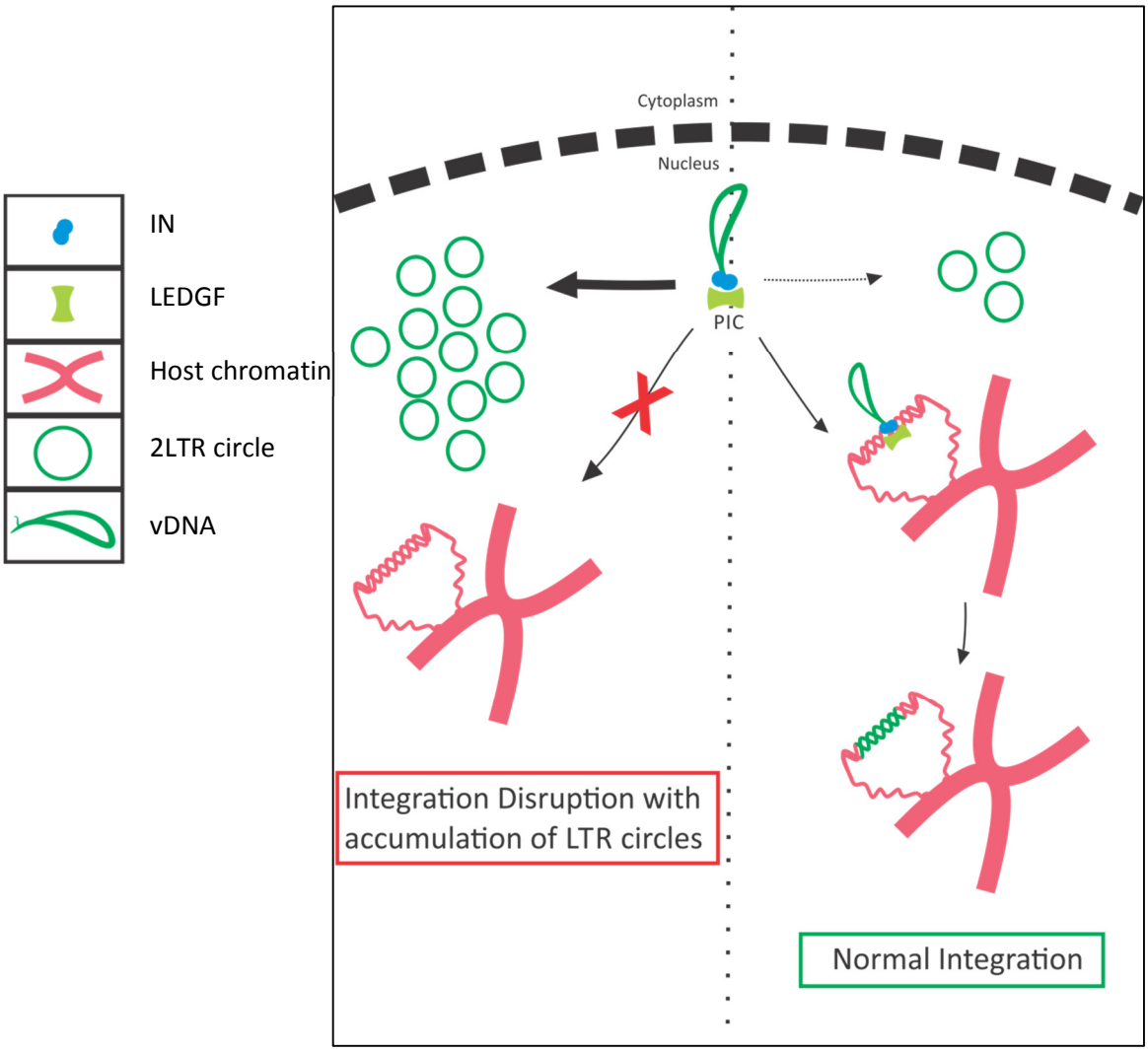


Figure 1.8: Factors affecting HIV integration. The pre-integration complex (PIC) is formed in the cytoplasm at the nuclear periphery and includes host factors such as lens epithelium derived growth factor (LEDGF). Upon nuclear entry two scenarios are presented in left and right panels. The left panel indicates the effects of Class I mutations and integrase strand transfer inhibitors (INSTI's), which both stall integration before the STC is formed. This leads to the accumulation of 2LTR circles with the subsequent disruption of integration. The right panel indicates the normal integrase process, which includes minor auto-integration (2LTR circle formation), formation of the STC and subsequent proviral formation [115].

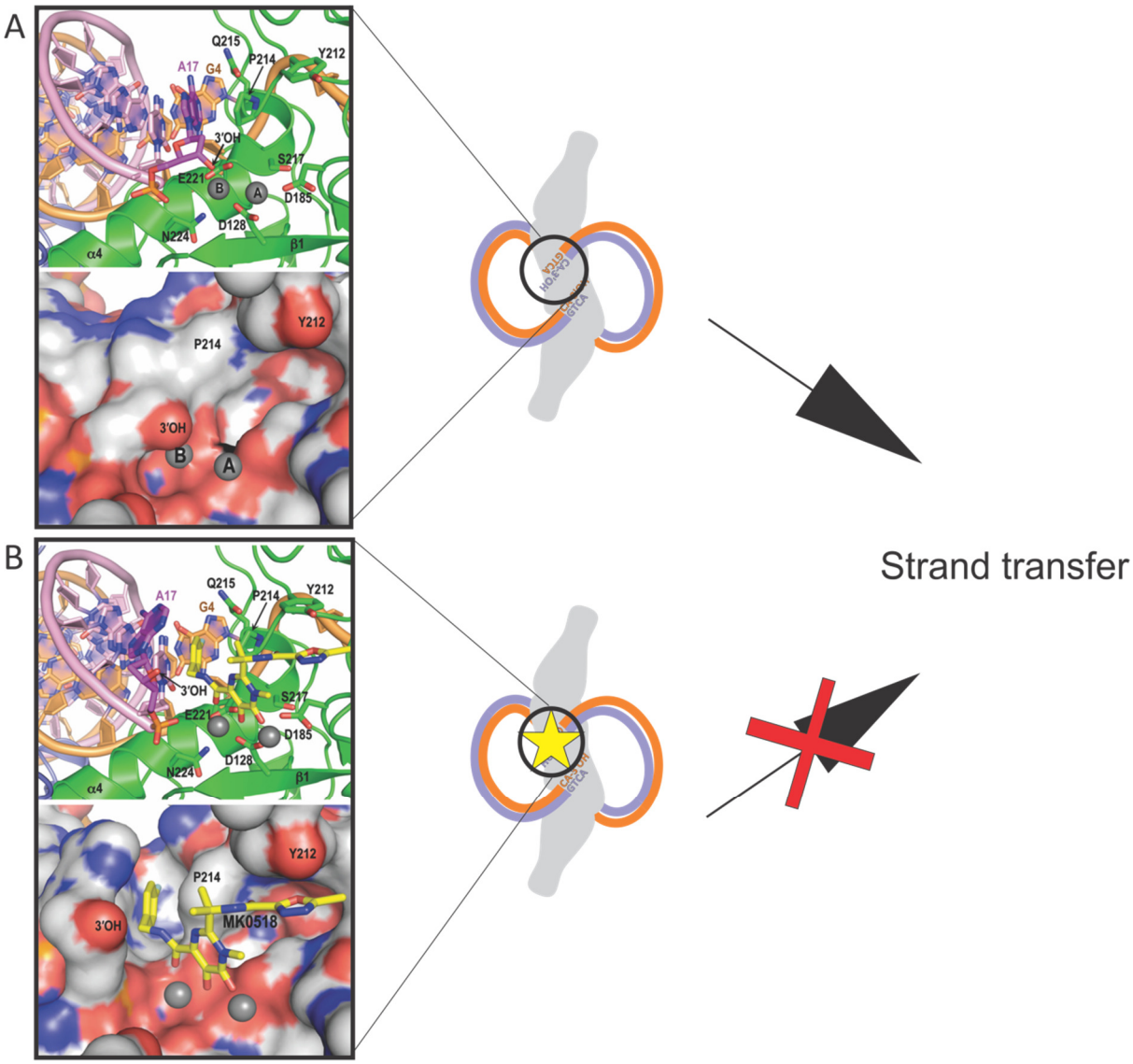


Figure 1.9: The IN/vDNA complex and its inhibition by RAL. Structural diagrams of the PFV intasome crystal structures are shown in the inserts and schematic representations are also presented. In each insert, the top panel indicates ribbon and ball-and-stick diagrams while the bottom panel shows the solvent accessible surface. The green ribbon represents the PFV IN amino acid backbone while the vDNA is presented as ball and stick diagrams in light purple (transferred strand) and orange (non-transferred strand). RAL, also known as MK0518, is coloured in yellow. The surface diagrams are coloured according to atoms, with carbon, oxygen and nitrogen coloured as light grey, red and blue respectively. Metals are shown in dark grey. A) After 3'-P the vDNA's free 3'-OH is in close proximity to the active site metals and is ready for the final nucleophilic attack during strand transfer. B) RAL binds the active site through its interactions with Q215, P214 and Y212, the purple CA dinucleotide and orange guanine 4. RAL's co-ordination with the metal ions effectively displaces the free hydroxyl of the terminal purple adenosine, ultimately inhibiting strand transfer. The inserts were copied from Hare *et al*. 2010 [94].

Diketo acids (DKAs), the precursors to RAL, have been extensively characterised and their inhibitory activity was determined to be dependent on metal chelation [94, 96, 124-130]. DKAs were among the first selective INSTIs discovered [131]. Although structurally distinct, RAL's mechanism of action was recognised to follow DKAs [132, 133], with its reactive oxygen species in the β-keto enol group coordinating either one or two metal ions [130], and due to the β-keto enol moiety, RAL has a predicted activity towards both Mg^{2+} and Mn^{2+} [134]. This was supported with PFV structures, indicating that there was no difference between Mg^{2+} or Mn^{2+} co-ordination by inhibitors at the active site, and inhibition was experienced in the presence of both metal cofactors [94]. It was predicted that a binding mode for RAL at the active site is only available after 3'-P of vDNA [135]. This mechanism of action (Figure 1.9 B), that displaces vDNA after 3'-P, was supported by PFV crystal structures [94, 96], and explains the selective inhibition of RAL towards the ST reaction. RAL does not form hydrogen bonds with any amino acids in the PFV active site [94], however, it has Van der Waals interactions with P214, Q215, the terminal CA dinucleotide of the transferred strand and guanine 4 of the non-transferred strand [94]. In addition, RAL has a hydrophobic interaction and a π-π stacking interaction with P214 and Y212 (corresponding to P145 and Y143 in HIV IN) respectively [94]. RAL is therefore believed to: 1) co-ordinate metal ions at the active site, 2) require vDNA to be present for the formation of RAL's binding site, and 3) competitively interacts with the active site of IN and displaces the 3' adenosine of vDNA after 3'-P [94, 96, 122, 135].

Clinical trials indicated that RAL was highly potent against HIV-1 in treatment experienced patients that presented with high levels of drug resistance to RT and PR inhibitors [136]. This confirmed RAL importance in the treatment of HIV. However, as is common in HAART treatments, RAL treated patients rapidly developed resistance; with documented cases appearing as early as phase II clinical trials [137]. These resistance mutations were analysed from patient samples, and the primary resistance pathways identified were at positions Q148, N155 and Y143 [136]. Secondary mutations, G140S and E92Q, are associated with positions Q148 and N155 respectively, and allow for increased viral fitness as well as higher resistance. PFV IN mutations at positions S217 and N224 correspond to HIV IN mutations at position Q148 and N155, respectively [96]. PFV IN mutants resistant to RAL, when compared to the wild type, required a pronounced changed in active site conformation to allow inhibitor binding

and this was energetically unfavourable, thus reducing RAL binding at the active site [96]. This indicated that these resistance mutations cause a rearrangement at the active site, thereby reducing efficient metal co-ordination and activity of the enzyme [96]. This activity tax, however, allows for a general steric hindrance that disrupts binding of the INSTIs [96].

1.5. Mechanisms of allosteric inhibition: Implications of modulating integrase oligomerization

Allosteric inhibitors target important structural features distal to the active site and may directly disrupt effector binding, or bind a serendipitous pocket that leads to unfavourable structural changes [138]. The success of developing a viable allosteric inhibitory approach is dependent on high quality structural information of the target. This is a significant challenge in the case of HIV IN, as its specific intasome structure is lacking. Mutagenesis studies have revealed IN's pleiotropic nature describes through the evidence of Class II mutations. These are important for IN's structural integrity, as will be highlighted below. These phenotypes present with a severely attenuated viral replication as a result of detrimental effects on virus assembly, maturation, reverse transcription, and nuclear localization [63] (Figure 1.3). Exploiting IN's pleotropic nature as a target is a focus of current inhibitor development. Promising candidates have a number of different names in the literature, but as a blanket term will be referred to here as allosteric integrase inhibitors (ALLINIs), as is practiced by Engelmann *et al*. [139-142].

1.5.1. Allosteric disruption of the integrase interaction with lens epithelium derived growth factor leads to aberrant multimerization and late stage replicative defects

The lens epithelium derived growth factor splice variant p75 (LEDGF/p75 or just LEDGF) interacts with a number of lentiviral INs through interaction with its integrase binding domain (IBD) [143] (Figure 1.6), and this host factor interaction is significant during the HIV life cycle [143, 144]. The LEDGF transcription factor is a member of the hepatoma-derived growth factor family, and contains a chromatin binding element defined by its PWWP domain [145]. LEDGF binds to the PIC and directs it towards, and tethers it to, transcriptionally active sites on the chromatin [146-149] (Figure 1.8). Initially LEDGF was found in complex with high order

IN oligomers, either tetrameric or higher order, within human cell culture [92]. The structural basis for this recognition involves the interaction of two LEDGF IBD loops with IN's CCD dimer interface [110]. LEDGF loop one interacted with IN's W131, W132 and loop 2 interacted with IN's I161-E170 [110] (Figure 1.6), and the significance of these residues was confirmed by biochemical testing [87]. The knock down of LEDGF in cell culture caused disruption in chromatin targeting and viral integration [148-151], while the overexpression of the transdominant IBD peptide likewise disrupted integration [152], implying good candidacy for therapeutic targeting. Interestingly, previous work highlighted the importance of class II mutations, IN_{V166A} and IN_{R166A}, which were shown to be primarily defective in 3'-P and ST at post nuclear entry, while still complementary to class I IN mutations [69], thereby providing an independent verification of the importance of the IN/LEDGF interaction, specifically during integration.

The minimal interaction of LEDGF with IN is mediated through the LEDGF IBD as determined by mutagenesis and structural studies [87, 110], and reveals important direct contacts particularly between LEDGF's D344 and IN's E170 and H171 (Figure 1.6). Amino acid residues that are indirectly involved in the IN/LEDGF include V165, R166, Q168, L172 and K173 [153], and although they do not contact LEDGF IBD directly, they are important in the conformational integrity of the IBD binding site [154]. The primary interaction of LEDGF IBD is with the CCD dimer interface [110], yet the LEDGF IBD in complex with a two domain IN of HIV-2 and maedi-visna virus revealed additional contacts with the NTD [155, 156]. Although a direct link between LEDGF and the NTD of HIV-1 IN has not been definitively established, the IN/LEDGF interaction was shown to protect amino acid residues at the proposed CCD-NTD interface [79, 88], involving the IN NTD's H14 and CCD's L107, R166, K186, R187 and K188 [88]. This confirmed previous tests where IN with a defective H12N mutation did not interact with LEDGF [148]. Considering the tetrameric model of IN depicted above (Figure 1.6), it appears that the CCD-NTD interaction between two core subunits of the IN intasome are either contacted directly by LEDGF, and this appears to be the consensus view, yet IN's NTD may also be allosterically altered by the IN/LEDGF interaction.

The chronology of LEDGF's interaction with the viral PIC during the viral life cycle is not well established. LEDGF is found within functional lentiviral PICs and small amounts may bind within the cytoplasm early during viral infection [157], however, LEDGF is not required for functional PIC formation and is rather required for transcriptional site targeting [149] and is similar to TNPO3 in this regard. The tetrameric form of IN is stabilized and promoted by its interaction with LEDGF [88, 92, 100, 158, 159]. LEDGF also stabilizes the interaction between IN's CCD-CCD dimeric interface and inhibits subunit exchange [88]. LEDGF's stabilization of IN tetramers alters catalytic activities by promoting 3'-P, while inhibiting concerted integration [158, 159]. Concerted integration requires at least the tetrameric form of IN [93, 100, 160], yet the specificity of the tetrameric conformation is paramount to IN activity [100]. IN pre-incubated with LEDGF forms inactive tetramers that are unable to bind vDNA, while IN/vDNA preassembled complexes readily interact with LEDGF [100]. These lines of evidence indicate that LEDGF likely contacts the PIC after SSC formation, as highlighted elsewhere [100, 158], and places its interaction with IN after reverse transcription.

Full length LEDGF promotes IN tetramerization, and similarly its IBD affects catalytic functions through modulating IN oligomerization [88]. Peptides derived from the specific IBD/IN interaction loops were shown to induce a shift of IN oligomerization into a non-functional tetramer [88, 161]. These peptides disrupted both 3'-P and ST catalytic functions, and inhibited viral replication at the integration stage without inhibiting reverse transcription [161]. This highlighted IN's I167-K178 sequence as an allosteric hotspot, and small molecules that could engage IN at this site could affect the IN/LEDGF interaction as well as potentially modulating IN multimerization.

Early design of small molecule inhibitors were initially targeted towards the disruption of IN/LEDGF [162], and some displayed an additional disruption of ST catalytic functions [163-166]. Nonetheless, there was increasing evidence suggesting that these molecules modulated IN oligomerization and resulted in stalling viral replication at stages other than integration [140, 167-170]. The best studied ALLINIs to date are the IN/LEDGF interaction disruptors comprising the quinoline-based acetic acid derivatives [139, 144]. The common biochemical

properties of these compounds was their non-selective inhibition of IN's 3'-P and ST catalytic activity's and the disruption of the IN/vDNA complex formation [167, 168, 171]. The LEDGIN 6 compound (Table 1.1), hereafter referred to simply as LEDGIN, had antiviral activity and selected for an A128T resistance mutation [162]. Although these compounds were designed for sites other than the active site, they disrupted viral replication at the integration step and caused an increase in 2-LTR circle formation [162], implying a similar mechanism of inhibition compared to the over expression of the transdominant LEDGF IBD [152]. Follow up studies indicated that additional late steps of the life cycle, namely viral maturation, were affected by these compounds [168, 172]. A number of similar compounds, with notable examples are: BI-D [140, 142, 173], Mut101 [169] and GS-A/B [170, 171]. These likewise exhibited similar late stage viral effects.

The ALLINI's MOA revealed that, similarly to IBD based peptides, they induce aberrant higher order multimerization of IN in biochemical assays as well as within virions, which leads to various deficiencies of IN in activity assays as well as during the HIV life cycle [140, 168-170, 172]. ALLINIs were shown to induce conformational changes at the IN dimeric interface that disrupted the IN/vDNA complexation [171]. Furthermore, ALLINIs modulate the multimerization of IN, prevents the formation of a functional SSC and also inhibits the SSC interaction with LEDGF [167]. Co-crystallization revealed that these ALLINIs directly interact with a CCD subunit at E170 and H171 [162], and additionally in some cases with T174 [140, 171], through hydrogen bonding via its carboxylic group. The quinoline is involved in hydrophobic interactions with the opposite subunit which forms a pocket around A128 [140, 162, 171]. The ALLINI interaction overlaps the critical binding site of LEDGF's D344 and likely explains the disruption of the IN/LEDGF interaction. Resistance mutations either affect ALLINI binding, or may reduce the compounds effects. For instance, the IN_{H171T} mutation was shown to disrupt BI-D binding while not affecting the IN/LEDGF interaction [173], thereby resulting in resistance to the disruption of the IN/LEDGF interaction as well as reduced aberrant multimerization [173]. On the other hand, IN_{A128T} selected by ALLINIs is a primary mutation and confers significant resistance against viral inhibition by the majority of ALLINIs [142, 162, 174]. The A128 forms part of the LEDGF IBD binding pocket and is found on the subunit opposite the 167-178 sequence [110]. In contrast to IN_{H171T}, the IN_{A128T} mechanism does not

confer resistance to the disruption of the IN/LEDGF interaction, but rather only reduces the ability of ALLINIs to induce aberrant multimerization [142]. The substitution resulted in an amino acid side chain shift, where the threonine was positioned away from the ALLINI's quinoline, thereby reducing the compounds stabilization effect at the CCD-CCD interface [142]. This highlighted that resistance selection is based on reducing the aberrant multimerization caused by ALLINIs, and this is likely their primary MOA. ALLINI binding modes, coupled with the oligomer modulating effects of LEDGF and LEDGF IBD, indicate that IN's CCD dimer interface significantly regulates multimerization, and binding to this site shifts IN towards higher order oligomers (Figure 1.10).

Although these ALLINIs appear to bind at the same site, resistance selection and conformational effects caused by ALLINIs are not always uniform, even among the same class of compounds. Mut101 and BI-D (Table 1.1) was shown to induced multimerization in both IN_{A128T} and IN_{WT} with little distinction [169], and BI-D has been mentioned as non- selective for IN_{A128T} resistance [173]. Differences in conformational changes resulting from ALLINI binding have also been reported. Minor side chain changes were noticed for co-crystal structures of LEDGIN and IN at residues Q168 and Q95 [162]. The similar pharmacophore, Mut101, caused side chain changes of E96 and K173 and a significant interaction between the two residues occured [169]. This likely explains the stabilization of the dimers [169] and also why the IN_{A128T} is ineffective against Mut101. Further conformational differences were noticed between two related compounds, BI-D [179] and ALLINI-2 (Table 1.1) [180]. Hydrogen exchange mass spectrophotometry indicated that ALLINI-2 caused significant changes at the CTD away from its primary CCD binding site [180]. This is achieved either through allosteric conformational effects or by directly contacting the CTD, however, an observation of direct CTD interaction was not made [180]. In addition to the CCD and CTD changes, BI-D was shown to protect the CCD-CTD linker region between residues 203-223 [179], and this change was not noticed with ALLINI-2. This linker region has been suggested to be involved in intermolecular interactions between core and peripheral subunits of the intasome [98], and could be significant for multimerization. These differences indicate that although these quinoline-based acetic acid derivatives have a generally related mechanism, they also exhibit differences that may relate to alternate resistance profiles.

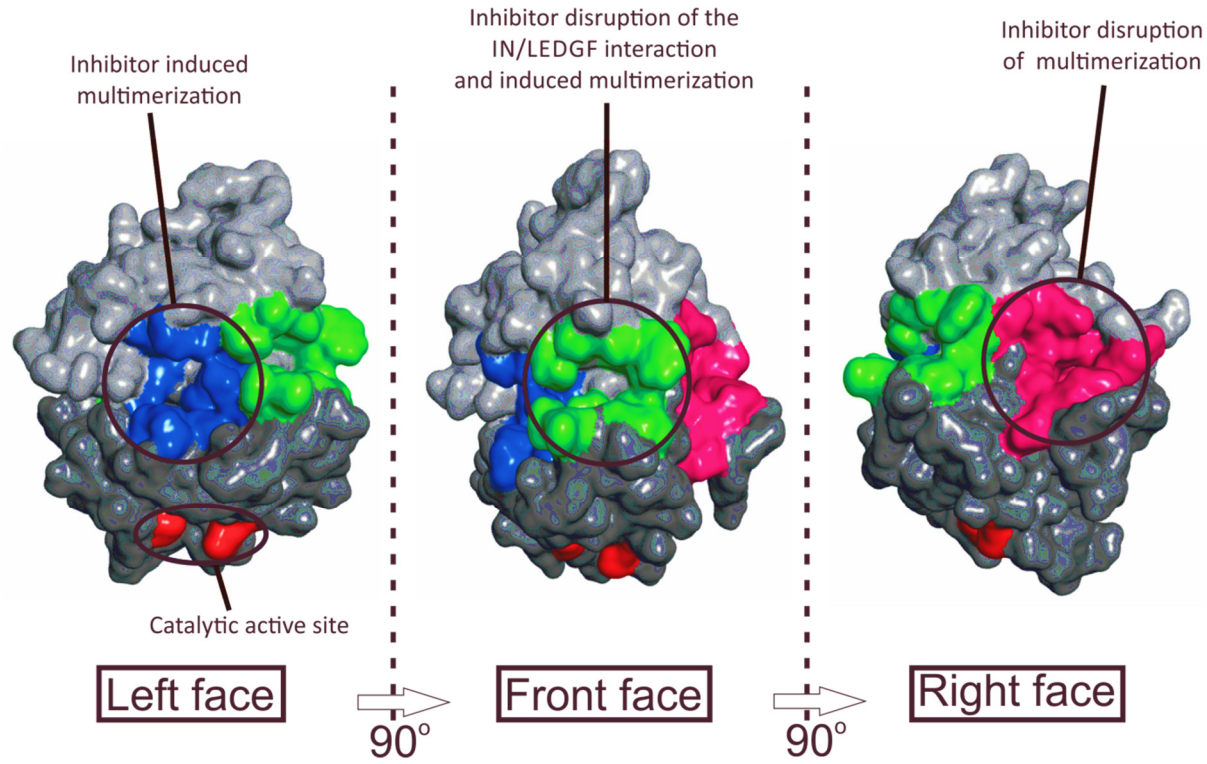


Figure 1.10: Compound binding sites responsible for the modulation of IN multimerization. HIV IN CCD dimer structure, 1QS4 [73], was obtained from the protein data base, and its surface was added in Discovery Studio version 4.0 (Accelrys Inc) with two shades of grey representing each CCD monomer. The sucrose site (blue) is represented on the left face of the orientation shown, and compounds binding at this site induce multimerization [175]. Most compounds and peptides that bind to the LEDGF binding site (green) have been reported to induce multimerization [139, 140, 161, 162, 169] as well as inhibit the IN/LEDGF interaction. IN interfacial peptides and compounds that bind near the F185 region of IN (magenta) have been shown to disrupt the multimerization of IN [176-178]. The active site DDE residues are shown (red).

ALLINIs ultimately cause disruptions in integration, irregular core morphogenesis, un-coating and reverse transcription deficiencies during the HIV life cycle [140, 168-170, 172]. As mentioned, ALLINIs were initially developed to affect the early life cycle phases and disrupt the IN/LEDGF interaction, yet, LEDGF was shown to out compete the binding of ALLINIs on the SSC and was able to reverse the multimerization defect during early viral replication, reducing the ALLINIs potency during the early stage of the viral life cycle [179]. Instead, the observation that ALLINIs are more effective if added to virus producer cells as opposed to the target cells, indicated that effects on late stage viral replication determine ALLINI potency [140, 169, 171]. LEDGF is packaged into virions during assembly by direct interaction with the Gag-Pol and is subsequently digested by PR [181] and contrary to early stage ALLINI antagonism [179], LEDGF's recruitment into the virion did not appear to affect ALLINI potency [181]. Nevertheless, this indicates that an ALLINI binding site is available on the Gag-Pol polyprotein. Late stage viral replication defects caused by ALLINIs are evident by the characteristic eccentric core morphology noticed in electron micrographs of virus produced in the presence of ALLINIs (Figure 1.11), accompanied by aberrant IN multimerization inside virions [140, 141, 170, 182], and the ALLINIs appear to selectively cause eccentric cores while showing similar levels of immature virions as compared to wildtype [141]. This phenotype is evocative of certain IN Class II mutations such as V165I, C43L and Q137R [63, 67, 140, 141, 183]. It has been shown recently that IN binding to the viral RNA genome during maturation is a vital requirement, and offers an explanation for the eccentric core morphology (Figure 1.11) [184]. Lysines in the IN's CTD were identified as responsible for the IN/vRNA interaction, and these include K264, K266, K273 [184]. Interestingly, the pleiotropic importance of the K264 and K266 have been suggested previously [62]. These ALLINI effects then result in early reverse transcription deficiencies during the subsequent round of viral infection [184].

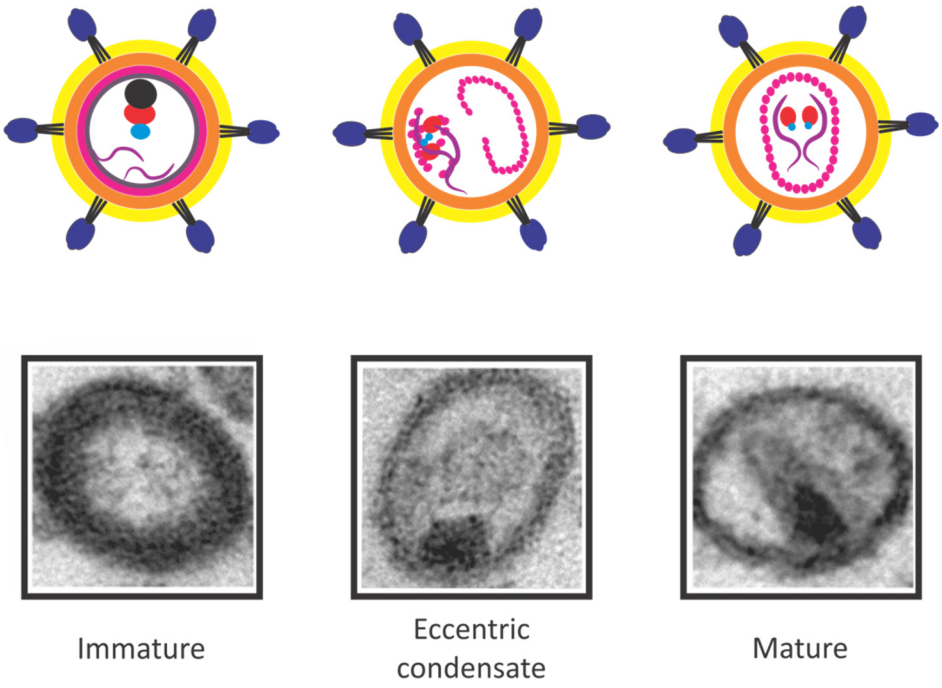


Figure 1.11: Various forms of the HIV virion as assessed by electron microscopy (adapted from [141]). Schematics of each type of form is represented above the electron micrographs and can be compared in other figures.

A separate IN allosteric site, adjacent to the LEDGF binding site, was identified by its interaction with sucrose (Figure 1.6) [111]. Small molecules interacting with this site have been identified [175, 185]. The tetra acetylated chicoric acid derivative (Table 1.1), hereafter referred to as ACA-1 (allosteric chicoric acid analogue compound 1), selectively acetylated IN's K173 [175, 186]. ACA-1 neither inhibited IN/vDNA complexation or the IN/LEDGF interaction, yet stabilized IN subunits and induced multimerization [175]. Modelling suggested a binding pocket adjacent to the LEDGF binding site involving residues E87, E96, Y99, K103 and K173 [175] which overlaps the sucrose binding site [111]. Interestingly, Mut101 binding at the LEDGF site, induced changes at the sucrose binding site [169]. Furthermore, mutagenesis of residues at this site negatively affected IN oligomerization [112]. This indicates that the sucrose binding site regulates IN multimerization, and compounds binding in this site leads to aberrant higher order IN oligomers (Figure 1.10). To support this view, reference can be made to a new class of multimerization selective ALLINIs [187]. These compounds are based on the quinoline acetic acid derivatives, but a pyridine ring replaces the quinoline and is either connected to a benzene or to a benzimidazole ring [187]. The changes resulted in compounds that stretched towards the sucrose binding site and formed interactions with T125, while overlapping the LEDGF binding site [187]. This resulted in at least one compound, KF116 (Table 1.1), that selectively caused aberrant IN multimerization with a reduced ability to inhibit the IN/LEDGF interaction, resulting in a viral inhibition profile that primarily affected late stage viral morphogenesis [187]. KF116 selected for the unusual V165I substitution, which is important for the IN/LEDGF interaction [87], yet the residue is not directly involved with either ALLINI or LEDGF binding, and is rather believed to cause conformational changes that disrupt the LEDGF binding site [154]. This suggests that the KF116 has a higher genetic barrier that selects for mutations distal to its binding site [188]. Thus, compounds binding at or near the sucrose binding site appear to selectively disrupt late stage viral replication by modulating multimerization. They may also have a significant barrier to resistance, which has been implied by mutagenesis studies [112].

1.5.2. Pleiotropic integrase mutations and integrase inhibition imply an allosteric mechanism involving un-coating and nuclear localization

If considering a karyophilic group as being responsible for the nuclear localization signal (NLS), it would suggest that there is a diverse and redundant mechanism for HIV nuclear localization [42]. Prima facie evidence connected IN as a possible viral NLS, as cell culture expression products migrated to the nucleus [116, 189, 190]. Certain regions on IN were identified as possible NLSs, such as residues I161-K173 [191], K186-K188 [192, 193], K211-K219 [192]. Nevertheless, describing IN's role in viral nuclear localization has been controversial, and many deletion studies indicate that a specific IN NLS is unnecessary [63, 65, 69, 194]. Furthermore, by constructing HIV chimeras with MLV IN, as well as removing the other viral NLS candidates such as matrix protein, Vpr or the central polypurine tract, did not affect the viral ability to infect non-dividing cells [50], indicating that none of the previously identified NLS pathways proposed are responsible for HIV nuclear localization. Interestingly, IN sites suggested as NLSs have been more recently identified as allosterically important and may indicate IN's structural significance in nuclear localization (see above section 1.5.1).

Virion morphology defects that present with eccentric condensates are noticed with pleiotropic IN mutations [141, 183]. Although this is sufficiently explained by the disruption of IN/vRNA by certain mutations such as IN_{K264A} [184], there is a significant proportion of IN mutant virus that presents with a non-conical core, indicating instability in CS oligomers which is possibly caused by IN [141]. Although IN has not been shown to directly interact with CypA, IN_{C130S} resulted in a lower incorporation of CypA into the virions [195]. The mutation was associated with a disruption in IN multimerization [116] and caused a significant structural alteration of IN [196]. This seems amino acid specific, as similar effects were not associated with IN_{C130A} [65, 196]. IN_{C130S} did not affect viral particle processing [68], and had a similar phenotype compared to IN_{V165I} [65, 69], IN_{K264A} [184, 197] and IN_{C43L} [198]. Although IN_{C130S} disrupted the packaging of CypA into virions, it is rather the target cell CypA that is utilised during infection [199]. Nevertheless, the IN_{C130S} effect likely persists in the target cell and subsequently presented an unstable core phenotype with premature disassembly in the cytosol, and an ultimate deficit in reverse transcription [195]. Although no direct evidence is yet available, this suggests that un-coating, and likely CS oligomerization, is sensitive to IN

structural organization. The exact MOA of IN influences on maturation is unclear, as IN has not been shown to interact with purified CS [200]. However IN may interact rather with the hexameric CS protein arranged in its lattice configuration, and the determination of this possible IN/CS interaction is a focus of future research proposed by Chow *et al*. [201].

1.5.3. Integrase pleiotropic phenotypes reveal direct allosteric involvement in viral release and reverse transcription

The late stage construction of the mature infectious virion is a complex multistep process that involves the transport of the Gag/Pr55 and Gag-Pol/Pr160 polyproteins to the cell membrane, assembly of the immature virion, budding and finally maturation of the viral core [44]. Normally, for cleavage to commence PR needs to form catalytically active dimers [202], which initially involves the oligomerization of Gag-Pol [203]. Proteolytic processing completes after budding, forming the mature virion core (Figure 1.11) [44]. The malfunction of virion and maturation due to pleiotropic IN mutations has been linked to the premature activation of PR in the context of the Gag-Pol polyprotein [44, 63, 204]. Evidence suggests that IN and RT plays a role in PR processing of Gag-Pol. Deletions of only IN from Gag-Pol caused a reduction of the virion incorporation of IN/p32 and RT/p66-p51 compared to CS/p24 [67]. A truncation of Gag-Pol RT-IN regions restored normal virion particle formation [204], suggesting that IN and RT act as opposing regulators of PR processing in the context of the complete Gag-Pol, and in the absence of IN there is a premature activation of PR. Further characterization of Gag-Pol elements revealed that the p51/p66 of RT was important for steady state PR processing, and that IN mainly influenced the RT heterodimer formation and also virion assembly and particle production [205]. The IN NTD and CTD appear significant in this relationship, where the IN NTD H12Q and H16Y mutants presented with immature virions [198], while the residues between 260-268 had defective processing and viral release [62]. The significance of this RT-IN relationship was, however, brought into question, and although the CTD mutation E246K in IN appear to be significant in regulating PR processing [66], the observed phenotype may be as a result of excessive RNA splicing introduced by the mutations [206]. Another line of enquiry indicated that, although the phenotype of HHCC mutations at H12 and H16 appear to have defects in viral assembly, they can be rescued by the trans-complementation of a Vpr-IN fusion construct [207-209], implying that IN independent of Gag-Pol has a distinct role in

reverse transcription. This evidence however indicates that in the context of Gag-Pol, IN/p32 plays a role in the regulation of PR processing.

The influence of IN in the context of the Gag-Pol on proteolytic processing may, in part, be explained by its interaction with integrase interactor 1 (INI1). There is significant sequestration of host factors into the HIV virion during assembly [210]. Of these, and specifically for HIV-1, INI1 directly binds to the IN portion of the Gag-Pol polyprotein, and this interaction is required for efficient viral assembly and particle release [211, 212]. A minimal binding peptide derived from the INI1 integrase binding domain (IBD), the INI1 transdominant mutant (S6), allosterically inhibited viral particle production through its interaction with IN [211]. The S6 peptide seems to primarily affect the early stages of viral assembly before budding, and has been suggested in the control of Gag-Pol shuttling to the cell membrane via its interaction with IN [213]. The endogenous expression of S6 peptide ultimately lead to deficiencies in Gag accumulation at the cell membrane, deficiencies in PR processing and viral release [213], a similar profile experienced with IN deletion mutants [67, 204]. Interestingly, an IN H12Y mutant presents with immature virions in electron microscopy analysis, and has anomalous PR processing [183], which is similarly seen for H12N [67]. These pleiotropic mutations disrupted the interaction between IN and S6 [211]. This evidence suggests that IN in the context of the Gag-Pol is a significant factor in the regulation of PR processing and viral assembly, and likely influences this process via interactions with both INI1 and RT.

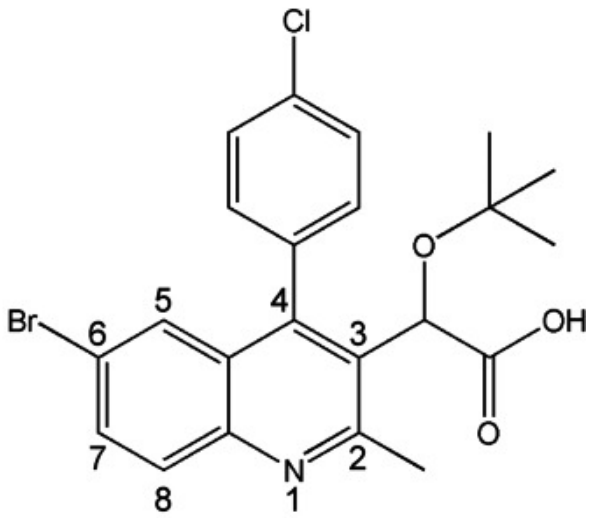
Most pleiotropic IN mutations have been implicated in the disruption of reverse transcription, however, these are overlapping phenotypes and usually present in conjunction with viral release, assembly, maturation, un-coating, nuclear localization and integration defects [63, 65, 67-70]. Nonetheless, certain mutations show greater deficiencies at specific points in the viral life cycle. The trans-complementation of Vpr-IN fusion proteins with mutations H12A, H16A, F185A and a 22 residue CTD truncation indicated that there was a specific interaction between IN and RT which was necessary for viral cDNA synthesis [209]. This was further classified, and this IN/RT interaction was confirmed as a requirement during viral replication [214, 215]. IN bound to the individual p66 and p51 monomeric as well as to the p66/p51

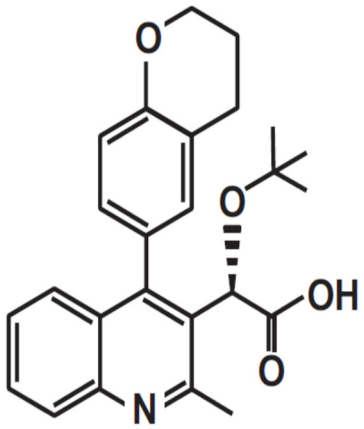
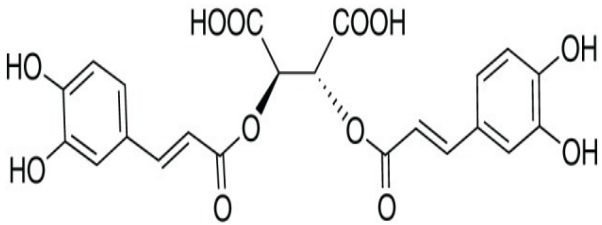
heterodimeric forms of RT in biochemical assays, [214]. IN multimerization was not necessary for the interaction [214], and the specific amino acid region 220-270 of IN CTD was identified as the sufficient requirement for the interaction [68, 214]. This, along with results indicating the importance of the HHCC and F185A mutations for viral cDNA synthesis [209], may indicate important sites on IN required for the formation of a RTC. The RTC contains multiple proteins in addition to IN and RT, and host factors that interact with IN such as INI1, Sin3a-complex and Gemin2 [216], form the RTC and these are required for reverse transcription. These host factors may facilitate IN/RT interactions in vivo by interacting with IN CCD and HHCC, as previously discussed [215].

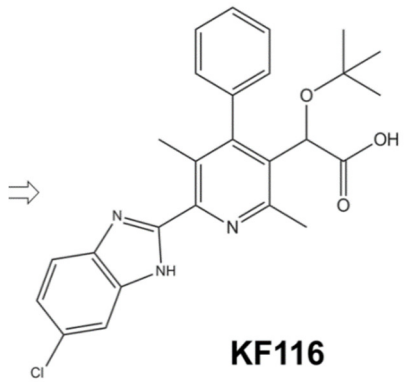
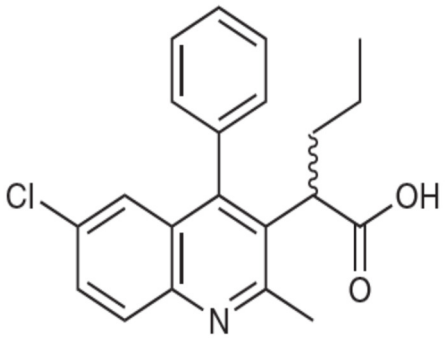
Small molecule ALLINIs that possibly disrupt the IN/RT interaction have been identified, and include polysulfonate dendrimers [217], pyranodipirimidines [218, 219] and styrylquinolines [220] (Table 1.1). The styrylquinoline, FZ41, and the pyranodipirimidine, V-165, were shown to disrupt the IN interaction with viral DNA [219, 221], and are more commonly referred to as integrase DNA binding inhibitors (INBIs). These compounds indicated a reverse transcription block during viral infection. Considering FZ41, this compound affected reverse transcription during first round viral infection without inhibiting RT directly and selected for a double V165I/V249I and single C280Y resistance mutations in IN [220]. The V249I residue mutation is important when considering the IN/RT interaction [215], while the effects of C280Y involves IN multimerization [84], and indicates that the molecules MOA is likely related to disruption of the RT/IN interaction and/or multimerization. The V165I is situated at the CCD dimer interface, it affects virion morphology and is significant for the interaction of IN with LEDGF [87, 140], and may also be important for IN multimerization. The V-165 compound (not to be confused with V165I), also inhibited RT during first round viral infection [219], and selected for T206S/S230N IN mutations [218], also suggesting a disruption of the IN/RT interaction. Interestingly, ARV treated patients, that were experienced for non-nucleoside RT inhibitors and were naïve for clinical INSTIs, presented with mutations in IN at residues M154I and V165I and T206S [222], indicating their significance in the IN/RT interaction.

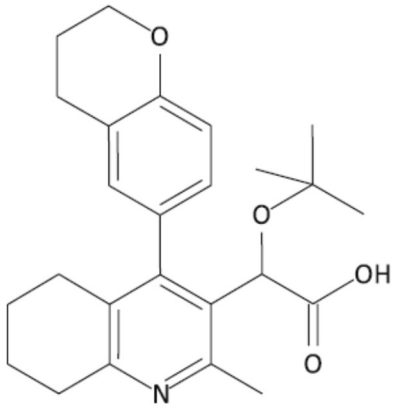
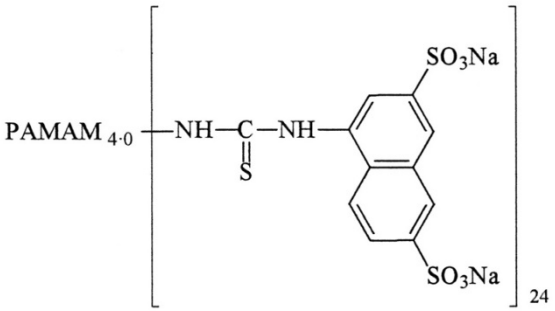
Although the FZ41 selected for mutations at the IN CTD, a model of the HIV IN dimer was designed based on the PVF arrangement of domains and the FZ41 was docked in a binding pocket comprising residues Y83, Y194, G193, I191, K188, E157, K186 and H183 [223], and this was similarly noticed for styrylquinolines docked into a CCD crystal structure [224]. Non-related compounds, that were specifically designed to interact with this binding site (Figure 1.10), were shown to disrupt IN dimer formation [178] and function in an alternative manner to LEDGF site and sucrose site binders. This binding site has also been described for interfacial IN peptides, which have been conceptualised to disrupt the oligomerization at the CCD dimeric interface [176, 177]. The α-helices (α) mainly contribute to the dimeric interface of the IN CCD, and a number of peptides that disrupted the IN dimer were derived from these [225-227]. Peptides designed for the IN CCD α1 (between residues 93-107) and α5 (between residues 167-187), called INH1 and INH5, respectively, were shown to disrupt IN oligomerization and affect catalytic activities [226, 228]. The IN α3 peptide, between residues 123-133, had been reported to have no inhibitory effect on a F185K/C280S soluble IN mutation in biochemical assays and no disruption of the dimer [229]. However, a separate study using wildtype IN, indicated inhibition by peptides derived from α3, but did not address dimerization [227]. If considering the homodimeric form of the CCD, the α3 of one CCD subunit lies adjacent to the α5 of the opposite subunit [73, 82, 83]. A study into the interaction between α3 and α5 revealed the importance of π-π stacking interactions, between α5 residues M178, F181 and F185 of one subunit and the α3 W132 of the other subunit, in viral infection and likely impacted multimerization [230]. This region is undoubtedly important in the formation of multimers as mutations at C130 [116], W131 [79], F185 [84], K186-K188 [101], have been implicated in affecting solubility of IN in biochemical assays as well as inside virions, and interfacial peptides binding to this region appear to disrupt IN dimerization.

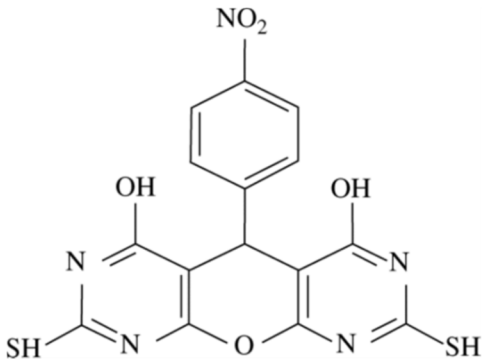
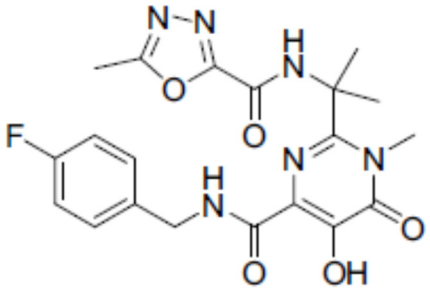
Table 1.1: A list of compounds mentioned in the text. The table contains the compound name as stated in the references, chemical structure, a description of its disruption pathway and the reference where the information was obtained.

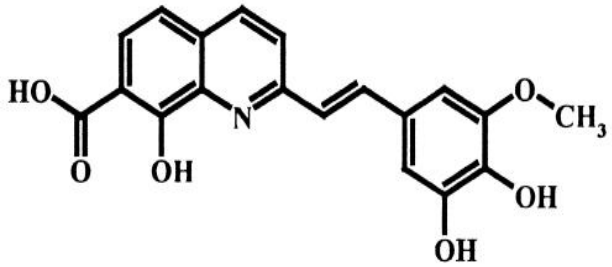
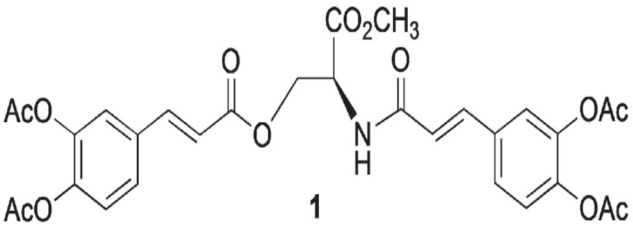
Name	Chemical Structure	Disruption	Reference
ALLINI-2		Non-specific for the 3'-P and ST reactions. IN/LEDGF interaction. Modulates multimerization of IN. Disrupts late and second round HIV infection	[142, 180]

Name	Chemical Structure	Disruption	Reference
BI-D		Non-specific for the 3'-P and ST reactions. IN/LEDGF interaction. Modulates multimerization of IN. Disrupts late and second round HIV infection	[140]
Chicoric acid (CA)		Viral entry, integration Non-specific for 3'-P and ST reactions.	[231-239]

Name	Chemical Structure	Disruption	Reference
KF116	 KF116	Non-specific for the 3'-P and ST reactions. Modulates multimerization of IN. Disrupts late and second round HIV infection	[187]
LEDGIN 6 (LEDGIN)		Non-specific for the 3'-P and ST reactions. IN/LEDGF interaction. Modulates multimerization of IN. Disrupts second round HIV infection	[162, 168]

Name	Chemical Structure	Disruption	Reference
Mut101		Non-specific for the 3'-P and ST reactions. IN/LEDGF interaction. Modulates multimerization of IN. Disrupts late and second round HIV infection	[169]
polysulfonate dendrimers		Viral entry, Reverse transcription, integration. Nonspecific for 3'-P and ST inhibition.	[217]

Name	Chemical Structure	Disruption	Reference
Pyranodipirimidines (V-165)		Viral entry, Reverse transcription, integration. Nonspecific for 3'-P and ST inhibition. IN/vDNA complex disruption	[218, 219]
Raltegravir (RAL)		Specific inhibition of ST during integration	[94, 121, 127]

Name	Chemical Structure	Disruption	Reference
Styrylquinoline (FZ41)		Reverse transcription, nuclear localization, integration. Non-specific for the 3'-P and ST reactions IN/vDNA complex disruption	[220, 221, 223]
tetra acetylated chicoric acid derivative (ACA-1)		Non-specific for the 3'-P and ST reactions. Modulates multimerization of IN.	[175, 235]

1.6. Integrase inhibition by chicoric acid

1.6.1. Integrase active site inhibition

Chicoric acid (CA) (Table 1.1) is a phytochemical found in a variety of natural sources and is described as a tartaric acid of two caffeic acids, or a hydroxycinnamic acid [157, 240]. Despite its natural sources, CA was first identified as an IN inhibitor when synthesised as an analogue of dicaffeoylquinic acid (DCQA) [238], and had since been shown to be a potent inhibitor of IN in biochemical as well as antiviral assays [231-233, 235, 236, 238, 239, 241]. CA and related aromatics were non-active against enzymes that did not require metal cofactors, while marginally effective against metalloenzymes other than IN [242]. This, along with the presence of the Bis-catechol moiety, indicates that CA's mechanism of action is possibly based on metal chelation. CA equally inhibited 3'-P and ST activity, and also inhibited the disintegration reaction, and the inhibitory concentration to reduce enzymatic activity by 50% (IC_{50}) ranged from ≈100nM up to 10μM for CA against IN, as various reports have suggested [231, 235, 238, 241]. This discrepancy has been noticed and commented on elsewhere [243], and is believed to be due to the variance in assay conditions used. The disintegration reaction only requires the CCD domain which CA likewise inhibited, indicating indirectly that CA specifically interacts with the CCD [238]. A resistant HIV mutation, G140S, was selected by CA in cell culture and is found in the IN flexible loop [232], further supporting CA's inhibition of IN through binding at (or close to) the active site. Docking simulations of CA binding to the IN CCD have also supported an active site binding mechanism of IN inhibition [244, 245]. These docking studies, along with mutational analysis [246], have highlighted the interaction of the carboxyl groups of CA and their interaction with K156 and K159 [234, 246], amino acids that are important for the interaction with viral DNA [247]. CA has also shown cross resistance with other ST inhibitors such as the T66I/M154I DKA resistance mutation [233] and the Q148A resistance pathway [234]. These data indicate that CA interacts with the active site for IN inhibition.

1.6.2. Active site binding cannot fully explain the inhibition mechanism of CA in viral cell culture

The antiviral activity of CA indicated varying inhibitory concentrations from 0.4-20µM with favourable cytotoxic selectivity, depending on the viral strain [232, 241], cell lines [236] and/or the antiviral assays used [235, 248]. Although there is evidence indicating CA's MOA involves the IN active site, other lines of enquiry have highlighted inconsistencies with this view. The antiviral activity of CA has been suggested by Pluymers et. al. [241] to primarily involve the blocking of gp120 to CD4, as indicated by the selection of resistant mutations in the variable region of gp120 [241]. No mutations in IN was selected, except for a C56Y, which was regarded as a polymorphism and excluded from the final analyses. Furthermore, HIV that was resistant to other entry inhibitors was also resistant to CA and DCQA, while virus selected by CA was resistant to dextran sulphate and bicyclam derivatives [241]. Time of addition experiments help in determining the stage of the virus life cycle that an inhibitor affects [249], and indicated that CA exerts its effects at a time corresponding to viral entry [241]. Compounds structurally related to CA, caffeoylquinic acid and a galloylquinic acid, were isolated from plant extracts and shown to bind gp120 irreversibly [158]. However, although CA showed evidence of interactions with RT and gp120 [242, 250], it was identified as one of the first selective inhibitors of IN [250] with preference over RT inhibition and binding to gp120 by 100 and 700 fold, respectively [242, 250]. The inhibition of viral entry by CA was confirmed by real time PCR analyses of RT products after viral infection [236], but this study also revealed that CA's MOA is dependent on the concentration of the compound. CA was found immobilised in the cell membrane at greater than 5µM concentrations, showing less than 2% cellular uptake. This is likely due to its abundance of highly reactive oxygen species and its interaction with the fatty-acid bilayer [251]. Real time PCR analysis indicated that CA selectively inhibited IN at concentrations less than 5µM in cell culture, yet behaved as an entry inhibitor at concentrations greater than 5µM [236]. Reinke *et al*. [236] suggest that, once CA gains entry into the cells, it acts as a selective IN inhibitor; while after a barrier concentration is reached in the environment around the cell, CA binds to the cell wall, blocking its own entry. This causes an accumulation of CA in the culture environment, and due to its affinity for gp120 at higher concentrations [250] the selectivity is shifted towards gp120. This relationship of a compound with a dependence on its concentration for a specific MOA has precedence. Polysulfonate dendrimers, which inhibit the viral entry step in TOA assays at concentrations

of 20µg/ml, was shown to inhibit the reverse transcription and integration steps at concentrations higher 500µg/ml [173]. Taking this concentration dependence into account, Pluymers *et al*. [241] used 5µM initial concentration of CA, doubling the concentration every two passages, to select for the mutations over ≈8 passages to a final of 63.2µM of CA. King and Robinson [232] however, selected for the G140S mutation with initial CA concentration of 2µM for several weeks, allowing for peak viral growth before doubling the CA concentration to 4µM, repeating the process before doubling and maintaining in 8µM CA. Pluymers *et al*. had used MT4 cells and the NL4-3 virus for CA induced mutation selection, while for the time of addition the HIV(III_B) strain was used. King and Robinson had used H9 cells and the HIV (NL4-3) strain for CA induced mutation selection, while Reinke *et al*. used HIV (LAI) strain for time of addition. These differences in methodology, especially the ramping of CA concentrations during selection, may have resulted in the disparate results between the two reports. In contrast to polysulfonate dendrimers, this suggests that CA is an IN inhibitor at low concentrations while an entry inhibitor at higher concentrations, and likely depends on the specific chemistry between these compounds. The entry inhibition of CA, although apparent, does not explain the low EC_{50} noticed for CA antiviral activity, under certain conditions [232, 233, 236].

The aromatic pyranodipirimidines, such as the V-165 derivative, inhibited IN/vDNA binding and was more selective for 3'-P than for ST, while TOA experiments indicated a block at reverse transcription and integration [219]. These were later shown to induce mutations in gp160 (gp120 and gp41), IN and RT [218]. These late stage mutations in the gp160, selected by V-165, conferred viral resistance to entry inhibitors. This cross reactivity of apparent IN inhibitors with gp120 may indicate a pleiotropic link, however, this is unlikely as there has been no indication of pleiotropic IN effects on viral entry [63]. Interestingly, the S230N mutation selected by V-165 in the IN CTD has been implicated as a requirement for the IN/RT interaction [68, 214]. The V-165 also selected for the T206S [218], which was a prevalent mutation of IN in patients treated with RT inhibitors while naïve for INSTIs [222]. The inhibition of reverse transcription during first round viral infection, accompanied by RT mutations, likely indicates an allosteric mechanism of action of V-165 that disrupts the IN/RT interaction, similar to Styrylquinolines [220, 221, 223]. This is in contrast to CA, which did not

inhibit reverse transcription during first round viral infection [236], and likely has a different MOA in cell culture. The attributes of V-165 offers credence to the idea that compounds may have a multimodal MOA that affects various points in the viral life cycle, which could likely be the case with CA and is bimodal inhibition of IN and gp120.

During viral replication the 3'-P reaction is achieved in the cytoplasm, likely upon PIC formation, just before nuclear entry. A single functional tetramer, or possibly higher order oligomers, of IN bind to the two LTRs of a single HIV cDNA genome. This IN/DNA complex is referred to as the stable synaptic complex (SSC) and this intermediate state has been duplicated *in vitro,* revealing a highly stable tetrameric IN/DNA complex that is resistant to high ionic strengths as well as detergents. Upon binding to the vDNA ends, the active site undergoes structural changes and it is only after 3'-P that specific INSTIs can recognise its binding site. The processed vDNA represents the integration competent PIC and as such, delivered to the nucleus.

Focusing on IN inhibition, the most prolific group of inhibitors are arguably those of the DKA variety. These are a diverse group with varying MOAs for IN inhibition, and the most successful of these are the INSTIs. A comparison of CA to INSTIs reveals a discordance between the two MOAs. The DKA, L-731,988, was shown to be 70 times more selective for ST than for 3'-P [131]. This is in contrast to CA which is non-selective for the inhibition of IN catalytic functions, and inhibits 3'-P, ST and disintegration at equal concentrations [238, 252]. Despite these differences, both the L-731,988 and CA inhibited viral replication at similar EC_{50} concentrations [236]. The mechanism of ST inhibition requires the formation of the intasome, and after 3'-P and fraying of the viral DNA ends [85], the INSTIs recognise this DNA bound IN active site conformation and they orient towards the metal ions and form a hydrogen bond with Y143. This ultimately causes the displacement of the single 3' adenosine, from the active site inhibiting ST, without disrupting the general IN/vDNA contacts [94]. The selectivity of INSTIs is directly linked to their recognition of the post 3'-P active site of an intasome [81, 253]. This explains the ineffective disruption of IN/vDNA interactions by highly selective INSTIs [252]. Contrary to this, CA inhibited the IN/vDNA contacts non-specifically in a Schiff Base

crosslinking assay that used abasic sites at various points in the DNA substrate. This indicates a global disruption of the IN/vDNA complex and cannot conclusively assume active site binding [252]. These types of crosslinking assays rely on the pre-incubation of the protein with inhibitor, followed by DNA addition, indicating that CA's interaction was with a free IN protein, unbound to DNA.

The binding mode of CA at the unbound active site was proposed to be similar to DKAs, such as 5CITEP [244]. A crystal of the IN CCD, named 1QS4, with 5CITEP bound at the active site was the first structure solved that indicated a binding mechanism of an IN inhibitor [83]. The binding mode of 5CITEP did not displace or chelate the Mg^{2+} ion at the active site and bound to the CCD in the absence of DNA. Similarly, CA's inhibition was not altered by varying the metal cofactor in catalytic assays, suggesting that its inhibition is not dependant on metal chelation [239]. Furthermore, CA was effective if added before and/or after IN/vDNA assembly [239]. The 5CITEP inhibitor had potential hydrogen bonds with the amino acid residues Q148, E152, N155, T66, K159, and K156. A similar binding mode of CA was proposed by modelling [244, 245] and due to the extra bulk of CA, it is proposed to take up the adjacent site involving the residues D116, C65 and H67, and mutational analysis appeared to support this binding mode [233, 234]. Despite these proposed similarities of binding between the compounds, CA had antiviral activity, while 5CITEP is inactive in cell culture [197].

The proposed binding mode of CA and 5CITEP assumes a free active site devoid of vDNA. This may be significant in biochemical assays, where isolated IN could conceivably have the required active site conformation to accommodate this interaction, however, conformational states of IN during the viral life cycle are variable and may not present a free active site conformation. The major criticism of CA's relevance in the inhibition of the ST reaction, during viral replication, is evident by the inability of CA to disrupt activity of isolated PICs [242, 254]. These PICs are competent for concerted integration into a target DNA strand, and it has been suggested that due to the pre-assembled conformation of these complexes, CA's recognition site is blocked [254]. This is in contrast to INSTIs, such as L-731,988, which have shown similar inhibition of the recombinant IN protein as well as isolated PICs [131]. These mechanistic

differences of CA compared to INSTIs suggest that the ST inhibition by CA is not physiologically relevant.

The *in vitro* PIC assays represent a static integration complex that is competent for concerted integration, and are useful in identifying INSTIs. However, this complex is highly dynamic during viral replication and undergoes significant changes after viral entry and subsequent nuclear localization. Allosteric inhibitors, that rely on structural conformations of IN and/or specific protein-protein interactions for binding site recognition, may then present with an apparent block of IN activity in cell culture, but not in PIC assays. It was shown that CA did not disrupt reverse transcription but led to an increase in the rate of 2-LTR circle formation during first round viral infection, specifically at low concentrations of compound [236]. The increase of 2-LTR circles represents the inhibition of viral replication after 3'-P and nuclear entry, and is caused by deficiencies of IN catalytic function [131, 255]. This has been used to determine the integration inhibition stage of antiviral compounds [140, 170, 236]. This apparent inhibition of IN catalytic functions during viral replication, despite CA's lack of PIC inhibition, indicate that CA's inhibition of HIV relies either on the specific conformations of IN or disruption of its interaction with host factors. The ALLINI compounds that inhibit the IN/LEDGF interaction were similarly shown to have no effect on reverse transcription, while increasing 2-LTR circle formation during viral infection [140, 170], however, less than that shown for INSTIs [172]. This was similar for CA, which was less effective at increasing 2-LTR circles compared to L-731,988 [236]. LEDGF was shown to outcompete ALLINI binding at the IN-dimer interface, and reduced the efficacy of these compounds during early stage viral replication [179]. Isolated PICs were shown to contain LEDGF [157], if assuming that CAs MOA relies on the IN/LEDGF interaction similar to other ALLINIs, this would explain its inefficacy in PIC assays. Another striking similarity of ALLINIs to that of CA, is that they are non-selective for 3'-P and ST and disrupt the IN/vDNA interaction in biochemical assays [167, 168, 171].

1.7. Hypothesis

CA active site binding has been extensively explored, and yet there is ambiguity in its antiviral activity of HIV. This is noticed in its inability to inhibit PICs [242, 254] while still causing a

disruption of integration during the viral lifecycle [236]. From this evidence, it is not clear whether CA affects the active site or a distal allosteric site during viral inhibition. ALLINI's that disrupt the IN/LEDGF interaction also inhibit integration during the viral lifecycle [140, 162, 172]. We thus hypothesise that CA has an allosteric mechanism for the inhibition of IN by binding to the IN/LEDGF interaction site.

The aims and objectives of this study are:

- To confirm a mechanistic difference between CA and RAL in the inhibition of IN. Both of these compounds have antiviral effects that appear to target the integration process. Our objective is to show a distinction between CA and exclusive active site binders such as RAL.
- To determine if CA disrupts the IN/LEDGF interaction. This would provide supporting evidence for CA binding at an allosteric site
- To determine if CA modulates the multimerization of IN. This would further support an allosteric inhibitory mechanism
- To determine the allosteric binding site and the interacting amino acids.

Chapter 2: Materials and methods

2.1. National Institutes of Health AIDS Research and Reference Reagent Program Reagents used in the study

The following reagents were obtained through the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH from: Dr. Robert Craigie: pINSD.His. (Cat# 2957) [77], pINSD.His.Sol from Dr. Robert Craigie (Cat# 2958) [84], pFT-1-LEDGF from Dr. Alan Engelman (Cat#11396) [256], pNL4-3 from Dr. Malcolm Martin (Cat# 114). Anti-HIV-1 HXB2 IN 1-16 aa Polyclonal from Dr. Duane P. Grandgenett (Cat# 756) [257].

2.2. Various inhibitor compounds used in the study

CA was purchased from Sigma with catalogue number C7243 (Sigma-Aldrich MO, USA). The LEDGIN compound (3-Quinolineacetic acid, 6-chloro-2-methyl-4-phenyl-a-propyl-) was the same as compound 6 developed by Christ *et al*. [162] and purchased from MedchemExpress with catalogue number HY-10522 (MedchemExpress NJ, USA). A tetra-acetylated chicoric acid derivative, with a single methyl group replacing the carboxyl, was the first described ALLINI of IN [175], referred to in this study as ACA-1, and was a kind synthesis gift from Thompo Jason Rashamuse, an in house chemist. Mut101, developed by Le Rouzic *et al.* [169], was a kind gift from Professor Richard Benarous from the Mutabilis institute in France. RAL was extracted from ISENTRESS 400mg tablets (Merck Sharp & Dohme, JHB, SA) by dissolving the tablets in purified water and thereafter removing the aggregate by centrifugation. The water soluble RAL was further purified by reverse phase HPLC. The final product was dried and the powder was weighed and reconstituted before use.

2.3. Recombinant IN vector modifications and synthesis

Recombinant plasmids expressing wild type and soluble IN (both with a 6X His tag) were used to introduce relevant resistance mutations by site directed mutagenesis, as well as subcloning of a wild type IN (FLAG tag) and truncated CCD (IN_{CCD}; 6x His tag) for protein purification and subsequent use in biochemical assays.

2.3.1. Resistance mutations at the active site and at the allosteric site was inserted into the IN coding sequence using site directed mutagenesis

Two plasmids, pINSD.His.sol and pINSD.His were used as templates for mutagenesis with the QuikChange Lightning Site Directed Mutagenesis kit (Agilent Technologies, CA, USA). These plasmids contain the full length, 288 amino acid, IN coding sequence. The pINSD.His.sol contains the F185K/C280S mutations in the IN sequence and increase the solubility of IN. The pINSD.His contains the wild type IN sequence. Both of these plasmids allow for the expression of a 6X His tagged IN protein. RAL resistant mutations were inserted into the pINSD.His.sol and pINSD.His backbone, and include Q148H and N155H with associated double mutations G140S/Q148H and E92Q/N155H, respectively. The A128T mutation was only introduced into pINSD.His as the allosteric mutant resistant to multimerization. Manufacturer's instructions were followed where applicable with details below.

2.3.1.1 ***Mutagenesis***

Primers (listed in Table 2.1) were designed with a single nucleic acid change to transcribe the appropriate codon. Primers, at 125ng final, were added to a 50µL reaction mixture [5µL 10X buffer, 1µL dNTP mix, 1.5µL QuikSolution, dH_2O to final volume and 1µL polymerase as per standard protocol] with 100ng final of either pINSD.His.sol or pINSD.His as template. Reactions were placed on a Mastercycler Gradient (Eppendorf, HH, DE) at 95^{o}C for one cycle; immediately followed by an 18X cycle at 95^{o}C for 20sec, 60^{o}C for 10sec and 68^{o}C for 3min; and a final single elongation cycle for 5 minutes at 68^{o}C. Restriction enzyme specific for non-nicked DNA, *Dpn*I, was added at 2µL to each reaction, and parent plasmids were digested and removed at 37^{o}C for 5min. Agarose (Sigma-Aldrich, MO, USA) was mixed at 1% w/v with Tris-borate buffer [89mM Tris base final pH 7.6 (Sigma-Aldrich, MO, USA), 89mM Boric acid (Sigma-Aldrich, MO, USA), 2mM EDTA (Sigma-Aldrich, MO, USA)]. The mixture was heated to a boil and cast into gel trays with gel combs until hardening. Reactions were visualized by agarose gel electrophoresis for successful amplification and compared to a non-amplified reaction mixture detected using SYBR Gold (Thermo Scientific, MA, USA) gel stain. Repetitions

performed where necessary with modification of the melting temperature by at most 3°C below 60°C.

Table 2.1: List of primers used in the study

Experimental purpose	Primer Name	Sequence 5'-3'
ELISA Assay	Donor top	*Biotin-ACCCTTTTAGTCAGTGTGGAAAATCTCTAGCA
	Donor bottom	ACTGCTAGAGATTTTCCACACTGACTAAAAG
	Target top	TGACCAAGGGCTAATTCACT-FITC*
	Target Bottom	AGTGAATTAGCCCTTGGTCA-FITC*
Mutagenesis[a]	A128T Fwd	CCAGTACTACAGTTAAGACCGCCTGTTGGTG
	A128T Rev	CACCAACAGGCGGTCTTAACTGTAGTACTGG
	Q148H Fwd	CCTACAATCCCCAAAGTCACGGGGTAATAGAATC
	Q148H Rev	GATTCTATTACCCCGTGACTTTGGGGATTGTAGG
	N155H Fwd	GGGGTAATAGAATCTATGCATAAAGAATTAAAGAAA
	N155H Rev	TTTCTTTAATTCTTTATGCATAGATTCTATTACCCC
	E92Q Fwd	GAAGTAATTCCAGCACAGACAGGGCAAGAAACAGC
	E92Q Rev	GCTGTTTCTTGCCCTGTCTGTGCTGGAATTACTTC
	G140S Fwd	GGGATCAAGCAGGAATTTAGCATTCCCTACAATCCC
	G140S Rev	GGGATTGTAGGGAATGCTAAATTCCTGCTTGATCCC
PCR[b]	FLAG IN FWD	GATGACGACAAGCTTTTTTTAGATGGAATAGATAAG
	FLAG IN REV	ATCTATCTAGATCTCTAATCCTCATCCTGTCTACTTGC
	CCD Fwd	GAAGGTCGTCATATGATGCATGGACAAGTAGACTGTAG
	CCD Rev	TTAGCAGCCGGATCCTTCTTTAGTTTGTATGTCTGTTGC

[a] underlines indicate the point mutations

[b] underlined sequence indicate the restriction site

2.3.1.2 ***Bacterial transformation***

Chemically competent *Escherichia (E.) coli* from either E. cloni 10G (Lucigen, WI, USA) or Stellar (Clontech Laboratories, CA, USA) were transformed with mutagenesis reactions. Competent bacteria were incubated with 2µL reaction mixtures on ice for 30min in 15mL conical tubes. This was followed by heat-shock at 37°C for 90sec in a water bath, with subsequent incubation on ice for 2min. The appropriate recovery media was added at 1mL

for E. cloni and 500µL for Stellar bacteria, followed by incubation for one hour at 37°C shaking in an orbital incubator at 250rpm. Transformants were then plated on LB-agar plates containing 100µg/mL ampicillin (AMP) (Sigma-Aldrich MO, USA) and incubated at 37°C overnight. Colonies were randomly selected and grown in 10mL LB-media containing 100µg/mL AMP overnight at 37°C, shaking at 250rpm in 50mL conical tubes.

2.3.1.3 ***Plasmid purification***

Plasmid was purified from overnight cultures using a NucleoSpin Plasmid kit (Macherey-Nagel, DN, DE). Bacteria were harvested from 5mL culture by centrifugation in 50mL conical tubes at 3250 x *g* on a Centrifuge 5810 (Eppendorf, HH,DE) using a swing-bucket rotor for three minutes. Bacteria were reconstituted in 250µL resuspension buffer A1 and transferred to 1.5mL micro-tubes. Thereafter, reconstituted bacteria were lysed in 250µL buffer A2 for 5min at ambient temperature, followed by neutralization in 300µL buffer A3. Soluble plasmid was separated from aggregated genomic DNA by centrifugation at 11000 x *g* for 5min on a MIKRO 200 (Hettich, TUT, DE) centrifuge. Supernatants were collected and bound to silica NucleoSpin columns by centrifugation at 11000 x *g* for one minute. The column was washed with 500µL AW buffer and then 600µL A4 buffer containing ethanol (Sigma-Aldrich, MO, USA). NucleoSpin columns were dried by centrifugation at 11000 x *g* for two minutes. Plasmids were eluted with Primer Resuspension Buffer [10mM Tris-HCl pH 7.2 (Sigma-Aldrich, MO, USA), 100mM NaCl (Merck Chemicals (Pty) LTD, TG, SA)] and quantified on a Nanodrop 2000 (Thermo Scientific, DE, USA). Mutations were confirmed by Sanger sequencing using an external commercial entity (Inqaba Biotech, PTA, SA).

Double mutations were accomplished by repeating the process from section 2.3.1.1-2.3.1.3 with either Q148H or N155H single mutated plasmids in-place of pINSD.His.sol or pINSD.His, with specific primers (Table 2.1), to generate G140S/Q148H and E92Q/N155H, respectively.

2.3.2. Generation of IN_{WT} and IN_{CCD} molecular clones

The AlphaScreen assays (see section 2.6 below) utilize donor and acceptor beads with specific affinities. Nickel-chelating donor beads and Anti-FLAG acceptor beads were available for testing. The pFT-1-LEDGF plasmid, obtained from the NIH reference reagent program, encodes a His-tagged LEDGF fusion protein. It was thus necessary to produce a FLAG-tagged IN protein for AlphaScreen testing. The pINSD.His plasmid was used as a template in the polymerase chain reaction (PCR) to amplify sequences of WT IN (IN_{WT}). PCR primers were designed to generate an insert to be used in the InFusion HD Cloning kit (Clontech Laboratories, CA, USA) with part complementarity to the intended vector and partly specific to the insert (Table 2.1, see primer features). The pT7-FLAG-1 generates a FLAG tagged fusion protein. IN_{WT} was first cloned into a pT7-FLAG-1 (Sigma-Aldrich, MO, USA) vector between *Hind*III and *Bgl*II sites. In consideration to protein purification of a FLAG tagged IN (F-IN), a double tagged strategy was used. Following initial cloning into pT7-FLAG-1, the IN_{WT} with FLAG-tag was inserted into a Novagen pET-16b vector. The pET-16B plasmid generates a 10X His tagged fusion protein which will be ideal for protein purification using nickel affinity chromatography. The pET-16B vector has a Factor Xa cleavage site located between the 10X-His-tag and the insert, and can be removed by serine protease cleavage.

Surface plasmon resonance (SPR) was utilized to obtain the binding kinetics of CA on its IN target site (see section 2.7 below). To gain an insight into the IN domain that CA binds to, the CCD was isolated and used in the SPR analysis. The pINSD.His plasmid was used as a template in the polymerase chain reaction (PCR) to amplify sequences for the truncated CCD (IN_{CCD}). Primers for IN_{CCD} flanked the 50-212 amino acid positions and were designed for its insertion into pET-16B (Merck Millipore, MA, USA) between *Nde*I and *Bam*HI restriction sites. PCR primers were designed to generate an insert to be used in the InFusion HD Cloning kit (Clontech Laboratories, CA, USA) with part complementarity to the intended vector and partly specific to the insert (Table 2.1, see primer features).

2.3.2.1 ***Polymerase chain reaction (PCR)***

PCR of the pINSD.His was performed with pT7-FLAG-1 specific primers (Table 2.1, pT7-IN Fwd/Rev) using the Takara Ex Taq DNA polymerase (Clontech Laboratories, CA, USA). Primers were added at 500nM final concentration each to the 50µL final reaction mixture [5µL 10X-Ex Taq buffer, 4µL 2.5mM final dNTP mix, dH_2O to final volume and finally 0.25µL Ex Tag enzyme as per standard protocol] with 100ng pINSD.His template DNA. Reactions were mixed and placed on the Mastercyler Gradient thermal cycler for 30 cycles at: 98°C for 10sec, 57°C for 30sec and 72°C for one minute. The same protocol was applied to the IN_{CCD} using the appropriate primers and pINSD.His.sol as a template source. Amplification was confirmed by 1% agarose gel electrophoresis, unless otherwise stated.

2.3.2.2 ***Restriction digests***

Double digests of 1µg pT7-FLAG-1, at 20µL volume, was performed with *Hind*III (Thermo Scientific, MA, USA) and *Bgl*II (Promega, WI, USA) in buffer D (Promega, WI, USA) at 37°C overnight. Similarly, double restriction enzyme digests for 1µg of pET-16B, using *Nde*I and *Bam*HI (Thermo Scientific, MA, USA) in FastDigest Buffer (Thermo Scientific, MA, USA), were carried out. Complete digestion was visualised by 1 % agarose gel electrophoresis, and digests were subsequently purified using the ISOLATE PCR and Gel Kit (Bioline, SG) following the recommended PCR Clean-up protocol. Briefly, binding buffer CB was added to digest reactions at 500µL and mixed. Mixtures were bound to silica membranes by centrifugation at 11000 x *g* for two minutes through the kit spin column. The column was washed with 700µL wash buffer A at 11000 x *g* for one minute, followed by a one minute drying and elution with 50µL PRB at the same speed.

2.3.2.3 ***Cloning***

PCR products at 5µL were treated with 2µL of Cloning Enhancer (Clontech Laboratories, CA, USA) at ambient temperature, 37°C and 80°C for 25min at each temperature. InFusion reactions, at 10µL final volume, were set up with 100ng of purified digested vector, 2µL of Cloning Enhancer treated PCR products, 2µL of 5X enzyme premix and dH_2O to final volume. Reactions were incubated at 50°C for 30min followed by transformations in chemically

competent Stellar E. *coli* as per section 2.3.1.2. Plasmids were purified as per section 2.3.1.3 and successful clones were determined by restriction digestion as per section 2.3.2.2 with use of restriction enzymes applicable to the clone, followed by agarose gel electrophoresis.

2.4. Protein purification of IN, IN_{CCD}, LEDGF and associated mutations

The recombinant clones produced in section 2.3 and the pFT-1-LEDGF (section 2.1) were all used to express recombinant proteins in bacterial cells. His-tagged IN proteins[128], IN CCD [111] and LEDGF [256] were purified as previously described using nickel affinity chromatography, with modifications to the methods [128, 256]. Two methods for purifying FLAG-tagged IN protein were attempted, where the purest protein yields were achieved from the double tagged His-FLAG-IN (HF-IN) and nickel affinity chromatography. IN with only FLAG-tag was purified using hydrophobic interaction chromatography (HIC) with a MacroPrep HIC resin (Bio Rad, CA, USA), followed by heparin column chromatography using a HiTrap Heparin HP (GE Healthcare Lifesciences, CA, USA), and this produced target protein with some contaminating bands.

2.4.1. Bacterial induction and lysis

E. cloni express BL21 (DE3) with or without pLyseS chemically competent cells (Thermo Scientific, MA,USA) were transformed with plasmids from the NIH, mutagenesis reactions and from cloning reactions, 2.3.1.2 except that in addition to AMP, 35µg/mL chloramphenicol (CAM) was included in LB-media and LB-agar. Colonies were selected and cultured overnight at 37^{o}C in LB media containing AMP and CAM. Overnight cultures were diluted 100 times in AMP/CAM LB broth and allowed to divide until an optical density (OD) of 0.4-0.6 at 600nM was obtained, and temperature was 37^{o}C and 30^{o}C for IN and LEDGF, respectively. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to 1mM final concentration to induce protein expression. Bacteria were induced for 3hr and thereafter harvested by centrifugation at 3250 x *g* for 30minutes at 4^{o}C and stored at -20^{o}C until further use. Bacterial pellets were defrosted on ice and re-suspended in 50ml Lysis Buffer [20mM HEPES pH 7.4 (Sigma-Aldrich, MO, USA), 1M NaCl, 0.25% CHAPS (Sigma-Aldrich, MO, USA), 10mM $MgCl_2$ (Merck Chemicals, TG, SA), 1mM PMSF (Sigma-Aldrich, MO, USA), 5mM Imidazole (Sigma-Aldrich, MO, USA), 10%

Glycerol (Melford, IPW, UK)] for IN or in Lysis Buffer [25mM Tris-HCl pH 7.4 (Sigma-Aldrich, MO, USA), 1M NaCl, 1mM PMSF, 5mM Imidazole] for LEDGF. Mixtures were briefly homogenised using a Bio-Gen Pro 200 Homogenizer (Pro Scientific, CT, USA), separated into 10mL volumes, followed by a 30 minute incubation on ice. Subsequently the lysates were sonicated at 70 hertz using 6 cycles for 30sec on and 30sec off, and repeated 5 times using a UP50H sonicator (Heilscher, TF, DE).

2.4.2. Nickel affinity chromatography

Nickel affinity chromatography was used to purify all recombinant His tagged proteins, since the Histadine tag has a high affinity for the nickel ions immobilized on the column. Lysates were pipetted into 1.5mL tubes and centrifuged at 15000 x *g* at 4^{o}C in a fixed angle rotor on the 5810r centrifuge. Cleared lysates were loaded onto a HisTrap (GE Healthcare Lifesciences, CA, USA) equilibrated with Binding Buffer [20mM HEPES pH7.4, 1M NaCl, 5mM Imidazole, 10% glycerol] for IN and [25mM Tris-HCl pH 7.4, 1M NaCl, 5mM Imidazole] for LEDGF at 1ml/minute at ambient temperature. The column was washed with Wash Buffer 1 and 2; which are the same as Binding Buffer except with 60mM imidazole and 150mM imidazole, respectively, for both IN and LEDGF buffers. His-tagged proteins were eluted at ambient temperature with a gradient from 150mM to 600mM imidazole. Eluted protein was concentrated using an ultrafiltration unit (Merck Millipore, MA, USA) with a 10kDa ultrafiltration disk (Merck Millipore, MA, USA). Buffer exchange was performed with a PD-10 G25 superdex column (GE Healthcare Lifesciences, CA, USA), and IN proteins were eluted in Storage Buffer [20mM HEPES pH 7.4, 1M NaCl, 10% Glycerol, 1mM DTT (Melford, IPW, UK), 0.1mM EDTA] while LEDGF was eluted in LEDGF Heparin-binding buffer [25mM Tris-HCl pH 7.4, 200mM NaCl, 6mM CHAPS].

Purification of the IN_{CCD} was essentially identical to IN except all purification buffer components, save the imidazole, were half diluted in water as per stated protocol [111]. The reasoning for the half dilution is not expressly stated in the reference, however, it is likely due to the increased solubility of the IN_{CCD} which would correspond to a lower requirement of

NaCl. In personal experience, the IN_{CCD} was more stable at 500mM NaCl, and when purified at 1M NaCl the protein would immediately aggregate in the elution tubes.

2.4.3. Factor Xa cleavage

Proteins cloned into the pET16B vector include the HF-IN and IN_{CCD}. Removal of the His-tag was attempted using Factor Xa protease (Qiagen, H, DE) according to manufacturer's instructions. Proteins were buffer exchanged into Cleavage buffer [20mM Tris-HCl pH 6.5, 50mM NaCl, 1mM $CaCl_2$ (Melford, IPW, UK)], and at least 0.25mg/ml in each reaction and incubations were performed at 23^{o}C for at least 6 hours. Factor Xa was then removed by loading the cleavage product onto a prepacked benzamine column (GE Healthcare Lifesciences, CA, USA) equilibrated with Cleavage buffer, and collecting the flow through. The protein containing flow through was buffer exchanged using PD-10 G25 superdex equilibrated with Storage Buffer [20mM HEPES pH 7.4, 1M NaCl, 10% Glycerol, 1mM DTT (Melford, IPW, UK), 0.1mM EDTA].

2.4.4. Hydrophobic interaction chromatography

The FLAG-tagged IN was purified using HIC, using the properties of hydrophobicity. Firstly, the FLAG-tagged IN was prepared for HIC similarly to Sinha *et al.*, 2002 [259]. Bacterial lysates were prepared as per section 2.4.1 for IN except that the Lysis Buffer contained only 100mM NaCl. At 100 mM NaCl, the IN protein is insoluble and would pellet at high centrifugal speeds. Thus, lysates were cleared at 95 000 *x g* on a Supra 30K high speed centrifuge (Hanil Sciences Industrial, W, SKR) using the A50S-6 rotor for 30 min to pellet IN protein. Supernatants were discarded, and pellets were re-suspended in Lysis Buffer and the centrifugation step was repeated. This low salt washing was performed three times. After washing, the pellet was re-suspended in 1M NaCl Lysis Buffer to increase IN solubility, centrifuged at 95 000 *x g* and the supernatant was retained for hydrophobic interaction chromatography purification.

IN with only a FLAG-tag was purified by HIC using the MacroPrep HIC resin (Bio Rad, CA, USA) similarly to Kessl et. al. 2012 [167] with modifications. Lysate prepared for HIC was

precipitated in HIC-binding Buffer [20mM HEPES pH7.4, 1M ammonium sulphate (Merck Chemicals, TG, SA), 200mM NaCl, 10% glycerol, 6mM CHAPS, 1mM DTT] in a 1:10 dilution of lysate to HIC-binding buffer. Protein was loaded onto a manually packed column at 1ml/min, and after loading, HIC-binding buffer was allowed to continuously run until baseline A_{280} was reached. A gradient of 1M-0M ammonium sulphate over a 200mL volume was used to elute the protein. Fractions containing IN were pooled, concentrated using a 10kDa ultrafiltration disk, and buffer exchanged into IN Heparin-binding buffer [20mM HEPES pH7.4, 200mM NaCl, 10% glycerol, 6mM CHAPS, 1mM DTT].

2.4.5. Heparin column purification

In an attempt to increase the protein purity, LEDGF and FLAG-tag IN was further processed by prepacked heparin column (GE Healthcare Lifesciences, CA, USA) purification in an attempt to increase the protein purity. The column was equilibrated with each protein's respective heparin-binding buffers prior to loading. Proteins were loaded at 1ml/min, followed by a heparin-binding buffer until the A_{280} baseline was achieved. Thereafter proteins were eluted over a 200mL volume with a gradient of 200mM-1M NaCl. Fractions containing target protein were pooled and concentrated. Buffer was exchanged using the PD-10 G25 superdex column, for LEDGF [25mM Tris-HCl pH 7.4, 200mM NaCl, 10% glycerol, 1mM DTT] and for FLAG-tagged IN [20mM HEPES pH7.4, 1M NaCl, 10% glycerol, 1mM DTT].

2.4.6. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and Western blot analysis

Purity of isolated proteins was determined by standard SDS-PAGE and specificity was determined by Western blot analysis. Pre-Cast any kD Mini- PROTEAN (Bio-Rad, CA, USA) or TGX 12.5% Fast Cast acrylamide solutions (Bio-Rad, CA, USA), with Stain Free technology, were used for all SDS-PAGE analyses. Gels were assembled on a Mini-PROTEAN Tetra Cell (Bio-Rad, CA, USA), with 10X Tris/Glycine/SDS (Bio-Rad, CA, USA) added to the tank. A 4x Laemmli Sample Buffer (Bio-Rad, CA, USA), either with or without 350mM final 2-mercaptoethanol (Sigma-Aldrich, MO, USA) was added to protein samples and loaded on the gel with a 10µL pipette. Gels were electrophoresed at 250 volts unless otherwise stated. On completion,

protein migration was viewed on a ChemiDoc MP System (Bio-Rad, CA,USA). Gels were then submerged in Transfer Buffer [25mM Tris base final pH 8.3, 190mM Glycine (Sigma-Aldrich, MO, USA), 20% Methanol (Merck Chemicals, TG, SA)] for 5min. Gels were then transferred to either Nitrocellulose or PVDF iBlot transfer stacks (Thermo Scientific, MA, USA) using the iBlot Dry Blotting System (Thermo Scientific, MA, USA). Membranes were blocked overnight in Tween-Tris Buffered Saline (TTBS) [20mM Tris-HCl pH 7.5, 150mM NaCl, 0.05% Tween 20 (Merck Chemicals, TG, SA)] containing 2.5% BSA (Highveld Biologicals, JHB, SA) or 5% Elite Skim Milk Powder (Clover, RD, SA). Membranes were then rinsed with TTBS and incubated with appropriate primary antibody for at least 2hr at ambient temperature. Thereafter, membranes were washed three times with TTBS, for 15min each wash on a shaker at ambient temperature. Membranes were then incubated with HRP-linked secondary antibody for at least 1hr at ambient temperature, and again similarly washed. Clarity Western ECL Blotting Substrate (Bio-Rad, CA, USA) was added directly to the blots, and finally, luminescent signal was visualised on the ChemiDoc MP System using the Chemi High-Resolution protocol.

2.5. Enzyme linked immunosorbent assay (ELISA) for the ST reaction used to determine resistance profiles and test metal dependence

IN enzyme activity, resistance profiles and metal ion activity dependence were determined using an ELISA. The ELISA was developed to specifically test the ST reaction of IN. This would entail the insertion of a donor substrate into a target substrate. Donor substrates act as a vDNA substitute while target substrates act as a substitute for the host genome. These artificial substrates are short oligonucleotide sequences of 32mer for the donor and 20mer for the target. The substrates are required to be double stranded with a chemical label such as biotin and fluorescein isothiocyanate (FITC) for the donor and target respectively. The donor substrate is conjugated to the ELISA plate via an interaction with streptavidin. IN or its associated mutations are then added, and bind to the donor DNA. The target DNA is then added in the presence of Mg^{2+} which commences the reaction. Thereafter an alkaline phosphatase conjugated antibody specific to FITC is added. Alkaline phosphatase substrate is added and the colour change is quantified spectrophotometrically.

Donor substrates (Top and bottom complementary strands, Table 2.1) and target substrates (Top and bottom complementary strands, Table 2.1) were heated to 95˚C for 5min and allowed to cool to room temperature to allow for annealing. Annealed double stranded donor substrate was diluted in a Donor Dilution Buffer [25mM Tris-HCl pH 7.5, 150mM NaCl, 0.05% Tween-20, 100µg/ml BSA] to a final concentration of 150nM, and added to Even Coat Streptavidin coated microplates (R&D Systems, MN, USA). These were incubated at ambient temperature for 1hr with gentle agitation. Plates were washed three times with 1X PBS (Sigma-Aldrich, MO, USA) at ambient temperature, and IN enzyme was then prepared in ST Reaction Buffer [20mM HEPES-NaOH pH 7.3 (Sigma-Aldrich, MO, USA), 75mM NaCl, 10mM $MgCl_2$ (Sigma-Aldrich, MO, USA), 10mM $MnCl_2$ (Merck Chemicals, TG, SA) 2µM $ZnCl_2$ (Merck Chemicals, TG, SA), 1% glycerol, 5mM DTT] to a final concentration of 1µM, and 100µL volume was added to each well and incubated at ambient temperature for 30min. For experiments testing metal dependence, either $MgCl_2$ or $MnCl_2$ was added to the ST Reaction Buffer. The plates were washed twice with ST Reaction Buffer at ambient temperature with no incubation. ST Reaction Buffer was then added to each well in a volume 89µL, plus 1µL of DMSO or inhibitor, at various concentrations, and the plates were incubated at 37˚C for 30min. Annealed Target substrate was diluted in ST Reaction Buffer to a concentration of 2.5µM and 10µL were added per well, giving a final concentration of 250nM Target DNA. Plates were incubated at 37˚C for 1hr, and thereafter, washed three times with 2X SSC buffer (Sigma-Aldrich, MO, USA). Monoclonal Alkaline Phosphatase conjugated anti-FITC antibody (Sigma-Aldrich, MO, USA) was prepared by 10000X dilution in TTBS and 200µL was added to the wells and incubated at ambient temperature for two hours. Plates were then washed three times with TTBS for 10min each at ambient temperature. BluePhos Microwell Substrate (KPL, MD, USA) was prepared by addition of equal parts Solution A and B, and added to each well at a volume of 100µL and incubated at 37˚C for 1hr with gentle agitation. The reactions were then stopped by addition of 100µL of AP stop solution (KPL, MD, USA). Absorbance of the wells was measured at 635nm using the xMark Microplate Absorbance Spectrophotometer (Bio-Rad, CA, USA).

2.6. AlphaScreen assay used to test the IN/LEDGF interaction

AlphaScreen utilized specialized beads which are coated with functional groups that allow for bioconjugation. IN the context of the current study, anti-His and anti-FLAG beads were used. Specific target molecules such as proteins can be bound to the surface of the beads. When a biological interaction between these protein brings the beads into proximity, a cascade of chemical reactions is initiated to produce a signal. A laser excites a photosensitizer in the "Donor" bead and converts ambient oxygen to a more excited singlet state. The singlet state oxygen molecules diffuse across to react with a chemiluminescer in the "Acceptor" bead that further activates fluorophores contained within the same acceptor bead. The fluorophores subsequently emit light at 520–620 nm. In this sense, AlphaScreen assays are similar to Forster resonance energy transfer (FRET) assays and are a useful platform to test protein-protein interactions. The stringency is less than conventional FRET and allows for a higher degree of flexibility between binding partners and requires little optimization. This is allowed, due to the relatively longer distance between donor and acceptor signals in AlphaScreen up to 200nM. This assay was an ideal choice for testing the IN/LEDGF interaction.

2.6.1. CA dose response AlphaScreen

Alpha Reaction Buffer [50mM Tris-HCl pH 7.4, 150mM NaCl, 1mM DTT, 1mM $MgCl_2$, 0.1% BSA, 0.01% Tween, 10% glycerol] was added to wells of a ½ Area 96 well plate (Perkin Elmer, MA, USA) at 40µL. Thereafter, CA at varying concentrations in DMSO was added to the wells at 2µL volumes. Either His-tagged or FLAG-tagged IN was added in a 2µL volume at a final concentration of 300nM. Reactions were incubated at ambient temperature for 30min with gentle agitation. His-tagged LEDGF/p75 was added at 2µL to a final concentration of 0.1 µM and reactions were incubated at ambient temperature for 1hr. Equal amounts of either Ni-chelating or anti-FLAG donor beads and Ni-chelating acceptor AlphaScreen beads (Perkin Elmer, MA, USA) were added at 2µl each to a final concentration of 10µg/ml. Final reaction volumes were 50µl. Reactions were incubated at 30^{o}C for 1hr and then read on an EnSpire Multimode plate reader (Perkin Elmer MA, USA).

2.6.2. Order of addition AlphaScreen

Alpha Reaction Buffer was added to wells of a ½ Area 96 well plate at 40µL. IN proteins at a final concentration of 300nM were added in a 2µL volume. For one set of experiments, annealed Donor DNA (as per section 2.6) at 150nM was first added to the wells and incubated for 30min at ambient temperature. Thereafter CA was added at 10µM final, at a 2µL volume and incubated at 30min at ambient temperature. For the second set of experiments, CA was incubated with IN for 30min before Donor DNA was added. LEDGF/p75 was added last for both and the experiment continued as per section 2.6.1.

2.7. Surface Plasmon Resonance (SPR) determination of binding kinetics of CA and LEDGIN compounds on IN_{WT} and IN_{CCD}

SPR kinetic results were obtained using the ProteOn XPR36 (Bio-Rad, MA, USA) using HTG sensor chips. The HTG sensor chips were prepared with conditioning buffer [0.5% SDS, 50mM NaOH, 100mM HCl and 300mM EDTA] and charged with 500mM $NiSO_4$. The system and sensor chips were equilibrated with SPR running buffer [50mM HEPES pH 7.4, 150mM NaCl, 0.05% Tween20, 5mM DTT, 10mM $MgCl_2$] at 25^oC. Full length IN_{Wt} and purified IN_{CCD}, both with His tags, were prepared in SPR running buffer and injected into flow cells of the HTG chips at 25µl/min until response units (RU) reached a maximum plateau. The final RU was on average ≈500 and ≈1500 for full length IN_{Wt} and IN_{CCD} respectively. Compounds, LEDGIN and CA, were diluted to various concentrations in SPR running buffer containing 5% DMSO and injected into the flow cells subsequent to each other. All data was fitted to a Langmuir 1:1 binding algorithm using the ProteOn Manager Software (Version 3.1.0.6; Bio-Rad, USA).

2.8. Determination of IN multimerization states

To determine the effect that CA has on multimerization, two separate assays were employed. Chemical covalent crosslinking of IN has been shown to stabilize the multimer forms when observed on SDS-PAGE [91]. This is particularly useful due to the disruption of these multimers due to SDS and reducing conditions in the gel. The IN cross-linking assay was performed as described previously [91]. This allowed for the capture of multimer forms and provided useful information in determining the effects that CA had on IN multimerization.

However, crosslinking isolates non-native forms of the protein. Therefore, size exclusion chromatography (SEC) is similar to SDS-PAGE and separates the proteins according to size, with the added advantage of sensitivity and allows the protein to be monitored under native conditions. SEC was used to monitor the formation of multimers of IN in response to compound incubation. Finally, the effects of CA on virion associated IN was examined by transfecting HEK293T cells with pNL4-3, a full length HIV clone, in the presence of compound. This was followed by non-reducing SDS-PAGE to detect an increase in IN dimers. This provided evidence that CA affects viral assembly similarly to results presented by Jurado *et al. (*2013) [140]

2.8.1. Gel crosslinking assays

Crosslinking reactions were prepared in Oligomerization Buffer [50mM HEPES, pH 7.1, 300mM NaCl, 2mM $MgCl_2$, 1mM DTT] with varying concentrations of ALLINI and IN protein. IN enzymes were incubated with compounds for 30 minutes at ambient temperature. The bi-functional reagent BS_3 (Thermo-Pierce, IL, USA) was added at 12.5µM final concentration, and the mixture was incubated for a further 30min at ambient temperature. Reactions were stopped by the addition of Tris-HCl to 50mM final concentration, and incubated at ambient temperature for a final 15 min. Samples were boiled in SDS-PAGE loading buffer under reducing conditions. IN multimers were visualized by SDS-PAGE and Western blot analysis as described previously (section 2.4.6).

2.8.2. Size exclusion chromatography

Multimerization caused by compounds was confirmed by SEC. Reactions were incubated in running buffer [20mM HEPES pH 7.4, 0.75 M NaCl, 1mM DTT, 10mM $MgCl_2$, 5% glycerol] for 30 minutes at room temperature prior to injection onto a TSKgel G3000SWXL column (TOSOH Bioscience, TKY, JAP). The column was attached to an NGC medium pressure chromatography system (Bio-Rad, MA, USA) and protein elution was monitored at 280nM, and protein was eluted with running buffer at 1 ml/minute. The column was calibrated with Thyroglobulin, Ferritin, Aldolase, Canalbumin, Ovalbumin, Carbonic Anhydrase and Ribonuclease A, available

in the Gel Filtrations Standards Kit (GE Healthcare BUX, UK). All SEC runs were at ambient temperature unless otherwise stated.

2.8.3. Multimerization studies in virus produced in HEK293T cells

HIV-1_{pNL4-3} virus was assembled in HEK293T cell in the presence of ALLINIs. HEK293T cells were seeded at $1X10^5$ cells/well in a 6 well format. Cells were maintained in DMEM, supplemented with 15% foetal calf serum (Thermo Scientific, MA, USA) and L-glutamine (Thermo Scientific, MA, USA). ALLINIs were added at 5µM in a 3ml volume of DMEM per well, and equivalent volume DMSO blank, was added to test wells, 24 hours before transfection. Virus was expressed from the pNL4-3 infectious viral clone after transfection in HEK293T cells by calcium phosphate transfection. A stock of 2M $CaCl_2$ (Sigma-Aldrich, MO, USA) was diluted to ≈300mM and a total of 10µg of the pNL4-3 plasmid was added to a final volume of 250µl. Non-transfected control mixtures that did not contain DNA were similarly prepared. Thereafter, 2X HBS buffer [40mM HEPES pH 7, D-Glucose 10mM (Sigma-Aldrich, MO, USA), 10mM KCl (Merck Chemicals, TG, SA), 270mM NaCl, 1.5mM Na_2HPO_4 (Sigma-Aldrich, MO, USA)] was slowly added to the DNA mixture to a final volume of 500µl. The HBS-$CaCl_2$-DNA mixture was then added at 250µl to each well. Supernatants were collected 48 hours post transfection. Samples were run on non-reducing SDS-PAGE and analysis of viral production were assayed on Western blot against the HXB2 antisera against IN_{24-36}. The Western blot detection of IN both confirmed viral production in the HEK293T cells and was used to test for multimerization.

2.9. In silico modelling for the prediction of possible binding modes of compounds

2.9.1. Receptor preparation and binding site selection

Three IN catalytic core domain (CCD) crystal structures, that were available on the RCSB protein database (http://www.rcsb.org/pdb/home/home.do), were prepared for ligand docking. These structures included: The CCD complexed with Mg^{2+} and 5CITEP (1QS4 [83]) at 2.1 Å resolution and the CCD complexed with LEDGF/p75 (2B4J [260]) at 2.02 Å. These structures were chosen as they represent the CCD in the various bound conformations with

an inhibitor and metal ion at the active site and with a critical host factor, respectively. Non-specific hetero atoms were manually removed and structures were further prepared in Discovery Studio 4.0 (ACCELRYS INC, CA, USA). Missing loop residues were built with the SEQRES definitions with a maximal loop residue set to 20 and loops were minimised using the CHARMm force-field. All proteins were then protonated at an energy cut-off of 0.9 (kcal/mol), with a dielectric constant of 10, at physiological pH 7.4. Water was removed and the entire structure was minimised using the CHARMm force-field.

2.9.2. Ligand preparation

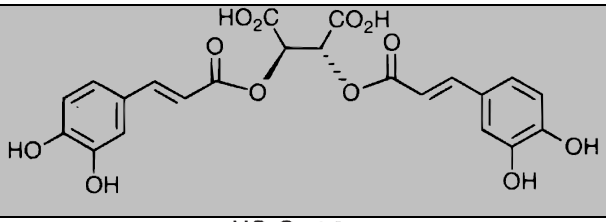
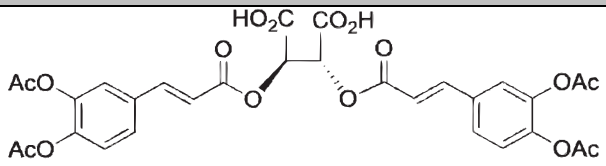
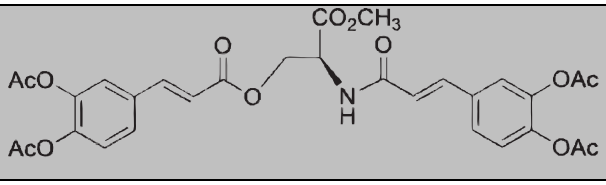
Three compounds: CA, TCA and ACA-1 (Table 2.2) were chosen based on their divergent chemistries and MOA [175, 235, 238] for comparisons in the docking studies. CA represents the parent compound, while TCA and ACA-1 are derivatives. Ligand ionisation was based on physiological pH set between 6.5 and 8.5. Tautomers and isomers were generated, with all tautomers generated up to a maximum of 10, and unknown stereoisomers and stereobonds generated. Bad valencies were fixed, and 3D coordinates of the ligand were generated. There was no parallel processing and duplicate structures that were generated were removed.

2.9.3. Docking, scoring and calculation of binding energies

Binding sites were defined manually at the IN/DNA interface as well as the IN/IN interface and their details are listed in Table 2.3. The active site definition was based on previous docking studies [244, 245] and these amino acid positions where supported by biochemical data [234, 246]. The allosteric site was chosen around the CCD amino acid residues involved in the interaction with LEDGF [92]. Prepared ligands were docked to the prepared protein binding sites using LibDock; with the number of polar and apolar hot spots set to 100, and a RMSD docking tolerance of 0.25. The default High Quality Docking Preference setting was used, and the “BEST” conformations algorithm was used and set at a threshold of 20kcal/mol. Docked poses generated by LibDock were scored using the default setting of the algorithms: LigScore1, LigScore2, PLP1, PLP2, Jain, PMF and PMF04. Minimizations of the docked and scored poses were performed last, using the Calculate Binding Energy function. The active site

residues were treated as flexible, and minimised using the Smart Minimiser function with 100000 max steps and a RMSD gradient cut-off of 0.001kcal/mol-Å. Entropy of the ligand pose conformations was calculated using the FAST algorithm and minimised using the Smart Minimiser function with 500 max steps and RMS gradient cut-off of 0.1kcal/(mol x Å). Binding energies were calculated under vacuum conditions, and the non-bond radius was set at 14Å. Translational and rotational entropy terms were estimated at 298.15K and partial charges were calculated by Momany-Rone estimation.

Table 2.2: Structures of Chicoric acid (CA), tetra-acetylated chicoric acid (TCA) and the tetra acetylated chicoric acid derivative (ACA-1) used in *in silico* modelling.

Name	Structure[1]	Reference
CA		[235, 238]
TCA[2]		[175, 235]
ACA-1[2]		[175, 235]

[1] Based on referenced data

[2] Ac = Acetyl

Table 2.3: Binding site details for *in silico* modelling. Amino acids were chosen around the binding site of interest and a sphere was created for ≈ 14.5 Å. The XYZ co-ordinates of the binding sphere are listed. Active site was only chosen for the 1QS4 structure. The allosteric site was chosen for both the 1QS4 and 2B4J structures.

Binding site name	**Amino Acid chosen**	**Binding site 3D coordinates**		**Radius (Å)**	
		1QS4	2B4J	1QS4	2B4J
Active site	Chain A: K159, K156, E92, Q148, E116	X=6.281, Y=9.338, Z=-20.124	na	14.4	na
Allosteric site*	Chain A: T125, D131 Chain B: D167, H171, K173	X=-16.653 Y=6.789 Z=49.879	X=-13.005 Y=4.917 Z=-7.53	14.53	14.39

*Allosteric site amino acids chosen for 2B4J were the same.

Chapter 3: Results

3.1. Synthesis and construction of expression vectors

3.1.1. Mutagenesis

Well defined [210, 258] RAL resistant mutations Q148H and N155H and the associated double mutations G140S/Q148H and E92Q/N155H were successfully introduced into the parent sequence by site directed mutagenesis. The pINSD.His.sol and pINSD.His expression vectors were both utilized as parent sequence and these vectors express soluble mutant IN (IN_{Sol}) and wild type IN (IN_{WT}), respectively. A pre-existing double F185K/C280S amino acid modification, which increases IN solubility [84], is found in the pINSD.His.sol sequence. These parent and modified vectors were used to express IN_{WT} and mutant enzymes for resistance profiling of CA. Additionally, an IN mutation resistant to multimerization, at position A128T, was introduced into pINSD.His only, and used in multimerization assays [142]. All mutagenesis products were confirmed by nucleotide sequencing as shown in the chromatograms in Figure 3.1.

3.1.2. Cloning the FLAG-tagged, His-FLAG-tagged IN and His-tagged CCD.

For the AlphaScreen assay it was desirable to generate a FLAG-tagged IN for interaction with a His-tagged LEDGF. To generate IN_{WT} with a joint His-FLAG-tag, pINSD.His was used as a template for PCR amplification and a two-step sub-cloning procedure was performed. IN_{WT} sequence is 864 base pairs long and was first cloned into pT7-FLAG-1 in frame with a FLAG tag (Figure 3.2 A) and hence named pIN-T7FLAG. Thereafter, the FLAG-tagged IN sequence was sub-cloned into pET16B in frame with the 10X His tag (Figure 3.2 B) and named pIN-His-FLAG. Successful clones were confirmed by restriction enzyme digests (Figure 3.2) and sequencing. The truncated CCD of IN, ranging from amino acid position 50 to 212, was PCR amplified from a pINSD.His.sol template sequence, and was successfully sub-cloned into pET16B (Figure 3.2 C) and named pCCD16B-His. A schematic of the expression systems is illustrated in Figure 3.3. Similarly, cloning success was confirmed for all vector modifications by restriction enzyme digest and DNA sequencing.

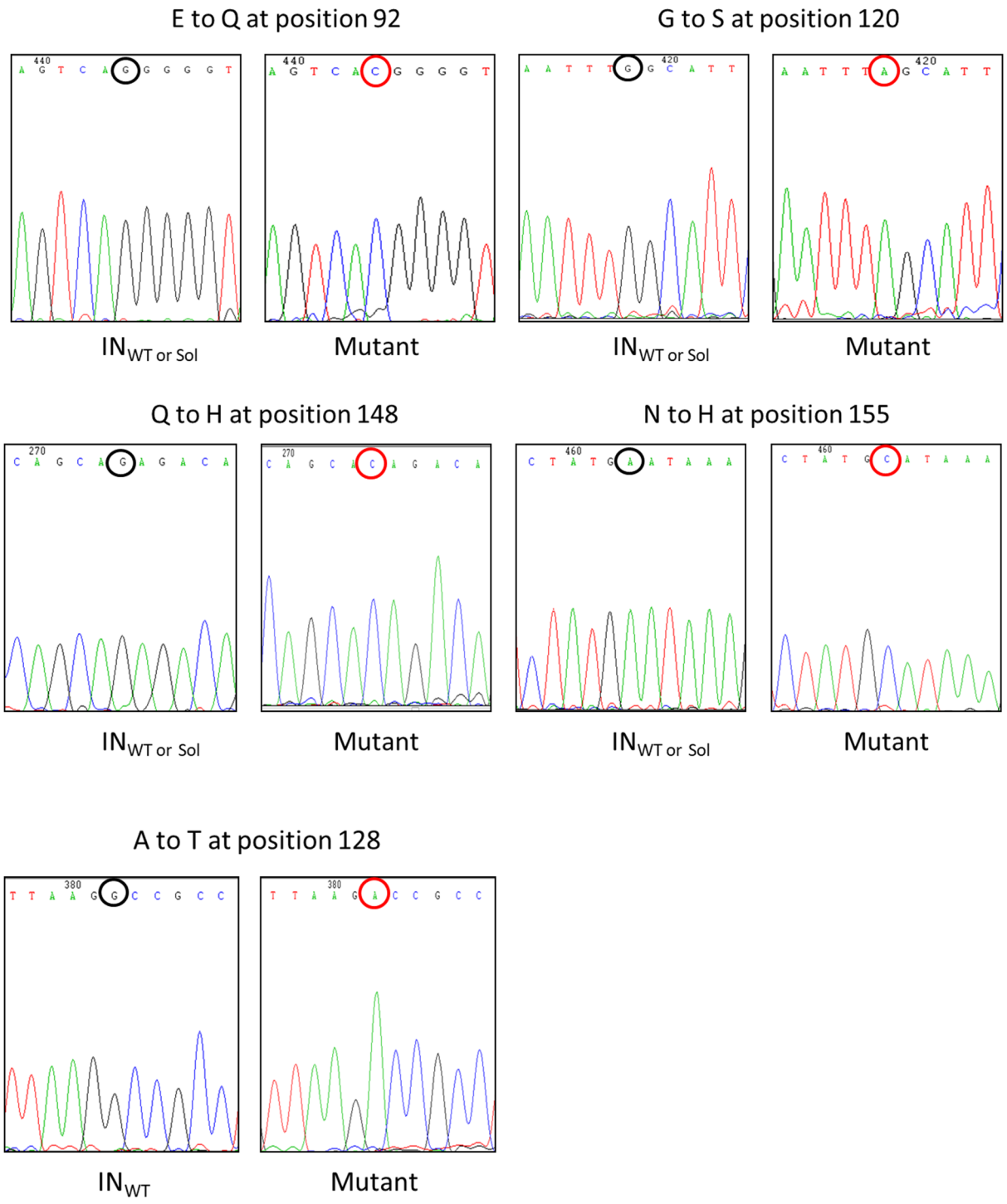


Figure 3.1: Sequence chromatograms of single base mutations. Each mutation is represented alongside the $IN_{WT\ or\ Sol}$ parent sequence for comparisons. The IN_{Q148H} and IN_{N155H} were single mutants, and the G140S and E92Q were mutations generated in context of the $IN_{G140S/Q148H}$ and $IN_{E92Q/N155H}$ double mutants. The IN_{A128T} was generated as a single mutation against only the IN_{WT} parent sequence.

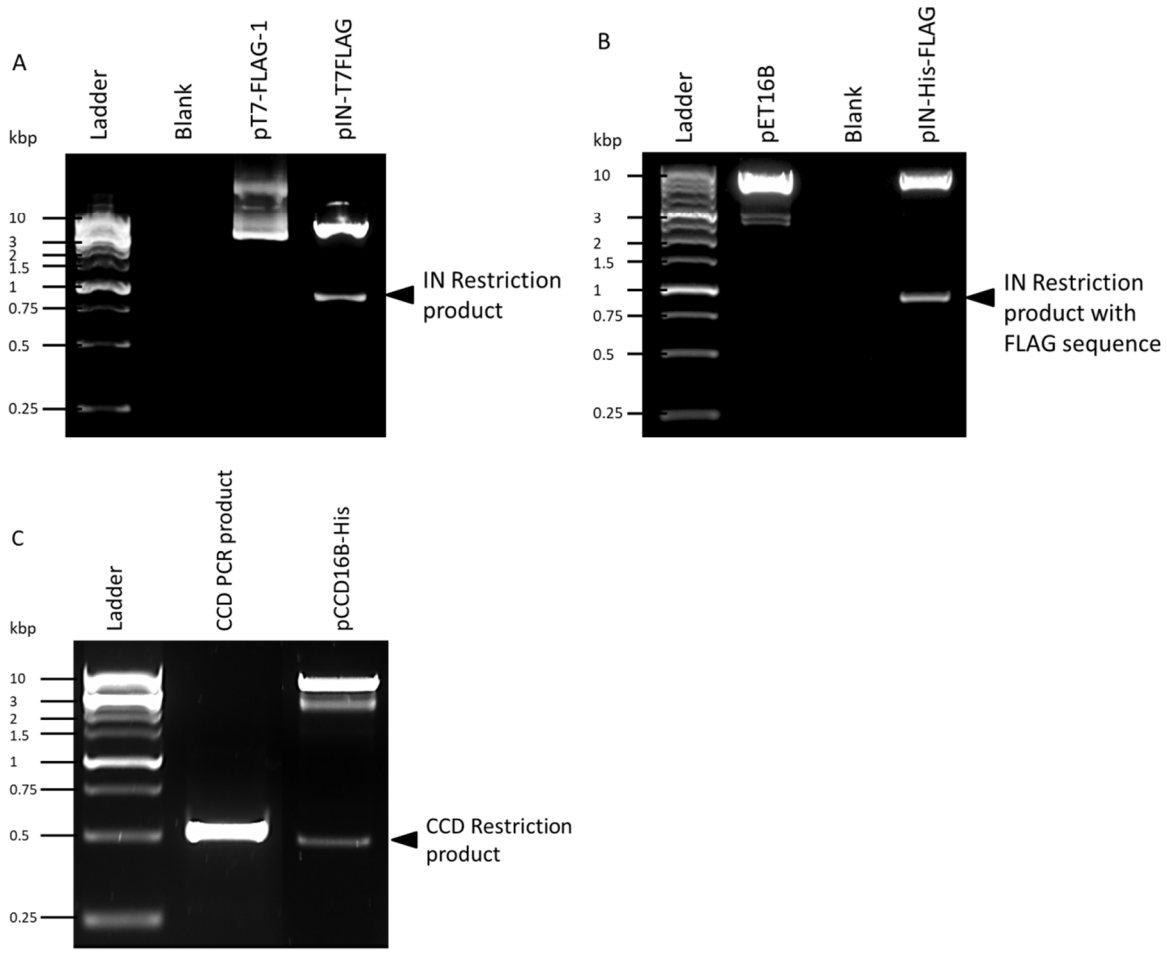


Figure 3.2: Restriction digests confirming IN_{WT} cloning. A) IN_{WT} was initially cloned into pT7-FLAG-1 to incorporate a FLAG-tag (pIN-T7FLAG), and this was confirmed by restriction digest of the cloning product with *Hind*III and *Bgl*II, yielding a ≈850 base pair DNA fragment (last lane). B) Following the cloning into the pT7-FLAG-1 vector, the resulting IN clone (pIN-His-FLAG) was used as a template for directional cloning into pET16B, and this effectively incorporated a His-tag into the pre-existing FLAG-tagged IN sequence. Cloning into pET16B was confirmed by restriction digests with *Nde*I and *Bam*HI, yielding a ≈850 base pair restriction product (last lane). C) The CCD represents a truncated IN product of 486 base pairs, and this sequence was cloned into pET16B between *Nde*I and *Bam*HI sites. A successful clone (pCCD16B-His) was confirmed by restriction digests and compared to its PCR products (last lane). Molecular weight marker sizes in kilobases (kb) are indicated on the left. The arrow depicts the amplicons of interest.

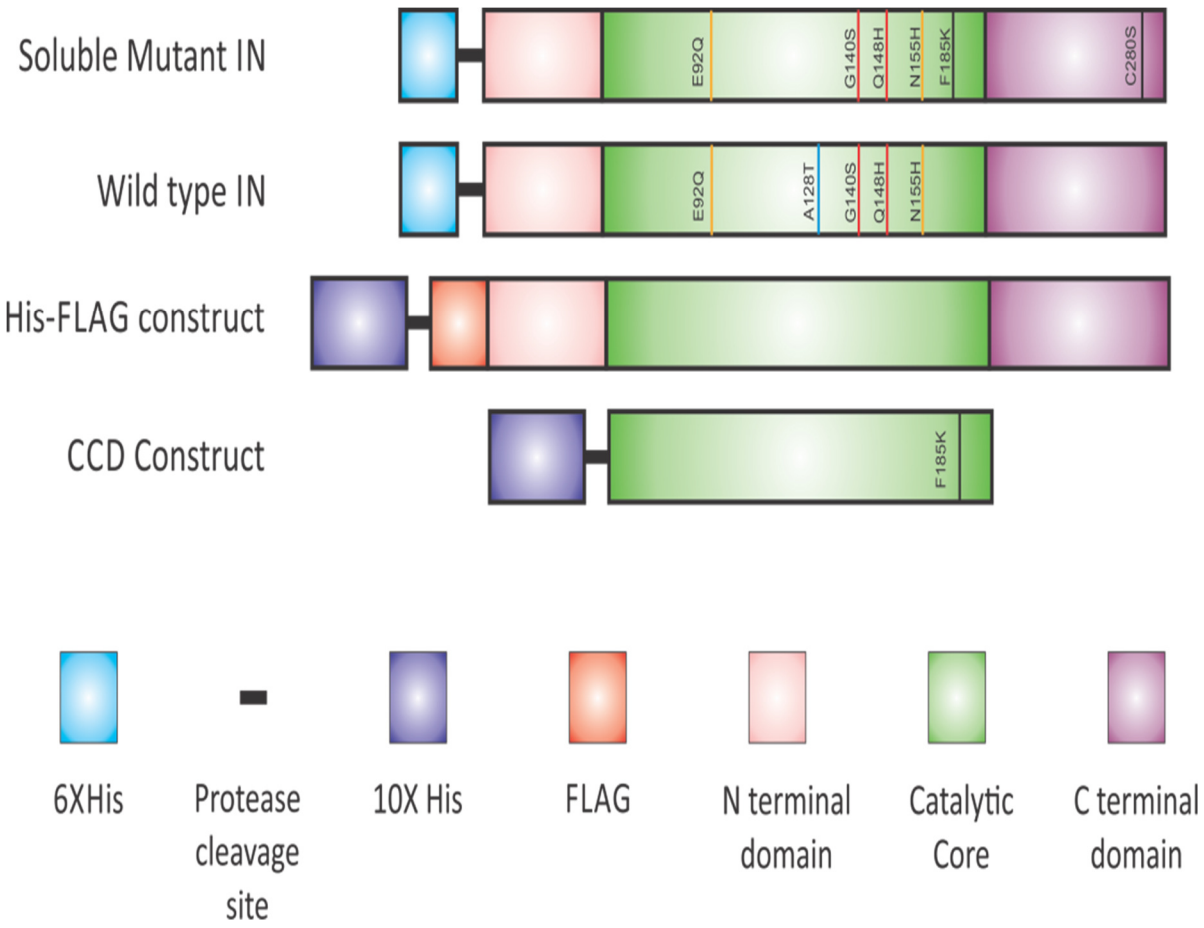


Figure 3.3: Schematic representation of the DNA constructs used or constructed in this study. Amino acid differences compared to the Wt pINSD.His are indicated by vertical lines. These are colour coded according to the grouping; the soluble double mutation (black) was a notable fixed feature of pINSD.His.sol. The modified positions of Q148H and its double mutation with G140S (red) and N155H and its double mutation with E92Q (yellow) are indicated. A128T was only modified in the wild type sequence. Wild type IN sequence from pINSD.His was used for the synthesis of the His-FLAG construct. The CCD was constructed using the soluble mutant IN sequence from pINSD.His.sol.

3.2. Protein expression and purification

3.2.1. IN_{WT} and IN_{Sol} and associated mutations

Expression of the His tagged IN_{WT} and IN_{Sol}, as well as all modified enzymes were induced in BL21 cells using IPTG. Inductions generally produced a distinctive increase in protein at the 30 kilo Dalton (kDa) range, corresponding to IN (Figure 3.4). Nickel affinity chromatography purifications produced sufficiently pure IN protein (Figure 3.5) and further purifications were not necessary. Enzymes resistant to RAL include, IN_{Q148H}, $IN_{G140S/Q148H}$, IN_{N155H} and $IN_{E92Q/N155H}$ which were modified for both IN_{WT} and IN_{Sol} parent sequences were purified by nickel affinity. Wild type sequence modifications (pINSD.His) IN_{A128T} and His-FLAG tagged (H-F) IN, and the Sol truncated CCD (IN_{CCD}) were similarly isolated. Final protein yields were less for the IN_{WT} compared to the IN_{Sol}. This was expected due to the increased solubility of the latter, with the exception of Sol $IN_{E92Q/N155H}$, which had consistently low protein yields in comparable final volumes. This reduced expression may be as a result of the cumulative effects of the introduced resistance mutations along with the double soluble mutations. The protein however could be adjusted to the desired working concentration before use by concentration with Amicon mini ultrafiltration columns. The SDS-PAGE result depicts a highly pure approximately 34kDa product for the full length IN (Figure 3.6). Final protein yields are listed in Table 3.1.

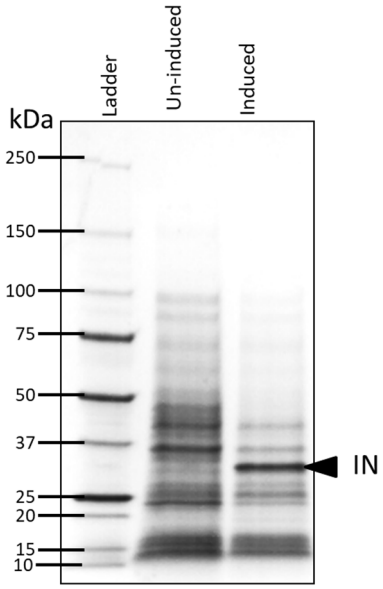


Figure 3.4: SDS-PAGE showing IPTG induction of IN expression, as compared to un-induced BL21 bacteria. This picture is representative of a typical induction for IN_{WT} and/or IN_{Sol} and all there associated mutations. A definite increase of a protein at approximately 34kDa is noticed, which corresponds to the expected size of IN. MW marker sizes (in kDa) are indicated on the left. The arrow shows the induced IN.

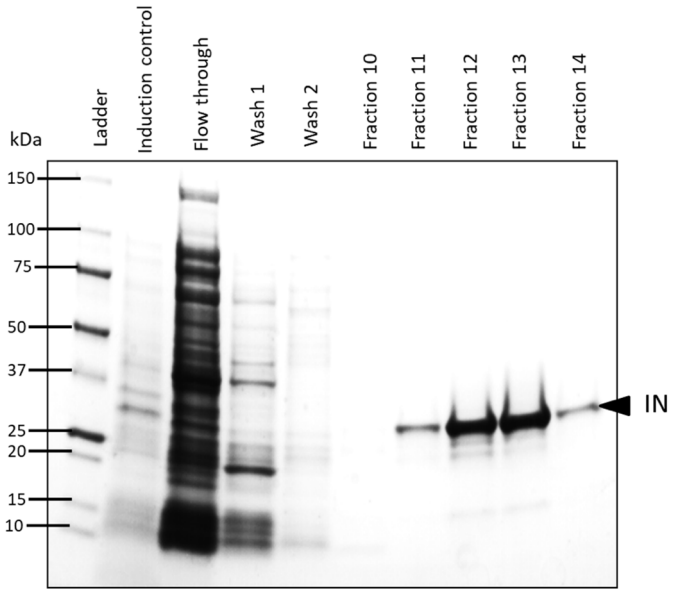


Figure 3.5: SDS-PAGE showing purification of IN_{Sol} by nickel affinity chromatography. Induced bacteria were lysed and loaded onto a nickel affinity column. The flow through and washes appeared to show little IN protein loss. Protein purification produced a pure protein with little to no contaminating bands, and most protein eluted of the column between fractions 11 to 14 using the current methodology. MW marker sizes (in kDa) are indicated on the left. The arrow shows the eluted IN in the various fractions. This figure serves as a representative purification to IN_{WT}, IN_{Sol} and their associated mutations.

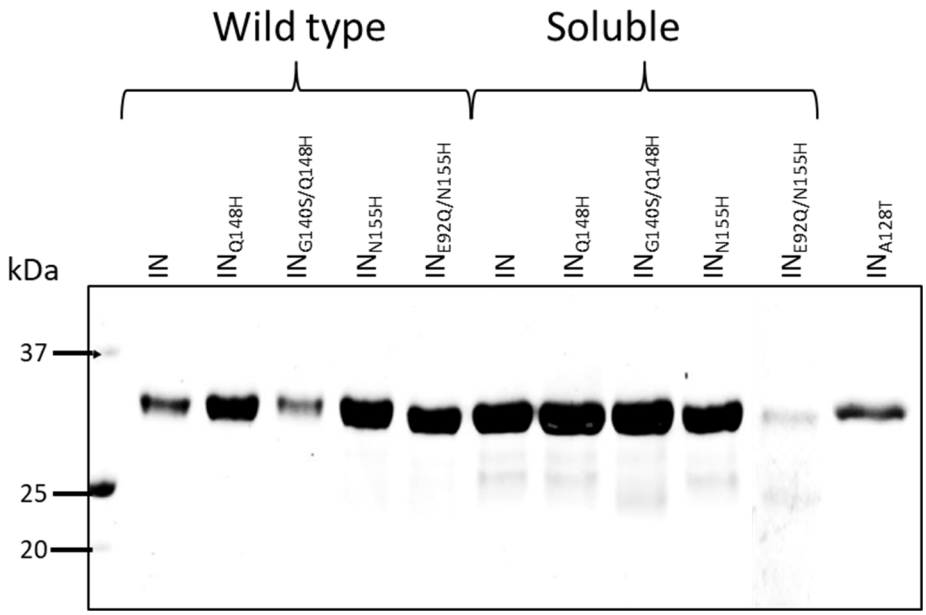


Figure 3.6: SDS-PAGE results of various purifications of IN_{WT} and IN_{Sol} with associated mutants. Proteins were purified using nickel affinity and produced highly pure products. All proteins migrated at the expected size of ≈34kDa. The picture is an amalgamation of final stocks of purified proteins. MW marker sizes (in kDa) are indicated on the left.

3.2.2. His-FLAG IN purification

The pIN-His-FLAG clone was used to produce a His-FLAG-tagged IN_{WT} protein (H-F IN). A Western blot was used to confirm that both FLAG and His tags were present before the subsequent purification steps (Figure 3.7). Bacterial transformants containing pIN-His-FLAG were induced with IPTG and SDS-PAGE resolved lysates were transferred to nitrocellulose. Membranes were subjected to incubations with primary Anti-His and Anti-FLAG antibodies, and the induced H-F IN tested positive for both tags (Figure 3.7). After the double tag confirmation, the H-F IN was purified using nickel affinity chromatography (Figure 3.8). For preparation of the AlphaScreen, the His tag needed to be removed, yielding a FLAG-tagged protein. The pET16B expression vector contains a Factor Xa cleavage site between the 10X His tag and the fusion protein. The H-F IN appeared stable after Factor Xa cleavage (Figure 3.9), however, during subsequent processing by benzamidine column and PD 10 buffer exchange most of the protein was lost (Figure 3.9). The protein loss is assumed to be due to aggregation and could not be corrected.

3.2.3. FLAG-IN purification

Unfortunately the H-F IN purification was unsuccessful after Factor Xa cleavage. The procedure could not be optimized as the Factor Xa cleavage protocol requires a low salt concentration and long incubations at 23^{o}C for cleavage. The IN_{WT} is especially prone to aggregation and therefore an alternative purification protocol was devised based on previously reported procedures [167, 259]. The pIN-T7FLAG clone is a bacterial expression system and was induced similarly to the pET expression systems used in this study. The FLAG-tagged IN_{WT} was then subject to an initial lysate wash by centrifugation at high speed, followed by ammonium sulphate precipitation and HIC column purification. Finally, the protein was purified by heparin column, pooled and concentrated (Figure 3.10). This procedure yielded a semi pure IN protein, relative to the nickel affinity chromatography procedure, with noticeable contaminating bands. The bulk of the protein eluted in fraction 6 during the gradient elution.

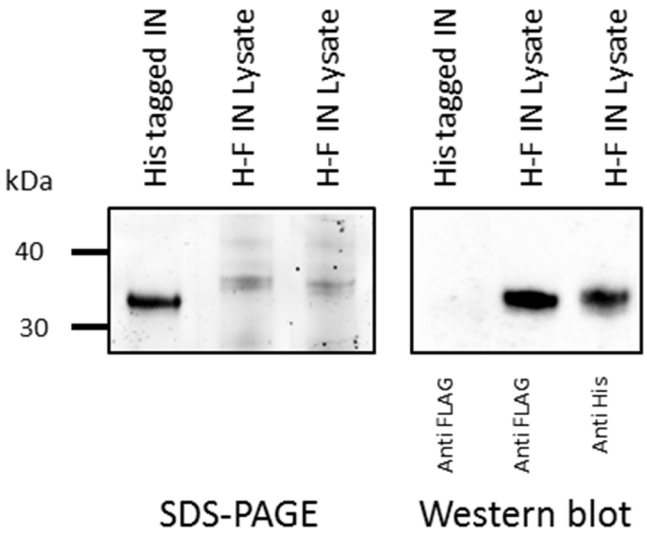


Figure 3.7: H-F IN tested positive for both the His tag and FLAG tag, as shown by SDS-PAGE and western blot. A His tagged IN_{WT} was used as a negative control. Both the His tag as well as FLAG tag was detected on H-F IN. The SDS-PAGE indicated a H-F IN band slightly higher than IN_{WT}. MW marker sizes (in kDa) are indicated on the left.

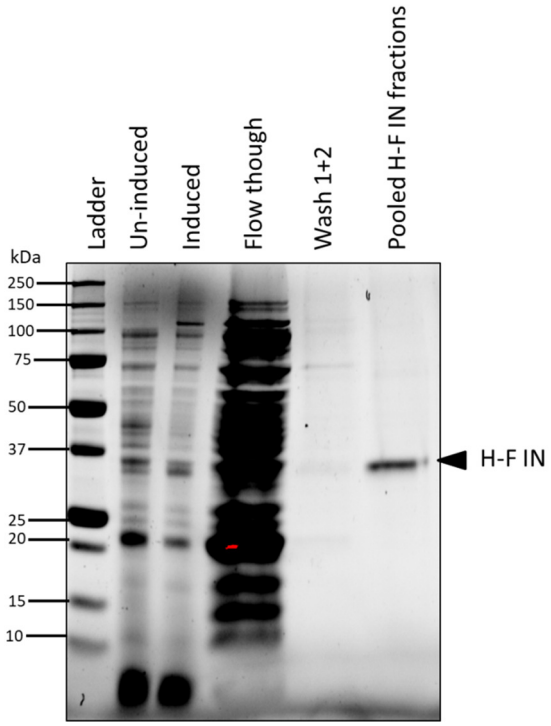


Figure 3.8: SDS-PAGE results of an H-F IN purification by nickel affinity chromatography. Induced H-F IN was not as clear as single His-tagged IN. Nevertheless, the protein eluted off the column as a relatively pure band with little to no contaminants. MW marker sizes (in kDa) are indicated on the left. The arrow shows the pooled IN fractions.

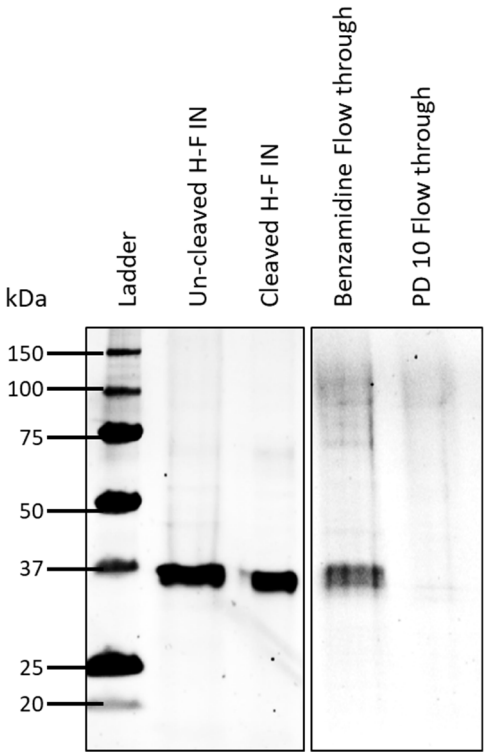


Figure 3.9: SDS-PAGE analysis of Factor Xa cleavage of H-F IN. An SDS-PAGE illustrates the H-F IN before and after a 16 hour cleavage at 23^{o}C. A benzamidine column was used to remove Factor Xa, and followed by a PD 10 buffer exchange. As noticed in the picture, most of the protein was lost after PD 10 buffer exchange, and this was as a result of suspected aggregation. MW marker sizes (in kDa) are indicated on the left.

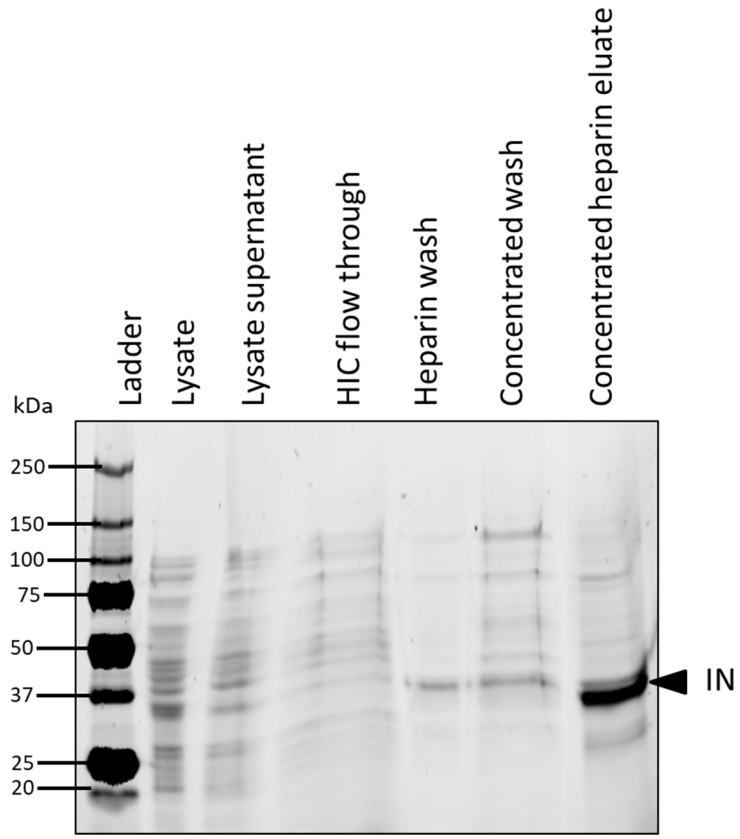


Figure 3.10: SDS-PAGE results of indicated FLAG-IN purification samples. The induced BL21 bacteria containing the pIN-T7FLAG clone was lysed in low salt (100mM NaCl) buffer. The lysates were washed by centrifugation at 95 000 *x g* and pellets were retained. The pellet was re-suspended in 1M NaCl buffer and loaded onto the HIC column. The HIC column flow through showed little loss of IN protein. After concetration and buffer exchange, the IN containing HIC eluate was loaded onto a heparin column. There was some IN protein loss in the washes, however, most FLAG-tagged IN protein eluted of the heparin column during gradient elution. MW marker sizes (in kDa) are indicated on the left. The arrow shows the concentrated IN.

3.2.4. LEDGF purification

A His-tagged LEDGF protein was prepared for the AlphaScreen. Initially the protein was purified by nickel affinity chromatography, which yielded a significantly contaminated protein mixture, which included LEDGF (Figure 3.11). Subsequently, LEDGF was purified by heparin column and some protein loss was experienced in the flow through and washes. During the gradient elution, LEDGF eluted in fraction 9, which was about 60% of the NaCl gradient, along with some contaminating bands (Figure 3.11).

Although the heparin purification of FLAG-tagged IN and His-tagged LEDGF did not produce pure protein, this was deemed an insignificant problem as the AlphaScreen assay has been successfully used with cell lysates previously [261, 262].

3.2.5. IN_{CCD} purification

The IN_{CCD} was purified from pCCD16B-His bacterial transformants. The procedure was similar to full length IN, however, purification buffer components were diluted two fold, as described elsewhere [169], save for the imidazole concentrations. Purification of IN_{CCD} yielded highly pure protein (Figure 3.12). The resulting protein along with the 10X His-tag was ≈20kDa, and the migration of the protein on SDS-PAGE confirmed IN_{CCD} presence (Figure 3.12).

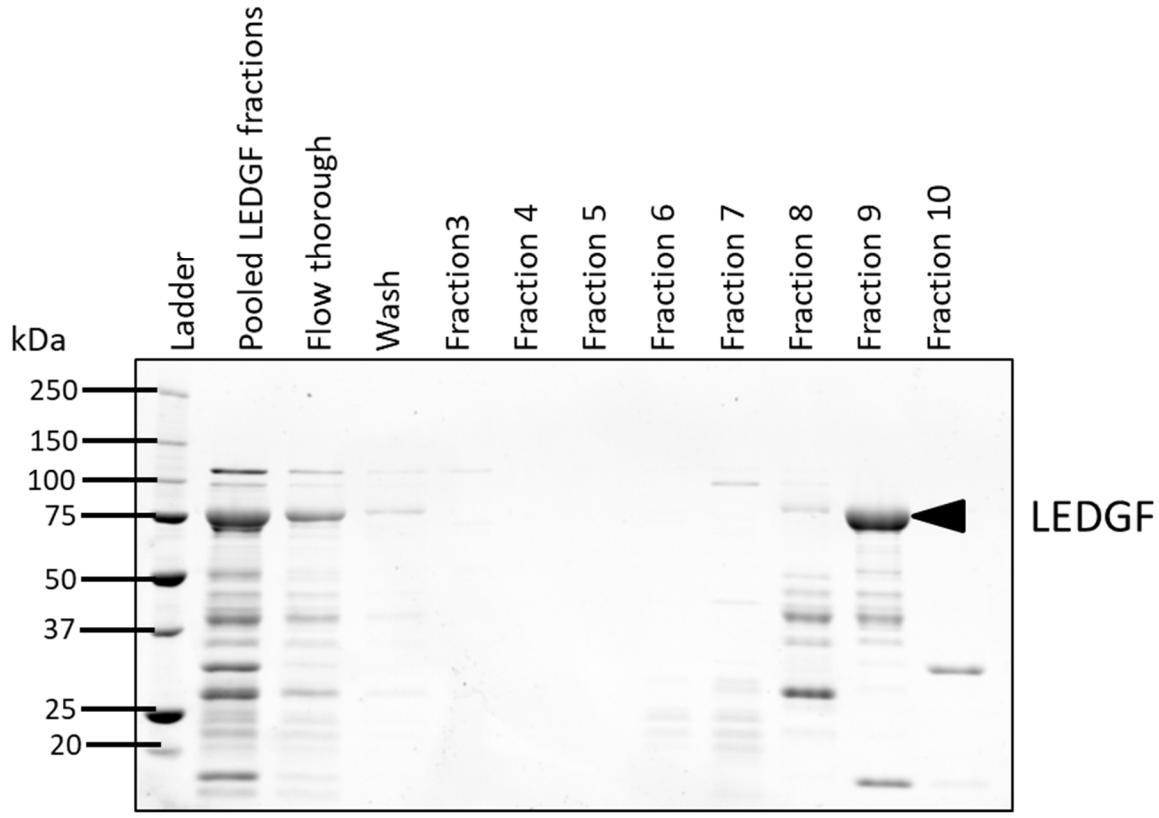


Figure 3.11: SDS-PAGE analysis of samples collected during the LEDGF purification by nickel affinity chromatography, followed by heparin purification. The nickel affinity eluate factions containing LEDGF were pooled, concentrated and buffer exchanged in preparation for the heparin column purification. Some LEDGF protein loses was experienced in the flow through and binding buffer wash, however bulk of the protein eluted in fraction 9 during the gradient elution. The LEDGF protein eluate was however noticeably contaminated with foreign bands. MW marker sizes (in kDa) are indicated on the left. The arrowhead points to LEDGF.

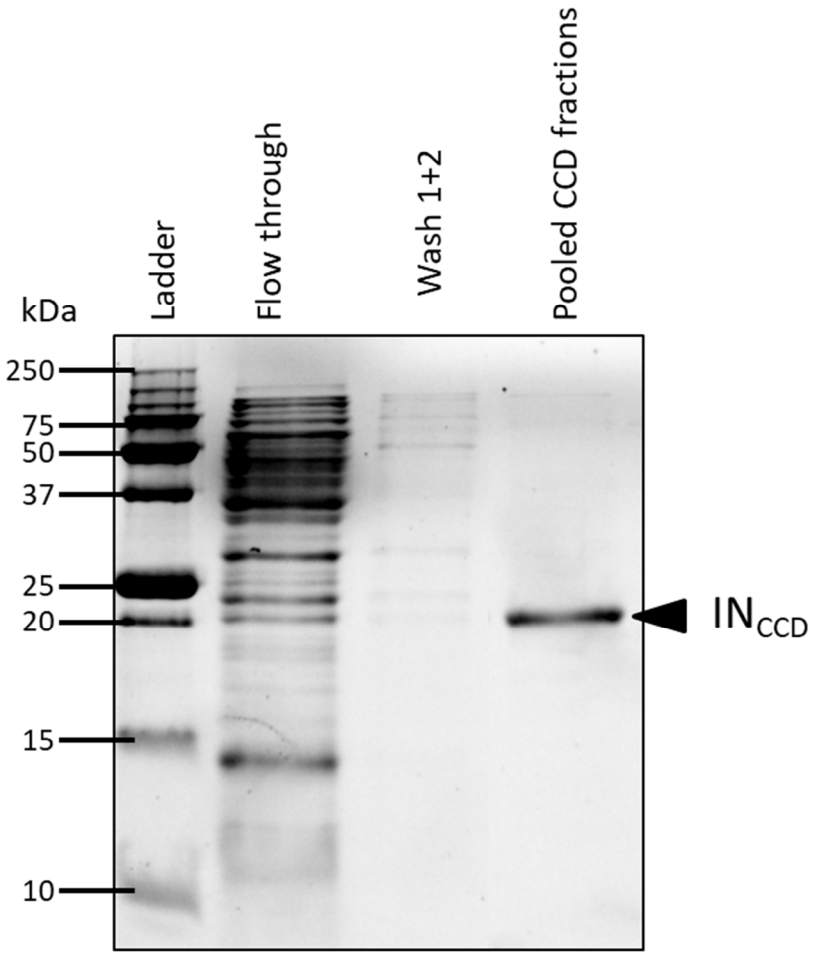


Figure 3.12: SDS-PAGE analysis of IN_{CCD} purification by nickel affinity chromatography. There was little protein loss in the flow though and the washes. A relatively pure protein at the ≈20kDa range was purified and corresponds to the expected size of His-tagged $IN_{CCD.}$ MW marker sizes (in kDa) are indicated on the left. The arrow shows the pooled IN CCD.

Table 3.1: Final protein stock concentrations and yields obtained from all recombinant integrase (IN) wild type, mutants and lens epithelium derived growth factor (LEDGF) protein purifications.

Parent Protein	**Protein purified**	**Average stock concentrations (µg/ml) in 3ml final volumes**	**Final yields (mg)**
IN_{WT}	IN	289	0.87
	IN_{Q148H}	442	1.33
	$IN_{G140S/Q148H}$	272	0.82
	IN_{N155H}	210.8	0.63
	$IN_{E92Q/N155H}$	612	1.84
	IN_{A128T}	714	2.14
	H-FIN	510	1.53
IN_{Sol}	IN	952	2.86
	IN_{Q148H}	1122	3.37
	$IN_{G140S/Q148H}$	1190	3.57
	IN_{N155H}	748	2.24
	$IN_{E92Q/N155H}$	68	0.20
	IN_{CCD}	803	2.41
LEDGF	LEDGF	432	1.30

3.3. Characterization of CA inhibition in functional assay of integrase and its mutants

3.3.1. ELISA functional assay characterization of enzymatic activity and resistance profiling

Comparisons of ALLINIs to INSTIs have indicated a difference in resistance profiling [168, 169]. To test this, we used an ELISA with prepossessed vDNA mimics to test the ST activity of the enzymes, using Mn^{2+} and Mg^{2+} in the reaction buffer. The ELISA confirmed the activities of the purified proteins. Activity of the IN_{Sol} RAL resistant mutants were normalized to IN_{sol} (Figure 3.13). Unfortunately, the RAL resistance mutations induced in IN_{WT} were not active under the conditions tested and were not included in the resistance profiling. The enzymatic activity of the IN_{Sol} and its associated mutations are summarized in Table 3.2. Both single mutations, IN_{Q148H} and IN_{N155H}, have significantly ($p<0.01$) higher activity compared to their double mutant counterparts, $IN_{G140S/Q148H}$ and $IN_{E92Q/N155H}$ respectively. IN_{N155H}, was slightly but not significantly higher than IN_{Sol}. The result was unexpected, as the single mutants should have impacted activity while the double mutants should have restored enzymatic activity [258, 263]. The IN_{WT} is represented separately, and in comparison to H-F IN (Figure 3.13). HF-IN had a 64% activity compared to the IN_{WT}.

IN_{Sol} mutants were screened against RAL and CA (Figure 3.14) and their fold change in IC_{50} ($FCIC_{50}$) values were compared to IN_{Sol}, (summarised in Table 3.2). There was approximately (≈)950 and ≈1260 fold resistance against RAL of IN_{Q148H} and $IN_{G140S/Q148H}$ respectively, while against CA the enzymes had a ≈3.4 and ≈1.4 fold resistance, respectively. Similarly for IN_{N155H} and $IN_{E92Q/N155H}$, RAL resistance was at ≈390 and ≈580 fold, respectively, while against CA it was ≈1 and ≈3 fold, respectively. This shows a striking difference of resistance profiling between RAL and CA, suggesting a separate binding site or possible distinct mode of action of CA as compared to RAL.

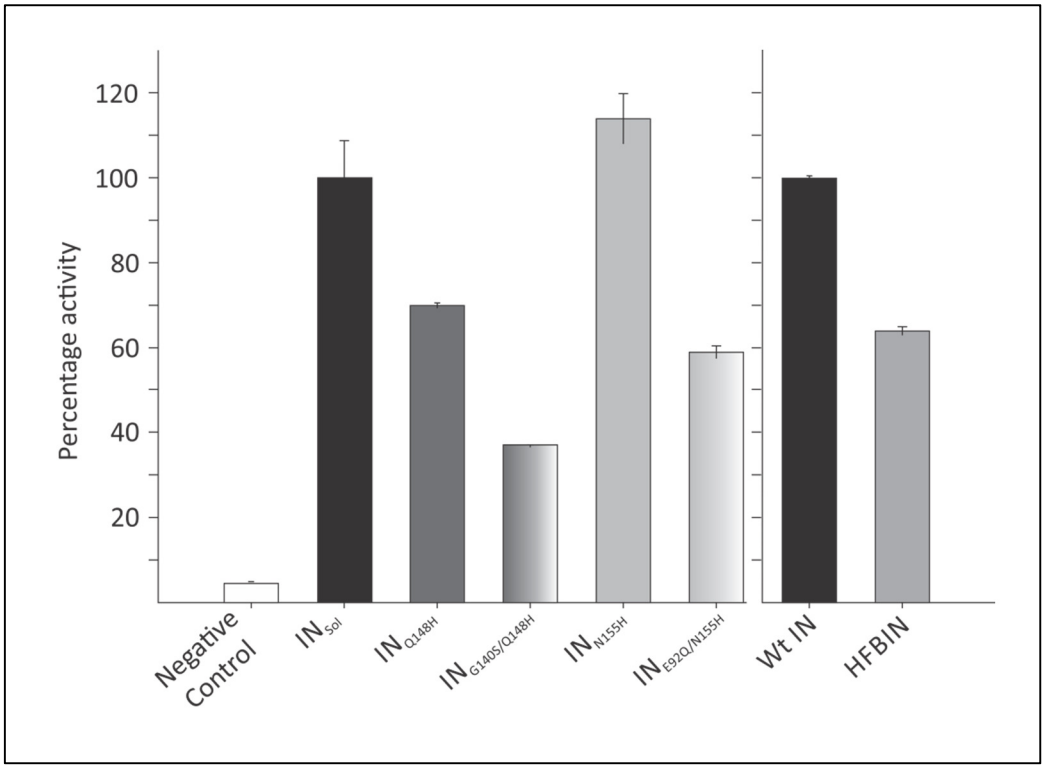


Figure 3.13: Enzymatic activity of all the purified recombinant IN as determined by ELISA. A 1µM concentration of IN was used for all reactions. IN_{Sol} amd IN_{WT} activities are compared (both normalized to 100%) to their associated mutations, and are represented on different axis in the graph.

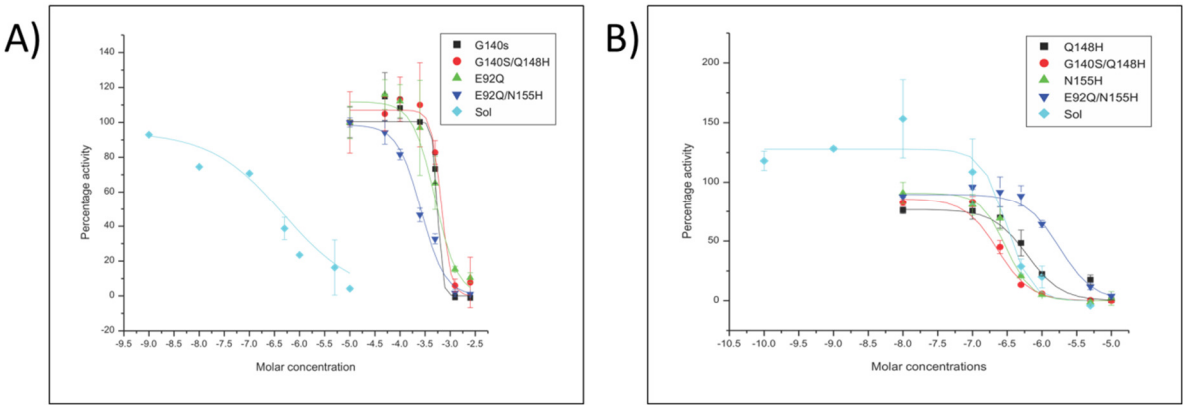


Figure 3.14: Dose response curves of IN_{Sol} and resistant mutants' activity against RAL (A) and CA (B). Results were analysed from ELISA obtained data. IC_{50} results are summarized in Table **3.2**.

Table 3.2: Activity, inhibition and fold change in IC_{50} ($FCIC_{50}$) summary for the RAL resistant mutants. All listed IN proteins were tested for activity, for inhibition against raltegravir (RAL) and chicoric acid (CA).

Mutant	**Activity**[a]	**Inhibition**						**$FCIC_{50}$**[b]	
		RAL			**CA**			**RAL**	**CA**
		IC_{50}[c]	**SD**[d]	**Hill**[e]	**IC_{50}**[c]	**SD**[d]	**Hill**[e]		
IN_{Sol}	100%	423nM	162nM	1,13	403nM	75nM	1,65	1	1
IN_{Q148H}	70%	401µM	27µM	6,83	1.4µM	273nM	2,42	947.99	3.42
$IN_{G140S/Q148H}$	37%	534µM	204µM	11,05	575nM	83nM	1,42	1262.41	1.43
IN_{N155H}	114%	166µM	20µM	2,77	441nM	364nM	0,76	392.44	1.09
$IN_{E92Q/N155H}$	59%	247µM	137µM	4,71	1.2µM	564nM	1,96	583.92	2.90

[a] enzymatic activity as per Figure 3.13

[b] comparisons made to IN_{Sol}

[c] 50% inhibitory concentration

[d] Standard deviation

[e] Hill coefficient as per reference [265, 266]

The Hill slope of a dose response curve is the measurement of cooperativity of inhibition and can be a useful indication of binding modes of compounds [167, 264, 265]. Enzymes may contain multiple substrate or regulator binding sites. When a single substrate or regulator interacts with a binding site, it may affect the substrate affinity of a different binding site. This relationship is known as cooperativity, and can be either non-cooperative, positively cooperative or negatively cooperative. Non-cooperative binding indicates that the initial substrate-binding event did not affect the affinity of subsequent binding events. Positive cooperativity is when the enzyme affinity for a substrate increases after the initial binding event, while negative cooperativity is the opposite. Cooperativity can thus also indicate multiple binding sites, as a cooperative relationship is dependent on multiple binding events. This relationship was used to explain the Hill slopes and cooperative nature of inhibitor binding to IN. A non-cooperative relationship is seen when the Hill slope is equal to or less than one and usually denotes a single binding site, while cooperative inhibition presents with a slope greater than one and may indicate more than one binding site. Inhibition of IN_{Sol} by RAL shows a non-cooperative relationship (Table 3.2). The inhibition of the resistance mutants by RAL shows an increase in cooperativity. A trend is noticed, indicating that the higher the resistance, the greater the cooperativity. In contrast to RAL, CA had a cooperative inhibition of IN_{Sol} with Hill slope = 1.65. CA followed a similar trend as RAL, where in cases of resistance and higher $FCIC_{50}$ values, the cooperativity was greatest, with the exception of $IN_{G140S/N155H}$.

3.3.2. Inhibition of IN_{Sol} and IN_{WT} by CA in the presence of $MgCl_2^+$ verses $MnCl_2^+$

Some DKAs are sensitive to the type of metal ion present in the reaction buffer, and were shown to be more effective in the presence of Mn^{2+} [128]. Modelling suggested that the active site is more flexible in the presence of Mn^{2+} accounting for the increased inhibition by the DKAs [267]. This was in contrast to what is noticed in ALLINIs [168] where they behave similarly in the presence of either metal ion. A comparison of CA activity in the presence of either Mg^{2+} or Mn^{2+}, was performed to determine compound potency under conditions exclusive for a single type of metal ion (Figure 3.15 A, Figure 3.16 A, and Table 3.3). Initially, the inhibition of IN_{sol} was compared in different metal ions (Figure 3.15 A). In the presence of Mn^{2+}, CA had an IC_{50} of 201nM (± 33nM), while in the presence of Mg^{2+} an IC_{50} of 65µM (±

1.23mM) is noticed, which indicates a $FCIC_{50}$ shift of 323 fold. This indicates that CA is far less potent in the presence of Mg^{2+} than it is in the presence of Mn^{2+}. IN_{Sol} activity is significantly higher in the presence of Mn^{2+} ions compared to Mg^{2+} (Figure 3.15 B).

It appears that CA was similarly affected by the metal ions as the DKAs and suggests binding at the active site. This metal dependent effect of the DKAs was noticed in the IN_{WT} [128], and since the IN_{WT} was active in the ELISA, a comparison of metal ions was determined in this system (Figure 3.16 A, and Table 3.3). The metal ion effect against IN_{WT} was not as prominent as what was noticed for the IN_{sol}, showing a 3.1 fold higher IC_{50} in the presence of Mg^{2+} compared to Mn^{2+}. The ALLINIs were also shown to have a 4 fold diminished inhibition if IN was incubated with DNA prior to addition of inhibitor [168]. When CA was added after IN/vDNA assembly, inhibition was not reduced in the presence of Mg^{2+}. However inhibition was reduced by 4 fold in the presence of Mn^{2+}.

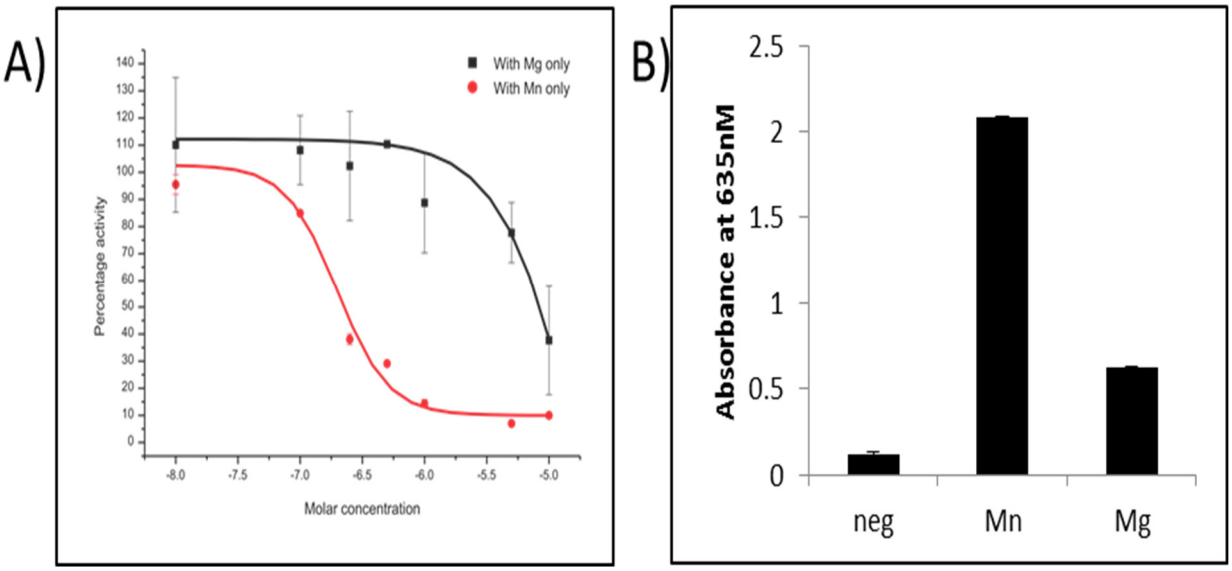


Figure 3.15: ELISA data of IN_{Sol} with different metal cofactors. A) Inhibition of IN_{Sol} by CA is preferred in the presence of Mn^{2+} compared to Mg^{2+} as indicated by an $FCIC_{50}$ of 323. ST activity was determined by ELISA, where either $MnCl_2$ (red) or $MgCl_2$ (black) was added to the reaction buffer. B) Activity of IN_{Sol} in the presence of Mn^{2+} was significantly higher compared to Mg^{2+}.

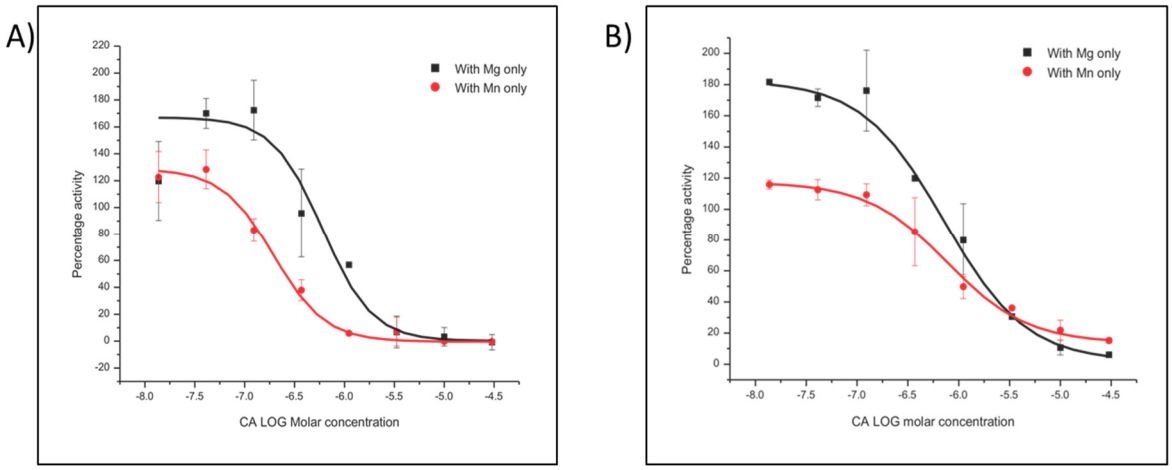


Figure 3.16: ELISA assay data of IN_{WT} with different metal cofactors and order of addition of assay components. The reduction of IN_{WT} ST activity by CA in the presence of different divalent metal cofactors. A) before DNA addition. B) After DNA addition.

Table 3.3: Summary of ST inhibition of IN_{WT} by CA with different orders of addition and different metal cofactors added. IC_{50} results are presented along with the Hill slope and $FCIC_{50}$. The same data used in Figure **3.16** was used to obtain IC_{50} concentrations. Standard deviations are represented in brackets.

	IN + CA, then DNA[a]	**Hill slope**	**IN + DNA, then CA[b]**	**Hill slope**	**$FCIC_{50}$[c]**
Reactions with Mg^{2+}	604nM (±242nM)	1.74	720nM (±266nM)	1.07	1.2
Reactions with Mn^{2+}	195nM (±28nM)	1.68	768nM (±200nM)	1.10	3.9
$FCIC_{50}$*	3.10	--	1.06	--	--

[a] IN and CA were incubated together before incubation with DNA

[b] IN and DNA were incubated together before incubation with CA

[c] Larger IC_{50} divided by smaller IC_{50}

3.4. Allosteric effects of CA on IN_{WT}

3.4.1. LEDGF-IN disruption by CA

Many ALLINIs identified thus far affect the IN/LEDGF interaction [139, 140, 162, 168-170]. AlphaScreen was used to determine the protein-protein interaction of IN_{WT}/LEDGF. Initially, before the FLAG-tagged IN was isolated, tests were conducted which used Ni-chelating donor and acceptor beads due to the IN and LEDGF both having a His-tag. CA was added at varying concentrations for IC_{50} determination in the absence of viral DNA. The dose response indicated an IC_{50} of 619 nM ±78 nM, and a positively cooperative binding with its target with a Hill slope of 2.2. This is a typical result of standard ALLINI's (Figure 3.17 A). The presence of Ni-chelating donors and acceptors is not ideal, as the usual accompanying effect of ALLINIs is that they cause multimerization of IN [139, 140, 168]. IN multimerization would bring two His-tagged IN proteins in close proximity, and if using both Ni-chelating donor and acceptor beads, the assay signal would increase and could possibly mask the effects of CA's disruption of the IN/LEDGF interaction. Using the FLAG-tagged IN and the corresponding anti-FLAG donor bead circumvents this problem by limiting the possible multimerization event to one bead group, which should not have an effect on assay signal. Indeed, when using a FLAG-tagged IN protein with anti-FLAG donor beads, an IC_{50} of 237nM was observed indicating a 2.6 fold decrease of required CA concentration (Figure 3.17 B). This confirmed our previous result and other ALLINIs, LEDGIN and Mut101 (Table 3.4), produced similar IC_{50} values in this assay compared to those documented [162, 169].

It was previously determined that CA was unable to inhibit the PIC [242, 254], and LEDGF was shown to reduce the potency of ALLINIs during early round viral infection [179]. The effects of CA on the IN/LEDGF interaction in the presence and absence of viral DNA was determined. Here a synthesised LTR homologue was added before and after CA incubation. When the DNA was added after CA incubation the IN/LEDGF interaction was disrupted. When DNA was added before CA incubation the interaction was similar to the positive control (Figure 3.18). This indicates that CA's inhibition of the IN/LEDGF is dependent on the order of addition, similar to other ALLINIs [168].

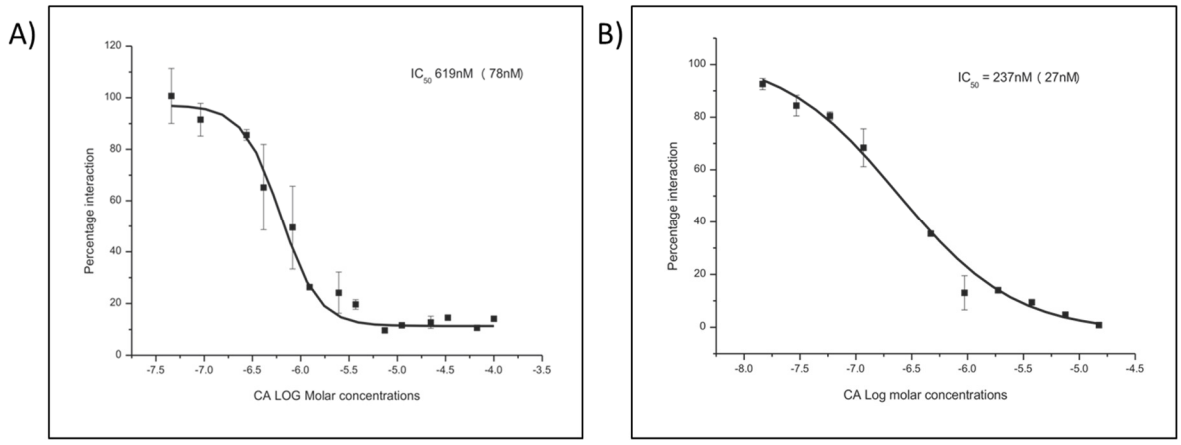


Figure 3.17: LEDGF interaction with IN as determined by AlphaScreen. A) The LEDGF/IN interaction was disrupted in a dose response fashion by CA. The graph represents percentage inhibition of the p-p interaction as CA concentrations are increased. Using both nickel chelating donor and acceptor beads, the IC_{50} was calculated to be ≈620nM with a cooperative Hill slope of 2.2. B) The assay was retested using FLAG-tagged IN_{WT} and anti-FLAG donor beads. The IC_{50} dropped to 237nM with a Hill slope of 0.83, showing a non-cooperative relationship.

Table 3.4: IN/LEDGF interaction disruption by chicoric acid and LEDGIN compounds. The table shows a summary of AlphaScreen assay results with CA, LEDGIN and Mut101 using anti-FLAG beads and FLAG-tagged IN_{WT}. Standard deviations are shown in brackets.

Compound	**IC_{50}**
Chicoric acid	237nM (±27)
LEDGIN	1600nM (±684)
Mut101	16nM (±222)

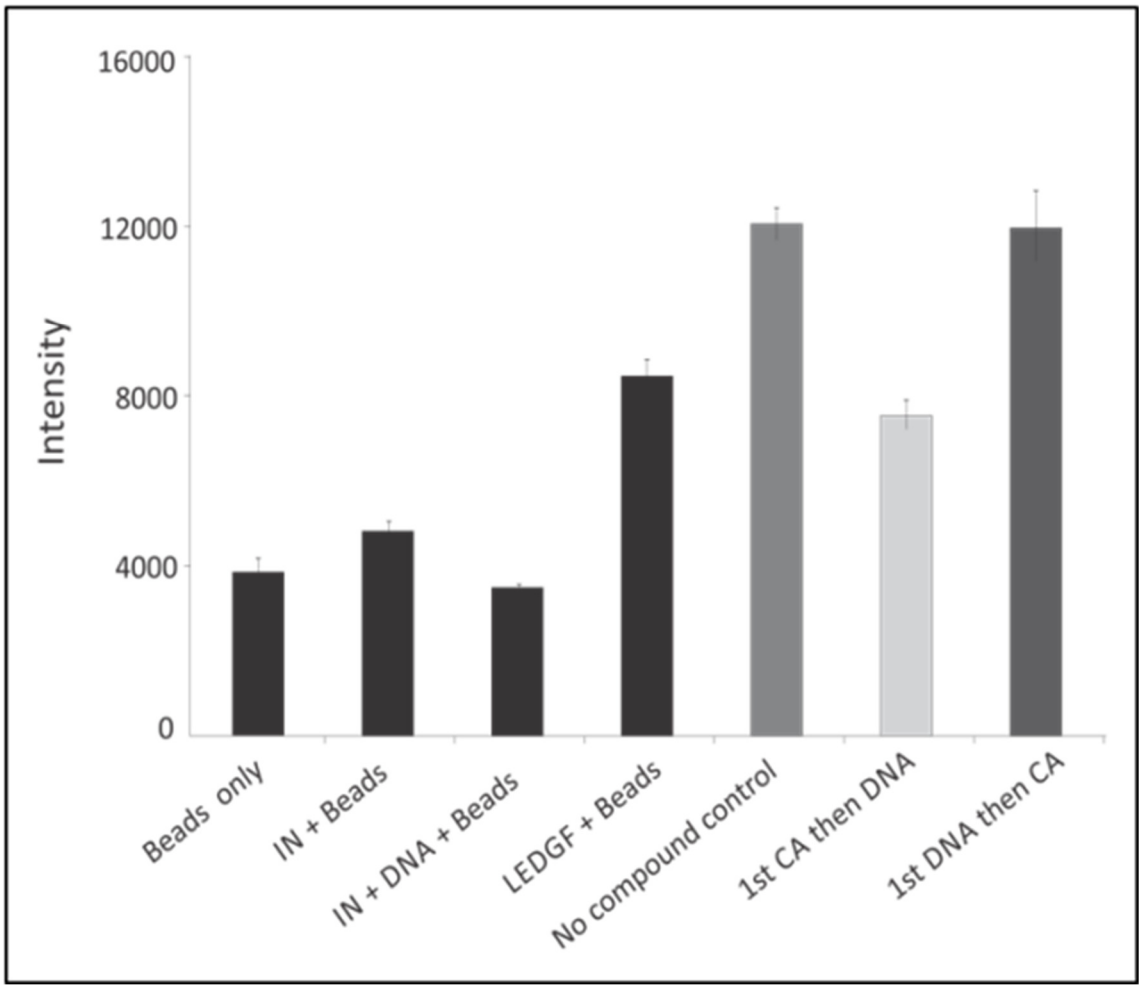


Figure 3.18: Disruption by CA is affected by order of addition in the AlphaScreen assay. Various negative controls were included, which contained components as indicated as well as reaction buffer. These are represented in the darkest shade. A no compound (DMSO) positive control had the highest intensity. If CA was added before vDNA, the LEDGF/IN interaction was disrupted, to a comparable level of the LEDGF+Beads negative control. IF vDNA was added first then CA did not disrupt the LEDGF/IN interaction. There was a significant difference ($p<0.05$) as determined by students p test between negative control (LEDGF+beads) and the positive (no compound control). Likewise, there was a significant difference between adding CA before DNA (1st CA then DNA) compared to adding CA after vDNA incubation (1st DNA then CA).

3.4.2. SPR confirms CA interaction with IN_{WT} and CCD

The binding of CA to the IN_{CCD} has either been assumed due to inhibition of catalytic assays [238] or by modelling [237, 245]. Therefore, binding kinetics of CA to IN_{WT} and/or IN_{CCD} was determined and compared to LEDGIN. SPR confirmed binding of CA to the IN_{CCD} and IN_{WT} His-tagged proteins (Figure 3.19) and kinetic data are summarised in Table 3.5. IN_{CCD} contains the active site, but also has the dimer-dimer interface and LEDGF binding site. Langmuir distribution was used to calculate the kinetic data of the compounds. LEDGIN bound to both the IN_{CCD} and IN_{WT} with similar kinetics, the affinity of LEDGIN for IN_{WT} was higher than the IN_{CCD} leading to a slightly lower K_D of 3.5µM on IN_{WT} compared to 7.6µM on IN_{CCD}. Similarly, CA bound to both the IN_{WT} and IN_{CCD} with comparable kinetics, and similar K_D of ≈2.3µM between the proteins. However, comparing CA binding kinetics to LEDGIN, the compounds behave differently. LEDGIN quickly binds to its target and saturates its binding site with its k_{on} at $1.9x10^4$ 1/Ms for IN_{CCD} and $6x10^4$ 1/Ms for IN_{WT}. This was an entire order of magnitude greater than CA, which had a k_{on} of $1.3x10^3$ 1/Ms for IN_{CCD} and $1.19x10^3$ 1/Ms for IN_{WT}, without reaching saturation at the concentrations used. LEDGIN also dissociates from its binding site at a rapid rate, with k_{off} at two orders of magnitude greater than CA at $1.5x10^{-1}$ 1/s and $2.1x10^{-1}$ 1/s for IN_{CCD} and $IN_{WT,}$ respectively. CA appears to strongly bind to IN_{CCD} and IN_{WT} with k_{off} of $3.4x10^{-3}$ 1/s and $2.5x10^{-3}$ 1/s respectively, which has been suggested previously [236, 239]. The dissociation of CA from the protein is about 1000 times slower than LEDGIN, and may stabilize and lock the protein in a specific conformation. Dissociation constants, K_D, are not very different between the two compounds, however, CA has a slightly lower K_D of 2.6 and 2.1 µM for IN_{CCD} and IN_{WT} respectively, compared to LEDGIN K_D of 7.6 µM for IN_{CCD} and 3.5 µM for IN_{WT}. CA has a tighter binding once bound, as judging from the kinetics profile and K_D. Direct binding of CA to the IN_{WT} with a similar binding kinetics as compared to IN_{CCD} indicates that the catalytic core domain is likely the target binding domain of CA.

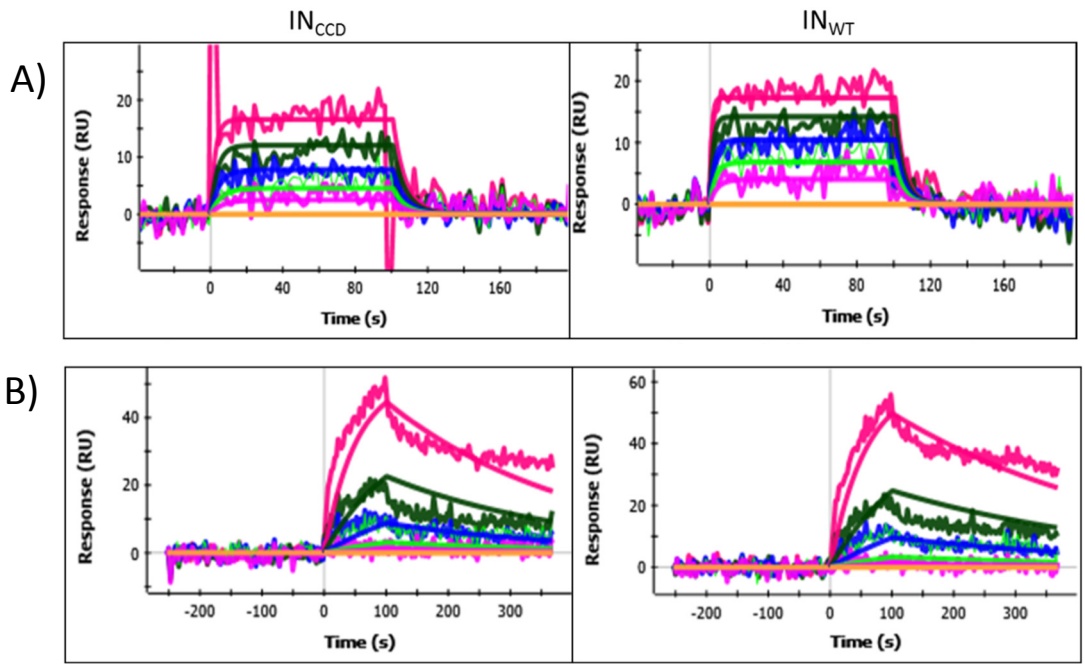


Figure 3.19: Binding of CA was confirmed by SPR analysis. His-tagged IN_{CCD} and IN_{WT} was immobilized on HGC Proteon chips and compounds at varying concentrations were injected for 100 seconds. A) LEDGIN and B) CA tested against IN_{CCD} and IN_{WT}, respectively. Langmuir distribution (1:1) was used to determine the kinetics and is summarised in Table **3.5**.

Table 3.5: List of kinetic data determined by SPR for IN binding of LEDGIN and CA.

Protein	Kinetics (units)	LEDGIN	CA
IN_{CCD}	k_{on} (1/Ms)	$1.9x10^4$	$1.3x10^3$
	k_{off} (1/s)	$1.5x10^{-1}$	$3.4x10^{-3}$
	K_D (µM)	7.6	2.6
	Rmax	26.68	62.66
	χ^2	3.04	15.35
IN_{WT}	k_{on} (1/Ms)	$6.0x10^4$	$1.19x10^3$
	k_{off} (1/s)	$2.1x10^{-1}$	$2.5x10^{-3}$
	K_D (µM)	3.5	2.1
	Rmax	22.11	70.50
	χ^2	3.31	15.22

3.4.3. The induction of IN multimerization by CA

Central to most ALLINI's MOA is their ability to induce multimerization [139, 140, 168]. To determine the induction of higher order IN oligomers by CA, BS3 crosslinking studies, SDS-PAGE and Western blot were performed. Compounds were incubated with IN_{WT} for 30 min unless otherwise stated. Initially IN_{WT} was tested over varying concentrations of enzyme against a fixed 200µM concentration of compound (Figure 3.20). High order oligomer formation was affected by the enzyme concentration, and seemed to naturally occur in high abundance at IN_{WT} concentrations greater than 2.5µM (Figure 3.20 A). At lower concentrations of enzyme, i.e. less than 0.63µM, the monomer is predominant. After adding LEDGIN at 200µM, the high order oligomers increase across all enzyme concentrations, but is however masked at greater than 2.5µM IN due to concentration dependent oligomerization and/or aggregation (Figure 3.20 B). A similar occurrence is noticed in the presence of 200µM CA (Figure 3.20 C). Comparing the formation of oligomers at IN_{WT} concentrations equal to and less than 1.25µM, between the negative control, LEDGIN and CA Western blots, there was more oligomer formation in the presence of the compound. This was further illustrated with intensity graphs of selected lanes (Figure 3.21), where at a 160nM IN_{WT} concentration, both LEDGIN and CA caused the rapid formation of higher order oligomers with no visible dimer transition. At 1.25µM IN_{WT} the graphs indicate the negative control has most forms in the higher order oligomers. In contrast, LEDGIN and CA seem to cause an ordered transition from monomer, to dimer, then to tetramer which was the most abundant form, and then finally moves to higher order oligomers. Lastly, at 5µM IN_{WT} there are no noticeable differences between negative and compound tested lanes; indicating the masking of the compound effects by high IN_{WT} concentrations. With possible exception, the curve shape left of the dimer peak in the CA panel slopes downward, and may represent a gradual transition of dimer towards higher order oligomers as affected by CA.

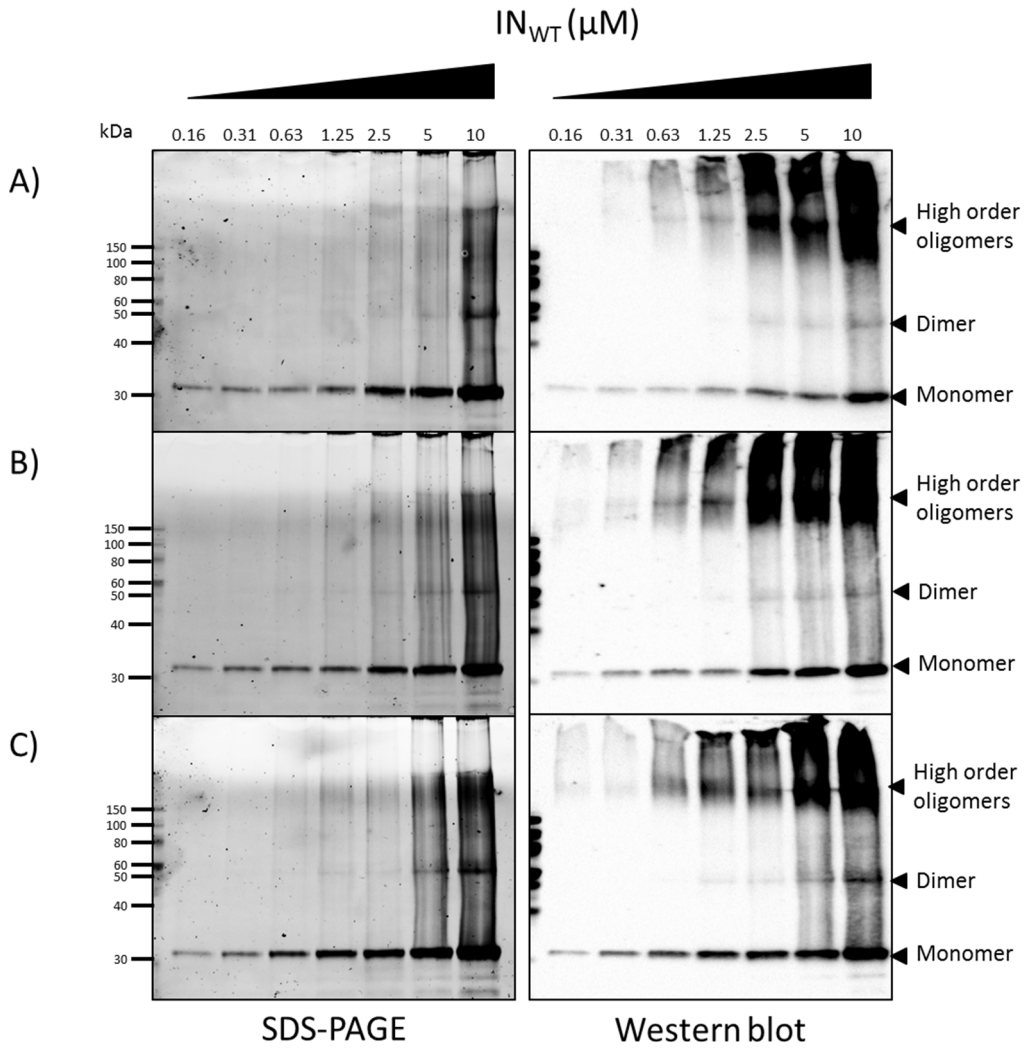


Figure 3.20: IN multimerization is affected by ALLINIs. IN$_{WT}$ was added at varying concentrations as indicated. SDS-PAGE and Western blot analysis was compared between A) DMSO blank, B) LEDGIN at 200µM and C) CA at 200µM. Exposure times for the SDS-PAGE and the Western blot was the same for each representation. Western blot analysis indicates that the IN$_{WT}$ undergoes natural concentration dependent oligomerization. This is affected by LEDGIN and CA with a shift in oligomer forms to higher order between 160nM and 1.25µM IN$_{WT}$ concentrations as compared to the negative. There after the natural oligomerization of IN overshadows the effect of the compounds. MW marker sizes (in kDa) are indicated on the left. The arrows show the different IN forms.

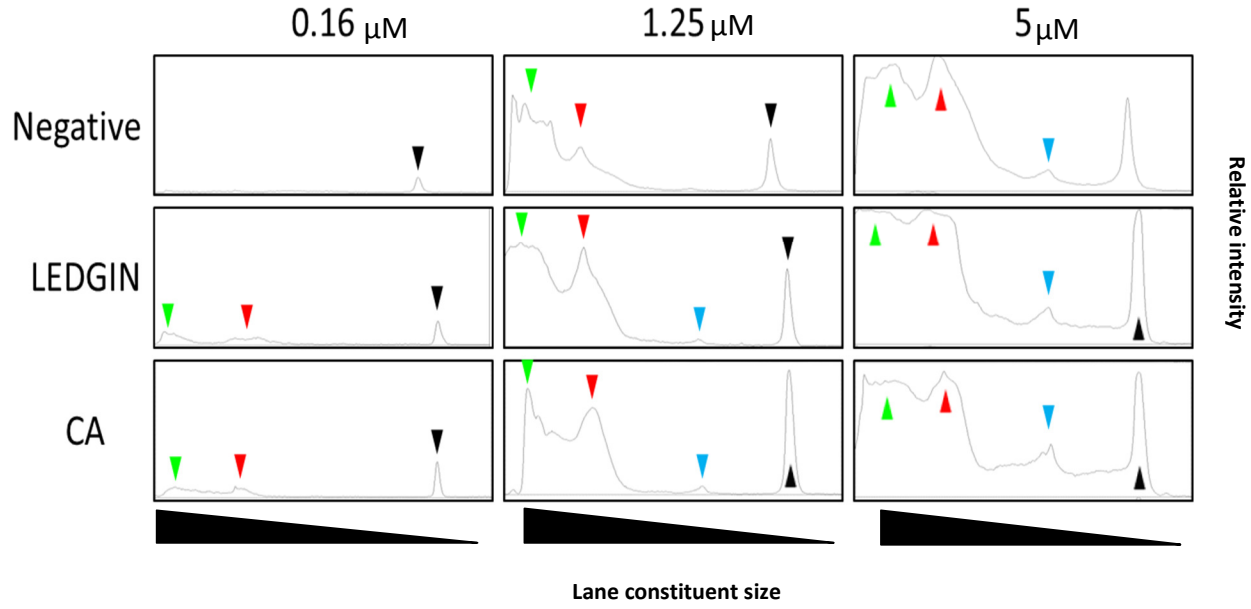


Figure 3.21: The actions of the compounds are affected by enzyme concentration as determined by band intensity (pixel densitometry) graphs. Graphs were obtained using ImageJ. Columns represent lanes from the Western blot analysis of Figure **3.20**. and are inspected from right to left, with smallest lane constituents towards the right and largest to the left margin of the graph. Arrowhead labels are: black for the monomer, blue for dimer, red for tetramer and green for higher order oligomer/aggregates. At 160nM IN_{WT} both LEDGIN and CA cause for the rapid formation of tetramer and higher order oligomer with no dimer transition visible. At 1.25µM IN_{WT} the negative control has undergone normal concentration dependent oligomerization and aggregation. LEDGIN and CA appear to cause oligomerization even at low, 0.16µM IN concentrations. Peak intensities should not be compared between graphs, but only between peaks in the same graph.

The concentration of IN may mask the appearance of oligomerization in response to the compounds. This then requires the use of low enzyme concentrations that are close to the limit of detection in the western blot. This was similar to what was experienced by Kessl *et al.* [175] where a subunit exchange experiment required low IN concentrations to be viable. To aid in the solubility of the enzymes, 0.1% Igepal detergent was added to reactions (Figure 3.22). This allowed for the detection of IN at 1.25µM and at 10µM that had less high order forms compared to the previous result. With the 1.25µM IN_{WT}, 200µM of both CA and ACA-1 increased oligomer formation compared to the negative control. LEDGIN however decreased overall protein amount of the reaction well. In contrast, at 10µM IN_{WT}, the presence of compounds did not have a notable increase in IN_{WT} oligomer formation. It was later realized that detergents cause anomalous oligomerization of IN [106, 268] which may account for the disparity of the results.

We therefore removed detergents from the reaction buffer and focused on increasing the sensitivity of detection in the western blot. The lack of dimer detection at low protein concentrations (Figure 3.20, Figure 3.21 Figure 3.22) was due to the low sensitivity of the 1:1000 anti His primary antibody. Anti His was not sensitive enough to detect IN_{WT} at these concentration and therefore the antisera developed against the 23-34 amino acid region of HXB2 IN was used for detection. IN_{WT} was tested at 300nM (Figure 3.23) with adequate detection using this antibody. All compounds tested, i.e. LEDGIN, Mut101, CA and ACA-1 increased the IN_{WT} oligomerization in a compound concentration dependent manner.

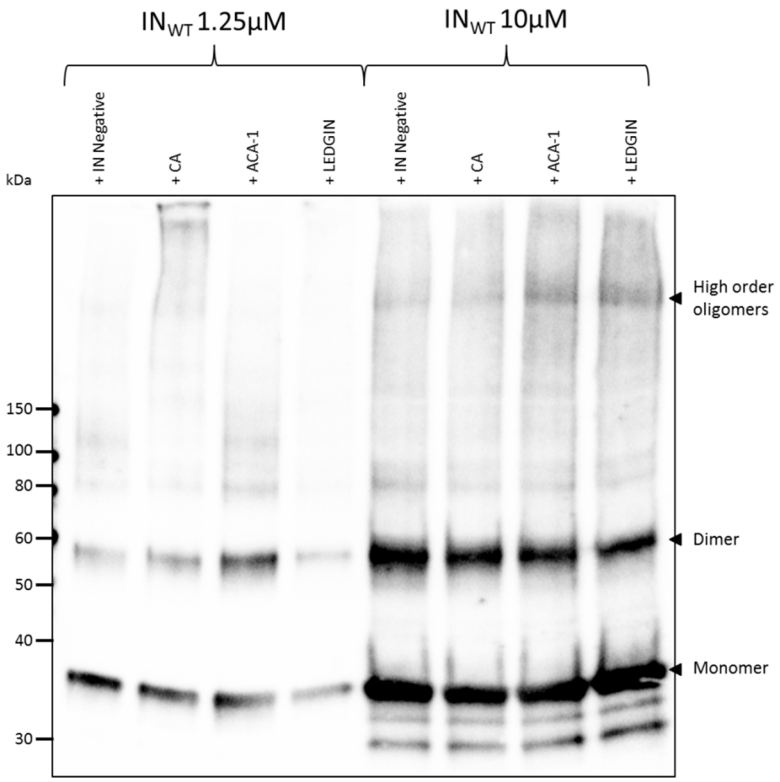


Figure 3.22: Western blot of IN$_{WT}$ depicting protein concentration dependence on the formation of multimers. Two concentration, 1.25µM and 10µM, were compared. At 1.25µM IN: a similar result for CA was obtained with prominent increases in higher order oligomers compared to the negative control; ACA-1 caused an increase of dimer, trimer and tetramer forms; while LEDGIN completely reduced the protein, which could be a result of rapid aggregation, or it may bury the anti-His epitope under certain conditions. At 10µM IN there was no distinguished differences between lanes. All compounds were tested at 200µM each. MW marker sizes (in kDa) are indicated on the left. The arrows show the different IN forms.

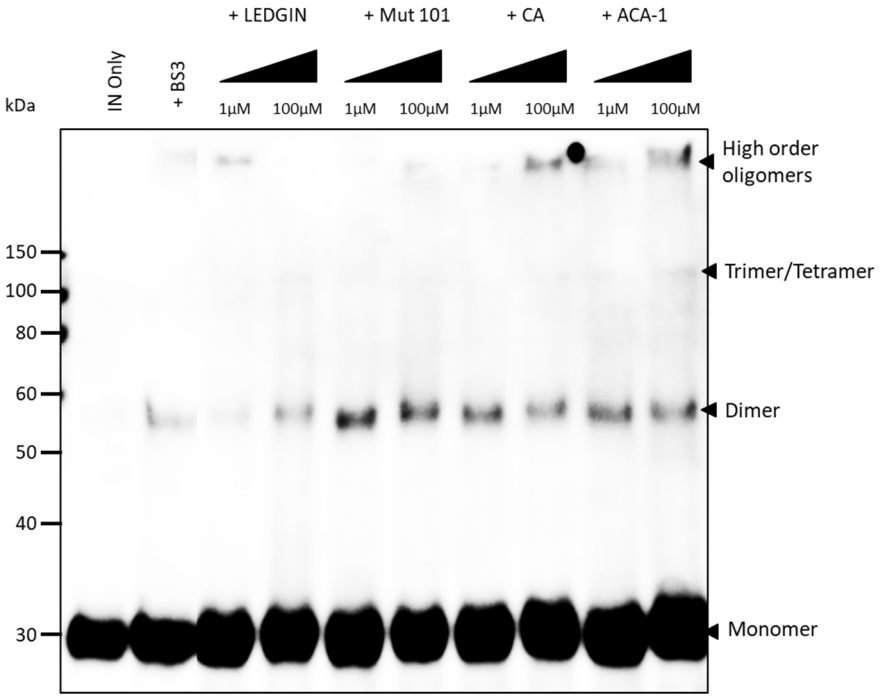


Figure 3.23: Western blot of IN_{WT} multimeric forms using IN antisera. Ideal concentration for IN testing in this assay was determined at 300nM $IN_{WT.}$ In this Western blot assay, IN_{WT} was incubated for 30 minutes with compounds at concentrations indicated. The negative control with cross-linker added (+BS3) added is used for comparisons. An increase in multimerization was experienced by all ALLINIs, and they seemed to increase the multimers in a dose response manner. This was also true for CA, which had a similar profile compared to Mut101 and ACA-1. MW marker sizes (in kDa) are indicated on the left. The arrows show the different IN forms.

The IN_{A128T} was tested at a 150nM concentration of protein and at 100µM concentration of each compound (Figure 3.24). The IN_{A128T} mutant is resistant to multimerization induced by ALLINIs [142], but of the compounds tested, LEDGIN was the only one found to be affected by this resistance [162], while Mut101 was fully effective against IN_{A128T} [169], and CA and ACA-1 are undetermined in the literature. Here in this study, it is noticed that Mut101 and CA induce comparable dimer formation, which is higher than that of LEDGIN (Figure 3.24). ACA-1 is the best multimerizer against IN_{A128T} and is likely due to its distinct binding mode compared to the other compounds [175].

Multimerization of IN_{WT} under native conditions was monitored by SEC using an analytical TSKgel G3000SWXL column that can separate proteins between 10 and 600 kDa. IN_{WT} concentration (Figure 3.25) was 10µM final concentration per reaction, and reactions were incubated for 30 min with compounds indicated. CA increased the amount of higher order oligomers and reduced the monomer as compared to a DMSO negative control. LEDGIN and ACA-1 both decreased the amount of monomer with no apparent increase in high order multimers. Aggregation caused by the compounds could remove IN_{WT} from solution, and reduce the amount of protein entering the column, which may limit detection. This was a significant problem with the HPLC experiments, especially with LEDGIN, and has been similarly described by other sources [142, 180]. CA has shown to cause a more stable intermediary oligomer formation leading up to aggregation, and allows the detection of these intermediates (Figure 3.25).

Virus produced in the presence of ALLINIs have shown an increase in multimerization of IN [140, 253]. Integrase from virus produced in HEK293T cells in the presence of CA, ACA-1 and Mut101 compounds was detected by non-reducing SDS-PAGE and western blot analysis (Figure 3.26). An increase of multimerization was noticed for each compound, while absent in the DMSO control transfection. There was some cross-rectivity of the IN antisera with suspected cellular factors, as is noticed in the non transfected, no DNA control. The size of the multimers were higher than that expected for dimers, but similar size ranges were observed elsewhere and described as dimers [140].

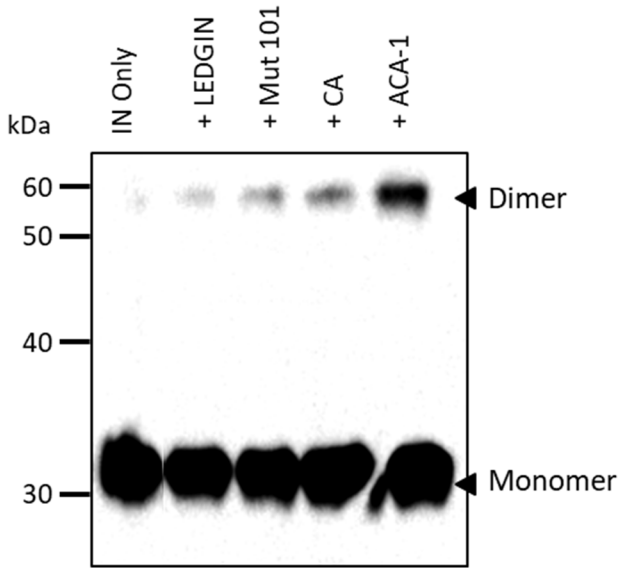


Figure 3.24: Western blot depicting IN_{A128T} multimeric formation in the presence of 100µM of indicated compounds. The IN_{A128T} was more resistant to multimerization and higher order forms, such as tetramers, were not observable. The dimeric forms were detected and these could be assessed to compare multimerization. All lanes had BS3 cross-linker added. The negative control (IN only) lane had virtually no dimeric forms of IN. Pixel densitometry was used to compare lanes. The LEDGIN test indicated the lowest increase in dimer, while the Mut101 and CA tests indicated comparably modest increases in dimer. ACA-1 caused the highest increase in dimer formation. MW marker sizes (in kDa) are indicated on the left. The arrows show the different IN forms.

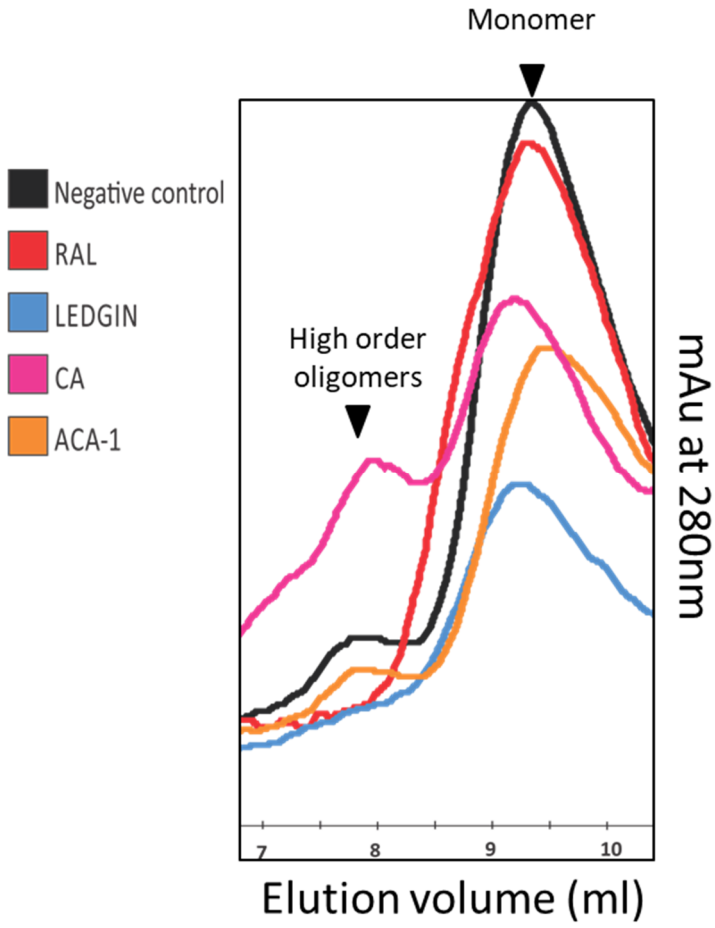


Figure 3.25: IN multimers were formed at native conditions in the presence of CA. SEC analysis was performed using 10µM IN_{WT} and 10µM compound. CA caused for the formation of stable High order IN_{WT} forms, while reducing the monomer. ACA-1 and LEDGIN similarly reduce the monomer, however no stable high order forms were noticed, and may have been removed from solution by aggregation. RAL was similar to the Negative control, and had no reduction in monomer, with no high order forms. The HPLC column used for the SEC was calibrated with the high and low molecular weight gel filtration kit standards (see appendix).

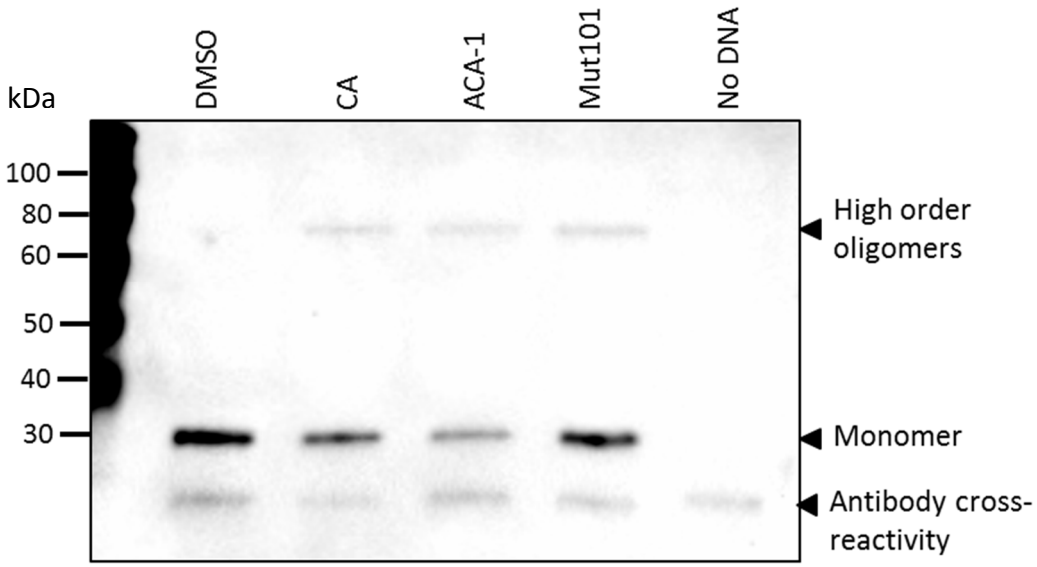


Figure 3.26: Anti IN western blot showing dimer formation in virus produced in the presence of compounds. HEK293T cells were transfected with a viral vector pNL4-3 and maintained in either 2% DMSO or 10µM compound of equivalent DMSO volume. A non-transfected (No DNA) control was included. Higher amounts of IN is noticed in virus produced in only DMSO, yet no higher order oligomers are present in this lane. The CA and ACA-1 tests both presented with increases in a band at ≈70 kDa, which is slightly larger than expected for dimeric IN and likely as a result of non-reducing conditions. This was comparable to the Mu101 ALLINI control. The non-transfected control (No DNA) indicates cross-reactivity of the IN antisera with possible HEK293T cell or media constituents. MW marker sizes (in kDa) are indicated on the left. The arrows show the different IN forms.

3.5. Modelling the structural interactions of CA with the CCD

3.5.1. Docking Pose selection

A Monte Carlo based docking approach was used to determine the best binders. Firstly, the site directed LibDock function [269] was used to generate random ligand conformations at the defined binding site. Thereafter, local energy minimization of the LibDock scored ligand poses, for each binding site, was performed with a CHARMm force field. The lowest energy poses of this final minimization would be the closest to native binding [270, 271], as far as predictive modelling would allow. The LibDock scores of the lowest binding scores, and thus most native binders, were then used to compare the docking at the various binding sites.

3.5.2. *In Silico* docking pose and binding site selection

Molecular docking is a theoretical approach and may indicate possible binding modes of a compound. Docking of CA, TCA and ACA-1 into IN_{CCD} crystal models, 1QS4 and 2B4J, returned hits for the active site and the allosteric LEDGF IBD interaction site, referred to simply as the allosteric site. CA was docked at various poses using LibDock scoring, and subsequent minimizations until the lowest binding energy poses where obtained. LibDock scoring gives an indication of the fitting and a higher score indicates a superior binder. The docked poses were initially ranked according to their CHARMm minimised binding energies, which give an indication of natural bound conformations, and the lowest binding energy pose was selected for each binding site. The LibDock score and Ligand RMSD of these poses were compared in Table 3.6. The Mg^{2+} coordinated active site of 1QS4 has been previously used to determine a theoretical binding site for CA [244, 245], and thus used for active site binding comparisons. The allosteric site of 1QS4 represents a free LEDGF interaction site of the IN_{CCD} [83], while the 2B4J is representative of an IN_{CCD} with interacting LEDGF [260]. LEDGF bound at this site may lead to significant changes, and therefore, the allosteric binding sites from both structures, an “closed” and “open” site, were used for comparison. All interacting amino acids, inclusive of hydrogen(H)-bonds and electrostatic interactions, are summarised in Table 3.7. The compounds used are represented in

Table 2.2 and differences described are: the CA has 1,2-catechol groups, while TCA and ACA-1 have a 1,2-diacetylbenzene group instead. The CA and TCA both have a double carboxylic acid central linker (CACL), while ACA-1 has a single methyl group with ester linker (MEL).

3.5.3. Docking at the active site

The lowest energy active site binding poses of the three compounds are depicted in Figure 3.27. One arm of CA appears to fit into the active site with its catechol having a direct H-bond interaction with the D116 catalytic triad residue (Figure 3.27, Left panel). The CACL region has interactions with K159 and K156, among others (Table 3.7). CA's other caffeic acid arm appears to fit into an adjacent active site groove, and notably has a hydrogen bond interaction between its catechol and Q148. CA coordinates the Mg^{2+} through a Pi-cation stack with a benzene ring. There are significant electrostatic interactions that stabilize all functional groups of CA at the active site indicated by pink and green shading in 2D diagrams (Figure 3.27, left panel, 2D interactions).

TCA's lowest energy binding pose at the active site shows a hydrogen bond interaction between a diacetylbenzene group and K159 (Figure 3.27, middle panel). A single benzene ring of TCA coordinates the Mg^{2+} ion through a Pi-cation stack with D116. The CACL region of TCA had an H-bond interaction with N117. TCA had significant electrostatic interactions between its diacetyl groups and the protein, however, its CACL is exposed to solvent as indicated by the dark shading around the N117 and lack of electrostatic interaction in the 2D diagram (Figure 3.27, middle panel). ACA-1 interacted with the protein residues with only one of its diacetylbenzene groups, interacting via an H-bond with H114 and an H-carbon bond with T115 (Figure 3.27, right panel). The methyl ester linker (MEL) region and ACA-1's second diacetylbenzene group are significantly exposed to solvent, as indicated by the dark shading around D116 and lack of surrounding electrostatic interactions. The involvement of the catechol and diacetylbenzene groups of each compound, indicate their importance for the interaction at the active site, and its placement in a grove formed around T115 is common.

Using the LibDock score to rank the best fitted compounds, CA had the highest with 76.014, TCA was second with 74.941 and ACA-1 was lowest with 72.426, and all had an RMSD less than two (Table 3.6). The LibDock scoring at this site was the lowest compared to the allosteric binding site, indicating that the active site may be the least favourable binding site.

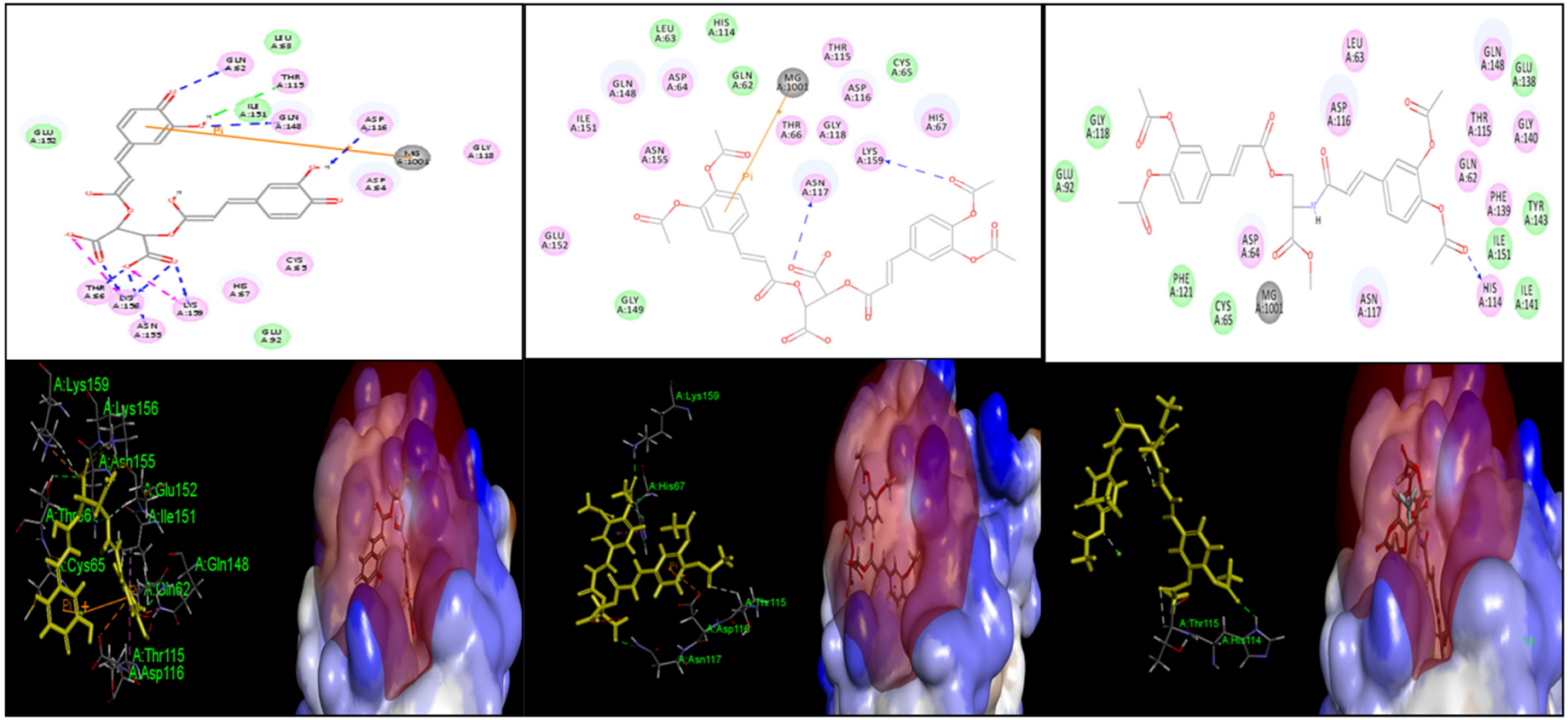


Figure 3.27: Docking of CA (Left), TCA (Middle), ACA-1 (Right) at the active site of 1QS4. The binding site was chosen around residues K159, K156, E92, Q148, E116. The upper-panel depicts the 2D interactions of each compound. The bottom panel is divided into left and right for each compound, showing the 3D interactions with residues labelled and the hydrophobic surface representation of the protein with compound bound. Interactions are detailed in the text section 3.5.3. (See appendix for enlarged panels.)

*Interactions: Green dashed line: H-bonds with amino acid main chain. Blue dashed line: H-bond with amino acid side chain. Magenta dashed line indicates charge interactions. Orange dashed line: Pi-cation. Purple dashed line: Pi-stacking. Red: Unfavourable bumps

*2D colouring of amino acids: Pink indicates=electrostatic interaction. Green=van der–Waals molecular interaction. High shading indicates more exposure to solvent.

3.5.4. Docking at the allosteric site of 1QS4 compared to 2B4J

3.5.4.1 ***CA***

The allosteric binding site was compared between the non-bound 1QS4 and LEDGF bound 2B4J structures (Figure 3.28 and Figure 3.29). CA docked into the 1QS4 allosteric site with a single catechol and CACL region significantly buried (Figure 3.28, left panel). The other catechol is exposed, however, it is stabilized by its H-bond interaction with chain A T125. Notable interactions are with LEDGF interaction chain A: A128, T125 and chain B: H117, M178 residues. The catechol stretches further in the crevice to reach H171, K173 and Q95. CA's positioning at the 2B4J allosteric site is not as buried as at the 1QS4 site (Figure 3.29), which is expected due to the "open" conformation of the LEDGF bound 2B4J allosteric site. Here, CA is anchored by a hydrogen bond between the CACL and the protein residues chain B: A128, T125. The catechol arms of CA are exposed to solvent, but involved in stabilizing interactions with chain A residues via the H- bond to E170 and π-cation with H171. A Pi-stacking interaction between the benzene rings and residue chain B: Y131 stabilize the other arm. Electrostatic forces are prevalent mainly around the CACL, and at least one catechol arm in both poses (Figure 3.28 and Figure 3.29, left, 2D representation).

3.5.4.2 ***TCA***

The TCA docking at the 1QS4 allosteric site predicted an H-bond interaction between the CACL and chain A: K127, Y132 and chain B: Q168 (Figure 3.28, middle panel) and is further stabilized by charge-charge and H-carbon interactions with chain A: K127, A128 respectively. A single diacetylbenzene has an H-carbon interaction with chain A: Q95, while its benzene ring is stacked between a π-anion interaction with chain B: E170 and Pi-Alkyl bonds with chain A: A98, A128 to further stabilize the arm in the allosteric binding site crevice. The other diacetylbenzene is folded towards the binding site by its H-carbon bond with chain B: E170. Similar binding is predicted for TCA in the 2B4J allosteric site, with H-bonds between the CACL region and Q95, H171, T174 and carbon-hydrogen interactions with K173 (Figure 3.29, middle panel). A single benzene ring is involved in Pi-stack interactions with chain B: residues A98 and A128, and an H- carbon with A129. The other diacetylbenzene is exposed to solvent with an unfavourable acceptor-acceptor distance from chain A: D167. Electrostatic forces are

mainly prevalent around the CACL and only one arm of TCA, while the other arm is exposed to solvent as is noticed by the dark shading around D167 and Q168 in the 2D diagram (Figure 3.28, middle panel, 2D diagram)

3.5.4.3 ***ACA-1***

ACA-1 binding pose at the 1QS4 allosteric site indicated significant interactions with one diacetylbenzene arm and MEL region, however the other arm is significantly exposed to the solvent (Figure 3.28, right panel). Notable interactions with the MEL and buried chain A: E96, Y99, T125. The diacetylbenzene of one arm stretches the length of the allosteric site and forms a π-stack with chain B: E170, M178 and chain A: A128, A129. The second arm is significantly exposed to solvent, with no important interaction with protein residues. Conversely, the 2B4J allosteric site pose of ACA-1 is not deeply buried in the binding site crevice (Figure 3.29, right panel). The MEL region interacts with the solvent exposed chain A:Q168, while one diacetylbenzene interacts with solvent exposed chain B: Y131. The acetyls of the other catechol are the only groups buried, forming H-bonds with Q99 and T174.

3.5.4.4 ***LibDock score comparisons***

The LibDock scores for CA and TCA were nearly indistinguishable for both structures at the allosteric site, however the ACA-1 poses were ≈30 and ≈10 points less for the 1QS4 and 2B4J poses, respectively (Table 3.6). LibDock scoring for CA and TCA was higher in the 1QS4 site compared to the 2B4J site, indicating that CA and TCA binding at this site is more efficient in the absence of LEDGF binding. ACA-1 had a similar LibDock score, and indicates no preference of the binding pose for either structure.

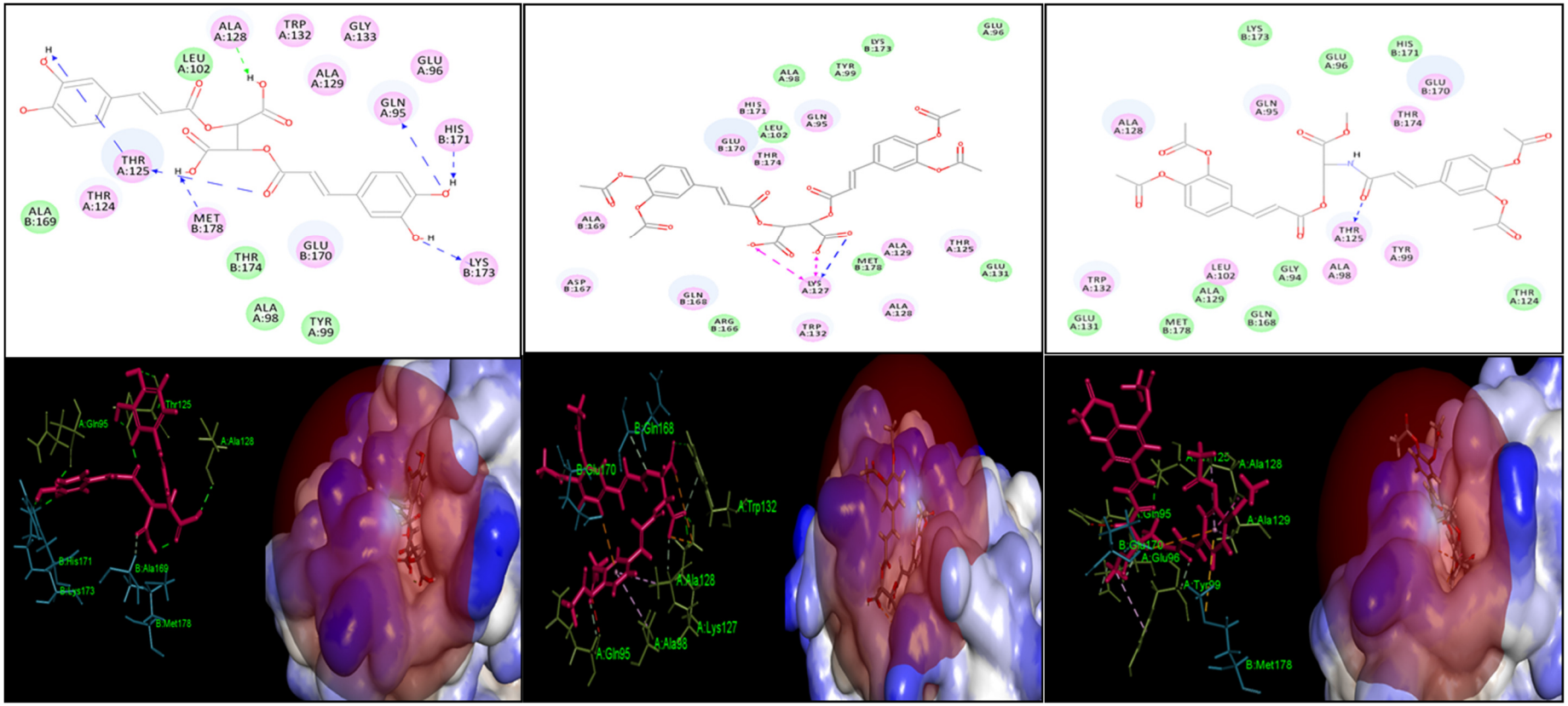


Figure 3.28: Docking of CA (Left), TCA (Middle), ACA-1 (Right) at the allosteric site on 1QS4. The binding site was chosen around residues Chain A: T125, D131 Chain B: D167, H171, K173. The upper-panel depicts the 2D interactions of each compound. The bottom panel is divided into left and right for each compound, showing the 3D interactions with residues labelled and the hydrophobic surface representation of the protein with compound bound. Interactions are detailed in the text section 3.5.4. (See appendix for enlarged panels.)

*Interactions: Green dashed line: H-bonds with amino acid main chain. Blue dashed line: H-bond with amino acid side chain. Magenta dashed line indicates charge interactions. Orange dashed line: Pi-cation. Purple dashed line: Pi-stacking. Red: Unfavourable bumps

*2D colouring of amino acids: Pink indicates=electrostatic interaction. Green=van der–Waals molecular interaction. High shading indicates more exposure to solvent.

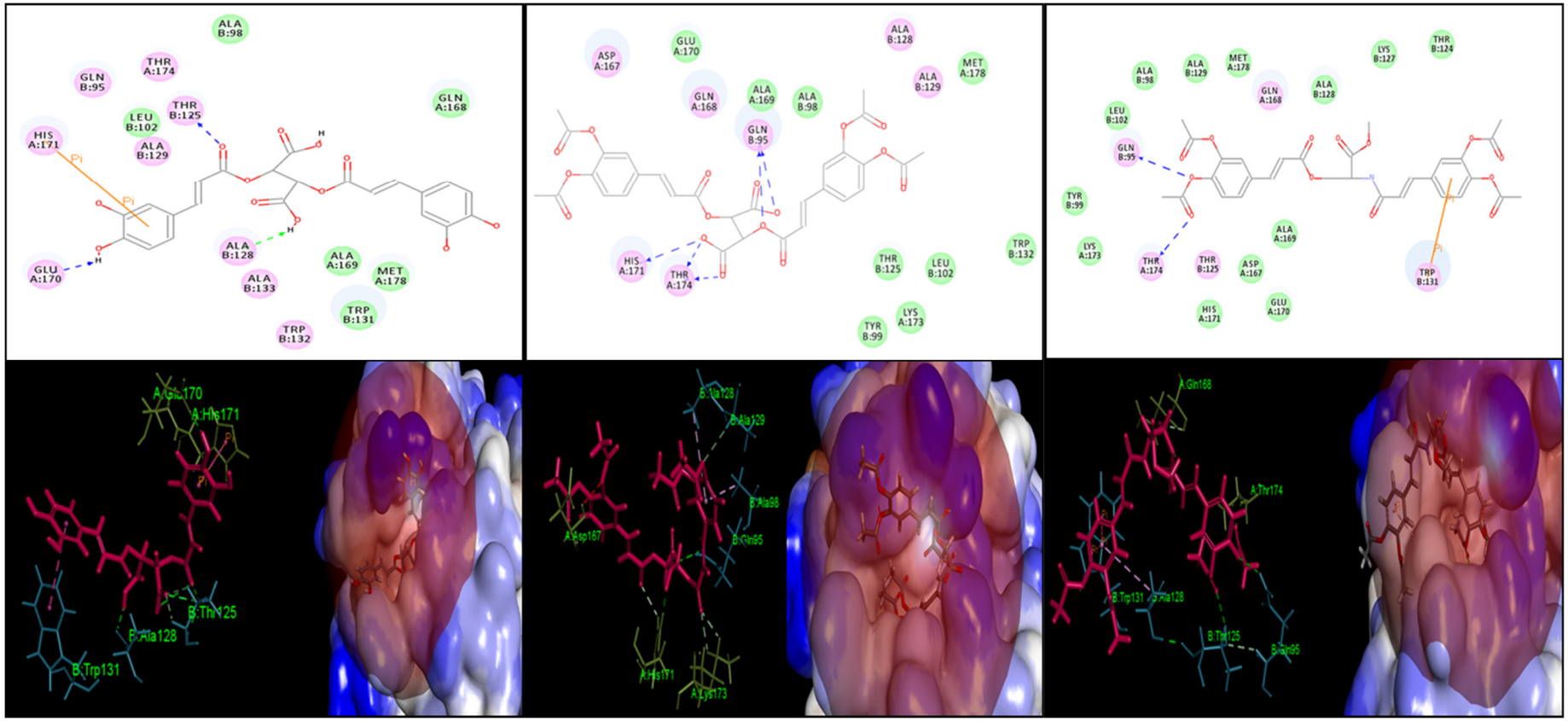


Figure 3.29: Docking of CA (Left), TCA (Middle), ACA-1 (Right) at the allosteric site on 2B4J. The binding site was chosen around residues Chain A: T125, D131 Chain B: D167, H171, K173. The upper-panel depicts the 2D interactions of each compound. The bottom panel is divided into left and right for each compound, showing the 3D interactions with residues labelled and the hydrophobic surface representation of the protein with compound bound. Interactions are detailed in the text section 3.5.4. (See appendix for enlarged panels.)

*Interactions: Green dashed line: H-bonds with amino acid main chain. Blue dashed line: H-bond with amino acid side chain. Magenta dashed line indicates charge interactions. Orange dashed line: Pi-cation. Purple dashed line: Pi-stacking. Red: Unfavourable bumps

*2D colouring of amino acids: Pink indicates=electrostatic interaction. Green=van der–Waals molecular interaction. High shading indicates more exposure to solvent.

Table 3.6: Summary of selected *in silico* binding parameters for CA, TCA and ACA-1 at the selected binding sites on the three IN models used.

Structure and binding site	**CA**			**TCA**			**ACA-1**		
	Binding Energy (kcal/mol)	LibDock Score	Ligand RMSD to reference	Binding Energy (kcal/mol)	LibDock Score	Ligand RMSD to reference	Binding Energy (kcal/mol)	LibDock Score	Ligand RMSD to reference
1QS4 Active site	-646.454	**76.014**	1.19524	-431.76	**74.941**	0.769635	-352.933	**72.426**	0.723391
1QS4 Allosteric site	-433.188	**120.321**	0.904109	-448.248	**121.009**	0.760606	-411.698	**92.685**	0.727392
2B4J Allosteric site	-405.328	**99.886**	1.3793	-599.685	**98.006**	0.879281	-445.854	**89.694**	0.805886

Table 3.7: Amino acid interactions of CA, TCA and ACA-1 determined by docking trails. These are hydrogen bond, hydrogen-carbon and π stacking interactions only. The central linker refers to either the CACL or MEL region of the respective compounds (see section 3.5.2 for details). Each compound has a catechol group, however, TCA and ACA-1 are acetylated. The active site was only chosen for 1QS4, while the allosteric site was chosen for 1QS4 and 2B4J.

PDB Structure	**CA**		**TCA**		**ACA-1**	
	Central linker	**Catechol**	**Central liker**	**Catechol (diacetyl)**	**Central Linker**	**Catechol (diacetyl)**
1QS4 Active site	T66, E152, N155, K159, K156	Q62, C65, T115, D116, Q148, I151	N117	H67, T115, D116, K159	none	H114, T115
1QS4 Allosteric site *	Chain A: T125*, A128, M178[1] Chain B: A169	Chain A: T125* Chain B:H171, K173	Chain A: K127, A128*, W132[2] Chain B: Q168	Chain A: Q95, A98, A128* Chain B: E170	Chain A: Q95, Y99, L102 Chain B: None	Chain A: A128, A129 Chain B: E170, M178
2B4J Allosteric site *	Chain A: None Chain B: T125, A128	Chain A: E170, H171 Chain B: W131	Chain A: H171, K173[3], T174[3] Chain B: Q95	Chain A: D167 Chain B: A98, A128, A129	Chain A: Q168 Chain B: None	Chain A: T174 Chain B: Q95, T125, A128, W131

* Central Linker shares some amino acid interactions with Catechol

[1] M178 interacts with the CACL in the 2D representation, but an intramolecular bond in the 3D representation takes preference.

[2]Y132 H-bond depicted only in 3D structure

[3]T174 H-bond with the CACL is favoured in the 2D diagram, while the K173 H-bond is favoured in the 3D diagram

Chapter 4: Discussion

In this chapter the potential MOA of CA as an ALLINI is discussed. This is based on the results obtained where CA showed a co-operative inhibition of IN_{WT}, it disrupted the IN/LEDGF interaction and induced IN multimerization. Thereafter, binding kinetics and the docking studies of CA onto IN are discussed.

4.1. CA does not have the same resistance profile compared to RAL

In the current study, an ELISA assay was used to determine IC_{50} of IN_{Sol} and associated resistance mutations against RAL and CA (Table 3.2). ST activity of IN was tested to determine the fold changes in IC_{50} ($FCIC_{50}$) of dose responses of resistant mutants compared to wild type IN (Table 3.2). The ST reaction was tested by the usage of 3'-pre-processed LTR mimic as a vDNA substrate (Table 3.2 and Figure 3.13). The resistant mutants were specific to RAL, and are clinically presented in patients failing IN-based therapy [272]. These resistance mutations have been proven to disrupt the binding of RAL, and other specific ST inhibitors [273], to the active site of IN.

A large $FCIC_{50}$ would indicate resistance, and is clearly noticed in RAL experiments ranging from ±400 - ±1300 fold. With single primary mutants IN_{Q148H} and IN_{N155H} having a $FCIC_{50}$ of 947.99 and 392.44 fold, respectively. The secondary/double mutations, $IN_{G140S/Q148H}$ and $IN_{E92Q/N155H}$, expectedly had a higher $FCIC_{50}$ at 1262.41 and 583.92, respectively. A comparison of these results with those documented for RAL resistance profiles in biochemical assays are different, with a 5-10 fold increase for the single primary mutants [210, 258, 263, 274]. The solubility of IN has been associated with various technical difficulties in IN protein preparation and enzymatic assays [73, 84]. Similar problems were experienced in the current study; while IN_{WT} was active in the ELISA, the RAL resistance mutations inserted into the IN_{WT} template (Figure 3.6) yielded proteins that were not active under the conditions used. To alleviate these issues an active soluble mutant containing the F185K and C280S mutations was used for the ELISA [84]. This IN_{Sol} mutant was previously determined to confer resistance to certain DKAs [128]. The general reported IC_{50} concentrations for RAL is <10nM against IN_{Wt} [210, 258, 263,

274], however IC_{50} values obtained in the current assay was at least 400nM for IN_{Sol}, indicating a possible resistance of 40 fold. This inherent resistance of IN_{Sol} may have a concerted effect on the RAL selective mutations (IN_{Q148H}, $IN_{G140S/Q148H}$, IN_{N155H} and $IN_{E92Q/N155H}$), thereby causing a higher resistance against RAL compared to those that are reported elsewhere [210, 258, 263, 274].

CA was compared in the same system, and contrastingly, CA does not present with these large $FCIC_{50}$, ranging from ±1 - ±3.5 fold (Table 3.2), and may indicate a reduced effect of these mutations. IN_{Q148H} and the $IN_{E92Q/155H}$ were moderately resistant to CA. Modelling data, presented in Figure 3.27, suggests that Q148 has significant hydrogen bond interactions of the best fit CA poses. This would explain the moderate inhibition shown with the Q148H resistant mutation against CA. The interaction of CA with Q148 and E92 has been suggested by computational modelling [234, 244], and supported by biochemical work which indicated a resistance of IN_{Q148A} and IN_{E92Q} single mutations towards CA [234]. The binding site obtained in the modelling data presented in Figure 3.3 is supported with other molecular docking studies of CA to the active site [244, 245] with variations of the interacting amino acids at the same site.

Hill slopes of the dose response curves used for resistance profiling were compared, and an interesting trend emerged. The Hill slope tends to indicate non-cooperative binding of the inhibitor at lower $FCIC_{50}$ values, while resistance steepens the slope thereby increasing cooperativity. This was noticed for both RAL and CA, and in the case of CA tested against IN_{N155H}, the $FCIC_{50}$ was 1 while the slope was 0.78. This relationship is in contrast to literature [266], which indicate that for INSTIs, the IC_{50} changes in response to a resistance mutation yet the slope remains the same. This may be as a result of differences in the methodology used, where in the current study an ELISA was utilized, and in the study by Sampah et. al. (2011) [266] the resistance profiling was conducted using antiviral assays.

Resistance is noticed between CA and some of the mutants tested. However, it was minor compared to what was experienced with RAL under the same conditions. These are all mutations specifically resistant to RAL, where double IN mutations were less susceptible to RAL compared to single mutations [258]. The trend of resistance against single mutants compared to double mutants followed an expected pattern for RAL (Figure 3.14 and Table 3.2) [258]. CA did not follow this pattern, and IN_{Q148H} presented with a higher resistance against CA compared to $IN_{G140S/Q148H.}$ Furthermore, IN_{N155H} was not resistant to CA. RAL is an established active site binder, and its mechanism has been well described (Figure 1.9). This difference of CA's resistance profile towards RAL selected resistance mutations, suggests that there is a mechanistic difference between CA and RAL, and this may indicate that CA binds differently to the IN active site compared to RAL.

4.2. The F185K/C280S is resistant to CA only in the presence of Mg^{2+}

Metal cofactors are required for IN function [75], and the MOA of INSTI inhibitors of IN function by chelation of metal ions. The order of addition of assay components of the ELISA assays performed in the current study depicts an initial incubation of IN with DNA, then followed by inhibitor incubation. As such, the effect of specific metal cofactors on CA inhibition was determined using IN_{sol} (Figure 3.15) and there was a 323 fold increase in IC_{50} in the presence of Mg^{2+} compared to Mn^{2+}. The metal preference of CA is presumed to be related to the F185K/C280S mutation contained in IN_{Sol}, as this was not noticed in IN_{WT} under the same conditions (see next subsection 4.3). As discussed above, the Hill slope tends to change in response to resistance under our assay conditions. Comparing the assays with the same order of addition, where DNA is incubated with IN before addition of compound, IN_{Sol} has a slope of 1.65 (Table 3.2) and IN_{WT} has a slope of 1.1 (Table 3.3), indicating a possible resistance of IN_{Sol} towards CA. If a comparison is made of IN_{WT} and IN_{Sol}, in the presence of Mg^{2+} only, an IC_{50} of 171 fold increase is recognized (380nM to 65µM).

The resistance of CA to F185K/C280S has been alluded to before [236, 241], but has not been explained in terms of metal preference. A previous study by Reinke et. al. [236] stated that IN_{F185K} was resistant to CA by about 10 fold for the disintegration reaction. The ELISA assay

tested in our study used pre-processed LTR mimics and would selectively test the ST reaction. In this regard, a study by Pluymers et. al. [241] depicting an assay specifically testing ST of the IN_{Sol} mutant, and only using Mg^{2+} in the reaction, found that CA did not inhibit ST. This supports our observation that in the presence of Mg^{2+}, the IN_{Sol} ST activity is resistant to inhibition by CA.

It is interesting that IN_{Sol} is only resistant to CA in the presence of Mg^{2+}. The character of CA in the presence of Mn^{2+} and Mg^{2+} may be better explained by the differences of IN oligomeric organization in the presence of each metal co-factor [80, 102, 105, 106, 113, 268]. Co-ordination of metal cofactors by the DDE acidic residues have shown no difference in conformation at the active site, both free of DNA as well as in the intasome [108]. It was determined that the IN oligomeric state was affected by the specific metal ion present in reaction buffers [106]. Metal co-factors are essential in the formation of a stable DNA bound IN multimer [105]. There was also a preference for Mn^{2+} when IN was assembled on short LTR substrates, while IN assembly on longer substrates preferred Mg^{2+} [104]. Furthermore, Mg^{+} and Mn^{2+}, bound at the IN active site has differential effects on the entire conformation of the protein [107, 108]. It was shown that the disruption of IN_{WT} multimerization was caused by detergents, and this disruption is corrected by the addition of Mn^{2+} but not Mg^{2+} [268], however, detergent free IN_{WT} was able to assemble on DNA normally in the presence of either metal cofactor. Although activity of the F185K/C280S was shown to be identical to IN_{WT} [84], others have reported that enzyme activity was attenuated in the presence of Mg^{2+} [128, 275] and this was likely due to the disruption of oligomerization by the F185K mutation [255]. It is clear that metal ions, presumably bound to the active site, have structural effects across the entire IN protein. However, these structural effects appear to be different when comparing the IN_{WT} with the F185K/C280S mutant activity, and likely related to the difference in oligomerization [84]. It is conceivable that Mn^{2+} may correct for the effects of F185K/C280S, while Mg^{2+} does not, and this may be the reason why resistance to CA was not noticed when IN_{Sol} was tested in the presence of Mn^{2+}. This supports the hypothesis that CA has an allosteric MOA that is reliant on the organization of the IN enzyme structure.

4.3. The order of addition of assay components reveal a multimodal cooperative relationship of IN_{WT} inhibition by CA, in the absence of vDNA

The order of assay components can give an indication of how the mechanism of inhibition unfolds in a native setting. In the context of IN allosteric inhibition, ALLINI's require a free binding site, and allosteric inhibition is reduced by either LEDGF binding or vDNA/IN complex formation {Al-Mawsawi, 2008 #277}. By simulating these conditions in biochemical assays, it is possible to gain mechanistic insights of a compounds inhibition. Due to the allosteric site of IN only being available at certain points in the life cycle, by using the order of addition of assay components, it may be indicative of the stage where the ALLINI is most affective. In terms of the viral life cycle, the ALLINI binging site is blocked from the point of viral un-coating upto viral integration. This is likely due to the vDNA/IN complexation that occurs during viral maturation [44]. Once un-coating and nuclear localization completes, LEDGF is then recruited and interacts with IN at the nuclear periphery or inside the nucleus [144]. There may be a small window for ALLINIs to bind during the integration event, when the intasome is in transition from a vDNA bound to unbound state. The main point of ALLINI interaction is when IN is free, during viral assembly and just before maturation. Time of addition studies have indicated that ALLINIs have a mild inhibition at viral integration, while the main inhibition event occurs at the late stage viral life cycle [140, 170]. Apart from indicating the possible stage of the viral life cycle that ALLINIs are effective, the order of addition can give an insight into the target site that is bound by the ALLINI. For IN, if DNA is present the allosteric site is blocked [168], and this can be indicative of binding mechanisms of CA.

Using the same conditions as IN_{Sol} above, the metal preference in IN_{WT} was non-existent when CA was added after DNA, as indicated by no change in IC_{50} (Figure 3.16B and Table 3.3). The metal preference in IN_{WT} was only apparent if CA was added before DNA, indicating that metal preferences in IN_{WT} are dependent on the order of addition of assay components. A 3 fold increase of the IN_{WT} IC_{50} is noticed if CA was incubated before DNA addition (Figure 3.16A and Table 3.3), which is lower compared to a 327 fold change in IC_{50} for IN_{Sol}.

Apart from the metal effect, the order of addition itself affected the IC_{50}. As seen for reactions with Mn^{2+} only, if CA was added before DNA the IC_{50} was lower compared to when DNA was added before CA, with an $FCIC_{50}$ of 3.9 (Table 3.3). The order of addition did not affect the IC_{50} of assays performed only in the presence of Mg^{2+}, indicating a 1.2 fold change (Table 3.3). The order of addition of ST assays has been indicated as an important factor in characterising ALLINIs [168]. If incubated with IN in the absence of DNA, the ALLINI tends to have a 4 fold reduction in IC_{50} compared to a DNA bound enzyme, irrespective of the metal ion.

Metals are required at the IN active site for DNA binding and enzyme activity [85, 104]. It appears that in the absence of vDNA, CA inhibition has a preference for Mn^{2+}, which may indirectly suggest active site binding similar to what has been observed for DKAs [128]. However unlike DKAs, when DNA is bound to the active site, CA's metal preference is lost indicating that CA does not interact with the DNA bound active site. To support this view, Hill slopes are listed (Table 3.3), and show a change in the cooperativity which is dependent on the order of addition. When CA is added before DNA the relationship is cooperative for both Mg^{2+} and Mn^{2+} with slope values of 1.74 and 1.68, respectively. When DNA was added before CA, the inhibition became non-cooperative with a slope value of 1.07 and 1.1 for Mg^{2+} and Mn^{2+} conditions, respectively. The cooperativity of dose response curves directly relates to the binding mode of the compounds tested, and a their interpretation has been used to conclude a multimodal binding mechanism of ALLINIs [167]. A cooperative Hill slope indicates that the inhibitor occupies more than one binding site while a non-cooperative relationship indicates a single inhibitor binding site. Therefore, in the absence of vDNA, CA occupies more than one binding site irrespective of the metal cofactor, while in the case of a DNA bound active site, CA only binds to a single site. This indicates that DNA blocks CA binding, likely at the allosteric site, and CA therefore inhibits the DNA bound IN through a dual binding at the active site.

4.4. CA's MOA is dependent on order of addition and disrupts the IN/LEDGF interaction

An order of addition effect was noticed in the ST activity assay when IN was either pre-incubated with inhibitor then DNA or DNA then inhibitor. A 2-4 fold increase in IC_{50} when IN was pre-incubated with DNA, and CA was able to reduce the activity of IN_{WT} down by a 100% under both conditions (Figure 3.16 and Table 3.3). This was similarly noticed elsewhere for certain ALLINI compounds resembling LEDGIN (used in this study) [168], however, these were not able to reduce catalytic activity down by 100% when IN was pre-incubated with DNA. This order of addition is not a feature of all ALLINIs, where inhibition of IN catalytic functions by Mut101 was the same regardless of the order of addition [169] and may be dependent on specific chemistry and binding modes.

ALLINIs are believed to affect the HIV-1 life cycle at various points. Its MOA involves either more than one target, or affects the same target in different ways, and is said to be multimodal [139, 140, 167]. One such mode of action that defines certain ALLINIs is the inhibition of the LEDGF/IN interaction [162, 168]. The inhibition of LEDGF/IN has been shown to hold merit as a therapeutic target [276], [277], [152]. These evidences indicate that the LEDGF/IN interaction is important for HIV replication. The inhibition of the LEDGF/IN interaction in a dose response manner by CA, with an IC_{50} of 237nM, has been determined (Figure 3.17 B). This level of inhibition was comparable to Mut101, LEDGIN (Table 3.4) [140, 162, 169]. The effects of CA on the IN/LEDGF interaction in the presence and absence of viral DNA was also investigated (Figure 3.18). Here a synthesised LTR homologue was added before and after CA incubation. When the DNA was added after CA incubation the IN/LEDGF interaction was disrupted. When DNA was added before CA incubation the interaction was similar to the positive control. This is confirmed by the observation that a peptide specific to the LEDGF/IN interface was blocked when the IN/vDNA complex was allowed to pre-assemble [276]. This result indicates that CA can be blocked by pre-assembled complexes, and is likely the reason why CA does not inhibit isolated PICs [242, 254]. This result is a novel finding of the study, and strongly supports the allosteric MOA of CA, and suggests a binding site on IN that overlaps with LEDGF binding.

4.5. CA and ACA-1 causes the formation of IN multimers *in vitro* as well as in pNL4-3 virus produced in their presence

Another well documented effect of ALLINIs is the induction of multimerization of IN [140, 167, 175] and reviewed by Jurado et. al (2013) [139] and Feng et. al. (2015) [188]. IN assembles into dimers, tetramers and higher order oligomers during multimerization, and an increase in these species due to addition of an inhibitor indicates allosteric effects. Multimerization is not mutually exclusive of the IN/LEDGF interaction. LEDGF binds to the IN dimeric interface and causes for the stabilization of IN tetramers [88, 92, 100, 158, 159]. LEDGF also stabilizes the interaction between IN's CCD-CCD dimeric interface and inhibits subunit exchange [88]. IN multimerization appears to be regulated by the allosteric LEDGF IBD binding site. This has been exploited though the design of LEDGF based peptides, and compounds that led to increases of non-functional IN oligomers.

Multimerization is believed to be the major inhibitory pathway of ALLINI inhibition and causes the late stage disruption of the virion assembly [140, 172], while in comparison, the disruption of the LEDGF/IN interaction has a more modest effect on viral replication. Multimerization ultimately leads to non-infectious progeny virus due to disruptions of viral assembly and maturation [140]. The results obtained in the current study indicate that factors affecting multimerization should be carefully considered when testing compounds. Factors such as high concentration of protein may mask the effect of the ALLINIs (Figure 3.20 and Figure 3.21), whereas assay components such as detergents may cause anomalous results (Figure 3.22). Nevertheless, CA caused the multimerization of IN_{WT} in a dose responsive manner (Figure 3.23) and was comparable to the other ALLINIs tested, i.e. LEDGIN, Mut101 and ACA-1. The IN_{A128T} mutant that reduces aberrant multimerization caused by ALLINIs [142] was likewise tested. This mutant conferred resistance to LEDGIN [162], and in contrast, Mut101 was unaffected [169]. In the current study, CA induced similar multimerization compared to Mut101 while LEDGIN induced the lowest. The ACA-1 compound is predicted to bind to a distinct site and still cause multimerization [175], and in agreement to that view, it was shown here (Figure 3.24) that ACA-1 was the least affected by the A128T mutation. To confirm the

results of the crosslinking gel, SEC was performed, and similar to the crosslinking gels, CA caused the increase in dimeric IN_{WT} forms and reduced the monomer (Figure 3.25). The LEDGIN and ACA-1 also reduced monomer, but did not cause the increase of high order forms, and is likely due to the aggregation of protein out of solution (Figure 3.25 blue and orange). This effect is similarly noticed for the ALLINI-1 [142] and the ALLINI-2 [180] compounds in SEC analysis. RAL did not have a noticeable effect on the monomer or dimeric forms of IN_{WT} as compared to the negative control. Finally, virus produced from a transfected pNL3-4 clone presented with increased IN multimerization when the virus was produced in the presence of CA, ACA-1 and Mut101. This was similar to results obtained elsewhere for the BI-D ALLINI [140]. The induction of IN multimerization by CA is a novel finding, and further strengthens the hypothesis that CA has allosteric inhibitory effects on HIV-1 IN, likely by binding to the LEDGF recognition site on IN.

4.6. IN_{CCD} binding by CA In the absence of DNA

Interaction of INSTIs at the active site of IN require a specific conformation created after 3'-P at the IN/vDNA interface [86, 94, 122, 123, 135] and RAL fails to bind to the either IN_{WT} or IN_{CCD} in the absence of DNA [108, 122]. The binding of CA to IN_{Wt} and IN_{CCD} in the absence of vDNA would indicate a significant difference compared to RAL. SPR binding kinetics of CA was determined (Figure 3.19), and indicated that CA bound to the IN_{Wt} and IN_{CCD} in the absence of vDNA with a K_D of 2.1 and 2.6, respectively. The lower K_D for CA binding of the IN_{Wt} was expected as the ligand would have preferential binding for a complete protein structure. The on (k_{on}) and off (k_{off}) rate kinetics (Table 3.5) were similar between the IN_{Wt} and IN_{CCD}, and suggested that the IN_{CCD} is in fact the target domain of CA. Previous kinetic analysis indicated that CA bound tightly, though reversibly, to IN [236], and supports the loe K_D obtained for CA on IN by SPR in the current study.

Thus far, the results obtained in this study implicate CA as an ALLINI with a possible binding site on IN that overlaps the LEDGF binding site. Compared to the LEDGIN tested (Figure 3.19), CA appears to have a slower k_{on} and k_{off} rate. There is a lack of information on the IN resistance selection by CA in viral cell culture, and the only mutations described are the G140S [232] and

C56Y [241]. The slow dissociation of CA may account for the lack of resistant viral selection at the expected binding sites on IN. To explain this, a comparison can be made to the INSTI, dolutegravir (DTG), keeping in mind its distinct MOA compared to CA. Mutations that are selected by RAL do not confer resistance to DGT, even though these compounds both bind to the active site [278]. This high barrier to resistance was attributed to the significantly slower dissociation of DTG from the IN/vDNA complexes compared to RAL [272]. This is also believed to be a reason why resistant mutations in clinical samples do not present in DTG treated patients, whereas selection in cell culture occurs at distal positions to the active site, and only rarely, with H51Y and R263K being notable examples [279]. Other examples of slow binding kinetics causing a high genetic barrier to resistance is found in HIV-1 PR inhibitors [280] and Hepatitis C virus inhibitors [281]. Taking this into account, the mutations G140S and C56Y that are selected by CA in viral cell culture may arise as a result of a genetic barrier that is caused by CA's slow dissociation rate, thereby causing mutations distal to the LEDGF IBD binding site on IN. The C56S mutation has been extensively used in structural studies to increase the solubility of the IN_{CCD} [108, 275, 282], and the C56Y likewise lacks a sulphur group which may be involved in salt bridges. The C56Y may reduce the effect of CA aberrant multimerization of IN by increasing the solubility of the protein. The G140S is part of the catalytic loop [258, 263] and its effects on multimerization is less apparent. Yet, an adjacent I141K mutation had a reduced high order oligomer formation compared to wildtype in the apo-protein, and indicates that the catalytic flexible loop can be involved in IN multimerization [184]. Therefore the slow dissociation of CA from IN may cause for the selection of mutations distal to its binding site, and supports the observations of mutant selection in IN made thus far [232, 241]. This line of reasoning depends on the assumption that CA binds to the LEDGF site on IN, however, active site binding cannot be ruled out, especially since the unbound apo-protein was used for SPR. Therefore further investigation is required to confirm the slow dissociation kinetics of CA.

4.7. CA binding to IN at the active site

Active site binding of CA is suggested by previously determined empirical data. CA inhibition is reduced in active site mutations that gives resistance to DKAs and INSTIs [233] and it disrupted the IN/DNA complex formation [252]. Predictive docking of CA at the active site has

been previously undertaken [244, 245], and mutations based on this model have shown resistance towards CA [234]. In the predictive modelling of the current study, the lowest energy binding pose of CA at the active site (Figure 3.27 and Table 3.6) had a similar conformation and binding to those predicted previously [244]. Its CACL is involved in hydrogen bond interactions with K159 and K156, which was confirmed by mutagenesis [234, 246], and these amino acids have been implicated in the binding of vDNA by IN [82]. Other important DNA binding residues that are predicted to interact with CA's CACL are T66, which has been modelled before [244, 245], and the T66I mutations was shown to confer resistance to DKAs as well as CA [233]. Interestingly, the N155 residue, which is significant in RAL resistance, was also shown to interact with the CACL via hydrogen bonds (Figure 3.27 and Table 3.6). The N155H mutation was not found to have increased resistance to CA (Table 3.2), however, the hydrogen bond association of CA with neighbouring residues, may reduce the need for its interaction with N155. The catechol groups of CA are predicted to interact with the catalytic triad D116, and important DNA binding residue Q148, each interaction with a separate catechol. The binding pocket of the catechol arm, involved in the binding to D116, is likely stabilized by electrostatic and Van de Waals interactions with C65, H67 and E92. The mutation of E92Q may displace this arm of CA causing for the slight increase in resistance noticed (Table 3.2). Similar mutations of E92K and C65S have previously been implicated in resistance to CA [234], and this arm was predicted to be in a binding pocket adjacent to 5CITEP, with slight increased exposure to solvent [244] and may be more volatile. The catechol involved in hydrogen bonding with Q148 and Q62 fits into a pocket formed around T115, likewise with a hydrogen bond interaction, and this site overlaps with the 5CITEP binding site [83, 244]. The benzene ring at this binding site has a π-stacking interaction with the metal ion, and does not directly involve the catechol groups (Figure 3.27). Due to the extensive hydrogen bonding of CA and its reduced exposure to solvent predicted at the active site in this model, and in addition to the resistant profile observed in this study as well as in others [233, 234], we are unable to exclude CA active site binding as a possible MOA for its inhibition of IN. However, this binding mode requires for the active site to be free from DNA binding and is may only be relevant in certain biochemical assays. A dual inhibitory mechanism of CA where it binds to the active site as well as to the LEDGF binding site may be plausible. Dual binding of IN has been proposed for indole-containing DKAs [165, 166] which

bind to the active site as well as inhibit the IN/LEDGF interaction. However, these DKA's have a selectivity towards the ST reaction [165] suggesting that a related MOA to CA is unlikely.

4.8. Docking at the IN/LEDGF allosteric site predicts a new binding site for CA

Using the same docking parameters as the active site, CA was predicted to bind to the IN/LEDGF allosteric binding site, or just allosteric site, in both 1QS4 and 2B4J structures (Figure 3.28 and Figure 3.29). The 1QS4 pose interacts through hydrogen bonds with residues T125, A128 and Q95 of IN subunit A and also with M178, H171 and K173 of subunit B (Table 3.7) and this overlaps with the proposed binding site of the LEDGF IBD [260]. A similar binding site is noticed when CA was docked in the 2B4J allosteric site (Table 3.7), but had fewer hydrogen bond interaction and bound to T125 and A128 of subunit A and E170 of subunit B.

Important residues involved in the IN/LEDGF interaction are the E170 and H171 [139], and the observation that CA inhibits the IN/LEDGF interaction (Section 3.4.1) supports these binding modes. The reduction in hydrogen bond interactions between the two models can be attributed to the different conformation of the two sites. The 2B4J was co-crystalized with the LEDGF IBD, while 1QS4 was not, and this may represent an "open" verses "closed" conformation of the binding site. The binding pose depicted in the 1QS4 allosteric site appears to be mainly stabilized by extensive hydrogen bonding, and is surrounded by strong electrostatic interactions (pick shaded residues). The 2B4J pose is stabilized by hydrogen bonding and π stacking of both the catechol groups with W131 of subunit B and E170 of subunit A, while electrostatic interactions only surround half the compound. The 1QS4 pose was generally more exposed to solvent than the 2B4J pose as indicated by the grey shading in the 2D diagram. In contrast to the CA binding mode, The binding of ALLINIs such as ALLINI-2 [142], LEDGIN [162] and Mut101 [169] show hydrogen bond interactions with T174, H171 and E170 of only a single subunit, while its interaction with the opposite subunit is mainly reliant on electrostatic interactions. The A128 and H171 have been implicated in resistance to ALLINIs [142, 173]. The A128T mutation displaces the rigidly planar quinolone ring of some ALLINIs, without affecting compound binding, as is the case with ALLINI-2 [142]. The biding mode of CA suggests that these residues are less significant, and CA's binding can be

compensated by the hydrogen bonding with both IN subunits. In support of this view, the A128T is not resistant to the KF116 compound, which lacks the quinolone and interacts with both IN subunits by hydrogen bonding [187]. A common feature of both the CA binding poses involves the buried insertion of the CACL region into the binding site interacting with T125 and A128, and this may be its mode of recognition of the allosteric site. This is similar to other ALLINIs where their carboxylic groups commonly identify the H171 and E170 [139, 162].

Finally, in poses for both 1QS4 and 2B4J, CA interacts with M178 and W132, which is uncommon for many ALLINIs [139]. With the 1QS4 pose, CA's CACL fits into an electrostatic pocket which is includes both M178 and W132, while one of the carboxyl groups form a hydrogen bond with M178. In the 2B4J pose, CA's catechol group extends towards M178 and W132 forming Van de Vaal's and electrostatic interactions respectively. This interaction is interesting as the M178 and W132 is part of a four tier π interaction along with F181 and F185 [230]. This π stack is essential for ST functions *in vitro*, while its disruption led to pleiotropic deficiencies during viral replication [230]. This may offer an alternative to the active site inhibition pathway of IN catalytic functions by CA, while also providing an explanation for the observed resistance of the F185K mutation in the ST activity assays. To elaborate, any changes to IN's four tier π stack, which includes F185, would affect all four residues [230] and the lysine substitution in the F185K mutation may lead to changes at the M178 and W132, which in turn would disrupt CA binding, leading to resistance.

4.9. Differences of CA, TCA and ACA-1 elude to chemical preferences to distinct binding modes

Small molecule binding at the IN dimer-dimer interface was observed with a tetracetylated chicoric acid derivative, referred to in this paper as ACA-1 [175, 186]. ACA-1 is a significantly altered derivative compared to CA, with a methyl and amino ester groups in the central linker, while catechol groups are acetylated [235]. It was approximately 3 fold less potent at inhibiting IN catalytic functions compared to CA [235]. ACA-1 does not inhibit the IN/DNA complex formation [175]. The compound specifically interacted with IN at K173 at inhibitory concentrations as determined by an affinity acetylation mass spectrometry approach [175,

186], which was later revealed to be the sucrose binding site [111]. The effects of this interaction led to the stabilization of the IN dimer, indicated by the reduction of IN subunit exchange and promotion of multimerization [175], which directly supports the multimerization results obtained in the current study. The docking results presented here, with regards to ACA-1, also support these assertions. ACA-1 docking at the active site predicted an unfavourable binding pose conformation, where a single diacetylbenzene was involved with the T115 binding pocket. The MEL region and the second diacetylbenzene arm are significantly exposed to the solvent, with unlikely coordination of the Mg^{2+} ion. This binding pose supports the assumption that ACA-1 did not bind to IN active site. The 1QS4 allosteric site pose of ACA-1 indicated that the MEL region had interaction with the E96 and Y99, which forms part of the sucrose site, and a single arm stretched into the allosteric site to interact with A128. The interaction of ACA-1's MEL with similar residues was predicted previously [175] and the MEL was postulated to be important in ACA-1's selectivity for the sucrose site. The other arm of the 1QS4 pose was exposed to solvent. The 2B4J pose of ACA-1 at the allosteric site was completely exposed to solvent (Figure 3.29), and its binding to the allosteric site was unfavourable. ACA-1 appeared to dock with unfavourable exposure to solvent at all three sites used in the docking analysis, and consistently scored the lowest for each binding pose. These results support ACA-1's preferential interaction at the sucrose binding site.

A second tetracetylated compound, referred to here as TCA, was proposed to have the same MOA as CA [175]. TCA, while structurally similar to CA, was designed to shield the reactive catechol groups by acetylation, and shown to be approximately 10 and 7 fold less potent for 3'-P and ST compared to CA [235]. TCAs function of IN/DNA complex disruption was determined by DNA-crosslinking studies [175] and has a similar function to CA in this regard [252]. It was shown that TCA did not cause the stabilization of IN multimers and based on results obtained with TCA, the authors proposed that CA would function similarly [175]. Our docking results, although showing similar binding modes between TCA and CA, indicate some differences. TCA docking at the active site predicted a displacement of the CACL, contrary to its previously predicted conformation [244]. However, Healy *et al.* [244], used a rigid binding site, and predicted that TCA's catechol arm involved in E92 interactions was exposed to

solvent. In the current study, a flexible binding site was used, which may account for the differences observed. However, the common interaction with the 5CITEP overlapping region persists, and TCA is still predicted to interact with K159 through its diacetylbenzene. TCA has a similar binding conformation to CA at the allosteric binding site (Figure 3.28and Figure 3.29). In the 1QS4 pose of TCA, biding for the CACL and a single diacetyl-catechol arm was predicted to span from W132 to Q95 of the same chain. The other arm was significantly exposed to solvent, but brought close to the binding site via its interaction with opposite chain E170. In the 2B4J binding pose, the CACL and arm of TCA interacts with the K173 of the sucrose site, and spans the allosteric site to A128. The other arm, again, is exposed to solvent, with unfavourable distances from solvent exposed E167. The bulky addition of the acetyls may be responsible for the displacement of the diacetyl-catechol arm of TCA from the binding site resulting in less favourable binding, while the smaller CA fits better and is less exposed to the solvent. This may explain TCA's lack of inducing multimerization in IN.

4.10. Conclusions

The inhibition of IN by CA, in the context of the active site, has been extensively explored in the past. Previous studies have alluded to a CA active site binding mode in biochemical assays [233, 234, 252] as well as through molecular docking [244, 245]. The current study was unable to exclude this type of binding mechanism for CA and, in fact, metal dependency for catalytic inhibition (section 3.3.2) and docking results (sections 3.5.3 and 4.7.) support a CA active site binding mode.

The novelty that the current study unambiguously shows, are the additional allosteric effects of IN caused by CA (see sections 3.3.2, 3.4, 3.5.4, 4.2, 4.3, 4.4, 4.5 and 4.8). Firstly, CA has a cooperative inhibition of the free IN_{WT}, yet after vDNA is bound at the active site, the inhibition is non-cooperative and the $FCIC_{50}$ increases. This suggests that CA binds to more than one site on the free IN, and after vDNA complexes at the active site, CA's binding stoichiometry is reduced to 1:1. The inhibition of the IN/LEDGF interaction suggests that this additional CA binding site is analogous to the LEDGF allosteric interaction site of IN. Consequently, there was a non-cooperative inhibition of CA for the IN/LEDGF interaction,

suggesting that CA binds the allosteric site at a 1:1 ratio. The modulation of IN multimers by CA further supports an allosteric binding mode, and CA was shown to shift IN monomers to higher order oligomers similar to other IN/LEDGF disruptors. In the past, viral resistance selection by CA did not yield mutations at or close to the LEDGF allosteric site [232, 241], however, binding kinetics shown in the current study (section 3.4.2and 4.6) specify a strong interaction between IN and CA with slow on and off rates, similarly to what has been previously noticed [236], indicating that CA may have a high barrier to resistance and thereby select for mutations distal to its binding sites.

Our results indicate that CA binds to a free active site as well as to the LEDGF allosteric site of IN, and this was supported by docking studies indicating favourable CA binding at both sites. This mechanism is similar to what was observed for indole containing DKAs [165, 166]. However, in contrast to CA, these compounds are selective for ST and preferentially bind an IN active site that is complexed to vDNA. IN allosteric inhibition by CA, compared to active site binding, serves as a better explanation of the observed anti-viral activity [236], which appeared to affect integration. Especially due to CA's lack of inhibition against isolated PICs [242, 254], which suggest that CA inhibits integration via a different mechanism compared to INSTIs. ALLINIs that disrupt the IN/LEDGF interaction have been shown to exhibit an antiviral profile affecting integration, as well as cause the aberrant multimerization of IN within virions [140]. Therefore, results presented here further support CA's allosteric inhibitory mechanism due to the increase in IN multimerization observed in assembled virions in tissue culture (see sections 3.4.3 and 4.5). We have thus completed the aims and objectives of this study and conclude that CA has indeed an allosteric inhibitory mechanism against HIV-1 IN in addition to its active site inhibition.

CA may not be an ideal drug candidate due to several reasons, including chemical stability of the ester linkages, rotatable bond number impacting oral bioavailability, anionic character impeding membrane permeability, and toxicity generally associated with catechols [243, 283]. However, our results show it has a unique binding mode at the allosteric site, and potential for a high barrier to resistance. This warrants further investigation and development

of compounds with similar pharmacophores, which may yield second generation ALLINIs of greater potency with low risk of resistance development.

REFERENCES

1. Barre-Sinoussi F, Ross AL, Delfraissy J-F: **Past, present and future: 30 years of HIV research**. *Nat Rev Micro* 2013, **11**(12):877-883.
2. Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, Cummins LB, Arthur LO, Peeters M, Shaw GM *et al*: **Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes**. *Nature* 1999, **397**(6718):436-441.
3. Sharp PM, Bailes E, Chaudhuri RR, Rodenburg CM, Santiago MO, Hahn BH: **The origins of acquired immune deficiency syndrome viruses: where and when?** *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 2001, **356**(1410):867-876.
4. Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, Bunce M, Muyembe J-J, Kabongo J-MM, Kalengayi RM, Van Marck E *et al*: **Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960**. *Nature* 2008, **455**(7213):661-664.
5. Sousa JDd, Müller V, Lemey P, Vandamme A-M: **High GUD Incidence in the Early 20th Century Created a Particularly Permissive Time Window for the Origin and Initial Spread of Epidemic HIV Strains**. *PLoS ONE* 2010, **5**(4):e9936.
6. **Aids: Origin of pandemic 'was 1920s Kinshasa** [http://www.bbc.com/news/health-29442642]
7. Zhu T, Korber BT, Nahmias AJ, Hooper E, Sharp PM, Ho DD: **An African HIV-1 sequence from 1959 and implications for the origin of the epidemic**. *Nature* 1998, **391**(6667):594-597.
8. Saxinger W, Levine P, Dean A, de The G, Lange-Wantzin G, Moghissi J, Laurent F, Hoh M, Sarngadharan M, Gallo R: **Evidence for exposure to HTLV-III in Uganda before 1973**. *Science* 1985, **227**(4690):1036-1038.
9. Centers_for_Disease_Control: **Pneumocystis pneumonia--Los Angeles**. *MMWR Morb Mortal Wkly Rep* 1981, **30**(21):250-252.
10. Gilbert MTP, Rambaut A, Wlasiuk G, Spira TJ, Pitchenik AE, Worobey M: **The emergence of HIV/AIDS in the Americas and beyond**. *Proceedings of the National Academy of Sciences* 2007, **104**(47):18566-18570.
11. Palmisano L, Vella S: **A brief history of antiretroviral therapy of HIV infection: success and challenges**. *Ann Ist Super Sanita* 2011, **47**(1):44-48.
12. Barre-Sinoussi F, Chermann J, Rey F, Nugeyre M, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C *et al*: **Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS)**. *Science* 1983, **220**(4599):868-871.
13. Popovic M, Sarngadharan M, Read E, Gallo R: **Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS**. *Science* 1984, **224**(4648):497-500.
14. Wain-Hobson S, Sonigo P, Danos O, Cole S, Alizon M: **Nucleotide sequence of the AIDS virus, LAV**. *Cell* 1985, **40**(1):9-17.
15. Ratner L, Haseltine W, Patarca R, Livak KJ, Starcich B, Josephs SF, Doran ER, Rafalski JA, Whitehorn EA, Baumeister K *et al*: **Complete nucleotide sequence of the AIDS virus, HTLV-III**. *Nature* 1985, **313**(6000):277-284.
16. Campbell-Yesufu OT, Gandhi RT: **Update on Human Immunodeficiency Virus (HIV)-2 Infection**. *Clinical Infectious Diseases* 2011, **52**(6):780-787.
17. **HIV and AIDS estimates: Global, regional and country fact sheets** [http://www.unaids.org/en/regionscountries/countries/southafrica]
18. Nyindo M: **Complementary factors contributing to the rapid spread of HIV-I in sub-Saharan Africa: a review**. *East Afr Med J* 2005, **82**(1):40-46.
19. Chigwedere P, Seage GRI, Gruskin S, Lee T-H, Essex M: **Estimating the Lost Benefits of Antiretroviral Drug Use in South Africa**. *JAIDS Journal of Acquired Immune Deficiency Syndromes* 2008, **49**(4):410-415.
20. **SA's Zuma 'showered to avoid HIV'** [http://news.bbc.co.uk/2/hi/africa/4879822.stm]

21. **Global AIDS response progress report** [http://www.unaids.org/sites/default/files/en/dataanalysis/knowyourresponse/countryprogressreports/2012countries/ce_ZA_Narrative_Report.pdf]
22. **South Africa HIV and AIDS estimates (2013)** [http://www.unaids.org/sites/default/files/epidocuments/ZAF.pdf]
23. Maddon PJ, Dalgleish AG, McDougal JS, Clapham PR, Weiss RA, Axel R: **The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain**. *Cell* 1986, **47**(3):333-348.
24. Dalgleish AG, Beverley PC, Clapham PR, Crawford DH, Greaves MF, Weiss RA: **The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus**. *Nature* 1984, **312**(5996):763-767.
25. Kawamura T, Gulden FO, Sugaya M, McNamara DT, Borris DL, Lederman MM, Orenstein JM, Zimmerman PA, Blauvelt A: **R5 HIV productively infects Langerhans cells, and infection levels are regulated by compound CCR5 polymorphisms**. *Proceedings of the National Academy of Sciences* 2003, **100**(14):8401-8406.
26. Roy S, Fitz-Gibbon L, Poulin L, Wainberg MA: **Infection of human monocytes/macrophages by HIV-1: effect on secretion of IL-1 activity**. *Immunology* 1988, **64**(2):233-239.
27. Weinberg JB, Matthews TJ, Cullen BR, Malim MH: **Productive human immunodeficiency virus type 1 (HIV-1) infection of nonproliferating human monocytes**. *The Journal of Experimental Medicine* 1991, **174**(6):1477-1482.
28. Yamashita M, Emerman M: **Retroviral infection of non-dividing cells: Old and new perspectives**. *Virology* 2006, **344**(1):88-93.
29. De Clercq E: **HIV Life Cycle: Targets for Anti-HIV Agents**. In: *HIV-1 Integrase.* John Wiley & Sons, Inc.; 2011: 1-14.
30. Foli A, Maiocchi MA, Lisziewicz J, Lori F: **A Checkpoint in the Cell Cycle Progression as a Therapeutic Target to Inhibit HIV Replication**. *Journal of Infectious Diseases* 2007, **196**(9):1409-1415.
31. Groschel B, Bushman F: **Cell Cycle Arrest in G2/M Promotes Early Steps of Infection by Human Immunodeficiency Virus**. *Journal of Virology* 2005, **79**(9):5695-5704.
32. Sugaya M, Loré K, Koup RA, Douek DC, Blauvelt A: **HIV-Infected Langerhans Cells Preferentially Transmit Virus to Proliferating Autologous CD4+ Memory T Cells Located within Langerhans Cell-T Cell Clusters**. *The Journal of Immunology* 2004, **172**(4):2219-2224.
33. Li Q, Duan L, Estes JD, Ma Z-M, Rourke T, Wang Y, Reilly C, Carlis J, Miller CJ, Haase AT: **Peak SIV replication in resting memory CD4+ T cells depletes gut lamina propria CD4+ T cells**. *Nature* 2005, **434**(7037):1148-1152.
34. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA: **Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody**. *Nature* 1998, **393**(6686):648-659.
35. Wyatt R, Sodroski J: **The HIV-1 Envelope Glycoproteins: Fusogens, Antigens, and Immunogens**. *Science* 1998, **280**(5371):1884-1888.
36. Feng Y, Broder CC, Kennedy PE, Berger EA: **HIV-1 Entry Cofactor: Functional cDNA Cloning of a Seven-Transmembrane, G Protein-Coupled Receptor**. *Science* 1996, **272**(5263):872-877.
37. Deng H, Liu R, Ellmeier W, Choe S, Unutmaz D, Burkhart M, Marzio PD, Marmon S, Sutton RE, Hill CM *et al*: **Identification of a major co-receptor for primary isolates of HIV-1**. *Nature* 1996, **381**(6584):661-666.
38. Yang Y, Fricke T, Diaz-Griffero F: **Inhibition of Reverse Transcriptase Activity Increases Stability of the HIV-1 Core**. *Journal of Virology* 2013, **87**(1):683-687.
39. Arhel N: **Revisiting HIV-1 uncoating**. *Retrovirology* 2010, **7**(1):1-10.
40. Ambrose Z, Aiken C: **HIV-1 uncoating: connection to nuclear entry and regulation by host proteins**. *Virology* 2014, **454–455**:371-379.

41. Schaller T, Ocwieja KE, Rasaiyaah J, Price AJ, Brady TL, Roth SL, Hué S, Fletcher AJ, Lee K, KewalRamani VN *et al*: **HIV-1 Capsid-Cyclophilin Interactions Determine Nuclear Import Pathway, Integration Targeting and Replication Efficiency**. *PLoS Pathog* 2011, **7**(12):e1002439.
42. Suzuki Y, Craigie R: **The road to chromatin — nuclear entry of retroviruses**. *Nature Reviews Microbiology* 2007, **5**(3):187-196.
43. Craigie R, Bushman FD: **HIV DNA Integration**. *Cold Spring Harbor Perspectives in Medicine* 2012, **2**(7):a006890.
44. Sundquist WI, Kräusslich H-G: **HIV-1 Assembly, Budding, and Maturation**. *Cold Spring Harbor Perspectives in Medicine* 2012, **2**(7):a006924.
45. Briggs JAG, Wilk T, Welker R, Kräusslich HG, Fuller SD: **Structural organization of authentic, mature HIV-1 virions and cores**. *The EMBO Journal* 2003, **22**(7):1707-1715.
46. Ganser-Pornillos BK, Cheng A, Yeager M: **Structure of Full-Length HIV-1 CA: A Model for the Mature Capsid Lattice**. *Cell* 2007, **131**(1):70-79.
47. Forshey BM, von Schwedler U, Sundquist WI, Aiken C: **Formation of a Human Immunodeficiency Virus Type 1 Core of Optimal Stability Is Crucial for Viral Replication**. *Journal of Virology* 2002, **76**(11):5667-5677.
48. Hu W-S, Hughes SH: **HIV-1 Reverse Transcription**. *Cold Spring Harbor Perspectives in Medicine* 2012, **2**(10).
49. Jacobo-Molina A, Ding J, Nanni RG, Clark AD, Lu X, Tantillo C, Williams RL, Kamer G, Ferris AL, Clark P: **Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA**. *Proceedings of the National Academy of Sciences of the United States of America* 1993, **90**(13):6320-6324.
50. Yamashita M, Emerman M: **The Cell Cycle Independence of HIV Infections Is Not Determined by Known Karyophilic Viral Elements**. *PLoS Pathog* 2005, **1**(3):e18.
51. Yamashita M, Emerman M: **Capsid Is a Dominant Determinant of Retrovirus Infectivity in Nondividing Cells**. *Journal of Virology* 2004, **78**(11):5670-5678.
52. Diaz-Griffero F: **The Role of TNPO3 in HIV-1 Replication**. *Molecular Biology International* 2012, **2012**:6.
53. Miller MD, Farnet CM, Bushman FD: **Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition**. *Journal of Virology* 1997, **71**(7):5382-5390.
54. Franke EK, Yuan HEH, Luban J: **Specific incorporation of cyclophilin A into HIV-1 virions**. *Nature* 1994, **372**(6504):359-362.
55. Gamble TR, Vajdos FF, Yoo S, Worthylake DK, Houseweart M, Sundquist WI, Hill CP: **Crystal Structure of Human Cyclophilin A Bound to the Amino-Terminal Domain of HIV-1 Capsid**. *Cell* 1996, **87**(7):1285-1294.
56. Shah VB, Shi J, Hout DR, Oztop I, Krishnan L, Ahn J, Shotwell MS, Engelman A, Aiken C: **The Host Proteins Transportin SR2/TNPO3 and Cyclophilin A Exert Opposing Effects on HIV-1 Uncoating**. *Journal of Virology* 2013, **87**(1):422-432.
57. Christ F, Thys W, De Rijck J, Gijsbers R, Albanese A, Arosio D, Emiliani S, Rain J-C, Benarous R, Cereseto A *et al*: **Transportin-SR2 Imports HIV into the Nucleus**. *Current Biology* 2008, **18**(16):1192-1202.
58. Ocwieja KE, Brady TL, Ronen K, Huegel A, Roth SL, Schaller T, James LC, Towers GJ, Young JAT, Chanda SK *et al*: **HIV Integration Targeting: A Pathway Involving Transportin-3 and the Nuclear Pore Protein RanBP2**. *PLoS Pathog* 2011, **7**(3):e1001313.
59. De Iaco A, Luban J: **Inhibition of HIV-1 infection by TNPO3 depletion is determined by capsid and detectable after viral cDNA enters the nucleus**. *Retrovirology* 2011, **8**:98-98.
60. Krishnan L, Matreyek KA, Oztop I, Lee K, Tipper CH, Li X, Dar MJ, KewalRamani VN, Engelman A: **The Requirement for Cellular Transportin 3 (TNPO3 or TRN-SR2) during Infection Maps to**

Human Immunodeficiency Virus Type 1 Capsid and Not Integrase. *Journal of Virology* 2010, **84**(1):397-406.
61. Cribier A, Ségéral E, Delelis O, Parissi V, Simon A, Ruff M, Benarous R, Emiliani S: **Mutations affecting interaction of integrase with TNPO3 do not prevent HIV-1 cDNA nuclear import**. *Retrovirology* 2011, **8**(1):104.
62. Mohammed KD, Topper MB, Muesing MA: **Sequential Deletion of the Integrase (Gag-Pol) Carboxyl Terminus Reveals Distinct Phenotypic Classes of Defective HIV-1**. *Journal of Virology* 2011, **85**(10):4654-4666.
63. Engelman A: **Pleiotropic Nature of HIV-1 Integrase Mutations**. In: *HIV-1 Integrase.* John Wiley & Sons, Inc.; 2011: 67-81.
64. Lu R, Limón A, Ghory HZ, Engelman A: **Genetic Analyses of DNA-Binding Mutants in the Catalytic Core Domain of Human Immunodeficiency Virus Type 1 Integrase**. *Journal of Virology* 2005, **79**(4):2493-2505.
65. Lu R, Limón A, Devroe E, Silver PA, Cherepanov P, Engelman A: **Class II Integrase Mutants with Changes in Putative Nuclear Localization Signals Are Primarily Blocked at a Postnuclear Entry Step of Human Immunodeficiency Virus Type 1 Replication**. *Journal of Virology* 2004, **78**(23):12735-12746.
66. Lu R, Ghory HZ, Engelman A: **Genetic Analyses of Conserved Residues in the Carboxyl-Terminal Domain of Human Immunodeficiency Virus Type 1 Integrase**. *Journal of Virology* 2005, **79**(16):10356-10368.
67. Engelman A, Englund G, Orenstein JM, Martin MA, Craigie R: **Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication**. *Journal of Virology* 1995, **69**(5):2729-2736.
68. Zhu K, Dobard C, Chow SA: **Requirement for Integrase during Reverse Transcription of Human Immunodeficiency Virus Type 1 and the Effect of Cysteine Mutations of Integrase on Its Interactions with Reverse Transcriptase**. *Journal of Virology* 2004, **78**(10):5045-5055.
69. Limón A, Devroe E, Lu R, Ghory HZ, Silver PA, Engelman A: **Nuclear Localization of Human Immunodeficiency Virus Type 1 Preintegration Complexes (PICs): V165A and R166A Are Pleiotropic Integrase Mutants Primarily Defective for Integration, Not PIC Nuclear Import**. *Journal of Virology* 2002, **76**(21):10598-10607.
70. Leavitt AD, Robles G, Alesandro N, Varmus HE: **Human immunodeficiency virus type 1 integrase mutants retain in vitro integrase activity yet fail to integrate viral DNA efficiently during infection**. *Journal of Virology* 1996, **70**(2):721-728.
71. Krishnan L, Engelman A: **Retroviral Integrase Proteins and HIV-1 DNA Integration**. *Journal of Biological Chemistry* 2012, **287**(49):40858-40866.
72. Nowotny M: **Retroviral integrase superfamily: the structural perspective**. *EMBO Reports* 2009, **10**(2):144-151.
73. Dyda F, Hickman A, Jenkins T, Engelman A, Craigie R, Davies D: **Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases**. *Science* 1994, **266**(5193):1981-1986.
74. Yang W, Steitz TA: **Recombining the structures of HIV integrase, RuvC and RNase H**. *Structure* 1995, **3**(2):131-134.
75. Engelman A, Mizuuchi K, Craigie R: **HIV-1 DNA integration: Mechanism of viral DNA cleavage and DNA strand transfer**. *Cell* 1991, **67**(6):1211-1221.
76. Zervos P, Hassell T, Van Frank R, Lai MT: **Characterization of an internally initiated integrase protein of HIV-1 produced in E. coli**. *Biochemical and Biophysical Research Communications* 1990, **170**(3):1061-1066.
77. Bushman FD, Engelman A, Palmer I, Wingfield P, Craigie R: **Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding**. *Proceedings of the National Academy of Sciences* 1993, **90**(8):3428-3432.

78. Woerner AM, Marcus-Sekura CJ: **Characterization of a DNA binding domain in the C-terminus of HIV-1 integrase by deletion mutagenesis**. *Nucleic Acids Research* 1993, **21**(15):3507-3511.
79. Wang JY, Ling H, Yang W, Craigie R: **Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein**. *The EMBO Journal* 2001, **20**(24):7333-7343.
80. Zheng R, Jenkins TM, Craigie R: **Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity**. *Proceedings of the National Academy of Sciences of the United States of America* 1996, **93**(24):13659-13664.
81. van den Ent FMI, Vos A, Plasterk RHA: **Dissecting the Role of the N-Terminal Domain of Human Immunodeficiency Virus Integrase bytrans-Complementation Analysis**. *Journal of Virology* 1999, **73**(4):3176-3183.
82. Goldgur Y, Dyda F, Hickman AB, Jenkins TM, Craigie R, Davies DR: **Three new structures of the core domain of HIV-1 integrase: An active site that binds magnesium**. *Proceedings of the National Academy of Sciences* 1998, **95**(16):9150-9154.
83. Goldgur Y, Craigie R, Cohen GH, Fujiwara T, Yoshinaga T, Fujishita T, Sugimoto H, Endo T, Murai H, Davies DR: **Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: a platform for antiviral drug design**. *Proc Natl Acad Sci U S A* 1999, **96**(23):13040-13043.
84. Jenkins TM, Engelman A, Ghirlando R, Craigie R: **A Soluble Active Mutant of HIV-1 Integrase: INVOLVEMENT OF BOTH THE CORE AND CARBOXYL-TERMINAL DOMAINS IN MULTIMERIZATION**. *Journal of Biological Chemistry* 1996, **271**(13):7712-7718.
85. Hare S, Maertens GN, Cherepanov P: **3'-Processing and strand transfer catalysed by retroviral integrase in crystallo**. *The EMBO Journal* 2012, **31**(13):3020-3028.
86. Lee MC, Deng J, Briggs JM, Duan Y: **Large-Scale Conformational Dynamics of the HIV-1 Integrase Core Domain and Its Catalytic Loop Mutants**. *Biophysical Journal* 2005, **88**(5):3133-3146.
87. Busschots K, Voet A, De Maeyer M, Rain J-C, Emiliani S, Benarous R, Desender L, Debyser Z, Christ F: **Identification of the LEDGF/p75 Binding Site in HIV-1 Integrase**. *Journal of Molecular Biology* 2007, **365**(5):1480-1492.
88. McKee CJ, Kessl JJ, Shkriabai N, Dar MJ, Engelman A, Kvaratskhelia M: **Dynamic Modulation of HIV-1 Integrase Structure and Function by Cellular Lens Epithelium-derived Growth Factor (LEDGF) Protein**. *Journal of Biological Chemistry* 2008, **283**(46):31802-31812.
89. Chen JC-H, Krucinski J, Miercke LJW, Finer-Moore JS, Tang AH, Leavitt AD, Stroud RM: **Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: A model for viral DNA binding**. *Proceedings of the National Academy of Sciences* 2000, **97**(15):8233-8238.
90. Jaskolski M, Alexandratos JN, Bujacz G, Wlodawer A: **Structural Studies of Retroviral Integrases**. In: *HIV-1 Integrase.* John Wiley & Sons, Inc.; 2011: 35-49.
91. Li M, Mizuuchi M, Burke TR, Craigie R: **Retroviral DNA integration: reaction pathway and critical intermediates**. *The EMBO Journal* 2006, **25**(6):1295-1304.
92. Cherepanov P, Maertens G, Proost P, Devreese B, Van Beeumen J, Engelborghs Y, De Clercq E, Debyser Z: **HIV-1 Integrase Forms Stable Tetramers and Associates with LEDGF/p75 Protein in Human Cells**. *Journal of Biological Chemistry* 2003, **278**(1):372-381.
93. Kotova S, Li M, Dimitriadis EK, Craigie R: **Nucleoprotein intermediates in HIV-1 DNA integration visualized by atomic force microscopy**. *Journal of Molecular Biology* 2010, **399**(3):491-500.
94. Hare S, Gupta SS, Valkov E, Engelman A, Cherepanov P: **Retroviral intasome assembly and inhibition of DNA strand transfer**. *Nature* 2010, **464**(7286):232-236.
95. Liao C, Nicklaus MC: **HIV-1 Integrase–DNA Models**. In: *HIV-1 Integrase.* John Wiley & Sons, Inc.; 2011: 429-455.
96. Hare S, Vos AM, Clayton RF, Thuring JW, Cummings MD, Cherepanov P: **Molecular mechanisms of retroviral integrase inhibition and the evolution of viral resistance**. *Proceedings of the National Academy of Sciences* 2010, **107**(46):20057-20062.

97. Krishnan L, Li X, Naraharisetty HL, Hare S, Cherepanov P, Engelman A: **Structure-based modeling of the functional HIV-1 intasome and its inhibition**. *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**(36):15910-15915.
98. Johnson BC, Métifiot M, Ferris A, Pommier Y, Hughes SH: **A Homology Model of HIV-1 Integrase and Analysis of Mutations Designed to Test the Model**. *Journal of Molecular Biology* 2013, **425**(12):2133-2146.
99. Yin Z, Shi K, Banerjee S, Pandey KK, Bera S, Grandgenett DP, Aihara H: **Crystal structure of the Rous sarcoma virus intasome**. *Nature* 2016, **530**(7590):362-366.
100. Kessl JJ, Li M, Ignatov M, Shkriabai N, Eidahl JO, Feng L, Musier-Forsyth K, Craigie R, Kvaratskhelia M: **FRET analysis reveals distinct conformations of IN tetramers in the presence of viral DNA or LEDGF/p75**. *Nucleic Acids Research* 2011, **39**(20):9009-9022.
101. Berthoux L, Sebastian S, Muesing MA, Luban J: **The Role of Lysine 186 in HIV-1 Integrase Multimerization**. *Virology* 2007, **364**(1):227-236.
102. Ellison V, Gerton J, Vincent KA, Brown PO: **An Essential Interaction between Distinct Domains of HIV-1 Integrase Mediates Assembly of the Active Multimer**. *Journal of Biological Chemistry* 1995, **270**(7):3320-3326.
103. Lee SP, Xiao J, Knutson JR, Lewis MS, Han MK: **Zn2+ Promotes the Self-Association of Human Immunodeficiency Virus Type-1 Integrase in Vitro**. *Biochemistry* 1997, **36**(1):173-180.
104. Lee SP, Kim HG, Censullo ML, Han MK: **Characterization of Mg2+-Dependent 3'-Processing Activity for Human Immunodeficiency Virus Type 1 Integrase in Vitro: Real-Time Kinetic Studies Using Fluorescence Resonance Energy Transfer**. *Biochemistry* 1995, **34**(32):10205-10214.
105. Wolfe AL, Felock PJ, Hastings JC, Blau CU, Hazuda DJ: **The role of manganese in promoting multimerization and assembly of human immunodeficiency virus type 1 integrase as a catalytically active complex on immobilized long terminal repeat substrates**. *Journal of Virology* 1996, **70**(3):1424-1432.
106. Deprez E, Tauc P, Leh H, Mouscadet J-F, Auclair C, Brochon J-C: **Oligomeric States of the HIV-1 Integrase As Measured by Time-Resolved Fluorescence Anisotropy†**. *Biochemistry* 2000, **39**(31):9275-9284.
107. Asante-Appiah E, Seeholzer SH, Skalka AM: **Structural Determinants of Metal-induced Conformational Changes in HIV-1 Integrase**. *Journal of Biological Chemistry* 1998, **273**(52):35078-35087.
108. Fitzkee NC, Masse JE, Shen Y, Davies DR, Bax A: **Solution Conformation and Dynamics of the HIV-1 Integrase Core Domain**. *Journal of Biological Chemistry* 2010, **285**(23):18072-18084.
109. Cellier C, Moreau K, Gallay K, Ballandras A, Gouet P, Ronfort C: **In vitro functional analyses of the human immunodeficiency virus type 1 (HIV-1) integrase mutants give new insights into the intasome assembly**. *Virology* 2013, **439**(2):97-104.
110. Cherepanov P, Ambrosio ALB, Rahman S, Ellenberger T, Engelman A: **Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75**. *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**(48):17308-17313.
111. Wielens J, Headey SJ, Jeevarajah D, Rhodes DI, Deadman J, Chalmers DK, Scanlon MJ, Parker MW: **Crystal structure of the HIV-1 integrase core domain in complex with sucrose reveals details of an allosteric inhibitory binding site**. *FEBS Lett* 2010, **584**(8):1455-1462.
112. Serrao E, Thys W, Demeulemeester J, Al-Mawsawi LQ, Christ F, Debyser Z, Neamati N: **A Symmetric Region of the HIV-1 Integrase Dimerization Interface Is Essential for Viral Replication**. *PLoS ONE* 2012, **7**(9):e45177.
113. Pemberton IK, Buckle M, Buc H: **The Metal Ion-induced Cooperative Binding of HIV-1 Integrase to DNA Exhibits a Marked Preference for Mn(II) Rather than Mg(II)**. *Journal of Biological Chemistry* 1996, **271**(3):1498-1506.

114. Carayon K, Leh H, Henry E, Simon F, Mouscadet J-F, Deprez E: **A cooperative and specific DNA-binding mode of HIV-1 integrase depends on the nature of the metallic cofactor and involves the zinc-containing N-terminal domain**. *Nucleic Acids Research* 2010, **38**(11):3692-3708.
115. Craigie R: **Integrase Mechanism and Function**. In: *HIV-1 Integrase.* John Wiley & Sons, Inc.; 2011: 23-33.
116. Petit C, Schwartz O, Mammano F: **Oligomerization within Virions and Subcellular Localization of Human Immunodeficiency Virus Type 1 Integrase**. *Journal of Virology* 1999, **73**(6):5079-5088.
117. Baranova S, Tuzikov FV, Zakharova OD, Tuzikova NA, Calmels C, Litvak S, Tarrago-Litvak L, Parissi V, Nevinsky GA: **Small-angle X-ray characterization of the nucleoprotein complexes resulting from DNA-induced oligomerization of HIV-1 integrase**. *Nucleic Acids Research* 2007, **35**(3):975-987.
118. Stevenson M, Haggerty S, Lamonica CA, Meier CM, Welch SK, Wasiak AJ: **Integration is not necessary for expression of human immunodeficiency virus type 1 protein products**. *Journal of Virology* 1990, **64**(5):2421-2425.
119. Shin CG, Taddeo B, Haseltine WA, Farnet CM: **Genetic analysis of the human immunodeficiency virus type 1 integrase protein**. *Journal of Virology* 1994, **68**(3):1633-1642.
120. Taddeo B, Haseltine WA, Farnet CM: **Integrase mutants of human immunodeficiency virus type 1 with a specific defect in integration**. *Journal of Virology* 1994, **68**(12):8401-8405.
121. Temesgen Z, Siraj DS: **Raltegravir: first in class HIV integrase inhibitor**. *Therapeutics and Clinical Risk Management* 2008, **4**(2):493-500.
122. Espeseth AS, Felock P, Wolfe A, Witmer M, Grobler J, Anthony N, Egbertson M, Melamed JY, Young S, Hamill T *et al*: **HIV-1 integrase inhibitors that compete with the target DNA substrate define a unique strand transfer conformation for integrase**. *Proceedings of the National Academy of Sciences* 2000, **97**(21):11244-11249.
123. Arora R, de Beauchene IC, Polanski J, Laine E, Tchertanov L: **Raltegravir flexibility and its impact on recognition by the HIV-1 IN targets**. *Journal of Molecular Recognition* 2013, **26**(9):383-401.
124. Neamati N: **HIV-1 Integrase Inhibitor Design: Overview and Historical Perspectives**. In: *HIV-1 Integrase.* John Wiley & Sons, Inc.; 2011: 165-196.
125. Bacchi A, Carcelli M, Compari C, Fisicaro E, Pala N, Rispoli G, Rogolino D, Sanchez TW, Sechi M, Neamati N: **HIV-1 IN Strand Transfer Chelating Inhibitors: A Focus on Metal Binding**. *Molecular Pharmaceutics* 2011, **8**(2):507-519.
126. Grobler JA, Stillmock K, Hu B, Witmer M, Felock P, Espeseth AS, Wolfe A, Egbertson M, Bourgeois M, Melamed J *et al*: **Diketo acid inhibitor mechanism and HIV-1 integrase: implications for metal binding in the active site of phosphotransferase enzymes**. *Proc Natl Acad Sci U S A* 2002, **99**(10):6661-6666.
127. Hicks C, Gulick RM: **Raltegravir: The First HIV Type 1 Integrase Inhibitor**. *Clinical Infectious Diseases* 2009, **48**(7):931-939.
128. Marchand C, Johnson AA, Karki RG, Pais GCG, Zhang X, Cowansage K, Patel TA, Nicklaus MC, Burke TR, Pommier Y: **Metal-Dependent Inhibition of HIV-1 Integrase by β-Diketo Acids and Resistance of the Soluble Double-Mutant (F185K/C280S)**. *Molecular Pharmacology* 2003, **64**(3):600-609.
129. Bacchi A, Biemmi M, Carcelli M, Carta F, Compari C, Fisicaro E, Rogolino D, Sechi M, Sippel M, Sotriffer CA *et al*: **From Ligand to Complexes. Part 2. Remarks on Human Immunodeficiency Virus type 1 Integrase Inhibition by β-Diketo Acid Metal Complexes**. *Journal of Medicinal Chemistry* 2008, **51**(22):7253-7264.
130. Sechi M, Bacchi A, Carcelli M, Compari C, Duce E, Fisicaro E, Rogolino D, Gates P, Derudas M, Al-Mawsawi LQ *et al*: **From Ligand to Complexes: Inhibition of Human Immunodeficiency Virus Type 1 Integrase by β-Diketo Acid Metal Complexes**. *Journal of Medicinal Chemistry* 2006, **49**(14):4248-4260.

131. Hazuda DJ, Felock P, Witmer M, Wolfe A, Stillmock K, Grobler JA, Espeseth A, Gabryelski L, Schleif W, Blau C *et al*: **Inhibitors of Strand Transfer That Prevent Integration and Inhibit HIV-1 Replication in Cells**. *Science* 2000, **287**(5453):646-650.
132. Loizidou EZ, Zeinalipour-Yazdi CD, Christofides T, Kostrikis LG: **Analysis of binding parameters of HIV-1 integrase inhibitors: Correlates of drug inhibition and resistance**. *Bioorganic & Medicinal Chemistry* 2009, **17**(13):4806-4818.
133. Mouscadet J-F, Tchertanov L: **Raltegravir: molecular basis of its mechanism of action**. *European Journal of Medical Research* 2009, **14**(Suppl 3):5-16.
134. Tchertanov L, Mouscadet J-F: **Target Recognition by Catechols and β-Ketoenols: Potential Contribution of Hydrogen Bonding and Mn/Mg Chelation to HIV-1 Integrase Inhibition**. *Journal of Medicinal Chemistry* 2007, **50**(6):1133-1145.
135. Chen X, Tsiang M, Yu F, Hung M, Jones GS, Zeynalzadegan A, Qi X, Jin H, Kim CU, Swaminathan S *et al*: **Modeling, Analysis, and Validation of a Novel HIV Integrase Structure Provide Insights into the Binding Modes of Potent Integrase Inhibitors**. *Journal of Molecular Biology* 2008, **380**(3):504-519.
136. Cooper DA, Steigbigel RT, Gatell JM, Rockstroh JK, Katlama C, Yeni P, Lazzarin A, Clotet B, Kumar PN, Eron JE *et al*: **Subgroup and Resistance Analyses of Raltegravir for Resistant HIV-1 Infection**. *New England Journal of Medicine* 2008, **359**(4):355-365.
137. Hazuda DJ, Miller MD, Nguyen BY, Zhao J: **Resistance to the HIV-Integrase Inhibitor Raltegravir: Analysis of Protocol 005, a Phase 2 Study in Patients with Triple-Class Resistant HIV-1 Infection** In: *16th International HIV Drug Resistance Workshop: June 12-16 2007; Barbados, West Indies*.
138. Hardy JA, Wells JA: **Searching for new allosteric sites in enzymes**. *Current Opinion in Structural Biology* 2004, **14**(6):706-715.
139. Jurado KA, Engelman A: **Multimodal mechanism of action of allosteric HIV-1 integrase inhibitors**. *Expert Reviews in Molecular Medicine* 2013, **15**:null-null.
140. Jurado KA, Wang H, Slaughter A, Feng L, Kessl JJ, Koh Y, Wang W, Ballandras-Colas A, Patel PA, Fuchs JR *et al*: **Allosteric integrase inhibitor potency is determined through the inhibition of HIV-1 particle maturation**. *Proceedings of the National Academy of Sciences* 2013, **110**(21):8690-8695.
141. Fontana J, Jurado KA, Cheng N, Ly NL, Fuchs JR, Gorelick RJ, Engelman AN, Steven AC: **Distribution and Redistribution of HIV-1 Nucleocapsid Protein in Immature, Mature, and Integrase-Inhibited Virions: a Role for Integrase in Maturation**. *Journal of Virology* 2015, **89**(19):9765-9780.
142. Feng L, Sharma A, Slaughter A, Jena N, Koh Y, Shkriabai N, Larue RC, Patel PA, Mitsuya H, Kessl JJ *et al*: **The A128T Resistance Mutation Reveals Aberrant Protein Multimerization as the Primary Mechanism of Action of Allosteric HIV-1 Integrase Inhibitors**. *Journal of Biological Chemistry* 2013, **288**(22):15813-15820.
143. Engelman A, Cherepanov P: **The Lentiviral Integrase Binding Protein LEDGF/p75 and HIV-1 Replication**. *PLoS Pathogens* 2008, **4**(3):e1000046.
144. Christ F, Debyser Z: **The LEDGF/p75 integrase interaction, a novel target for anti-HIV therapy**. *Virology* 2013, **435**(1):102-109.
145. Dietz F, Franken S, Yoshida K, Nakamura H, Kappler J, Gieselmann V: **The family of hepatoma-derived growth factor proteins: characterization of a new member HRP-4 and classification of its subfamilies**. *Biochemical Journal* 2002, **366**(2):491-500.
146. Ciuffi A, Llano M, Poeschla E, Hoffmann C, Leipzig J, Shinn P, Ecker JR, Bushman F: **A role for LEDGF/p75 in targeting HIV DNA integration**. *Nat Med* 2005, **11**(12):1287-1289.
147. Gijsbers R, Vets S, De Rijck J, Ocwieja KE, Ronen K, Malani N, Bushman FD, Debyser Z: **Role of the PWWP Domain of Lens Epithelium-derived Growth Factor (LEDGF)/p75 Cofactor in Lentiviral Integration Targeting**. *The Journal of Biological Chemistry* 2011, **286**(48):41812-41825.

148. Maertens G, Cherepanov P, Pluymers W, Busschots K, De Clercq E, Debyser Z, Engelborghs Y: **LEDGF/p75 Is Essential for Nuclear and Chromosomal Targeting of HIV-1 Integrase in Human Cells**. *Journal of Biological Chemistry* 2003, **278**(35):33528-33539.
149. Shun M-C, Raghavendra NK, Vandegraaff N, Daigle JE, Hughes S, Kellam P, Cherepanov P, Engelman A: **LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration**. *Genes & Development* 2007, **21**(14):1767-1778.
150. Marshall HM, Ronen K, Berry C, Llano M, Sutherland H, Saenz D, Bickmore W, Poeschla E, Bushman FD: **Role of PSIP1/LEDGF/p75 in Lentiviral Infectivity and Integration Targeting**. *PLoS ONE* 2007, **2**(12):e1340.
151. Ferris AL, Wu X, Hughes CM, Stewart C, Smith SJ, Milne TA, Wang GG, Shun M-C, Allis CD, Engelman A *et al*: **Lens epithelium-derived growth factor fusion proteins redirect HIV-1 DNA integration**. *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**(7):3135-3140.
152. De Rijck J, Vandekerckhove L, Gijsbers R, Hombrouck A, Hendrix J, Vercammen J, Engelborghs Y, Christ F, Debyser Z: **Overexpression of the Lens Epithelium-Derived Growth Factor/p75 Integrase Binding Domain Inhibits Human Immunodeficiency Virus Replication**. *Journal of Virology* 2006, **80**(23):11498-11509.
153. Cherepanov P, Sun Z-YJ, Rahman S, Maertens G, Wagner G, Engelman A: **Solution structure of the HIV-1 integrase-binding domain in LEDGF/p75**. *Nat Struct Mol Biol* 2005, **12**(6):526-532.
154. Hare S, Cherepanov P: **The Interaction Between Lentiviral Integrase and LEDGF: Structural and Functional Insights**. *Viruses* 2009, **1**(3):780-801.
155. Hare S, Shun M-C, Gupta SS, Valkov E, Engelman A, Cherepanov P: **A Novel Co-Crystal Structure Affords the Design of Gain-of-Function Lentiviral Integrase Mutants in the Presence of Modified PSIP1/LEDGF/p75**. *PLoS Pathogens* 2009, **5**(1):e1000259.
156. Hare S, Di Nunzio F, Labeja A, Wang J, Engelman A, Cherepanov P: **Structural Basis for Functional Tetramerization of Lentiviral Integrase**. *PLoS Pathogens* 2009, **5**(7):e1000515.
157. Llano M, Vanegas M, Fregoso O, Saenz D, Chung S, Peretz M, Poeschla EM: **LEDGF/p75 Determines Cellular Trafficking of Diverse Lentiviral but Not Murine Oncoretroviral Integrase Proteins and Is a Component of Functional Lentiviral Preintegration Complexes**. *Journal of Virology* 2004, **78**(17):9524-9537.
158. Raghavendra NK, Engelman A: **LEDGF/p75 interferes with the formation of synaptic nucleoprotein complexes that catalyze full-site HIV-1 DNA integration in vitro: Implications for the mechanism of viral cDNA integration**. *Virology* 2007, **360**(1):1-5.
159. Pandey KK, Sinha S, Grandgenett DP: **Transcriptional Coactivator LEDGF/p75 Modulates Human Immunodeficiency Virus Type 1 Integrase-Mediated Concerted Integration**. *Journal of Virology* 2007, **81**(8):3969-3979.
160. Guy S, Rotem D, Hayouka Z, Gabizon R, Levin A, Zemel L, Loyter A, Porath D, Friedler A: **Monitoring the HIV-1 integrase enzymatic activity using atomic force microscopy in a 2LTR system**. *Chemical Communications* 2013, **49**(30):3113-3115.
161. Hayouka Z, Rosenbluh J, Levin A, Loya S, Lebendiker M, Veprintsev D, Kotler M, Hizi A, Loyter A, Friedler A: **Inhibiting HIV-1 integrase by shifting its oligomerization equilibrium**. *Proceedings of the National Academy of Sciences* 2007, **104**(20):8316-8321.
162. Christ F, Voet A, Marchand A, Nicolet S, Desimmie BA, Marchand D, Bardiot D, Van der Veken NJ, Van Remoortel B, Strelkov SV *et al*: **Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication**. *Nat Chem Biol* 2010, **6**(6):442-448.
163. De Luca L, Gitto R, Christ F, Ferro S, De Grazia S, Morreale F, Debyser Z, Chimirri A: **4-[1-(4-Fluorobenzyl)-4-hydroxy-1H-indol-3-yl]-2-hydroxy-4-oxobut-2-enoic acid as a prototype to develop dual inhibitors of HIV-1 integration process**. *Antiviral Research* 2011, **92**(1):102-107.
164. Fan X, Zhang F-H, Al-Safi RI, Zeng L-F, Shabaik Y, Debnath B, Sanchez TW, Odde S, Neamati N, Long Y-Q: **Design of HIV-1 integrase inhibitors targeting the catalytic domain as well as its**

interaction with LEDGF/p75: A scaffold hopping approach using salicylate and catechol groups. *Bioorganic & Medicinal Chemistry* 2011, **19**(16):4935-4952.
165. Rogolino D, Carcelli M, Compari C, De Luca L, Ferro S, Fisicaro E, Rispoli G, Neamati N, Debyser Z, Christ F *et al*: **Diketoacid chelating ligands as dual inhibitors of HIV-1 integration process**. *European Journal of Medicinal Chemistry* 2014, **78**(0):425-430.
166. Ferro S, De Luca L, Lo Surdo G, Morreale F, Christ F, Debyser Z, Gitto R, Chimirri A: **A new potential approach to block HIV-1 replication via protein–protein interaction and strand-transfer inhibition**. *Bioorganic & Medicinal Chemistry* 2014, **22**(7):2269-2279.
167. Kessl JJ, Jena N, Koh Y, Taskent-Sezgin H, Slaughter A, Feng L, de Silva S, Wu L, Le Grice SF, Engelman A *et al*: **Multimode, cooperative mechanism of action of allosteric HIV-1 integrase inhibitors**. *J Biol Chem* 2012, **287**(20):16801-16811.
168. Christ F, Shaw S, Demeulemeester J, Desimmie BA, Marchand A, Butler S, Smets W, Chaltin P, Westby M, Debyser Z *et al*: **Small molecule inhibitors of the LEDGF/p75 binding site of integrase (LEDGINs) block HIV replication and modulate integrase multimerization**. *Antimicrobial Agents and Chemotherapy* 2012.
169. Le Rouzic E, Bonnard D, Chasset S, Bruneau J-M, Chevreuil F, Le Strat F, Nguyen J, Beauvoir R, Amadori C, Brias J *et al*: **Dual inhibition of HIV-1 replication by integrase-LEDGF allosteric inhibitors is predominant at the post-integration stage**. *Retrovirology* 2013, **10**(1):144.
170. Balakrishnan M, Yant SR, Tsai L, O'Sullivan C, Bam RA, Tsai A, Niedziela-Majka A, Stray KM, Sakowicz R, Cihlar T: **Non-Catalytic Site HIV-1 Integrase Inhibitors Disrupt Core Maturation and Induce a Reverse Transcription Block in Target Cells**. *PLoS ONE* 2013, **8**(9):e74163.
171. Tsiang M, Jones GS, Niedziela-Majka A, Kan E, Lansdon EB, Huang W, Hung M, Samuel D, Novikov N, Xu Y *et al*: **New Class of HIV-1 Integrase (IN) Inhibitors with a Dual Mode of Action**. *Journal of Biological Chemistry* 2012, **287**(25):21189-21203.
172. Desimmie B, Schrijvers R, Demeulemeester J, Borrenberghs D, Weydert C, Thys W, Vets S, Van Remoortel B, Hofkens J, De Rijck J *et al*: **LEDGINs inhibit late stage HIV-1 replication by modulating integrase multimerization in the virions**. *Retrovirology* 2013, **10**(1):57.
173. Slaughter A, Jurado KA, Deng N, Feng L, Kessl JJ, Shkriabai N, Larue RC, Fadel HJ, Patel PA, Jena N *et al*: **The mechanism of H171T resistance reveals the importance of Nδ-protonated His171 for the binding of allosteric inhibitor BI-D to HIV-1 integrase**. *Retrovirology* 2014, **11**(1):100.
174. Fenwick C, Amad Ma, Bailey MD, Bethell R, Bös M, Bonneau P, Cordingley M, Coulombe R, Duan J, Edwards P *et al*: **Preclinical Profile of BI 224436, a Novel HIV-1 Non-Catalytic-Site Integrase Inhibitor**. *Antimicrobial Agents and Chemotherapy* 2014, **58**(6):3233-3244.
175. Kessl JJ, Eidahl JO, Shkriabai N, Zhao Z, McKee CJ, Hess S, Burke TR, Jr., Kvaratskhelia M: **An allosteric mechanism for inhibiting HIV-1 integrase with a small molecule**. *Mol Pharmacol* 2009, **76**(4):824-832.
176. Camarasa M-J, Velázquez S, San-Félix A, Pérez-Pérez M-J, Gago F: **Dimerization inhibitors of HIV-1 reverse transcriptase, protease and integrase: A single mode of inhibition for the three HIV enzymes?** *Antiviral Research* 2006, **71**(2–3):260-267.
177. Maes M, Loyter A, Friedler A: **Peptides that inhibit HIV-1 integrase by blocking its protein–protein interactions**. *FEBS Journal* 2012, **279**(16):2795-2809.
178. Tintori C, Demeulemeester J, Franchi L, Massa S, Debyser Z, Christ F, Botta M: **Discovery of small molecule HIV-1 integrase dimerization inhibitors**. *Bioorganic & Medicinal Chemistry Letters* 2012, **22**(9):3109-3114.
179. Feng L, Dharmarajan V, Serrao E, Hoyte A, Larue RC, Slaughter A, Sharma A, Plumb MR, Kessl JJ, Fuchs JR *et al*: **The Competitive Interplay between Allosteric HIV-1 Integrase Inhibitor BI/D and LEDGF/p75 during the Early Stage of HIV-1 Replication Adversely Affects Inhibitor Potency**. *ACS Chemical Biology* 2016, **11**(5):1313-1321.
180. Shkriabai N, Dharmarajan V, Slaughter A, Kessl JJ, Larue RC, Feng L, Fuchs JR, Griffin PR, Kvaratskhelia M: **A Critical Role of the C-terminal Segment for Allosteric Inhibitor-induced**

Aberrant Multimerization of HIV-1 Integrase. *The Journal of Biological Chemistry* 2014, **289**(38):26430-26440.
181. Desimmie BA, Weydert C, Schrijvers R, Vets S, Demeulemeester J, Proost P, Paron I, De Rijck J, Mast J, Bannert N *et al*: **HIV-1 IN/Pol recruits LEDGF/p75 into viral particles**. *Retrovirology* 2015, **12**:16.
182. van Bel N, van der Velden Y, Bonnard D, Le Rouzic E, Das AT, Benarous R, Berkhout B: **The Allosteric HIV-1 Integrase Inhibitor BI-D Affects Virion Maturation but Does Not Influence Packaging of a Functional RNA Genome**. *PLoS ONE* 2014, **9**(7):e103552.
183. Mathew S, Nguyen M, Wu X, Pal A, Shah VB, Prasad VR, Aiken C, Kalpana GV: **INI1/hSNF5-interaction defective HIV-1 IN mutants exhibit impaired particle morphology, reverse transcription and integration in vivo**. *Retrovirology* 2013, **10**:66-66.
184. Kessl Jacques J, Kutluay Sebla B, Townsend D, Rebensburg S, Slaughter A, Larue Ross C, Shkriabai N, Bakouche N, Fuchs James R, Bieniasz Paul D *et al*: **HIV-1 Integrase Binds the Viral RNA Genome and Is Essential during Virion Morphogenesis**. *Cell* 2016, **166**(5):1257-1268.e1212.
185. Du L, Zhao Y-x, Yang L-m, Zheng Y-t, Tang Y, Shen X, Jiang H-l: **Symmetrical 1-pyrrolidineacetamide showing anti-HIV activity through a new binding site on HIV-1 integrase**. *Acta Pharmacol Sin* 2008, **29**(10):1261-1267.
186. Shkriabai N, Patil SS, Hess S, Budihas SR, Craigie R, Burke TR, Le Grice SFJ, Kvaratskhelia M: **Identification of an inhibitor-binding site to HIV-1 integrase with affinity acetylation and mass spectrometry**. *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**(18):6894-6899.
187. Sharma A, Slaughter A, Jena N, Feng L, Kessl JJ, Fadel HJ, Malani N, Male F, Wu L, Poeschla E *et al*: **A New Class of Multimerization Selective Inhibitors of HIV-1 Integrase**. *PLoS Pathogens* 2014, **10**(5):e1004171.
188. Feng L, Larue RC, Slaughter A, Kessl JJ, Kvaratskhelia M: **HIV-1 Integrase Multimerization as a Therapeutic Target**. *Current topics in microbiology and immunology* 2015, **389**:93-119.
189. Tsurutani N, Kubo M, Maeda Y, Ohashi T, Yamamoto N, Kannagi M, Masuda T: **Identification of Critical Amino Acid Residues in Human Immunodeficiency Virus Type 1 IN Required for Efficient Proviral DNA Formation at Steps prior to Integration in Dividing and Nondividing Cells**. *Journal of Virology* 2000, **74**(10):4795-4806.
190. Bouyac-Bertoia M, Dvorin JD, Fouchier RAM, Jenkins Y, Meyer BE, Wu LI, Emerman M, Malim MH: **HIV-1 Infection Requires a Functional Integrase NLS**. *Molecular Cell*, **7**(5):1025-1035.
191. Bouyac-Bertoia M, Dvorin JD, Fouchier RAM, Jenkins Y, Meyer BE, Wu LI, Emerman M, Malim MH: **HIV-1 Infection Requires a Functional Integrase NLS**. *Molecular Cell* 2001, **7**(5):1025-1035.
192. Gallay P, Hope T, Chin D, Trono D: **HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway**. *Proceedings of the National Academy of Sciences of the United States of America* 1997, **94**(18):9825-9830.
193. Hearps Anna C, Jans David A: **HIV-1 integrase is capable of targeting DNA to the nucleus via an Importin α/β-dependent mechanism**. *Biochemical Journal* 2006, **398**(Pt 3):475-484.
194. Petit C, Schwartz O, Mammano F: **The Karyophilic Properties of Human Immunodeficiency Virus Type 1 Integrase Are Not Required for Nuclear Import of Proviral DNA**. *Journal of Virology* 2000, **74**(15):7119-7126.
195. Briones MS, Dobard CW, Chow SA: **Role of Human Immunodeficiency Virus Type 1 Integrase in Uncoating of the Viral Core**. *Journal of Virology* 2010, **84**(10):5181-5190.
196. Al-Mawsawi LQ, Sechi M, Neamati N: **Single amino acid substitution in HIV-1 integrase catalytic core causes a dramatic shift in inhibitor selectivity**. *FEBS Letters* 2007, **581**(6):1151-1156.

197. Ao Z, Fowke KR, Cohen ÉA, Yao X: **Contribution of the C-terminal tri-lysine regions of human immunodeficiency virus type 1 integrase for efficient reverse transcription and viral DNA nuclear import**. *Retrovirology* 2005, **2**(1):62.
198. Nakamura T, Masuda T, Goto T, Sano K, Nakai M, Harada S: **Lack of Infectivity of HIV-1 Integrase Zinc Finger-like Domain Mutant with Morphologically Normal Maturation**. *Biochemical and Biophysical Research Communications* 1997, **239**(3):715-722.
199. Hatziioannou T, Perez-Caballero D, Cowan S, Bieniasz PD: **Cyclophilin Interactions with Incoming Human Immunodeficiency Virus Type 1 Capsids with Opposing Effects on Infectivity in Human Cells**. *Journal of Virology* 2005, **79**(1):176-183.
200. Briones MS, Chow SA: **A new functional role of HIV-1 integrase during uncoating of the viral core**. *Immunologic Research* 2010, **48**(1):14-26.
201. **Characterizing the binding between HIV-1 integrase and capsid and the role of the interaction in modulating the uncoating process** [http://grantome.com/grant/NIH/R21-AI118540-01]
202. Brik A, Wong C-H: **HIV-1 protease: mechanism and drug discovery**. *Organic & Biomolecular Chemistry* 2003, **1**(1):5-14.
203. Tang C, Louis JM, Aniana A, Suh J-Y, Clore GM: **Visualizing transient events in amino-terminal autoprocessing of HIV-1 protease**. *Nature* 2008, **455**(7213):693-696.
204. Bukovsky A, Göttlinger H: **Lack of integrase can markedly affect human immunodeficiency virus type 1 particle production in the presence of an active viral protease**. *Journal of Virology* 1996, **70**(10):6820-6825.
205. Liao W-H, Wang C-T: **Characterization of human immunodeficiency virus type 1 Pr160gag-pol mutants with truncations downstream of the protease domain**. *Virology* 2004, **329**(1):180-188.
206. Mandal D, Feng Z, Stoltzfus CM: **Gag-Processing Defect of Human Immunodeficiency Virus Type 1 Integrase E246 and G247 Mutants Is Caused by Activation of an Overlapping 5′ Splice Site**. *Journal of Virology* 2008, **82**(3):1600-1604.
207. Liu H, Wu X, Xiao H, Conway JA, Kappes JC: **Incorporation of functional human immunodeficiency virus type 1 integrase into virions independent of the Gag-Pol precursor protein**. *Journal of Virology* 1997, **71**(10):7704-7710.
208. Wu X, Liu H, Xiao H, Conway JA, Hunter E, Kappes JC: **Functional RT and IN incorporated into HIV-1 particles independently of the Gag/Pol precursor protein**. *The EMBO Journal* 1997, **16**(16):5113-5122.
209. Wu X, Liu H, Xiao H, Conway JA, Hehl E, Kalpana GV, Prasad V, Kappes JC: **Human Immunodeficiency Virus Type 1 Integrase Protein Promotes Reverse Transcription through Specific Interactions with the Nucleoprotein Reverse Transcription Complex**. *Journal of Virology* 1999, **73**(3):2126-2135.
210. Marinello J, Marchand C, Mott B, Bain A, Thomas CJ, Pommier Y: **Comparison of Raltegravir and Elvitagravir on HIV-1 Integrase Catalytic Reactions and on a Series of Drug-Resistant Integrase Mutants**. *Biochemistry* 2008, **47**(36):9345-9354.
211. Yung E, Sorin M, Pal A, Craig E, Morozov A, Delattre O, Kappes J, Ott D, Kalpana GV: **Inhibition of HIV-1 virion production by a transdominant mutant of integrase interactor 1**. *Nat Med* 2001, **7**(8):920-926.
212. Yung E, Sorin M, Wang E-J, Perumal S, Ott D, Kalpana GV: **Specificity of Interaction of INI1/hSNF5 with Retroviral Integrases and Its Functional Significance**. *Journal of Virology* 2004, **78**(5):2222-2231.
213. Cano J, Kalpana GV: **Inhibition of Early Stages of HIV-1 Assembly by INI1/hSNF5 Transdominant Negative Mutant S6**. *Journal of Virology* 2011, **85**(5):2254-2265.
214. Hehl EA, Joshi P, Kalpana GV, Prasad VR: **Interaction between Human Immunodeficiency Virus Type 1 Reverse Transcriptase and Integrase Proteins**. *Journal of Virology* 2004, **78**(10):5056-5067.

215. Tekeste SS, Wilkinson TA, Weiner EM, Xu X, Miller JT, Le Grice SFJ, Clubb RT, Chow SA: **Interaction Between Reverse Transcriptase and Integrase is Required for Reverse Transcription During HIV-1 Replication**. *Journal of Virology* 2015.
216. Warren K, Warrilow D, Meredith L, Harrich D: **Reverse Transcriptase and Cellular Factors: Regulators of HIV-1 Reverse Transcription**. *Viruses* 2009, **1**(3):873.
217. Witvrouw M, Fikkert V, Pluymers W, Matthews B, Mardel K, Schols D, Raff J, Debyser Z, De Clercq E, Holan G *et al*: **Polyanionic (i.e., Polysulfonate) Dendrimers Can Inhibit the Replication of Human Immunodeficiency Virus by Interfering with Both Virus Adsorption and Later Steps (Reverse Transcriptase/Integrase) in the Virus Replicative Cycle**. *Molecular Pharmacology* 2000, **58**(5):1100-1108.
218. Hombrouck A, Hantson A, van Remoortel B, Michiels M, Vercammen J, Rhodes D, Tetz V, Engelborghs Y, Christ F, Debyser Z *et al*: **Selection of human immunodeficiency virus type 1 resistance against the pyranodipyrimidine V-165 points to a multimodal mechanism of action**. *Journal of Antimicrobial Chemotherapy* 2007, **59**(6):1084-1095.
219. Pannecouque C, Pluymers W, Van Maele B, Tetz V, Cherepanov P, De Clercq E, Witvrouw M, Debyser Z: **New Class of HIV Integrase Inhibitors that Block Viral Replication in Cell Culture**. *Current Biology* 2002, **12**(14):1169-1177.
220. Bonnenfant S, Thomas CM, Vita C, Subra F, Deprez E, Zouhiri F, Desmaële D, d'Angelo J, Mouscadet JF, Leh H: **Styrylquinolines, Integrase Inhibitors Acting Prior to Integration: a New Mechanism of Action for Anti-Integrase Agents**. *Journal of Virology* 2004, **78**(11):5728-5736.
221. Han Y-S, Xiao W-L, Quashie PK, Mesplède T, Xu H, Deprez E, Delelis O, Pu J-X, Sun H-D, Wainberg MA: **Development of a fluorescence-based HIV-1 integrase DNA binding assay for identification of novel HIV-1 integrase inhibitors**. *Antiviral Research* 2013, **98**(3):441-448.
222. Ceccherini-Silberstein F, Malet I, Fabeni L, Dimonte S, Svicher V, D'Arrigo R, Artese A, Costa G, Bono S, Alcaro S *et al*: **Specific HIV-1 integrase polymorphisms change their prevalence in untreated versus antiretroviral-treated HIV-1-infected patients, all naive to integrase inhibitors**. *Journal of Antimicrobial Chemotherapy* 2010, **65**(11):2305-2318.
223. Quashie PK, Han Y-S, Hassounah S, Mesplède T, Wainberg MA: **Structural Studies of the HIV-1 Integrase Protein: Compound Screening and Characterization of a DNA-Binding Inhibitor**. *PLoS ONE* 2015, **10**(6):e0128310.
224. Deprez E, Barbe S, Kolaski M, Leh H, Zouhiri F, Auclair C, Brochon J-C, Le Bret M, Mouscadet J-F: **Mechanism of HIV-1 Integrase Inhibition by Styrylquinoline Derivatives in Vitro**. *Molecular Pharmacology* 2004, **65**(1):85-98.
225. Sourgen F, Maroun RG, Frère V, Bouziane M, Auclair C, Troalen F, Fermandjian S: **A Synthetic Peptide from the Human Immunodeficiency Virus Type-1 Integrase Exhibits Coiled-Coil Properties and Interferes with the in Vitro Integration Activity of The Enzyme**. *European Journal of Biochemistry* 1996, **240**(3):765-773.
226. Maroun RG, Gayet S, Benleulmi MS, Porumb H, Zargarian L, Merad H, Leh H, Mouscadet J-F, Troalen F, Fermandjian S: **Peptide Inhibitors of HIV-1 Integrase Dissociate the Enzyme Oligomers**. *Biochemistry* 2001, **40**(46):13840-13848.
227. Li H-Y, Zawahir Z, Song L-D, Long Y-Q, Neamati N: **Sequence-Based Design and Discovery of Peptide Inhibitors of HIV-1 Integrase: Insight into the Binding Mode of the Enzyme**. *Journal of Medicinal Chemistry* 2006, **49**(15):4477-4486.
228. Demeulemeester J, Tintori C, Botta M, Debyser Z, Christ F: **Development of an AlphaScreen-Based HIV-1 Integrase Dimerization Assay for Discovery of Novel Allosteric Inhibitors**. *Journal of Biomolecular Screening* 2012, **17**(5):618-628.
229. Zhao L, O'Reilly MK, Shultz MD, Chmielewski J: **Interfacial peptide inhibitors of HIV-1 integrase activity and dimerization**. *Bioorganic & Medicinal Chemistry Letters* 2003, **13**(6):1175-1177.

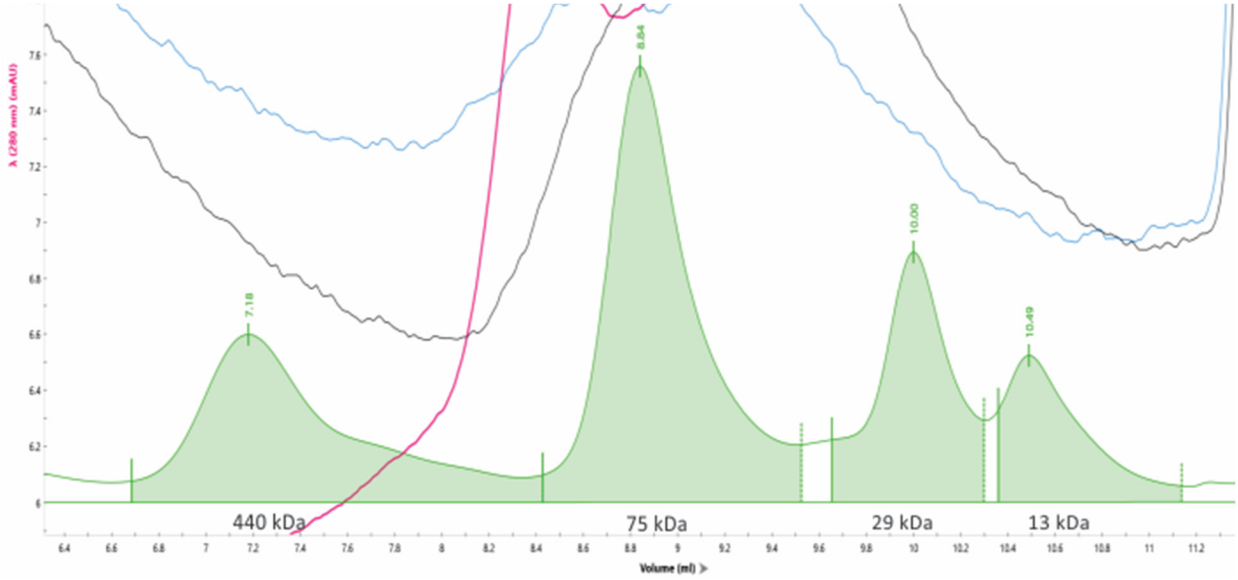
230. Al-Mawsawi LQ, Hombrouck A, Dayam R, Debyser Z, Neamati N: **Four-tiered π interaction at the dimeric interface of HIV-1 integrase critical for DNA integration and viral infectivity**. *Virology* 2008, **377**(2):355-363.
231. King PJ, Ma G, Miao W, Jia Q, McDougall BR, Reinecke MG, Cornell C, Kuan J, Kim TR, Robinson WE: **Structure−Activity Relationships: Analogues of the Dicaffeoylquinic and Dicaffeoyltartaric Acids as Potent Inhibitors of Human Immunodeficiency Virus Type 1 Integrase and Replication†**. *Journal of Medicinal Chemistry* 1999, **42**(3):497-509.
232. King PJ, Robinson WE: **Resistance to the Anti-Human Immunodeficiency Virus Type 1 Compound l-Chicoric Acid Results from a Single Mutation at Amino Acid 140 of Integrase**. *Journal of Virology* 1998, **72**(10):8420-8424.
233. Lee DJ, Robinson WE: **Human Immunodeficiency Virus Type 1 (HIV-1) Integrase: Resistance to Diketo Acid Integrase Inhibitors Impairs HIV-1 Replication and Integration and Confers Cross-Resistance to l-Chicoric Acid**. *Journal of Virology* 2004, **78**(11):5835-5847.
234. Lee DJ, Robinson WE: **Preliminary Mapping of a Putative Inhibitor-Binding Pocket for Human Immunodeficiency Virus Type 1 Integrase Inhibitors**. *Antimicrobial Agents and Chemotherapy* 2006, **50**(1):134-142.
235. Lin Z, Neamati N, Zhao H, Kiryu Y, Turpin JA, Aberham C, Strebel K, Kohn K, Witvrouw M, Pannecouque C *et al*: **Chicoric Acid Analogues as HIV-1 Integrase Inhibitors**. *Journal of Medicinal Chemistry* 1999, **42**(8):1401-1414.
236. Reinke RA, Lee DJ, McDougall BR, King PJ, Victoria J, Mao Y, Lei X, Reinecke MG, Robinson WE, Jr.: **L-chicoric acid inhibits human immunodeficiency virus type 1 integration in vivo and is a noncompetitive but reversible inhibitor of HIV-1 integrase in vitro**. *Virology* 2004, **326**(2):203-219.
237. Robinson WE, Cordeiro M, Abdel-Malek S, Jia Q, Chow SA, Reinecke MG, Mitchell WM: **Dicaffeoylquinic acid inhibitors of human immunodeficiency virus integrase: inhibition of the core catalytic domain of human immunodeficiency virus integrase**. *Molecular Pharmacology* 1996, **50**(4):846-855.
238. Robinson WE, Reinecke MG, Abdel-Malek S, Jia Q, Chow SA: **Inhibitors of HIV-1 replication that inhibit HIV integrase**. *Proceedings of the National Academy of Sciences* 1996, **93**(13):6326-6331.
239. Zhu K, Cordeiro ML, Atienza J, Robinson WE, Chow SA: **Irreversible Inhibition of Human Immunodeficiency Virus Type 1 Integrase by Dicaffeoylquinic Acids**. *Journal of Virology* 1999, **73**(4):3309-3316.
240. Lee J, Scagel CF: **Chicoric acid: chemistry, distribution, and production**. *Frontiers in Chemistry* 2013, **1**:40.
241. Pluymers W, Neamati N, Pannecouque C, Fikkert V, Marchand C, Burke TR, Pommier Y, Schols D, De Clercq E, Debyser Z *et al*: **Viral Entry as the Primary Target for the Anti-HIV Activity of Chicoric Acid and Its Tetra-Acetyl Esters**. *Molecular Pharmacology* 2000, **58**(3):641-648.
242. Farnet CM, Wang B, Hansen M, Lipford JR, Zalkow L, Robinson WE, Siegel J, Bushman F: **Human Immunodeficiency Virus Type 1 cDNA Integration: New Aromatic Hydroxylated Inhibitors and Studies of the Inhibition Mechanism**. *Antimicrobial Agents and Chemotherapy* 1998, **42**(9):2245-2253.
243. Crosby DC, Robinson WE: **Dicaffeoyltartaric Acid and Dicaffeoylquinic Acid HIV Integrase Inhibitors**. In: *HIV-1 Integrase*. John Wiley & Sons, Inc.; 2011: 341-362.
244. Healy EF, Sanders J, King PJ, Robinson Jr WE: **A docking study of l-chicoric acid with HIV-1 integrase**. *Journal of Molecular Graphics and Modelling* 2009, **27**(5):584-589.
245. Sotriffer CA, Ni, McCammon JA: **Active Site Binding Modes of HIV-1 Integrase Inhibitors**. *Journal of Medicinal Chemistry* 2000, **43**(22):4109-4117.
246. Crosby DC, Lei X, Gibbs CG, Reinecke MG, Robinson WE, Jr.: **Mutagenesis of lysines 156 and 159 in human immunodeficiency virus type 1 integrase (IN) reveals differential interactions between these residues and different IN inhibitors**. *Nat Prod Commun* 2015, **10**(1):117-124.

247. Jenkins TM, Esposito D, Engelman A, Craigie R: **Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking**. *The EMBO Journal* 1997, **16**(22):6849-6859.
248. Reinke RA, King PJ, Victoria JG, McDougall BR, Ma G, Mao Y, Reinecke MG, Robinson WE: **Dicaffeoyltartaric Acid Analogues Inhibit Human Immunodeficiency Virus Type 1 (HIV-1) Integrase and HIV-1 Replication at Nontoxic Concentrations**. *Journal of Medicinal Chemistry* 2002, **45**(17):3669-3683.
249. Daelemans D, Pauwels R, De Clercq E, Pannecouque C: **A time-of-drug addition approach to target identification of antiviral compounds**. *Nat Protocols* 2011, **6**(6):925-933.
250. McDougall B, King PJ, Wu BW, Hostomsky Z, Reinecke MG, Robinson WE: **Dicaffeoylquinic and Dicaffeoyltartaric Acids Are Selective Inhibitors of Human Immunodeficiency Virus Type 1 Integrase**. *Antimicrobial Agents and Chemotherapy* 1998, **42**(1):140-146.
251. Ingólfsson HI, Thakur P, Herold KF, Hobart EA, Ramsey NB, Periole X, de Jong DH, Zwama M, Yilmaz D, Hall K *et al*: **Phytochemicals Perturb Membranes and Promiscuously Alter Protein Function**. *ACS Chemical Biology* 2014, **9**(8):1788-1798.
252. Johnson AA, Marchand C, Patil SS, Costi R, Di Santo R, Burke TR, Pommier Y: **Probing HIV-1 Integrase Inhibitor Binding Sites with Position-Specific Integrase-DNA Cross-Linking Assays**. *Molecular Pharmacology* 2007, **71**(3):893-901.
253. Métifiot M, Johnson BC, Kiselev E, Marler L, Zhao XZ, Burke TR, Marchand C, Hughes SH, Pommier Y: **Selectivity for strand-transfer over 3'-processing and susceptibility to clinical resistance of HIV-1 integrase inhibitors are driven by key enzyme–DNA interactions in the active site**. *Nucleic Acids Research* 2016, **44**(14):6896-6906.
254. Hansen MST, Smith GJ, Kafri T, Molteni V, Siegel JS, Bushman FD: **Integration complexes derived from HIV vectors for rapid assays in vitro**. *Nature Biotechnology* 1999, **17**(6):578.
255. Sloan RD, Wainberg MA: **The role of unintegrated DNA in HIV infection**. *Retrovirology* 2011, **8**(1):52.
256. Vandegraaff N, Devroe E, Turlure F, Silver PA, Engelman A: **Biochemical and genetic analyses of integrase-interacting proteins lens epithelium-derived growth factor (LEDGF)/p75 and hepatoma-derived growth factor related protein 2 (HRP2) in preintegration complex function and HIV-1 replication**. *Virology* 2006, **346**(2):415-426.
257. Grandgenett DP, Goodarzi G: **Folding of the multidomain human immunodeficiency virus type-I integrase**. *Protein Science* 1994, **3**(6):888-897.
258. Métifiot M, Maddali K, Naumova A, Zhang X, Marchand C, Pommier Y: **Biochemical and Pharmacological Analyses of HIV-1 Integrase Flexible Loop Mutants Resistant To Raltegravir**. *Biochemistry* 2010, **49**(17):3715-3722.
259. Sinha S, Pursley MH, Grandgenett DP: **Efficient Concerted Integration by Recombinant Human Immunodeficiency Virus Type 1 Integrase without Cellular or Viral Cofactors**. *Journal of Virology* 2002, **76**(7):3105-3113.
260. Cherepanov P, Ambrosio AL, Rahman S, Ellenberger T, Engelman A: **Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75**. *Proc Natl Acad Sci U S A* 2005, **102**(48):17308-17313.
261. Bielefeld-Sevigny M: **AlphaLISA Immunoassay Platform— The "No-Wash" High-Throughput Alternative to ELISA**. *ASSAY and Drug Development Technologies* 2009, **7**(1):90-92.
262. Guenat S, Rouleau N, Bielmann C, Bedard J, Maurer F, Allaman-Pillet N, Nicod P, Bielefeld-Sévigny M, Beckmann JS, Bonny C *et al*: **Homogeneous and Nonradioactive High-Throughput Screening Platform for the Characterization of Kinase Inhibitors in Cell Lysates**. *Journal of Biomolecular Screening* 2006, **11**(8):1015-1026.
263. Delelis O, Malet I, Na L, Tchertanov L, Calvez V, Marcelin A-G, Subra F, Deprez E, Mouscadet J-F: **The G140S mutation in HIV integrases from raltegravir-resistant patients rescues catalytic defect due to the resistance Q148H mutation**. *Nucleic Acids Research* 2009, **37**(4):1193-1201.

264. Weiss JN: **The Hill equation revisited: uses and misuses**. *The FASEB Journal* 1997, **11**(11):835-841.
265. Hill AV: **The Combinations of Haemoglobin with Oxygen and with Carbon Monoxide. I**. *Biochem J* 1913, **7**(5):471-480.
266. Sampah MES, Shen L, Jilek BL, Siliciano RF: **Dose–response curve slope is a missing dimension in the analysis of HIV-1 drug resistance**. *Proceedings of the National Academy of Sciences* 2011, **108**(18):7613-7618.
267. do Nascimento JP, Araújo Silva JR, Lameira J, Alves CN: **Metal-dependent inhibition of HIV-1 integrase by 5CITEP inhibitor: A theoretical QM/MM approach**. *Chemical Physics Letters* 2013, **583**:175-179.
268. Leh H, Brodin P, Bischerour J, Deprez E, Tauc P, Brochon J-C, LeCam E, Coulaud D, Auclair C, Mouscadet J-F: **Determinants of Mg2+-Dependent Activities of Recombinant Human Immunodeficiency Virus Type 1 Integrase†**. *Biochemistry* 2000, **39**(31):9285-9294.
269. Rao SN, Head MS, Kulkarni A, LaLonde JM: **Validation Studies of the Site-Directed Docking Program LibDock**. *Journal of Chemical Information and Modeling* 2007, **47**(6):2159-2171.
270. Trosset J-Y, Scheraga HA: **Reaching the global minimum in docking simulations: A Monte Carlo energy minimization approach using Bezier splines**. *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95**(14):8011-8015.
271. Totrov M, Abagyan R: **Flexible protein-ligand docking by global energy optimization in internal coordinates**. *Proteins* 1997, **1**:215-220.
272. Hightower KE, Wang R, DeAnda F, Johns BA, Weaver K, Shen Y, Tomberlin GH, Carter HL, Broderick T, Sigethy S *et al*: **Dolutegravir (S/GSK1349572) Exhibits Significantly Slower Dissociation than Raltegravir and Elvitegravir from Wild-Type and Integrase Inhibitor-Resistant HIV-1 Integrase-DNA Complexes**. *Antimicrobial Agents and Chemotherapy* 2011, **55**(10):4552-4559.
273. Canducci F, Ceresola ER, Boeri E, Spagnuolo V, Cossarini F, Castagna A, Lazzarin A, Clementi M: **Cross-resistance Profile of the Novel Integrase Inhibitor Dolutegravir (S/GSK1349572) Using Clonal Viral Variants Selected in Patients Failing Raltegravir**. *Journal of Infectious Diseases* 2011, **204**(11):1811-1815.
274. Bellecave P, Malato L, Calmels C, Reigadas S, Parissi V, Andreola M-L, Fleury H: **In vitro analysis of the susceptibility of HIV-1 subtype A and CRF01_AE integrases to raltegravir**. *International Journal of Antimicrobial Agents* 2014, **44**(2):168-172.
275. Podtelezhnikov AA, Gao K, Bushman FD, McCammon JA: **Modeling HIV-1 integrase complexes based on their hydrodynamic properties**. *Biopolymers* 2003, **68**(1):110-120.
276. Al-Mawsawi LQ, Christ F, Dayam R, Debyser Z, Neamati N: **Inhibitory profile of a LEDGF/p75 peptide against HIV-1 integrase: Insight into integrase–DNA complex formation and catalysis**. *FEBS Letters* 2008, **582**(10):1425-1430.
277. Schrijvers R, De Rijck J, Demeulemeester J, Adachi N, Vets S, Ronen K, Christ F, Bushman FD, Debyser Z, Gijsbers R: **LEDGF/p75-Independent HIV-1 Replication Demonstrates a Role for HRP-2 and Remains Sensitive to Inhibition by LEDGINs**. *PLoS Pathog* 2012, **8**(3):e1002558.
278. Kobayashi M, Yoshinaga T, Seki T, Wakasa-Morimoto C, Brown KW, Ferris R, Foster SA, Hazen RJ, Miki S, Suyama-Kagitani A *et al*: **In Vitro Antiretroviral Properties of S/GSK1349572, a Next-Generation HIV Integrase Inhibitor**. *Antimicrobial Agents and Chemotherapy* 2011, **55**(2):813-821.
279. Wainberg MA, Han Y-S: **Will drug resistance against dolutegravir in initial therapy ever occur?** *Frontiers in Pharmacology* 2015, **6**:90.
280. Dierynck I, De Wit M, Gustin E, Keuleers I, Vandersmissen J, Hallenberger S, Hertogs K: **Binding Kinetics of Darunavir to Human Immunodeficiency Virus Type 1 Protease Explain the Potent Antiviral Activity and High Genetic Barrier**. *Journal of Virology* 2007, **81**(24):13845-13851.
281. Hang JQ, Yang Y, Harris SF, Leveque V, Whittington HJ, Rajyaguru S, Ao-Ieong G, McCown MF, Wong A, Giannetti AM *et al*: **Slow Binding Inhibition and Mechanism of Resistance of Non-**

nucleoside Polymerase Inhibitors of Hepatitis C Virus. *Journal of Biological Chemistry* 2009, **284**(23):15517-15529.
282. Alian A, Griner SL, Chiang V, Tsiang M, Jones G, Birkus G, Geleziunas R, Leavitt AD, Stroud RM: **Catalytically-active complex of HIV-1 integrase with a viral DNA substrate binds anti-integrase drugs**. *Proceedings of the National Academy of Sciences* 2009, **106**(20):8192-8197.
283. Charvat TT, Lee DJ, Robinson WE, Chamberlin AR: **Design, synthesis, and biological evaluation of chicoric acid analogs as inhibitors of HIV-1 integrase**. *Bioorganic & Medicinal Chemistry* 2006, **14**(13):4552-4567.

Appendix



Standards Run for SEC with kDa sizes indicated below the peaks.

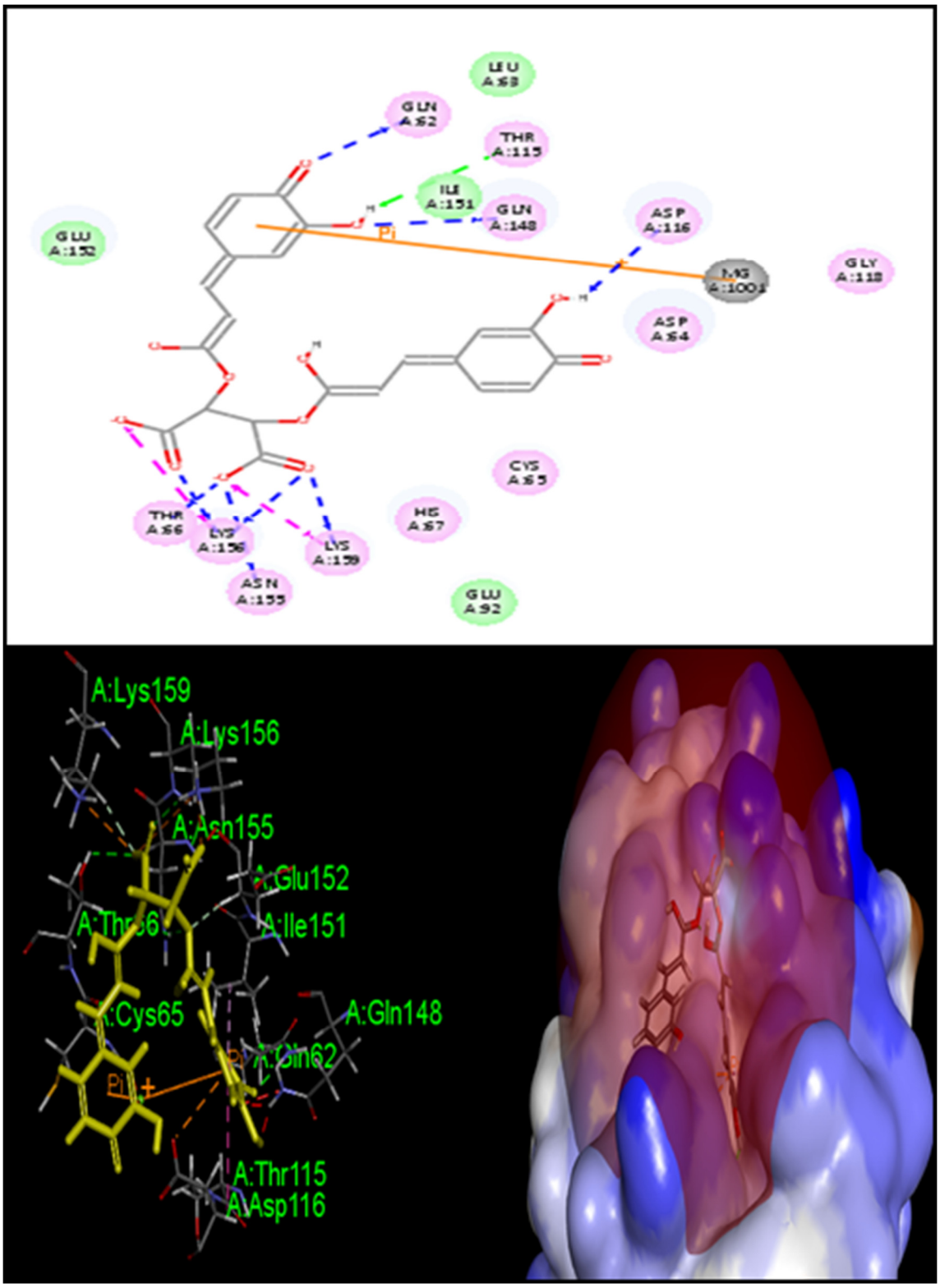


Figure 3.27 left panel. CA at active site of 1QS4

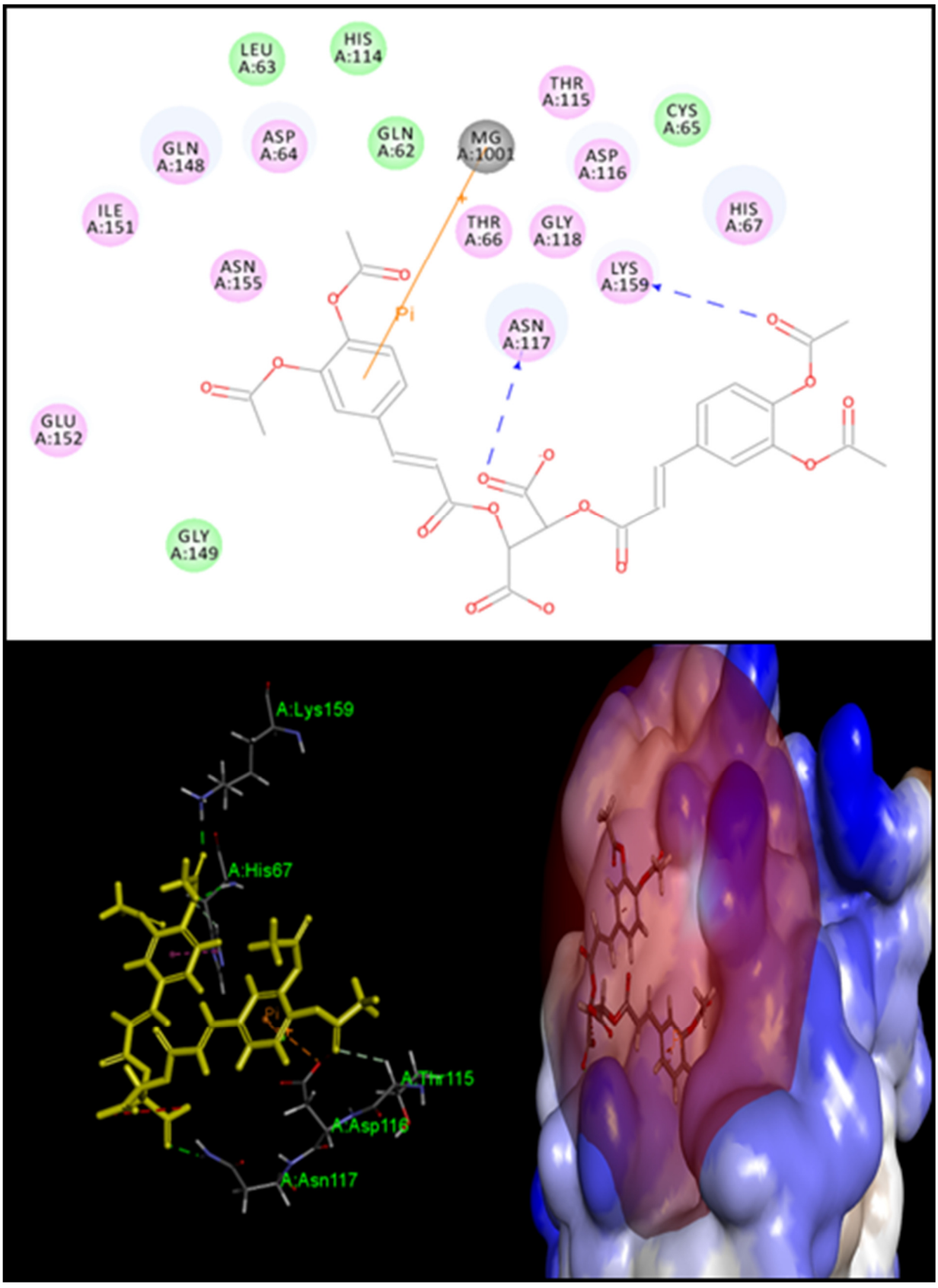


Figure 3.27 middle panel. TCA at active site of 1QS4

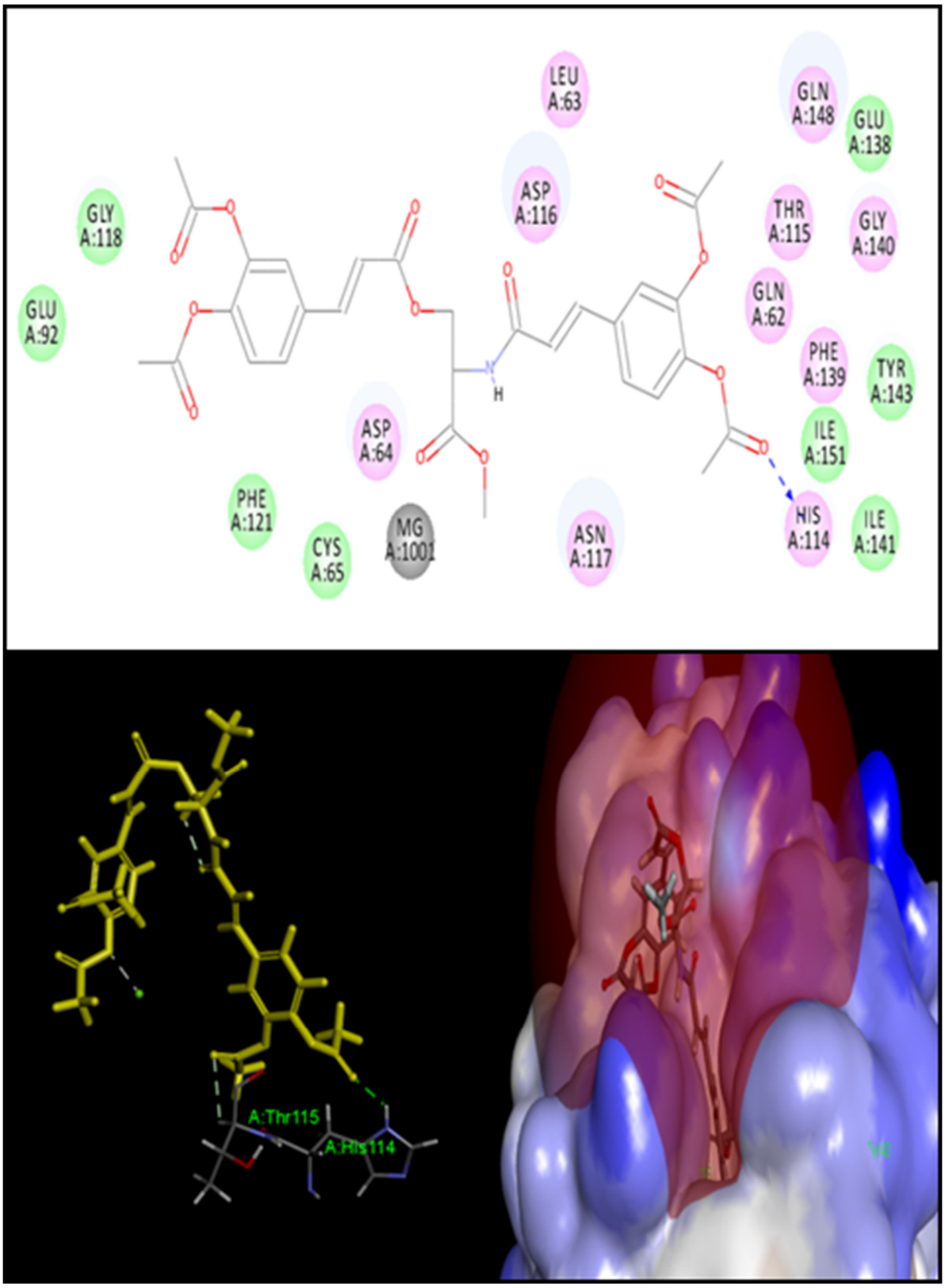


Figure 3.27 right panel. ACA-1 at active site of 1QS4

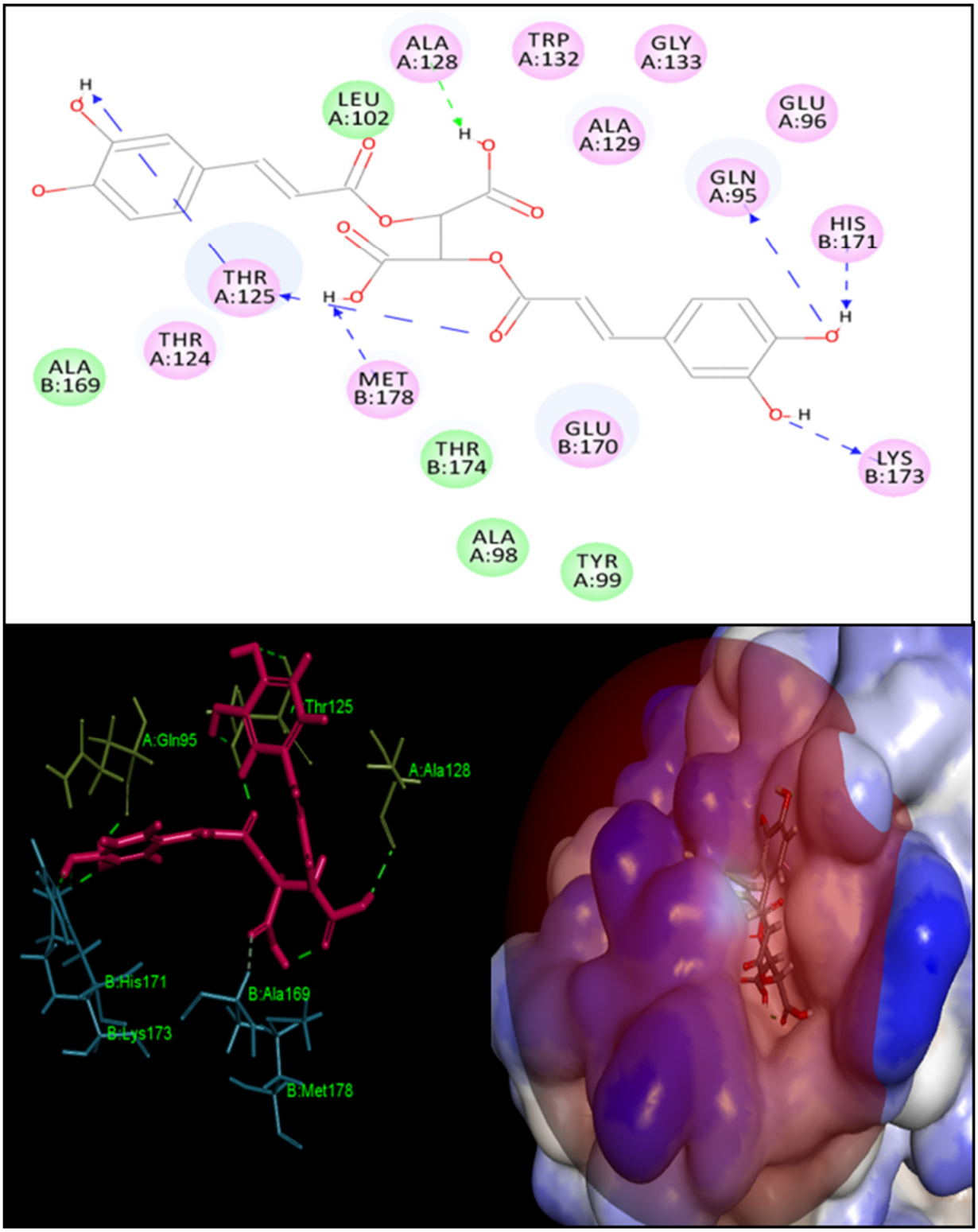


Figure 3.28 left panel. CA at allosteric site of 1QS4

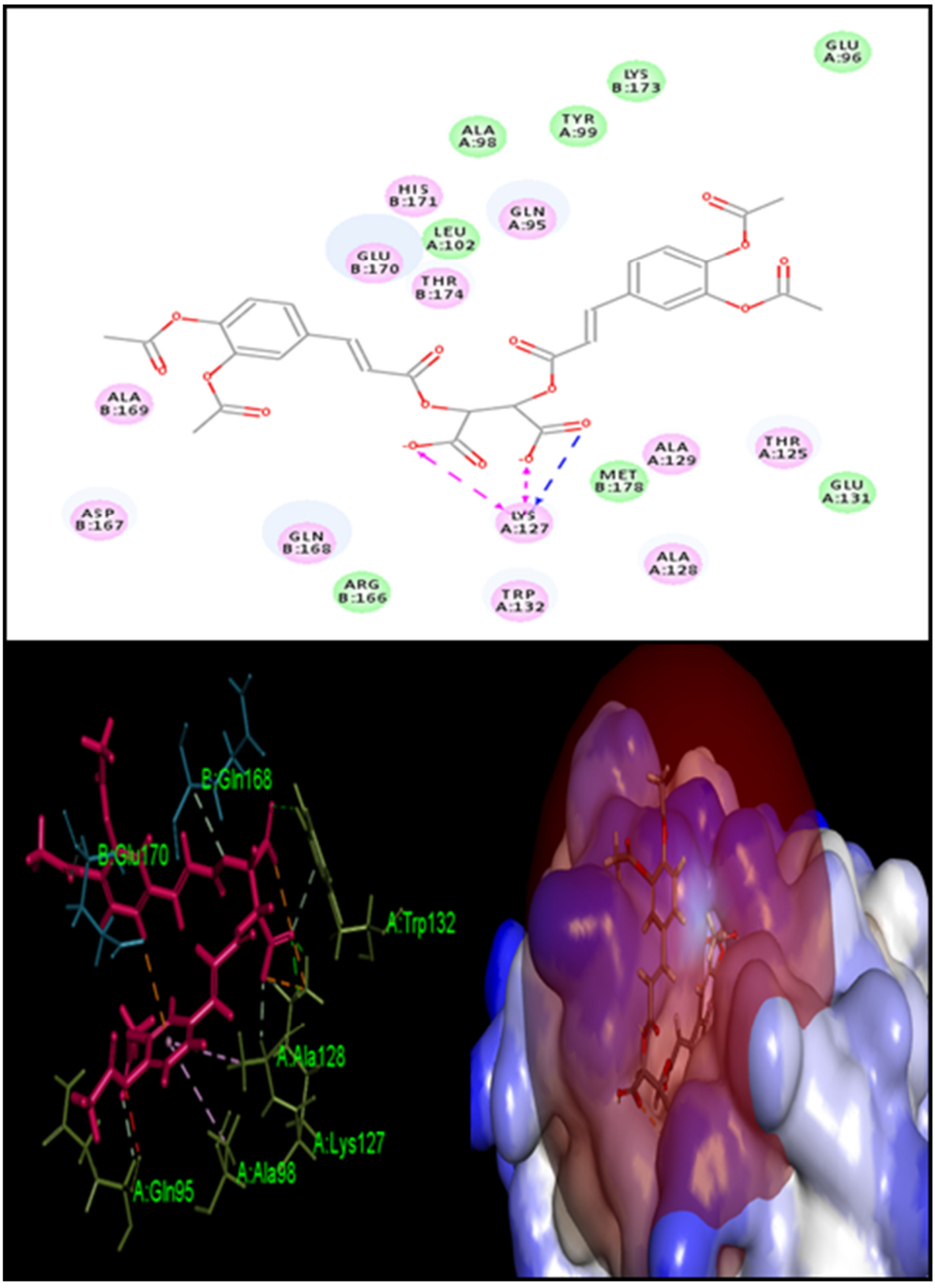


Figure 3.28 middle panel. TCA at allosteric site of 1QS4

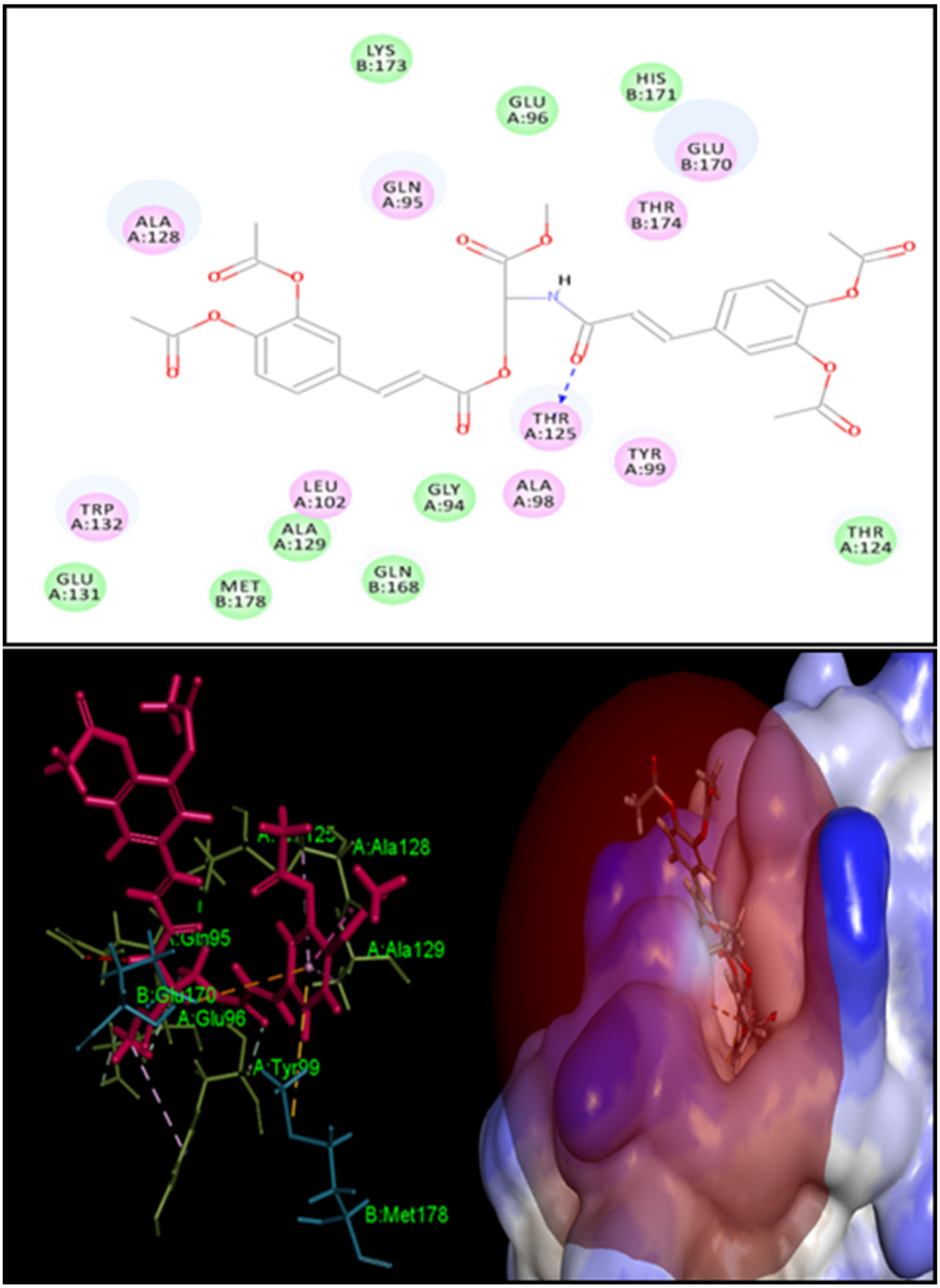


Figure 3.28 right panel. ACA-1 at allosteric site of 1QS4

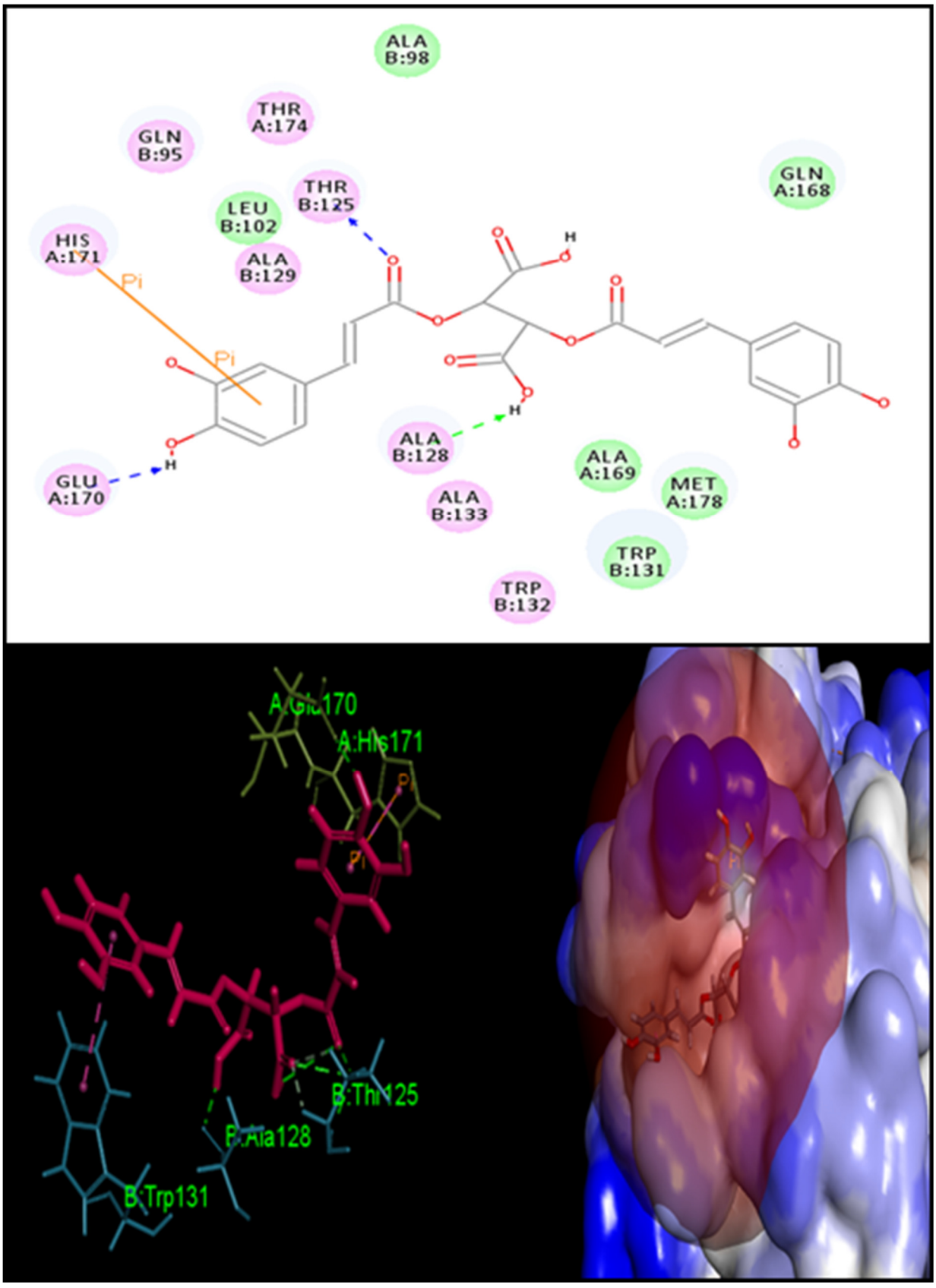


Figure 3.29 left panel. CA at allosteric site of 2B4J

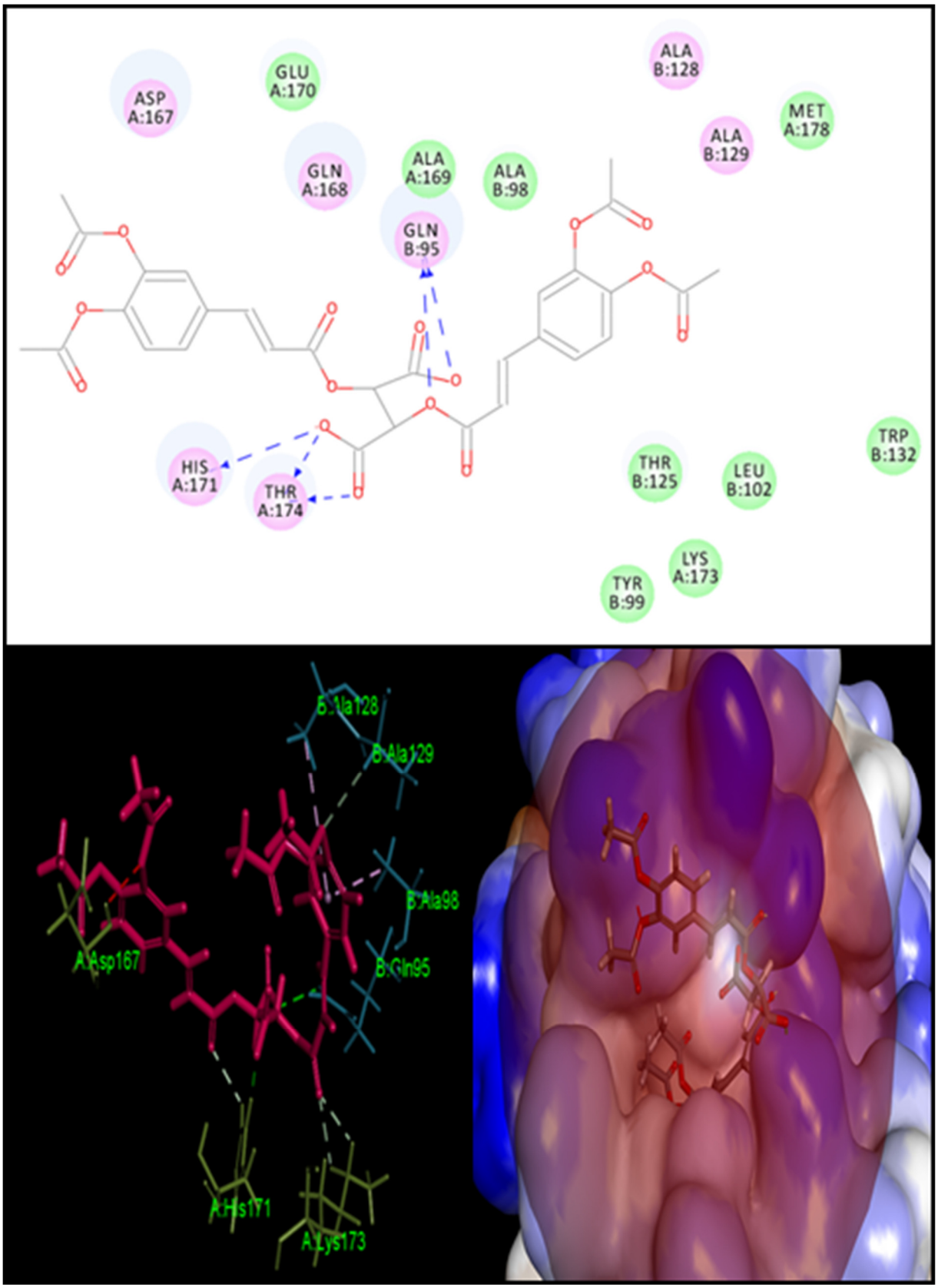


Figure 3.29 middle panel. TCA at allosteric site of 2B4J

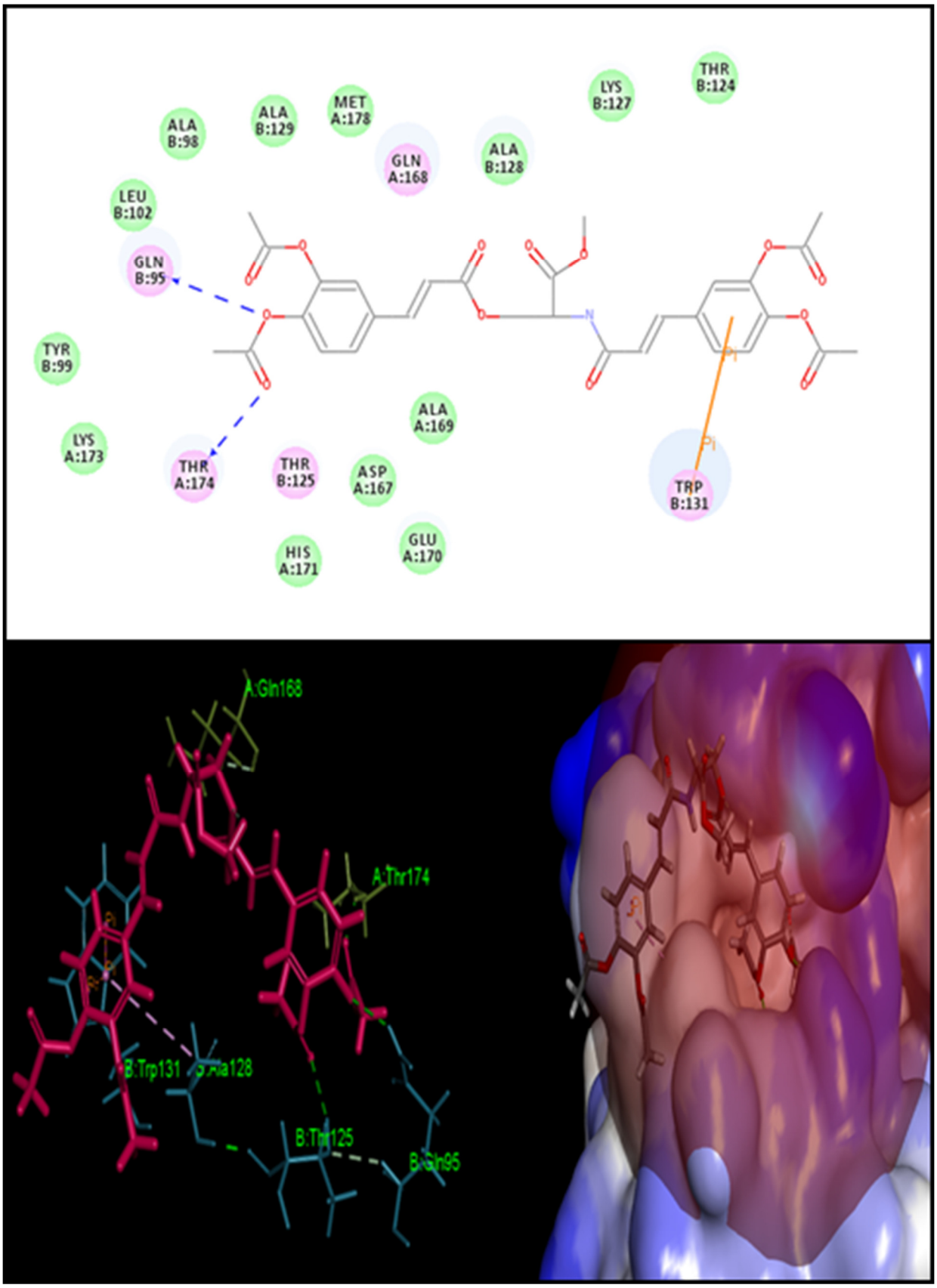


Figure 3.29 right panel. ACA-1 at allosteric site of 2B4J

R14/49 Mr Muhammad Qasim Fish

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M130474

NAME: (Principal Investigator) Mr Muhammad Qasim Fish

DEPARTMENT: Molecular Medicine & Haematology
National Health Laboratory Services

PROJECT TITLE: Determination of the Importance of the Vpr-Dynein Interaction in HIV-1 Motility towards the Nucleus

DATE CONSIDERED: Ad hoc

DECISION: Approved unconditionally

CONDITIONS:

SUPERVISOR: Dr M Papathanasopoulos

APPROVED BY:

Professor PE Cleaton-Jones, Chairperson, HREC (Medical)

DATE OF APPROVAL: 13/05/2013

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Secretary in Room 10004, 10th floor, Senate House, University.

I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. **I agree to submit a yearly progress report**.

Principal Investigator Signature Date

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES