All models are wrong, same are useful
(Anonymous)
Quannu lu diaulu te n'carizza l'anima ne vole
(La Mamma)

Ai miei genitori

Università degli Studi di Roma "*La Sapienza*"

Facoltà di Farmacia Dipartimento di Studi Farmaceutici

Istituto Pasteur-Fondazione Cenci Bolognetti

Dottorato in Scienze Pasteuriane XX Ciclo

Indolyl Aryl Sulfones, HIV-1 Non-Nucleoside Reverse Transcriptase Inhibitors: Docking and 3-D QSAR studies

Relatore:
Prof. R. Silvestri
Coordinatore
Prof. Marco Tripodi

Dottorando
Dott. Antonio Coluccia

1. Introduction

Acquired immunodeficiency syndrome (AIDS) was firstly recognized as a new disease in the United States when clinicians in New York, Los Angeles and San Francisco began to see young, homosexual men with *Pneumocystis carinii* (now *Pneumocystis jiroveci*) pneumonia and Kaposi's sarcoma, unusual diseases for young adults not known to be immunosuppressed. The first report in the medical literature that alerted the world to this new immunodeficiency syndrome appeared in June of 1981 and described five young, homosexual men in Los Angeles with *Pneumocystis carinii* pneumonia.[1] That observation was followed a few weeks later by a report of 26 homosexual men, from both New York and San Francisco, with Kaposi's sarcoma (four of whom also had *Pneumocystis carinii* pneumonia).[2] Other reports followed of a similar syndrome in injecting drug users.[3] All of these individuals shared a profound immunodeficiency, the hallmark of which was a depletion of CD4-positive, or T-helper, lymphocytes. In mid-1982, the Centers for Disease Control (CDC) reported on *Pneumocystis carinii* pneumonia among persons with hemophilia.[5] For a short time, the new disease was called gay-related immunodeficiency syndrome (GRIDS), but by September of 1982, the CDC had published a case definition, using the current designation of acquired immune deficiency syndrome (AIDS) in print, and it was rapidly adopted by researchers.[6]

The prominence of homosexual men and injecting drug users in the early cases of AIDS suggested an agent that was both blood borne and sexually transmitted, although early speculation about the etiology of AIDS included the hypothesis that all the patients were immunosuppressed because they had a history of drug use or multiple sexually transmitted

diseases or malnutrition (the "immune overload" hypothesis).[7] The majority of researchers thought that the likely agent was a sexually transmitted virus that would be found in the peripheral blood. In Pasteur institute research group, guided by Luc Montagnier, on a lymph node biopsy, from one young gay man with lymphadenopathy in the neck, François Brun-Vezinet with François Barrè-Sinoussi, after dissociated the node into single cells, and cultured the T lymphocytes discovered Reverse Transcriptase (RT) trace. The only then known were the human T cell leukemia viruses HTLV-1 HTLV-2, identified by Gallo's group. Interestingly new isolated virus doesn't precipitated by HTLV antibodies. This was the first evidence of retrovirus was agent of AIDS disease. Repeating analysis on biopsy's cells of young hemophiliac man with evidence Kaposi's sarcoma, a virus was isolated and cultured in patient's T lymphocytes killing them as T lymphocytes cells from blood donors. Electron microscopic analysis showed that HTLV and new virus weren't the same.

HIV was first isolated in France in 1983 by Françoise Barré-Sinoussi as lymphadenopathy-associated virus (LAV),[8] but strong evidence that it was the AIDS virus did not appear until 1984, when four papers were published in one issue of Science by Robert Gallo and colleagues, who designated their isolate HTLV-III.[9] The International Committee on the Taxonomy of Viruses chose the designation human immunodeficiency virus (HIV) in 1986. With the discovery by Montaignier's group in late 1986 of the related HIV-2 virus in West Africa, the original virus became HIV-1.[11]

Four years ago blood test becomes commercially available, reducing almost to zero AIDS transfusion transmission risks. In 1978 the first anti-HIV drug, AZT, was introduced and with the discovery of protease inhibitors and HAART therapy (1996) many patient are living

today[12].

In the twenty-five years since the initial discovery that AIDS is caused by HIV, the disease has grown into a global pandemic. At the end of 2007, the Joint United Nation Programme on HIV/AIDS (UNAIDS) reported more than 4 millions infection new cases, bringing the world-wide number of HIV-infected individuals up to more than 40 million[13].

2. Molecular virology of HIV

HIV is a lentivirus, from the family of retroviruses, which characteristically have an RNA genome contained within a capsid and a lipid envelope (*Figure 1*).

2.1 Viral structure

The viral envelope, a bi-layered membrane derived from the host cell, contains 2 major viral glycoproteins, gp41 and gp120. These originate from enzymatic cleavage of the larger viral pre-protein gp160; they mediate viral entry and syncytium formation. Gp120 has a variable protein domain containing the V3 loop, which elicits a strong immune response.[14] The HIV core is composed of 3 structural proteins, p24, p16, and p9. The p24 protein forms the capsid that encloses 2 genomic RNA strands and the viral enzymes reverse transcriptase, protease, ribonuclease, and integrase. The matrix protein, gp16, is anchored to the internal face of the envelope. P9 is a nucleocapsid protein not covalently attached to the viral RNA.[15]

The genome of HIV, similar to retroviruses in general, contains three major genes: gag, pol, and env. These genes code for the major structural and functional components of HIV, including envelope proteins and reverse transcriptase. The major structural components coded by env include the envelope glycoproteins, including the outer envelope glycoprotein gp120 and transmembrane glycoprotein gp41 derived from glycoprotein precursor gp160. Major components coded by the gag gene include core nucleocapsid proteins p55, p40, p24 (capsid, or "core" antigen), p17(matrix), and p7 (nucleocapsid); the important proteins coded by pol are the enzyme proteins p66 and p51 (reverse transcriptase), p11 (protease), and p32 (integrase).[16]

Accessory genes carried by HIV include tat, rev, nef, vif, vpr, and vpu (for HIV-1) or vpx (for HIV-2). The functions of some of these genes are known. The tat gene produces a regulatory protein that speeds up transcription of the HIV provirus. The rev gene encodes for a regulatory protein which switches the processing of viral RNA transcripts to a pattern that predominates with established infection, leading to production of viral structural and enzymatic proteins. The nef gene produces a regulatory protein that modifies the infected cell to make it more suitable for producing HIV virions. The vif, vpr, and vpu genes encode proteins that appear to play a role in generating infectivity and pathologic effects. *Figure 2* shows an overview of the organization of the (approximately) 9-kilobase genome of the HIV provirus and a summary of the functions of its nine genes encoding 15 proteins[17].
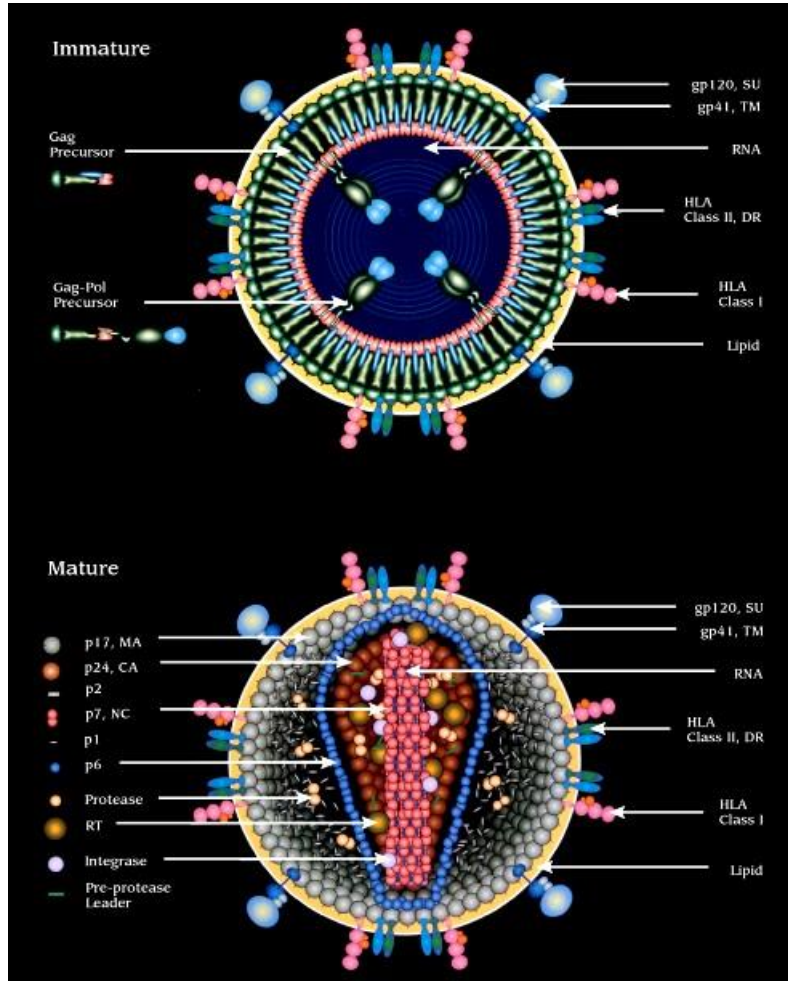
Figure 1. HIV structure.

2.2 Mechanism of replication

HIV life cycle (*Figure 3)* starts when the virus enters lymphocytes and monocytes through

cognate recognition of the viral glycoprotein gp120 with the cell surface CD4 molecule and a chemokine receptor (either CXCR4 or CCR5).[18,19] The two viral envelope proteins, gp120 and gp41, are conformationally associated to form a trimeric functional unit consisting of three molecules of gp120 exposed on the virion surface and associated with three molecules of gp41 inserted into the viral lipid membrane. Trimeric gp120 on the surface of the virion binds CD4 on the surface of the target cell, inducing a conformational change in the envelope proteins that in turn allows binding of the virion to a specific subset of chemokine receptors on the cell surface (*Figure 4*). These receptors normally play a role in chemoattraction, in which hematopoietic cells move along chemokine gradients to specific sites. Although these receptors, which contain seven membrane-spanning domains, normally transducer signals through G proteins, signaling is not required for HIV infection.[20]

The binding of surface gp120, CD4, and the chemokine coreceptors produces an additional radical conformational change in gp41. Assembled as a trimer on the virion membrane, this coiled-coil protein springs open, projecting three peptide fusion domains that "harpoon" the lipid bilayer of the target cell. The fusion domains then form hairpin-like structures that draw the virion and cell membranes together to promote fusion, leading to the release of the viral core into the cell interior.[21]

HIV virions can also enter cells by endocytosis. Usually, productive infection does not result, presumably reflecting inactivation of these virions within endosomes. However, a special form of endocytosis has been demonstrated in submucosal dendritic cells. These cells, which normally process and present antigens to immune cells, express a specialized attachment structure termed DC-SIGN.[22] This C-type lectin binds HIV gp120 with high

affinity but does not trigger the conformational changes required for fusion. Instead, virions bound to DC-SIGN are internalized into an acidic compartment and subsequently displayed on the cell surface after the dendritic cell has matured and migrated to regional lymph nodes, where it engages T cells. Thus, dendritic cells expressing DC-SIGN appear to act as "Trojan horses" facilitating the spread of HIV from mucosal surfaces to T cells in lymphatic organs.[23]
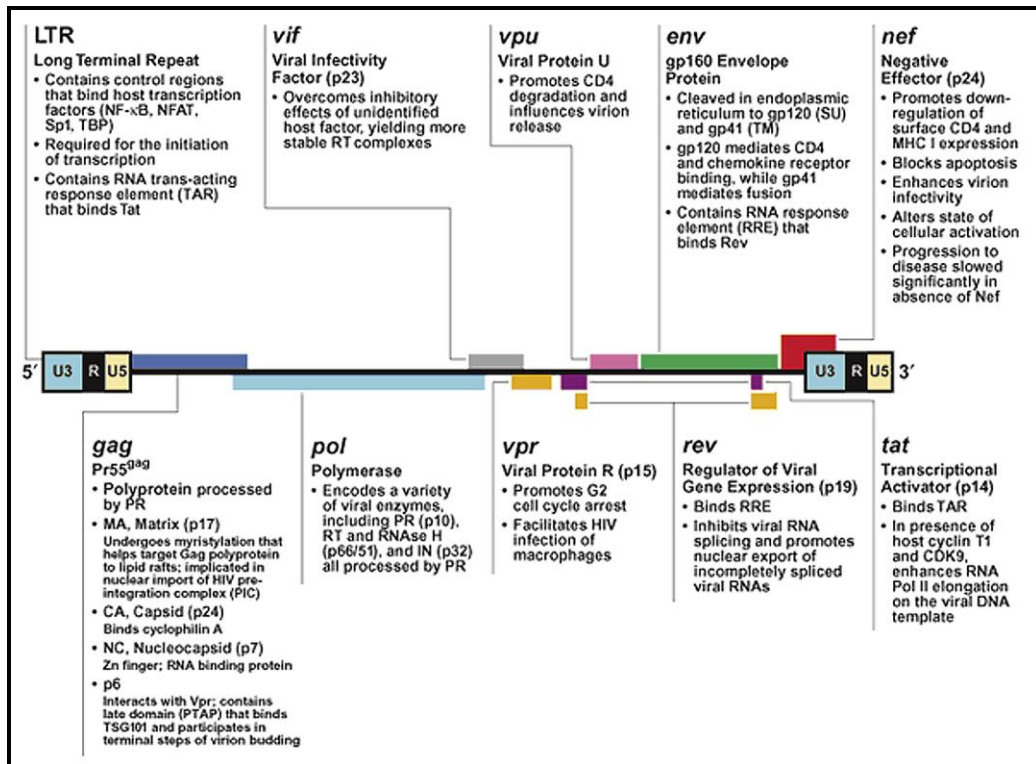

Figure 2. Organization of the HIV proviral genome and summary of gene product functions.

On the basis of cell tropism, HIV strains can be broadly divided into 2 categories, macrophage-tropic (M-tropic) and T-cell tropic (T-tropic).

M-tropic strains use CCR5 as a coreceptor and are referred as R5 viruses. They primarily infect macrophages and primary T cells and infect poorly CD4[+] T-cell lines. In addition, these viruses tend to be transmitted sexually more easily.[24] T-tropic strains use the CXCR4 coreceptor, which is most expressed in CD4[+] T cells. Also referred as X4 viruses, they induce the formation of syncytia in the infected cells[25]. Early in the course of HIV infection, the R5 strain viruses predominate, but eventually both X4 and R5 strains are recovered. Mutations in the CCR5 gene protect cells from HIV infection. These mutations have not shown to be deleterious, probably because other chemokine receptors replace CCR5 functions. The level of expression of chemokine receptors also determines the ability of HIV strains to infect host cells. Recently, a primary HIV isolate has been reported to enter T cells using the CD8 molecule, not requiring CD4 or chemokine receptor expression.[26]
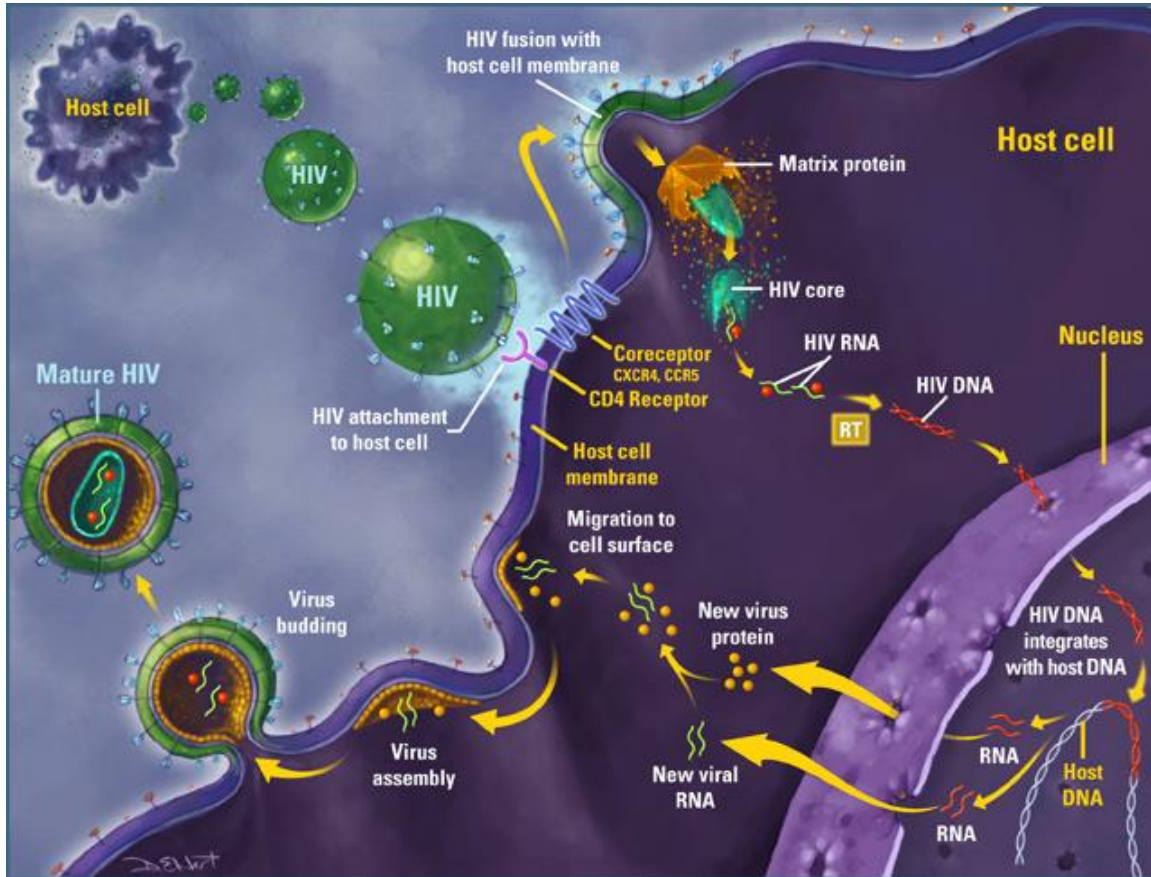
Figure 3. HIV life cycle.

Once inside the cell, the virion undergoes uncoating, likely while still associated with the plasma membrane (*Figure 4*). This poorly understood process may involve phosphorylation of viral matrix proteins by a mitogen-activated protein (MAP) kinase and additional actions of

cyclophilin A and the viral proteins Nef and Vif. Nef associates with a universal proton pump, V-ATPase, which could promote uncoating by inducing local changes in pH in a manner similar to that of the M2 protein of influenza.[18] After the virion is uncoated, the viral reverse transcription complex is released from the plasma membrane.[26] This complex includes the diploid viral RNA genome, lysine transfer RNA (tRNALys) which acts as a primer for reverse transcription, viral reverse transcriptase, integrase, matrix and nucleocapsid proteins, viral protein R (Vpr), and various host proteins. The reverse transcription complex docks with actin microfilaments[28]. This interaction, mediated by the phosphorylated matrix, is required for efficient viral DNA synthesis. By overcoming destabilizing effects of a recently identified protein termed CEM15/APOBEC3G, Vif stabilizes the reverse transcription complex in most human cells.[26-28]
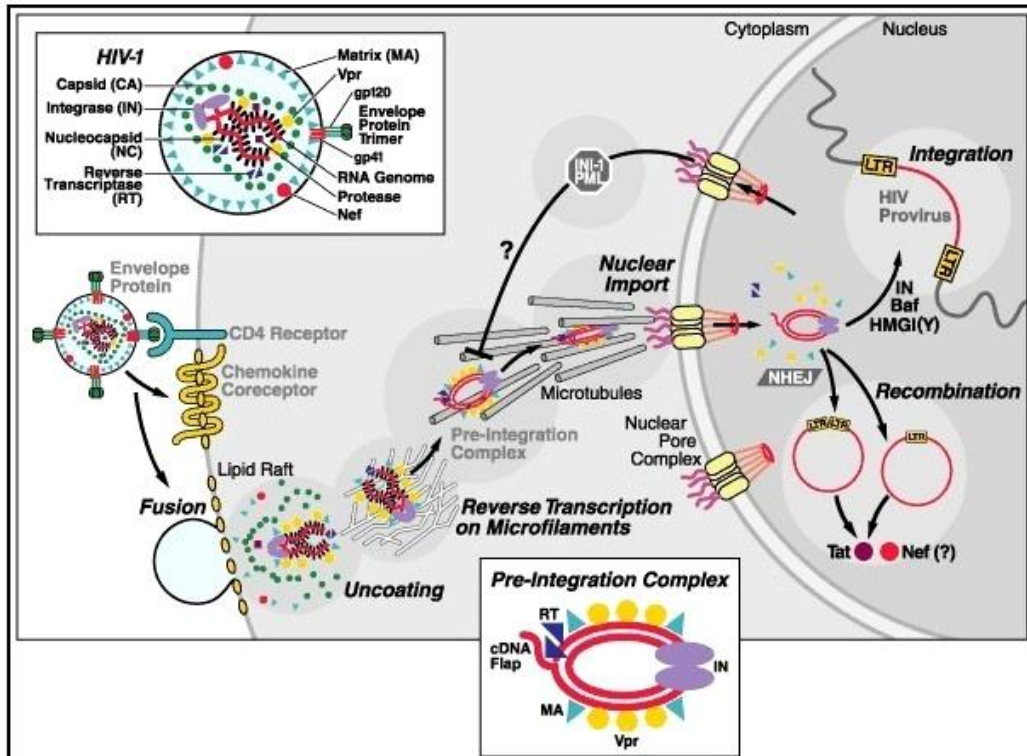
Figure 4. Early events occurring after HIV infection of a susceptible target cell.

The viral RNA is reverse-transcribed into complementary DNA (cDNA) by the virus reverse transcriptase through use of a cellular lysine tRNA molecule as a primer; subsequently, the RNAase activity of the reverse transcriptase degrades the viral RNA template. The reverse transcriptase incorporates an incorrect nucleotide every 1500 to 4000 bases, which explains the rapid occurrence of mutations. Some of the resulting mutations

provide a survival advantage, leading to drug-resistant strains.[23]

Reverse transcription yields the HIV preintegration complex (PIC), composed of double-stranded viral cDNA, integrase, matrix, Vpr, reverse transcriptase, and the high mobility group DNA-binding cellular protein HMGI(Y). The PIC may move toward the nucleus by using microtubules as a conduit.[30]

Unlike most animal retroviruses, HIV can infect nondividing cells, such as terminally differentiated macrophages. This requires an ability to cross the intact nuclear membrane. With a Stokes radius of approximately 28 nm or roughly the size of a ribosome, the PIC is roughly twice as large as the maximal diameter of the central aqueous channel in the nuclear pore. The 3 µm contour length of viral DNA must undergo significant compaction, and the import process must involve considerable molecular gymnastics.[31]

One of the most contentious areas of HIV research involves the identification of key viral proteins that mediate the nuclear import of the PIC. Integrase, matrix, and Vpr have been implicated (*Figure 4*). Because plus-strand synthesis is discontinuous in reverse transcription, a triple helical DNA domain or "DNA flap" results that may bind a host protein containing a nuclear targeting signal.[32] Matrix contains a canonical nuclear localization signal that is recognized by the importins α and β, which are components of the classical nuclear import pathway. However, a recent publication calls into question the contributions both of the nuclear import signal in integrase and of the DNA flap to the nuclear uptake of the PIC.[31] The HIV Vpr gene product contains at least three noncanonical nuclear targeting signals. Vpr may bypass the importin system altogether, perhaps mediating the direct docking of the PIC with one or more components of the nuclear pore complex. The multiple nuclear targeting signals

within the PIC may function in a cooperative manner or play larger roles individually in different target cells. For example, while Vpr is not needed for infection of nondividing, resting T cells, it enhances viral infection in nondividing macrophages.[35] The finding that both matrix and Vpr shuttle between the nucleus and cytoplasm explains their availability for incorporation into new virions.[36]

Once inside the nucleus, the viral PIC can establish a functional provirus (*Figure 4*). Integration of double-stranded viral DNA into the host chromosome is mediated by integrase, which binds the ends of the viral DNA. The host proteins HMGI(Y) and barrier to autointegration (BAF) are required for efficient integration, although their precise functions remain unknown. Integrase removes terminal nucleotides from the viral DNA, producing a two-base recess and thereby correcting the ragged ends generated by the terminal transferase activity of reverse transcriptase. Integrase also catalyzes the subsequent joining reaction that establishes the HIV provirus within the chromosome.[29]

Not all PICs that enter the nucleus result in a functional provirus. The ends of the viral DNA may be joined to form a 2-LTR circle containing long terminal repeat sequences from both ends of the viral genome, or the viral genome may undergo homologous recombination yielding a single-LTR circle. Finally, the viral DNA may auto-integrate into itself, producing a rearranged circular structure. Although some circular forms may direct the synthesis of the transcriptional transactivator Tat or the accessory protein Nef, none produces infectious virus.[37] In a normal cellular response to DNA fragments, the nonhomologous end-joining (NHEJ) system may form 2-LTR circles to protect the cell.[38] This system is responsible for rapid repair of double-strand breaks, thereby preventing an apoptotic response. A single

double-strand break within the cell can induce G1 cell-cycle arrest. The ability of the free ends of the viral DNA to mimic such double-strand chromosomal breaks may contribute to the direct cytopathic effects observed with HIV.

Integration can lead to latent or transcriptionally active forms of infection. Moreover, despite a vigorous immune response early in infection, these silent proviruses are a reservoir that allows reemergence of HIV when the body's defenses grow weaker.

The chromosomal environment likely shapes the transcriptional activity of the provirus.[39] For example; proviral integration into repressed heterochromatin might result in latency (*Figure 5*). Other causes of latency may include cell type differences in the availability of activators that bind to the transcriptional enhancer in the HIV LTR or the lack of Tat. However, of the multiple copies of provirus that are usually integrated in a given infected cell, at least one is likely to be transcriptionally active. This fact may explain why the number of latently infected cells in infected patients is small.

In the host genome, the 5´ LTR functions like other eukaryotic transcriptional units. It contains downstream and upstream promoter elements, which include the initiator (Inr), TATA-box (T), and three Sp1 sites [40]. These regions help position the RNA polymerase II (RNAPII) at the site of initiation of transcription and to assemble the preinitiation complex. Slightly upstream of the promoter is the transcriptional enhancer, which in HIV-1 binds nuclear factor kB (NFkB), nuclear factor of activated T cells (NFAT), and Ets family members.[41] NFkB and NFAT relocalize to the nucleus after cellular activation. NFkB is liberated from its cytoplasmic inhibitor, IkB, by stimulus-coupled phosphorylation, ubiquitination, and proteosomal degradation of the inhibitor. NFAT is

dephosphorylated by calcineurin (a reaction inhibited by cyclosporin A) and, after its nuclear import, assembles with AP1 to form the fully active transcriptional complex. NFkB, which is composed of p50 and p65 (RelA) subunits, increases the rates of initiation and elongation of viral transcription.[42] Since NFkB is activated after several antigen-specific and cytokine-mediated events, it may play a key role in rousing transcriptionally silent proviruses.
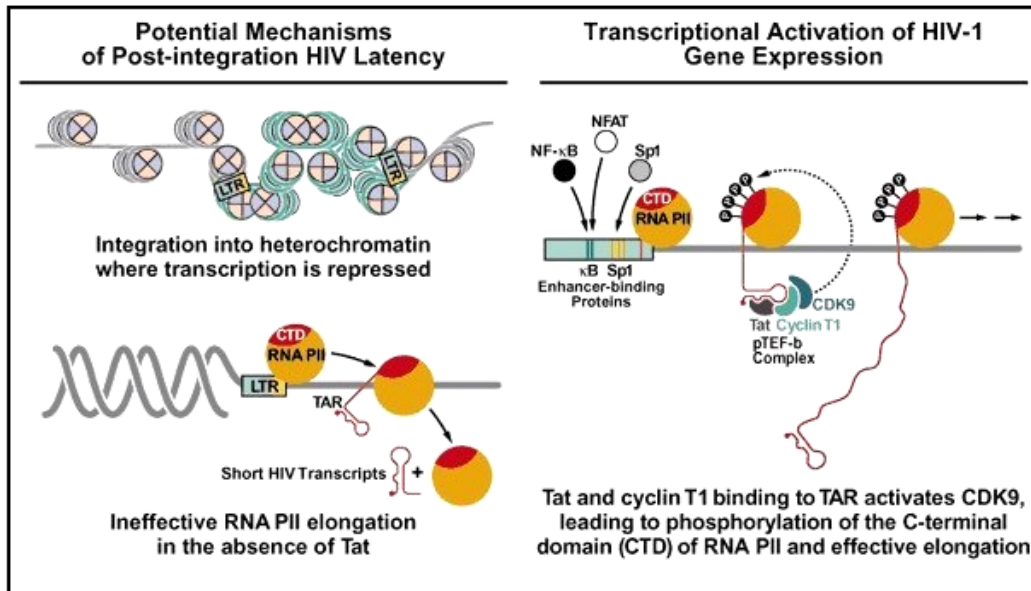

Figure 5. Mechanisms of post-integration latency.

When these factors engage the LTR, transcription begins, but in the absence of Tat described below the polymerase fails to elongate efficiently along the viral genome (*Figure 5*). In the process, short nonpolyadenylated transcripts are synthesized, which are stable and persist in cells due to the formation of an RNA stem loop called the transactivation response (TAR) element.[43]

Tat significantly increases the rate of viral gene expression. With cyclin T1 (CycT1), Tat binds to the TAR RNA stem-loop structure and recruits the cellular cyclin-dependent kinase 9 (Cdk9) to the HIV LTR (*Figure 5*). Within the positive transcription elongation factor b (P-TEFb) complex, Cdk9 phosphorylates the C-terminal domain of RNAPII, marking the transition from initiation to elongation of eukaryotic transcription.[44] Other targets of P-TEFb include negative transcription elongation factors (N-TEF), such as the DRB-sensitivity inducing (DSIF) and negative elongation (NELF) factors.[44] The high efficiency with which the HIV LTR attracts these negative transcription factors *in vivo* may explain why the LTR is a poor promoter in the absence of Tat. The arginine-rich motif (ARM) within Tat binds the 5' bulge region in TAR. A shorter ARM in cyclin T1, which is also called the Tat-TAR recognition motif (TRM), binds the central loop of TAR.

Binding of the Tat cyclin T1 complex to both the bulge and loop regions of TAR strengthens the affinity of this interaction. All of these components are required for Tat transactivation. In the presence of the complex between Tat and P-TEFb, the RNAPII elongates efficiently. Because murine CycT1 contains a cysteine at position 261, the complex between Tat and murine P-TEFb binds TAR weakly.[45] Thus, Tat transactivation is severely compromised in murine cells. Cdk9 also must undergo autophosphorylation of

several serine and threonine residues near its C-terminus to allow productive interactions between Tat, P-TEFb, and TAR. Additionally, basal levels of P-TEFb may be low in resting cells or only weakly active due to the interaction between P-TEFb and 7SK RNA.[46] All of these events may contribute to postintegration latency.

Transcription of the viral genome results in more than a dozen different HIV-specific transcripts. Some are processed cotranscriptionally and, in the absence of inhibitory RNA sequences (IRS), transported rapidly into the cytoplasm. These multiply spliced transcripts encode Nef, Tat, and Rev. Other singly spliced or unspliced viral transcripts remain in the nucleus and are relatively stable. These viral transcripts encode the structural, enzymatic, and accessory proteins and represent viral genomic RNAs that are needed for the assembly of fully infectious virions.

Incomplete splicing likely results from suboptimal splice donor and acceptor sites in viral transcripts. In addition, the regulator of virion gene expression, Rev, may inhibit splicing by its interaction with alternate splicing factor/splicing factor 2 (ASF/SF2) and its associated p32 protein.[47]

Transport of the incompletely spliced viral transcripts to the cytoplasm depends on an adequate supply of Rev. Rev is a small shuttling protein that binds a complex RNA stem-loop termed the Rev response element (RRE), which is located in the env gene. Rev binds first with high affinity to a small region of the RRE termed the stem-loop IIB (*Figure 6*).[48] This binding leads to the multimerization of Rev on the remainder of the RRE. In addition to a nuclear localization signal, Rev contains a leucine-rich nuclear export sequence (NES). Of note, the study of Rev was the catalyst for the discovery of such NES in many

cellular proteins and led to identification of the complex formed between CRM1/exportin-1 and this sequence.[49]

The nuclear export of this assembly (viral RNA transcript, Rev, and CRM1/exportin 1) depends critically on yet another host factor, RanGTP. Ran is a small guanine nucleotide-binding protein that switches between GTP- and GDP-bound states. RanGDP is found predominantly in the cytoplasm because the GTPase activating protein specific for Ran (RanGAP) is expressed in this cellular compartment. Conversely, the Ran nucleotide exchange factor, RCC1, which charges Ran with GTP, is expressed predominantly in the nucleus. The inverse nucleocytoplasmic gradients of RanGTP and RanGDP produced by the subcellular localization of these enzymes likely plays a major role in determining the directional transport of proteins into and out of the nucleus. Outbound cargo is only effectively loaded onto CRM1/exportin-1 in the presence of RanGTP. However, when the complex reaches the cytoplasm, GTP is hydrolyzed to GDP, resulting in release of the bound cargo. The opposite relationship regulates the nuclear import by importing α and β, where nuclear RanGTP stimulates cargo release.[51]

For HIV infection to spread, a balance between splicing and transport of viral mRNA species must be achieved. If splicing is too efficient, then only the multiply spliced transcripts appear in the cytoplasm. Although required, the regulatory proteins encoded by multiply spliced transcripts are insufficient to support full viral replication. However, if splicing is impaired, adequate synthesis of Tat, Rev, and Nef will not occur. In many non-primate cells, HIV transcripts may be overly spliced, effectively preventing viral replication in this hosts.[50]
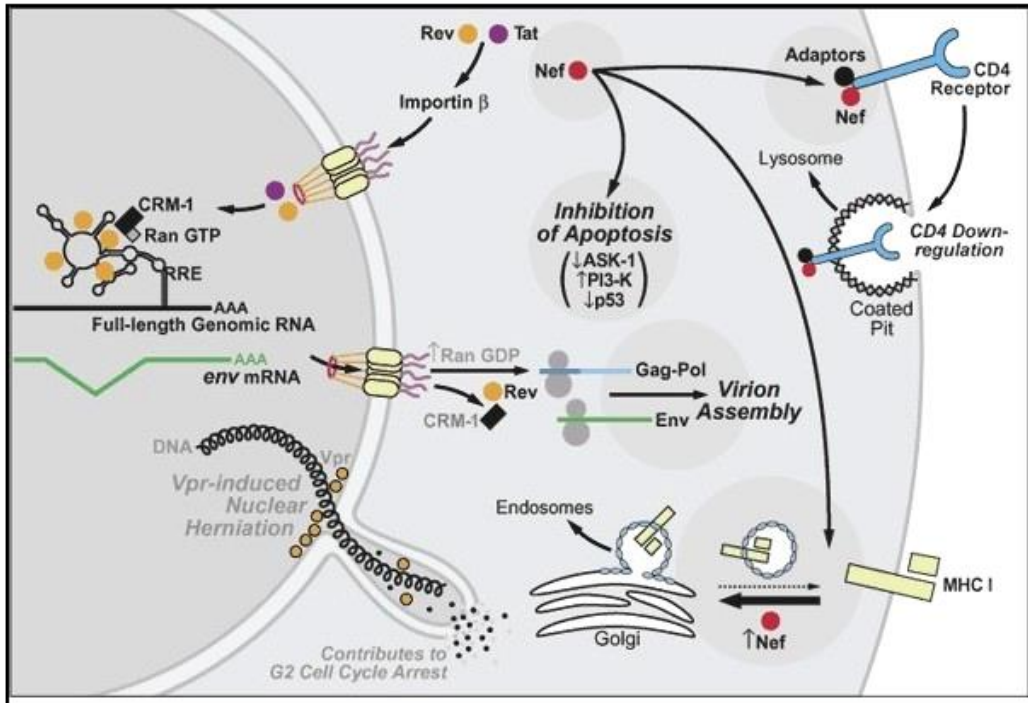
Figure 6. Late events in the HIV-infected cell culminating in the assembly of new infectious virions.

In contrast to Tat and Rev, which act directly on viral RNA structures, Nef modifies the environment of the infected cell to optimize viral replication (*Figure 6*). The absence of Nef in infected monkeys and humans is associated with much slower clinical progression to AIDS. This virulence caused by Nef appears to be associated with its ability to affect signaling cascades, including the activation of T-cell antigen receptor, and to decrease the expression of

CD4 on the cell surface.[51] Nef also promotes the production and release of virions that are more infectious.[52] Effects of Nef on the PI3-K signaling cascade - which involves the guanine nucleotide exchange factor Vav, the small GTPases Cdc42 and Rac1, and p21-activated kinase PAK - cause marked changes in the intracellular actin network, promoting lipid raft movement and the formation of larger raft structures that have been implicated in T-cell receptor signaling.[53] Indeed, Nef and viral structural proteins colocalize in lipid rafts.[54] Two other HIV proteins assist Nef in downregulating expression of CD4.[55] The envelope protein gp120 binds CD4 in the endoplasmic reticulum, slowing its export to the plasma membrane, and Vpu binds the cytoplasmic tail of CD4, promoting recruitment of TrCP and Skp1p (*Figure 7*). These events target CD4 for ubiquitination and proteasomal degradation before it reaches the cell surface.[56]

Nef acts by several mechanisms to impair immunological responses to HIV. In T cells, Nef activates the expression of FasL, which induces apoptosis in bystander cells that express Fas, thereby killing cytotoxic T
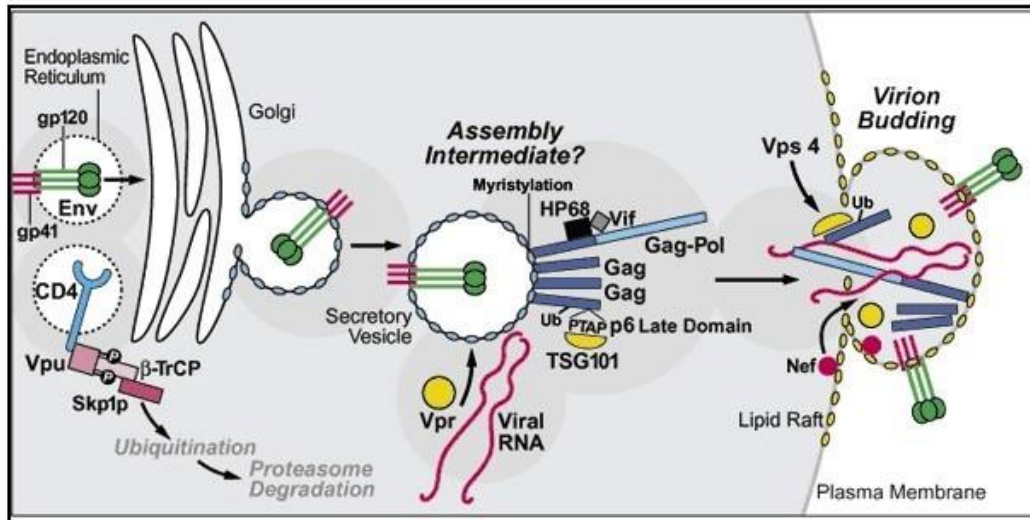
Figure 7. Late steps in the assembly of new virions.

cells that might otherwise eliminate HIV-1 infected cells. Nef also reduces the expression of MHC I determinants on the surface of the infected cell (*Figure 6*) and so decreases the recognition and killing of infected cells by CD8 cytotoxic T cells.[58] However, Nef does not decrease the expression of HLA-C,[59] which prevents recognition and killing of these infected cells by natural killer cells.

Nef also inhibits apoptosis. It binds and inhibits the intermediate apoptosis signal regulating kinase-1 (ASK-1) that functions in the Fas and TNFR death signaling pathways and stimulates the phosphorylation of Bad leading to its sequestration by 14-3-3 proteins (*Figure 6*).[60] Nef also binds the tumor suppressor protein p53, inhibiting another potiential initator of apoptosis.

Via these different mechanisms, Nef prolongs the life of the infected host cell, thereby optimizing viral replication. Other viral proteins also participate in the modification of the environment in infected cells. Rev-dependent expression of Vpr induces the arrest of proliferating infected cells at the G2/M phase of the cell cycle.[61] Since the viral LTR is more active during G2, this arrest likely enhances viral gene expression.[62] These cell-cycle arresting properties involve localized defects in the structure of the nuclear lamina that lead to dynamic, DNA-filled herniations that project from the nuclear envelope into the cytoplasm (*Figure 6*).[63] Intermittently, these herniations rupture, causing the mixing of soluble nuclear and cytoplasmic proteins. Either alterations in the lamina structure or the inappropriate mixing of cell cycle regulators that are normally sequestered in specific cellular compartments could explain the G2 arresting properties of Vpr.

New viral particles are assembled at the plasma membrane (*Figure 7*). Each virion consists of roughly 1500 molecules of Gag and 100 Gag-Pol polyproteins,[64] two copies of the viral RNA genome, and Vpr.[65] Several proteins participate in the assembly process, including Gag polyproteins and Gag-Pol, as well as Nef and Env. A human ATP-binding protein, HP68 (previously identified as an RNase L inhibitor), likely acts as a molecular chaperone, facilitating conformational changes in Gag needed for the assembly of viral capsids.[66] In primary CD4 T lymphocytes, Vif plays a key but poorly understood role in the assembly of infectious virions. In the absence of Vif, normal levels of virus are produced, but these virions are noninfectious, displaying arrest at the level of reverse transcription in the subsequent target cell. Heterokaryon analyses of cells formed by the fusion of nonpermissive (requiring Vif for viral growth) and permissive (supporting growth of Vif-deficient viruses) cells have revealed

that Vif overcomes the effects of a natural inhibitor of HIV replication.[67] Recently this factor, initially termed CEM15/APOBEC3G, was identified and shown to share homology with APOBEC1, an enzyme involved in RNA editing.[68] Whether the intrinsic antiviral activity of CEM15 involves such an RNA editing function remains unknown. CEM15 is expressed in nonpermissive but not in permissive cells and when introduced alone is sufficient to render permissive cells nonpermissive.

The Gag polyproteins are subject to myristylation, and thus associate preferentially with cholesterol- and glycolipid-enriched membrane microdomains [69]. Virion budding occurs through these specialized regions in the lipid bilayer, yielding virions with cholesterol-rich membranes. This lipid composition likely favors release, stability, and fusion of virions with the subsequent target cell.

The budding reaction involves the action of several proteins, including the "late domain" sequence (PTAP) present in the p6 portion of Gag (*Figure 7*).[70] The p6 protein also appears to be modified by ubiquitination. The product of the tumor suppressor gene 101 (TSG101) binds the PTAP motif of p6 Gag and also recognizes ubiquitin through its ubiquitin enzyme 2 (UEV) domain. The TSG101 protein normally associates with other cellular proteins in the vacuolar protein sorting pathway to form the ESCRT-1 complex that selects cargo for incorporation into the multivesicular body (MVB).[71] The MVB is produced when surface patches on late endosomes bud away from the cytoplasm and fuse with lysosomes, releasing their contents for degradation within this organelle. In the case of HIV, TSG101 appears to be "hijacked" to participate in the budding of virions into the extracellular space away from the cytoplasm.

3 Treatment of AIDS

3.1 Diagnosis

HIV infection was detected by two different ways. The first based on antibody engender by immunitary system, (serologic manner) or by the finding of virus molecules and antigenic (virus manner). The diagnostic analysis was based on ELISA and Western-Blot tests.

ELISA is an enzymatic process widely used as screening method thanks to its low costs and easy implementation. With this method is possible to find out the antibody versus viral′s proteins, for example gp41 e gp120. Sensibility of ELISA test is among 95%, but sometimes could be bias answer, called "lie positive" in witch the test result shows positive reply but without infection, or "lie negative" in with test result shows negative reply but with infection. One of causes of "lie positive" could be the presence of other pathologies that give production of different antibody. The "lie negative" for example could happen in persons that take the infection very yet.

The Western Blot (WB) is a more sensible and specific than ELISA test used to confirm the result of ELISA test. WB method gives evidence of presence of viral proteins antibody. The test′s results are positive when at least two antibodies are found.

In advance to these tests there are other methods based again on the research of viral proteins but not for a diagnostic use but for the screening of infection, in particular during

HAART therapy. For example HIV-RNA quantity, directly associated to the speed of virus replication, give the opportunity to control the effect of the therapy. One of these tests the Q-PCR (Quantitative Polymerase Chain Reaction)

3.2. Treatment for AIDS

The therapy most effective in the treatment of AIDS is the highly active antiretroviral therapy (HAART). This kind of therapy for the treatment of infection by HIV is quite similar to chemotherapy against cancer. The efficiency is evaluated by the increasing of the number of CD4 cells and by decreasing of blood's levels of viral RNA. HAART is concerned to use at the same time of one or more nucleosidic reverse transcriptase inhibitors, non-nucleosidic inhibitors and protease inhibitors.

With the use of HAART the number of CD4 cells usually increases, and there is a partial re-building of immunitary system. In some cases there is no increasing of CD4 cells although we can register a suppression of viral replication. In some people, functional improvement is associated to a worsening of opportunistic infections due to an increasing of inflammatory response to a non physiological level.

The HAART therapy is recommended to all people who are in an advanced stage of infection by HIV. Problems involved with this treatment are about drugs interaction with the agents involved in the cure of opportunistic infections, and also problems concerning toxicity and tolerability[72].

Compounds that inhibiting viral replication can be divided in some classes:

a) viral binding inhibitors

b) virus-cell fusion inhibitors

c) reverse transcriptase inhibitors

d) protease inhibitors

To these compounds we can add a possible vaccine that is recently studied as a possible therapeutic way.

3.3 Viral binding inhibitors

3.3.1 Soluble derivatives of CD4.

The assumption that the viral glycoprotein gp120 binds to the receptor CD4 protein who we can find on the surface of various cells such as T-Helper lymphocytes has suggested the production of a soluble derivatives who can get competition whit the cell's CD4 protein.

The first problem of these derivatives is the short half-life which has been successfully resolved by the linkage of an immunoglobulinic portion and that has demonstrated that these compounds are able to inhibit macrophage infection.

3.3.2 Small anionic molecules

Sodic suramine (*chart 1*) an anionic compound is being used for a long time in the treatment of trypanosomiasis and as a potent inhibitor of revers transcriptase of RNA carcinogenic virus has now shown active against HIV-1 but it has failed in clinical trials. Therefore, there are being considered some of its derivatives belonging to the class of azoic colorants like the Blue Evans (*chart 1*).

Their character of colorants and the fact that in vivo they were reduced to carcinogenic amines has excluded them from the use in mono therapy.

By the substitution of the diazo- groups whit other moiety who doesn't create carcinogenic groups and by the elimination of electronical conjugations (who assign the colorant propriety) there are being synthesized compounds effective in concentration relatively not toxic for the host cell. A potent example of this class is the napthalensulfonic derivatives (*chart 1*) although these modifications have conducted to a decreasing in activity.

3.3.3 Polimerics anionics molecules

The dextrans sulphate (*chart 1*) polisaccharidics molecules esteriphicated whit sulphuric acid, cause their in vivo instability and their low oral absorption there were left out. Some new compounds have been synthesized and those, to prevent the depolimerization, present a hydrocarbon skeleton and to prevent the desolphatation, they present sulfonic acid residues

instead of sulphuric esters. An example is poly(vinilsulphonic) acid (PVS) (*chart 1*) who interacts whit the protein gp120.
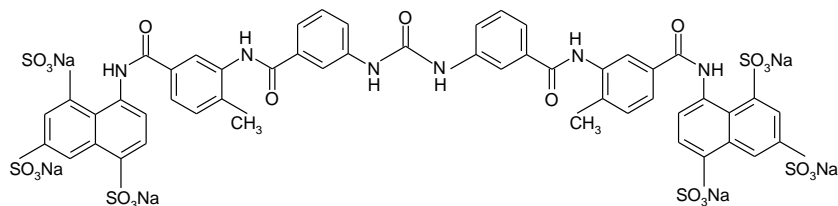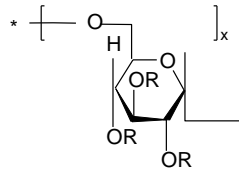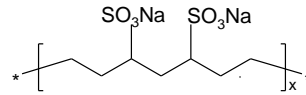

Sodic Suramin
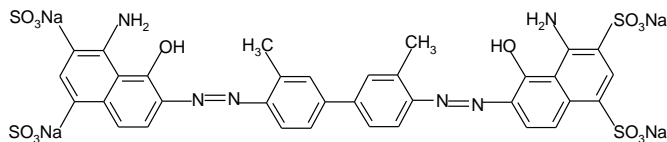
Blue Evans

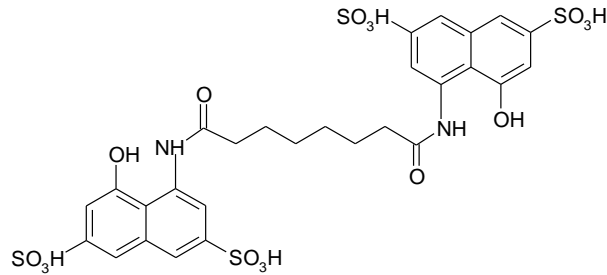**Chart 1**: Viral binding inhibitors (to be continued)

R=SO$_3$H

Dextran Sulphate

PVS

simmetric bis-naphtalenesulphonic derivative
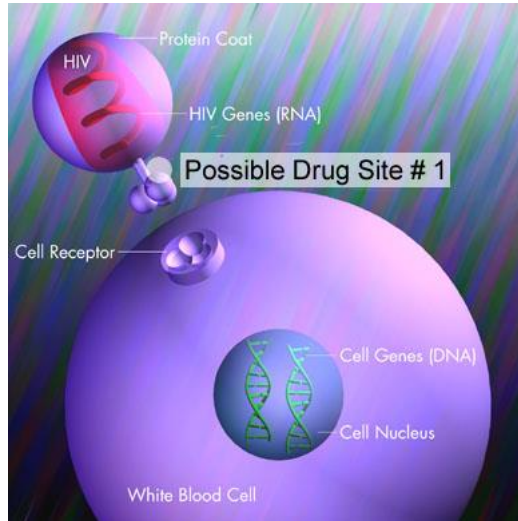

Chart 1: Viral binding inhibitors

Figure 8*:* Possibile drug site of action

3.4 Fusion inhibitors

The fusion cell-virus depends from the interactions between viral glycoproteins gp120 and gp41 and CD4 receptor, a protein that is localized on the surface of various cells. (P.M. 55 Kd).

The fusion real mechanism is not been exactly comprised but is widely studied but the possibility of inhibiting this bound by molecules such as lectyne, albumine, and triterpen derivatives is very interesting.

A new drug belonging to this class of antiretroviral is the enfuvirtide, synthetic peptide formed by 36 amino acidic residues in which the C-terminal portion interacts whit the viral membrane glycoprotein gp41 blocking the conformational changes very important for the interaction virus-cell. The drug is administered by under skin injection. (*chart 2*)[73].
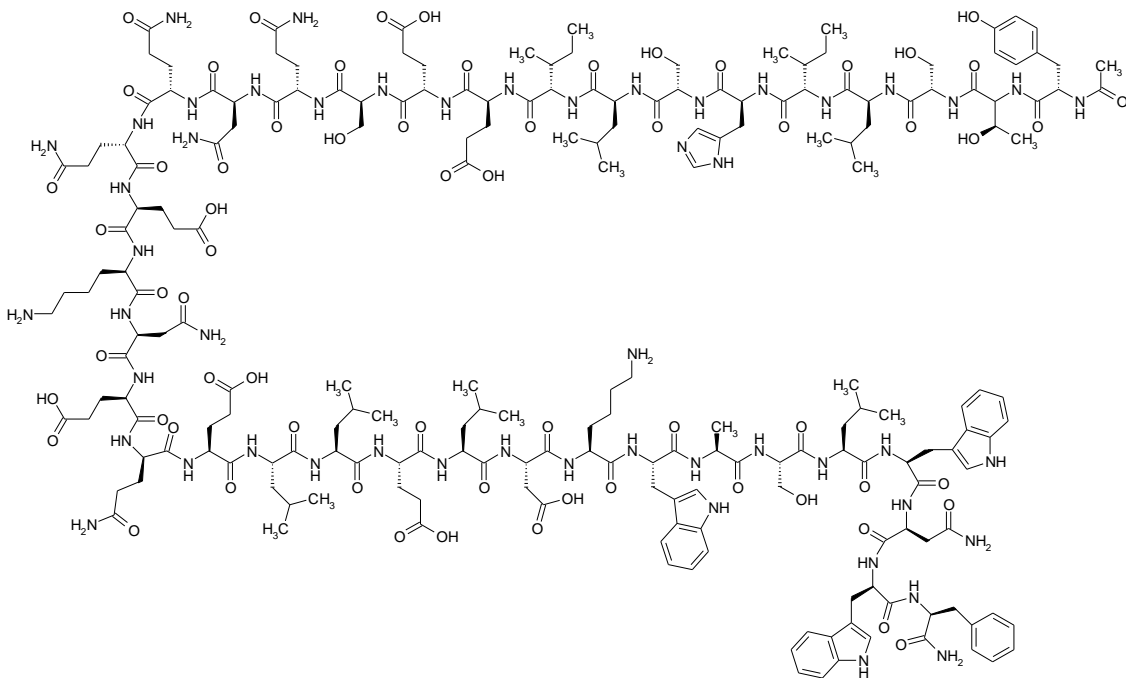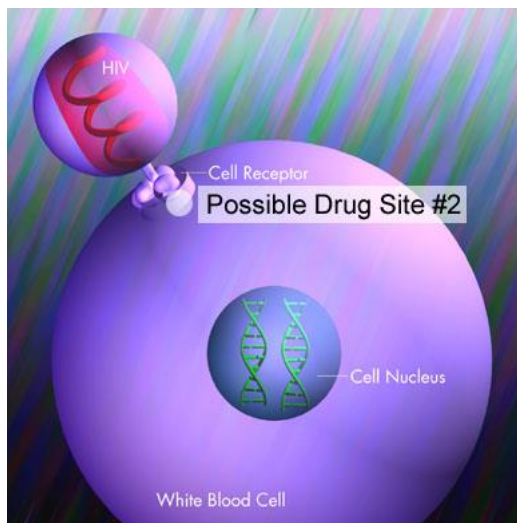
Chart 2: enfuvirtide

Figure 9: Possible site of interaction virus-cell

3.5 ReverseTranscriptase Inhibitors

The most important molecules against HIV are these who inhibiting the viral reverse transcriptase (RT), a DNA polymerase RNA dependent very important for virus replication.

3'-azido-2',3'-Didesossitimidine (zidovudine or AZT)[74] *(chart 3)* who is known active against a murine retrovirus, has shown his activity against HIV-1 *in vitro* . after clinical experimentation it becomes the first drug approved by FDA.

The screening among nucleosidic analogue compounds lead to the discovery of a new series of derivatives: 2',3'-dideoxycitidine (zalcitabine, ddC),[75] 2',3'-dideoxyinosine

(didanosine, ddI),[75] 3'-deoxy-2',3'-didehydrothymidine (stavudine, d4T),[76] (-)(1*S*,4*R*)-4-(2-amino-6-(cyclopropyamino)9*H*-purine-9-yl)-2-cyclopentene-1-methanol succinate (abacavir, ABC)[77] (-)-β-2',3'-dideoxy-3'-thiacytidine (lamuvidine, 3TC).[78]

These molecules were phosphorilated in position 5' of sugar ring from selective kinases becoming the corresponding triphosphate activated form. Once incorporated into the borning nucleic acid, they can't interact by forming a phospodiesteric bound because they are lacking of a hydroxile in position 3': in such way they act as a chain terminator.

Considering the highly efficiency demonstrate by these compounds, all the efforts are concentrated in the research of acyclic analogues derivatives by phosphonilmethoxialkilic of purine and pyrimidines, a class of a wide spread antiviral compounds, who acts as nucleotidic inhibitors of RT, which are true drugs because they don't need bioactivation by phosphorylation such as 9-[(*R*)-2-[[bis[[isopropoxycarbonyl)oxa]fosphynyl]methoxy] propyl]adenine, fumarate (1:1) (tenofovir disoproxil, PMPA)

Among these adeninics derivatives PMEA and PMEDAP are the most active more than AZT in a large number of animal models, present an immunomodulatory action and they also inhibit other viruses who can infect patients who affected by AIDS. The research extended to more than 600 class of drugs has lead to the discover of derivatives thetraimidazo[4,5,1,jk][1,4]benzodiazein-2(1H)-onici e –tionici (TIBO) which present a potency quite similar to AZT but better selectivity.

Among benzodiazepinic derivatives, very important is Nevirapine[79], Delavirdine[80] and Efavirenz[81] (*figure 10*)[82]. These molecules block RT act as a non-competitive inhibitor on an allosteric site of the enzyme, the *Non-Nucleosidic Binding Site* (NNBS).

Because of their activity is strictly connected to the enzyme structure, gene mutations induce resistance against these classes of compounds; but at the same time the selectivity against HIV-1's RT is very important.
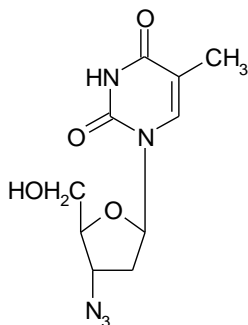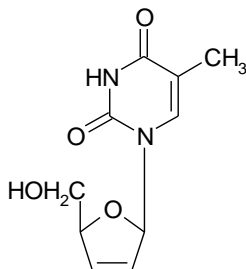
Zidovudine	Stavudine	Lamividine
Zalcitabine	Didanosin	Abacavir
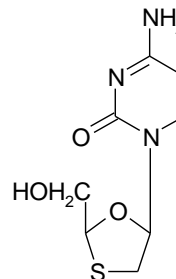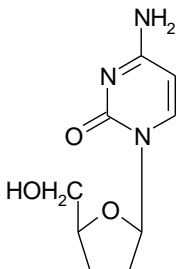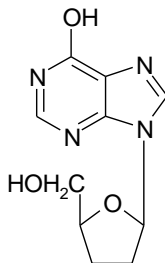
Chart 3: Nucleosidic inhibitors of RT (To be continued)

Tenofovir
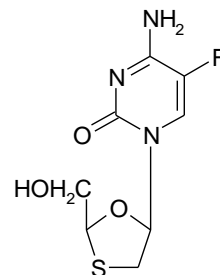
Emcitabine

Chart 3: Nucleosidic inhibitors of RT
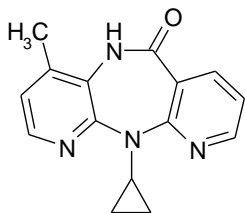
Nevirapine
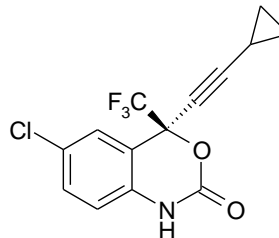
Efavirenz

Chart 4*:* non-nucleosidic inhibitors of RT (To be continued)
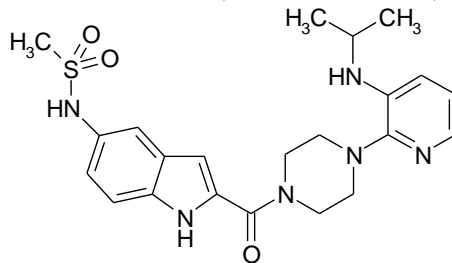
Delarvidine

Chart 4*: non-nucleosidic inhibitors of RT

Table 1. Anti-HIV Food & Drug Administration (FDA) approved drugs.

Nucleosidic and Nucleotidic inhibitors of RT (NRTIs, NtRTI)

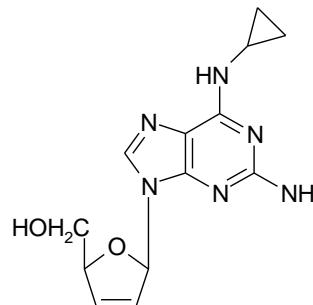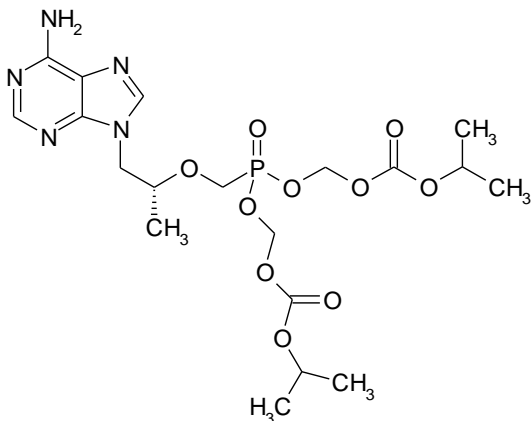| Year	Name	Commercial Name	Producer
1987	Zidovudine	Retrovir®	GlaxoSmithKline
1991	Didanosine	Videx® Videx EC®	Bristol - Myers Squibb Co
1992	Zalcitabine	Hivid®	Roche Laboratories
1994	Stavudine	Zerit®	Bristol - Myers Squibb Co
1995	Lamivudine	Epivir®	GlaxoSmithKline
1997	Lamivudine/Zidovudine	Combivir®	GlaxoSmithKline
1998	Abacavir	Ziagen®	GlaxoSmithKline
2000	Abacavir /Lamivudine/Zidovudine	Trizivir®	GlaxoSmithKline
2001	Tenofovir	Viread®	Gilead Sciences Inc
2003	Emcitrabine	Emtriva™	Gilead Sciences Inc
2004	Abacavir/Lamivudine	Epzicom®	GlaxoSmithKline

Nucleosidic and Nucleotidic inhibitors of RT (NRTIs, NtRTI)

Year	Name	Commercial Name	Producer
2004	Emcitrabine/Tenofovir	Truvada®	Gilead Sciences Inc

non nucleosidic RT inhibitors (NNRTIs)

Year	Name	Commercial Name	Producer
1996	Nevirapine	Viramune®	Boehringer Ingelheim Pharmaceuticals Inc
1997	Delavirdine	Rescriptor®	Pfizer Inc
1998	Efavirenz	Sustiva®	Bristol - Myers Squibb Co

Protease inhibitors (PIs)

Year	Name	Commercial Name	Producer
1995	Saquinavir	Invirase®	Roche Laboratories
1996	Ritonavir	Norvir®	Abbott Laboratories
1996	Indinavir	Crixivan®	Merck & Company, Inc
1997	Nelfinavir	Viracept®	Agouron Pharmaceuticals Inc
1997	Saquinavir	Fortovase®	Roche Laboratories
1999	Amprenavir	Agenerase®	Vertex Pharmaceuticals Inc
2000	Lopinavir/Ritonavir	Kaletra®	Abbott Laboratories
2003	Atazanavir	Reyataz™	Bristol - Myers Squibb Co
2003	Fosamprenavir	Lexiva™	GlaxoSmithKline

Fusion inhibitors (FIs)

Year	Name	Commercial Name	Producer
2003	Enfuvirtide	Fuzeon™	Roche Laboratories
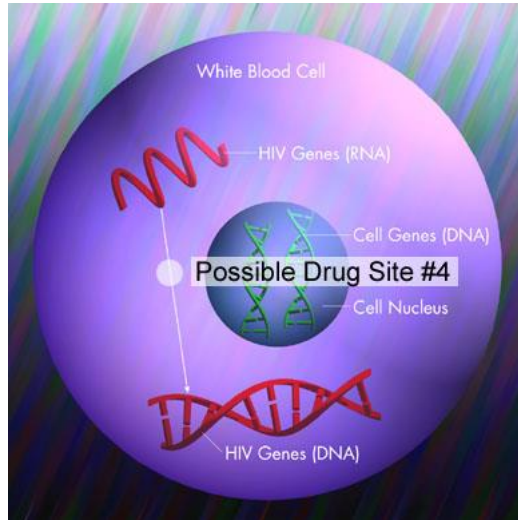
Figure 10: Possibile drug site of action of RT's inhibitors

3.6 Protease inhibitors

The HIV-1 protease belongs to the family of aspartilprotease and is an enzyme that plays a relevant role in a post integrational phase *(figure 11)*, before and during viral gemmation. At this level the retroviral protease cuts the poliprotein gp160, viral precursor, in proteins that can creates the mature virion there are more evidences that the protease can attack some host's proteins.The active site of this enzyme is formed by the three-peptide aspartic acid-threonine- glycine that is a typical sequence of the central core of some aspartic proteases. There were prepared and tested many compounds which structure was miming the sequence of three-peptide asparagines-phenilalanine-proline. Compounds such as Ampenavir[84],

Atazanavir[85], Fosamprenavir[86], Indinavir[86], Lopinavir[87], Nelfinavir[88], Ritonavir[89], and Saquinavir[90] *(chart 5)* show a good anti-protease activity, but problems involved in ADME limit their use.
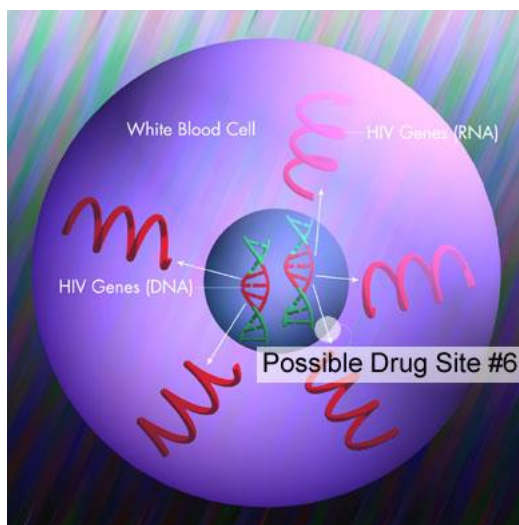
Figure 11: Possibile drug site of action of Protease's inhibitors

Amprenavir

Atazanavir

Chart 5: Protease inhibitors
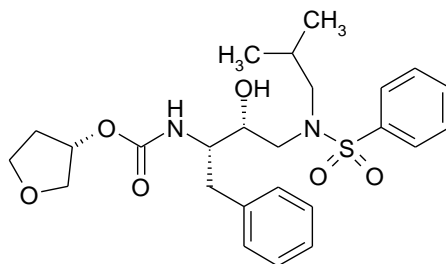
Fosamprenavir

Indinavir
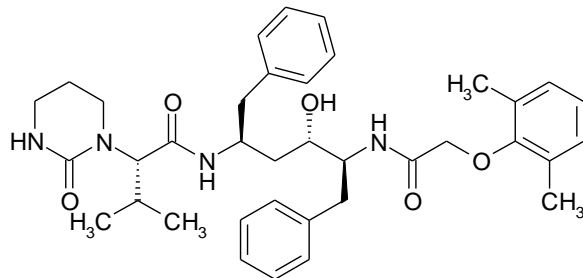
Lopinavir

Chart 5: Protease inhibitors
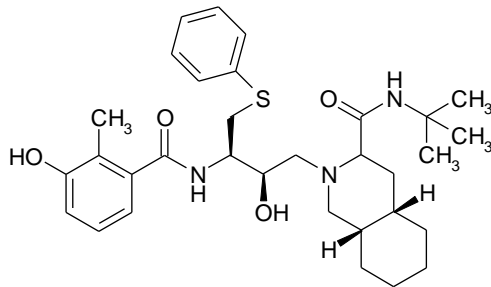
Nelfinavir
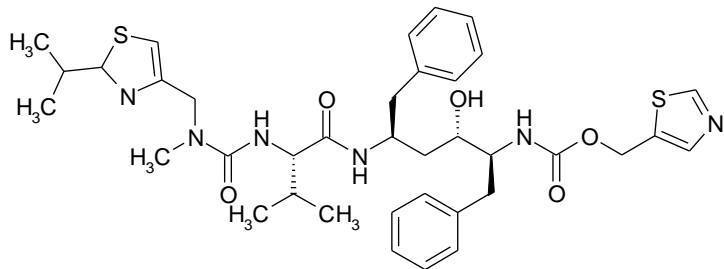
Ritonavir
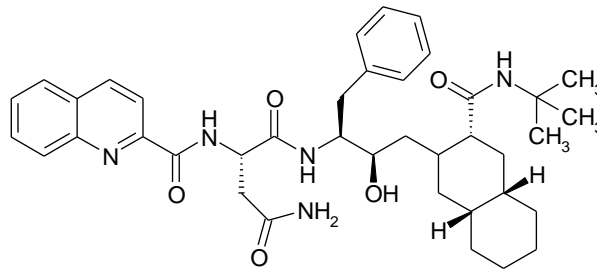
Saquinavir

Chart 5: Protease inhibitors

3.7 Anti-HIV-1 Vaccine

The needle to develop a vaccine for therapeutic scope or preventive is now very evident.

An efficient vaccine that can limit or prevent the diffusion of the infection is very far to be discovered because of many factors.

Virus shows *in vivo* a great variability, in fact there were isolated more than 100 variants of HIV-1. Mutations in one only amino acid in the variable region of envelope give to the virus the capability to survive into the body and to avoid the immunitary system defences.

The phenomenon of *enhancing antibodies,* instead to play a neutralizing effect on virus, play a role in the entrance in macrophages cells by the Fc fragment of Ig. Particularly the vaccine's studies are point out to obtain a recombinant protein subunit. Because of their highly dangerousness vaccine based on inactive viruses are not used.

Actually, research is concentrate in the variable region of gp120, but very important are the results by the PCR studies. This method has lead to identify a high conserved sequence of gp120 against which was obtained a higher antibodies response.

There are so synthesized some peptides with the same sequence with the aim to obtain a T Helper reply with production of IL-2.

4. 3-D QSAR: an introduction

In Drug Discovery, synthesis, biological evaluation and screening of hundred and thousand of potential drug candidates is an extremely expensive and laborious procedure, more efficient strategy are necessaries. A rational mode is the derivation of structure-activity hypothesis and their quantitative evaluation without biological assay or laboratory synthesis.

However, in doing so, it is important to better understand the aim and the limitations of the applied approach in order to use it in appropriate and correct manner.

The important computers contribution to science is their capability to quantify and evaluate hypotheses much faster.

With the availability of powerful workstation, molecular modeling and drugs three dimensional quantitative activity relationships (3-D QSAR) their interaction complexes became increasingly important.

Major contribution to 3D QSAR method development were Hansch-Dieter Höltie's calculation of interaction energies in hypothetical drug receptor complexes[92], Garland Marshall's active analogues approach[92], Peter Goodford's program GRID[93] and Richard Cramer's comparative molecule fields analysis (CoMFA)[94]. In parallel, powerful statistical tools as Partial Least Square (PLS)[95] analysis and Gabriele Cruciani's software GOLPE[96], had to be developed to correlate a large number of data points with the biological activity values of a limited number of objects.

4.1 Partial Least squares

Partial least squares (PLS) is a general a powerful tool for deriving multi linear relationships among columns of data. Compared to classical approaches to the same problem, notably multiple regression, PLS has been show to be the much less susceptible to change correlation, that is, reporting as scientific fact random pattern among data values that are unrelated. PLS may also be applied to a much wider range of problems, from simple traditional least-squares Y versus X line fitting to the COMFA approach of fitting thousand of columns descriptor of molecular characteristics to same biological property. PLS result are evaluated for their ability to predict new information, using cross validation Most often PLS is performed in two stages: the first with cross validation to determine how rich or complex a model is appropriate for the data values (how many components to use), the second without cross validation and the optimum number of components establish the single model which best represent the data.

Nowadays PLS represent a new mathematical approach to analyze a great number of descriptors and to create a correlation between them and biological activity (as in my model) or a different characteristic. For example this type of analysis was use in cosmetic or alimentary industry to obtain an equation that represents a particular product (wines, cheeses and creams).

PLS can be interpreted in many ways: conceptually, mathematically etc. In this contest the most important thing is that PLS finds combination of the structural descriptors that a) well

predict biological activity and b) well model the structure descriptor matrix.

Mathematically summarize the X-variables in a few orthogonal new variables called scores, or latent variable (LVs). The score vectors are collected in a matrix T (N= number of object) (N x A).This summary is achieved by means of a projection of X on an A-dimensional plane. The coordinates of each object into this plane are the score.

If there is more than one Y-variable, PLS simultaneously summarize the Y-data model. PLS proceeds by calculating one component (model dimension) at a time and one-stop when cross-validation becomes significant. The score vector (T) are then used as new predictor variables to model Y. Since the score are few and orthogonal, and for this reasons independent there aren't problem of spurious correlations. In practice PLS give us a function without bias correlation between Y and T **Y= f(T)**

4.2 The history and use of Quantitative Structure Activity Relationships (QSAR).

More than century years ago, in 1893 Charles Richet published a note on the relationship between the toxicity and the physicochemical properties of a series of organic compounds. From this analysis Richet concluded that: more soluble was less toxic[97]. In the same year Emil Fisher investigate the influence of α- and β- configurations of different glycosides on their enzymatic cleavage formulating that enzyme and glycoside must fit together as lock and key, in order to obtain different effect on each one[98]. Only fifty years later Paul Ehrlich' investigation on anti-bacterial dyestuffs led to the concept of receptor.[99.]These fundamental

theory stimulating chemical researcher and more attention was attributed to the effect of pharmacophoric groups on selectivity and activity. The situation changed in 1964 when Hansch and Fujuta[100]) published their studies on quantitative relationship between physicochemical properties and biological activities and Free and Wilson developed a model of additive group contribution to biological activity values[101] However the three-dimensional structures of molecules not be considered in these approaches.

Hansch analysis is a relatively easy method. It assumed that all series compounds interact with their biological target, be it an enzyme, a nucleic acid or a channel protein, in the same manner and with the identical conformation. The situation is more complex in 3-D QSAR analysis where the series of compounds are often dissimilar, alignment can introduces bias and structure of biological receptor are not available. Therefore Hansch analysis guided to the development of 3-D QSAR, prior to 1979 no general concept was developed how to implement 3-D features in a QSAR analysis. In 1988 R. Cramer, introduced 3-D QSAR involving the analysis of the quantitative relationship between the biological activity of a set of compounds and their three-dimensional properties using statistical correlation methods[102]

From the very first formulation of a lattice model to compare molecules by alignment them in 3D space and by mapping their molecular fields a 3-D grid, it took nearly a decade until CoMFA[103] and commercial software [104-105] were developed.

The success of a QSAR development is based on three fundaments: the biological data, the selection of compounds and the selection of structural descriptors. The selection and calculation of structural descriptors plays a crucial role in the ability of resulting model. Any mathematical modeling needs homogeneity of the variables, each X and Y variable, must

mean the same thing for all objects (molecules). This necessitates some kind of alignment of the compounds. Also important is to note that is not necessary to map all molecule structure but only changing moiety. A constant backbone or large fragment does not contribute any information and hence no parameterizations. The selection of compounds for the learning set of the QSAR is equally important. It must be remember, however, that an unbalanced and unrepresentative selection of compounds will always yield a low quality QSAR, difficult to interpret and perhaps misleading prediction.

4.2.1 Design

When we developing a QSAR model we are investigating direction in an abstract multidimensional *structure space*, and how activity changes when we move in these different direction. This structure space has one axis for each possible structural factor that can be changed in our investigated molecules. The variation along each axis in bounded by a lowest and highest possible value: in a peptide, for instance, we cannot have amino acid smaller than glycine or larger than tryptophan. To be able to investigate well all directions in the space, we need a set of compounds that well spans this space.

4.2.2 Missing Data

Although data analysis works better when the data are complete, PLS modeling tolerate

moderate amounts of missing data both in X and in Y, provided that the missing elements are fairly randomly spread out over all objects and variables. A reasonable percentage is 10% maximum missing data in X and Y. Naturally, Y must be multivariate for PLS to tolerate missing data in Y. Presently no working method exist for the proper analysis of such data. The better result could be obtained using a combined analysis of two partial data set; one with all molecules (object) and only screening Y-values, and a second with just the highly active compounds and all Y-variables. The latter may give information about the correlation pattern between the various tests that might be used to interpret the first "rich" model with respect to more interesting high-cost Y-variables.

4.2.3. Scaling

The result of PLS and PCA projection depend of the scaling of the data. Using an appropriate scaling of the data, one can focus the model on more important Y-variables, and use experience and fundamental knowledge to increase the weights of X containing much information and decrease the weights of other parts containing little information.

4.2.4. Data analysis and model interpretation

With PLS one needs to estimate also the model complexity: for example the number of significant PLS components (model dimension A). Particularly with the large number of X-

variables, it is essential to strictly test the meaning of each models dimension to obtain that the final model include only significant components. This is fundamental to limit overfit and to give validity to the model. Cross-Validation provides a very reliable way for testing this significance.

With CV one estimates the predictive ability of the model by:

1-   calculating an ensemble of modes from reduced data where parts of the original data have been deleted, and then

2-   2- using each model to predict the data the where deleted from its modeling calculation.

These deletion and calculation are made in such a way that each object is deleted once and once only over the model ensemble. One then computes the sum of all objects of squared differences between predicted (by the model without the deleted object) and experimental Y-value. This sum is called r^2. Moreover the CV result is expressed also by SDEP.

Sometimes CV of PLS indicates that the first or second dimension is insignificant while there are one or several significant model dimensions beyond these. Hence, if only a small part of Y has been explained, one should look also at the model dimension after the apparently insignificant one, and continue the modeling if this is indicated to be significant by CV. This is completely safe as long as the number of model dimension.

Conceptually, we can see the scores as estimates of same molecules characteristics.

Hence one should make efforts to understand and interpret these scores, witch is best done by considering plots and their patterns. Each score vector is a linear combination of the X-variables with a coefficient w_a. Similarly, the Y-score is linear combinations of the Y-

variables with another coefficient t_a. Hence the size of element of w_a informs about the importance of that variable. For example is possible to find out that a second score vector is composed mainly of hydrophilic descriptors, and also witch part of molecules contribute much. With the 3 D-QSAR color coded plots of coefficients are informative in this respect, making interpretation easier.

Because PLS is a projection method outlier and other inhomogeneities are easily detected.

Once the 3-D QSAR model has been developed based on the training set of compounds, it can be used to predict the activity profile Y_{pred} of new, not yet synthesized molecules. This prediction is done by first translating the structure of a new compound into a vector of structure descriptor values. This vector is trans-formed and scaled in the same way of training set data. Thereafter the resulting vector is inserted into the QSAR model giving predicted values of Y-variable. Usually a test set is used to evaluate the prediction ability of the model and to evaluate the standard error of prediction. Also is important to evaluate the structurally similar among data set and prediction set. If this similar characteristic is greater then a value prediction is unreliable.

4.2.5. QSAR Plots

The QSAR-PLS model provides two kinds of plots: score plots showing the objects (here molecules), and loading/weight plots showing the variables of X and/or Y. Plotting the Y-scores versus the corresponding X-score shows the connection between structure (X) and

activity (Y).

The use of PLS for 3 D- QSAR model and data analysis is very advantageous.

4.3 Application of GOLPE

The use of GOLPE software allows the developing of 3-D QSAR model implementing PLS technique previously analyzed and to obtain plot representation of it. Obtained the 3D descriptor by the GRID the GOLPE procedure is given by the following steps:

- A variable selection phase aimed at deleting redundancy

- Evaluations of the effects of each individual variable on the predictive power of the model, obtained by the comparison with the apparent effect of dummy variables on a representative number of variable combinations

- Evaluation of the predictive power of the supposed optimal equation.

Hence, the reference method uses a PLS implementation in which the validation is made in leave-one-out and which does not include any variable selection. Consequently, it seems appropriate to consider GOLPE a culturally advance tool which gives better reliability in terms of prediction.

The predictive power of the model can be evaluated by the SDEP parameter computed by GOLPE both in term of self-consistency or on a validation set. Although this latter approach gives results dependent upon the selection of the test set, it seems, whenever possible, more reliable. Nevertheless, it is very fast in computation.

GOLPE seems to be a powerful tool to improve the quality of the result of the regression step (PLS) of any 3 D-QSAR models.

Data pretreatment is very important in effect different data pretreatment generate different PLS models and consequently, different results, both in terms of predictive power and in term of interpretation. An appropriate data pretreatment should take into account, the fact that interactions relevant to the ligand receptor binding can occur only in the area of the molecule surface, but, on the other hand, all grid point should in principle be given the same opportunity of showing up as important. Another critical point is the cut-off value to be used on fields or energies. In order to avoid chemometric model depending too heavily on points of high leverage it is appropriate that these range are roughly centered.

A series of crystallized and modeled data set were used and studied to establish if is more useful or not to autoscaling variables scores and after or before data agreement between predicted and experimental region of interaction is obtained on preselecting variables and autoscaling the remaining.

GOLPE plots makes models interpretation easer but is important to take in to account that when transforming the PLS loading into the pseudo coefficient of the polynomial in the original variables, it is not appropriate to give too much credit to the individual coefficient: they suffer of the same instability of Multi Linear Regression, although prediction are stable since they are derived by PLS.

5. Design of Indolyl Aryl Sulphones (IASs)

5.1 Towards the design of new Indolyl Aryl Sulphones

While searching for new anti-AIDS agents, we focused our attention on Nitrophenyl Phenyl Sulphones (NPPS, compound **1**)[106], a new class of NNRTIs discovered at the National Cancer Institute (NCI) during a large-scale drug-screening program. The most active compound, the dinitro derivative **2** (EC_{50} 0.89 µM), was discarded due to its poor bioavailability, while the (4-methoxy-2- nitrophenyl)phenylsulphone **3** (EC_{50} 3.4 µM) was selected for further studies. In our laboratories we selected [107] NPPS **1** (*Chart 6*) as lead compound for the design of new isosteres characterized by the presence of a pyrrole nucleus. The new Pyrryl Aryl Sulphones (PASs) were less potent than NPPS, although they were also less cytotoxic (PAS **4**, EC_{50} 15.08 µM, CC_{50} = >308 µM). Extensive SARs studies led to disclose new PAS derivatives active at submicromolar concentration characterized by the *para*-chloroaniline moiety as a featuring pharmacophoric group [108-109] (PAS **5**, EC_{50} 0.18 µM).
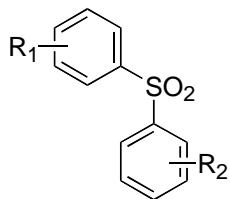
NPPSs

1: R_1 = 2-NO_2, R_2 = H
2: R_1 = R_2 = 2-NO_2
3: R_1 = 2-NO_2,4-OCH_3, R_2 = H
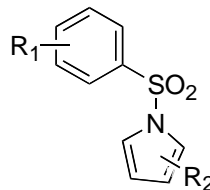
PASs

4: R_1 = 2-NO_2, R_2 = 2-COOEt
5: R_1 = 2-NH_2,5-Cl, R_2 = 2-COOEt

Chart 6

Considering the tricyclic structure of known NNRTIs suh as NVP and TIBO [110-111] **(6)**[112-113], the PAS derivatives were intramoleculary cyclized to obtain new pyrrolo[1,2 *b*][1,2,5]benzothiadiazepine (PBTDs), and some related pyrrolo[2,3-*b*][1,5]benzothiazepine and pyrrolo[3,2-*b*][1,5]benzothiazepine derivatives (general structures **7-9**) [114-115] (*Chart 7*). PBTDs derivatives showed submicromolar anti-HIV activity, comparable to that of NVP (**10b**: EC_{50} = 0.5 μM; CC_{50} = > ≥ 300 μM).

6 (TIBO)
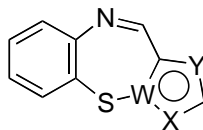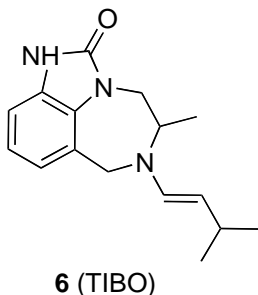
7: (X=N, Y=W=C)
8: (X=W =C, Y=N)
9: (X=Y=C, W=N)
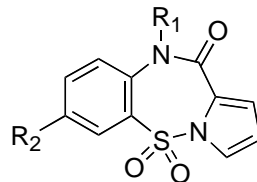
10a: R₁ = R₂ = H
10b: R₁ = H, R₂ = Cl
10c: R₁ = CH₃, R₂ = Cl

Chart 7

In 1993 Merck Research Laboratories reported the anti-HIV-1 activity of the potent RT inhibitor L-737,126 (**11**, EC_{50} 0.001 μM, CC_{50} = 200 μM)[108] (*Chart 8*). Similarly to the PAS ➔ PBTD transformation, Silvestri *et al.* [155] synthesized two series of **11**'s analogues. Representative examples are the conformationally constrained 2-chloro-5*H*-indolo[3,2-*b*][1,5]benzothiazepi-6(7*H*)-one 12,12-dioxide **(12)** and the acyclic variant 3-[(2-amino-5-chlorophenyl)sulphonyl]-5-chloroindolo-2-carboxamide **(13)** (*Chart 3*)[155]. Unfortunately both **12** and **13** related series were less active than reference compound **11**. These results pulsed us to do molecular modelling studies to rationalize the experimental data. Docking studies [155] of **11**, **12** and **13** revealed different binding mode for each molecule: the compounds docked into the NNBS of RT showed dissimilar orientations of the bound conformations (*Figure 12*).
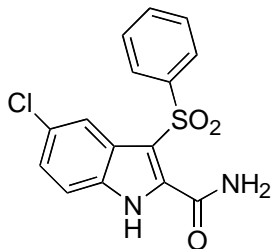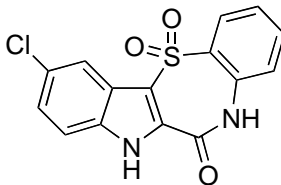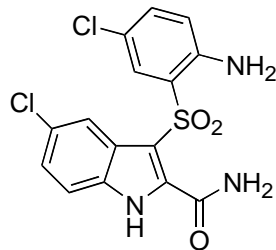
11 (L-737,126)   **12**   **13**

Chart 8.

As reported by Silvestri *et al*. [116]the different binding modes of compounds **12** and **13** would account for non-parallel SARs of PBTDs and their related PASs derivatives. Using the three-dimensional structure of RT complexed with Cl_2-α-APA **(14)** [117](*Chart 9*) and **5**'s X-ray structure (*Figure 12A*), a model of RT/**5** complex was derived [118-119] (Figure 12B). This model was used to design new PASs derivatives and some related analogues (*Chart 9*).
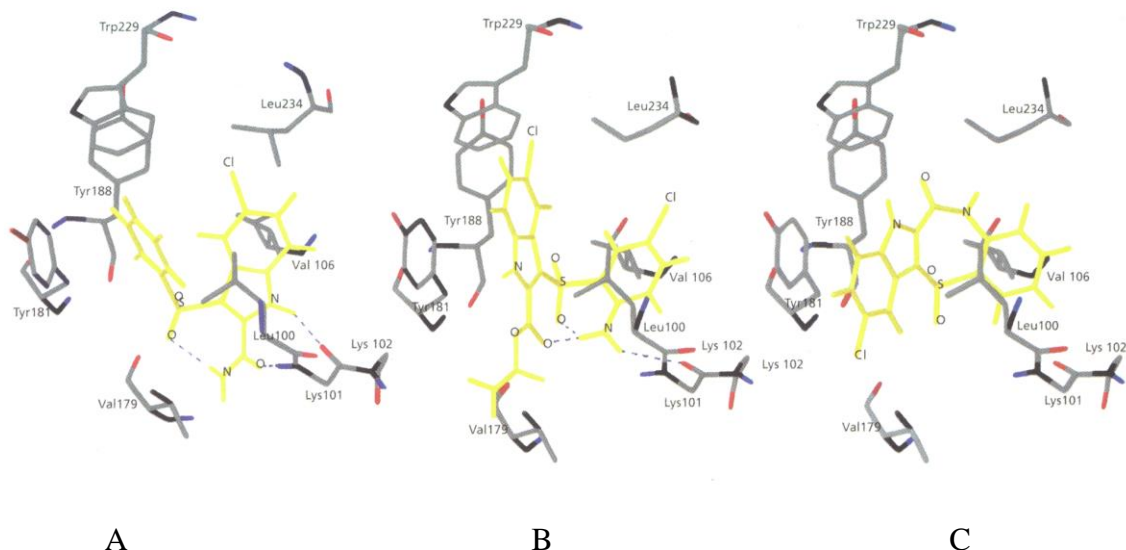
Figure 12. Molecular models of L-737,126 (A), **13** (B) and **12** (C) docked into the non-nucleoside binding site of RT. Hydrogen bonds are highlighted by dashed lines.

Among these compounds, derivative **15** *(Chart 9)* was found the most active (EC_{50} 0.045 μM, IC_{50} 0.05 μM and CC_{50} 240 μM). Compared with those of lead compound **5**, these values represented significant improvements in the cell-based and enzyme assays. Molecular docking studies led to design new derivatives **16**, **17** and **18** (*Chart 9*) based on the hypothesis that additional hydrogen bond(s) would favour the binding of the inhibitors into the NNBS (*Figure 13*). Although compounds **16-18** were found inactive in cell-based assay, they inhibited the RT enzyme at micromolar concentrations (**16**: IC_{50} 2 μM; **17**: IC_{50} 3 μM; **18**: IC_{50} = 9 μM).
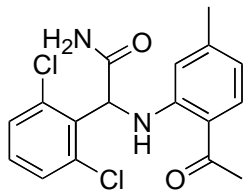
14 (α-APA)
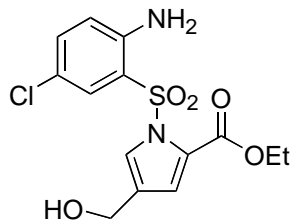
15
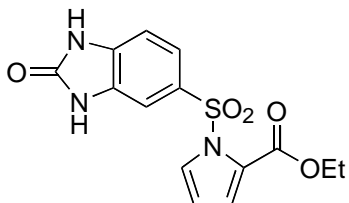
16
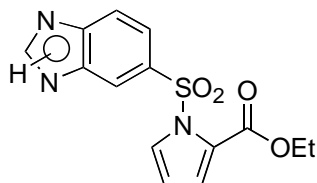
17
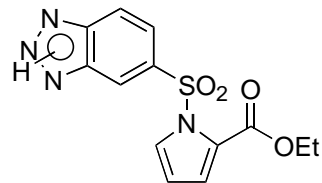
18

Chart 9.
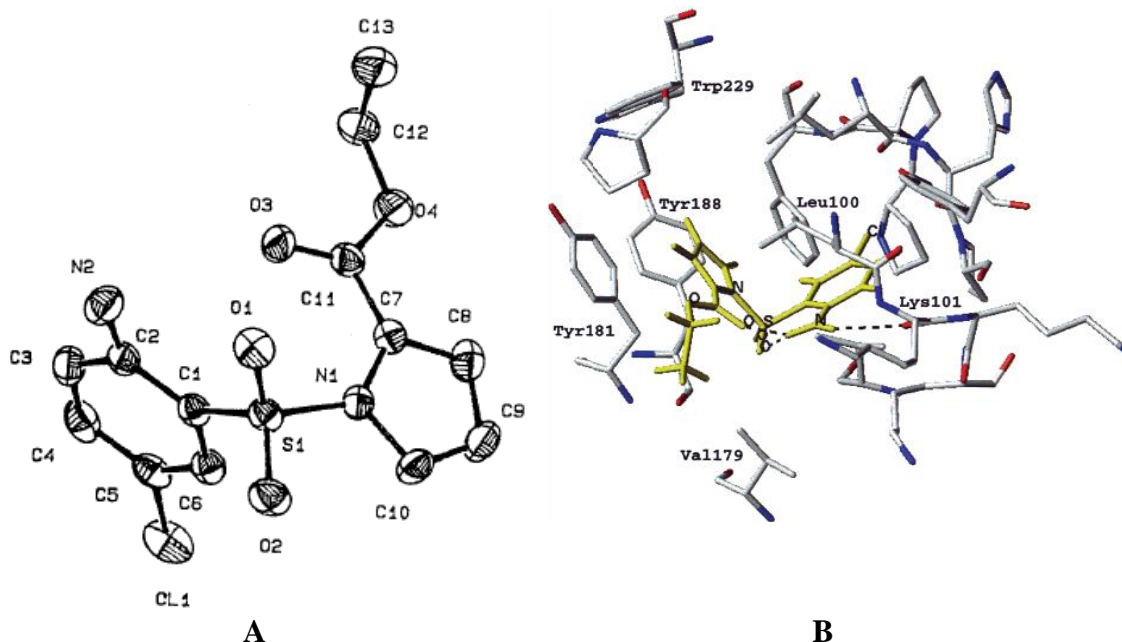
Figure 13. (A) Crystal structure of compound **5** showing the thermal ellipsoids at 40% probability and the labelling of non-H atoms. (B) Derivative **5** (yellow) docked into the RT NNBS. Only a subset of residues closest to the ligand is displayed for sake of clarity.

Docking studies of compound **11** revealed positive attractive π-π interactions between the phenyl ring and an aromatic rich pocket formed by Tyr188, Tyr188 and Trp229. These findings prompted us to design and synthesize 1- and 3-arylsulphonyl-5-chloroindoles (for example compounds **18a-d** and **19a-e**) (*Chart 10*) [108,120]

A

B
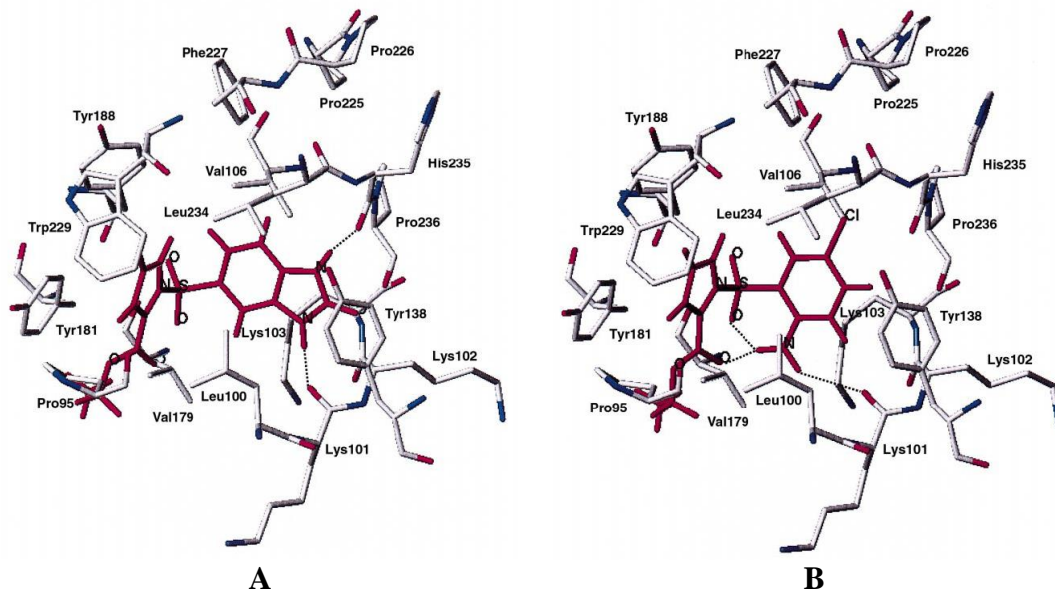
Figure 13 to be continued
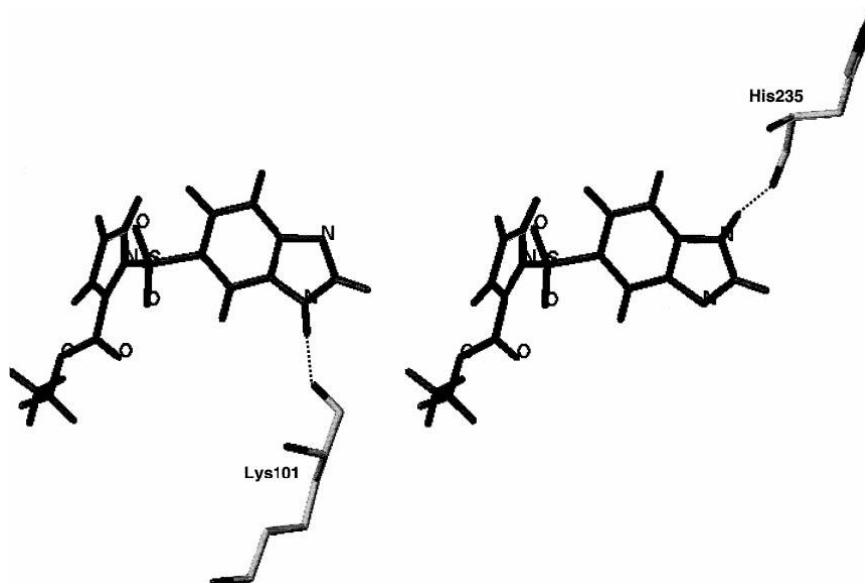
C

Figure 13. (A) Theoretical model of the RT/**16** complex. Two hydrogen bonds are hypothesized to occur between the ureidic NH groups of the docked molecule and the carbonyl oxygen's of Lys101 and His235. Only a subset of residues of the non-nucleoside binding site closest to the ligand is displayed. For comparison in (B) is reported the theoretical model of the RT/**5** maintaining the same orientation as in (A). In (C) is reported the theoretical model of the RT/**17** complex. Two tautomeric forms of the benzimidazole derivative **17** were considered differing in the position of the NH group. This latter donates a hydrogen bond to the carbonyl oxygen of Lys101 (tautomer on the left) or His235 (tautomer on the right). Only residues Lys101 and His 235 of the NNBS are displayed.
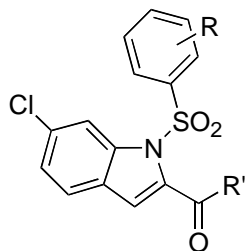
18a-d
18a: R = H, R' = OEt (EC$_{50}$ = >200 μM)
18b: R = 4-Me, R' = OEt, (EC$_{50}$ = >200 μM)
18c: R = 2-NH$_2$,5-Cl, R' = OEt (EC$_{50}$ = 8.3 μM)
18d: R = H, R' = NH$_2$ (EC$_{50}$ = 66 μM)
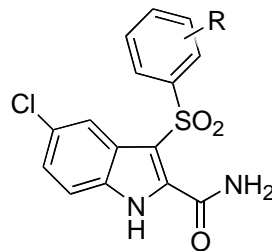
19a-e
19a: R = 2-Me (EC$_{50}$ = 0.001 μM)
19b: R = 3-Me (EC$_{50}$ = 0.001 μM)
19c: R = 4-Me (EC$_{50}$ = 0.003 μM)
19d: R = 2,4-Me$_2$ (EC$_{50}$ = 0.004 μM)
19e: R = 3,5-Me$_2$ (EC$_{50}$ = 0.004 μM)

Chart 10

5.2 IASs: the use of molecular modelling techniques to design new and improved anti-HIV derivatives

5.2.1 The first development

The indol-1-yl aryl sulphones were found poorly active or inactive against HIV-1. In contrast, the indol-3-yl aryl sulphone series showed highly activity both in cell-based and in enzyme assays.[110]. SAR analysis showed that both the carboxamide function and the methyl substituents on the 3-phenyl ring were crucial for the high activity and selectivity of derivatives **19a-e** (*Chart 10*). These compounds showed EC$_{50}$ values comparable to that of reference compound **11** against HIV-1 WT. Compounds **19a-e** were also active [110] against the most common and clinically relevant HIV-1 drug-resistant mutant strains at concentrations

comparable to that of EFV.

5.2.2 Combining SBDD (docking) and LBDD (3-D QSAR) Studies

To better understand the SAR of the synthesized IASs, extensive docking and 3-D QSAR studies were undertaken. The Autodock3 program [121] proved to be a valid tool in structure based drug design (SBDD) [122-123] that, in some extent in cross-docking experiments using 40 different experimentally determined co-crystallized RTs conformations[124] was also able to include the induced fit. Therefore Autodock3 was used for all IASs' docking studies. For the ligand-based drug design (LBDD) approach, the GRID/GOLPE [125] combination was adopted.

6. Molecular Modeling on Indolyl Aryl Sulphones (IASs)

6.1 Docking Studies

The first utilized approach has the aim to explain binding mode of most active derivative **11.** It was analyzed by means of a cross-docking procedure using fourteen experimentally determined 3-D structures of WT RTs (pdb codes 1dtq,[126] 1dtt,[126] 1eet,[127] 1fk9,[128] 1hni,[129] 1hnv,[130] 1jlq,[131] 1rt1,[132] 1rt2,[132] 1rt3,[133] 1rt4,[134] 1rt5,[135] 1rt7,[135] 1vrt,[136] and 1vru,[130])[137] In all cases the docked conformations were in good agreement with each other. Only the carboxamide function displayed some uncertainty because it was oriented to the same side of indole NH or rotated about 180° (*Figure 14, Chart 11*).

The binding conformation of compound **11** proposed by docking studies, was characterized by four different distinct interactions (*Figure 14*): (i) an hydrogen bond between the indole NH or the carboxyamide NH with the Lys101 carbonyl oxygen (blue residue in *Figure 15*); (ii) a π-π hydrophobic interaction between the phenyl ring and an hydrophobic aromatic-rich pocket formed mainly by the side chains of Tyr181, Tyr188, Phe227 and Tyr229 (red surface in *Figure 15*); (iii) a purely hydrophobic interaction between the sulphonyl group and α, β carbons of Val106, Val179 and the Lys103 side chains that describe a small pocket into the NNBS (yellow surface in *Figure 15*); (iv) the indole chlorine atom was found to make selective positive contacts with Pro236 .(magenta residue in *Figure 15*)
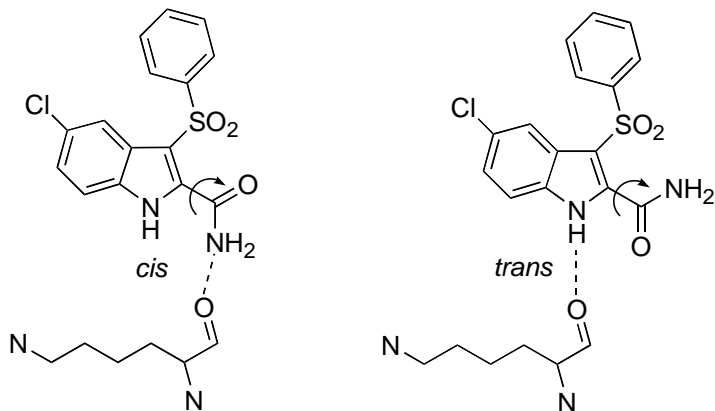
Chart 11. (A) The carboxamide function is in *cis* position with respect to the indole NH. The amide NH_2 forms a hydrogen bond with the carbonyl of Lys101. (B) The carboxamide is in *trans* position. A hydrogen bond is formed between the indole NH and the carbonyl of Lys101.
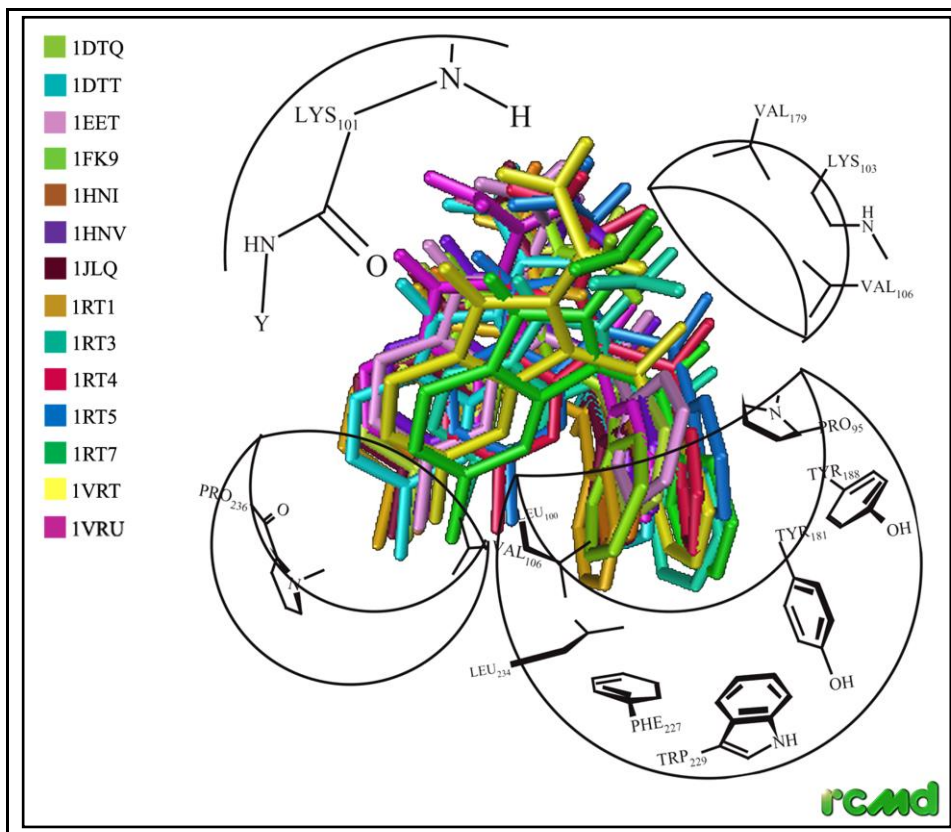
Figure 14 Graphical result of the dockings of reference compound **11** into fourteen NNBSs. Conformations docked in the different NNBS are colour coded; in the legend are reported the PDB entry codes of the used RTs. A sketch of the NNBS sub pockets is also reported
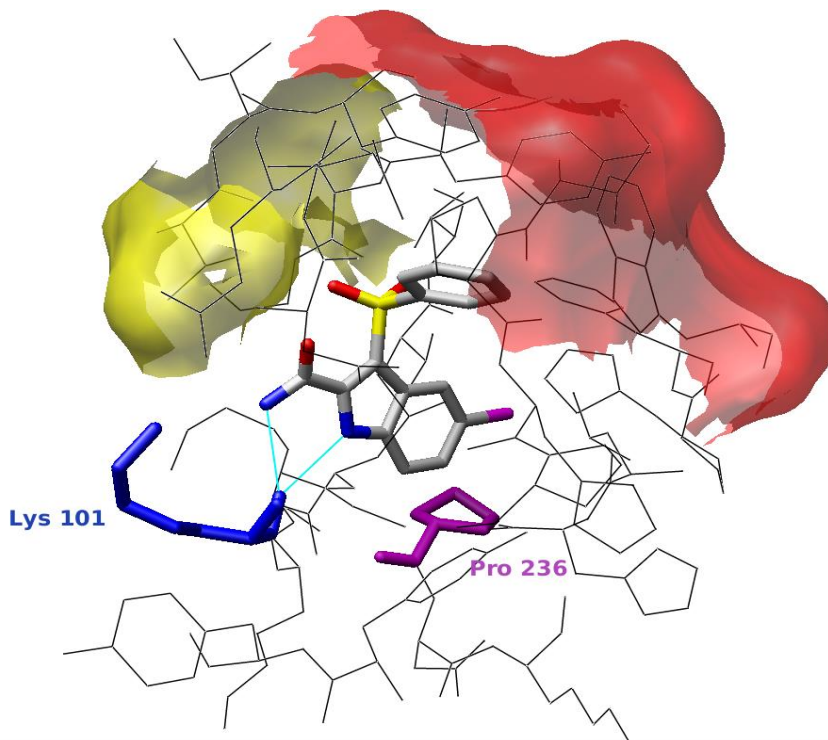
Figure 15. Binding mode of **11** (atom type colour). in red the hydrophobic pocket formed by Tyr181, Tyr188, Phe227 and Trp229 that accommodate the benzene moiety; in yellow the small hydrophobic cleft formed by Val106, Val179 and α, β carbons of Lys103 side chain that accommodate the sulphone group. In cyan are represented the H-bonds between **11** and Lys101 (blue). In magenta is depicted the Pro236 residue. 4Å RT NNBS are also reported (black line) for clarity.

To validate the docking procedure, parallel docking and cross-docking studies were conducted on all the complexes of the ligand co-crystallized within the NNBS of the 14 RTs. Autodock3 was able to reproduce with a minimal error the experimental binding mode of all

the ligands.[124]

6.2 3-D QSAR Studies

The above docking procedure was coupled with three-dimensional structure-activity relationship (3-D QSAR) techniques with the aim to develop useful descriptive models for the design of new IASs. The synthesis of new derivatives would prioritise activity predictions.

6.2.1 Definition of the first 3-D QSAR model

A training set of 70 molecules (*Table 2*), compiled using fifty indol-3-yl and twenty indol-1-yl aryl sulphones, was used to develop a predictive model [137]. Both active and inactive compounds were included to better analyze favourable and unfavourable ligand/receptor interactions. Moreover, the inactive derivatives proved to be necessary in order to code, in the model, those information (as grid variables) needed to discriminate active versus inactive compounds. Using the 3-D coordinates of the RT extracted from the complex with the IASs related di-aryl sulphone 739W94 [131] (*Figure 16*) the training set molecules were structure-based aligned by a docking procedure into NNBS. In several cases, different binding modes were found among the conformations suggested by the Autodock3 program. Since the molecular alignment of the training set is a crucial step in a 3-D QSAR, two different alignment rules were used: the lowest energy docked conformations of the most populated clusters ("best cluster alignment"), and the first ranked docked conformations as obtained from the Autodock3 run ("best docked alignment").

Table 2. Structure and anti-HIV-1 activities of Training Set
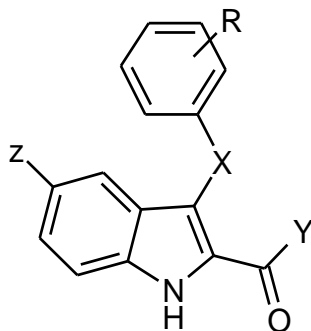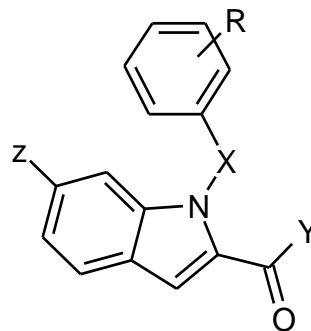
		A			B	
Compd	structure	X	Y	Z	R	wt$_{IIIB}$ EC$_{50}$ (μM)
11	A	SO$_2$	NH2	Cl	H	0.001
20	A	S	OEt	H	H	1.4
21	A	S	OEt	H	2-NH$_2$	>200
22	A	S	OEt	H	2-NH$_2$-5-Cl	≥200

Compd	structure	X	Y	Z	R	wt$_{IIIB}$ EC$_{50}$ (µM)
23	A	S	OEt	Cl	2-NH$_2$	2.3
24	A	S	OEt	Cl	2-NH$_2$-5-Cl	2.5
25	A	SO$_2$	OEt	H	H	3.7
26	A	SO$_2$	OEt	Cl	2-NH$_2$	>200
26	A	SO$_2$	OEt	H	2-NH$_2$-5-Cl	2.5
28	A	SO$_2$	OEt	Cl	2-NH$_2$-,5-Cl	1.9
29	A	S	NH$_2$	H	H	1.4
30	A	S	NH$_2$	H	2-NH$_2$-5-Cl	9
31	A	S	NH$_2$	Cl	H	0.02
32	A	S	NH$_2$	Cl	2-Me	0.3
33	A	S	NH$_2$	Cl	4-Me	>0.45
34	A	S	NH$_2$	Cl	4-F	1.4
35	A	S	NH$_2$	Cl	4-Cl	3.1
Compd	structure	X	Y	Z	R	wt$_{IIIB}$ EC$_{50}$ (µM)
36	A	S	NH$_2$	Cl	4-*iso*-Pr	1.9
37	A	S	NH$_2$	Cl	4-*tert*-Bu	8
38	A	S	NH$_2$	Cl	3,5-Me$_2$	0.006
39	A	S	NH$_2$	Cl	2,6-Cl$_2$	1.2
40	A	S	NH$_2$	Cl	2-NH$_2$-5-Cl	1.2
41	A	SO$_2$	NH$_2$	H	H	0.18
42	A	SO$_2$	NH$_2$	H	2-NH$_2$-5-Cl	0.3
19a	A	SO$_2$	NH$_2$	Cl	2-Me	0.001
19b	A	SO$_2$	NH$_2$	Cl	3-Me	0.001
19c	A	SO$_2$	NH$_2$	Cl	4-Me	0.003
43	A	SO$_2$	NH$_2$	Cl	4-F	0.014
44	A	SO$_2$	NH$_2$	Cl	4-Cl	0.011
45	A	SO$_2$	NH$_2$	Cl	4-*iso*-Pr	0.08
46	A	SO$_2$	NH$_2$	Cl	4-*tert*-Bu	0.13
19d	A	SO$_2$	NH$_2$	Cl	2,4-Me$_2$	0.004
19e	A	SO$_2$	NHy	Cl	3,5-Me$_2$	0.001
47	A	SO$_2$	NH$_2$	Cl	2,6-Cl$_2$	0.1
48	A	SO$_2$	NH$_2$	Cl	2-NH$_2$-5-Cl	0.04
49	A	SO$_2$	NH$_2$	Br	3,5-Me$_2$	0.002
50	A	SO$_2$	NHy	COMe	3,5-Me$_2$	0.015
51	A	SO$_2$	NH$_2$	CH(OH)Me	3,5-Me$_2$	0.025
52	A	S	NHNH$_2$	Cl	H	0.55
53	A	S	NHNH$_2$	Cl	4-Me	1.5

Compd	structure	X	Y	Z	R	wt$_{IIIB}$ EC$_{50}$ (μM)
54	A	S	NHNH$_2$	Cl	4-F	5
55	A	S	NHNH$_2$	Cl	4-Cl	10
56	A	S	NHNH$_2$	Cl	2-NH$_2$-5-Cl	>13
57	A	SO$_2$	NHNH$_2$	H	H	0.53
58	A	SO$_2$	NHNH$_2$	Cl	H	0.01
59	A	SO$_2$	NHNH$_2$	Cl	4-Me	0.05
60	A	SO$_2$	NHNH$_2$	Cl	4-F	0.32
61	A	SO$_2$	NHNH$_2$	Cl	4-Cl	0.19
Compd	structure	X	Y	Z	R	wt$_{IIIB}$ EC$_{50}$ (μM)
62	A	SO$_2$	NHNH$_2$	Cl	3,5-Me$_2$	0.13
63	A	SO$_2$	NHNH$_2$	Cl	2-NH$_2$-5-Cl	0.3
64	B	SO$_2$	H	H	H	>150
65	B	SO$_2$	COOEt	H	4-Cl	>100
66	B	SO$_2$	COOEt	H	2-NO$_2$	1.8
67	B	SO$_2$	COOEt	H	2-NO$_2$-5-Cl	>100
68	B	SO$_2$	COOEt	H	2-NH$_2$-5-Cl	1.8
69	B	SO$_2$	H	3-COOEt	H	>32
70	B	SO$_2$	H	3-COO-*iso*-Pr	H	>54
71	B	SO$_2$	H	5-Cl	H	>75
72	B	SO$_2$	COOEt	5-Cl	H	>200
73	B	SO$_2$	COOEt	5-Cl	4-Me	>200
74	B	SO$_2$	COOEt	5-Cl	4-Cl	>200
75	B	SO$_2$	COOEt	5-Cl	2-NO$_2$-5-Cl	>31
76	B	SO$_2$	COOEt	5-Cl	2-NH$_2$-5-Cl	8.3
77	B	SO$_2$	H	5-Cl,3-COOEt	H	>42
78	B	SO$_2$	H	5-Cl,3-COO-*iso*-Pr	H	>200
79	B	SO$_2$	CONH$_2$	H	H	15
80	B	SO$_2$	CONHNH$_2$	H	H	>200
81	B	SO$_2$	CONH$_2$	5-Cl	H	66.6
82	B	SO$_2$	H	3-CONH$_2$	H	>32
83	B	SO$_2$	H	5-Cl,3-CONH$_2$	H	≥200

Table 2. Structures and anti-HIV-1 activities of compounds used as the 3-D QSAR training set.
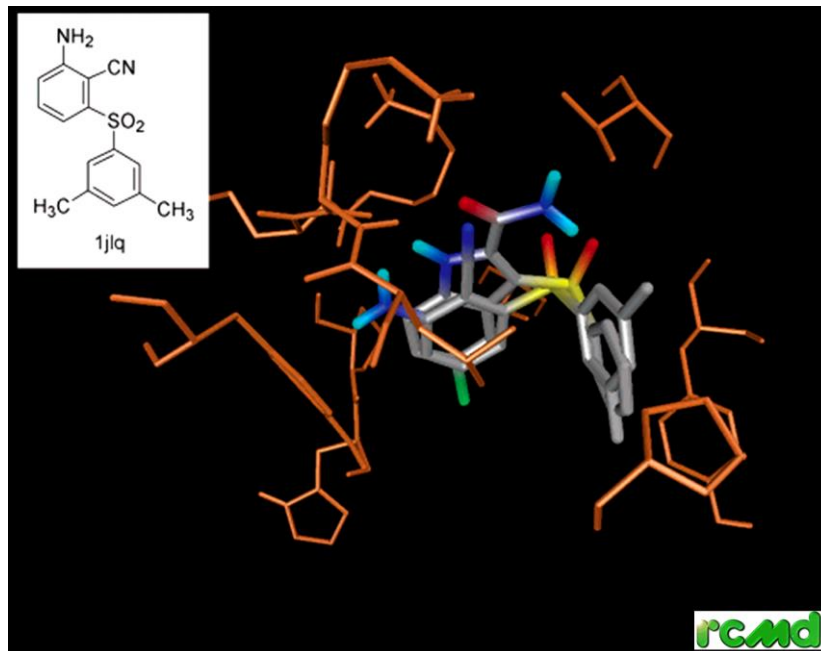
Figure 16. Structure of 739W94 (left). Superimposition of the experimental conformation found in the 739W94/RT with the docked pose of **90** (right).

On the aligned molecules, the molecular interaction fields (MIF) were calculated by the GRID software [93] using either the aromatic carbon (C1) or water (OH2) as probes, alone and in combination.

The MIFs of the training set were imported in GOLPE along with the inverse of the logarithm of the experimental anti-HIV-1 activities (pEC_{50}s). It could be argued that cell-based activities would not be suitable for such a 3-D QSAR study, enzyme-based assays (IC_{50}s) being a more appropriate choice. Nevertheless, it has been proved that there is a good

correlation between enzyme-based (IC_{50}) and cell-based (EC_{50}) assays, and therefore, data derived from cell-based activities were expected not to affect the 3-D QSAR.

By combining two different alignment strategies (best docked or best cluster docked conformation) and three probes arrangements (C1, OH2 and C1+OH2), six 3-D QSAR models were developed by means of the PLS algorithm implemented in the GOLPE program[92]. After data pretreatment, the smart region definition (SRD) was used in the fractional factorial design (FFD) variable selection[138]. The FFD selections were continued until no further statistical improvements were observed. To measure the goodness of the model, the statistical indices r^2, q^2, and SDEP were employed. The six final 3-D QSAR models were characterized by correlation coefficient (r^2), predictive correlation coefficient (q^2), and cross-validated standard deviation of errors of prediction ($SDEP_{CV}$) values falling in the ranges 0.79-0.93, 0.59-0.84, and 0.69-1.07, respectively (*Table 3*). For both the FFD selections and the cross-validations, the group method was used, setting to five the number of groups to be used.

Table 3. Statistical results of the 3-D QSAR models

Model	Alignment	Grid probes	vars[a]	PC[b]	r^2	q^2	$SDEP_{CV}$
1	best cluster	C1=	665	3	0.88	0.72	0.88
2	best docked	C1=	683	3	0.93	0.84	0.69
3	best cluster	OH2	752	2	0.85	0.68	0.95
4	best docked	OH2	669	3	0.92	0.81	0.74
5	best cluster	C1= and OH2	1140	2	0.79	0.59	1.07

6	best docked	C1= and OH2	709	3	0.91	0.79	0.76

Table 3 [a]Number of selected variables. [b]Number of principal components which showed maximum q^2 value.

As the six 3-D QSAR models were endowed of similar statistical profiles, an external test set (compounds **334**-**341**) taken from a previously reported series of indolyl aryl sulfones was used to measure their predictive ability (*Table 4*). On the basis of the standard deviation error of prediction value on the test set (SDEP$_{Test-Set}$), the external validation proved the best cluster alignment/OH2 GRID probe combination (model 3, SDEP$_{Test-Set}$ = 0.88, *Figure 17*) as the most predictive. Selected for description of the 3-D QSAR maps and it was chosen as a tool for design and prediction of pEC$_{50}$s of new IAS derivatives. Apart from the optimal combination between the alignment and the probe choice, the higher predictive ability of model 3 can also be attributed to the fact that the model is defined with two principal components (PC) only. In fact, it seems that the extraction of a further PC (models 1, 2, 4, and 5) enhances the self-consistency (cross-validation) of the model, but less predictive 3-D QSAR are obtained (*Table 4*).

6.2.2 Model interpretation

3-D QSAR models are built with the aim to predict the biological activity of new untested compounds[139]. The prediction ability of the 3-D QSAR model depends on the PLS coefficients associated to each variable. The regression PLS coefficients 164 are useful for the

interpretation of the PLS model and for predictions of external objects (test set). In order to interpret the model, the PLS coefficients are plotted as contour plots (or polyhedrons) connecting grid points with similar values.

$$y_E = c_{pls}\, x_1 + c_{pls}\, x_2 + c_{pls1}\, x_3 + \ldots + c_{plsn}\, x_n$$

Equation 1. y_E is the estimation of y; n is the number of grid points; c_{plsn} is the PLS coefficient in the nth grid point; x_n is the actual field in the nth grid point.

Table 4. Test set predictions and SDEP values obtained with models 1-6
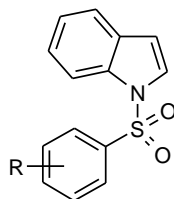
Compd	R	exp[a]	M1	M2	M3	M4	M5	M6
					pEC$_{50}$ predicted[b]			
132	2-NO$_2$	5.30	5.37	5.87	4.42	4.59	4.77	4.08
133	2-NH$_2$	4.96	5.96	5.89	5.33	5.58	5.76	5.39

134	2-NO$_2$-5-Cl	5.40	5.11	3.26	4.42	3.68	4.33	4.32
135	2-NH$_2$-5-Cl	6.00	6.43	2.70	5.28	4.52	5.74	3.82
136	2-NO$_2$-4-Cl	4.80	3.43	5.58	4.19	5.18	3.71	4.61
137	2-NH$_2$-4-Cl	4.13	3.79	5.56	3.81	5.35	3.54	5.28
138	2-Cl-5-NO$_2$	5.22	5.19	5.35	4.68	5.50	5.12	4.53
139	2-Cl-5-NH$_2$	3.82	6.31	6.23	5.56	6.39	5.96	5.55
SDEP$_{Test-Set}$			1.09	1.31	0.88	1.29	1.08	1.25

Table 4 [a]Experimental pEC$_{50}$ (pEC50 = -log EC$_{50}$ (M)). [b]Predicted pEC$_{50}$: M1 = model 1, M2 = model 2, M3 = model 3, M4 = model 4, M5 = model 5, M6 = model 6.

Equation 1 may help the interpretation of Figure 18. The c_{pls} coefficients are needed for predictions of the biological activity of new molecules. However, since the sizes and signs of the coefficients reveal the relative influence of each grid point on y, they are also suitable for the interpretation. That is, a new compound with a substituents protruding into a region with negative c_{plsr} produces a positive (repulsive) field x_r in this region.
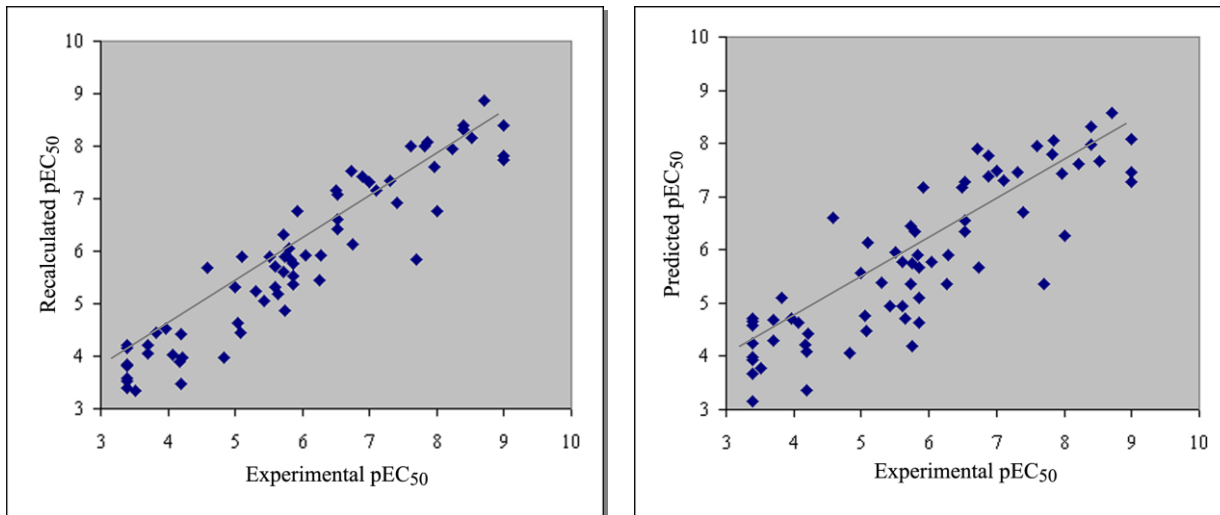
Figure 17. Fitting (up) and cross-validation (down) plots of the model 3 with two principal components.

Consequently $c_{pls}\, x_r$ is negative, thus indicating a negative influence on y_E (a low pEC$_{50}$ corresponds to low affinity). The opposite is valid if the region has positive or negative c_{pls} coupled with positive or negative fields, respectively.

One important feature of 3-D QSARs is the graphical representation of the model, which makes its interpretation easier. In the GOLPE software, several options are available to display the final model. Among these, the PLS Coefficients and the Activity contribution plots are very useful. The PLS Coefficients plot (c_{pls}) allows the visualization of the selected grid points at determined molecules/probe interaction energy level, indicating graphical information of the whole training set. However, in this plot the signs of the coefficients can

induce errors, as reported with *eq.1*, coefficients have opposite meaning depending on the fact that the compound produces positive or negative field values in a specific area ($c_{pls}\,x_1 + c_{pls}\,x_2$). On the other hand, the Activity contribution plot gives the possibility to display spatial regions that are individually important for the selected molecule. The activity contribution plot, different for every molecule within the training set, results from multiplying the values of the coefficients by the actual values of the field for that molecule (y_A) ($y_A = c_{pls}\,x_1 + c_{pls}\,x_2$)

In *Figure 18* the negative (red contours) and positive (blue contours) PLS coefficients are superimposed to **19e** (the most active derivative, EC_{50} = 0.001 μM) and **40** (one of less active compounds, EC_{50} = 1.2 μM) activity contribution plots (positive contribution in yellow polyhedron and negative contribution in cyan) and a 4Å core of the NNBS.

The contemporary representation allows an easy appreciation of important interactions on the basis of polyhedrons' disposition and size. The most important differences were associated with the polyhedrons around the 3,5-dimethylphenyl moiety and the sulphonyl bridge at position 3 of the indole. The presence of two yellow areas around **19e**'s 3,5-dimethylphenyl moiety (positive activity contribution) highlighted the importance of hydrophobic interactions with this group. In contrast, unfavourable (yellow polyhedrons) interactions were recorded for the poorly active derivative **40**. A further inspection of the PLS plots revealed the S → SO₂ interchange as second important feature. Sulphone derivatives displayed a yellow area (positive contribution) around the sulphonyl group, while the sulphur counterparts lacked any positive contribution in the same region. Other positively contributing areas were observed in the proximity of the indole NH and the C-2 side chain. The presence of these polyhedrons in both highly and poorly active derivatives proved that these

interactions are a minimum requisite in order to display some biological activity.
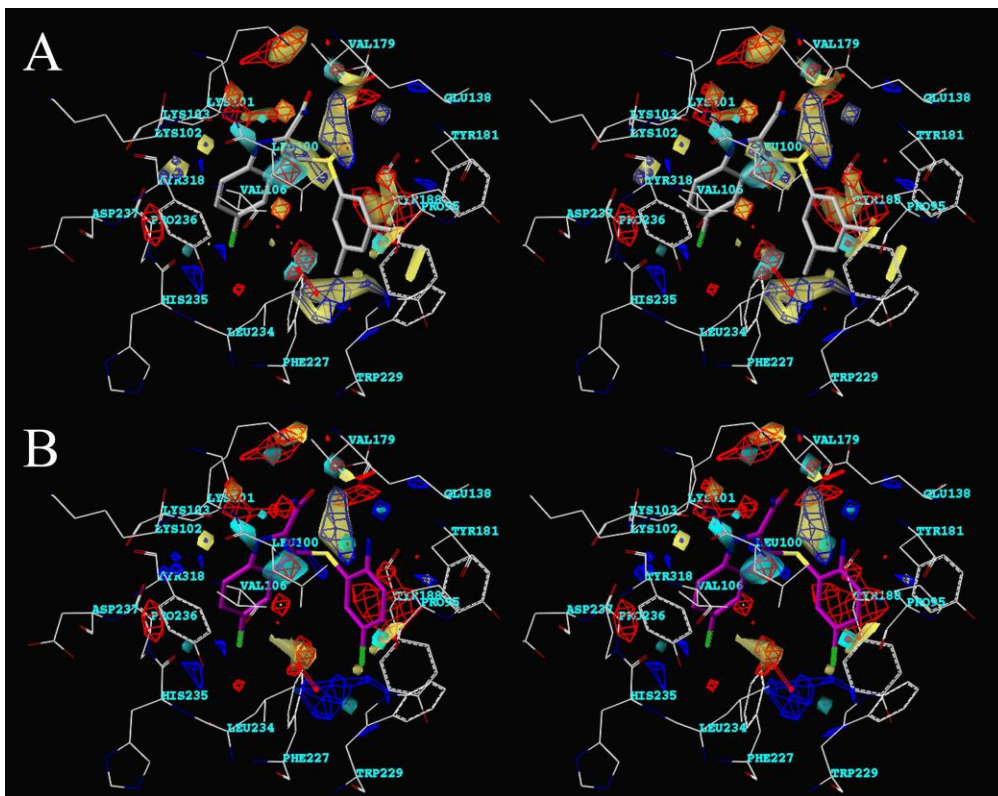
Figura 18. Stereo view of the PLS coefficients. Polyhedron in red are negative coeff. (data level = -0.008), polyhedron in blue are positive coeff (data level = +0.008). Are also reported the activity contribution plots for **19e** (Figure 7A, carbon atoms in white) and **40** (Figure 7B, carbon atoms in magenta). Positive contribution (data level = +0.035) to the activity is yellow polyhedron. Negative contribution (data level = -0.035) to the activity is cyan polyhedron.

6.2.3 Design of new IASs through molecular modelling and 3-D QSAR studies

Merck's findings[147] suggested that little chemical modification at the indole-2-carboxamide may lead to derivatives endowed with improved anti-HIV-1 activity against the drug-resistant mutant strains[140]. In addition, we found that insertion of 1-3 aminoacid units as chain elongators of the 2-carboxamide led to new potent IAS derivatives[141]. Therefore, we modelled 24 new IAS structures characterized by the presence of 2-[*N*-(2-hydroyethyl)]aminocarbonyl or 2-[*N'*-(2-hydroyethyl)]hydrazinocarbonyl groups at the indole-2-position (**92-115**) (*Table 5*).

The modelled IAS derivatives **92-115** were structure-based aligned as above described, to predict the associated pEC_{50} values. Among them, four compounds predicted as the most active (carboxamides **102** and **103**, and carbohydrazides **112** and **115**) along with carboxamide **92** and carbohydrazide **108** predicted as the least active derivatives, were synthesized and tested for their anti-HIV activity. Although the 3-D QSAR model was slightly under-predictive (*Table 5*), the predicted pEC_{50}s showed the same trend of the experimental values. Deeper inspection of the **92**, **102**, **103**, **108**, **112** and **115** docked conformations allowed disclosing an alternate binding mode compared with that of **11**. In particular, although we observed a 180° rotation of the indole nucleus (*Figure 19*), the interaction pattern was very similar to that already observed for the previous IASs. The benzene rings of the new compounds **92**, **102**, **103**, **108**, **112** and **115** were practically overlapped to that of **11** inside the NNBS aromatic rich pocket, and the hydrogen bond with

Lys101 was also maintained. These observations confirmed our hypothesis claiming the correct accommodation of the aryl ring and the hydrogen bond with Lys101 crucial requirements for the anti HIV-1 activity. Furthermore, we found that the alternative binding mode of the new IASs was associated with a hydrogen bond formed between the terminal OH and the carbonyl of Leu234 (that is one of the most conserved residues of NNBS[142]) (*Figure 19*). We correlated the under-predictive behaviour of the 3-D QSAR model to the limited chemical variability of the training set, and consequently to the presence of unexplored spaces around the training set and inside the NNBS.

Regarding the low predicted (pEC$_{50}$ < 7) designed IAS, inspection of derivatives **93**, **96**, **108** and **114** Autodock3 poses revealed different and not optimal binding modes that could account for the lower predicted pEC$_{50}$. On the other hand the low pEC$_{50}$ predicted sulphurs compounds **92**, **94-95**, **97**, **104-107** and **109**, although displaying binding modes similar to the corresponding higher pEC$_{50}$ predicted sulphones (**100-10**3 and **110-115**), lack of some interactions with the Val106, Val179 and Lys103 small hydrophobic pocket.

Table 5. pEC$_{50}$ Predictions of 2-[*N'*-(2hydroyethyl)]carboxamide and 2-[*N'*-(2-hydroyethyl)]carbohydrazide IAS Derivative
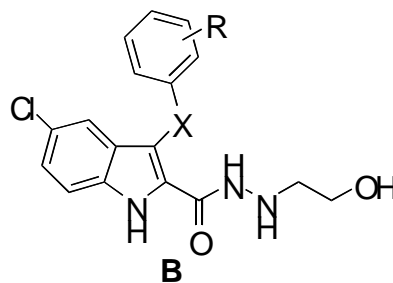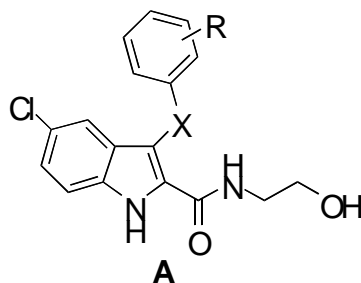
Compd	Structure	X	R	predicted pEC$_{50}$
92	A	S	H	5.93
93	A	S	2-Me	5.96
94	A	S	3-Me	6.00
95	A	S	4-Me	6.54
96	A	S	2,3-Me	5.97
97	A	S	3,5-Me	6.41
98	A	SO$_2$	H	6.76
99	A	SO$_2$	2-Me	6.32
100	A	SO$_2$	3-Me	7.05
101	A	SO$_2$	4-Me	7.11
102	A	SO$_2$	2,4-Me	7.13
103	A	SO$_2$	3,5-Me	7.53
104	B	S	H	6.49
Compd	Structure	X	R	predicted pEC$_{50}$
105	B	S	2-Me	6.70
106	B	S	3-Me	6.12
107	B	S	4-Me	6.55
108	B	S	2,4-Me	6.16
109	B	S	3,5-Me	6.62
110	B	SO$_2$	H	7.06

111	B	SO_2	2-Me	7.10
112	B	SO_2	3-Me	7.36
113	B	SO_2	4-Me	6.98
114	B	SO_2	2,4-Me	6.65
115	B	SO_2	3,5-Me	7.79

Table 5

Table 6. Predicted pEC_{50} and $SDEP_{Pred-Set}$ value of selected compounds

Compd	pEC_{50}	
	exp[a]	pred[b]
92	6.98	5.93
102	8.00	7.13
103	9.00	7.53
108	5.64	6.16
112	8.00	7.336
115	8.00	7.73
$SDEP_{Pred-Set}$	-	1.00

Table 6 [a]Test Set experimental values. [b]Test Set prediction values
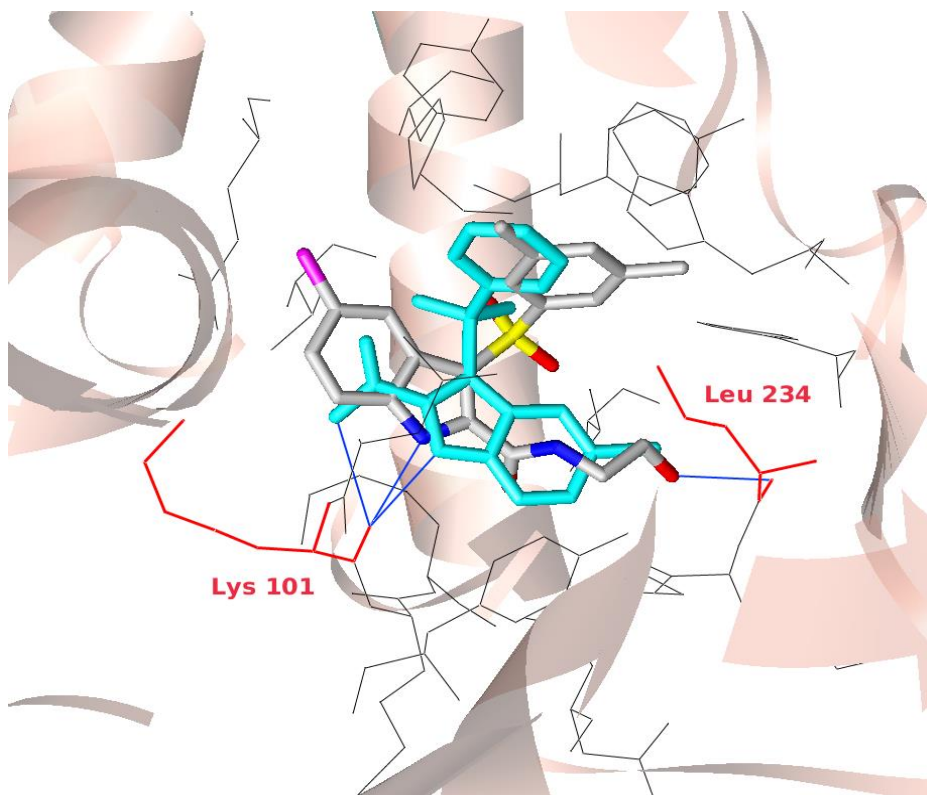
Figure 19. Comparison between **11** (cyan) and **103** (atom type) binding modes. Blue lines represent the hydrogen-bonds with Lys101 (in red, left side) and Leu234 (in red, right side). A 4 Å core of NNBS is displayed (black lines). Hydrogen atoms are omitted for sake of clarity.

Table 7. Antiviral activities

Compd	EC$_{50}$ (μM)				
	wt$_{IIIB}$	Y181C	Y181C-	112	AB1

	K130N				
103	0.0008	0.009	0.3	0.05	0.01
115	0.001	0.04	3.2	0.055	0.07
NVP	0.05	14	>20	>20	>20
EFV	0.003	0.01	0.2	>20	>20

Table 7. Antiviral activities of **103** and **115** against HIV WT, and Y181C and K103N-Y181C mutant strains. The activities against the clinical isolates 112 and AB1 are also reported.

6.3 Cross-docking experiments into mutated RTs

Progress made in binding mode of IAS derivatives suggested to extend these studies towards the mutated forms of RT. Docking and cross-docking studies were carried out of the most active derivatives **103** and **115** into the WT RT, and the single (Y181C, K103N) and double (Y181C-K103N) mutated forms [143]. WT (PDB entry code 1VRT$_{WT/NVP}$[136]; PDB entry code 1FK9$_{WT/EFV}$ [128]), Y181C (PDB entry code 1JLB$_{Y181C/NVP}$[136]; PDB entry code 1JKH$_{Y181C/EFV}$[144]) and K103N (PDB entry code 1FKP$_{K103N/NVP}$[145]; PDB entry code 1FKO$_{K103N/EFV}$[146]) experimental RT coordinates in complex with either NVP or EFV were promptly available. Y181C-K103N RT was extracted from 1JKH (K103N RT) structure by a virtual point mutation of the Tyr181 to Cys181. It was obtained introducing a virtual point mutation in an Y181C mutated RT.(PDB entry code 1jkh) Coordinates of second mutated residue was selecting from a K103N mutated RT (PDB entry code 1FKO)[146] Both RT used were obtained by crystallography technique, it present high value of resolution (1FKO 2,5Å; 1FKO 2.9 Å) and it were co-crystallized with the same inhibitor (efavirenz).

After alignment, performed by ProFit, 1jkh Lys103 coordinates residue was substituted with 1fko Arg103 coordinates. At this point alignment procedure was repeated. The minimization, necessary to optimized complex structures was performed with Sander module; the partial atomic charges were derived by Antechamber module.

The root mean square of deviation (RMSD) for all the heavy atoms between the binding modes of **103** and **115** in either 1VRT (NVP/RT$_{WT}$ complex) or 1FK9 (EFV/RT$_{WT}$ complex) extracted NNBSs, and those in 1JLQ (739W94/RT$_{WT}$ complex)[131] confirmed our findings (**103**: RMSD1VRT-1JLQ = 1.918; RMSD1VRT-1JLQ = 1.646. **115**: RMSD1VRT-1JLQ_11 = 1.098; RMSD1VRT-1JLQ = 1.613). The cross-docked binding modes of **103** and **115** in RT$_{WT}$ did not change with respect to those previously reported [137]. Again, crucial interactions were observed between Leu234 and OH of the 2-hydroxyethyl chain (*Figure 19*). Either **103**'s or **115**'s 2-hydroxyethyl chains fitted into the NNBS entrance channel formed by Val106, Pro225, Phe227, Leu234, His235, Pro236 and Tyr318, thus making both steric and electrostatic contacts.

The Tyr181 ➜ Cys181 mutation weakly affected the Autodock proposed **103**/**115** binding modes, and showed conformations close to those observed for the RT$_{WT}$ (*Figure 20*). The **103** bound conformations in either NVP/RT$_{Y181C}$ or EFV/RT$_{Y181C}$ derived NNBSs were closely related to those obtained from the corresponding RT$_{WT}$. Differently, the docked conformation of **115** into NVP /RT$_{Y181C}$ enzyme was slightly shifted, likely due to a less effective anchorage of the molecule inside the NNBS. We assumed that this difference may contribute to the poorer activity of compound **115** against the Y181C mutation (compare **103** and **115**

EC$_{50}$s in *Table 7*).

Lipophylic calculations revealed higher ALOGPs [145] value for **103** (2.31) compared to that of **115** (ALOGPs = 1.66). These observations suggested addressing their different biological activity to other factors, such as membrane permeability. Switching Lys103 ➔ Arg103 led to different results with respetct to the Y181C mutation. In NVP/RT$_{K103N}$, both compounds showed only little modification of the RT$_{WT}$ binding modes (RMSD $_{NVP /K103N/}$**103** =1.167; RMSD $_{NVP /K103N/}$**115** = 1.947). On the contrary, the in EFV/RT$_{K103N}$ mutation thse compounds did not show RT$_{WT}$-like binding modes (as showed by high RMSD values (>5.5) for both derivatives) (*Figure 21*); a wider cross-docking study is ongoing to confirm these data, taking in to account also the new mutated structure deposited in Protein Data Bank..

Figure 19. Docked conformations of **103** (green) and **115** (white) in the 1VRT (A) and 1FK9 (B) WT NNRTI binding pockets. Coloured in atom type and displayed in stick are the non-mutated

Lys103 and Tyr181. Hydrogen's atoms are not displayed for sake of clarity.

The contemporary mutation of Tyr181 ➔ Cys181 and Lys103 ➔ Arg103, displayed different docked conformations from those observed for the single mutated Y181C or K103N RTs. In the double mutant, IAS **103** displayed a binding mode similar to that of L-737,126 in RT$_{WT}$, while derivative **115** showed a different disposition (*Figure 22*). Such differences may account of **103**'s higher activity against this double mutant strain. In addition these findings are in agreement with our observations[123]: a fixed binding mode of the molecule positively affects the activity against mutated RTs.

A	B

Figure 20. Docked conformations of **103** (green) and **115** (white) as found by Autodock in the 1JLB (A) and 1JKH (B) Y181C mutated NNRTI binding pockets. Colored in atom type and displayed in stick are the non-mutated Lys103 and the mutated Tyr181➔Cys181 residues. Hydrogen's atoms are not displayed for sake of clarity.

Compounds **103** and **115** were also evaluated against the HIV-112 and HIV-AB1 strains. The 112 strains was isolated from patient treated with NRTIs/NNRTIs based regimens, and carried K103N-V108I-M184V RT mutations that lead to >99% resistance to NVP and EFV. The AB1 strain was obtained after HAART treatment and carried the L100I-V108I RT mutations responsible of >90% and >85% resistance to NVP and EFV, respectively. Interestingly, both compounds were active at nano molar concentration against either HIV-112 or HIV-AB1 strains (*Table 7*).
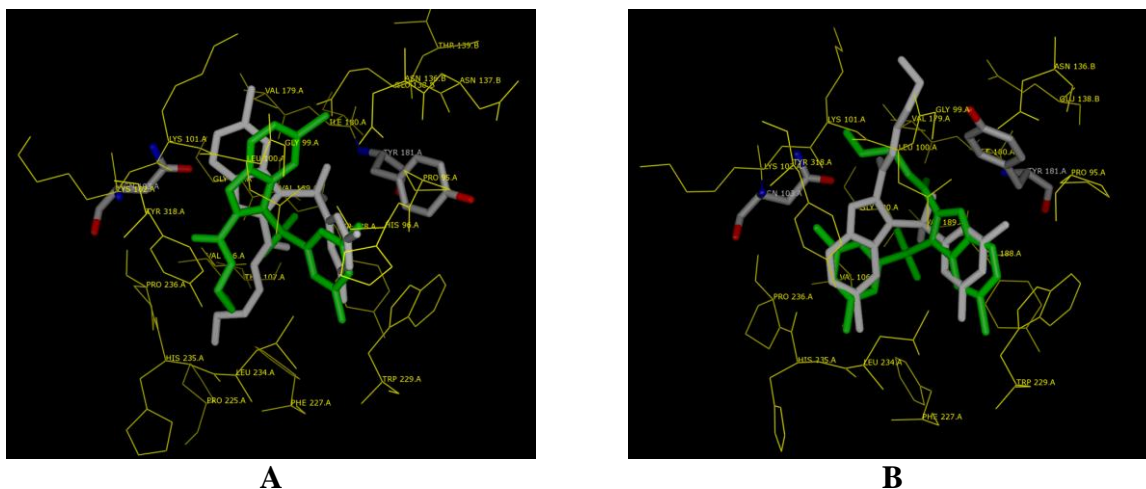
Figure 21. Docked conformations of **103** (green) and **115** (white) in the 1FKP (A) and 1FKO (B) K103N mutated NNRTI binding pockets. Coloured in atom type and displayed in stick are the mutated Lys103➜Arg103 and the unchanged Tyr181 residues. Hydrogen's atoms are not displayed for sake of clarity.
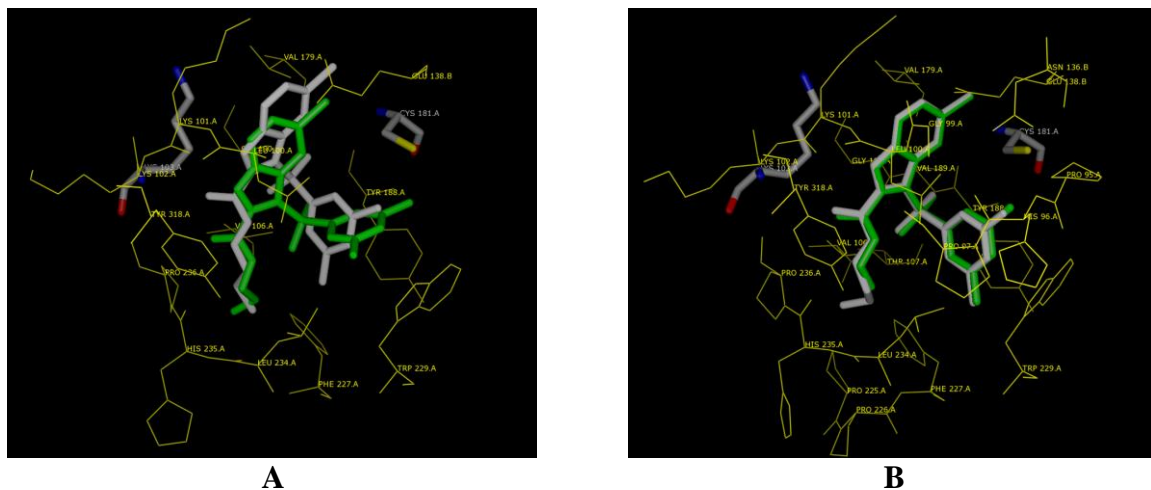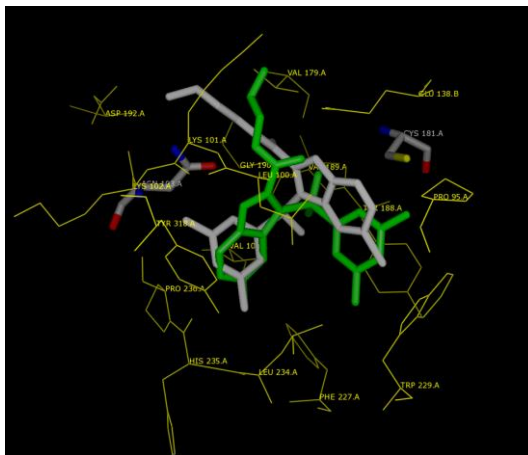
Figure 22. Docked conformations of **103** (green) and **115** (white) as found by Autodock in the K103N-Y181C modeled NNRTI binding pocket. Colored in atom type and displayed in stick are the mutated Lys103 ➔ Arg103 and Tyr181 ➔ Cys181 residues. Hydrogen's atoms are not displayed for sake of clarity.

6.4 The second improved 3-D QSAR model

One of the main goals of the medicinal chemists is to improve the predictive ability of the model. To this end, the initial training set is enriched with new experimental information, thus starting over the cycle of designing new potential drugs.

6.4.1 Definition, prediction ability and interpretation

The above reported 3-D QSAR model was characterized by an under-predictive behaviour (*Figure 13A*). The limited number of molecules in the training set (limited chemical diversity) may be a possible explanation. While attempting to overcome this problem, we developed a new 3-D QSAR model by increasing the number of molecules in the training set from 70 to 101[148]. In particular the new training set was obtained by using the old test and prediction sets (eight and twenty-three compounds respectively, *Tables 3 and 4*) The addition of molecules bearing different substituents at the 2-carboxamide function boosted the predictive ability of the previous model.

As for the first 3-D QSAR studies, the statistical indexes r^2, q^2 and SDEP were employed to measure the goodness of the update model (*Table* 7). To determine the model dimensionality (number of optimal principal components, PCs) and hence evaluate its robustness, three cross-validation levels were employed: the leave one out (LOO) method and two leave some out (LSO) methods. Either five randomly selected groups (leave five out, L5O) or the extreme two randomly selected groups (leave half out, LHO) were used. The final new 101 objects composed a 3-D QSAR model which was defined by three PCs and r^2, q^2 LOO, q^2 LFO and q^2 LHO values equal to 0.90, 0.79, 0.78 and 0.73, respectively. Comparing these statistical values with those of the precedent model ($r^2 = 0.85$ and $q^2 = 0.63$), the updated model highlighted its higher robustness and predictivity power (*Table 8*). For any 3-D QSAR model, an external test taken in addition to the internal validation (cross-validation) is mandatory to really evaluate its predictive ability. To this end, a test set of eight hybrid IAS-peptide derivatives was compiled [141] (*Table 9*). The chemical composition

of the TS was suitable (i) to measure the predictive power of the new model, and (ii) to compare its under-predictivity with that of the previous model. As expected, the new 3-D QSAR model was highly predictive and the under-predictive behaviour almost disappeared (*Figure 23B, Table 9*).

Table 8. Statistical Values of the 2nd 3-D QSAR Model

FFD	Vars	PC	r2	LOO q2	LOO SDEP	L5O q2	L5O SDEP	LHO q2	LHO SDEP
0	3834	2	0.69	0.51	1.16	0.49	1.18	0.43	1.25
1	1690	2	0.77	0.66	0.96	0.65	0.97	0.60	1.05
2	1178	2	0.81	0.72	0.88	0.71	0.90	0.66	0.97
3	942	2	0.84	0.76	0.81	0.75	0.83	0.70	0.91
4	840	2	0.86	0.78	0.77	0.77	0.78	0.72	0.88
5	802	3	0.90	0.79	0.75	0.78	0.77	0.73	0.86
6	740	3	0.90	0.81	0.72	0.80	0.74	0.75	0.83
7	729	3	0.90	0.81	0.72	0.80	0.74	0.75	0.83

Table 9. Test set structures, ALOGpS values, and pEC$_{50}$ predictions
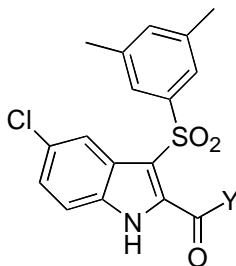
Compd	Y	ALOGpS (mg/L)	Exp. pEC_{50}	New Pred. pEC_{50}	Old Pred pEC_{50}
116	GlyNH$_2$	-4.72	8.222	9.093	7.584
117	GlyNHNH$_2$	-4.53	8.000	9.044	7.543
118	GlyGlyNH$_2$	-4.69	7.222	8.081	4.082
119	GlyGlyNHNH$_2$	-4.50	7.097	5.933	7.649
120	GlyAlaNHNH$_2$	-4.64	7.854	7.641	4.875
121	GlyAlaNH$_2$	-5.07	6.921	7.346	7.503
122	GlyGlyGlyNH$_2$	-4.69	6.745	5.988	6.388
123	GlyGlyGlyNHNH$_2$	-4.43	9.155	7.897	5.386

Table 9. Test set structures, ALOGpS values, and pEC_{50} predictions using the updated 3-D QSAR model (New Pred). For comparison purpose the predictions of the previous model are shown (Old Pred).

As a measure of the improved prediction ability, the test set standard deviation error of prediction ($SDEP_{TS}$) was calculated for both the old and new models; the updated 3-D QSAR showed a $SDEP_{TS}$ value of only 0.89, while for the old one the same coefficient was 2.05 (*Table 10*). In particular the old model reported **118** and **120** as false negatives (*Figure 23B*) in the high micromolar range of concentration. In contrast, the new model correctly predicted the antiviral activity at low (**120**) and sub-micromolar (**118**) concentrations. It is worthy to note, using the IAS-peptide set the new model correctly predicted the overall activity trend, while the old one did not (*Figure 23B, Table 9*).
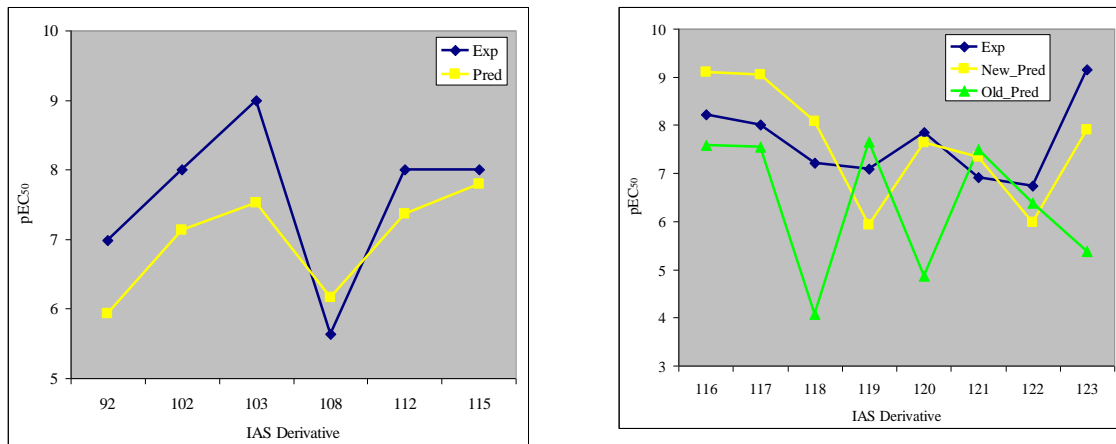
Figure 23. A: plot of prediction set pEC_{50} experimental and predicted values of the old 3-D QSAR model. B: plot of test set pEC_{50} experimental and predicted values of the new 3-D QSAR model.

Table 10. Comparison of statistical coefficients and predictive ability

Model	PC	r^2	q^2 L5O	$SDEP_{CV-L5O}$	$SDEP_{TS}{}^a$
Old	2	0.85	0.68	0.95	2.05
New	3	0.90	0.78	0.77	0.89

Table 10. Comparison of statistical coefficients and predictive ability between old and new 3-D QSAR models   [a]The same test set reported in table 13 was used for either models

The graphical potentiality of the GOLPE program was used for model interpretation. In *Figures 23 and 24* are shown the PLS coefficients, present field (*A and B of Figure 23*) and activity contribution (*C and D of Figure 23*) plots for compounds **95** and **103**. From Figures

23 (PLS coefficients plot) it is evident the homogeneous distribution of the positive (cyan) and negative (yellow) FFD selected grid points around the training set, almost without any uncovered areas. This result is an important source of information although such a crumbling of the PLS polyhedrons makes model interpretation hard. To overcome this difficulty, the model analysis was focused on the molecular structure, and the PLS coefficient polyhedrons (*Figures 23*) were divided into three areas: (i) a group around the 3-aryl moiety (blue circle); (ii) a group near the 5-chloroindole nucleus (magenta circle) and (iii) a group surrounding the 2-carboxamido/hydrazido group (green circle). Although **95** and **103** exhibited some similarity in chemical structure, both the structure-based alignment and the PLS algorithm successfully contributed to rationalize their 150-fold EC_{50}'s difference. From both the PLS coefficients and present field plots, **103**'s 3,5-dimethylbenzene moiety showed polyhedrons that correlated with favourable interactions into the NNBS aromatic rich sub-pocket (yellow polyhedrons with positive activity contribution in *Figures 24*). On the other hand, **95**'s 4-methylbenzene was slightly shifted and turned the same interaction into less favourable contacts (less yellow polyhedrons in *Figure 24B*). Regarding the 5-chloroindole moiety (magenta circled areas of *Figure 23*) no differences were observable for both **95** and **103**; the two indole rings were perfectly superimposed each other, and the 5-chloroindole of **95** showed favourable interactions through both the chlorine atom and the indole ring (compare PLS coefficients and activity contribution plots of *Figures 23 and 24*). The most important feature of the new 3-D QSAR model resided in the third group of polyhedrons (green circle in *Figures 23*). In fact, a comparison with the previous model in this area showed the greater differences that shed light on the possible role of the hydroxyethyl groups of either **95** or **103**

and all their related analogues.
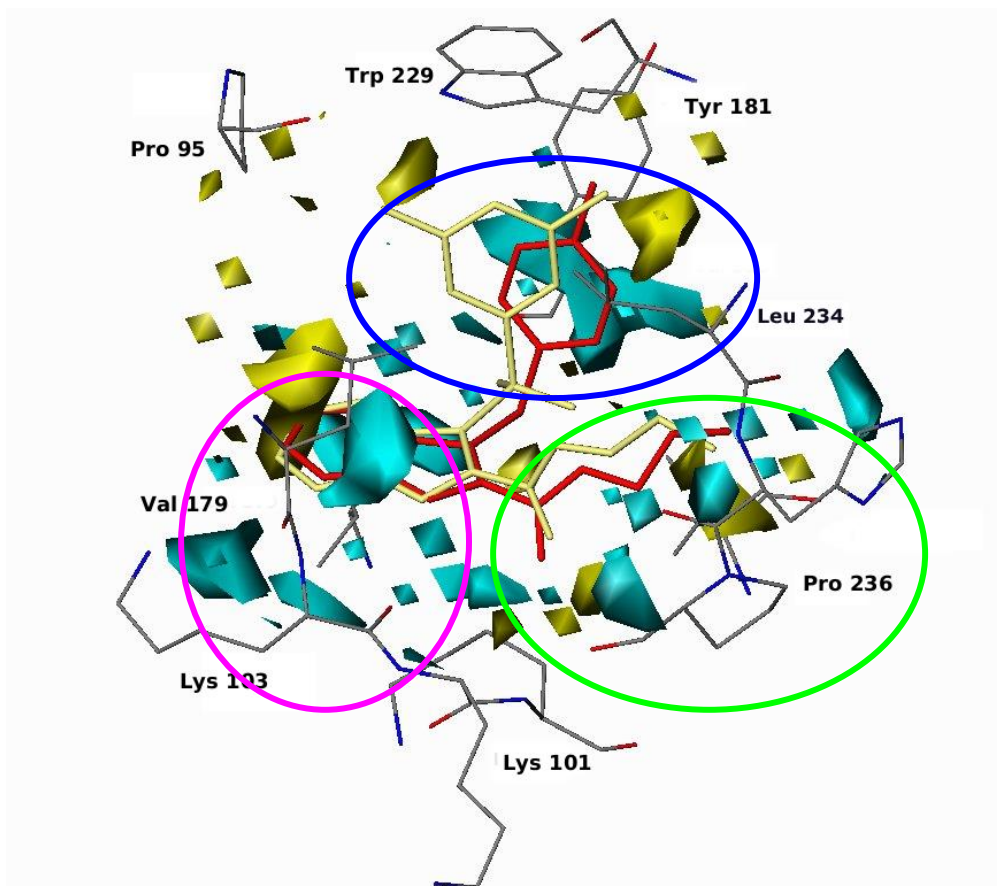
Figure 23. PLS-coefficient plot. Polyhedrons in yellow are negative coefficients (data level = -0.008); polyhedrons in cyan are positive coefficients (data level = +0.008); Compounds **103** (yellow) and **95** (red), and some relevant RT key residues are also displayed

Both the GOLPE[125] plots and Autodock3[121] conformations attached importance not only

to the 2-hydroxyethyl group, but also to its relative orientation into the NNBS. Again, the PLS coefficients correctly predicted that the OH group of **103** was optimally oriented to make electrostatic interactions (yellow polyhedrons in right side of *Figure 24A*; we hypothesized H bond between **103**'s OH and Leu234; $OH_{105}\ldots O=C_{Leu234} = 2.8\text{Å}$, *Figure 25*). While the hydroxyl group of **95** displayed a different orientation. This group avoided the formation of any hydrogen bond and made unfavourable interaction with Pro225; it is underlined by the small yellow field preceding the big cyan one (*B, D vs C, E in Figure 25*).
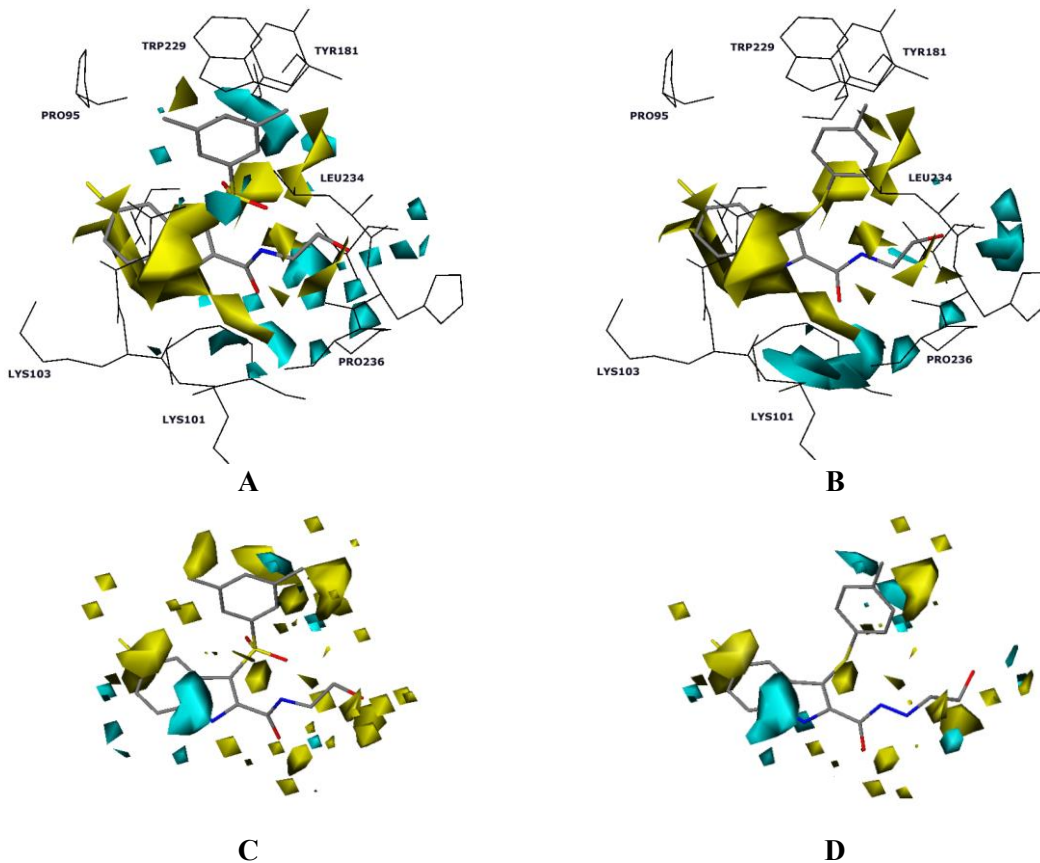
Figure 24. A and B**:** Present field plots. In yellow are negative fields (data level = 2.4), in cyan are positive fields (data level = +5.0). C and D: Activity contribution plots. In yellow are negative contributions (data level = -0.035), while in cyan are positive (data level = +0.035).A and C is the high active representative **103**, in B and D the less active **95**.

A

B

C
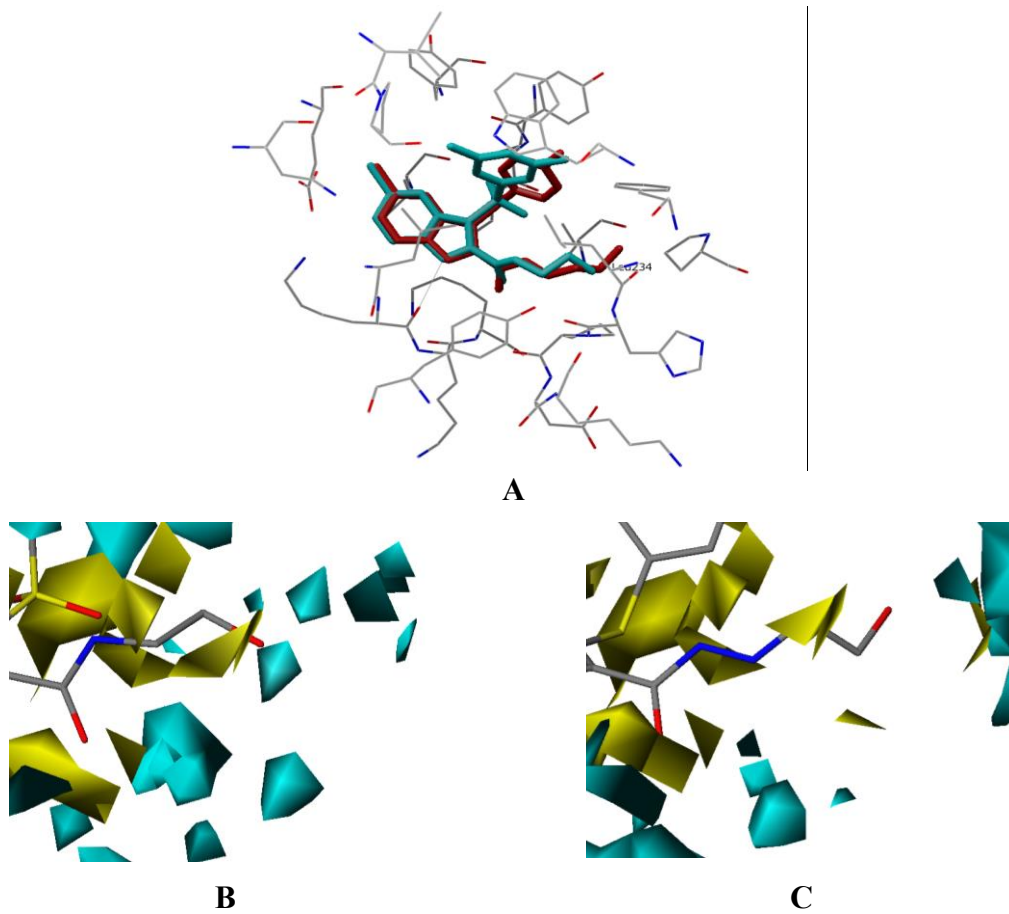
Figure 25. (To be continued**)**
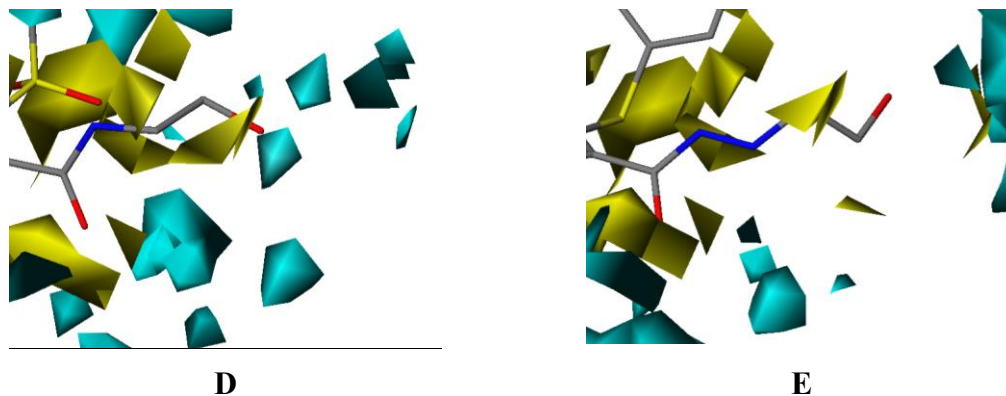
Figure 25. A: Comparison of **103** (cyan) and **95** (red) binding modes. B and C: Present fields plots focalized on the 2-hydroxyethyl groups of **103** (B, left) and **95** (C, right). D and E: Activity contribution plots focalized on the 2-hydroxyethyl groups of **103** (D, left) and **95** (E, right).

6.4.2 Design and activity prediction of new IASs

To ascertain whether the new model would predict untested derivatives, we designed eight new IASs (**124**-**131**) characterized by unreported structural modifications at the 2-carbohydrazide group of **115** (*Char t12 and Table 11*)[148].
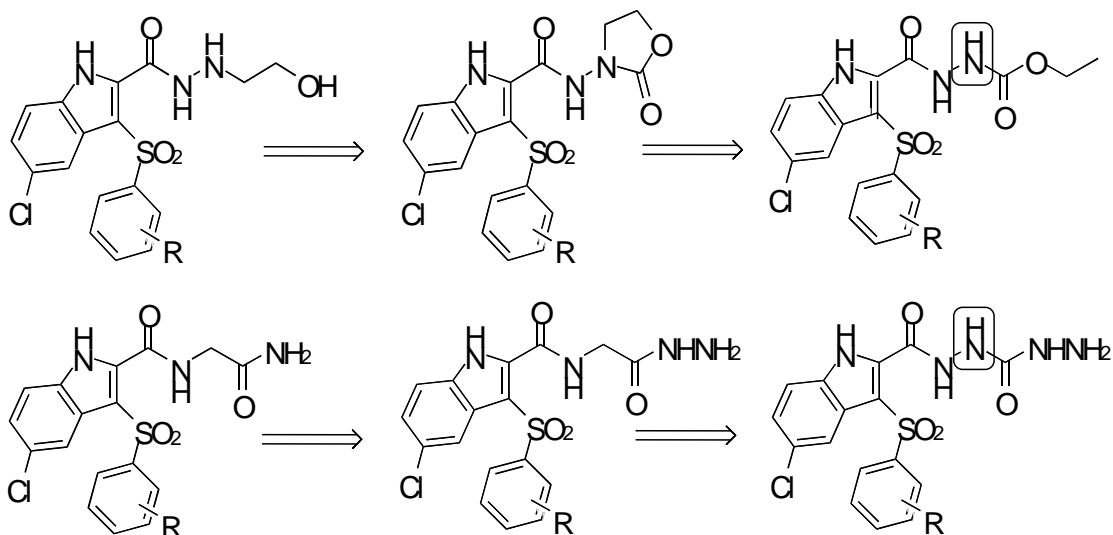
Chart 12. Design of new IAS derivatives.

Apart from the goal of validating the new procedure, the new compounds **124-131** were also designed to confirm the importance of the free hydroxyl groups of either **103** or **115**. In fact, binding mode analyses indicated that their 2-hydroxyethyl moieties lay in an hydrophobic pocket mainly formed by the side chains Val106, Pro225, Phe227, Leu234 and Pro236 (see *Figures, 19 and 25A*). Such an observation suggested the introduction of less hydrophilic substituents at the 2-carbohydrazide moiety of **5** (*Table 11*). Taking into account the contribution of sterical interactions, the size of the substituents was varied from the small methyl to the bulky cyclohexyl groups. Interestingly, prior of any quantitative application, the

binding modes proposed by Autodock gave some anticipation of their potential anti-HIV-1 activities. Weaker anti-HIV activities were qualitatively predicted for **126** which showed a novel alternative binding mode different from either the initial reference compound L-737,126 or **103** and **115**

This different binding mode led to hypothesize loss of important favourable interactions. The other seven designed IASs displayed binding modes comparable to that of **103** that allowed easier qualitative considerations. Derivative **124** was expected to be less active than **103** due to the unsubstituted 3-phenyl ring[120], even if the oxazolidinone ring of this compound fulfilled the hydrophobic pocket (*Figure 25A*).

Table 11. Structure, predicted and experimental activity

A   B
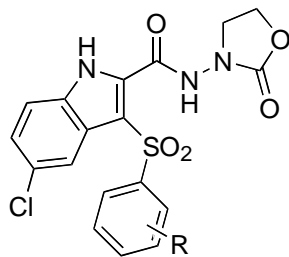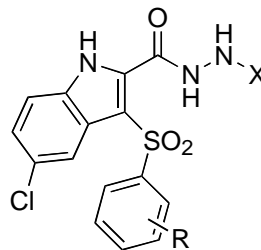
| | | | | pEC50 | |
Compd	Structure	R	X	Pred.	Exp.
124	A	H	-	7.446	7.824
125	A	3,5-Me$_2$	-	7.442	9.046
126	B	3,5-Me$_2$	Me	6.540	7.699
127	B	3,5-Me$_2$	*iso*-Pr	8.455	9.155
128	B	3,5-Me$_2$	*cyclo*-He	6.578	7.301
129	B	3,5-Me$_2$	COMe	8.139	8.301
130	B	3,5-Me$_2$	COOEt	6.234	7.699
131	B	3,5-Me$_2$	CONHNH$_2$	7.134	8.000

Table 11. Structure, predicted and experimental activity (pEC$_{50}$) of designed compounds **124-131.**

The corresponding 3,5-dimethyl derivative **125** assumed almost the same binding mode of **124**, and according previous SARs, it was expected to be more active than **124**. Designed derivatives **127**, **129** and **131** (isopropyl, acetyl and hydrazinocarbonyl groups, respectively) optimally fulfilled the pocket formed by Val106, Pro225, Phe227, Leu234 and Pro236 residues. Consequently these compounds were expected to be highly active.

Regarding compounds **128** and **130**, some unfavourable steric clashes were observed. The cyclohexyl or carboxyethyl residue (not shown) forced of the whole molecule to shift, and

caused the disruptions of favourable interactions, such as the π−π stacking interaction of the 3,5-dimethylbenzene with Tyr181 and Tyr188 and the hydrogen bond between the indole NH and Lys101.

The qualitative prediction of **124-131**'s activity by SBDD was successfully confirmed by the application of the second IAS 3-D QSAR model. Interestingly, the average predicted anti-HIV-1 activity was about 60 nM and pressed their synthesis and biological evaluation. As qualitative (SBDD) and quantitative (LBDD) predicted, the designed new IASs revealed to be effectively highly active against HIV-1 acutely infected MT4 cells (*Table 11*). Again, the new derivatives were found more active than those predicted with an average pEC_{50} of 8.12 (EC_{50} about 10 nM), and interestingly two compounds (**125** and **127**) showed sub-nanomolar anti-HIV-1 activity (*Table 11*). In conclusion the new and updated 3-D QSAR model was fully successful. While expanding the IAS chemical diversity, the model increased its prediction ability and proved to be an efficient tool to discover new more potent and selective IASs.
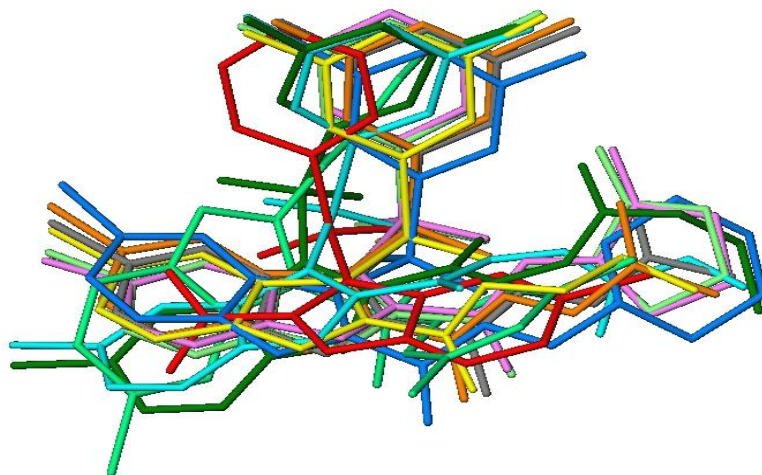
Figure 26. Docked conformations of the designed IASs **124-131** (**124**: purple; **125**: light green; **126**: green; **127**: orange; **128**: blue; **129**: grey; **130**: dark green; **131**: cyan). For comparison purposes the conformation of reference compounds con **103** (red) e **115** (yellow) are also reported.

6.5 The third 3-D QSAR model on HIV WT IASs and the first models on Y181C and K103N-Y181C HIV strains

By the means of classical and computational medicinal chemistry techniques, our research

group made extensive studies on sulphone containing NNRTIs that led to the discovery of very potent anti-HIV IASs (*Chart 13*)
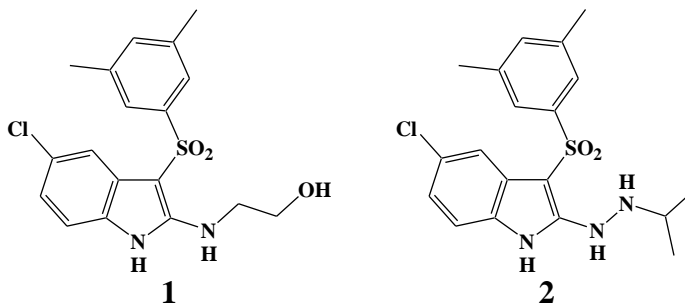
	1	**2**
WT	0.0008	0.0007
Y181C	0.009	0.04
K103N-Y181C	0.3	8.6

Chart 13. The most potent published IASs. The EC_{50} is reported in μM against the WT, Y181C and K103N-Y181C HIV strains..
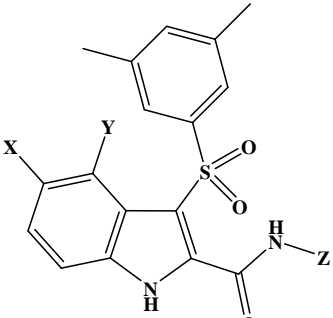
Continuing our research on new anti-HIV agents and making a strong use of the in house biological data, we have developed new 3-D QSAR models specific for the wild type and two mutated RT isoforms (Y181C and K103N-Y181C). The new models revealed highly predictive and supported the synthesis of new unreported IASs derivatives characterized by the presence of various substitution patterns either at the indole nucleus or at the side 2 carboxyl chain (compounds **3-10**, *Table 12*). Experimental data, in agreement with the computational predictions, revealed the new IASs **3-10** to be very potent against both wild type and drug resistant mutants, and disclosed new lead compounds as starting point for the

development of potential more potent anti-HIV IAS generations.

6.5.1 3-D QSAR studies open strategies

Continuing from the previous reported studies[148,143] a more robust and predictive 3-D QSAR model specific for the HIV-1 wild type (3-D $QSAR_{wt}$) strain was developed. In parallel, two further 3-D QSAR models were also built for the Y181C mutant (3-D $QSAR_{Y181C}$) and K103N-Y181C double mutant (3-D $QSAR_{K103N-Y181C}$) strains, respectively.

Table 12. Newly Designed IASs Derivatives.

Comp	X	Y	Z
3	NO_2	-	$CH_2CH_2CH_2SO_2NH_2$
4	Cl	-	$CH_2(CH_3)CONHNH_2$
5	Cl		$CH_2(CH_3)CONH_2$
6	F	F	H
7	Cl	-	$CH_2(CH_2OH)CONHNH_2$

8	Cl	-	$CH_2(CH_2OH)CONH_2$
9	Cl	-	CH_2CH_2-1midazol-1-yl
10	NO_2	-	$CH_2(CH_2OH)CONH_2$

6.5.2 Training Sets Composition and Alignment Rules

For the 3-D QSAR$_{wt}$ model a training set (TrSet) of 117 IASs derivatives was compiled by adding to the last previous training set (101 derivatives) the test (8 hybrid IAS-peptide derivatives) and prediction (8 IASs characterized by the presence of pyrrolidyn-2-one nucleus at the indole-2-carboxamide or bearing some substituents at the indole-2-carbohydrazide) data sets (*Table 13*). For the 3-D QSAR$_{Y181C}$ and 3-D QSAR$_{K103N-Y181C}$ models the same TrSet was used *(Table 13)* comprising all IAS derivatives (40 for both models) for which biological anti-HIV activities against the single (Y181C) and the double (K103N-Y181C) mutants strains were available**.**

The training sets were structure-based (SB) aligned by means of docking experiments. The wild type TrSet was docked into the NNBS extracted from the 739W94/RT complex; the Y181C TrSet was docked into the NNBS extracted from the Efavirenz/RT$_{Y181C}$ complex[144] (pdb entry code 1jkh) and finally for the K103N-Y181C TrSet SB alignment a modeled RT double mutant was used.

The good results obtained with old models were for us the better confirm to continue to use program GOLPE[92] using as dependent variable the inverse of the logarithm of the experimental anti-HIV-1 activities (pEC$_{50}$s) and as independent variables the molecular fields generated by the means of the GRID[93] software using the water (OH2) probe.

Table 13. Structures, WT, Y181C and K103N-Y181C Anti-HIV-1 µM Activities of Training Set Compounds Used for the 3-D QSARs.
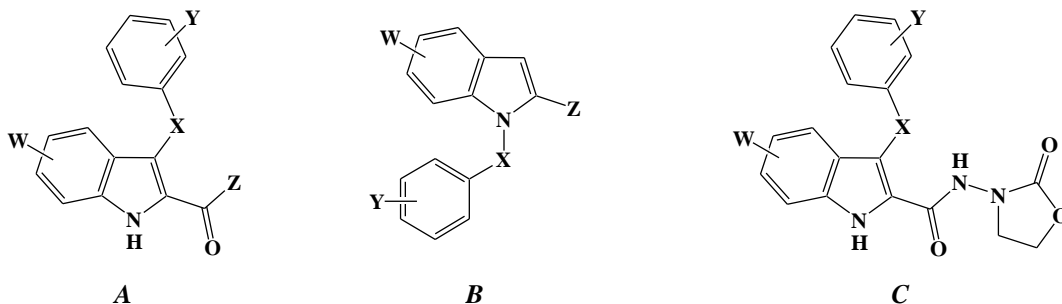
								EC$_{50}$ (µM)		
Comp.		X	Y	Z		W		WT	Y181C	K103N Y181C
11	A	SO$_2$	H	NH$_2$		Cl		0.001	0.02	8
20	A	S	H	OEt		H		1.4	-	-
21	A	S	2-NH$_2$	OEt		H		>200	-	-
22	A	S	2-NH$_2$-5-Cl	OEt		H		≥200	-	-
23	A	S	2-NH$_2$	OEt		Cl		2.3	-	-

Comp.	X	Y	Z	W	EC$_{50}$ (μM)		
					WT	*Y181C*	*K103N Y181C*
24	A	S	2-NH$_2$-5-Cl	OEt	Cl	2.5	-	-
25	A	SO$_2$	H	OEt	H	3.7	-	-
26	A	SO$_2$	2-NH$_2$	OEt	Cl	>200	-	-
27	A	SO$_2$	2-NH$_2$-5-Cl	OEt	H	2.5	-	-
28	A	SO$_2$	2-NH$_2$-5-Cl	OEt	Cl	1.9	-	-
29	A	S	H	NH$_2$	H	1.4	-	-
30	A	S	2-NH$_2$-5-Cl	NH$_2$	H	9	-	-
31	A	S	H	NH$_2$	Cl	0.02	4	>9
32	A	S	2-Me	NH$_2$	Cl	0.3	-	-
33	A	S	4-Me	NH$_2$	Cl	>0.45	-	-
34	A	S	4-F	NH$_2$	Cl	1.4	-	-
35	A	S	4-Cl	NH$_2$	Cl	3.1	-	-
36	A	S	4-iso-Pr	NH$_2$	Cl	1.9	-	-
37	A	S	4-tert-Bu	NH$_2$	Cl	8	-	-
38	A	S	3,5-Me$_2$	NH$_2$	Cl	0.006	>0.7	>0.7
39	A	S	2,6-Cl$_2$	NH$_2$	Cl	1.2	-	-
40	A	S	2-NH$_2$-5-Cl	NH$_2$	Cl	1.6	-	-
41	A	SO$_2$	H	NH$_2$	H	0.18	-	-
42	A	SO$_2$	2-NH$_2$-5-Cl	NH$_2$	H	0.3	-	-
19a	A	SO$_2$	2-Me	NH$_2$	Cl	0.001	0.16	10
19b	A	SO$_2$	3-Me	NH$_2$	Cl	0.001	0.006	7
19c	A	SO$_2$	4-Me	NH$_2$	Cl	0.003	0.02	>100
43	A	SO$_2$	4-F	NH$_2$	Cl	0.014	-	-
44	A	SO$_2$	4-Cl	NH$_2$	Cl	0.011	-	-
45	A	SO$_2$	4-iso-Pr	NH$_2$	Cl	0.08	-	-
46	A	SO$_2$	4-tert-Bu	NH$_2$	Cl	0.13	-	-
Com	X	Y	Z	W	EC$_{50}$ (μM)		

Comp.		X	Y	Z	W	WT	Y181C	K103N Y181C
19d	A	SO$_2$	2,4-Me$_2$	NH$_2$	Cl	0.004	0.15	10
19e	A	SO$_2$	3,5-Me$_2$	NH$_2$	Cl	0.004	0.03	0.6
47	A	SO$_2$	2,6-Cl$_2$	NH$_2$	Cl	0.1	-	-
48	A	SO$_2$	2-NH$_2$-5-Cl	NH$_2$	Cl	0.04	-	-
49	A	SO$_2$	3,5-Me$_2$	NH$_2$	Br	0.002	0.015	0.55
50	A	SO$_2$	3,5-Me$_2$	NH$_2$	COMe	0.015	0.18	>100
51	A	SO$_2$	3,5-Me$_2$	NH$_2$	CH(OH)Me	0.025	0.05	>100
52	A	S	H	NHNH$_2$	Cl	0.55	-	-
53	A	S	4-Me	NHNH$_2$	Cl	1.5	-	-
54	A	S	4-F	NHNH$_2$	Cl	5	-	-
55	A	S	4-Cl	NHNH$_2$	Cl	10	-	-
56	A	S	2-NH$_2$-5-Cl	NHNH$_2$	Cl	>13	-	-
57	A	SO$_2$	H	NHNH$_2$	H	0.53	-	-
58	A	SO$_2$	H	NHNH$_2$	Cl	0.01	0.4	>100
59	A	SO$_2$	4-Me	NHNH$_2$	Cl	0.05	-	-
60	A	SO$_2$	4-F	NHNH$_2$	Cl	0.32	-	-
61	A	SO$_2$	4-Cl	NHNH$_2$	Cl	0.19	-	-
62	A	SO$_2$	3,5-Me$_2$	NHNH$_2$	Cl	0.13	2.5	>10
63	A	SO$_2$	2-NH$_2$-5-Cl	NHNH$_2$	Cl	0.3	-	-
64	B	SO$_2$	H	H	H	>150	-	-
65	B	SO$_2$	4-Cl	COOEt	H	>100	-	-
66	B	SO$_2$	2-NO$_2$	COOEt	H	1.8	-	-
67	B	SO$_2$	2-NO$_2$-5-Cl	COOEt	H	>100	-	-
68	B	SO$_2$	2-NH$_2$-5-Cl	COOEt	H	1.8	-	-

Comp.		X	Y	Z	W	EC$_{50}$ (μM)		
						WT	Y181C	K103N Y181C
69	B	SO$_2$	H	H	3-COOEt	>32	-	-
70	B	SO$_2$	H	H	3-COO-i-Pr	>54	-	-

Comp.		X	Y	Z	W	EC₅₀ (μM) WT	Y181C	K103N Y181C
71	B	SO$_2$	H	H	5-Cl	>75	-	-
72	B	SO$_2$	H	COOEt	5-Cl	>200	-	-
73	B	SO$_2$	4-Me	COOEt	5-Cl	>200	-	-
74	B	SO$_2$	4-Cl	COOEt	5-Cl	>200	-	-
75	B	SO$_2$	2-NO$_2$-5-Cl	COOEt	5-Cl	>31	-	-
76	B	SO$_2$	2-NH2-5-Cl	COOEt	5-Cl	8.3	-	-
77	B	SO$_2$	H	H	5-Cl-3-COOEt	>42	-	-
78	B	SO$_2$	H	H	5-Cl-3-COO-iPr	>200	-	-
79	B	SO$_2$	H	CONH$_2$	H	15	-	-
80	B	SO$_2$	H	CONHNH$_2$	H	>200	-	-
81	B	SO$_2$	H	CONH$_2$	5-Cl	66.6	-	-
82	B	SO$_2$	H	H	3-CONH$_2$	>32	-	-
83	B	SO$_2$	H	H	5-Cl-CONH$_2$	≥200	-	-
132	B	-	2-NO$_2$	-	-	5.30	-	-
133	B	-	2-NH$_2$	-	-	4.96	-	-
134	B	-	2-NO$_2$,5-Cl	-	-	5.40	-	-
135	B	-	2-NH$_2$, 5-Cl	-	-	6.00	-	-
136	B	-	2-NO$_2$, 4-Cl	-	-	4.80	-	-
137	B	-	2-NH$_2$, 4-Cl	-	-	4.13	-	-
138	B	-	2-Cl, 5-NO$_2$	-	-	5.22	-	-
139	B	-	2-Cl, 5-NH$_2$	-	-	3.82	-	-
92	A	S	H	NHCH$_2$CH$_2$OH	5-Cl	0.04	-	-

Comp.		X	Y	Z	W	EC₅₀ (μM) WT	Y181C	K103N Y181C
93	A	S	2-Me	NHCH$_2$CH$_2$OH	5-Cl	0.05	-	-
94	A	S	3-Me	NHCH$_2$CH$_2$OH	5-Cl	0.08	-	-
95	A	S	4-Me	NHCH$_2$CH$_2$OH	5-Cl	0.12	-	-
96	A	S	2,3-Me$_2$	NHCH$_2$CH$_2$OH	5-Cl	0.10	-	-
97	A	S	3,5-Me$_2$	NHCH$_2$CH$_2$OH	5-Cl	0.012	0.6	1.2
98	A	SO$_2$	H	NHCH$_2$CH$_2$OH	5-Cl	0.001	0.07	4.2

Comp.	X	Y	Z	W	WT	Y181C	K103N Y181C
99	A	SO_2	2-Me	$NHCH_2CH_2OH$	5-Cl	0.031	1.2	12
100	A	SO_2	3-Me	$NHCH_2CH_2OH$	5-Cl	0.003	0.08	1.1
101	A	SO_2	4-Me	$NHCH_2CH_2OH$	5-Cl	0.043	-	-
102	A	SO_2	2,4-Me_2	$NHCH_2CH_2OH$	5-Cl	0.008	-	-
103	A	SO_2	3,5-Me_2	$NHCH_2CH_2OH$	5-Cl	0.0008	0.009	0.3
104	A	S	H	$NHNHCH_2CH_2OH$	5-Cl	0.2	-	-
105	A	S	2-Me	$NHNHCH_2CH_2OH$	5-Cl	0.1	-	-
106	A	S	3-Me	$NHNHCH_2CH_2OH$	5-Cl	0.17	-	-
107	A	S	4-Me	$NHNHCH_2CH_2OH$	5-Cl	0.6	-	-
108	A	S	2,4-Me_2	$NHNHCH_2CH_2OH$	5-Cl	1.2	-	-
109	A	S	3,5-Me_2	$NHNHCH_2CH_2OH$	5-Cl	0.03	1.0	6.7
110	A	SO_2	H	$NHNHCH_2CH_2OH$	5-Cl	0.01	1.2	>20
111	A	SO_2	2-Me	$NHNHCH_2CH_2OH$	5-Cl	0.05	0.1	>20
112	A	SO_2	3-Me	$NHNHCH_2CH_2OH$	5-Cl	0.007	0.2	20
113	A	SO_2	4-Me	$NHNHCH_2CH_2OH$	5-Cl	0.04	1	>20
114	A	SO_2	2,4-Me_2	$NHNHCH_2CH_2OH$	5-Cl	0.08	0.04	3.2
115	A	SO_2	3,5-Me_2	$NHNHCH_2CH_2OH$	5-Cl	0.001	0.05	0.8
116	A	SO_2	3,5-Me_2	$Gly-NH_2$	5-Cl	0.006	0.03	0.8
Comp.	X	Y	Z	W	**EC_{50} (µM)**		
					WT	*Y181C*	*K103N Y181C*
117	A	SO_2	3,5-Me_2	$Gly-NHNH_2$	5-Cl	0.01	0.05	2
118	A	SO_2	3,5-Me_2	$GlyGly-NH_2$	5-Cl	0.0007	0.005	1.2
119	A	SO_2	3,5-Me_2	$GlyGly-NHNH_2$	5-Cl	0.06	0.08	2.8
120	A	SO_2	3,5-Me_2	$GlyAla-NH_2$	5-Cl	0.014	0.03	2.2
121	A	SO_2	3,5-Me_2	$GlyAla-NHNH_2$	5-Cl	0.08	0.16	3.55
122	A	SO_2	3,5-Me_2	$GlyGlyGly-NH_2$	5-Cl	0.12	0.23	5.9
123	A	SO_2	3,5-Me_2	$GlyGlyGly-NHNH_2$	5-Cl	0.18	0.33	>11.1
124	C	SO_2	H		5-Cl	0.015	0.3	20
125	C	SO_2	3,5-Me_2		5-Cl	0.001	0.01	0.9
126	A	SO_2	3,5-Me_2	$NHNHCH_3$	5-Cl	0.05	0.13	>20
127	A	SO_2	3,5-Me_2	NHNH-i-Pr	5-Cl	0.0007	0.04	8.6

128	A	SO_2	$3,5\text{-Me}_2$	NHNH-c-He	5-Cl	0.1	1.2	>20
129	A	SO_2	$3,5\text{-Me}_2$	$NHNHCONHNH_2$	5-Cl	0.04	0.02	5.7
130	A	SO_2	$3,5\text{-Me}_2$	$NHNHCOCH_3$	5-Cl	0.005	0.02	>20
131	A	SO_2	$3,5\text{-Me}_2$	NHNHCOOEt	5-Cl	0.02	0.04	9.2

6.5.3 3-D QSAR$_{wt}$ model

A final robust 3-D QSAR model was achieved with only 3 principal components (PCs) characterized by r^2 value of 0.90 and leave one out (LOO), five group (L5O) and leave half out (LHO) q^2 values (q^2_{LOO}, q^2_{L5O} and q^2_{LHO}) corresponding to 0.83, 0.81 and 0.76 and LOO L5O and LHO SDEP (SDEP$_{LOO}$ SDEP$_{L5O}$ SDEP$_{LHO}$) values of 0.70, 0.73 and 0.82, respectively (*Table 14*). Compared to the 3-D QSAR model reported on 2006[148] an improvement of the statistical coefficient was obtained, such an improvement is even greater if one compare the actual model statistical coefficients with those of the very first model reported on 2005[137] (*Table 14*)

Table 14. Statistical coefficients of the actual 3-D QSAR model compared to those of the previously reported models.

	Vars	PCs	r^2	LOO		L5O		LHO	
				q^2	SDEP	q^2	SDEP	q^2	SDEP
Actual Model	1024	3	0.90	0.83	0.70	0.81	0.73	0.76	0.82
JMC 2005 Model[164]	669	2	0.85	0.69	0.94	0.68	0.95	0.64	1.00
JMC 2006 Model[187]	802	3	0.90	0.79	0.75	0.78	0.77	0.73	0.86

One important feature of 3-D QSAR studies is the graphical representation of the models, which allows making its easier interpretation. In the actual model, the variable number to get the optimal 3-D QSAR was greater than in the precedent two models and this features (*Table 14*) allowed to display more defined PLS polyhedrons leading to an even easier interpretation of the PLS maps that the program GOLPE generates. Similarly as previously reported [148,137] the PLS pseudo-coefficients and the activity contribution plots were used for the model interpretation.
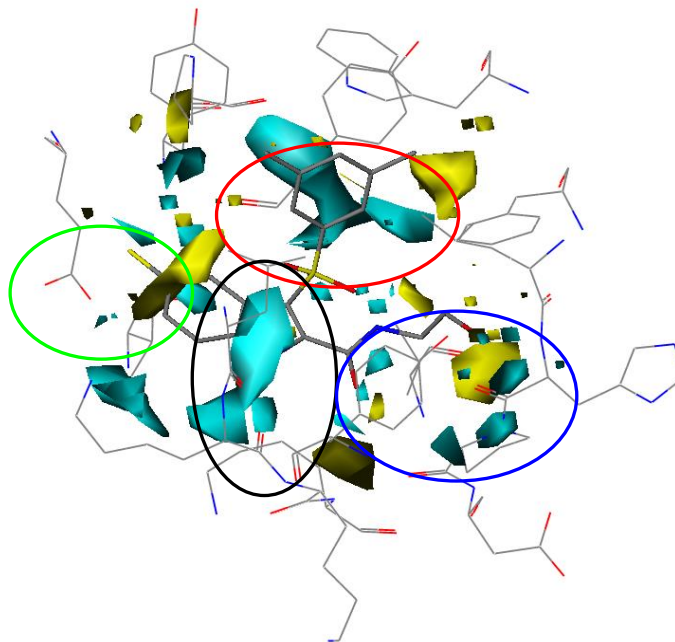
Figure 27.Two rotated views (top and bottom) of PLS coefficients (Polyhedrons in yellow are negative fields (data level = -0.008), polyhedrons in cyan are positive fields (data level = +0.008). The high active IAS representative X (displayed as by atom type

colour) is also reported. A 4Å core of the NNBS is also displayed.

Derivative **103** one of the most active IASs, was used as reference compound to interpret the 3-D QSAR maps. The analysis of the PLS coefficients and the activity contribution plots lead the polyhedrons to be reduced into four key groups (*Figure 27*):

1. a group around the dimethylphenyl portion (red circles in *Figure 27*);
2. a group around the chlorine atom in position 5 of indole ring (green circles in *Figure27*);
3. a group around N1 of indole nucleus(black circles in *Figure 27*); likely this field could be referred to the H-bond with Lys101carbonyl oxygen (*Figure 29*).
4. a group around the side chain attached in indole position 2 (blue circles in *Figure 27*).

This analysis is similar to that previously reported[148] in which only three main polyhedron groups were identified. Nevertheless, the greater number of variables led to better defined maps relative to the 5-Cl and N1 atoms that could clearly be divided in two groups (green and black circles *in Figure 27*).

Respect to the previous model (JMC 2006 Model)[148] the main polyhedrons shape and dimension differences were recorded around the carbonyl and hydroxyl oxygens of indole side chains and closely to the 5-Cl atom (*Figure 28B*). Recalling that the TrSet was SB aligned the use of the NNBS helped the correlation of the 3-D QSAR results with the potential ligand/receptor interactions. The role of the electrostatic interaction described[109] between the terminal **103** hydroxyl and the conserved Leu234[137]main chain carbonyl was also

observed for the very most active IAS **118** (EC$_{50}$ 0.0007 μM, *Table 13 and Figure 27 and 29*).
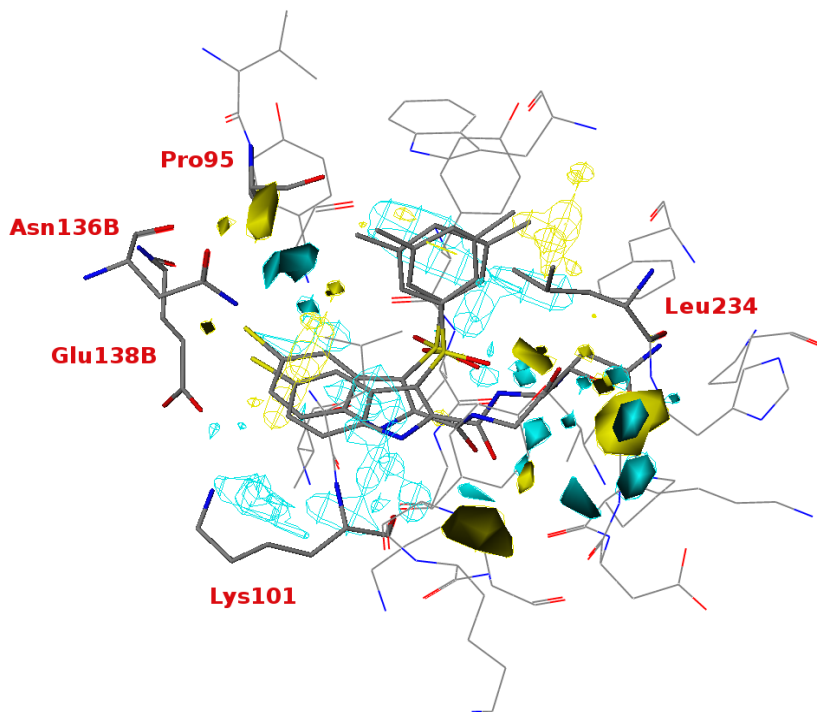
Figure 28. **119** and **103** PLS coefficient plots. Polyhedrons (contours) in yellow are negative fields (data level = -0.008), polyhedrons in cyan are positive fields (data level = +0.008). In surface style are highlighted the fields relative to indole side chains carbonyl and hydroxyl oxygen and relative to chlorine atom.

Interestingly the **118** hydrazide analogue **119** showed a different binding mode (BM) (*Figure 29*). This **119** BM adjustment could be ascribed to the length of the side chain and the

presence of a "wall" formed by Leu234, His235, Pro225 and Phe227 residues. In **119** the further nitrogen atom makes the side chain too long to be accommodated in the same space of **118** glycine moiety thus forcing the molecule to an alternative BM, that while keeping the dimethylphenyl moiety supeimposable to that of **119** and the hydrogen bond between the indole NH with the Lys101 carbonyl oxygen, disrupt a strong hydrogen bond (*Figure 29*), leading to an activity decrease (**119** EC_{50} = 0.06 μM) correctly incorporated in the 3-D QSAR model. Regarding the above further two PLS coefficients differences (see areas around 5-Cl and 2-C=O in *Figures 29 and Figure 30*.) between the actual and previous model a detailed inspection did not identify any relevant interaction. For the **118** carbonyl oxygen's no hydrogen bonds were visible with any residue around a distance of 4Å. The same space is occupied by the **119** chlorine atom, so that the interactions are likely steric in nature. Similar conclusion can be drawn out for the type of interaction occurring in the Pro95, Glu138B and Asn136B residues region. In this area either the **118** 5-Cl or the **119** terminal portion of the C2 side chain of indole are not involved in any relevant interaction.
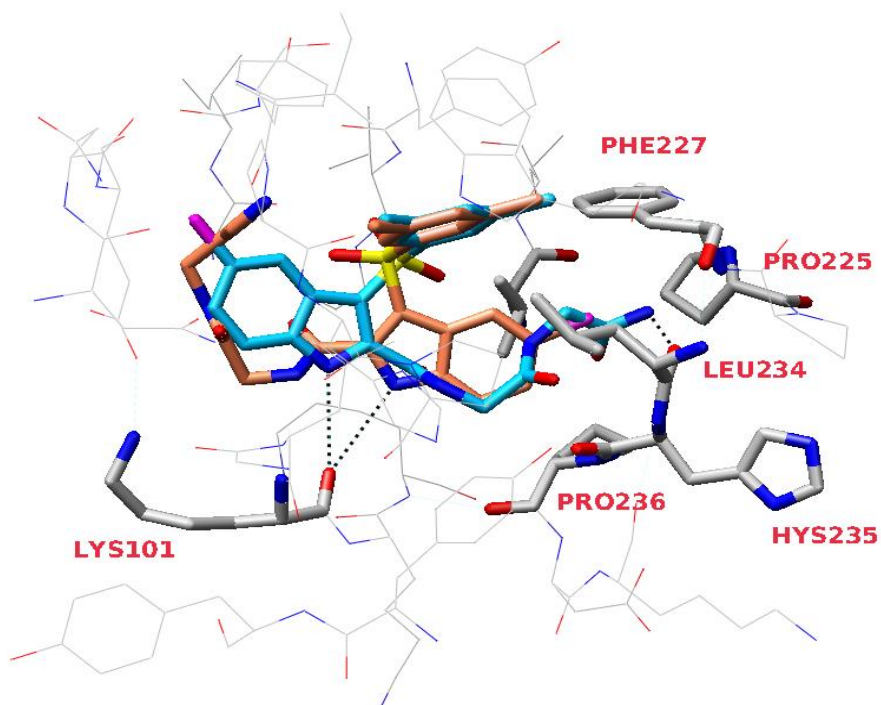
Figure 29. **118** (carbon atoms in cyan) and **119** (carbon atoms in orange) docked conformations. A 5Å core of the 1jlq NNBS is also reported in wire style. "Wall" residues (see text) are reported in stick. Potential H-bond are reported in black dashed lines ($NH_{RS1919} \cdots O=C_{Lys101} = 3.3$ Å; $CONH2_{RS1919} \cdots O=C_{Leu234} = 2.3$ Å; $NH_{RS2068} \cdots O=C_{Lys101} = 3.3$ Å.
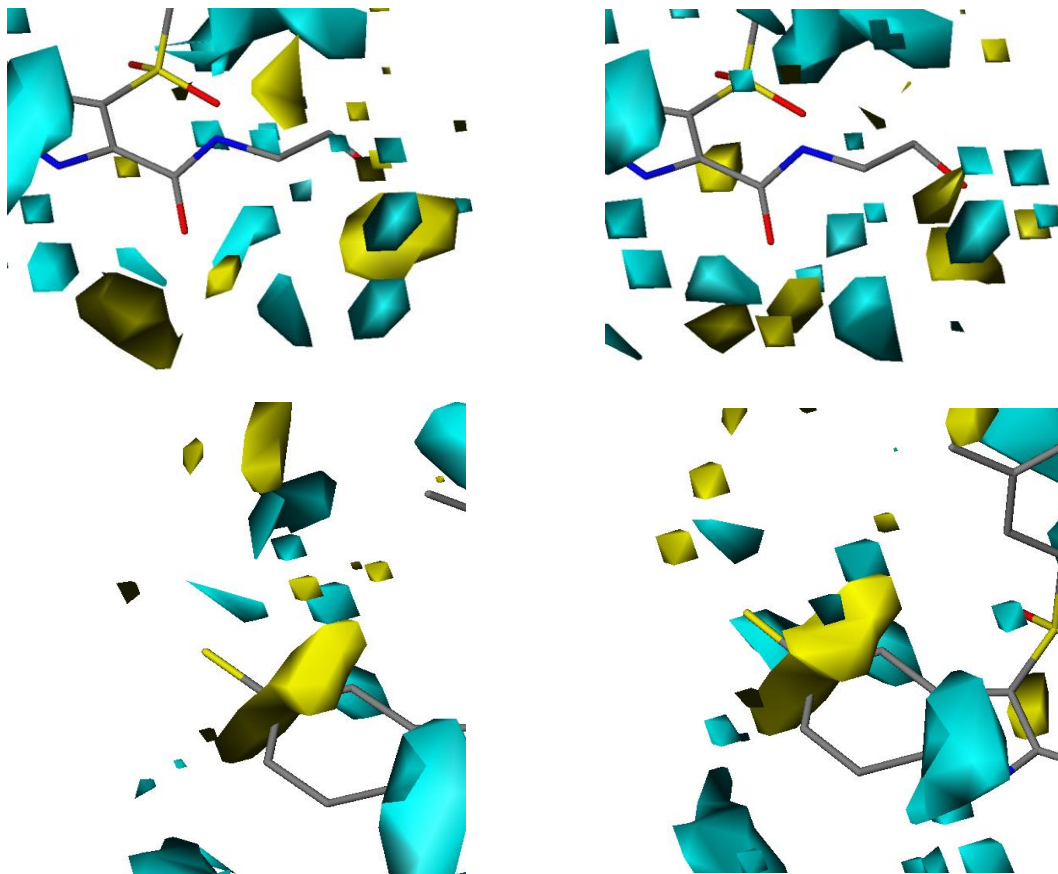
Figure 30. Details on the differences between the previous (right column) and the actual models (left column) PLS coefficients plots. The RS1675 (**1**) structure is displayed for interpretation.

6.5.4 3-D QSAR$_{Y181C}$ model characteristic and interpretation

The same procedure used to build the 3-D QSAR$_{wt}$ model was adopted. Optimal variables

selection was achieved after only three FFDs, leading to 602 active variables. The final model was characterized by high-quality statistical indexes values with only 3 PCs ($r^2 = 0.94$, $q^2_{L5O} = 0.68$, $SDEP_{L5O} = 0.42$; *Figure 31*). Additional FFD selections led to slight statistical improvements, nevertheless a much reduced containing variables model often prove to be useless in predicting the dependent variable (biological activity)

For the 3-D QSAR$_{Y181C}$ model analysis and interpretation the PLS pseudo coefficients (*Figure 31*), activity contribution (*Figure 32*) and present field (*Figure 33*) plots relative to **19e** (EC$_{50-Y181C}$ 0.006 μM), **19c** (EC$_{50-Y181C}$ 0.03 μM), **19d** (EC$_{50-Y181C}$ 0.02μM) and **19a** (EC$_{50-Y181C}$ 0.16μM) were used. Differently from the well characterized WT model, in this one the use of a series of compounds differently substituted on the phenyl ring is of some utility in highlighting dissimilar polyhedrons spatial arrangements that correlate with the role of mutation Tyr181 to Cys181 and hence with the activity differences against the RT$_{Y181C}$ variant. Comparing the **19a**, **19e**, **19d** and **19c** structure based aligned conformations (*Figure 31*) the inhibitor binding modes seem highly dependent from both the Tyr181 → Cys181 mutation and the position of the methyl group on the phenyl ring.

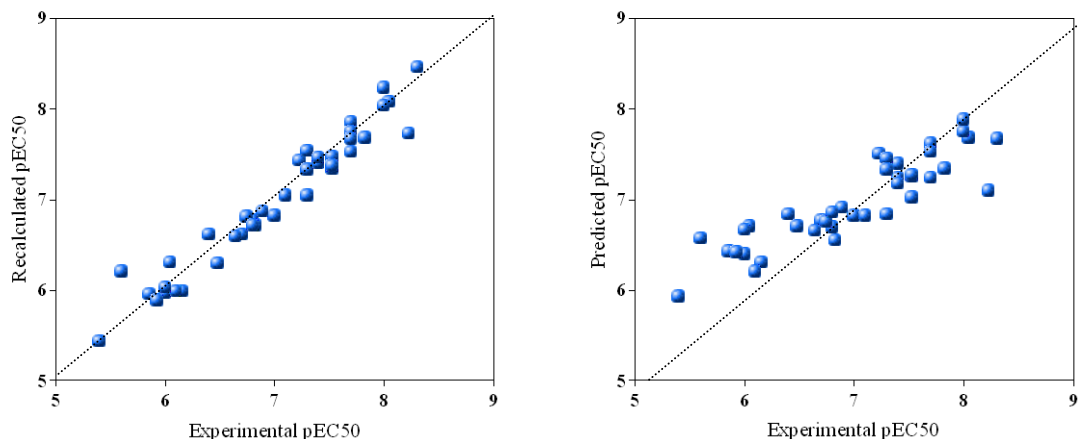|	*Vars*	**PCs**	r^2	**L5O**	
				q^2	**SDEP**
3-D QSARY181C	602	3	0,94	0,68	0,42

Figure 31. Fitting (left side) and Cross-validation (right side) Plots of the Y181C Model with 3 PCs.

The Y181C NNBS cavity is therefore able to discriminate the different phenyl substitutions and this was correctly identified by the 3-D QSAR$_{Y181C}$ model. Differently as observed for the WT model the PLS coefficients polyhedrons cannot be easily divided into groups, in fact the structure based alignment were rather different and not comparable each to other and, as a consequence a different PLS polyhedron distribution was observed. At the ±0.0045 data level the displayed PLS coefficients are mainly negative (cyan polyhedrons, *Figure 31*) and only three main small yellow maps groups (positive coefficients) are present. Two of the latter's (around the methyl's groups on the **19a**, **19e**, **19d** and **19c** benzene rings and close to the indole moiety) are associated to positive values of the OH2 GRID fields and

could explain some sterical repulsive interactions between the IASs methyl substituents and the Pro95, Cys181 and Trp229 residues and the C6-C7 indole carbon atoms with the Lys101 methylene side chain (*Figure 31*).
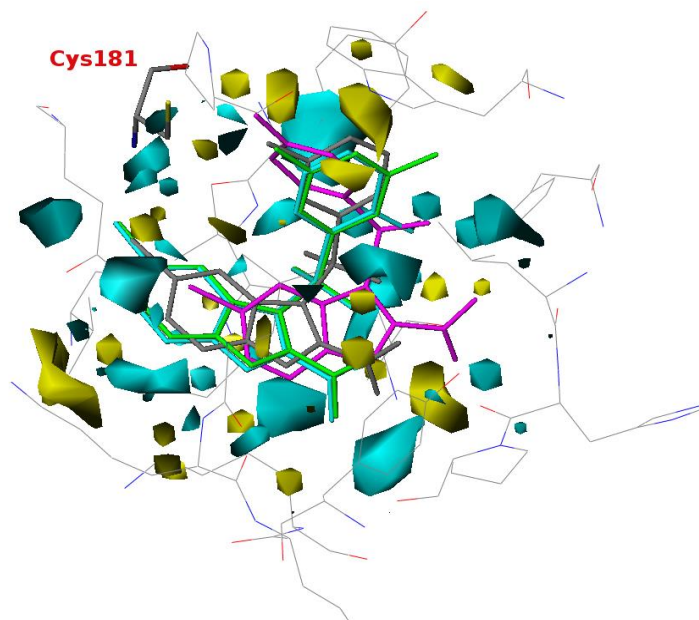

Figure 31. 3-D QSAR$_{Y181C}$ PLS coefficients plot. **19a** (cyan), **19e** (grey), **19d** (magenta) and **19c** (green) docked conformations into RT$_{Y181C}$ are displayed. Polyhedrons in yellow are negative fields (data level = -0.0045), polyhedrons in cyan are positive fields (data level = +0.0045). A 4Å core of the Y181C NNBS is also displayed for easer interpretation.

On the other hand, the third yellow coloured maps group, associated to negative GRID

field values, situated in the proximity of the 2-carboxyamide groups, are involved in hydrogen bonding interactions with His235 and/or Pro236 main chain carbonyl oxygen's and/or Tyr318 hydroxyl (electrostatic attractive interactions). Others small yellow polyhedron are also visible in *Figure 31*, but any trials to interpret them was useless. Regarding the cyan polyhedrons in *Figure 31* (negative PLS coefficients), the most important are a couple around the area that correspond to the two amidic hydrogens. In fact, these two polyhedrons are characterized by lower values of PLS coefficients that modulate the corresponding GRID fields; depending on the signs of the latter's a positive (negative GRID fields) or negative (positive GRID fields) contribution to the activity is associated to the above areas (*Figure 32*). In this scenario for the four representative IASs **19a**, **19e**, **19d** and **19c** as displayed in *Figures 31, 32, 33 and 34*, the methyl substituents seem to play a crucial role.

   In particular from *Figure 31*, taking as reference the docked structure the most active IAS **19e,** which bear a methyl in position *meta*, it is possible to observe that a second *meta* methyl group or a single *ortho* methyl substitution are responsible of the inhibitor molecule shift (to the left in *Figure 9*) likely caused by the steric clashes with either Tyr188 and Phe227 (**19a**) or Trp229 (**19c**) side chains.

      A more severe influences is exerted by the *meta* → *para* modification (**19e** → **19d**); the **19d** *para*-methylphenyl group being less bulky than the **19e** *meta*-methylphenyl can enter in a NNBS small sub-pocked formed by Pro95, Cys181 and Trp229 residues with an overall binding mode that is the result of even greater molecular shift (to the right of *Figure 31*).
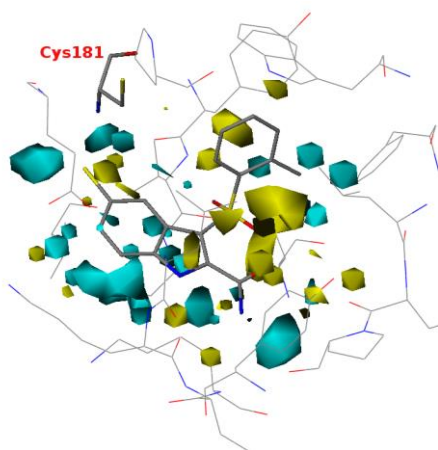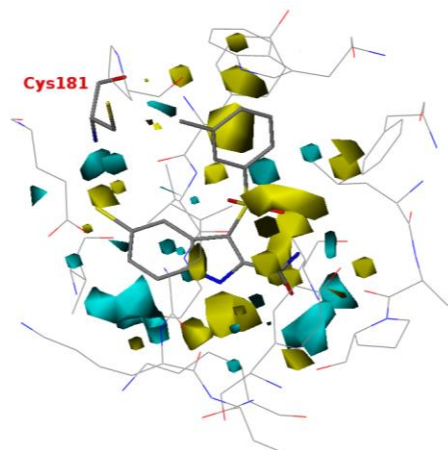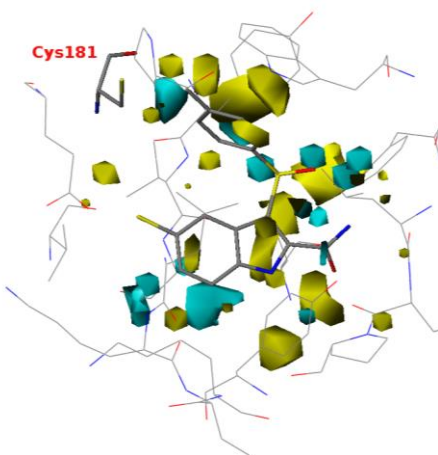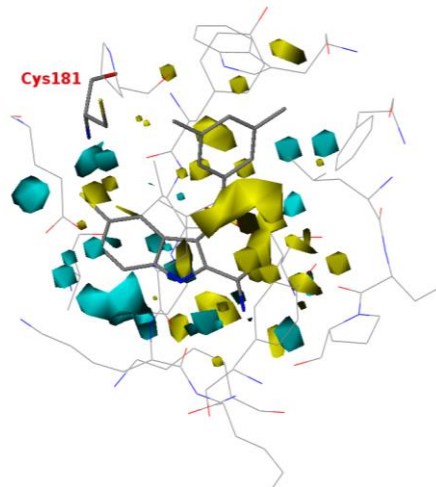
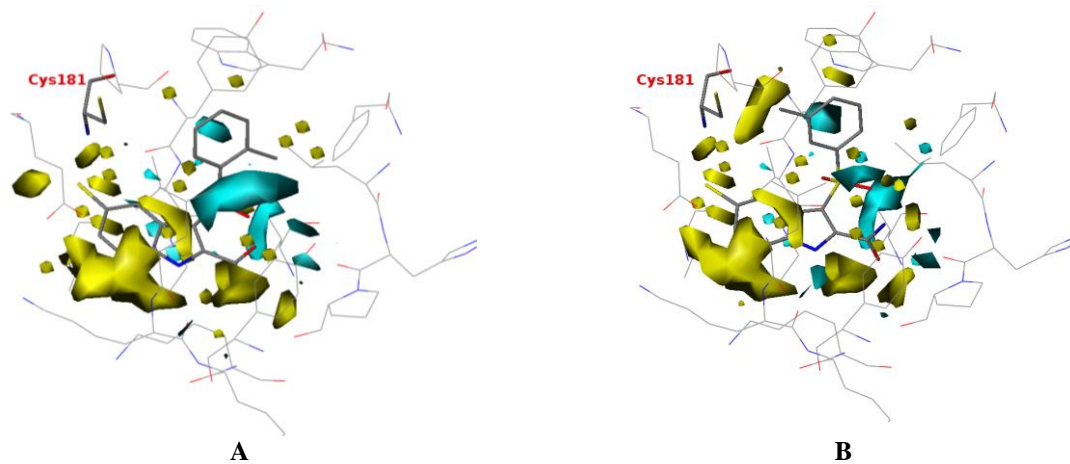**A**   **B**

Figure 32 To be continued

C   D

Figure 32. Activity contribution plots relative to **19a** (A), **19e** (B), **19d** (C) and **19c** (D). A 4Å core of the Y181C NNBS is also displayed for interpretation.
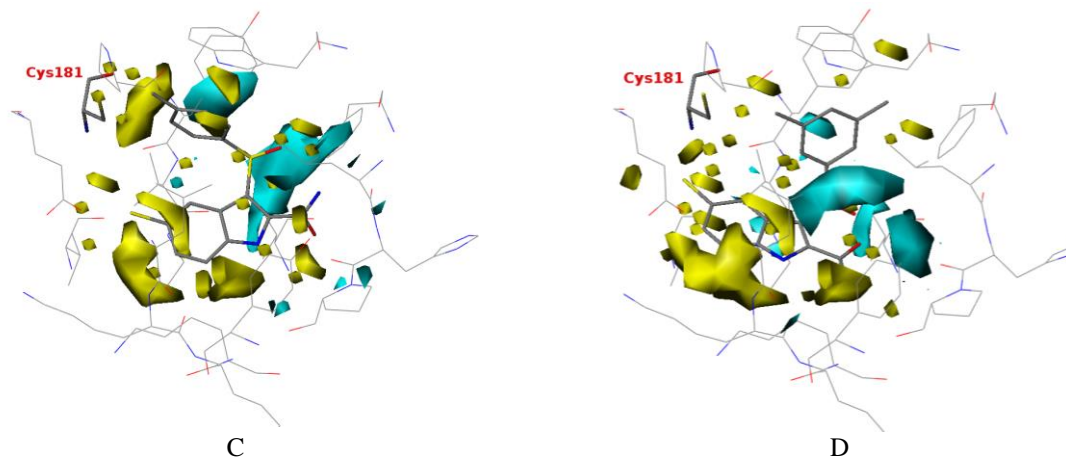
C   D

Figure 33: Present field plots relative to **35**(A), **36** (B), **37** (C) and **43** (D). Polyhedrons are associated to OH2 GRID field values. Sign is positive in yellow polyhedrons, while are negative in cyan polyhedrons.

   The **19a**, **19e**, **19d** and **19c** different binding modes, once encoded into the 3-D QSAR model (alignment rules) were quantitatively described and diverse activity contribution plots have been obtained for each compound (*Figure 32*), thus reflecting the experimental biological activity values. A deeper inspections of the activity contribution plots in *Figure 32* led to point the attention around the carboxyamide moiety, were the maximum polyhedron location and dimension variability can be observed consistently with the above PLS coefficients plot interpretation. In fact, while for **19e** two hydrogen bonding distances are present between the amide nitrogen and HIS235 main chain carbonyl oxygen and Tyr318 hydroxyl group ($N_{RS1720}{\cdots}O{=}C_{His235}$ = 3.0 Å; $NH_{2RS1720}{\cdots}O{-}Phenyl_{Tyr318}$ = 3.0 Å; *Figure 35*), only one possible hydrogen bond can be recorded for **19a**, **19d** and **19c** (*Figure 35*). It could be argued that **19c** is more than five times more active than **19a** while sharing the same

binding mode; nevertheless the *orto*-methyl is involved in repulsive interactions with Phe227 as confirmed by the negative contribution polyhedrons (*Figure 32B*).
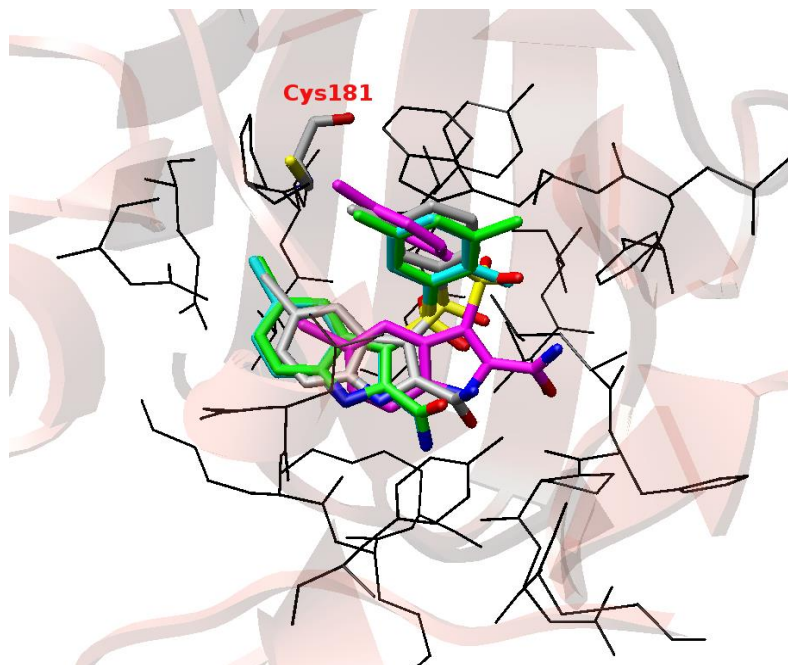

Figure 34. View of **19a** (cyan), 19e (grey), **19d** (magenta) and **19c** (green) docked conformations into the Y181C NNBS (1jkh)
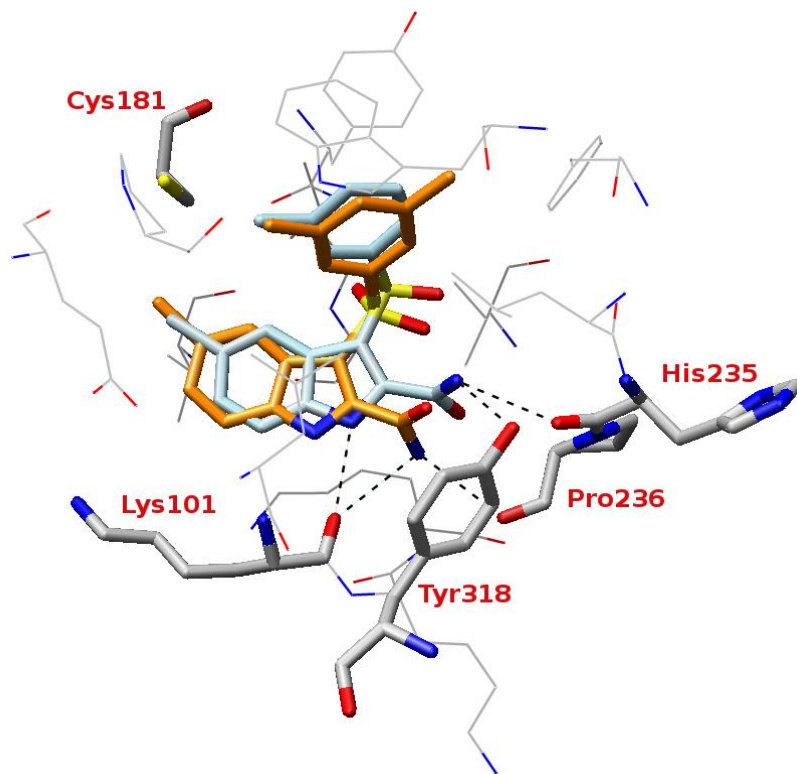
Figure 35. 19e (carbon atoms in cyan) and **19c** (carbon atoms in orange) docked conformations. A 4Å core of the 1jlq NNBS is also reported in wire style. Potential H-bond are reported in black dashed lines

6.5.5 3-D QSAR$_{K103N-Y181C}$ model characteristic and interpretation

The same Y181C model (3-D QSAR$_{Y181C}$) training set of 40 molecules was used to build

the 3-D QSAR model relative to the K103N-Y181C double mutant anti-HIV activities. A protocol similar to those used for the two previous described models (3-D QSAR$_{WT}$ and 3-D QSAR$_{Y181C}$) was employed for the molecule alignment, GRID calculations and model refinements (data pre-treatment, FFD selections). TS molecules were aligned by means of docking experiments using as receptor a modelled K103N-Y181C mutated RT. The final 3-D QSAR$_{K103N-Y181C}$ model was derived after a number of FFD selections with only 1 PCs and displaying r^2, q^2_{L5O} and SDEP values of 0.74, 0.55 and 0.52, respectively (*Figure. 36*).

   The docked conformation of derivative **103**, the most active IAS against the double mutant HIV strain (*Table 1*), was used to aid the interpretation of the subsequent reported PLS coefficients and activity contributions plots (*Figure 37*). Differently from the previous two models a straightforward interpretation can be drawn out from the plots in *Figure 37*. In the PLS coefficients plot three groups of polyhedrons can be related to the di-methylphenyl, 5-halogen indole and 2-hydroxyethyl carboxyamide side chain respectively (*Figure 38A*). These three moieties are related to positive activity contributions (*Figure 37B*) that well correlate with the corresponding ligand/receptor interactions obtained by the SB alignment.

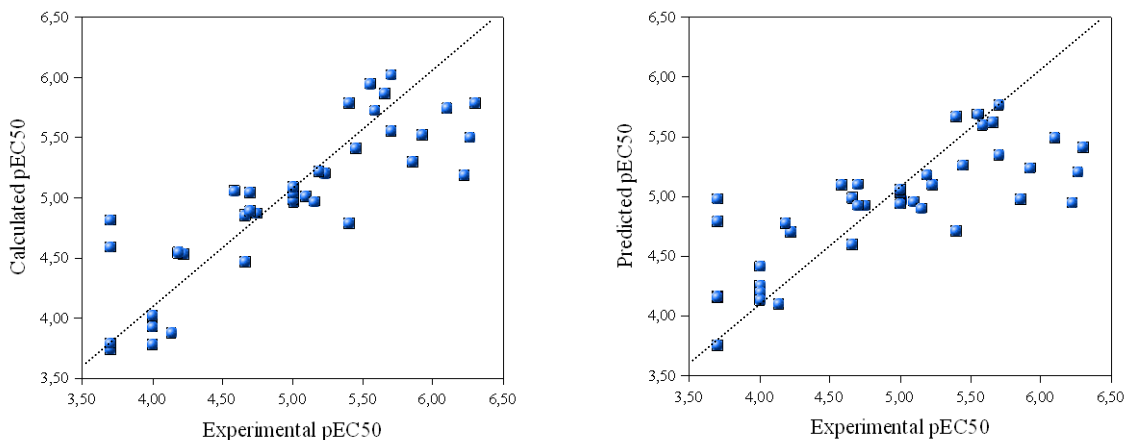| | *Vars* | **PCs** | r^2 | **L5O** | |
| | | | | q^2 | **SDEP** |
| *3-D QSARK103N-Y181C* | 673 | 1 | 0,74 | 0,55 | 0,52 |

Figure 36: Fitting (left side) and Cross validation (right side) Plots of the K103N-Y181C Model with 1 PCs.

In particular the di-methylphenyl moiety lay in a aromatic rich hydrophobic pocket formed by Cys181, Tyr188, and Trp229, the 5-chloroindolyl fulfils another pocket formed by Pro235 Tyr318 and Leu234 side chains (*Figure 37*) and finally two hydrogen bonding distances were recorded between the hydroxyethyl OH group from the Lys101 and the side chain ε–N and between the indole N1 and Lys101 carbonyl oxygen (*Figure 37*). A support of the importance of the latter interactions, compound **19c**, lacking of the hydroxyethyl group and having a similar docked conformation is less active.
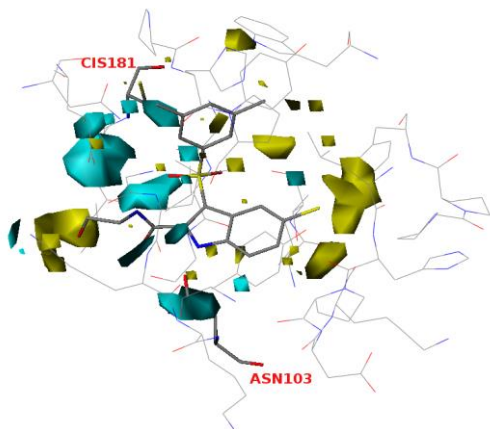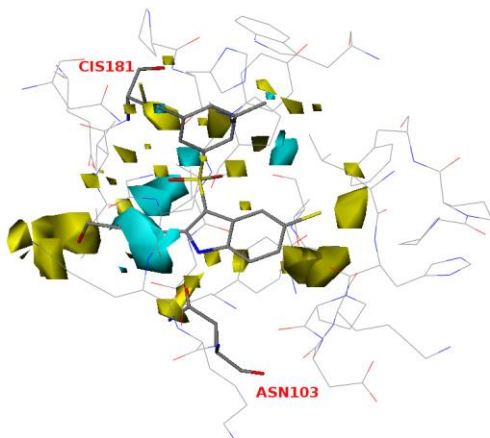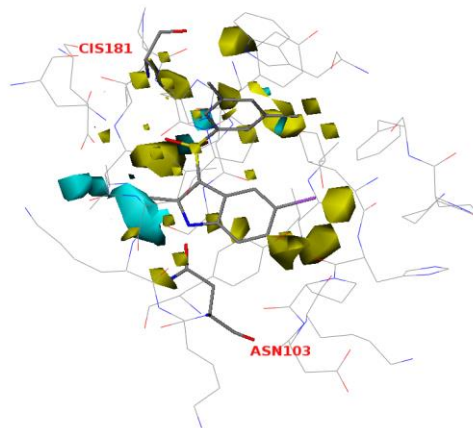
A1
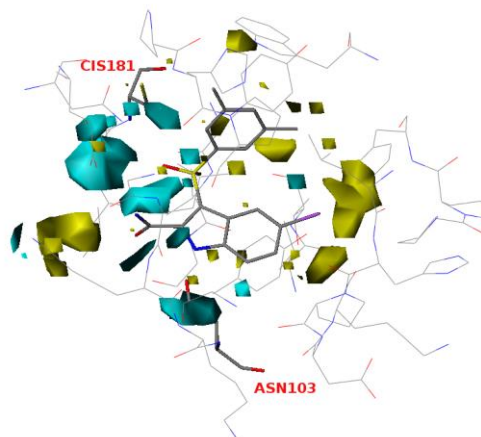
B1

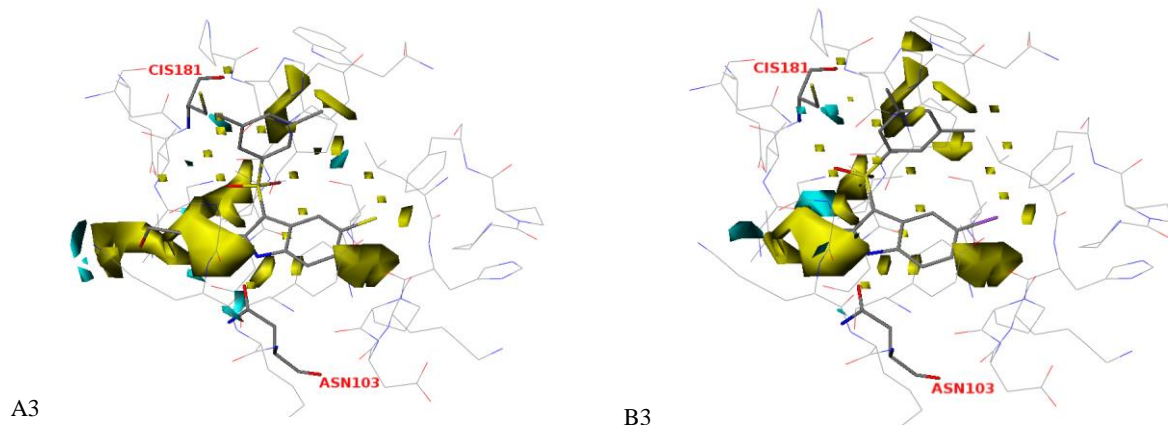**Figure37** (To be continued)

A2

B2

A3

B3

Figure37 PLS coeff (1), Activity Contribution (2), and Field (3) of **103** (A) and **19c** (B).

6.5.6 Prediction

Thanks to the 3-D QSAR analysis, we have designed and predict different series of new IASs. The plots generated by GOLPE suggest us to modify molecular function were polyhedrons interpretation got the great number of information. For the tree models analyzed these moieties are: the side chain in position 2 and chlorine atom in position 5 of indole. In particular our attention was focused on the substitution of chlorine atom with bromine or nitro group, and the contemporary di-substitution in position 4-5, 4-6, 5-6 and 5-7 on indole nucleus with chlorine and fluorine atoms. On the other hand we have to evaluate the effect of modification of side chain with introduction of one or more units of both natural and non-natural amino-acids and the substitution with heterocyclic aromatic and non aromatic. The prediction set was about 100 compounds. The conformation of each was obtained repeating

docking procedure used for the training set

Table15.1: Di-Halogen Derivatives
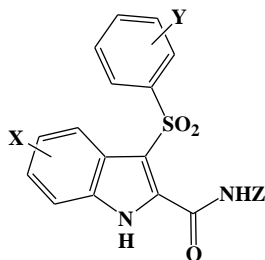
Comp	X	Y	Z	WT	Y181C	K103N-Y181C
140	5,6-Cl$_2$	H	H	5.91	7.15	3.84
141	5,6-Cl$_2$	3,5-Me$_2$	H	6.1	6.9	4.6
142	5,6-Cl$_2$	3,5-Me$_2$	CH$_2$CH$_2$OH	8.25	7,17	5.27
143	5,7-F$_2$	3,5-Me$_2$	H	8.16	7.49	5.06
144	5,6-F$_2$	3,5-Me$_2$	H	7.55	7.5	5.5
145	4,5-F$_2$	3,5-Me$_2$	H	8.06	7.75	5.34
146	6-Cl,5-F	3,5-Me$_2$	H	6.43	7.2	5.57
147	4-Cl,5-F	3,5-Me$_2$	H	7.37	8.08	5.61
148	5-Cl,6-F	3,5-Me$_2$	H	6.42	6.71	5.64
149	5-Cl,4-F	3,5-Me$_2$	H	7.07	7.21	5.78
150	5,6-F$_2$	H	H	6.07	6.88	4.81
151	4,5-F$_2$	H	H	7.16	7.42	4.88

New IASs molecules were separated into four different groups (T*able15.1 15.5*) using new introduced moiety to classify it.

Table 15.2: Heterocyclic Derivatives 1

A   B
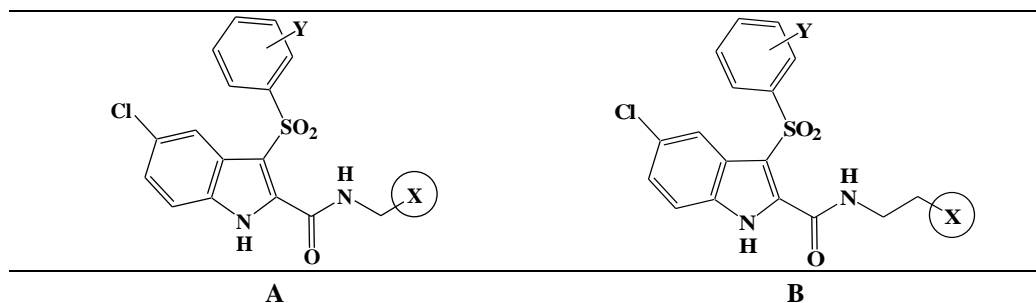
Comp		X	Y	WT	Y181C	K103N-Y181C
152	A	pyrrolidin-1-yl	H	7.22	7.3	4.24
153	A	piperidin-1-yl	H	8.12	6.46	4.85
154	A	pyrrolidin-1-yl	3,5-Me$_2$	6,97	6,4	4,73
155	A	piperidin-1-yl	3,5-Me$_2$	8,75	6,45	3,33
156	A	morpholin-4-yl	3,5-Me$_2$	6,31	7,04	5,41
157	A	morpholin-4-yl	3,5-Me$_2$	8,24	6,94	4,75
158	B	2-Me-5-NO$_2$-imidazol-1-yl	3,5-Me$_2$	6,99	6,43	4,18
159	B	2-Me-5-NO$_2$-imidazol-1-yl	H	6,77	6,88	5,29
160	B	2-Me-5-NO$_2$-imidazol-1-yl	3-F	7,41	6,27	5,46
161	B	2-Me-5-NO$_2$-imidazol-1-yl	3-Me	7,87	6,81	5,19
162	B	pyrrol-1-yl	3,5-Me$_2$	8,17	6,71	5,84
Comp		X	Y	WT	Y181C	K103N-Y181C
163	B	2-Me-5-NO$_2$-imidazol-1-yl	3-Cl	7,85	6,5	3,91
164	B	2-Me-5-NO$_2$-imidazol-1-yl	3-OMe	7,29	6,43	4,83

165	B	phenyl	3,5-Me$_2$	8,26	6,42	4,29
166	B	imidazol-1-yl	3,5-Me$_2$	8,29	7,47	5,18
167	B	thiophen-2-yl	3,5-Me$_2$	8,13	6,96	5,39
168	B	4-Cl-phenyl	3,5-Me$_2$	7,8	6,72	4,87
169	B	2-Cl-phenyl	3,5-Me$_2$	7,3	6,92	5,09
170	B	3-Cl-phenyl	3,5-Me$_2$	7,75	6,41	5,18
171	B	2-Me-phenyl	3,5-Me$_2$	8,03	6,48	5,78
172	B	3-Me-phenyl	3,5-Me$_2$	7,8	6,44	4,55
173	B	4-Me-phenyl	3,5-Me$_2$	7,78	6,23	5,08
174	B	morpholin-4-yl	3,5-Me$_2$	7,8	6,84	5,12
175	B	pyrrolidin-1-yl	3,5-Me$_2$	7,71	7,29	4,54

All the predicted pEC$_{50}$ values were used to build a consensus model. Molecules for in vitro testing were selected with a rank-by-rank approach based on the predicted EC$_{50}$ values of the three models. Compounds with lower rank sum should be active versus not only wild type but also against two analyzed mutant strains. These derivatives were selected to be synthesized and tested (*Table 16*)

The analysis of wild type 3-D QSAR predicted biological activity values confirmed the goodness of model resulting a very powerful tool for activity prediction (*Table 20A*), Standard Deviation Errors of Prediction (SDEP) was 0.41. In particular compound RS1980S (Exp 8.20; Pre 8.82; EC$_{50}$ 0.003 μM) result active against wild type RT at η-molar concentration. S refers to amino acid configuration that doesn't change in condensation reaction.

For not enantiomeric pure amino-acids, after condensation on indole nucleus, racemic moisture was resolved by chiral HPLC and single enantiomer was tested. New IASs binding mode into wild type RT NNBS study show two different
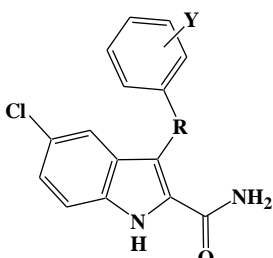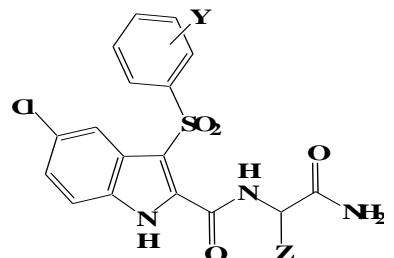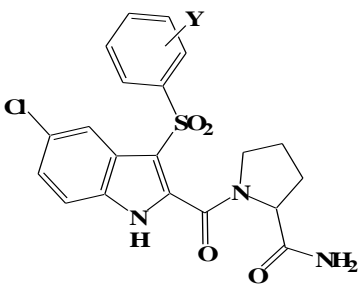
Table 2 15.3: Heterocyclic derivatives 2

A

B
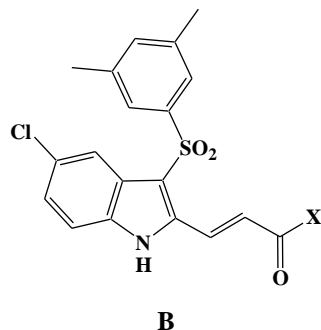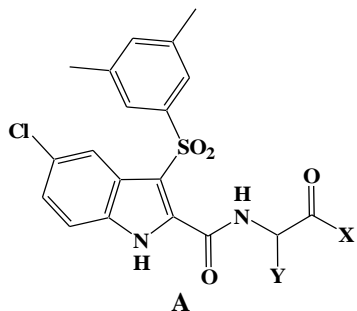
Comp		X	Y	WT	Y181 C	K103N Y181C
176	A	Morpholin-4-yl	H	5.36	6.66	4.48
177	A	Cyclo-propylamino	H	7.94	6.51	4.55
178	A	*Tert*-butylamino	H	7.02	7.46	5.42
179	A	Anilino	H	7.62	6.77	4.8

Comp			X	Y	WT	Y181 C	K103N Y181C
180	A	R	NH_2	CH_2SH	7.97	6.69	5.22
		S			7.46	6.08	5.27
181	A	R	$NHNH_2$	CH_3	8.10	7.25	4.82
		S			7.87	7.33	5.16
182	A	R	$NHNH_2$	CH_2SH	8.06	6.74	5.15
		S			7.36	6.56	4,75
183	A	R	NH_2	CH_2OH	7,86	7.48	5,47
		S			8.37	7.22	4.98
184	A	R	$NHNH_2$	CH_2OH	8.55	7.04	4.24
		S			8.19	7.73	5.13
185	A		Gly-$NHCH_2CH_2OH$	H	8.14	6.67	4.67

|						
186	A	Gly-N(CH$_2$CH$_2$OH)$_2$	H	7.76	5.8	5.18
187	A	NHCH$_2$CH$_2$-1-(2-Me-5-NO$_2$-imidazolyl)	H	5,71	6.84	4,82
188	A	Gly-NHCH$_2$CH$_2$-1-(2-Me-5-NO$_2$-imidazolyl)	H	5.46	6.58	4.31
189	A	NHNHCH$_2$CONH$_2$	H	6.82	7.14	4.59
190	A	NHNHCH$_2$CONHNH$_2$	H	7.1	7.14	5.16
191	B	NH$_2$	-	6.4	6.74	5.23
192	B	NHCH$_2$CONH$_2$	-	6.42	6.86	4.89

spatial dispositions into WT RT NNBS: 6 of 8 derivatives have the same BM of reference compound **103**. Only derivatives (unexpected) **145** and **262** (one of less active molecule) showed indole nucleus rotated (*Figure 38 A1-A2*).
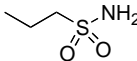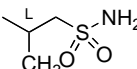
Table 15.4: Nitro Derivatives

A

B

C

Comp		R	X	Y	Z	WT	Y181 C	K103N Y181C
193	A	S	5-NO$_2$	H	-	5.24	6.54	5.48
194	A	SO$_2$	5-NO$_2$	H	-	6.33	7.19	5.74
195	A	SO$_2$	5-NO$_2$	3,5-Me	-	8.17	6.95	5.57
196	B		5-NO$_2$	3,5-Me	CH$_2$CH$_2$SCH$_3$	5.74	6.23	5.42

197	B	5-NO$_2$	3,5-Me	CH$_2$PH	6.91	6.13	4.68
198	B	5-NO$_2$	3,5-Me	CH$_2$-indol-3-yl	3.88	6.13	5.49
199	B	5-NO$_2$	3,5-Me	CH$_2$OH	7.85	6.92	4.42
200	B	5-NO$_2$	3,5-Me	CH(CH3)2	6.77	6.78	4.76
201	B	5-NO$_2$	3,5-Me	CH$_2$CH(CH$_3$)$_2$	5.94	6.9	5.58
202	B	5-NO$_2$	3,5-Me		4.3	6.13	4.57
203	C	-	-	-	5	6.99	4.6
204	B	5-NO$_2$	3,5-Me	(CH$_2$)$_2$	6.38	7.02	5.32
205	B	5-NO$_2$	3,5-Me	CH(CH$_3$)	7.4	6.43	4.88
206	A	5-pyrryl	3,5-Me	-	5.93	6.53	5.35
207	A	5-pyrryl	3,5-Me	-	6.68	5.6	4.3

Table 15.5: Amino Acids Derivatives

R$_1$ — (indole scaffold) — SO$_2$ — (3,5-dimethylphenyl); 2-position: C(=O)–NHR$_2$; N–H
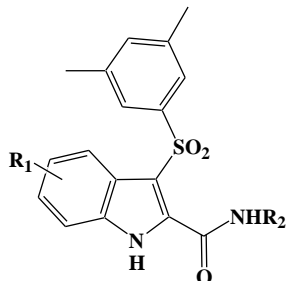
Comp	Abs. Conf	R$_1$	R$_2$	WT	Y181C	K103N Y181C
208	-	5-Br	C(=O)NH$_2$ (propanamide)	6.28	7.26	4.15
209	-	5-NO$_2$		6.9	6.54	5.57
210	R	5-Cl		8.92	7.32	4.9
	S			8.19	6.73	5.4
211	R	5-Br	C(=O)NH$_2$, CH$_3$ (D,L, 2-methylpropanamide)	5.45	6.74	5.06
	S			6.05	6.69	4.88
212	R	5-NO$_2$		4.5	7.45	5.44
	S			4.78	7.90	4.49
213	R	5-Cl,4-F		5.14	6.74	4.68
	S			6.03	6.39	4.99
214	-	5-Cl	C(=O)NH$_2$ (butanamide)	5.96	7.3	4.26
215	-	5-Br		7.89	7.87	5.18
216	-	5-NO$_2$		6.38	7.02	5.32
217	-	5-Cl,4-F		5.77	8.02	5.26
218	R	5-Cl		5.18	6.73	4.79

Comp	Abs. Conf	R$_1$	R$_2$	WT	Y181C	K103N Y181C
	S		C(=O)NH$_2$, CH$_3$ (D,L)	3.4	6.74	4.35
219	R	5-Br		5.33	6.05	5.09
	S			6.96	6.72	4.86

No.	Config.	R	Structure			
220	R S	5-NO₂		5.83	6.90	4.64
221		5-Cl		5.19	6.95	5.16
222		5-Br		4.46	6.63	5.35
223		5-NO₂		7.22	7.82	5.04
224	R S	5-Cl		4.88 5.24	6.75 7.25	5.30 5.82
225	R S	5-Br		6.13 7.56	6.79 6.27	4.89 4.56
226		5-NO₂		7.22	7.82	5.04

The analysis of experimental and predicted anti-HIV activity values for Y181C model (*Table 16B*) showed that it was over-predictive with SDEP 1.27. A wide training set, (with new tested IASs), should improved model prediction ability. In spite of this the screening method permitted to select two compounds **262** and **210S** active at the same concentration of reference **19e**.

Table 16: Predicted Compounds
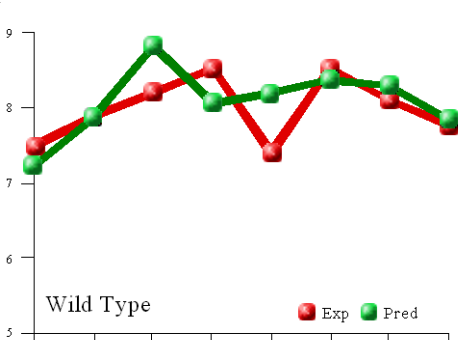
Comp	WT EC$_{50}$	
	Exp	Pred
262	7.45	7.22
2043S	7.87	7.89
210S	8.82	8.20
145	7.52	8.06
184S	7.40	8.19
183S	8.52	8.37
166	8.10	8.29
199	7.77	7.85

A: Wild Type Model
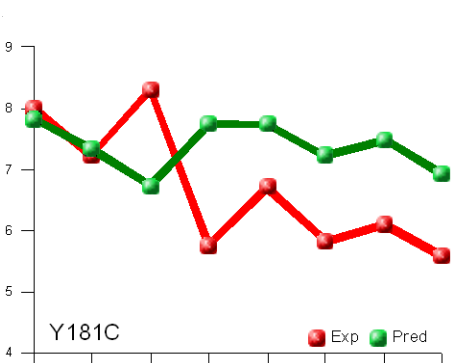
Comp	WT EC$_{50}$	
	Exp	Pred
262	8	7.82
2043S	7.22	7.33
210S	8.30	6.73
145	5.77	7.75
184S	6.70	7.73
183S	5.82	7.22
166	6.10	7.47
199	5.59	6.92

B: Y181C Model

Table 16: Predicted Compounds (Continued)
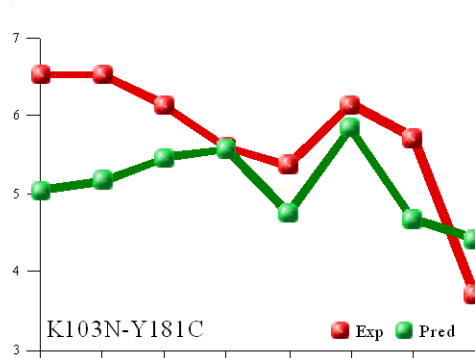
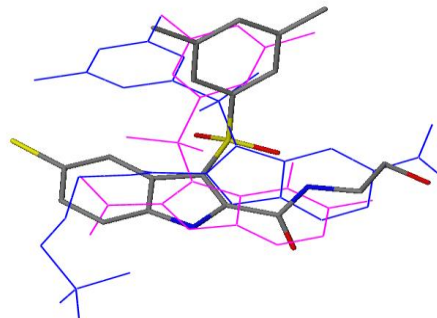| Comp	WT EC$_{50}$	
	Exp	**Pred**
262	6.52	5.04
2043S	6.52	5.16
210S	6.14	5.40
145	5.60	5.57
184S	5.36	4.75
183S	6.14	5.84
166	5.72	4.67
199	3.70	4.42

C: K103N-Y181C Model

The BM exploration of the new IASs showed two different spatial dispositions: 6 of 8 derivatives poses as reference **19e** and only two have indole nucleus rotated. These two compounds **262** and **199** were characterized by a nitro group in position 5 of indole. Probably, in NNBS of Y181C-RT isoforms, this function constrained molecules to this spatial disposition.
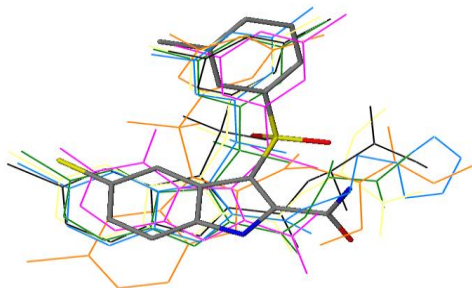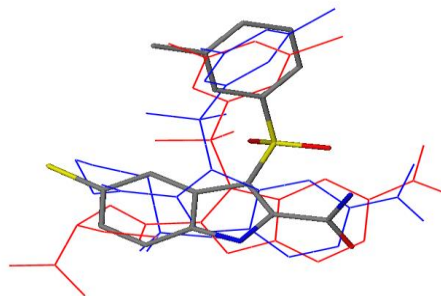
A1   A2

B1   B2

C1   C2
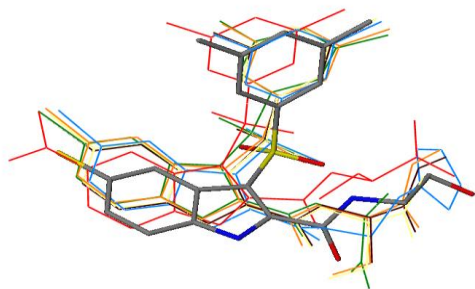
Figure 38 (To be continued)

Figure 38: Binding Mode of new IASs in wild Type, Y181C and K103N-Y181C

The K103N-Y181C model showed under-predictive ability. Also for this model an extended training set should enhance the prediction capability, although among the selected derivatives three have biological activity values close to the reference compound. The BM analysis has given a dominant conformation (5 of 8) common to the most active molecules (**103**).

7. Conclusions.

This thesis describes the progress made in the development of anti-HIV-1 indolyl aryl sulphones. Molecular modelling, docking and 3-D QSAR studies played a key role in this research. New and updated software allowed more refined SBDD and LBDD studies that gave a great contribution in understanding the possible binding mode of known IASs and the definition of highly predictive 3-D QSAR models. These findings guided the disclosure of new and more potent IAS series. The new IASs were potent inhibitors of the HIV-1 WT and clinically relevant mutant strains in both cellular-based and enzyme assays at nanomolar concentrations. These encouraging results did not quench our enthusiasm in searching for more and more new updated SBDD and LBDD based models. In particular new studies are ongoing aimed to design new compounds endowed with activity against the HIV-1 mutant strains. Therefore, specific SBDD and/or LBDD models will be soon developed using our accumulated activity data.

References

1. *Pneumocystis* pneumonia, Los Angeles. MMWR Morb Mortal Wkly Rep. **1981**, *30*, 250-252.

2. Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men, New York City and California. MMWR Morb Mortal Wkly Rep. **1981**, *30*, 305-308.

3. Masur, H.; Michelis, M. A.; Greene, J. B.; Onorato, I.; Stouwe, R. A.; Holzman, R. S.; Wormser, G.; Brettman, L.; Lange, M.; Murray, H. W.; Cunningham-Rundles, S. An outbreak of community-acquired *Pneumocystis carinii* pneumonia: initial manifestation of cellular immune dysfunction. *N. Engl. J. Med.* **1981**, *305*, 1431-1438.

4. Opportunistic infections and Kaposi's sarcoma among Haitians in the United States. MMWR Morb Mortal Wkly Rep. **1982**, *31*, 353-354.

5. Update on acquired immune deficiency syndrome (AIDS) among patients with hemophilia A. MMWR Morb Mortal Wkly Rep. **1982**, *31*, 644-646.

6. CDC. Update on acquired immune deficiency syndrome (AIDS), United States. MMWR Morb Mortal Wkly Rep. **1982**, *31*, 507-508.

7. Sonnabend, J.; Witkin, S. S.; Purtilo, D. T. Acquired immunodeficiency syndrome, opportunistic infections, and malignancies in male homosexuals. A hypothesis of etiologic factors in pathogenesis. *JAMA* **1983**, *249*, 2370-2374.

8. Barre-Sinoussi, F.; Chermann, J. C.; Rey, F.; Nugeyre, M. T.; Chamaret, S.; Gruest, J.; Dauguet, C.; Axler-Blin, C.; Vezinet-Brun, F.; Rouzioux, C.; Rozenbaum, W.; Montagnier, L. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **1983**, *220*, 868-871.

9. Gallo, R. C.; Salahuddin, S. Z.; Popovic, M.; Shearer, G. M.; Kaplan, M.; Haynes, B. F.; Palker, T. J.; Redfield, R.; Oleske, J.; Safai, B. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* **1984**, *224*, 500-503.

10. Levy, J. A.; Hoffman, A. D.; Kramer, S. M.; Landis, J. A.; Shimabukuro, J. M.; Oshiro, L. S. Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. *Science* **1984**, *225*, 840-842.

11. Guyader, M.; Emerman, M.; Sonigo, P.; Clavel, F.; Montagnier, L.; Alizon, M. Genome organization and transactivation of the human immunodeficiency virus type 2. *Nature* **1987**, *326*, 662-669.

12. Bozzette, S. A.; Joyce, G.; McCaffrey, D. F. Expenditures for the care of HIV-infected patients in the era of highly active antiretroviral therapy. *N. Engl. J. Med*. **2001**, *344*, 817-823

13. UNAIDS forecasts future of global progression of HIV/AIDS. *AIDS Policy Law* (2006), **20**(8):1, 4.

14. Freed, E. O.; Martin, M. A. The role of human immunodeficiency virus type 1 envelope glycoproteins in virus infection. *J. Biol. Chem.* **1995**, *270*, 23883-23886.

15. Luciw, P. A. Human immunodeficiency virus and its replication. In: Fields BM, editor. Fields virology. 3rd ed. Philadelphia: Lippincott-Raven Publishers, **1996**.

16. Levy, J. A. Pathogenesis of HIV infection. *Microbiol. Rev*. **1993**, *57*, 183-289.

17. Ferguson, M. R.; Rojo, D. R.; von Lindern, J.J.; O' Brien, W. A. HIV-1 replication cycle. *Clin. Lab. Med*. **2002**, *22*, 611-635.

18. Dalgleish, A. G.; Beverley, P. R.; Clapham, D. H.; Crawford, D.H.; Greaves, M. F.; Weiss, R. A. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* **1984**, *312*, 763-767.

19. Littman, D. Chemokine receptors: keys to AIDS pathogenesis. *Cell* **1998**, *93*, 677-680.

20. Kwong, P. D.; Wyatt, R.; Robinson, J.; Sweet, R. W.; Sodroski, J.; Hendrickson, W. A. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **1998**, *393*, 648-59.

21. Chan, D. C.; Kim, P. S. HIV entry and its inhibition. *Cell* **1998**, 93, 681-684.

22. Geijtenbeek, T. B.; Kwon, D. S.; Torensma, R.; van Vliet, S. J.; van Duijnhoven, G. C.; Middel, J.; Cornelissen, I. L.; Nottet, H. S.; KewalRamani, V. N.; Littman, D. R.; Figdor, C. G.; van Kooyk, Y. DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell* **2000**, *100*, 587-597.

23. Kwon, D. S.; Gregorio, G.; Bitton, N.; Hendrickson, W. A.; Littman, D. R. DC-SIGN-mediated internalization of HIV is required for trans-enhancement of T cell infection. *Immun.* **2002**, *16*, 135-144.

24. Franke, E. K.; Yuan, H. E.; Luban, J. Specific incorporation of cyclophilin A into HIV-1 virions. *Nature* **1994**, *372*, 359-362.

25. Schaeffer, E.; Geleziunas, R.; Greene, W. C. Human immunodeficiency virus type 1 Nef functions at the level of virus entry by enhancing cytoplasmic delivery of virions. *J. Virol.* **2001**, *75*, 2993-3000.

26. Karageorgos, L.; Li, P.; Burrell, C. Characterization of HIV replication complexes early after cell-to-cell infection. *AIDS Res.Hum. Retrovir.* **1993**, *9*, 817-823.

27. Lu, X.; Yu, H.; Liu, S. H.; Brodsky, F. M.; Peterlin, B. M. Interactions between HIV1 Nef and vacuolar ATPase facilitate the internalization of CD4. *Immun.* **1998**, *8*, 647-656.

28. Bukrinskaya, A.; Brichacek, B.; Mann, A.; Stevenson, M. Establishment of a functional human immunodeficiency virus type 1 (HIV-1) reverse transcription complex involves the cytoskeleton. *J. Exp Med.* **1998**, *188*, 2113-2125.

29. Miller, M. D.; Farnet, C. M.; Bushman, F. D. Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J. Virol.* **1997**, *71*, 5382-5390.

30. McDonald, D.; Vodicka, M. A.; Lucero, G.; Svitkina, T. M.; Borisy, G. G.; Emerman, M.; Hope, T. J. Visualization of the intracellular behavior of HIV in living cells. *J. Cell Biol.* **2002**, *159*, 441-452.

31. Pemberton, L. F.; Blobel, G.; Rosenblum, J. S. Transport routes through the nuclear pore complex. *Curr. Opin Cell. Biol.* **1998**, *10*, 392-399.

32. Zennou, V.; Petit, C.; Guetard, D.; Nerhbass, U.; Montagnier, L.; Charneau, P. HIV-1 genome nuclear import is mediated by a central DNA flap. *Cell* **2000**, *101*, 173-185.

33. Dvorin, J. D.; Bell, P.; Maul, G. G.; Yamashita, M.; Emerman, M.; Malim, M. H. Reassessment of the roles of integrase and the central DNA flap in human immunodeficiency virus type 1 nuclear import. *J. Virol.* **2002**, *76*, 12087-12096.

34. Sherman, M. P.; de Noronha, C. M.; Heusch, M. I.; Greene, S.; Greene, W. C. Nucleocytoplasmic shuttling by human immunodeficiency virus type 1 Vpr. *J. Virol.* **2001**, *75*, 1522-1532.

35. Vodicka, M. A.; Koepp, D. M.; Silver, P. A.; Emerman, M. HIV-1 Vpr interacts with the nuclear transport pathway to promote macrophage infection. *Genes Dev.* **1998**, *12*, 175-185.

36. Dupont, S.; Sharova, N.; DeHoratius, C.; Virbasius, C. M.; Zhu, X.; Bukrinskaya, A. G.; Stevenson, M.; Green, M. R. A novel nuclear export activity in HIV-1 matrix protein required for viral replication. *Nature* **1999**, *402*, 681-685.

37. Wu, Y.; Marsh, J. W. Selective transcription and modulation of resting T cell activity by preintegrated HIV DNA. *Science* **2001**, *293*, 1503-1506.

38. Li, L.; Olvera, J. M.; Yoder, K. E.; Mitchell, R. S.; Butler, S. L.; Lieber, M.; Martin, S. L.; Bushman, F. D. Role of the non-homologous DNA end joining pathway in the early steps of retroviral infection. *EMBO J.* **2001**, *20*, 3272-3281.

39. Jordan, A.; Defechereux, P.; Verdin, E. The site of HIV-1 integration in the human genome determines basal transcriptional activity and response to Tat transactivation. *EMBO J.* **2001**, *20*, 1726-1738.

40. Taube, R.; Fujinaga, K.; Wimmer, J.; Barboric, M.; Peterlin, B. M. Tat transactivation: a model for the regulation of eukaryotic transcriptional elongation. *Virol.* **1999**, *264*, 245-253.

41. Jones, K. A.; Peterlin, B. M. Control of RNA initiation and elongation at the HIV-1 promoter. *Ann. Rev. Biochem.* **1994**, *63*, 717-743.

42. Barboric, M.; Nissen, R. M.; Kanazawa, S.; Jabrane-Ferrat, N.; Peterlin, B. M. NFkB binds P-TEFb to stimulate transcriptional elongation by RNA polymerase II. *Mol. Cell.* **2001**, *8*, 327-337.

43. Kao, S. Y.; Calman, A. F.; Luciw, P. A.; Peterlin, B. M Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. *Nature* **1987**, *330*, 489-493.

44. Price, D. H. P-TEFb, a cyclin-dependent kinase controlling elongation by RNA polymerase II. *Mol. Cell. Biol.* **2000**, *20*, 2629-2634.

45. Garber, M. E.; Wei, P.; KewalRamani, V. N.; Mayall, T. P.; Herrmann, C. H.; Rice, A. P.; Littman, D. R.; Jones, K. A. The interaction between HIV-1 Tat and human cyclin T1 requires zinc and a critical cysteine residue that is not conserved in the murine CycT1 protein. *Genes Dev.* **1998**, *12*, 3512-3527.

46. Yang, Z.; Zhu, Q.; Luo, K.; Zhou Q. The 7SK small nuclear RNA inhibits the CDK9/cyclin T1 kinase to control transcription. *Nature* **2001**, *414*, 317-322.

47. Luo, Y.; Yu, H.; Peterlin, B. M. Cellular protein modulates effects of human immunodeficiency virus type 1 Rev. *J. Virol.* **1994**, *68*, 3850-3856.

48. Malim, M. H.; Tiley, L. S.; McCarn, D. F.; Rusche, J. R.; Hauber, J.; Cullen, B. R. HIV-1 structural gene expression requires binding of the Rev trans-activator to its RNA target sequence. *Cell* **1990**, *60*, 675-683.

49. Cullen, B. R. Retroviruses as model systems for the study of nuclear RNA export pathways. *Virol.* **1998**, *249*, 203-210.

50. Malim, M. H.; Cullen, B. R. Rev and the fate of pre-mRNA in the nucleus: implications for the regulation of RNA processing in eukaryotes. *Mol. Cell. Biol.* **1993**, *13*, 6180-6189.

51. Khan, I. H.; Sawai, E. T.; Antonio, E.; Weber, C. J.; Mandell, C. P.; Montbriand, P.; Luciw, P. A. Role of the SH3-ligand domain of simian immunodeficiency virus Nef in

interaction with Nef-associated kinase and simian AIDS in rhesus macaques. *J. Virol.* **1998**, *72*, 5820-5830.

52. Lama, J.; Mangasarian, A.; Trono, D. Cell-surface expression of CD4 reduces HIV-1 infectivity by blocking Env incorporation in a Nef- and Vpu-inhibitable manner. *Curr. Biol.* **1999**, *9*, 622-631.

53. Geyer, M.; Fackler, O. T.; Peterlin, B. M. Structure-function relationships in HIV-1 Nef. *EMBO Rep.* **2001**, *2*, 580-585.

54. Kiyokawa, E.; Verdin, E.; Trono, D. The Nef protein of HIV-1 associates with rafts and primes T cells for activation. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 394-399.

55. Chen, B. K.; Gandhi, R. T.; Baltimore, D. CD4 down-modulation during infection of human T cells with human immunodeficiency virus type 1 involves independent activities of vpu, env, and nef. *J. Virol.* **1996**, *70*, 6044-6053.

56. Margottin, F.; Bour, S. P.; Durand, H.; Selig, L.; Benichou, S.; Richard, V.; Thomas, D.; Strebel, K.; Benarous, R. A novel human WD protein, h-beta TrCp, that interacts with HIV-1 Vpu connects CD4 to the ER degradation pathway through an F-box motif. *Mol. Cell.* **1998,** *1*, 565-574.

57. Xu, X. N.; Laffert, B.; Screaton, G. R.; Kraft, M.; Wolf, D.; Kolanus, W.; Mongkolsapay, J.; McMichael, A. J.; Baur, A. S. Induction of Fas ligand expression by HIV involves the interaction of Nef with the T cell receptor zeta chain. *J. Exp. Med.* **1999**, *189*, 1489-1496.

58. Collins, K. L.; Chen, B. K.; Kalams, S. A.; Walker, B. D.; Baltimore, D. HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes. *Nature* **1998**, *391*, 397-401.

59. Le Gall, S.; Erdtmann, L.; Benichou, S.; Berlioz-Torrent, C.; Liu, L.; Benarous, R.; Heard, J. M.; Schwartz, O. Nef interacts with the μ subunit of clathrin adaptor complexes and reveals a cryptic sorting signal in MHC I molecules. *Immun.* **1998**, *8*, 483-495.

60. Wolf, D.; Witte, V.; Laffert, B.; Blume, K.; Stromer, E.; Trapp, S.; d'Aloja, P.; Schurmann, A.; Baur, A. S. HIV-1 Nef associated PAK and PI3-kinases stimulate Akt-independent Bad-phosphorylation to induce anti-apoptotic signals. *Nat. Med.* **2001**, *7*, 1217-1224.

61. Jowett, J. B.; Planelles, V.; Poon, B.; Shah, N. P.; Chen, M. L.; Chen, I. S. The human immunodeficiency virus type 1 vpr gene arrests infected T cells in the G2 + M phase of the cell cycle. *J. Virol.* **1995**, *69*, 6304-6313.

62. Goh, W. C.; Rogel, M. E.; Kinsey, C. M.; Michael, S. F.; Fultz, P. N.; Nowak, M. A.; Hahn, B. H.; Emerman, M. HIV-1 Vpr increases viral expression by manipulation of the cell cycle: a mechanism for selection of Vpr in vivo. *Nat. Med.* **1998**, *4*, 65-71.

63. de Noronha, C. M.; Sherman, M. P.; Lin, H. W.; Cavrois, M. V.; Moir, R. D.; Goldman, R. D.; Greene, W. C. Dynamic disruptions in nuclear envelope architecture and integrity induced by HIV-1 Vpr. *Science* **2001**, *294*, 1105-1108.

64. Wilk, T.; Gross, I.; Gowen, B. E.; Rutten, T.; de Haas, F., Welker, R.; Krausslich, H. G.; Boulanger, P., Fuller, S. D. Organization of immature human immunodeficiency virus type 1. *J. Virol.* **2001**, *75*, 759-771.

65. Freed, E. O. HIV-1 gag proteins: diverse functions in the virus life cycle. *Virol.* **1998**, *251*, 1-15.

66. Zimmerman, C.; Klein, K. C.; Kiser, P. K.; Singh, A. R.; Firestein, B. L.; Riba, S. C.; Lingappa, J. R. Identification of a host protein essential for assembly of immature HIV-1 capsids. *Nature* **2002,** *415*, 88-92.

67. Madani, N.; Kabat, D. An endogenous inhibitor of human immunodeficiency virus in human lymphocytes is overcome by the viral Vif protein. *J. Virol.* **1998**, *72*, 10251-10255.

68. Sheehy, A. M.; Gaddis, N. C.; Choi, J. D.; Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **2002**, *418*, 646-650.

69. Gottlinger, H. G.; Sodroski, J. G.; Haseltine, W. A. Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5781-5785.

70. Strack, B.; Calistri, A.; Accola, M. A., Palu, G., Gottlinger, H. G. A role for ubiquitin ligase recruitment in retrovirus release. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 13063-13068.

71. Katzmann, D. J.; Babst, M.; Emr, S. D. Ubiquitin-dependent sorting into the multivesicular body pathway requires the function of a conserved endosomal protein sorting complex, ESCRT-I. *Cell* **2001**, *106*, 145-155.

72. Centers for Disease Control. Report of the NIH Panel to define principles of therapy of HIV infection and guidelines for the use of antiretroviral agents in HIV-infected adults and adolescents. *MMWR* **1998**, *47* (RR-5), 1-83.

73. Lalezari, J. P.; Henry, K.; O' Hearn, M. Enfuvirtide, an HIV-1 fusion inhibitor, for drug resistant HIV infection in North and South America. *N. Engl. J. Med*. **2003**, *348*, 2175-2185.

74. Mitsuya, H.; Weinhold, K. J.; Furman P. A.; Clair, M. H.; Nusinofff-Lehrman, S.; Gallo, R. C.; Bolognesi, D.; Barry, D. W.; Broder, S. 3'-azido-3'-deoxythymidine (bwa509u): an antiviral agent that inhibits the infectivity and cytopathic effect of

human T-lymphotropicvirus type III/lymphadenopathy-associated virus in vitro. *Proc. Natl. Acad. Sci.* USA **1985**, *82*, 7096-7100

75. . Mitsuya, H.; Broder, S. Inhibition of the in vitro infectivity and cytopathic effect of human T-lymphotrophic virus type III/lymphadenopathy-associated virus (HTLV-III/LAV) by 2',3'-dideoxynucleosides. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 1911-1915.

76. Lin, T.-S.; Schinazi, R. F.; Prusoff, W. H. Potent and selective in vitro activity of 3'-deoxythymidin-2'-ene(3'-deoxy-2',3'-didehydrothymidine) against human immunodefiency virus. *Biochem. Phar.* **1987**, *36*, 2713-2718.

77. Arimilli, M. N.; Kim, C. U.; Dougherty, J.; Mulato, A.; Oliyai, R.; Shaw, J. P.; Cundy, K. C.; Bischofberger, N. Synthesis, in vitro biological evaluation and oral bioavailability of 9-[2-[(phosphonomethoxy)propyl]adenine (PMPA) prodrugs. *Antivir. Chem. Chemother.* **1997**, *8*, 557-564.

78. Coates, J. A. V.; Cammack, N.; Jenkinson, H. J.; Mutton, I. M.; Pearson, B. A.; Storer, R.; Cameron, J. M.; Penn, C. R. The separated enantiomers of 2'-deoxy-3'-thiacytidine (BCH 189) both inhibit HIV replication in vitro. *Antimicrob. Agents Chemother.* **1992**, *36*, 202-205

79. Merluzzi, V. J.; Hargrave, K. D.; Labadia, M.; Grozinger, K.; Skoog, M.; Wu, J. C.; Shih, C.-K.; Eckner, K.; Hattox, S.; Adams, J.; Rosehthal, A. S.; Faanes, R.; Eckner, R. J.; Koup, R. A.; Sullivan, J. L. Inhibition of HIV-1 replication by a non-nucleoside reverse transcriptase inhibitor. *Science* **1990**, *250*, 1411-1413.

80. Romero, D. L.; Morge, R. A.; Genin, M. J.; Biles, C.; Busso, M.; Resnick, L.; Althaus, I. W.; Reusser, F.; Thomas, R. C.; Tarpley, W. G. Bis(heteroaryl)piperazine (BHAP) RT inhibs: struct-act. Relationships of novel substituted indole analogues and the identification of monomethanesulfonate (U-90152S). *J. Med. Chem.* **1993**, *36*, 1505-1508.

81. Young, S. D.; Britcher, S. F.; Tran, L. O.; Payne, L. S.; Lumma, W. C.; Lyle, T. A.; Huff, J. R.; Anderson, P. S.; Olsen, D. B.; Carroll, S. S.; Pettibone, D J.; o' Brien, J. A.; Ball, R. G.; Balani, S. K.; Lin, J. H.; Long, W. J.; Byrnes, V. W.; Emini, E. A. L-743,726 (DMP-266): a novel, highly potent non nucleoside inhibitor of the human immunodeficiency virus type 1 reverse transcriptase. *Antimicrob. Agents Chemother.* **1995**, *39*, 2602-2605.

82. Kohlstaedt, L. A.; Wang, J.; Friedman, J. M.; Rice, P. A.; Steitz, T. A. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **1992**, *256*, 1783-1790.

83. Navia, M. A.; Sato, V. L.; Tung, R. D. Design of VX-478, a potent inhibitor of HIV protease. *Int. Antiviral News* **1995**, *3*, 143-145.

84. Gong, Y. F.; Robinson, B. S.; Rose, R. E.; Deminie, C.; Spicer, T. P.; Stock, D.; Colonno, R. J.; Lin, P. F. In vitro resistance profile of the human immunodeficiency virus type 1 protease inhibitor BMS-232632. *Antimicrob. Agents Chemother.* **2000**, *44*, 2319-26.

85. Sorbera, L. A.; Martin, L.; Castaner, J.; Castaner, R. M. Fosamprenavir. Anti-HIV HIV protease inhibitor. Drugs of the Future 2001, 26, 224-231.

86. Vacca, J. P.; Dorsey, B. D.; Schleif, W. A.; Levin, R. B.; McDaniel, S. L.; Darke, P. L.; Zugay, J.; Quintero, J. C.; Blahy, O. M.; Roth, E.; Sardana V. V.; Schlabach A. J.; Graham, P. I.; Condra J. H.; Gotlib, L.; Anderson, P. S.; Emini, E. A.; Huff, J. R. L-735,524: an orally bioavailable HIV-1 protease inhibitor. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 4096-4100.

87. Sham, H.; Kempf, D.; Molla, A.; Marsh, K.; Wideburg, N.; Chen, C.; Kati, W.; Kumar, M.; Korneyeva, M.; Vasavanonda, S.; McDonald, E.; Saldivar, A.; Chernyavskiy, T.; Carillo, A.; Lyons, N.; Park, C.; Stewart, K.; Plattner, J.; Norbeck, D. Design, synthesis

and biological properties of ABT-378, a highly potent HIV protease inhibitor (Session 5, Abstract 14). *4th Conf. Retrovir. and Opportun. Infect.* **1997**, *4*, 67.

88. Kalish, V.; Kaldor, S.; Shetty, B.; Tatlock, J.; Davies, J.; Hammond, M.; Dressman, B.; Fritz, J.; Appelt, K.; Reich, S.; Musick, L.; Wu, B.; Su, K. Iterative protein structure-based drug design and synthesis of HIV protease inhibitors. *Eur. J. Med. Chem.* **1995**, *30* (suppl), S201-S214.

89. Kempf, D. J.; Marsh, K. C.; Denissen, J. F.; McDonald, E.; Vasavanonda, S.; Flentge, C. A.; Green, B. E.; Fino, L.; Park, C. H.; Kong, X. P.; Plattner, J. J.; Leonard, J. M.; Norbeck, D. W. ABT-538 is a potent inhibitor of human immunodeficiency virus protease and has high oral bioavailability in humans. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 2484-2488.

90. Roberts, N. A.; Martin, J. A.; Kinchington, D.; Broadhurst, A. V.;

91. Craig, J. C.; Duncan, I. B.; Galpin, S. A.; Handa, B. K.; Kay, J.; Krohn, A.; Lambert, R. W.; Merrett, J. H.; Mills, J. S.; Parkes, K. E. B.; Redshaw, S.; Ritchie, A. J.; Taylor, D. L.; Thomas, G. J.; Machin, P. J. Rational design of peptide-based HIV proteinase inhibitors. *Science* **1990**, *248*, 358-361

92. Kubinj, H. 3D QSAR in drug design. Theory, methods and applications, ESCOM, Science publishers B.V., Leiden, **1993.**

93. Goodford P.J.: A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem* **1985**, *28*(7):849-857.

94. Cramer, R., D., III Patterson, D.E., Bunce, J.D. Kubinyi H., Comparative Molecular Field Analysis (CoMFA). *J. Am. Chem. Soc*. **1988** 110, 5959-5967.

95. Haenlein M., Kaplan A.M., A beginner's guide to partial least square analysis understanding statistics

96. Baroni M, Costantino G, Cruciani G, Riganelli D, Valigi R, Clementi S: Generating optimal linear PLS estimations (GOLPE): An advanced chemometric tool for handling 3D-QSAR problems. *Quant. Struct.-Act. Relat.* **1993**, *12*:9-20.

97. Richet, M.C.. Notè sur le rapport entre la toxicitè et les propietes physiques des corps, *Compt. Rend. Soc. Biol*. **1983** *45* 775-776

98. Fischer, E., Einfluss der Configuration auf die Wirkung der Enzyme, Ber.Dtsch. Chem. Ges., **1894** *27*, 2985-2993.

99. Ehrlich P., Address in pathology on chemotherapeutics: Scientific principles, methods and result. *Lancett II* **1913**, 445-451

100. Hansch, C., and Fujita T., ρ-σ-π Analysis. A method for correlation of biological activity and chemical structure. *J. Am. Chem. Soc.* **1964**, *86* 1616-1626.

101. Free Jr., S.M. and Wilson, J.W., A mathematical contribution to structure activity. *J. Med. Chem*. **1964**, *7* 395-399

102. . H. vad de Waterbeemd , Glossary of Terms Used in Computational Drug Design, IUPAC*: http://www.iupac.org/reports/1997/6905vandewaterbeemd*

103. Cramer III, R.D., Patterson, D.E., Bunce, J.D., Comparative Molecular Field Analysis (COMFA) 1. Effect of shape on binding of steroids to carrier protein. *J.Am. Chem. Soc* **1988**, *110* 5959-5967

104. SYBYL/QSAR. Molecular modeling Software Tripos Associates, Inc.St. Louis, MO.

105. Cramer III, R.D. and Wold, S.B., Comparative Molecular Field Analysis (COMFA) U.S. Patent 5,025,388 June 18 **1991**.

106. McMahon J.B., Gulakowski R.J., Weislow O.S., Schultz R.J., Narayanan V.L., Clanton D.J., Pedemonte R, Wassmundt F.W., Buckheit R.W., JR., Decker WD, et al..: Diarylsulfones, a new chemical class of nonnucleoside antiviral inhibitors of human immunodeficiency virus type 1 reverse transcriptase. *Antimicrob Agents Chemother* **1993**, *37*(4):754-760.

107. Artico M., Silvestri R., Stefancich G., Massa S., Pagnozzi E., Musu D., Scintu F., Pinna E., Tinti E., La Colla P: Synthesis of pyrryl aryl sulfones targeted at the HIV-1 reverse transcriptase. *Arch Pharm (Weinheim)* **1995**, *328*(3):223-229.

108. Williams T.M., Ciccarone T.M., MacTough S.C., Rooney C.S., Balani S.K., Condra J.H., Emini E.A., Goldman M.E., Greenlee W.J., Kauffman L.R., et al.: 5-chloro-3-(phenylsulfonyl)indole-2-carboxamide: a novel, non-nucleoside inhibitor of HIV-1 reverse transcriptase. *J Med Chem* **1993**), *36*(9):1291-1294.

109. Artico M., Silvestri R., Massa S., Loi A.G., Corrias S., Piras G., La Colla P.: 2-Sulfonyl-4-chloroanilino moiety: a potent pharmacophore for the anti-human immunodeficiency virus type 1 activity of pyrrolyl aryl sulfones. *J Med Chem* **1996**, 39(2):522-530.

110. Kukla M.J., Breslin H.J., Diamond C.J., Grous P.P., Miranda M., Rodgers J.D., Sherril R.G., De Clercq E., Pauwels R., et al.: Synthesis and anti-HIV-1 activity of 4,5,6,7-tetrahydro-5-methylimidazo[4,5,1-jk][1,4]benzodiazepin-2(1H)-on e (TIBO) derivatives. 2. *J Med Chem* **1991**, *34*(11):3187-3197.

111. Kukla M.J., Breslin H.J., Pauwels R., Fedde C.L., Miranda M., Scott M.K., Sherrill R.G., Raeymaekers A., Van Gelder J., Andries K., et al.: Synthesis and anti-HIV-1

activity of 4,5,6,7-tetrahydro-5-methylimidazo[4,5,1-jk][1,4]benzodiazepin- 2(1H)-one (TIBO) derivatives. *J Med Chem* **1991**, *34*(2):746-751.

112. Cohen K.A., Hopkins J., Ingraham R.H., Pargellis C., Wu J.C., Palladino D.E., Kinkade P., Warren T.C., Rogers S., Adams J., et al.: Characterization of the binding site for nevirapine (BI-RG-587), a nonnucleoside inhibitor of human immunodeficiency virus type-1 reverse transcriptase. *J Biol Chem* **1991**, *266*(22):14670-14674.

113. Merluzzi V.J., Hargrave K.D., Labadia M., Grozinger K., Skoog M., Wu J.C., Shih C.K., Eckner K., Hattox S., Adams J., et al.: Inhibition of HIV-1 replication by a nonnucleoside reverse transcriptase inhibitor. *Science* **1990**, *250*(4986):1411-1413.

114. Artico M., Silvestri R., Pagnozzi E, Stefancich G., Massa S., Loi A.G., Putzolu M., Corrias S., Spiga M.G., La Colla P.: 5H-pyrrolo[1,2-b] [1,2,5]benzothiadiazepines (PBTDs): a novel class of non-nucleoside reverse transcriptase inhibitors. *Bioorg Med Chem* **1996**, *4*(6):837-850.

115. Artico M., Stefancich G., Silvetsri R., Massa S., Pagnozzi E., Loi A.G., Musu D., Doa M., Scano P., La Colla P.: Pyrrolobenzothiazepines: a new class of nonnucleoside HIV-1 reverse transcriptase inhibitors. *Med Chem Res* **1994**, *4*(4):283-290.

116. Silvestri R., Artico M., De Martino G., Ragno R., Massa S., Loddo R, Murgioni C., Loi A.G., La Colla P., Pani A.: Synthesis, Biological Evaluation, and Binding Mode of Novel 1-[2-(Diarylmethoxy)ethyl]-2-methyl-5-nitroimidazoles Targeted at the HIV-1 Reverse Transcriptase. *J Med Chem* **2002**, *45*(8):1567-1576.

117. Ding J., Das K., Moereels H., Koymans L, Andries K., Janssen P.A., Hughes S.H., Aarnold E.: Structure of HIV-1 RT/TIBO R 86183 complex reveals similarity in the binding of diverse nonnucleoside inhibitors. *Nat Struct Biol* **1995**, *2*(5):407-415.

118. Artico M, Silvestri R., Pagnozzi E., Bruno B., Novellino E., Greco G., Massa S, Ettorre A., Loi A.G., Scintu F., La Colla P.: Structure-based design, synthesis, and biological evaluation of novel pyrrolyl aryl sulfones: HIV-1 non-nucleoside reverse transcriptase inhibitors active at nanomolar concentrations. *J Med Chem* **2000**, *43*(9):1886-1891.

119. Silvestri R., Artico M., De Martino G., Novellino E., Greco G., Lavecchia A., Massa S., Loi A.G., Doratiotto S., La Colla P.: Computer-assisted design, synthesis and biological evaluation of novel pyrrolyl heteroaryl sulfones targeted at HIV-1 reverse transcriptase as non-nucleoside inhibitors. *Bioorg Med Chem* **2000**, *8*(9):2305-2309.

120. Silvestri R., De Martino G., La Regina G., Artico M., Massa S., Vargiu L., Mura M., La Colla P.: Novel indolyl aryl sulfones active against HIV-1 carrying NNRTI resistance mutations: synthesis and SAR studies. *J Med Chem* **2003**, *46*(12):2482-2493.

121. Goodsell D.S., Morris G.M., Olson A.J.: Automated docking of flexible ligands: applications of AutoDock. *J Mol Recognit* **1996**, *9*(1):1-5.

122. Ragno R., Mai A., Massa S., Cerbara I., Valente S., Bottoni P., Scatena R., Jesacher F., Loidl P., Brosch G.: 3-(4-Aroyl-1-methyl-1H-pyrrol-2-yl)-N-hydroxy-2-propenamides as a New Class of Synthetic Histone Deacetylase Inhibitors. 3. Discovery of Novel Lead Compounds through Structure-Based Drug Design and Docking Studies. *J Med Chem* **2004**, *47*(6):1351-1359.

123. Ragno R., Mai A., Sbardella G., Aartico M., Massa S., Musiu C., Mura M., Marturana F., Cadeddu A., La Colla P.: Computer-aided design, synthesis, and anti-HIV-1 activity in vitro of 2-alkylamino-6-[1-(2,6-difluorophenyl)alkyl]-3,4-dihydro-5-alkylpyrimidin-4(3H)-ones as novel potent non-nucleoside reverse transcriptase inhibitors, also active against the Y181C variant. *J Med Chem* 2004, *47*(4):928-934.

124. Ragno R., Frasca S., Manetti F., Brizzi A., Massa S.: HIV-Reverse Transcriptase Inhibition: Inclusion of Ligand-Induced Fit by Cross-Docking Studies. *J Med Chem* **2005**, *48*(1):200-212.

125. Cruciani G., Watson K.A.: Comparative molecular field analysis using GRID force-field and GOLPE variable selection methods in a study of inhibitors of glycogen phosphorylase b. *J Med Chem* **1994**, *37*(16):2589-2601.

126. Ren, J.; Diprose, J.; Warren, J.; Esnouf, R. M.; Bird, L. E.; Ikemizu,
S.; Slater, M.; Milton, J.; Balzarini, J.; Stuart, D. I.; Stammers, D. K. Phenylethylthiazolylthiourea (PETT) nonnucleoside inhibitors of HIV-1 and HIV-2 reverse transcriptases. Structural and biochemical analyses. *J. Biol. Chem.* **2000**, *275*, 5633-5639.

127. Hogberg, M.; Sahlberg, C.; Engelhardt, P.; Noreen, R.; Kangasmetsa, J.; Johansson, N. G.; Oberg, B.; Vrang, L.; Zhang, H.; Unge, T.; Lovgren, S.; Fridborg, K. Urea-PETT compounds as a new class of HIV-1 reverse transcriptase inhibitors. 3. Synthesis and further structure-activity relationship studies of PETT analogues. *J. Med. Chem.* **1999**, *42*, 4150-4160.

128. Ren, J.; Milton, J.; Weaver, K. L.; Short, S. A.; Stuart, D. I.; Stammers, D. K. Structural basis for the resilience of efavirenz (DMP-266) to drug resistance mutations in HIV-1 reverse transcriptase. *Struct. Fold Des.* **2000**, *8*, 1089-1094.

129. Pauwels, R.; Andries, K.; Debyser, Z.; Van Daele, P.; Schols, D.; Stoffels, P.; De Vreese, K.; Woestenborghs, R.; Vandamme, A. M.; Janssen, C. G. Potent and highly selective human immunodeficiency virus type 1 (HIV-1) inhibition by a series of alpha-anilinophenylacetamide derivatives targeted at HIV-1 reverse transcriptase. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 1711-1715.

130. Ding, J.; Das, K.; Moereels, H.; Koymans, L.; Andries, K.; Janssen, P. A.; Hughes, S. H.; Arnold, E. Structure of HIV-1 RT/TIBO R 86183 complex reveals similarity in the binding of diverse nonnucleoside inhibitors. *Nat. Struct. Biol.* **1995**, *2*, 407-415.

131. Chan, J. H.; Hong, J. S.; Hunter, R. N. r.; Orr, G. F.; Cowan, J. R.; Sherman, D. B.; Sparks, S. M.; Reitter, B. E.; Andrews, C. W.; Hazen, R. J.; St Clair, M.; Boone, L. R.; Ferris, R. G.; Creech, K. L.; Roberts, G. B.; Short, S. A.; Weaver, K.; Ott, R. J.; Ren, J.; Hopkins, A.; Stuart, D. I.; Stammers, D. K. 2-Amino-6-arylsulfonylbenzonitriles as non-nucleoside reverse transcriptase inhibitors of HIV-1. *J. Med. Chem.* **2001**, *44*, 1866-1882.

132. Hopkins, A. L.; Ren, J.; Esnouf, R. M.; Willcox, B. E.; Jones, E. Y.; Ross, C.; Miyasaka, T.; Walker, R. T.; Tanaka, H.; Stammers, D. K.; Stuart, D. I. Complexes of HIV-1 reverse transcriptase with inhibitors of the HEPT series reveal conformational changes relevant to the design of potent non-nucleoside inhibitors. *J. Med. Chem.* **1996**, *39*, 1589-1600.

133. Ren, J.; Esnouf, R. M.; Hopkins, A. L.; Jones, E. Y.; Kirby, I.; Keeling, J.; Ross, C. K.; Larder, B. A.; Stuart, D. I.; Stammers, D. K. 3'-Azido-3'-deoxythymidine drug resistance mutations in HIV-1 reverse transcriptase can induce long range conformational changes. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9518-9523.

134. Barnard, J.; Borkow, G.; Parniak, M. A. The thiocarboxanilide nonnucleoside UC781 is a tight-binding inhibitor of HIV-1 reverse transcriptase. *Biochemistry* **1997**, *36*, 7786-7792.

135. Ren, J.; Esnouf, R. M.; Hopkins, A. L.; Warren, J.; Balzarini, J.; Stuart, D. I.; Stammers, D. K. Crystal structures of HIV-1 reverse transcriptase in complex with carboxanilide derivatives. *Biochemistry* **1998**, *37*, 14394-14403.

136. Kohlstaedt, L. A.; Wang, J.; Friedman, J. M.; Rice, P. A.; Steitz, T. A. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **1992**, *256*, 1783-1790.

137. Ragno R., Artico M., De Martino G., La Regina G., Coluccia A., Di Pasquali A., Silvetsri R.: Docking and 3-D QSAR Studies on Indolyl Aryl Sulfones. Binding Mode Exploration at the HIV-1 Reverse Transcriptase Non-Nucleoside Binding Site and Design of Highly Active N-(2-Hydroxyethyl)carboxamide and N-(2-Hydroxyethyl)carbohydrazide Derivatives. *J Med Chem* **2005**, *48*(1):213-223.

138. Pastor M., Cruciani G., Clementi S.: Smart region definition: a new way to improve the predictive ability and interpretability of three-dimensional quantitative structure-activity relationships. *J Med Chem* **1997**, *40*(10):1455-1464.

139. Wasley J.W.F.: *Molecular Interaction Fields, edited by Gabriele Cruciani*, vol 49; **2006.**

140. Young S.D., Amblard M.C., Brichter S.F., Grey V.E., Tran L.O., Lumma W.C.J., Huff J.R., Schleif W.A., Emini E.E., O'Brien J.A., Pettibone D.J.: 2-Heterocyclic indole-3-sulfones as inhibitors of HIV-1 reverse transcriptase. *Bioorg. & Med. Chem. Lett.* **1995**, *5*:491-496.

141. Silvestri R., Artico M., De Martino G., La Regina G., Loddo R., La Colla M., La Colla P.: Simple, Short Peptide Derivatives of a Sulfonylindolecarboxamide (L-737,126) Active in Vitro against HIV-1 Wild Type and Variants Carrying Non-Nucleoside Reverse Transcriptase Inhibitor Resistance Mutations. *J Med Chem* **2004**, *47*(15):3892-3896.

142. Hopkins A.L., Ren J., Milton J., Hazen R.J., Chan J.H., Stuart D.I., Stammers D.K.: Design of Non-Nucleoside Inhibitors of HIV-1 Reverse Transcriptase with Improved Drug Resistance Properties. 1. *J Med Chem* **2004**, *47*(24):5912-5922.

143. De Martino G., La Regina G., Ragno R., Coluccia A., Bergamini A., Ciaprini C., Sinistro A., Maga G., Crespan E., Artico M., Silvestri R.: Indolyl aryl sulphones as HIV-1 non-nucleoside reverse transcriptase inhibitors: synthesis, biological evaluation and binding mode studies of new derivatives at indole-2-carboxamide. *Antiviral Chemistry & Chemotherapy* **2006**, *17*(2):59-77.

144. Ren J., Esnouf R., Garman E., Somers D., Ross C., Kirby I., Keeling J., Darby G., Jones Y., Stuart D., et al.: High resolution structures of HIV-1 RT from four RT-inhibitor complexes. *Nat Struct Biol* **1995**, *2*(4):293-302.

145. Ren J., Milton J., Weaver K.L., Short S.A., Stuart D.I., Stammers D.K.: Structural basis for the resilience of efavirenz (DMP-266) to drug resistance mutations in HIV-1 reverse transcriptase. *Structure Fold Des* **2000**, *8*(10):1089-1094.

146. Ren J., Nichols C., Bird L., Chamberlain P., Weaver K., Short S., Stuart D.I., Stammers D.K.: Structural mechanisms of drug resistance for mutations at codons 181 and 188 in HIV-1 reverse transcriptase and the improved resilience of second generation non-nucleoside inhibitors. *J Mol Biol* **2001**, *312*(4):795-805.

147. Tetko I.V., Tanchuk V.Y., Kasheva T.N., Villa A.E.: Estimation of aqueous solubility of chemical compounds using E-state indices. *J Chem Inf Comput Sci* **2001**, *41*(6):1488-1493.

148. Ragno R., Coluccia A., La Regina G., De Martino G., Lavecchia A., Novellino E., Bergamini A., Ciaprini C., Sinistro A., Maga G., Crespan E., Artico M., Silvestri R. : Design, molecular modeling, synthesis, and anti-HIV-1 activity of new indolyl aryl sulfones. Novel derivatives of the indole-2-carboxamide. *J Med Chem* **2006**, *49*(11):3172-3184.

INDEX

10. Acknowledgments